Petra Carina Edel

Justus-Liebig-Universität Gießen
Institut für Didaktik der Mathematik

JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

Beweisversuche von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade

Entwicklung eines Analysemodells und empirische Ergebnisse zu den Klassenstufen 5 bis 8

Dissertation zur Erlangung
des Doktorgrades der naturwissenschaftlichen Fachbereiche

– 2024 –

1. Gutachterin / 1st reviewer	Prof. Dr. Katja Lengnink (Justus-Liebig-Universität Gießen)
2. Gutachterin / 2nd reviewer	Prof. Dr. Christine Knipping (Universität Bremen)
Eingereicht / submitted	16. Oktober 2023
Disputation / disputation	22. März 2024

Danksagung

Rückblickend betrachtet, ist diese Arbeit das Ergebnis eines langen Schaffensprozesses, in dem viele Personen involviert waren und mich unterstützt haben. Für all die kleinen und großen Beiträge zu dieser Dissertation möchte ich mich an dieser Stelle herzlich bedanken.

Mein besonderer Dank gilt meiner Doktormutter, Prof. Dr. Katja Lengnink. Danke für die intensive und stets konstruktive Unterstützung, das große Vertrauen und die schier endlose Geduld! Ohne deinen Zuspruch und den regen Austausch hätte diese Arbeit so nie beendet werden können. Ein besonderer Dank gilt ebenfalls meiner Zweitbetreuerin, Prof. Dr. Christine Knipping, für Ihr großes Interesse an meinem Thema, die richtungsweisenden Rückmeldungen und Gespräche sowie die Begutachtung der vorliegenden Arbeit. Außerdem danke ich Prof. Dr. Albrecht Beutelspacher und Prof. Dr. Nicole Graulich für ihre wohlwollende Kenntnisnahme dieser Arbeit.

Diese Dissertationsschrift präsentiert einen Ausschnitt eines Forschungsprojekts zur Mathematik-Olympiade, bei dem viele Ehrenamtliche des Landesverbands Mathematikwettbewerbe NRW e. V. mitgewirkt haben. Besonders hervorheben möchte ich Ralf Wambach und Michael Rüsing, die damaligen Vorsitzenden des Landesverbands, sowie alle beteiligten Regionalkoordinatorinnen und Regionalkoordinatoren. Ohne ihr Einverständnis und ihre Unterstützung wäre das Projekt nicht möglich gewesen. Zudem gilt mein Dank den Schülerinnen und Schülern, die an meiner Datenerhebung teilgenommen und somit die Entstehung eines sehr reichhaltigen Datensatzes ermöglicht haben.

Die Entwicklung dieser Arbeit wurde durch mein berufliches Umfeld an der Justus-Liebig-Universität Gießen begleitet und unterstützt. Einen Grundstein für die Entstehung dieser Arbeit hat Prof. Dr. Sebastian Rezat gelegt, der meine Mitarbeit im Institut für Mathematikdidaktik der Justus-Liebig-Universität Gießen in der Anfangszeit meines Studiums initiierte. Bei allen ehemaligen und aktuellen Mitarbeiterinnen und Mitarbeitern des Instituts für Didaktik der Mathematik möchte ich mich für das sehr angenehme, offene und konstruktive Umfeld bedanken, in dessen Wirkungskreis die Arbeit wachsen konnte. Insbesondere möchte ich meinen Mitdoktorandinnen und -doktoranden, Jacqueline Bonow, Eileen Baschek, Johanna Brück, Patrick Eckert, Dr. Friederike Heinz, Tabea Knobbe, Dr. Rebecca Klose, Sabine Kühn, Andreas Leinigen, Franziska Peters, Theresa Scholl, Prof. Dr. Sebastian Schorcht, Jonas Wagner, Dr. Nicole Weber-Koppitz und Moritz Zessin, sowie Prof. Dr. Christof Schreiber, Prof. Dr. Lothar Profke und Dr. Franziska Siebel für die anregenden Doktorandenkolloquien ein Dankeschön aussprechen. Zudem haben zahlreiche studentische Mitarbeiterinnen und Mitarbeiter bei der Datenerhebung, der Datenaufbereitung und der Datenanalyse mitgeholfen. Besonders zu erwähnen sind dabei Julia Ebert, Runa Förster, Isabelle Freising, Theresa Knöß, Martin Major, Paul Schäfer, Yannik Wehr, Jan Witzel und Moritz Zessin. Am Ende meines Schreibprozesses haben mich Dr. Rebecca Klose, Prof. Dr. Sebastian Schorcht und Prof. Dr. Rudolf Sträßer besonders unterstützt und die Vollendung dieser Arbeit ermöglicht.

Auch möchte ich mich universitätsübergreifend für den freundschaftlichen und zugleich hilfreichen Austausch mit Dr. Florian Füllgrabe, Dr. Kerstin Hein, Prof. Dr. Leander Kempen sowie Miriam Krieger bedanken. Darüber hinaus hat mir Prof. Dr. Regina Bruder durch die Bereitstellung von Literatur und ein produktives Expertengespräch geholfen. Prof. Dr. Benjamin Rott und seine Arbeitsgruppe haben zum Gelingen der Arbeit durch eine Besprechung meiner Untersuchungsinstrumente beigetragen. Daran anknüpfend gilt mein Dank allen Lehrerinnen und Lehrern, den Schulleitungen, den Eltern und den Schülerinnen und Schülern sowie den Verantwortlichen und Teilnehmenden des MatheClubs in Wetzlar, die es mir ermöglicht haben, meine Untersuchungsinstrumente zu pilotieren. Außerdem möchte ich Patrick Bauermann, dem Leiter des Bereichs „Bundesweite Mathematik-Wettbewerbe“ bei Bildung & Begabung, für den interessanten Austausch, die Bereitstellung von Materialien und die detailliert-konstruktiven Rückmeldungen ein herzliches Dankeschön aussprechen. Zudem danke ich Béatrice Gronau, Inge Hachtel, Petra Hesse, und Michael Rüsing, die mir als Mitglieder des Aufgabenausschusses des Mathematik-Olympiaden e. V. die Entwicklung der Wettbewerbsaufgaben erläutert haben. Von Johannes Herrmann wurde ich durch eine Statistikberatung unterstützt, wofür ich ihm dankbar bin.

Des Weiteren gebührt ein Dank der Justus-Liebig-Universität Gießen für die Bereitstellung von Räumlichkeiten, Materialien und finanzieller Mittel. Der Stiftung der Deutschen Wirtschaft (SDW) danke ich für die finanzielle und insbesondere für die wertvolle ideelle Unterstützung im lehramtsbezogenen Programm des Studienkollegs. Den Verantwortlichen von Mentoring Hessen danke ich für die gute Förderung, sowie meiner Mentorin, Prof. Dr. Adriana del Rey, und den anderen Mentees für den interessanten Austausch.

Familie und Freunde sind unbezahlbar, sie machen gute Zeiten noch besser und lassen uns Herausforderungen überwinden. Die Herausforderungen der letzten Jahre waren nur überbrückbar mit der Hilfe meines Ehemanns, meiner Eltern, meiner Schwiegereltern und meiner Schwiegeroma. Ihre Toleranz, die umfangreiche private Unterstützung sowie die schönen Zeiten der Abwechslung und Erholung waren enorm wichtig für die Fertigstellung dieser Arbeit. Meinen Kindern danke ich vor allem für all die Freude, die sie mir Tag für Tag geben, und für ihre Selbstständigkeit, die mir Freiräume zur kreativen Arbeit geschaffen hat. Meinen Freundinnen und Freunden danke ich für ihr offenes Ohr und ihr Verständnis, die sie mir geschenkt haben, und für die wertvolle gemeinsame Zeit.

Petra Carina Edel

Zusammenfassung

Aus einer ehrenamtlichen Tätigkeit im Bereich der Förderung von Teilnehmenden der Mathematik-Olympiade heraus ist ein Forschungsprojekt zur Mathematik-Olympiade in Deutschland entstanden. Dieses Projekt beschäftigt sich mit dem Führen von Beweisen, das bei der Mathematik-Olympiade von großer Bedeutung ist. Darauf fokussiert werden zuerst ausgewählte Aufgaben aus Mathematik-Olympiaden in Hinblick auf die durch sie gestellten Anforderungen untersucht. Danach werden Beweisversuche von Teilnehmenden der Mathematik-Olympiade analysiert sowie Bearbeitungen von Aufgaben zum Validieren und Bewerten fiktiver Beweisversuche ergänzend ausgewertet. Die Ergebnisse der Anforderungsanalyse und der Analyse vorhandener Beweiskompetenzen lassen sich mit Blick auf Ansatzpunkte für eine Förderung im Führen von Beweisen von Teilnehmenden der Mathematik-Olympiade interpretieren.

Die vorliegende Arbeit beschreibt einen Ausschnitt dieses Forschungsprojekts. Sie beschäftigt sich mit Beweisversuchen von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den Klassenstufen 5 bis 8. Als Beweisversuche werden dabei schriftlich vorliegende Bearbeitungen von Beweisaufgaben bezeichnet. In dieser Arbeit werden mehr als 1000 Beweisversuche zu einer Allaussage aus dem Themenbereich der Teilbarkeit ganzer Zahlen vergleichend untersucht, die im Rahmen der Wettbewerbsklausur der Regionalrunde der Mathematik-Olympiade 2016/17 in Nordrhein-Westfalen erhoben wurden. Dazu wird das Verfahren der qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) angewendet. Es werden ein Kategoriensystem sowie zwei Klassifikationen entwickelt, die aus verschiedenen Perspektiven systematische Einblicke in mögliche und vorhandene Ausprägungen von Beweisversuchen geben können. Die Forschungsergebnisse legen dar, inwieweit es den Teilnehmenden in den einzelnen Klassenstufen 5 bis 8 gelungen ist, eine bestimmte Allaussage aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen (1). Außerdem beschreiben die Ergebnisse, welche Charakteristika deren Beweisversuche u. a. bezüglich der Beispielnutzung und der Darstellungsmittel auszeichnen (2) und inwieweit Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind (3). In Bezug auf alle drei Punkte wird aufgezeigt, welche Gemeinsamkeiten und Unterschiede mit Blick auf die verschiedenen Klassenstufen bestehen.

Die vorgelegte Arbeit bearbeitet auf diese Weise zwei Forschungsgegenstände: Zum einen können vor dem Hintergrund anderer mathematikdidaktischer Studien zum Führen von Beweisen das Analysemodell und die darauf aufbauenden Klassifikationen als die wesentlichen Beiträge der vorliegenden Dissertation zur mathematikdidaktischen Theoriebildung angesehen werden. Zum anderen und mit Blick auf die Mathematik-Olympiade ergänzen die empirischen Ergebnisse dieser Arbeit die vorhandenen Erkenntnisse zu Beweisversuchen von Schülerinnen und Schülern. Zusätzlich leisten die entwickelten theoretischen Elemente und die empirischen Ergebnisse einen wichtigen Beitrag, um zukünftig Angebote zur Förderung im Führen von Beweisen für Teilnehmende der Mathematik-Olympiade zu konzipieren.

Abstract

A research project on the Mathematical Olympiad in Germany has emerged from a voluntary activity in the field of supporting Mathematical Olympiad participants. This project is about constructing proofs, which is of great importance in the Mathematical Olympiad. In a first step, selected tasks from Mathematical Olympiads are examined with regard to the demands they set in the area of constructing proofs. In a second step, proof constructions of participants and their evaluations of fictitious proof constructions are analysed. In a third step, the results of the task analysis and the investigation of participants' competencies can contribute to the design of a support in constructing proofs.

This thesis describes a part of the research project. It deals with proof constructions of Mathematical Olympiad participants from grades 5 to 8. Thereby only written proof attempts are considered as proof constructions. In this paper, more than 1000 proof constructions on an universal statement on the divisibility of integers are comparatively examined. These proof constructions were collected in the context of the competition exam of the regional round of the Mathematical Olympiad 2016/17 in North Rhine-Westphalia. For the thesis, the method of qualitative content analysis according to Schreier (2012, 2014) is applied. A category system as well as two classifications are developed, which can provide systematic insight into possible and existing proof constructions from different perspectives. The research results show to what extent the participants in the individual grades 5 to 8 succeeded in proving the universal statement from the area of divisibility of integers (1). Furthermore, the results describe, which properties characterize their proof constructions (2) and to what extent starting points for their further development into a proof can be identified (3). For all three issues, similarities and differences between the different grade levels exist and are shown.

The thesis as outlined above offers two research results: On the one hand, the analytical model and the classifications contribute to theory building in the field of mathematics didactics, especially to studies on proof. On the other hand, and especially with regard to the Mathematical Olympiad, the empirical results complement the existing data and findings on students' proof constructions. The developed theoretical elements and the empirical results could be of help when it comes to support Mathematical Olympiad participants in constructing proofs.

Inhaltsverzeichnis

1 Einleitung ... 1
1.1 Motivation, Fokussierungen und Fragestellungen ... 2
1.2 Aufbau der vorliegenden Arbeit ... 6
1.3 Für die Leserinnen und Leser ... 7

2 Die Mathematik-Olympiade ... 9
2.1 Entwicklung der Mathematik-Olympiade ... 9
2.2 Heutige Konzeption und Durchführungspraxis ... 11
2.2.1 Bundesweite Konzeption ... 11
2.2.2 Wettbewerbspraxis in Nordrhein-Westfalen ... 13
2.2.3 Die Verbindung von Wettbewerb und Förderung ... 17
2.3 Aufgaben der Mathematik-Olympiade ... 18
2.3.1 Entwicklung der Aufgaben ... 18
2.3.2 Anforderungen an Aufgabenbearbeitungen ... 20
2.4 Zusammenfassung des zweiten Kapitels ... 23

3 Annäherung an den Beweisbegriff ... 25
3.1 Der Beweisbegriff innerhalb der Mathematik ... 25
3.1.1 Mathematische Aussagen ... 26
3.1.2 Beschreibungsansätze zum mathematischen Beweisbegriff ... 27
3.2 Problematisierung des Beweisbegriffs innerhalb der Mathematik für die Arbeit mit Schülerinnen und Schülern ... 34
3.3 Ein Modell zur Analyse von Beweisversuchen zu Allaussagen ... 36
3.3.1 Skizze des Modells ... 38
3.3.2 Begründender, prüfender oder sonstiger Ansatz ... 39
3.3.3 Begründender Ansatz: Beweisideen ... 53
3.3.4 Begründender Ansatz: Korrektheit und Vollständigkeit ... 68
3.3.5 Begründender Ansatz: Beispiele ... 83
3.3.6 Begründender Ansatz: Schlussweisen ... 94
3.3.7 Prüfender Ansatz: Beispiele und Angabe eines Wahrheitswerts ... 100
3.3.8 Prüfender Ansatz: Schlussweise ... 109
3.3.9 Weder begründender noch prüfender Ansatz: Angabe eines Wahrheitswerts und Art der Bearbeitung ... 115
3.3.10 Nicht eindeutig kategorisierbarer Ansatz ... 122
3.3.11 Darstellungsmittel und Darstellungsform ... 126
3.4 Der Beweisbegriff in der vorliegenden Arbeit ... 136
3.5 Zusammenfassung des dritten Kapitels ... 138

4 Beweiskompetenz: Begriffsannäherung und curriculare Vorgaben ... 141
4.1 Begriff „Beweiskompetenz“ ... 141
4.2 Curriculare Vorgaben zu Beweiskompetenzen ... 145
4.2.1 Vorgaben in den Bildungsstandards ... 145

4.2.2 Vorgaben in Nordrhein-Westfalen 149
4.2.3 Erwartungen an Teilnehmende der Mathematik-Olympiade 151
4.3 Zusammenfassung des vierten Kapitels 152

5 Ergebnisse empirischer Studien zu Beweiskompetenzen von Schülerinnen und Schülern 153
5.1 Gelingen von Beweisversuchen.... 154
5.1.1 Bearbeitungen von Aufgaben mit Beweisnotwendigkeit 154
5.1.2 Zusammenhänge zwischen Bearbeitungen von Aufgaben mit Beweisnotwendigkeit und weiteren Faktoren 163
5.2 Verwendung von Beispielen.... 168
5.2.1 Empirisch-induktive Ansätze 168
5.2.2 Generische Beispiele und beispielgebundene Beweise 171
5.2.3 Beispiele in Aufgabenbearbeitungen mit begründendem Ansatz 171
5.2.4 Zusammenfassung....173
5.3 Verwendung von Darstellungsmitteln.... 173
5.3.1 Symbolische und verbale Darstellungsmittel....173
5.3.2 Ikonische Darstellungsmittel 174
5.3.3 Zusammenfassung....175
5.4 Diskussion 175
5.5 Zusammenfassung des fünften Kapitels.... 177

6 Forschungsdesiderat, Fragestellungen und Design 179
6.1 Forschungsdesiderat und Fragestellungen.... 179
6.2 Design der Studie 182
6.2.1 Anlage und Untersuchungsinstrumente der Studie....183
6.2.2 Aufgabenkonstruktion....185
6.3 Zusammenfassung des sechsten Kapitels 199

7 Datenerhebung und Auswertung.... 201
7.1 Datenerhebung 201
7.2 Auswertungsmethoden.... 201
7.2.1 Qualitative Inhaltsanalyse 202
7.2.2 Statistische Grundlagen....210
7.3 Anwendung der Auswertungsmethoden 218
7.3.1 Überblick über die gewählte Vorgehensweise 218
7.3.2 Konkretisierung der Forschungsfragen....219
7.3.3 Auswahl des Materials....219
7.3.4 Unterteilung des Materials in Einheiten....221
7.3.5 Erstellung des Kategoriensystems 221
7.3.6 Pilotierung des Kategoriensystems....223
7.3.7 Hauptcodierung....224
7.3.8 Weitere Auswertung und Ergebnispräsentation 228
7.4 Zusammenfassung des siebten Kapitels.... 230

8 Ergebnisse 231
8.1 Begründender, prüfender oder sonstiger Ansatz 232
8.2 Beweisversuche mit begründendem Ansatz 233
8.2.1 Korrektheit, Vollständigkeit und Schlussweise 234
8.2.2 Beweisideen 245
8.2.3 Beispiele 255
8.2.4 Darstellungsmittel und Darstellungsformen 273
8.2.5 Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage 278
8.2.6 Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf deren Weiterentwicklung zu einem Beweis 283
8.3 Beweisversuche mit prüfendem Ansatz 341
8.3.1 Passung eines Beispiels zu der zu zeigenden Aussage 341
8.3.2 Eigenschaften der Beweisversuche mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis 342
8.3.3 Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit 358
8.3.4 Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf deren Weiterentwicklung zu einem Beweis 361
8.4 Beweisversuche mit weder begründendem noch prüfendem Ansatz 365
8.4.1 Angabe der Allgemeingültigkeit 365
8.4.2 Arten der Bearbeitung und Darstellungsmittel, Darstellungsformen 366
8.4.3 Klassifikation der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit 367
8.4.4 Klassifikation der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf deren Weiterentwicklung zu einem Beweis 368
8.5 Eigenschaften aller Beweisversuche 370
8.5.1 Fokus: Nachweis der Gültigkeit 371
8.5.2 Fokus: Charakteristika der Beweisversuche 381
8.5.3 Fokus: Anknüpfungspunkte für die Weiterentwicklung 386
8.6 Zusammenfassung des achten Kapitels 405

9 Diskussion 407
9.1 Diskussion des Designs und der Durchführung der Studie 407
9.1.1 Die Untersuchungspopulation: Teilnehmerinnen und Teilnehmer der Mathematik-Olympiade in den Klassenstufen 5 bis 8 407
9.1.2 Ort, Umfang, Ablauf und Zeitpunkt der Datenerhebung 408
9.1.3 Die Fokussierung auf eine Beweisaufgabe aus dem Themenbereich der Teilbarkeit ganzer Zahlen und die Konzeption dieser Aufgabe 410
9.1.4 Die Fokussierung auf Beweisversuche 413
9.2 Diskussion der Vorgehensweise bei der Datenauswertung 414
9.2.1 Diskussion des Kategoriensystems 415
9.2.2 Diskussion der Klassifikationen 417
9.3 Diskussion der erhaltenen Ergebnisse 418
9.3.1 Gelingen von Beweisversuchen 419

9.3.2 Verwendung von Beispielen 423
9.3.3 Darstellungsmittel und Darstellungsformen 426
9.3.4 Ansatzpunkte für Förderung 429
9.4 Zusammenfassung des neunten Kapitels 433

10 Ausblick 435
10.1 Anregungen für zukünftige Forschungsarbeiten 435
10.1.1 Anforderungen bei der Mathematik-Olympiade 435
10.1.2 Beweisversuche unterschiedlicher Teilgruppen der Teilnehmenden der Mathematik-Olympiade 436
10.1.3 Beweisversuche unterschiedlicher Zielgruppen 437
10.1.4 Beweisversuche zu unterschiedlichen Aufgabenstellungen 438
10.1.5 Beweisvorstellungen und Beweisprozesse 439
10.2 Überblick über ein Forschungsprojekt zum Führen von Beweisen bei der Mathematik-Olympiade 440
10.2.1 Anforderungen bei der Mathematik-Olympiade 440
10.2.2 Vorhandene Kompetenzen von Teilnehmenden der Mathematik-Olympiade 441
10.2.3 Ansatzpunkte für eine Förderung von Teilnehmenden der Mathematik-Olympiade 446
10.3 Anforderungen bei der Mathematik-Olympiade 447
10.3.1 Anlage und Datengrundlage der Anforderungsanalyse 447
10.3.2 Ein Modell zur Analyse objektiver Anforderungsstrukturen von Aufgaben mit besonderem Fokus auf dem Führen von Beweisen 452
10.3.3 Ergebnisse der Anforderungsanalyse 455
10.4 Beweisversuche unterschiedlicher Teilgruppen der Teilnehmenden der Mathematik-Olympiade 457
10.4.1 Beweisversuche von Teilnehmenden in der Klassenstufe 9 458
10.4.2 Beweisversuche, Wettbewerbserfolg und Förderung 472
10.5 Zusammenfassung des zehnten Kapitels 486

Schlusswort 487

Literaturverzeichnis 489

Abbildungsverzeichnis 507

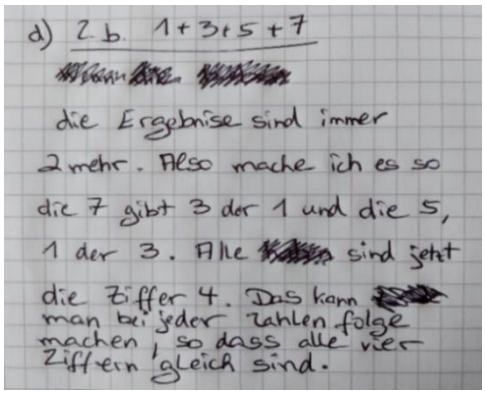
Tabellenverzeichnis 521

1 Einleitung

Aus einer ehrenamtlichen Tätigkeit heraus ist ein Forschungsprojekt zum Führen von Beweisen bei der Mathematik-Olympiade entstanden. Die vorliegende Arbeit präsentiert einen Ausschnitt dieses Forschungsprojekts. Sie beschäftigt sich mit Beweisversuchen von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den Klassenstufen 5 bis 8. Als Beweisversuche werden dabei in schriftlicher Form vorliegende Bearbeitungen von Beweisaufgaben bezeichnet. Beispielsweise zeigt die Abbildung 1 zwei Beweisversuche[1] von Fünftklässlern, die im Rahmen einer Regionalrunde der Mathematik-Olympiade in Deutschland erhoben wurden.

Aufgabe: Defne berechnet die Summe von vier aufeinanderfolgenden ungeraden Zahlen (z. B. $5 + 7 + 9 + 11 = 32$). Sie behauptet: „Die Summe von **vier aufeinanderfolgenden ungeraden** Zahlen ist immer durch 4 teilbar." Hilf Defne zu begründen, warum ihre Behauptung richtig ist.

Bearbeitung 1a: Kl. 5, männlich (Code: 0521720)



Bearbeitung 1b: Kl. 5, männlich (Code: 0510208)

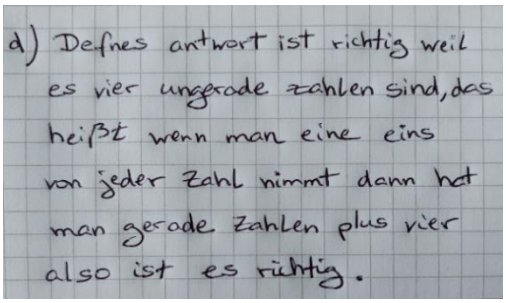


Abbildung 1: Zwei Beweisversuche von Fünftklässlern aus dem Datensatz dieser Arbeit

Diese und mehr als 1000 weitere Beweisversuche von Teilnehmenden der Mathematik-Olympiade werden in der vorliegenden Arbeit vergleichend untersucht. Dazu wird das Verfahren der qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) angewendet. Es werden ein Kategoriensystem sowie zwei Klassifikationen entwickelt, die aus verschiedenen Perspektiven systematische Einblicke in mögliche und vorhandene Ausprägungen von Beweisversuchen geben können. Die Forschungsergebnisse legen dar, inwieweit es Teilnehmenden der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 gelungen ist, eine bestimmte Allaussage aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen (1), welche Charakteristika deren Beweisversuche u. a. bezüglich der Beispielnutzung und der Darstellungsmittel auszeichnen (2) und inwieweit Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind (3). In Bezug auf alle drei Punkte wird aufgezeigt, welche Gemeinsamkeiten und Unterschiede mit Blick auf die verschiedenen Klassenstufen bestehen.

Im Folgenden werden die Motivation, Fokussierungen und Fragestellungen sowie der Aufbau dieser Arbeit ausführlich beschrieben.

[1] Alle Beweisversuche von Schülerinnen und Schülern, die in dieser Arbeit abgebildet sind, wurden aus Gründen der Anonymisierung handschriftlich transkribiert. Die Ausführungen der Lernenden wurden wortgetreu inklusive sämtlicher Durchstreichungen und visueller Hervorhebungen übernommen. Die Rechtschreibung und Grammatik sowie die Anordnung der Ausführungen wurden ebenfalls beibehalten.

1.1 Motivation, Fokussierungen und Fragestellungen

Schon seit meiner Schulzeit engagierte ich mich in der Förderung mathematisch interessierter, leistungsstarker Kinder und Jugendlicher. Unter anderem unterrichtete ich als Dozentin des Landesverbands Mathematikwettbewerbe NRW e. V. regelmäßig Lernende der Sekundarstufen im Rahmen von mathematischen Akademien, Mathematikwochenenden und weiteren mathematischen Förderangeboten. Bei der Vorbereitung auf diese Dozententätigkeit stellte sich mir immer wieder die Frage, welche Kompetenzen ich in den Unterrichtseinheiten fördern möchte. Die **Förderangebote** richteten sich vor allem an **Teilnehmerinnen und Teilnehmer der Mathematik-Olympiade** in Nordrhein-Westfalen (NRW) ab der Klassenstufe 5. Die Mathematik-Olympiade ist ein Stufenwettbewerb für mathematisch interessierte Schülerinnen und Schüler, in dem sie eigenständig auf ihre Klassenstufe abgestimmte Aufgaben bearbeiten. Dabei sollen sie vorgegebene Aussagen und selbstständig erarbeitete Lösungen beweisen. Die gestellten Anforderungen in Bezug auf das Führen von Beweisen gehen teils deutlich über die Vorgaben in den Bildungsstandards und den Kernlehrplänen für NRW hinaus. So kam mir die Idee, in meinen Unterrichtseinheiten im Rahmen mathematischer Förderangebote einen Fokus auf das Führen von Beweisen zu legen.

„Beweise sind das Herz der Mathematik“, so Ziegler (2008, S. 409). Das Führen von Beweisen, insbesondere das Beweisen von Allaussagen mit unendlichem Gültigkeitsbereich, ist charakteristisch für diese Wissenschaft und unterscheidet sie von vielen anderen Wissenschaften (u. a. Davis & Hersh, 1986; Heintz, 2000). Beweise sind eine Voraussetzung für die Gültigkeit mathematischer Aussagen. Sie erklären, warum eine Aussage wahr ist, und können helfen, Beziehungen zwischen bekannten und neuen Definitionen, Begriffen sowie Sätzen herzustellen (de Villiers, 1990). Daran anknüpfend sind aus mathematikdidaktischer Sicht Beweise und die Tätigkeit des Beweisens auch beim Mathematiklernen wichtig. Sie können das Lernen durch Verstehen fördern, das Ordnen, Vernetzen und Vertiefen mathematischen Wissens unterstützen, Problemlöse- und Argumentationskompetenzen fördern und zu einem authentischen Bild der Wissenschaft Mathematik beitragen (u. a. Hanna et al., 2012; Malle, 2002; Reiss, 2009).

In den Bildungsstandards wurde das Beweisen unter dem Oberbegriff „Argumentieren“ (KMK, 2004, S. 8) bzw. „Mathematisch argumentieren“ (KMK, 2003, S. 8; KMK, 2012, S. 15) als allgemeine mathematische Kompetenz aufgenommen. Schülerinnen und Schüler sollen am Ende der Sekundarstufen mehrschrittige Argumentgefüge entwickeln können (KMK, 2003, 2012). Allerdings haben Studien zufolge viele Schülerinnen und Schüler Schwierigkeiten, Beweise zu führen. In Studien mit deutschen Gymnasiastinnen und Gymnasiasten stehen starken Leistungen im Basiswissen schwache Leistungen bei mehrschrittigen Beweisaufgaben gegenüber (u. a. Reiss et al., 2002). Im britischen Forschungsprojekt von Healy und Hoyles (2000) gelingt unter fünf Prozent der mathematisch leistungsstarken Teilnehmenden am Ende der zehnten Klassenstufe der Beweis einer nicht aus dem Schulunterricht bekannten Allaussage.

Auch viele Studienanfängerinnen und -anfänger haben Probleme, mathematische Aussagen zu beweisen (z. B. Hemmi, 2006; Herrmann, 2012; Weber, 2001). Dabei sind die Relevanz von Beweisaktivitäten und die dabei gestellten Anforderungen im Mathematikstudium höher als im Schulunterricht Mathematik, was zu den ausgeprägten Studienabbruchquoten beiträgt (u. a.

Fischer et al., 2009; Ostsieker & Biehler, 2012; Selden, 2012). Es gibt vielerorts Maßnahmen wie Vorkurse, Brückenvorlesungen und Lernzentren, um den Studienbeginn in Mathematik zu erleichtern (z. B. Frischemeier et al., 2016; Grieser, 2016; Kempen, 2019; Moore, 1994). In Ergänzung zu diesen Maßnahmen erscheint es als sinnvoll, Lernende, die potenziell ein Studium mit Schwerpunkt Mathematik anstreben, frühzeitig und kontinuierlich zu fördern.

Bei den Teilnehmenden der Mathematik-Olympiade handelt es sich um Schülerinnen und Schüler, denen man mit Blick auf den wissenschaftlichen Nachwuchs viel Beachtung schenken sollte. Schließlich ist es ein Ziel des Wettbewerbs, leistungsstarke und potenziell besonders leistungsfähige Schülerinnen und Schüler zu finden und zu fördern, u. a. um akademischen Nachwuchs zu gewinnen (u. a. Bauermann, 2018; Huawei, 2002; Kanel-Beloy & Yavich, 2014; KMK, 2015). Eine Förderung von Teilnehmenden der Mathematik-Olympiade mit einem besonderen Fokus auf mathematisches Beweisen erscheint für die Schülerinnen und Schüler demnach auch aus akademischer Perspektive als lohnend.

Bei der Konzeption einer solchen Förderung stellt sich die Frage, was Teilnehmende der Mathematik-Olympiade in Bezug auf mathematisches Beweisen können sollen, d. h. was diesbezüglich mit Blick auf die verschiedenen Klassenstufen sinnvolle und realistische Ziele sind. Zudem ist bisher weitgehend ungeklärt, inwieweit es Teilnehmenden der Mathematik-Olympiade gelingt, mathematische Aussagen zu beweisen und welche Anknüpfungspunkte ihre Beweisversuche für eine weitere Förderung bieten.

Das Führen von Beweisen ist ein seit vielen Jahren und häufig untersuchtes Forschungsfeld (z. B. Balacheff, 1988; Healy & Hoyles, 2000; Kirsten, 2021; Lin, 2005; Recio & Godino, 2001; Reiss et al., 2006; Selden & Selden, 2003; Ufer, 2008; Walsch, 1975). Studien zu Anforderungen zum Beweisen bei der Mathematik-Olympiade bzw. zu Beweiskompetenzen von Teilnehmenden der Mathematik-Olympiade gibt es bisher jedoch kaum.[2] Daher habe ich im Rahmen der wissenschaftlichen Hausarbeit im Lehramtsstudium eine Untersuchung mit erfolgreichen Teilnehmenden der Mathematik-Olympiade ab der Klassenstufe 5 durchgeführt (Tebaartz, 2013; Tebaartz, 2015; Tebaartz & Lengnink, 2015). Die Ergebnisse geben erste Einblicke in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit[3] dieser Schülerinnen und Schüler und motivierten mich dazu, die Anforderungen zum Beweisen bei der Mathematik-Olympiade und Beweiskompetenzen von Teilnehmenden der Mathematik-Olympiade zu beforschen.

Also begann ich mit einem **Forschungsprojekt**, um Informationen zu erhalten, welche Ziele im Rahmen einer Förderung von Teilnehmenden der Mathematik-Olympiade in Bezug auf das

[2] Der Großteil der Literatur zur Mathematik-Olympiade ist konzeptioneller oder beschreibender Art (z. B. Andreescu & Enescu, 2011; Bin & Yee, 2009; Cáceres-Duque et al., 2017; Engel, 1987, 1998; Engel et al., 2009; Soifer, 2009; Taylor, 2012; WFNMC, 2017). Zur Mathematik-Olympiade sind der Autorin dieser Arbeit ausschließlich forschungsbasierte Arbeiten aus Lettland bekannt. Veilande et al. (2018) analysieren schriftliche Bearbeitungen von Aufgaben der Open Mathematical Olympiads of Latvia von Teilnehmenden in der Klassenstufe 9. Sie kommen zu dem grundlegenden Ergebnis, dass den meisten Teilnehmenden keine vollständigen Lösungen gelingen. Zudem stellen die Autorinnen fest, dass bei vielen der wiederholt teilnehmenden Lernenden die durchschnittliche Lösungsqualität in der Klassenstufe 9 niedriger ist als in den Klassenstufen 6, 8 und führen dies auf Schwierigkeiten im Umgang mit Algebra zurück.

[3] Als Aufgaben mit Beweisnotwendigkeit gelten im Rahmen dieser Arbeit, kurz gesagt, Aufgaben, deren korrekte und vollständige Bearbeitung den Beweis einer mathematischen Aussage erfordert. Dabei kann die zu zeigende Aussage in der Aufgabenstellung vorgegeben sein, muss es aber nicht (Abschn. 10.3.2).

Führen von Beweisen sinnvoll und realistisch sind sowie welche Beweiskompetenzen Teilnehmende der Mathematik-Olympiade in Deutschland mitbringen (Abschn. 10.2). In diesem Forschungsprojekt werden zuerst Anforderungen, die ab der Klassenstufe 5 der Mathematik-Olympiade gestellt werden, herausgearbeitet (Teil 1). Danach werden Beweisversuche von Teilnehmenden der Mathematik-Olympiade untersucht sowie deren Bearbeitungen von Aufgaben zum Validieren und Bewerten fiktiver Beweisversuche ergänzend ausgewertet (Teil 2). Die Ergebnisse der Anforderungsanalyse und der Analyse vorhandener Beweiskompetenzen lassen sich mit Blick auf Ansatzpunkte für eine Förderung im Führen von Beweisen von Teilnehmenden der Mathematik-Olympiade interpretieren (Teil 3). Fachinhaltlich konzentriert sich dieses Forschungsprojekt auf den elementarmathematisch reichhaltigen Themenbereich der Teilbarkeit ganzer Zahlen[4]. Aufgaben aus diesem Bereich werden vom Aufgabenausschuss der Mathematik-Olympiade regelmäßig in allen Klassenstufen gestellt, der gewählte Themenbereich passt somit sehr gut zum Wettbewerb. Zudem erscheint der Bereich der Teilbarkeit ganzer Zahlen für den Vergleich von Beweisversuchen aus verschiedenen Klassenstufen sehr geeignet, da er mathematisch voraussetzungsarm ist. Wesentliche inhaltliche Grundlagen, wie das dezimale Stellenwertsystem sowie Teiler und Vielfache von Zahlen, werden im schulischen Mathematikunterricht früh behandelt und danach kaum vertieft, so dass Wettbewerbsteilnehmende ab der Klassenstufe 5 diesbezüglich ähnliche schulische Vorkenntnisse mitbringen sollten (KMK, 2003, 2004a, 2012). Im Bereich der Teilbarkeit ganzer Zahlen ist es möglich, bereits auf sehr elementarem Niveau und ohne weitreichende fachinhaltliche Vorkenntnisse Beweise zu führen und vielfältige, unterschiedlich anspruchsvolle Aufgaben zu entwickeln, sowohl für eine Erhebung vorhandener Beweiskompetenzen als auch, perspektivisch, im Rahmen einer möglichen Förderung.

In der vorliegenden Arbeit beschreibe ich einen Ausschnitt dieses Forschungsprojekts zur Mathematik-Olympiade. Die Arbeit beschäftigt sich mit Beweisversuchen zu einer Allaussage mit unendlichem Gültigkeitsbereich aus dem Themenbereich der Teilbarkeit ganzer Zahlen, die von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den Klassenstufen 5 bis 8 verfasst wurden.

Ausgehend vom Anspruch, Anknüpfungspunkte für eine Förderung zu beschreiben, wurde der Fokus auf die Analyse von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade gelegt. Darauf aufbauend wurden Beweisversuche zu einer Allaussage mit unendlichem Gültigkeitsbereich (Aufgabenteil d)) fokussiert, weil Allaussagen[5] mit unendlichem Gültigkeitsbereich für die Wissenschaft Mathematik besonders charakteristisch sind. Im Falle der Analyse von Beweisversuchen zu einer Allaussage mit unendlichem Gültigkeitsbereich ist es zudem möglich, an viele vorhandene mathematikdidaktische Theorien und Studien anzuknüpfen und

[4] Unter dem Bereich der Teilbarkeit ganzer Zahlen werden in dieser Arbeit alle Begriffe und mathematischen Aussagen zusammengefasst, die sich auf ganze Zahlen beziehen und die Teilbarkeit von Zahlen oder Termen, Teilbarkeitsregeln, Teiler-, Vielfachen- oder Potenzmengen, Querprodukte, Potenzzahlen, Primzahlen, Primfaktorzerlegung, teilerfremde Zahlen, das kleinste gemeinsame Vielfache oder den größten gemeinsamen Teiler, diophantische Gleichungen, Positionssysteme oder modulares Rechnen betreffen.

[5] Als Allaussagen werden an dieser Stelle Aussagen über alle Elemente einer gewissen Menge verstanden. Im Verlauf dieser Arbeit wird auf die Analyse von Beweisversuchen zu einer Implikation, der elementarsten Form einer Allaussage, fokussiert (Abschn. 3.1.1, 6.1).

die Ergebnisse dieser Arbeit mit denen anderer Studien zu vergleichen. In vorhandenen empirischen Studien, in denen Beweisversuche untersucht werden, stehen in der Regel frühestens Schülerinnen und Schüler ab der Klassenstufe 7 im Fokus, sodass es für eine Einordnung der eigenen Ergebnisse sinnvoll ist, Teilnehmende der Mathematik-Olympiade in der Klassenstufe 7 oder in höheren Klassenstufen zu berücksichtigen. Zugleich erscheint der Einbezug der Klassenstufen 5 und 6 besonders sinnvoll, weil in der Mathematik-Olympiade ab der Klassenstufe 5 Beweise zu Allaussagen gefordert werden und Lernende in diesen Klassenstufen in der Regel noch nicht mit Algebra vertraut sind. Ergebnisse einer Analyse von Beweisversuchen von Teilnehmenden ab der Klassenstufe 5 wären zusätzlich zur mathematikdidaktischen Theoriebildung auch mit auf die Förderung der Teilkompetenz des Führens von Beweisen interessant. Vielleicht wäre es möglich und sinnvoll, bereits in den Klassenstufen 5 und 6 verstärkt in mathematisches Beweisen einzuführen, um präalgebraische, inhaltlich-anschauliche Denkweisen zu bewahren und elementare Herangehensweisen zu fördern. Studien mit Schülerinnen und Schülern höherer Klassenstufen deuten darauf hin, dass einigen die Verwendung symbolischer Darstellungsmittel bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit Schwierigkeiten bereitet und ihnen dabei Fehler passieren (u. a. Healy & Hoyles, 2000; Küchemann, 2008; Veilande et al., 2018). Vor dem Hintergrund dieser Überlegungen wurden die Klassenstufen 5 bis 8 ausgewählt, sodass es sowohl möglich ist, Einblicke in Beweisversuche von Teilnehmenden der Mathematik-Olympiade zu Beginn der Sekundarstufe I zu erhalten, als auch Beweisversuche von Teilnehmenden aus unterschiedlichen Klassenstufen und mit unterschiedlichen Darstellungsmöglichkeiten zu beschreiben und die Ergebnisse dieser Arbeit mit denen anderer Studien zu vergleichen.

Ausgehend von der Forschungslage wird in dieser Arbeit deduktiv-induktiv ein Modell entwickelt, um Beweisversuche von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 präzise, trennscharf und systematisch differenziert analysieren zu können (Kap. 3). Dieses Analysemodell stellt ein wesentliches Entwicklungsprodukt dieser Arbeit dar. Es kann auch über die vorliegende Arbeit hinaus als Grundlage für Analysen von Beweisversuchen zu Allaussagen aus dem Bereich der Teilbarkeit ganzer Zahlen und angrenzender Themenbereiche verwendet werden. Die vorliegende Arbeit leistet folglich einen Beitrag zur mathematikdidaktischen Theoriebildung.

Zusätzlich beschreibt sie empirische Ergebnisse zu Beweisversuchen aus den Klassenstufen 5 bis 8. Die empirischen Ergebnisse ergänzen den bisherigen Forschungsstand und lassen sich mit Blick auf Ansatzpunkte für eine Förderung im Führen von Beweisen interpretieren. Konkret wird im Rahmen dieser Arbeit mithilfe des Analysemodells untersucht, inwieweit Teilnehmenden der Mathematik-Olympiade aus den Klassenstufen 5 bis 8 der Beweis einer mathematischen Allaussage gelungen ist (1), welche Charakteristika deren Beweisversuche u. a. bezüglich der Beispielnutzung und der Darstellungsmittel auszeichnen (2) und inwieweit Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind (3). Dabei wird in Bezug auf alle drei Punkte aufgezeigt, welche Gemeinsamkeiten und Unterschiede mit Blick auf die verschiedenen Klassenstufen bestehen. Die Forschungsfragen lauten dementsprechend:

1. Inwieweit gelingt es Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 eine mathematische Allaussage mit unendlichem Gültigkeitsbereich aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen, welche Charakteristika zeichnen deren Beweisversuche aus und inwieweit sind Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen?
2. Welche Gemeinsamkeiten und Unterschiede bestehen diesbezüglich zwischen Beweisversuchen von Teilnehmenden der Mathematik-Olympiade mit Blick auf verschiedene Klassenstufen?

1.2 Aufbau der vorliegenden Arbeit

Im folgenden Kapitel 2 wird zunächst die Mathematik-Olympiade vorgestellt. Es werden die Konzeption und Durchführungspraxis dieses Wettbewerbs beschrieben. Zudem werden Einblicke in reguläre Wettbewerbsaufgaben und damit verbundene Anforderungen gegeben. Die Konzeption und Durchführungspraxis des Wettbewerbs sowie die regulären Wettbewerbsanforderungen werden beim Design der vorliegenden Studie und der Auswertung und Interpretation ihrer Ergebnisse berücksichtigt.

Kapitel 3 widmet sich dem Beweisbegriff in der Mathematik und seiner Diskussion in der Mathematikdidaktik. Es wird hier als wesentliches Ergebnis ein Modell zur Analyse von Beweisversuchen zu Allaussagen entwickelt, das an zahlreiche mathematikdidaktische Theorien und Forschungsergebnisse anknüpft und zugleich Besonderheiten von Beweisversuchen junger Lernender berücksichtigt. Das Analysemodell ermöglicht eine differenzierte und kompetenzorientierte Deskription einer großen Vielfalt an Beweisversuchen zu Allaussagen aus dem Themenbereich der Teilbarkeit ganzer Zahlen.

In Kapitel 4 wird der im Rahmen dieser Arbeit verwendete Begriff der Beweiskompetenz eingegrenzt und das Führen von Beweisen als eine Teilkompetenz der Beweiskompetenz konzeptualisiert. Danach werden bezüglich des Führens von Beweisen relevante curriculare Vorgaben vorgestellt und Erwartungen an Beweisversuche von Teilnehmenden der Mathematik-Olympiade formuliert. Diese werden in Kapitel 5 durch bereits vorliegende Ergebnisse empirischer Studien zu Beweiskompetenzen von Schülerinnen und Schülern ergänzt. Es werden Forschungsergebnisse präsentiert, die für das Design der Analyse von Beweisversuchen im Rahmen dieser Arbeit oder für die Diskussion deren Ergebnisse von Bedeutung sind.

Anschließend wird in Kapitel 6 das Design der vorliegenden Studie beschrieben und begründet. Dabei liegt ein besonderer Fokus auf den Aufgaben, die zur Erhebung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade verwendet wurden. Anknüpfend an diese Überlegungen zur Datenerhebung werden in Kapitel 7 die für die Auswertung zur Verfügung stehenden Daten präsentiert. Zudem werden die methodischen Grundlagen, die zur Auswertung dieser Daten verwendet werden, erläutert. Zu den methodischen Grundlagen gehören insbesondere das Verfahren der qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) und statistische Verfahren zur Analyse von Unterschieds- und Zusammenhangshypothesen (Benesch, 2013; Bortz & Schuster, 2010; Busing et al., 2015).

Die Ergebnisse der Analyse von Beweisversuchen werden in Kapitel 8 präsentiert. Sie zeigen entsprechend den Forschungsfragen ein differenziertes Bild der Beweisversuchen von Teilnehmenden der Mathematik-Olympiade, das vor allem die Korrektheit und Vollständigkeit der Beweisversuche, Schlussweisen, Beweisideen, Beispiele und Darstellungsformen berücksichtigt. Bei der Präsentation der Ergebnisse werden häufige und seltene Ausprägungen von Beweisversuchen sowie Gemeinsamkeiten und Unterschiede zwischen Beweisversuchen aus den einzelnen Klassenstufen 5 bis 8 in systematischer Form identifiziert. Zudem werden alle Beweisversuche bezüglich des Nachweises der Gültigkeit einer zu zeigenden Aussage und mit Blick auf Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche klassifiziert. Die wichtigsten Ergebnisse der differenzierten Analyse von Beweisversuchen werden in Abschnitt 8.5 herausgearbeitet und resümiert.

In Kapitel 9 werden das Design und die Durchführung der im Rahmen dieser Arbeit erfolgten Studie, die Vorgehensweise bei der Auswertung der erhobenen Daten und die damit erzielten Ergebnisse diskutiert. Zudem werden ausgehend von den Ergebnissen Ansatzpunkte für eine Förderung der Teilkompetenz des Führens von Beweisen generiert. Auf den Überlegungen aus Kapitel 9 aufbauend werden in Kapitel 10 Untersuchungsaspekte, die im Anschluss an diese Arbeit bearbeitet werden könnten, gebündelt und strukturiert. Dabei wird dargelegt, welche dieser Untersuchungsaspekte auf Grundlage der im Rahmen des Forschungsprojekts bereits erhobenen Daten bearbeitet werden könnten. Abschließend wird die weitere Arbeit am Forschungsprojekt skizziert.

1.3 Für die Leserinnen und Leser

Mit dem Ziel, die vorliegende Arbeit leserfreundlich zu gestalten, befindet sich am Ende jedes Kapitels ein grau hinterlegter Kasten, in dem wesentliche Aussagen zusammengefasst sind. Eine Möglichkeit der Annäherung an diese Arbeit besteht somit darin, zunächst diese Zusammenfassungen zu lesen. Bei der detaillierten Lektüre können die Einleitungen zu Beginn jedes Kapitels als Orientierung dienen. Zudem sind innerhalb des Textes zahlreiche Untergliederungen vorgenommen und strukturierende Ausdrücke per Fettdruck hervorgehoben.

Ich wünsche Ihnen eine interessante und angenehme Lektüre.

2 Die Mathematik-Olympiade

In diesem Kapitel wird die Mathematik-Olympiade in Deutschland, die den Rahmen der vorliegenden Arbeit bildet, näher vorgestellt. Es werden Charakteristika und Strukturen der Mathematik-Olympiade beschrieben und Einblicke in die Anforderungen zum Beweisen bei der Mathematik-Olympiade gegeben.

Die Mathematik-Olympiade in Deutschland ist mit bis zu 200 000 Teilnehmerinnen und Teilnehmern pro Jahr ein großer Schülerwettbewerb. Wesentliches Ziel dieses Wettbewerbs ist das Finden und Fördern mathematisch interessierter und leistungsstarker Schülerinnen und Schüler in den Klassenstufen drei bis dreizehn.

> Die Mathematik-Olympiade bietet allen interessierten Schülerinnen und Schülern die Möglichkeit, ihre besondere Leistungsfähigkeit auf mathematischem Gebiet unter Beweis zu stellen. Der nach Altersstufen gegliederte Wettbewerb für Schülerinnen und Schüler ab Klasse 3 erfordert logisches Denken, Kombinationsfähigkeit und kreativen Umgang mit mathematischen Methoden. Die Teilnahme am Wettbewerb regt häufig zu einer weit über den Unterricht hinausreichenden Beschäftigung mit der Mathematik an. (Mathematik-Olympiaden e. V., o. J., o. S.)

Damit viele interessierte Schülerinnen und Schüler aller Klassenstufen ihre mathematische Leistungsfähigkeit erproben und auf diesem Wege besonders leistungsstarke Lernende gefunden werden können, sind ein mehrstufiger Wettbewerbsverlauf und elaborierte Organisationsstrukturen sowie gute Wettbewerbsaufgaben erforderlich. In Abschnitt 2.1 werden wesentliche Aspekte der Entwicklung der Mathematik-Olympiade beschrieben, die diese Abläufe und Organisationsstrukturen bis heute prägen. Daran anknüpfend werden in Abschnitt 2.2 die derzeitige Konzeption und Durchführungspraxis der Mathematik-Olympiade dargestellt. Der Abschnitt 2.3 gibt Einblicke in die Entwicklung der Wettbewerbsaufgaben und die Anforderungen zum Beweisen, die den Wettbewerbsteilnehmenden gestellt werden. Diese Ausführungen zur Konzeption der Mathematik-Olympiade und insbesondere ihrer Aufgaben sind beim Design der Studie im Rahmen dieser Arbeit sowie der Auswertung und Interpretation ihrer Ergebnisse zu berücksichtigen (Kap. 6, 8 – 10).

2.1 Entwicklung der Mathematik-Olympiade

Im Schuljahr 1961/62 wurde in der **Deutschen Demokratischen Republik (DDR)**[6] die erste Olympiade für das Fach Mathematik durchgeführt. In den beiden Jahren zuvor hatte Rumänien alle europäischen Ostblockstaaten zu einer Internationalen Mathematik-Olympiade (IMO) eingeladen. Die Schülerinnen und Schüler aus der DDR schnitten dort jedoch in beiden Jahren sehr schwach ab, weshalb die Regierung von da an alle Maßnahmen zur Verbesserung der mathematischen Bildung unterstützte. So wurde Anfang der sechziger Jahre die Olympiade Junger Mathematiker als Teil eines vom Ministerium für Volksbildung unterstützten, umfassenden Programms zur Verbesserung der mathematischen Bildung in der DDR etabliert. Durch diesen

[6] Dieser Abschnitt zur DDR basiert auf den Erläuterungen von Engel (2011) und Sewerin (1979).

Wettbewerb sollten möglichst viele Schülerinnen und Schüler zur Beschäftigung mit Mathematik angeregt, deren Interesse an Mathematik gestärkt, leistungsstarke Lernende früh gefunden und Teilnehmende für die IMO ausgewählt werden. Um möglichst viele Schülerinnen und Schüler zur Beschäftigung mit Mathematik anzuregen und zugleich besonders leistungsstarke Lernende zu identifizieren, wurde die Olympiade Junger Mathematiker als vierstufiger Wettbewerb gestaltet. Sie startete für alle Klassenstufen[7] 5 bis 12 stets im September mit einer Schulrunde, so dass viele und vor allem auch junge Lernende teilnehmen konnten. Über die fortlaufende Qualifikation für die darauffolgenden, zunehmend anspruchsvollen Runden, konnten sich besonders leistungsstarke Schülerinnen und Schüler profilieren. Nach der Schulrunde folgte für die besten Teilnehmenden im November ein Klausurwettbewerb auf Kreisebene. Die erfolgreichsten Lernenden ab der Klassenstufe 7 nahmen anschließend im Februar an einem zweitägigen Klausurwettbewerb auf Bezirksebene teil. Schülerinnen und Schüler ab der Klassenstufe 10 konnten sich dabei für die DDR-Olympiade, einem ebenfalls zweitägigen Klausurwettbewerb im März oder April, qualifizieren.[8] „Aus den Preisträgerinnen und -trägern dieser vierten Runde wurden 12 bis 14 Kandidaten für einen zehntägigen Vorbereitungskurs zur Internationalen Mathematik-Olympiade ausgewählt, von denen 8 (später 6) die DDR-Mannschaft letztendlich bildeten“ (Engel, 2011, S. 18).

Die finanzielle Unterstützung der Olympiade Junger Mathematiker endete mit der **Wiedervereinigung**, und eine weiterführende Unterstützungsmaßnahme war nicht zu erwarten (Gronau, 2011). Die Olympiade wurde als Konkurrenz zum in der Bundesrepublik Deutschland ab 1970 durchgeführten Bundeswettbewerb Mathematik[9] angesehen. Darüber hinaus wollte die neue Regierung nur einen Wettbewerb pro Fächerkomplex. In einigen Bundesländern wurde der Klausurwettbewerb als Eliteförderung abgelehnt, die Notwendigkeit eines solchen Wettbewerbs nicht gesehen, oder die Verfügbarkeit ausreichend vieler ehrenamtlich engagierter Lehrkräfte und damit die Durchführbarkeit bezweifelt (Gronau, 2011). Trotz dieser schwierigen Situation gelang es, dank des großen und anhaltenden Engagements einer wachsenden Personengruppe[10], die Olympiaden fortzuführen und auf die westdeutschen Bundesländer auszudehnen. In den Schuljahren 1990/91 bis 1994/95 lud jeweils Thüringen, Sachsen-Anhalt oder

[7] An dieser Stelle präziser wäre der Ausdruck „Olympiadeklassen“, weil manche Schülerinnen und Schüler an einer höheren Olympiadeklasse als ihrer Klassenstufe teilgenommen haben, etwa um die Chance zu haben, an der DDR-Olympiade teilzunehmen. Grundsätzlich stimmt die Olympiadeklasse mit der Klassenstufe der teilnehmenden Schülerinnen und Schüler überein, insbesondere mit Blick auf Lernende in den Klassenstufen 5 bis 8. Zugunsten einer guten Lesbarkeit wird in dieser Arbeit durchgehend der Begriff „Klassenstufe“ verwendet.

[8] Für das Schuljahr 1966/67 sind die Teilnehmerzahlen bekannt. Rund eine Million Schülerinnen und Schüler nahmen an der ersten Runde der Olympiade Junger Mathematiker teil, das waren über 70% aller zur Teilnahme berechtigten Lernenden. In der Kreisrunde waren es dann rund 50 000, in der Bezirksrunde knapp 2800 und bei der DDR-Olympiade 240 Schülerinnen und Schüler.

[9] Der Bundeswettbewerb Mathematik ist ein mathematischer Schülerwettbewerb, der sich primär an die Klassenstufen 9 bis 13 richtet. Er besteht aus zwei Hausaufgabenrunden, in denen jeweils vier anspruchsvolle Problemlöseaufgaben zu bearbeiten sind, und einem mathematischen Fachgespräch (Kolloquium) in der abschließenden dritten Runde (Bildung und Begabung gGmbH, 2018a).

[10] Besonders hervorzuheben sind an dieser Stelle Prof. Dr. Wolfgang Engel, spiritus rector und langjähriger Verantwortlicher der Olympiade Junger Mathematiker, Prof. Dr. Hans-Dietrich Gronau, ab 1986 Vorsitzender des Zentralen Komitees der Olympiaden Junger Mathematiker sowie Initiator und langjähriger Vorsitzender des Mathematik-Olympiaden e. V., Dr. Ludwig Stammler, von 1986 bis 2001 Leiter des Aufgabenausschusses der Olympiaden, und Dr. Wolfgang Moldenhauer, der nach der Wiedervereinigung die ersten Bundesrunden in Erfurt ermöglichte und die Arbeit des Aufgabenausschusses wesentlich mit prägte.

Sachsen zu einer Bundesrunde (zuvor: DDR-Olympiade) ein, im Jahr 1995/96 fand die erste Bundesrunde in einem westlichen Bundesland und unter Beteiligung aller 16 Bundesländer statt (Gronau, 2011). In der Zwischenzeit wurde ein Koordinierungskomitee für die **Mathematik-Olympiaden** und davon ausgehend der Mathematik-Olympiaden e. V. gegründet. Um der Gleichstellungsdebatte gerecht zu werden und die Weiterentwicklung der Olympiaden zum Ausdruck zu bringen, wurde dabei der Name „Olympiade Junger Mathematiker“ in „Mathematik-Olympiaden“ geändert (Gronau, 2011). Mitte der 90er Jahre wurde die Mathematik-Olympiade zu einem vierstufigen Wettbewerb für alle Lernenden ab der Klassenstufe 5 ausgebaut und die Bundesrunde um die Klassenstufen 8, 9 erweitert (Kugel, o. J.). Im Jahr 2005 wurden zusätzlich die Klassenstufen 3, 4 in den Wettbewerb aufgenommen (Mathematik-Olympiaden e. V., 2005, 2006).

Die Idee, die Mathematik-Olympiade in die bestehenden **Bundesweiten Mathematik-Wettbewerbe** einzubinden, ermöglichte schließlich auch eine finanzielle Absicherung des Wettbewerbs. Zusammen mit dem Bundeswettbewerb Mathematik und dem Auswahlwettbewerb zur Internationalen Mathematik-Olympiade wird die deutsche Mathematik-Olympiade seit 2007 unter dem Dach der Bildung und Begabung gGmbH von einer Geschäftsstelle verwaltet. Zudem werden die Bundesrunden und die Aufgabenentwicklung vom Bundesministerium für Bildung und Forschung, vom Stifterverband, von der Kultusministerkonferenz und von der Hector-Stiftung II gefördert. Seit der Einbindung in die Bundesweiten Mathematik-Wettbewerbe 2007 wird die Mathematik-Olympiade jedes Schuljahr für die Klassenstufen 3 bis 12 in ähnlicher Form durchgeführt. Diese inzwischen fest etablierten Strukturen und Abläufe werden im Folgenden näher erläutert.

2.2 Heutige Konzeption und Durchführungspraxis

Die heutigen Organisationsstrukturen der Mathematik-Olympiade sind im Wesentlichen in zwei Ebenen gegliedert. Der bereits vorgestellte bundesweite Verein Mathematik-Olympiaden e. V. befindet sich auf der übergeordneten Ebene. Er unterstützt die Durchführung von Bundesrunden und stellt Aufgaben, Lösungen und Punktverteilungsvorschläge für alle Wettbewerbsrunden zur Verfügung (Mathematik-Olympiaden e. V., 2016). Auf der zweiten Ebene befinden sich die Verantwortlichen der einzelnen Bundesländer. Sie sind für die Durchführung der ersten drei Runden und die Auswahl des eigenen Länderteams für die Bundesrunde der Mathematik-Olympiade verantwortlich.

2.2.1 Bundesweite Konzeption

Die Mathematik-Olympiade ist auf nationaler Ebene in vier aufeinander aufbauende Wettbewerbsrunden strukturiert (Abb. 2). Ähnlich wie bereits in der DDR beginnt die Olympiade jedes Jahr im September mit einer **Schulrunde**. Diese ist als Hausaufgabenwettbewerb mit einer Bearbeitungszeit von rund vier Wochen konzipiert, kann aber auch als Klausur in der Schule durchgeführt werden (Bildung und Begabung gGmbH, 2018c).

Die weiteren Runden sind Klausurwettbewerbe. Nach der Schulrunde folgt zuerst Mitte November die **Regionalrunde**. Laut den Empfehlungen des Mathematik-Olympiaden e. V. (2018) soll in den Klassenstufen 5, 6 eine vierstündige Klausur geschrieben werden, für alle höheren Klassenstufen wird eine Bearbeitungszeit von vier bis sechs Stunden vorgeschlagen. In dieser Klausur sollen nur Schreib- und Zeichenutensilien verwendet werden. Die Nutzung von Hilfsmitteln, wie Formelsammlungen oder Taschenrechnern, ist nicht vorgesehen. Außerdem wird auf eine eigenständige Bearbeitung und einen fairen Ablauf Wert gelegt. Inhaltliche Fragen sollen deshalb ausschließlich in der Anfangsphase der Klausur und durch eine verantwortliche Person beantwortet werden.

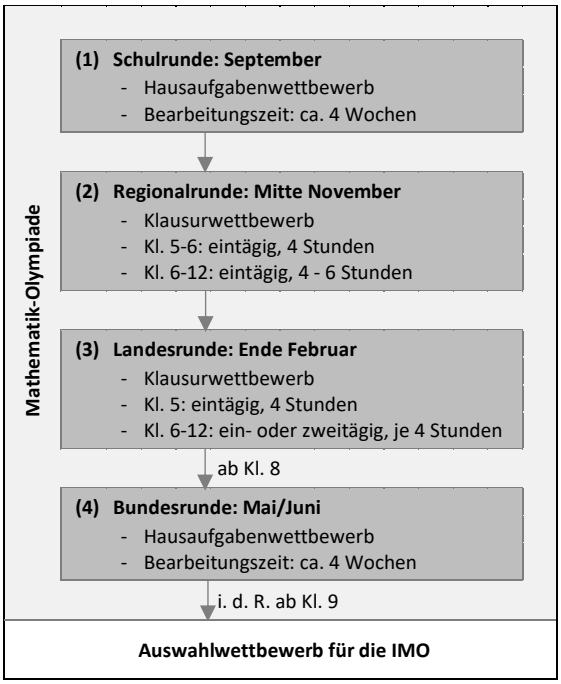


Abbildung 2: Bundesweite Konzeption der Mathematik-Olympiade für Schülerinnen und Schüler in den Sekundarstufen

Die Erfolgreichsten dieses regionalen Klausurwettbewerbs dürfen Ende Februar die **Landesrunde** bestreiten. Die Landesrunde ist für die Klassenstufe 5 als eintägiger Klausurwettbewerb, für alle höheren Klassenstufen als ein- oder zweitägiger Klausurwettbewerb mit einer je vierstündigen Bearbeitungszeit konzipiert. Auch hier ist eine eigenständige und hilfsmittelfreie Bearbeitung der Klausur vorgesehen (Mathematik-Olympiaden e. V., 2018). In der Landesrunde haben Schülerinnen und Schüler ab der Klassenstufe 8 die Chance, sich für die Bundesrunde zu qualifizieren.

Die **Bundesrunde** wird im Mai oder Juni in einem jährlich wechselnden Bundesland durchgeführt. Eingebettet in ein buntes Rahmenprogramm schreiben die Teilnehmenden an zwei aufeinanderfolgenden Vormittagen viereinhalbstündige, hilfsmittelfreie Klausuren. Dabei dürfen sie innerhalb der ersten beiden Stunden einer jeden Klausur schriftlich Fragen zum Aufgabentext stellen, die dann von drei Hauptverantwortlichen schriftlich beantwortet werden (Mathematik-Olympiaden e. V., 2012b). Die Korrektur der Klausuren erfolgt direkt vor Ort, sodass am letzten Tag der Bundesrunde die Preisverleihung stattfinden kann. In jeder Klassenstufe werden erste, zweite und dritte Preise sowie Anerkennungen vergeben. Die entsprechenden Punktegrenzen werden im Anschluss an die Korrektur separat für jede Klassenstufe von der Jury festgelegt. Dabei wird darauf geachtet, dass die Anzahlen an ersten, zweiten und dritten Preisen ungefähr im Verhältnis $1:2:3$ stehen, und etwa 40% der Teilnehmenden einer Klassenstufe einen Preis erhalten (Mathematik-Olympiaden e. V., 2012c). Wie viele Punkte für einen Preis oder eine Anerkennung erforderlich sind, hängt also von den erreichten Punktzahlen aller Teilnehmenden einer Klassenstufe ab. Auch für einen ersten Preis ist folglich nicht unbedingt die volle Punktzahl erforderlich, aber eine der höchsten Punktzahlen innerhalb der jeweiligen Klassenstufe.

Nach der Bundesrunde erhalten viele der Preisträgerinnen und Preisträger die Möglichkeit, sich über einen **Auswahlwettbewerb für die Internationale Mathematik-Olympiade** zu qualifizieren. Lernende aus der Klassenstufe 8 werden jedoch wegen des sehr hohen Anforderungsniveaus in der Regel nicht eingeladen (Mathematik-Olympiaden e. V., 2012b).[11]

Für Schülerinnen und Schüler der Klassenstufen 5 bis 8 umfasst die Mathematik-Olympiade folglich i. A. maximal vier Wettbewerbsrunden. In jeder dieser Wettbewerbsrunden sind die Lernenden aufgefordert, drei bis sechs auf ihre Klassenstufe abgestimmte Problemlöseaufgaben zu bearbeiten. Diese Aufgaben, dazugehörige Lösungen und Punktverteilungsvorschläge werden von einem **bundesweiten Aufgabenausschuss** des Mathematik-Olympiade e. V. entwickelt (Abschn. 2.3). In der Bundesrunde werden genau diese vom Aufgabenausschuss entwickelten Aufgaben gestellt und die Aufgabenbearbeitungen entsprechend der Vorgaben bewertet. In den ersten drei Wettbewerbsrunden hingegen wird von den Vorschlägen mancherorts abgewichen. Während die Bundesrunde vom Mathematik-Olympiaden e. V., der Geschäftsstelle und lokalen Ausrichtern organisiert wird, liegt die Verantwortung für die ersten drei Wettbewerbsrunden, d. h. für die Schul-, Regional- und Landesrunde, bei den jeweiligen **Bundesländern**. Für die Durchführung dieser Wettbewerbsrunden werden die vom Aufgabenausschuss entwickelten Aufgaben, Lösungen und Punktverteilungsvorschläge an die Landesbeauftragten weitergeleitet. Die Verwendung dieser Dokumente ist üblich, aber nicht bindend. Außerdem sind weitere Variationen in der Durchführung, wie beispielsweise die Reduzierung von drei auf zwei Runden oder die Veränderung der Klausurzeit, möglich.[12] Da die vorliegende Arbeit auf Nordrhein-Westfalen (NRW) fokussiert, wird im Folgenden die dortige Wettbewerbspraxis im Detail vorgestellt. Die Anlage des Wettbewerbs in NRW wird den Rahmen der in dieser Arbeit vorgenommenen Untersuchung von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade bilden (Kap. 6).

2.2.2 Wettbewerbspraxis in Nordrhein-Westfalen

In Nordrhein-Westfalen (NRW) findet entsprechend der vorgestellten Konzeption der Mathematik-Olympiade jedes Schuljahr ein dreistufiger Wettbewerb statt. Dieser Wettbewerb wird von einem landesweiten Verein, dem **Landesverband Mathematikwettbewerbe NRW** e. V., koordiniert und in Teilen organisiert. Der Landesverband organisiert die jährliche NRW-Landesrunde, er koordiniert und unterstützt die Durchführung der Schul- und Regionalrunden.

Für diese ersten beiden Runden wurde das Land nach den Kreisen und kreisfreien Städten NRWs in 54 Regionen aufgeteilt, die von je ein bis zwei **Regionalkoordinatoren** betreut werden. Zu Beginn jedes Schuljahres lädt der Landesverband alle Koordinatorinnen und Koordinatoren zu einem Treffen ein. Hier werden die organisatorischen Rahmenbedingungen für die ersten zwei Runden besprochen und Erfahrungen ausgetauscht (Ernst-Brandt, 2011). Über die

[11] Für weiterführende Informationen zum Auswahlwettbewerb und zur Internationalen Mathematik-Olympiade siehe Bildung und Begabung gGmbH (2018a, 2018b).

[12] Dieser Gestaltungsspielraum ermöglichte nach der Wiedervereinigung, die Mathematik-Olympiade in den westlichen Bundesländern in bereits gewachsene Strukturen zu integrieren. Auch heutzutage ist diese gewisse Freiheit der Bundesländer wegen des föderalen Bildungssystems, der unterschiedlichen demographischen Strukturen und der gewachsenen Organisationsstrukturen nicht wegzudenken.

konkrete organisatorische und inhaltliche Ausgestaltung der **Schul- und Regionalrunden** dürfen die Regionalkoordinatorinnen und -koordinatoren entscheiden.[13] Diese Gestaltungsfreiheit hat Auswirkungen auf die Erhebung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade, die im Rahmen einer Regionalrunde in zahlreichen Regionen NRWs durchgeführt wird (Abschn. 6.2.1). Deshalb werden nun die Teilnahmebedingungen der Regionalrunde, die Aufgabenauswahl, die Klausurbedingungen, sowie die Korrektur und Bewertung der Klausurbearbeitungen mit Blick auf die Regionalrunde beschrieben, und aufgezeigt, wo Gemeinsamkeiten oder aber kleine Unterschiede zwischen den Praktiken in den Regionen bestehen. Danach werden Charakteristika der Landesrunde der Mathematik-Olympiade in NRW vorgestellt, die für die Untersuchung der Beweisversuche relevant sind.

2.2.2.1 Regionalrunde der Mathematik-Olympiade in NRW

An der Regionalrunde der Mathematik-Olympiade nehmen in der Regel Schülerinnen und Schüler teil, die die Aufgaben der Schulrunde gut bearbeitet haben.[14] Für die Durchführung der Schulrunde und die anschließende **Selektion der Teilnehmenden der Regionalrunde** leiten die meisten Regionalkoordinatorinnen und -koordinatoren die Aufgaben, die vom Aufgabenausschuss für die Schulrunde herausgegeben wurden, an die Verantwortlichen der einzelnen Schulen weiter. Den Schulen wird dabei empfohlen, nur Lernende zur Regionalrunde zu schicken, die in der Schulrunde mindestens die Hälfte der möglichen Punkte erreichen. Da die Aufgaben der Schulrunde normalerweise zu Hause bearbeitet werden dürfen, kann es gerade in den unteren Klassenstufen in Einzelfällen vorkommen, dass Eltern oder Geschwister die Lernenden unterstützen. Auch können sich die Strenge der Korrektur und die vorhandene Konkurrenz an den einzelnen Schulen unterscheiden, was zu Unterschieden in den Eingangsvoraussetzungen der Teilnehmenden der Regionalrunde beitragen kann. In manchen Regionen werden diese Effekte durch eine schulübergreifende Korrektur oder die Durchführung eines Klausurwettbewerbs verringert. Über alle Regionen hinweg nehmen also überwiegend interessierte und leistungsstarke Schülerinnen und Schüler an der Regionalrunde teil. Das Spektrum der Teilnehmenden kann jedoch von mathematisch eher leistungsschwachen und allenfalls extrinsisch motivierten Lernenden bis hin zu mathematikbegeisterten und mathematisch außerordentlich leistungsstarken Schülerinnen und Schülern reichen.[15]

[13] Diese Gestaltungsfreiheit ermöglichte auch auf regionaler Ebene, die bereits vor der Einführung der Mathematik-Olympiade bestehenden Strukturen zu integrieren, und erlaubt den Verantwortlichen die Organisationsstrukturen an die räumlichen und personellen Gegebenheiten anzupassen.

[14] Ausschließlich in einer Region dürfen alle interessierten Schülerinnen und Schüler ohne Vorauswahl die Regionalrunde bestreiten, die Aufgaben der Schulrunde werden hier zur Übung ausgegeben. In allen anderen Regionen müssen sich die Schülerinnen und Schüler zunächst in der Schulrunde qualifizieren.

[15] In NRW nehmen jedes Jahr $14\,000$ bis $17\,000$ Lernende aus den Klassenstufen fünf bis dreizehn an der Schulrunde der Mathematik-Olympiade teil. Rund 40% dieser Lernenden dürfen anschließend die Regionalrunde bestreiten, in den zehn Schuljahren 2006/07 bis 2015/16 waren es stets 5600 bis 7100 Teilnehmende. Die Anzahl an Teilnehmenden pro Region differiert dabei zwischen etwa 5 und über 350 Schülerinnen und Schülern, wobei Teilnehmerzahlen unter 30 und über 250 nur in wenigen Regionen vorkommen. Wie sich die Lernenden auf die einzelnen Olympiadeklassen verteilen, wird nicht erfasst. Der Erfahrung nach ist bei der Regionalrunde die Anzahl an Fünftklässlerinnen und -klässlern besonders hoch. Ab der Klassenstufe sechs sind die Zahlen niedriger.

In der Regionalrunde schreiben alle Teilnehmenden eine **Klausur**, die drei bis vier Aufgaben umfasst. Für diese Klausur werden die vom Mathematik-Olympiaden e. V. herausgegebenen Aufgabensätze genutzt. Diese umfassen pro Klassenstufe vier Aufgaben. Beispielsweise in Bonn und Wuppertal werden diese Aufgabensätze unverändert übernommen, während in anderen Regionen kleinere Modifikationen erfolgen. Da für die Klausur vielerorts weniger als die vom Mathematik-Olympiaden e. V. vorgeschlagenen vier Zeitstunden zur Verfügung stehen, wird oftmals eine der vier Aufgabe gestrichen oder den Lernenden die Auswahl dreier Aufgaben überlassen. Zudem werden mancherorts entsprechend den Kernlehrplänen NRWs Aufgaben zwischen den Klassenstufen verschoben, Formulierungen angepasst oder Begriffserklärungen in den Aufgabenstellungen ergänzt. In Essen gibt es einen Verein, der die Aufgabensätze für die Essener Klausurdauern jedes Jahr sehr sorgfältig zusammenstellt und diese Aufgabensätze dann an andere Regionen weiterleitet, weshalb trotz der Abweichungen eine gewisse Einheitlichkeit besteht.

In jeder Region findet die Klausur am Donnerstag, Freitag oder Samstag nach Beginn des bundesweiten Wettbewerbszeitraums Mitte November statt. Die Dauer der Klausur wird innerhalb der einzelnen Regionen festgelegt und differiert deutlich. In den Klassen 5 und 6 erhalten die Lernenden zwischen eineinhalb und vier Zeitstunden, in den Klassen 7 und 8 zwei bis vier Zeitstunden. In der Klausur werden höchstens Taschenrechner zugelassen, die weder über eine Computeralgebra-Funktionalität noch über grafische Darstellungsmöglichkeiten verfügen. Weitere Hilfsmittel sind nirgends erlaubt. Am Ende der Klausur sollen die Schülerinnen und Schüler alle Dokumente abgeben. Der Landesverband bittet die Regionalkoordinatorinnen und -koordinatoren zudem darum, auch anwesenden Lehrkräften die Aufgabenblätter, Lösungen und Schülerbearbeitungen nicht vor Samstagnachmittag mit nach Hause zu geben, um das Bekanntwerden der Dokumente vor Abschluss des Wettbewerbszeitraums zu vermeiden.

Die **Korrektur** der Klausurbearbeitungen erfolgt innerhalb der einzelnen Regionen, bis spätestens Mitte Dezember. Danach werden die Klausuren den Lernenden in einigen Regionen zurückgegeben, in anderen werden die Klausuren einbehalten, und mancherorts werden sie für eine kurze Einsicht ausgehändigt. Außerdem werden nach der Korrektur die Preisträgerinnen und Preisträger der Regionalrunde sowie die Teilnehmenden der Landesrunde festgelegt. Die Preisvergabe wird in den Regionen individuell gehandhabt, und auch für die **Auswahl der Teilnehmenden der Landesrunde**[16] werden unterschiedliche Kriterien verwendet. Die Auswahl hängt zum einen von der Anzahl an Plätzen ab, die eine Region vergeben darf. Diese Anzahl an Plätzen wird vor der Regionalrunde vom Landesverband anhand der Anzahlen der teilnehmenden Schulen, der Anzahlen der teilnehmenden Schülerinnen und Schüler und den Erfolgen in den letzten drei Landesrunden berechnet. Jede Region darf mindestens zwei und höchstens zwanzig Lernende zur Landesrunde entsenden. In manchen Regionen werden die Schülerinnen und Schüler mit der höchsten Gesamtpunktzahl ausgewählt. In anderen Regionen erhalten die besten Teilnehmenden jeder Klassenstufe einen Platz. Mancherorts werden zudem besondere

[16] Von allen Teilnehmenden der Regionalrunde dürfen rund fünf Prozent zur Landesrunde fahren. Die Teilnehmerzahl der Landesrunde wurde vom Landesverband auf 350 festgelegt und schwankt allenfalls geringfügig z. B. durch Krankheitsfälle. Von diesen 350 Lernenden sind meistens mehr als 150 aus den Klassenstufen fünf bis acht. Die genauen Anzahlen hängen jedoch von der Auswahl der Teilnehmenden in den Regionen ab und variieren deutlich.

Erfolge im Vorjahr berücksichtigt, oder z. B. die Fünftklässlerinnen und Fünftklässler von der Auswahl für die Landesrunde ausgeschlossen. Teilnehmerin oder Teilnehmer bei der Landesrunde zu sein, heißt folglich nicht unbedingt bei der Regionalrunde die volle Punktzahl erreicht zu haben, weist aber in den meisten Fällen auf eine im Vergleich besonders erfolgreiche Bearbeitung der Wettbewerbsklausur hin. Zugleich ist es möglich, dass Schülerinnen und Schüler trotz einer hohen Punktzahl keinen Platz für die Landesrunde erhalten.

2.2.2.2 Landesrunde der Mathematik-Olympiade in NRW

Alle ausgewählten Schülerinnen und Schüler dürfen Ende Februar an der **Landesrunde** teilnehmen. Diese wird in NRW für Lernende aller Klassenstufen fünf bis dreizehn als eintägiger Wettbewerb gestaltet. Am Vormittag schreiben die Schülerinnen und Schüler die Wettbewerbsklausur. Den Lernenden in den Klassenstufen 5 bis 8 stehen dafür stets zwei bis drei Stunden Zeit zur Verfügung, wobei die Siebt- und Achtklässler meistens 15 bis 30 Minuten länger als die Jüngeren schreiben dürfen. In der Klausur erhalten die Lernenden drei bis vier auf ihre Klassenstufe abgestimmte Aufgaben. Für jede Klassenstufe sind zwei Aufgabenkoordinatoren verantwortlich. Sie stellen die Aufgabenblätter ähnlich wie bei der Regionalrunde in Essen ausgehend von den Vorschlägen des Mathematik-Olympiaden e. V. zusammen und sind zu Beginn der Klausur anwesend, um Fragen zu beantworten und einen fairen Wettbewerb zu gewährleisten. Hilfsmittel wie Taschenrechner sind normalerweise nicht erlaubt.

Die Korrektur erfolgt unter der Leitung der Aufgabenkoordinatoren getrennt nach Klassenstufen. Als Orientierung bei der Korrektur dienen die vom Mathematik-Olympiaden e. V. herausgegebenen Musterlösungen mit Punktverteilungsvorschlägen für alle Aufgabenteile. Für jede Aufgabe sind immer mehrere Personen zuständig, die sich zu Beginn nach der Durchsicht einzelner Schülerbearbeitungen auf eine endgültige Bewertungsgrundlage einigen und bei Bedarf während der Korrektur kurz Rücksprache halten. Es ist insbesondere möglich, dass eine Schülerin oder ein Schüler für eine Aufgabe die volle Punktzahl erhält, obwohl die Bearbeitung vom vorgeschlagenen Lösungsweg abweicht. Wenn alle Klausuren bepunktet sind, werden ähnlich wie bei der Bundesrunde erste, zweite und dritte Preise sowie Anerkennungen vergeben (Abschn. 2.2.1). Alle Preisträgerinnen und Preisträger werden zu einer einwöchigen Mathematischen Sommerakademie eingeladen. Wegen der begrenzten Plätze bei der Sommerakademie wird die Anzahl an Preisträgerinnen und Preisträgern auf rund 30% der Schülerinnen und Schüler einer jeden Klassenstufe beschränkt. Bei den ersten, zweiten und dritten Preisen wird ein Verhältnis von $1 : 2 : 3$ angestrebt. Alle ersten Preisträgerinnen und Preisträger ab der Klassenstufe 8 sind für die Bundesrunde qualifiziert. Die nach diesem Kriterium übrigbleibenden Plätze für die Bundesrunde werden gegebenenfalls von einem kleinen Auswahlteam an weitere erfolgreiche Teilnehmerinnen und Teilnehmer der Landesrunde vergeben.[17]

[17] Aus der Olympiadeklasse acht durften in den Schuljahren 2006/07 und 2015/16 immer ein bis drei Schülerinnen und Schüler an der Bundesrunde teilnehmen. Insgesamt durfte NRW stets dreizehn bis fünfzehn Teilnehmende zur Bundesrunde anmelden, das heißt ca. vier Prozent der Teilnehmenden der Landesrunde bzw. ca. ein Tausendstel aller Schülerinnen und Schüler in den Sekundarstufen, die mit der Schulrunde beginnen.

2.2.3 Die Verbindung von Wettbewerb und Förderung

Die Teilnahme an den verschiedenen Runden der Mathematik-Olympiade ist zum Teil mit mathematischen Förderangeboten verbunden, die im Folgenden mit besonderem Fokus auf NRW skizziert werden. Die Schul- und die Regionalrunde werden mancherorts durch schulische oder überschulische Förderangebote, wie Arbeitsgemeinschaften, begleitet. Nach der Regionalrunde finden normalerweise in über zehn Regionen NRWs Mathematikwochenenden speziell für erfolgreiche Teilnehmende dieser zweiten Wettbewerbsrunde statt.[18] Hinzu kommen weitere, i. A. regionale Angebote zur Vorbereitung auf die Landesrunde. Nach der Landesrunde dürfen alle Preisträgerinnen und Preisträger die einwöchige Mathematische Sommerakademie NRW besuchen. Die Teilnehmenden der Bundesrunde werden zudem gemeinsam mit einigen weiteren, besonders leistungsstarken Lernenden zu zwei Vorbereitungswochenenden auf diesen bundesweiten Wettbewerb eingeladen. Außerdem gibt es seit 2007 das bundesweite Spitzenförderprogramm „Jugend trainiert Mathematik“ (JuMa).

> Jugend trainiert Mathematik (JuMa) ist als bundesweite Fördermaßnahme für Schülerinnen und Schüler ab Klasse 7 konzipiert. JuMa führt die Teilnehmenden an das Niveau der bundesweiten Mathematikwettbewerbe, den Bundeswettbewerb Mathematik und die Mathematik-Olympiade in Deutschland in den höheren Stufen, heran. (Bildung und Begabung gGmbH, o. D, o. S.)

Der Landesverband Mathematikwettbewerbe NRW e. V. darf jedes Frühjahr zwölf Schülerinnen und Schüler der Klassenstufe 7 für JuMa anmelden. Die Auswahl der Siebtklässlerinnen und Siebtklässler erfolgt ausgehend von ihren Ergebnissen bei der Mathematik-Olympiade in den Klassen fünf bis sieben. Zudem werden Eindrücke aus den beschriebenen Förderangeboten und das Interesse der Schülerinnen und Schüler an JuMa berücksichtigt.

Über diese Enrichment-Angebote innerhalb der Mathematik-Olympiade hinaus ist es denkbar, dass sich Lernende selbstständig beispielweise auf die Regionalrunde vorbereiten und auf diese Weise ihre mathematischen Kompetenzen weiterentwickeln. Dafür werden vom Mathematik-Olympiaden e. V. die Aufgabensätze der vorhergehenden Wettbewerbsdurchläufe im Downloadbereich als Materialien bereitgestellt (Mathematik-Olympiaden e. V., o. J.).

Außerdem gibt es Förderangebote für mathematisch Interessierte in den Klassenstufen 5 bis 8, die in keinem direkten Zusammenhang zur Mathematik-Olympiade stehen. Dazu gehören zahlreiche schulinterne und schulübergreifende mathematische Arbeitsgemeinschaften, sowie Formate, die zusätzlich zur Mathematik auch andere MINT-Fächer berücksichtigen. Hinzu kommen vielfältige Angebote von Universitäten, mehrere landesweite mathematische Akademien des Landesverbands Mathematikwettbewerbe NRW e. V., Deutsche JuniorAkademien, Korrespondenzzirkel, wie die Fernschule für Mathematik, und Lernangebote im Internet. Die Mathematik-Olympiade und der damit verbundene Austausch mit Gleichgesinnten können dazu beitragen, dass Schülerinnen und Schüler von diesen Angeboten erfahren. Zudem können die Aufgaben der Mathematik-Olympiade z. B. in Arbeitsgemeinschaften genutzt werden und somit deren Durchführung unterstützen (QUA-LiS NRW, 2018).

[18] Während der COVID-19-Pandemie wurden weniger Mathematikwochenenden durchgeführt.

2.3 Aufgaben der Mathematik-Olympiade

Die Aufgaben der Mathematik-Olympiade sind das Herzstück des Wettbewerbs (Bauermann, 2018; Langmann, 2016). Neben einer möglichen Verwendung in Enrichment-Angeboten bestimmen die mit ihnen verbundenen Anforderungen wesentlich, welche Schülerinnen und Schüler über die Mathematik-Olympiade gefunden werden. Dafür sollen die Wettbewerbsaufgaben zur Teilnahme an der Schulrunde und somit an der Mathematik-Olympiade motivieren. Der Schwierigkeitsgrad der gestellten Anforderungen bestimmt anschließend mit, welchen Teilnehmenden es gelingt, sich für fortgeschrittene Wettbewerbsrunden zu qualifizieren und welche Teilnehmenden somit auch eher Zugang zu Förderangeboten erhalten. Die Aufgaben und Anforderungen erhalten noch mehr Gewicht, wenn man davon ausgeht, dass der Spaß und Erfolg im Wettbewerb das mathematikspezifische Fähigkeitsselbstkonzept der Lernenden und deren Wille, auch in Zukunft bei der Mathematik-Olympiade mitzumachen oder sich intensiver mit Mathematik zu beschäftigen, beeinflussen.

Im Folgenden werden zuerst Aspekte beschrieben, auf die im Rahmen der Entwicklung der Aufgaben der Mathematik-Olympiade Wert gelegt wird, und die folglich auch bei der eigenen Konstruktion einer Aufgabe für die Erhebung von Beweisversuchen beachtet werden (Abschn. 2.3.1, 6.2). Anschließend werden drei in vergangenen Wettbewerben eingesetzte Aufgaben vorgestellt, um ergänzend die regulären Wettbewerbsanforderungen zum Beweisen einzukreisen (Abschn. 2.3.2).

2.3.1 Entwicklung der Aufgaben

Die Aufgaben der Mathematik-Olympiade werden von einem **bundesweiten Aufgabenausschuss** des Mathematik-Olympiaden e. V. auf ehrenamtlicher Basis entwickelt. Dabei arbeiten über 50 Lehrinnen und Lehrer, Mitarbeiterinnen und Mitarbeiter von Hochschulen, Studierende und ehemals erfolgreiche Teilnehmende der Mathematik-Olympiade in mehreren kleinen Gruppen zusammen (Gronau et al., 2011). Jede dieser Gruppen ist für zwei aufeinanderfolgende Klassenstufen zuständig und entwickelt für diese Klassenstufen alle Aufgaben, Lösungen und Punktverteilungsvorschläge. Für die Klassenstufen 5 bis 8 werden jedes Jahr 60 Aufgaben samt Lösungen und Punktverteilungsvorschlägen veröffentlicht.[19] Insbesondere wird für die Regionalrunde, in deren Rahmen die eigene Erhebung erfolgt, pro Klassenstufe ein Arbeitsblatt mit vier Aufgaben herausgegeben. Die Entwicklung und Auswahl dieser Aufgaben ist ein aufwendiger Prozess[20], in dem auf zahlreiche Aspekte Wert gelegt wird. Diese werden im Folgenden mit besonderem Fokus auf den Klassenstufen 5 bis 8 vorgestellt.

Bei der Entwicklung und Auswahl der Aufgaben von zentraler Bedeutung ist die Passung zur Klassenstufe. Dementsprechend wird versucht, die verschiedenen **curricularen Vorgaben** der Bundesländer angemessen zu berücksichtigen. Auch Schülerinnen und Schüler, die noch nicht

[19] Diese Aussage gilt ab dem Schuljahr 2009/2010. Zuvor wurden z. B. in den Schulrunden der Klassenstufen 5, 6 mehr Aufgaben gestellt.

[20] Da diese Prozesse bisher ausschließlich ansatzweise in Publikationen Erwähnung fanden, wurden die vorhandenen Informationen in Gesprächen mit Mitgliedern des Aufgabenausschusses ergänzt. Für die Offenheit gegenüber meinen Fragen und die sehr interessanten und ausgesprochen angenehmen Gespräche bedanke ich mich herzlich.

gefördert wurden, sollen sich für weiterführende Wettbewerbsrunden qualifizieren und Preise gewinnen können. Insbesondere werden zu Beginn der Klassenstufe 5 nur Inhalte vorausgesetzt, die am Ende der Klassenstufe 4 bekannt sind. In den Klassenstufen 7, 8 wird in den Aufgaben der Regionalrunde, die von einem großen Spektrum unterschiedlich leistungsstarker Lernender bearbeitet werden, mit symbolischen Elementen zurückhaltend umgegangen.

Die für die einzelnen Klassenstufen am besten geeigneten Aufgaben werden in der Regel für die **Regionalrunde** ausgewählt. Da die Schulrunde als Hausaufgabenwettbewerb konzipiert ist, bei dem die Lernenden mehr Zeit für die Bearbeitung haben und sich auch Hilfe holen können, dürfen Aufgaben der Schulrunde schwieriger sein als die der Regionalrunde. Sie werden gerne genutzt, um auf die Anforderungen in weiterführenden Wettbewerbsrunden auf inhaltlicher oder heuristischer Ebene vorzubereiten und trotz der curricularen Einschränkungen hohe Anforderungen stellen zu können. Da die Mathematik-Olympiade in den Klassenstufen 5 bis 7 mit der Landesrunde endet, werden hier auch einzelne Aufgaben gestellt, bei denen viele Teilnehmende sicherlich nicht zur Lösung kommen und dann eventuell nach dem Wettbewerb weiter darüber nachdenken. Ab der achten Klassenstufe qualifizieren sich die Besten der Landesrunde für die Bundesrunde. Daher werden die letzten ein bis zwei Aufgaben der Landesrunde ähnlich schwierig, wie die der Bundesrunde gestaltet. Von den Teilnehmenden der Bundesrunde wird erwartet, dass sie sich bereits deutlich über den schulischen Rahmen hinaus mit Mathematik beschäftigt haben und dementsprechend anspruchsvolle Aufgaben bearbeiten können.

Zusätzlich zu diesen Anforderungsunterschieden zwischen den Wettbewerbsrunden wird eine **Progression der Anforderungen innerhalb jeder Wettbewerbsrunde** angestrebt. Die Aufgabensätze für eine jede Wettbewerbsrunde werden so zusammengestellt, dass die erste Aufgabe möglichst für alle Teilnehmenden dieser Runde lösbar ist, die letzte Aufgabe hingegen auch besonders leistungsstarke Schülerinnen und Schüler herausfordert und differenziert.

Zudem wird versucht, in jeder Wettbewerbsrunde verschiedene **Themenbereiche**, insbesondere stets die Bereiche Geometrie und Arithmetik/Algebra, zu berücksichtigen und folglich die Aufgaben auch fachinhaltlich zu mischen. Insbesondere werden in allen Klassenstufen auch regelmäßig Aufgaben aus dem Bereich der Teilbarkeit ganzer Zahlen gestellt.

Jede einzelne Aufgabe soll altersgerecht und anregend gestaltet sein. Zudem sollen die Formulierungen der **Aufgabenstellungen** einerseits mathematisch präzise, andererseits für die teilnehmenden Schülerinnen und Schüler verständlich sein. Die Aufgaben sollen verschiedene **Lösungswege** und, vor allem in den Klassenstufen 5, 6, auch elementare Herangehensweisen zulassen. Auf anspruchsvolle Arbeitsaufträge, die z. B. keine elementare Herangehensweise zulassen, wird oftmals in einem einführenden Aufgabenteil hingeführt. Solche Arbeitsaufträge sind in der Regel der abschließende Aufgabenteil und fallen daher bei der Bepunktung weniger stark ins Gewicht. Insbesondere ab der Regionalrunde müssen alle Aufgaben prinzipiell im gegebenen **zeitlichen Rahmen** und **ohne Hilfsmittel** gelöst werden können. Darüber hinaus werden die Aufgaben möglichst so angepasst, dass die **Korrektur** reliabel und in einem vertretbaren Zeitrahmen möglich ist.

Zusammenfassend sollen die Aufgaben zu den einzelnen Klassenstufen passen, so dass alle Teilnehmenden Erfolgserlebnisse haben können, aber auch herausgefordert werden. Dazu

wird auf eine Progression der Anforderungen innerhalb des Wettbewerbs, der einzelnen Wettbewerbsrunden und der einzelnen Aufgaben geachtet. Vor allem bei den Aufgaben der Regionalrunde wird versucht, die verschiedenen curricularen Vorgaben der Bundesländer angemessen zu berücksichtigen. Zudem sollen sämtliche Aufgaben ansprechend, altersgerecht und innerhalb der gegebenen Zeit ohne Hilfsmittel bearbeitbar sein.

2.3.2 Anforderungen an Aufgabenbearbeitungen

In jeder Wettbewerbsrunde erhalten die Teilnehmenden der Mathematik-Olympiade, wie beschrieben, Arbeitsblätter mit mehreren Aufgaben. Die Abbildungen 3 bis 5 zeigen exemplarisch solche Aufgaben und die dazu vom Aufgabenausschuss entwickelten Lösungen. Anhand dieser Aufgaben werden im Folgenden Einblicke gegeben, inwieweit Anforderungen im Bereich des Führens von Beweisen gestellt werden.

Klassenstufe 6, Landesrunde, Aufgabe 520631

Eine natürliche Zahl kann die Eigenschaft haben, dass sie durch ihre Quersumme teilbar ist. Ein Beispiel ist 12.

a) Gib zwei zweistellige natürliche Zahlen an, deren größter gemeinsamer Teiler 1 ist und die jeweils durch ihre Quersumme teilbar sind.

b) Untersuche, ob alle durch 9 teilbaren zweistelligen natürlichen Zahlen durch ihre Quersumme teilbar sind.

c) Begründe durch allgemeine Feststellungen, dass alle durch 10 teilbaren zweistelligen natürlichen Zahlen durch ihre Quersumme teilbar sind.

(Mathematik-Olympiaden e. V., 2013, S. 35)

Lösung zu 520631

Teil a) Die ersten zweistelligen natürlichen Zahlen, die durch ihre Quersumme teilbar sind, lauten $10, 12, 18, 20, 21, 24, 27, 30, 36, 40, 42, 45, 48, 50$. Für das gesuchte teilerfremde Paar kommen zum Beispiel $(10, 21)$, $(10, 27)$, $(20, 21)$, $(20, 27)$ und $(27, 40)$ in Betracht.

Teil b) Das ist nicht der Fall. Die durch 9 teilbaren Zahlen 18, 27, 36, …, 90 haben die Quersumme 9 und sind demzufolge durch ihre Quersumme teilbar. Die Zahl 99 hat aber die Quersumme 18, durch die sie nicht teilbar ist.

Teil c) Eine durch 10 teilbare zweistellige natürliche Zahl setzt sich aus der Zehnerziffer a mit $1 \leq a \leq 9$ und der Einerziffer 0 zusammen. Ihre Quersumme ist a. Dividiert man eine solche Zahl durch ihre Quersumme a, so erhält man stets 10. Damit ist die behauptete Teilbarkeit nachgewiesen.

(Mathematik-Olympiaden e. V., 2013, S. 110, H. i. O.)

Abbildung 3: Aufgabe 520631 der Mathematik-Olympiade und dazugehörige Lösung

Die **Aufgabe 520631** wurde im Schuljahr 2012/13 als erste Aufgabe der Landesrunde in der Klassenstufe 6 gestellt. Die drei Aufgabenteile a), b) und c) beschäftigen sich alle mit natürlichen Zahlen, die durch ihre Quersumme teilbar sind. Im ersten Aufgabenteil sollen die Lernenden zwei teilerfremde zweistellige natürliche Zahlen angeben, die jeweils durch ihre Quersumme teilbar sind. Es geht also um den Beweis einer Existenzaussage, wobei anstatt des Existenzquantors der Operator „angeben“ verwendet wurde. In der Lösung sind für den Zahlenraum kleiner 50 alle zweistelligen Zahlen, die durch ihre Quersumme teilbar sind, und alle teilerfremden Paare solcher Zahlen aufgeführt. Dieser Lösungsvorschlag ist bei der Korrektur sehr hilfreich, weil er im Zahlenraum bis 50 ermöglicht, schnell zu beurteilen, ob es sich um ein passendes Paar handelt. Von den Schülerinnen und Schülern wird eine solche Darstellung nicht

verlangt. Die Angabe eines passendes Zahlenpaares und der Nachweis der Bedingungen sind auf jeden Fall ausreichend. Inwieweit bei der Angabe eines Zahlenpaares von den Lernenden erwartet wird, dass sie die gestellten Bedingungen wirklich noch rechnerisch überprüfen, oder dies als offensichtlich akzeptiert wird, bleibt den Korrektorinnen und Korrektoren überlassen. Ähnlich verhält es sich beim Aufgabenteil b). In diesem Aufgabenteil ist zu untersuchen, ob alle durch neun teilbaren zweistelligen natürlichen Zahlen durch ihre Quersumme teilbar sind. Die Lernenden sollen herausfinden, dass diese Allaussage falsch ist und sie mithilfe der Zahl 99 widerlegen. Hier ist also die Negation einer Allaussage, d. h. ebenfalls eine Existenzaussage zu beweisen. Im Lösungsvorschlag wird die gegebene Aussage verneint und gezeigt, dass die Aussage auf alle durch neun teilbaren zweistelligen Zahlen außer auf 99 zutrifft. Wenn Teilnehmende ausschließlich erklären, dass die durch neun teilbare Zahl 99 nicht durch ihre Quersumme 18 teilbar ist, und deshalb die Aussage verneinen, erhalten sie ebenfalls die volle Punktzahl. Im letzten Aufgabenteil c) sollen die Schülerinnen und Schüler dann durch allgemeine Feststellungen begründen, dass alle durch zehn teilbaren zweistelligen Zahlen durch ihre Quersumme teilbar sind. Trotz des etwas weicheren Operators „Begründe“ geht es hier um den Beweis der angegebenen Implikation. In der Lösung wird ausgehend von den Voraussetzungen mittels eines logisch stringenten Argumentgefüges und unter Verwendung einer Variablen die Behauptung deduziert.

Klassenstufe 5, Regionalrunde, Aufgabe 590523

Eine spezielle Rechenmaschine arbeitet nur mit natürlichen Zahlen, und sie ist so programmiert, dass sie bei Eingabe einer Zahl genau die folgenden Operationen hintereinander ausführt und dann das Ergebnis ausgibt:

An die Eingabezahl werden zunächst zwei Nullen gehängt, dann wird die erhaltene Zahl halbiert. Zu diesem Wert wird die Zahl 23 addiert uind danach wird diese Summe mit 3 multipliziert.

a) Welches Ergebnis wird die Rechenmaschine anzeigen, wenn man die Zahl 4 eingibt?

b) Bestimme die Startzahl, die eingegeben wurde, wenn die Maschine das Ergebnis 2019 anzeigt.

c) Begründe, dass alle Ergebnisse der Rechenmaschine auf die Ziffer 9 enden müssen.

(Mathematik-Olympiaden e. V., 2020, S. 29)

Lösung zu 590523c

Teil c) Die Multiplikation der Startzahl mit 100 und die anschließende Halbierung entsprechen einer Multiplikation der Startzahl mit 50. Jede Zahl, die mit 50 multipliziert wird, endet auf eine Null und nach Addition von 23 steht an der letzten Stelle des Zwischenergebnisses eine 3. Da $3 \cdot 3 = 9$ ist, folgt für die letzte Ziffer des Endergebnisses immer eine 9. (Mathematik-Olympiaden e. V., 2020, S. 95-96, H. i. O.)

Abbildung 4: Aufgabe 590523 der Mathematik-Olympiade und dazugehörige Lösung

Die **Aufgabe 590523** wurde im Schuljahr 2019/20 als dritte Aufgabe der Regionalrunde in der Klassenstufe 5 gestellt. In der Aufgabe ist ein Algorithmus angegeben, der zunächst vorwärts und anschließend rückwärts auf einen Spezialfall anzuwenden ist. Im letzten Aufgabenteil c) geht es darum, dass das Ergebnis immer auf die Ziffer neun endet. Die Schülerinnen und Schüler sollen eine Allaussage mit unendlichem Gültigkeitsbereich mittels eines mehrschrittigen Argumentgefüges beweisen. Bei der Entwicklung des Beweises können insbesondere das Invarianzprinzip und das Rechnen modulo zehn helfen.

Die beschriebenen **Aufgaben 520631, 590523** zeigen auf, dass in der Mathematik-Olympiade ab der Klassenstufe 5 Aufgabenteile mit Beweisnotwendigkeit gestellt werden. Die Formulierungen der Aufgabenstellungen wurden offenbar an die Klassenstufe angepasst, beispielsweise kommt der Operator „beweisen“ nicht explizit vor. Die in den vorgestellten Aufgabenteilen zu zeigenden Aussagen beziehen sich teils auf unendlich viele Objekte, teils ausschließlich auf zweistellige Zahlen. Bei beiden Aufgabenteilen c), in denen jeweils eine Allaussage zu beweisen ist, lässt sich primär über Invarianten argumentieren. Die erforderlichen Argumentgefüge sind in allen vorgestellten Aufgabenteilen geringer Komplexität.

Klassenstufe 8, Landesrunde, Aufgabe 510835

Ermittle alle geordneten Tripel $(\mathrm{p}_1,\ \mathrm{p}_2,\ \mathrm{p}_3)$ von Primzahlen p_1, p_2 und p_3, welche die Gleichung $\mathrm{p}_1 \cdot (\mathrm{p}_2 + \mathrm{p}_3) = 1683$ erfüllen und für die $\mathrm{p}_2 < \mathrm{p}_3$ gilt. (Mathematik-Olympiaden e. V., 2012a, S. 56)

Lösung zu 510835

I: Angenommen, es gibt ein Primzahltripel, das die Gleichung $p_1 \cdot (p_2 + p_3) = 1683$ erfüllt. Dann kann wegen $1683 = 3^2 \cdot 11 \cdot 17$ die Primzahl p_1 nur eine der Zahlen 3, 11 oder 17 sein.

Fall 1: Es gilt $p_1 = 3$. Dann gilt $3 \cdot (p_2 + p_3) = 1683$, also $p_2 + p_3 = 561$. Wenn die Summe zweier positiver ganzer Zahlen ungerade ist, dann muss eine der Zahlen gerade und die andere ungerade sein. Da 2 die einzige gerade Primzahl ist und da $p_2 < p_3$ gelten soll, müssen $p_2 = 2$ und $p_3 = 559$ gelten.

Fall 2: Es gilt $p_1 = 11$. Dann gilt $11 \cdot (p_2 + p_3) = 1683$, also $p_2 + p_3 = 153$. Analog wie im Fall 1 müssen dann $p_2 = 2$ und $p_3 = 151$ gelten.

Fall 3: Es gilt $p_1 = 17$. Dann gilt $17 \cdot (p_2 + p_3) = 1683$, also $p_2 + p_3 = 99$. Analog wie im Fall 1 müssen dann $p_2 = 2$ und $p_3 = 97$ gelten.

Folglich gilt: Wenn es ein geordnetes Tripel $(p_1,\ p_2,\ p_3)$ von Primzahlen p_1, p_2 und p_3 gibt, das die Gleichung $p_1 \cdot (p_2 + p_3) = 1683$ erfüllt, wobei $p_2 < p_3$ gilt, dann kann dies nur eines der Tripel $(3, 2, 559)$, $(11, 2, 151)$ oder $(17, 2, 97)$ sein.

II: Wegen $559 = 13 \cdot 43$ ist 559 keine Primzahl. Daher erfüllt das Tripel $(3, 2, 559)$ nicht alle gestellten Bedingungen.

Es gilt $151 < 16 = 13^2$. Wenn 151 keine Primzahl wäre, so müsste 151 durch eine Primzahl kleiner als 13 , also durch eine der Zahlen 2, 3, 5, 7, 11 teilbar sein. Nach den bekannten Teilbarkeitsregeln ist 151 durch keine der Zahlen 2, 3, 5, 11 teilbar. Wegen $151 = 147 + 4 = 7 \cdot 21 + 4$ ist 151 auch nicht durch 7 teilbar. Folglich ist 151 eine Primzahl. Bekanntlich ist 97 eine Primzahl.

Es gelten $11 \cdot (2 + 151) = 11 \cdot 153 = 1683$ und $17 \cdot (2 + 97) = 17 \cdot 99 = 1683$ sowie $2 < 151$ und $2 < 97$. Folglich erfüllen die beiden Tripel $(11, 2, 151)$ und $(17, 2, 97)$ alle gestellten Bedingungen.

Aus I. und II. folgt, dass genau die Zahlentripel $(11, 2, 151)$ und $(17, 2, 97)$ alle gestellten Bedingungen erfüllen.

(Mathematik-Olympiaden e. V., 2012a, S. 172)

Abbildung 5: Aufgabe 510835 der Mathematik-Olympiade und dazugehörig Lösung

Die **Aufgabe 510835** wurde im Schuljahr 2011/12 als fünfte Aufgabe der Landesrunde in der Klassenstufe 8 vorgeschlagen. Die Teilnehmenden sollen alle geordneten Tripel $(\mathrm{p}_1,\ \mathrm{p}_2,\ \mathrm{p}_3)$ von Primzahlen p_1, p_2 und p_3 ermitteln, welche die Gleichung $\mathrm{p}_1 \cdot (\mathrm{p}_2 + \mathrm{p}_3) = 1683$ erfüllen und für die $\mathrm{p}_2 < \mathrm{p}_3$ gilt. Dabei handelt es sich um eine Bestimmungsaufgabe. Es gilt herauszufinden, dass genau die Tripel $(11, 2, 151)$ und $(17, 2, 97)$ alle gestellten Bedingungen erfüllen. Für dieses korrekte Ergebnis erhalten die Lernenden nach den Punktverteilungsvorschlägen des Aufgabenausschusses einen von maximal sieben Punkten. Die restlichen sechs Punkte sind für die korrekte und vollständige Herleitung des Ergebnisses, das heißt für den Beweis dieser Lösung, vorgesehen.

Dazu passend steht auf jedem Aufgabenblatt der Schul-, Regional- und Landesrunden in den Klassenstufen 5 bis 8 folgender **Hinweis zur Aufgabenbearbeitung**:

> Hinweis: Der Lösungsweg mit Begründungen und Nebenrechnungen soll deutlich erkennbar sein. Du musst also auch erklären, wie du zu Ergebnissen und Teilergebnissen gelangt bist. Stelle deinen Lösungsweg logisch korrekt und in grammatisch einwandfreien Sätzen dar. (Aufgabenausschuss des Mathematik-Olympiaden e. V., 2015a, S. 1)

Der Begriff „Beweis“ wird bei diesen jungen Schülerinnen und Schülern im Hinweistext nicht verwendet. Stattdessen wird vom „Lösungsweg mit Begründungen und Nebenrechnungen“ (Aufgabenausschuss des Mathematik-Olympiaden e. V., 2015a, S. 1) gesprochen und zum Erklären des Lösungsweges aufgefordert. Diese Erklärung soll logisch korrekt und in grammatikalisch einwandfreien Sätzen erfolgen.

Auf den Aufgabenblättern der Klassenstufe 8 in der Bundesrunde sowie bei allen Wettbewerbsrunden der Klassenstufen 9 bis 12 heißt es dann:

> Hinweis: Der Lösungsweg mit Begründungen und Nebenrechnungen soll deutlich erkennbar in logisch und grammatisch einwandfreien Sätzen dargestellt werden. Zur Lösungsgewinnung herangezogene Aussagen sind zu beweisen, falls sie nicht aus dem Schulunterricht bekannt sind. Auf eine Beweisangabe kann außerdem verzichtet werden, wenn die Aussage einen eigenen Namen besitzt und dadurch als allgemein bekannt angesehen werden kann. (Aufgabenausschuss des Mathematik-Olympiaden e. V., 2015b, S. 1)

In den höheren Klassenstufen kommt folglich der Beweisbegriff explizit im Hinweis vor. Es wird klargestellt, dass als Wissensbasis nur aus dem Schulunterricht bekannte Aussagen und mathematische Sätze, die einen eigenen Namen besitzen, beispielsweise der kleine Satz von Fermat, benutzt werden dürfen. Alle weiteren verwendeten Aussagen sind zu beweisen.

So legen die drei Aufgaben und die Hinweise zur Aufgabenbearbeitung folgende **Tendenzen** nahe: Die Teilkompetenz des Führens von Beweisen ist zentral für Erfolg im Wettbewerb. Dabei gehen die Anforderungen zum Beweisen in der Klassenstufe 8 über die in den Klassenstufen 5 und 6 hinaus, es scheint eine gewisse Progression der Anforderungen im Laufe der Klassenstufen zu bestehen. Die Anforderungen zum Beweisen überschreiten zumindest in einzelnen Aufgabenteilen die Vorgaben in den Bildungsstandards und den Lehrplänen in NRW (KMK, 2003, 2004; MSW NRW 2007, 2008a; Abschn. 4.2).

2.4 Zusammenfassung des zweiten Kapitels

Die Mathematik-Olympiade

Die Mathematik-Olympiade ist ein 1961 in der DDR eingeführter und seit Mitte der 90er Jahre bundesweit jedes Schuljahr stattfindender **Stufenwettbewerb zum Finden und Fördern mathematisch interessierter und leistungsstarker Schülerinnen und Schüler**. Für alle interessierten Schülerinnen und Schüler der in dieser Arbeit fokussierten Klassenstufen 5 bis 8 startet der Wettbewerb stets im September mit einer Schulrunde. Die jeweils Besten nehmen anschließend an einer Regional-, einer Landes- und ab der Klassenstufe 8 an der Bundesrunde teil. Während die Schulrunde als Hausaufgabenwettbewerb durchgeführt

werden kann, sind die darauffolgenden Runden überall Klausurwettbewerbe. In jeder dieser Runden bearbeiten die Teilnehmenden eigenständig mehrere auf ihre Klassenstufe abgestimmte Aufgaben. Diese Aufgaben, dazugehörige Lösungen und Punktverteilungsvorschläge werden vom bundesweiten Aufgabenausschuss des Mathematik-Olympiaden e. V. entwickelt. Bei der Entwicklung und Zusammenstellung der Aufgaben wird versucht, die verschiedenen curricularen Vorgaben der Bundesländer angemessen zu berücksichtigen. Außerdem wird auf eine Progression der Anforderungen innerhalb des Wettbewerbs, der einzelnen Wettbewerbsrunden und auch innerhalb einzelner Aufgaben geachtet. Die Aufgaben sollen ansprechend und altersgerecht sein sowie innerhalb der gegebenen Zeit ohne Hilfsmittel bearbeitbar sein. Die Bearbeitung der Aufgaben erfordert u. a. gute Kompetenzen im Führen von Beweisen. In allen Klassenstufen 5 bis 8 werden Beweisaufgaben gestellt, bei anderen Aufgabenformaten wird eine logisch korrekte Herleitung der Lösung, das heißt ein Beweis derselben, gefordert. Insbesondere sollen die teilnehmenden Schülerinnen und Schüler ab der Klassenstufe 5 in einzelnen Aufgabenteilen Allaussagen mit unendlichem Gültigkeitsbereich beweisen.

Die vom Aufgabenausschuss entwickelten Aufgaben, Lösungen und Punktverteilungsvorschläge werden in der Bundesrunde stets in unveränderter Form eingesetzt und für alle vorhergehenden Wettbewerbsrunden den Verantwortlichen in den Bundesländern zur Verfügung gestellt. Mit Blick auf die Untersuchung von Beweisversuchen von Wettbewerbsteilnehmenden ist vor allem die **Regionalrunde in Nordrhein-Westfalen** relevant. In NRW gibt es 54 Regionen, für die jeweils ein bis zwei Personen verantwortlich sind. Bei der Durchführung der Regionalrunde gibt es zwischen den Regionen kleine Unterschiede, die die Teilnahmebedingungen, die Aufgabenauswahl, die Klausurbedingungen sowie die Korrektur und Bewertung der Klausurbearbeitungen betreffen. An der Regionalrunde nehmen jedes Jahr mehrere Tausend Schülerinnen und Schüler der Klassenstufen 5 bis 8 teil. Erfolgreiche Teilnehmende dieser und der darauffolgenden Landesrunde werden vielerorts zu mathematischen Enrichment-Angeboten eingeladen. Hinzu kommen zahlreiche weitere Förderangebote, die zum Teil mit dem Wettbewerb verbunden sind.

3 Annäherung an den Beweisbegriff

Die Einblicke in die Anforderungen bei der Mathematik-Olympiade aus Kapitel 2 verdeutlichen, dass im Rahmen dieser Arbeit ein weiter Beweisbegriff[21] sinnvoll ist. Ziel dieses Kapitels ist es, den dieser Arbeit zugrundeliegenden Beweisbegriff sowie eine dazugehörige Terminologie zur Analyse von Beweisversuchen von Schülerinnen und Schülern zu entfalten.

Dafür erfolgt als erster Schritt in Abschnitt 3.1 eine Annäherung an den in der Mathematik geteilten Beweisbegriff. Dieser wird die Grundlage für alle weiteren Überlegungen bilden. An den Beweisbegriff innerhalb der Mathematik anknüpfend kommen aus mathematikdidaktischer Perspektive die Lernenden mit ihren spezifischen Ausgangslagen ins Spiel. In Abschnitt 3.2 wird der in der Mathematik geteilte Beweisbegriff aus der Perspektive der Mathematikdidaktik problematisiert. Innerhalb der Mathematikdidaktik gibt es unterschiedliche Diskussionsstränge und Ansätze, was im Schulunterricht Mathematik unter dem Beweisbegriff gefasst werden sollte und wie Beweisversuche von Lernenden zu beschreiben sind. Unter Berücksichtigung dieser Ansätze und Diskussionsstränge wird in Abschnitt 3.3 ein Analysemodell entwickelt, das mit Blick auf die Forschungsfragen und Datengrundlage dieser Arbeit eine differenzierte und aussagekräftige vergleichende Deskription von Beweisversuchen ermöglicht. Ausgehend vom Beweisbegriff innerhalb der Mathematik, der Problematisierung des Beweisbegriffs aus mathematikdidaktischer Sicht, der Terminologie des Analysemodells und den in Kapitel 2 beschriebenen Charakteristika der Mathematik-Olympiade kann abschließend der Beweisbegriff der vorliegenden Arbeit eingegrenzt werden (Abschn. 3.4).

3.1 Der Beweisbegriff innerhalb der Mathematik

Beweise sind von zentraler Bedeutung innerhalb der Mathematik und ein Charakteristikum mathematischer Arbeit (z. B. Heintz, 2000). Trotzdem gibt es keine eindeutige und allgemein akzeptierte Beschreibung dafür, was ein Beweis ist und wann ein Beweis als solcher akzeptiert wird (Hanna & Jahnke, 1996). Diese fehlende Definition stellt in der Praxis kein großes Problem dar, es existiert ein impliziter Konsens darüber, wie und mit welchen Mitteln ein Beweis zu führen ist und was somit als Beweis anerkannt wird (Davis & Hersh, 1985; Heintz, 2000).

Dieser Konsens basiert u. a. auf einem bestimmten Verständnis des Begriffs „Aussage“ und damit einhergehend einer gewissen Fokussierung dessen, was Mathematikerinnen und Mathematiker beweisen. Zugleich beeinflusst die Art einer infrage stehenden Aussage, was als Beweis derselben ausreicht. Deshalb werden im Folgenden zuerst der Begriff „Aussage“ und verschiedene Aussagetypen beschrieben. Anschließend gibt der Abschnitt 3.1.2 Einblicke in die Frage, wie Mathematikerinnen und Mathematiker auf den Beweisbegriff schauen. Dabei werden wesentliche Merkmale des Beweisbegriffs herausgearbeitet.

[21] In dieser Arbeit wird mit dem Ausdruck „Beweisbegriff“ das geteilte begriffliche Verständnis des Terminus „Beweis“ bezeichnet.

3.1.1 Mathematische Aussagen

Mathematikerinnen und Mathematiker beweisen Aussagen. Als **Aussage** verstehen sie dabei einen „Ausdruck, der entweder wahr oder falsch sein kann“ (Grieser, 2013, S. 136).[22] Auf diesem Verständnis aufbauend können verschiedene Aussagetypen, wie Existenzaussagen und Allaussagen, unterschieden werden.

In einer **Existenzaussage** wird die Existenz eines, mehrerer oder unendlich vieler Elemente bestimmter Art postuliert. Diese Elemente können vieles sein, z. B. Quadratzahlen, Zahlentupel, Funktionen oder geometrische Figuren mit bestimmten Eigenschaften. Es kann sich unter anderem auch um ein Verfahren, z. B. einen Algorithmus, handeln. Wichtig ist dabei der Ausdruck „es gibt“ (bzw. „es existiert“, symbolisch: $\exists$ oder $\bigvee$), der als Existenzquantor bezeichnet wird. Existenzaussagen können auch ohne einen Existenzquantor ausgedrückt werden, lassen sich aber stets unter Verwendung eines Existenzquantors reformulieren. Mit dem Existenzquantor ist in der Mathematik immer gemeint, dass es mindestens die angegebene Anzahl der genannten Elemente gibt. Es interessiert also nicht, ob genau die angegebene Anzahl oder noch mehr Elemente dieser Sorte existieren. In der vorliegenden Arbeit wird zwischen Existenzaussagen, in denen die Existenz einer endlichen Anzahl bestimmter Elemente behauptet wird, und solchen, in denen es um die Existenz unendlich vieler Elemente mit gewissen Eigenschaften geht, differenziert. Erstere werden im Rahmen dieser Arbeit als **spezielle Existenzaussagen** bezeichnet, da sie sich potenziell durch Nachweis an einer endlichen Zahl an Beispielen beweisen lassen. Im Vergleich dazu kann bei Aussagen, in denen die Existenz unendlich vieler Elemente einer bestimmten Art behauptet wird, z. B. eine allgemeine Konstruktion notwendig sein. Diese Aussagen, in denen die Existenz unendlich vieler Elemente einer bestimmten Art behauptet wird, werden im Folgenden als **allgemeine Existenzaussagen** bezeichnet.

Bei **Allaussagen** handelt es sich um Aussagen über alle Elemente einer gewissen Menge, zum Beispiel über alle natürlichen Zahlen, alle Parallelogramme oder alle bijektiven Funktionen. Sie können folglich unter Verwendung des Allquantors („jede/r/s“, „alle“, auch: „für alle“, symbolisch: $\forall$ oder $\bigwedge$) formuliert werden. Die logische Struktur von Allaussagen ist häufig die einer Implikation oder einer Äquivalenz. **Implikationen** sind Aussagen der Form „Wenn A, dann B“ (symbolisch: $\mathrm{A} \Rightarrow \mathrm{B}$, bzw. $(\mathrm{A}_1 \wedge \ldots \wedge \mathrm{A}_\mathrm{j}) \Rightarrow (\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_\mathrm{k}), \mathrm{j}, \mathrm{k} \in \mathbb{Z}^+$). Dabei nennt man A die Prämisse (Voraussetzung, Bedingung) und B die Konklusion. Zwar kann eine Implikation auf verschiedene Weisen ausgedrückt sein, sie lässt sich aber meistens in eine Wenn-Dann-Aussage umformulieren. Im Falle einer Implikation ist folglich z. B. mittels eines direkten Beweises oder einer vollständigen Induktion zu beweisen, dass für alle Elemente, die die Prämisse erfüllen, auch die Konklusion zutrifft bzw. dass wenn für ein (beliebiges) Element die Prämisse gilt, dann auch die Konklusion zutrifft. Bei manchen Implikationen ist ebenfalls die Umkehrung eine wahre Aussage. Dann werden A und B als logisch äquivalent bezeichnet. **Äquivalenzaussagen** sind dementsprechend Aussagen der Art „A gilt genau dann, wenn B gilt“ (symbolisch: $\mathrm{A} \Leftrightarrow \mathrm{B}$, bzw. $(\mathrm{A}_1 \wedge \ldots \wedge \mathrm{A}_\mathrm{j}) \Leftrightarrow (\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_\mathrm{k}), \mathrm{j}, \mathrm{k} \in \mathbb{Z}^+$). Hier ist die Gültigkeit beider Implikationen, $\mathrm{A} \Rightarrow \mathrm{B}$ und $\mathrm{B} \Rightarrow \mathrm{A}$, nachzuweisen.

[22] Dabei wird ein Axiomensystem zugrunde gelegt.

Weiterhin ist es möglich, dass in einer Aussage ein All- und ein Existenzquantor vorkommen. Zwei typische Formulierungen solcher Aussagen lauten: „Es gibt ein $m \in M$, sodass für alle $n \in N$ gilt: …“ und „Für alle $m \in M$ gilt: Es gibt ein $n \in N$, sodass …“, wobei M in der Regel eine Menge unendlicher Mächtigkeit ist. Bei Aussagen, in denen die Existenz und Eindeutigkeit eines Elements behauptet wird, d. h. Aussagen der Art „Es gibt genau ein $m \in M$, für das gilt: …“, lassen sich ebenfalls ein Existenz- und ein Allquantor identifizieren. Um eine solche Aussage zu zeigen, ist es erforderlich, die Existenz des Elements mit den beschriebenen Eigenschaften zu zeigen und zu beweisen, dass alle anderen Elemente der betrachteten Grundmenge nicht alle geforderten Eigenschaften erfüllen. Solche Aussagen, in denen explizit oder implizit Existenz- und Allquantoren kombiniert sind, erscheinen von ihrer logischen Struktur her komplexer als reine Existenzaussagen bzw. Allaussagen. Der Beweis kann i. A. über eine spezielle oder allgemeine Konstruktion, wie es bei Existenzaussagen nahe liegt, oder mithilfe eines Beweisverfahrens, das eher für Allaussagen typisch ist, erfolgen. Teilweise ist eine Kombination verschiedener Beweisverfahren notwendig (Grieser, 2013). Da solche Aussagen zwar bei der Analyse der Wettbewerbsanforderungen zum Beweisen im Rahmen des dieser Arbeit übergeordneten Forschungsprojekts vorkommen, aber nicht primär im Fokus dieser Arbeit stehen, werden sie im Folgenden vereinfachend und zugleich in Abgrenzung zu den anderen Aussagetypen als **Existenz- und Allaussagen** bezeichnet.

Abschließend ist auf einen Zusammenhang zwischen speziellen Existenzaussagen und Allaussagen hinzuweisen. Die Negation einer Allaussage ist eine spezielle Existenzaussage. Eine Allaussage ist falsch, wenn sie auf ein Element der betrachteten Grundmenge nicht zutrifft. Um eine Allaussage zu widerlegen, ist folglich eine spezielle Existenzaussage zu beweisen. Andersherum ist eine spezielle Existenzaussage, in der die Existenz eines Elements mit gewissen Eigenschaften postuliert wird, falsch, wenn es kein Element dieser Art gibt. Die Negation einer solchen speziellen Existenzaussage ist also eine Allaussage. Um eine solche spezielle Existenzaussage zu widerlegen, ist eine Allaussage zu beweisen.

3.1.2 Beschreibungsansätze zum mathematischen Beweisbegriff

Der bereits vielfach verwendete mathematische Beweisbegriff ist ein komplexes Konstrukt und kaum in allen Facetten zu erfassen. Eine erste, humorvolle Annäherung bietet ein Gespräch zwischen einem idealen Mathematiker und einem Philosophiestudenten, das von den Mathematikern Philip J. Davis und Reuben Hersh (1985) verfasst wurde. In dem Gespräch fordert der Student den idealen Mathematiker (I. M.) auf, den Beweisbegriff zu definieren.[23]

> **Student:** Herr Professor, was ist eigentlich ein mathematischer Beweis? […] Was ich von Ihnen wissen möchte, sind aber nicht Beispiele von Beweisen, sondern eine Definition des Beweises.
>
> **I. M.** […] Also man geht folgendermaßen vor: man schreibt die Axiome seiner Theorie in einer formalen Sprache mit Hilfe einer Reihe von Symbolen oder einem Alphabet nieder. Dann schreibt man die Voraussetzungen des Satzes in denselben Symbolen auf. Dann zeigt man, daß sich diese Voraussetzung Schritt um Schritt nach den Regeln der Logik umwandeln läßt, bis man bei der Schlußfolgerung ankommt. Das ist ein Beweis.

[23] Der im Folgenden zitierte Gesprächsausschnitt wurde in der alten Rechtschreibung verfasst. Zur Gewährung der Lesbarkeit wird die aus heutiger Sicht fehlerhafte Verwendung des Eszett nicht gekennzeichnet.

> **Student:** Ach, wirklich? Das ist erstaunlich. Ich habe elementare und fortgeschrittene Differential- und Integralrechnung, elementare Algebra und Topologie gehört und habe noch nie so etwas gesehen.
>
> **I. M.** Natürlich tut das keiner! Es würde ja ewig dauern. Man zeigt einfach, daß man es tun könnte, wenn man wollte. Das reicht.
>
> **Student:** Aber auch das klingt nicht wie das, was in meinen Vorlesungen und Lehrbüchern praktiziert wurde. Das alles läuft darauf hinaus, daß die Mathematiker also doch nichts beweisen.
>
> **I. M.** Selbstverständlich tun wir das. Ein Satz, der nicht bewiesen ist, ist nichts. […]
>
> **Student:** Aber was ist nun ein Beweis wirklich?
>
> **I. M.** Also, das ist ein Argument, das den überzeugt, der sich auf dem Gebiet auskennt.
>
> **Student:** Der sich auf dem Gebiet auskennt? Das heißt aber, daß die Definition des Beweises subjektiv, von bestimmten Personen abhängig ist. Bevor ich entscheiden kann, ob etwas ein Beweis ist, muß ich entscheiden, wer die Experten sind. Was hat das damit zu tun, daß man etwas beweisen möchte?
>
> **I. M.** Aber nein. Mit subjektiv hat das überhaupt nichts zu tun! Jeder weiß doch, was ein Beweis ist. Lesen Sie ein paar Bücher, hören Sie Vorlesungen bei einem tüchtigen Mathematiker, und dann wird Ihnen schon ein Licht aufgehen. (Davis & Hersh, 1985, S. 35-37, H. i. O.)

Um den Studenten zu befriedigen, formuliert der ideale Mathematiker verschiedene Definitionsansätze, die im Verlauf dieses Textes sukzessive aufgegriffen werden. Als ersten Versuch, den Beweisbegriff zu definieren, beschreibt er das **Ideal eines formal-strengen Beweises**[24]. Ein solcher Beweis findet in einem formalen System statt, das auf Symbolen, Axiomen und Regeln der Logik basiert, die rein syntaktische Bedeutung tragen. Ausgehend von diesen Grundlagen und den spezifischen Voraussetzungen der infrage stehenden Behauptung werden neue Aussagen dem Anspruch nach ohne semantische Überlegungen[25] vollständig deduktiv abgeleitet. Bei einer deduktiven Herleitung handelt es sich um notwendiges Schließen. Den Begrifflichkeiten des amerikanischen Philosophen Charles S. Peirce (1839 - 1914) nach wird in einer Deduktion ausgehend von einem Fall und einem gegebenen Gesetz auf ein Resultat geschlossen (Meyer, 2007). Das Gesetz ist dabei eine allgemeine Regel, z. B. eine Implikation, die den betrachteten Fall umfasst. Wenn der Fall als gegeben angenommen wird und das Gesetz korrekt ist, führt eine Deduktion zu einem denknotwendigen, sicheren Resultat. Ein mathematischer Beweis kann idealerweise als eine Kette von Deduktionen dargestellt werden, in der die infrage stehende Behauptung ausgehend von ihren spezifischen Voraussetzungen und Axiomen Schritt für Schritt gefolgert wird. Mittels deduktiver Schlüsse wird also die Gültigkeit von als gültig angenommenen Axiomen auf die infrage stehende Behauptung übertragen. Dabei sind Axiome unbewiesene Aussagen, idealerweise „felsenfeste, offensichtliche Tatsachen“ (Davis & Hersh, 1985, S. 152), die als Setzungen der Menschen aber auch anders hätten ausfallen können (Jahnke, 2007). „Die Sicherheit mathematischer Aussagen hängt [somit] ab von der Sicherheit der Hypothesen (Axiome), die man voraussetzt“ (Jahnke, 2009, S. 27). Neue, bewiesene Aussagen sind folglich nur innerhalb der betrachteten Rahmentheorie gültig. Hanna und Barbeau

[24] Veranschaulichungen des Ideals formal-strenger Beweise sind bei Kempen (2019) sowie bei Reid und Knipping (2010) zu finden.

[25] Es bestehen Zweifel, ob das in der Praxis tatsächlich so praktiziert wird. Bender und Jahnke (1992) schreiben: „[T]he ‘average’ professional mathematician does not work with uninterpreted concepts. His or her language (about sets and functions) is not formalized but rather filled with interpreted concepts and phrases. Furthermore, mathematicians, too, need, whether they are aware of the fact or not, content-related, intuitive interpretations for the production and reception of proofs” (Bender & Jahnke, 1992, S. 261).

(2002) schreiben:

> In such a system [based on definitions and axioms and relying on rules of inference], a proposition is considered true if it can be derived from the axioms in a finite number of logical steps using the permitted rules of inference. (Hanna & Barbeau, 2002, S. 38)

Die zentrale Rolle der verwendeten Grundlagen und die Idee, dass jede (neue) mathematische Aussage ausgehend von den Axiomen der betrachteten Rahmentheorie Schritt für Schritt hergeleitet bzw. auf diese Grundlagen zurückgeführt werden kann, werden auch in den Ausführungen von Beutelspacher (2016) deutlich. Zu der Frage, was man in der Mathematik unter einem Beweis versteht, erklärt er:

> Eine Aussage wird bewiesen, indem sie aus schon bewiesenen Aussagen durch rein logische Argumente abgeleitet wird. [...] Um diese Sätze zu beweisen, darf man wiederum nur Sätze benutzen, die schon vorher bewiesen wurden. Und so weiter. Aber irgendwann ist dieser Rückgriff auf bereits bewiesene Sätze natürlich ausgeschöpft. Irgendwo muss jeder Beweis schließlich verankert werden. Das bedeutet: Man führt alles auf Sätze zurück, die nicht bewiesen werden. Weil sie nicht mehr bewiesen werden können. Diese Sätze nennt man Axiome. (Beutelspacher, 2016, S. 20-21)

Beutelspacher (2016) hebt hervor, dass im Beweis einer neuen Aussage auch bereits bewiesene Aussagen als Gesetze verwendet werden. Diese Aussagen bzw. Sätze könnten ihrerseits aber wiederum auf bereits bewiesene Aussagen zurückgeführt werden. Eine solche fortlaufende Explizierung würde schließlich zu den Grundlagen der betrachteten Rahmentheorie führen, weshalb auch bei einem Rückgriff auf bereits bewiesene Sätze im Kern nur Axiome die Basis der neuen Aussage bilden. In Einklang mit den bisherigen Überlegungen formulieren Jahnke und Ufer (2015) folgende kurze Definition:

> Unter einem mathematischen Beweis versteht man die deduktive Herleitung eines mathematischen Satzes aus Axiomen, Definitionen und zuvor bereits bewiesenen Sätzen nach spezifizierten Schlussregeln. Axiome sind unbewiesene Aussagen, die man an den Anfang einer Theorie stellt. (Jahnke & Ufer, 2015, S. 331)

In dieser Definition wird der zusammenfassende Fachterminus „deduktive Herleitung“ explizit verwendet. Es werden sowohl Axiome und Definitionen als auch zuvor bereits bewiesene Sätze als Grundlagen genannt. Hinsichtlich der Abstraktheit und Formalität der Objekte und Aussagen, mit denen Mathematikerinnen und Mathematiker arbeiten, wird dabei keine Einschränkung vorgenommen.

Zusätzlich zu solchen kurzen Beweisdefinitionen versuchen mehrere Autorinnen und Autoren Ansprüche, die üblicherweise von der mathematischen Gemeinschaft an einen **zur Publikation eingereichten Beweis** gestellt werden, begrifflich zu fassen, auch wenn diese eigentlich nicht vollständig klar sind. Im Ansatz von Wittmann (2009) werden diese Ansprüche mit folgenden Begriffen gefasst:

- *Lückenlosigkeit und Vollständigkeit:* Ein Beweis sollte mit Hilfe der logischen Schlussregeln lückenlos und vollständig darlegen, dass die Behauptung aus den Voraussetzungen sowie den Axiomen und Definitionen und anderen, schon bewiesenen Aussagen folgt.
- *Minimalität:* Ein Beweis sollte sich nur auf diejenigen Voraussetzungen stützen, die für die Gültigkeit der Behauptung unbedingt nötig sind. Ferner sollten die einzelnen Argumentationsschritte keine Redundanzen enthalten.

- *Formalisierung von Struktur, Sprache und Symbolik:* Beweise werden häufig in einer formalisierten Struktur präsentiert. Gleiches gilt für die gepflegte Fachsprache, die sich darüber hinaus durch ein hohes Maß an Präzision auszeichnet (z. B. „es existiert ein" versus „es existiert genau ein"); ferner werden häufig Symbole (etwa für Quantoren) verwendet.

(Wittmann, 2009, S. 36, H. i. O.)

Damit ein Beweis in der mathematischen Gemeinschaft als solcher akzeptiert wird, muss er demnach eine stringente Struktur aufweisen. Das Argumentgefüge sollte keine Lücken, aber auch keine unnötigen Ausführungen oder Redundanzen enthalten. Zudem würden Beweise möglichst strukturiert sowie in einer präzisen Fachsprache und mithilfe von Symbolen verfasst.

Nach Dawkins und Weber (2017) sind Minimalität und eine gewisse formale Abfassung Usus und von der mathematischen Gemeinschaft erwünscht, stellen jedoch keine notwendigen Kriterien für die prinzipielle Anerkennung eines Beweises dar. Dawkins und Weber (2017) haben unter Rückbezug auf andere Publikationen[26] eine Liste mit einigen Werten und Normen veröffentlicht, die von der mathematischen Gemeinschaft mit Blick auf Beweise gehalten würden. Demnach sollen Beweise keine Bezüge zum spezifischen Entstehungskontext, zum Autor oder zu Erfahrungen aufweisen. Sie sollen so geschrieben sein, dass Lesende die Überlegungen selbstständig wiederholen könnten. Deshalb sei es unter anderem wichtig, Begriffe und Konzepte, auf denen ein Beweis beruht, eindeutig zu definieren und offenzulegen. Die Schlüsse müssten deduktiv sein und die Deduktionen so in einer logisch korrekten Reihenfolge angeordnet werden, dass sie jeweils auf Vorhergehendem, aber nicht auf Nachfolgendem aufbauen. Alle Gesetze, die in den Deduktionen verwendet werden, müssten wahrheitsübertragend sein. Weiterhin fände der Entstehungsprozess eines Beweises im „fertigen" Beweis keine Erwähnung und sei i. d. R. ausgehend vom fertigen Produkt nicht mehr nachvollziehbar. Entgegen obiger Anforderungen an einen Beweis ist dessen Entwicklung i. A. ein kreativer, nicht linear verlaufender Prozess, bei dem vielfältige Wahrheitskriterien zusammenspielen.[27] Zusätzlich zur Unabhängigkeit von subjektiven, erfahrungsbasierten und kontextspezifischen Faktoren legten Mathematikerinnen und Mathematiker Wert darauf, zu verstehen, weshalb eine Aussage gilt und wie diese mit den bisher bekannten Aussagen zusammenhängt (Dawkins & Weber, 2017; Hersh, 1993; Thurston 1994). Deshalb gibt es nach Dawkins und Weber (2017) eine

[26] Zu diesen Quellen gehören unter anderem Durand-Guerrier (2008), Halmos (1970), Lai et al. (2012), Mariotti (2006), Rav (1999), Selden und Selden (2003, 2013) und Thurston (1994).

[27] Zur Beschreibung von Beweisprozessen wurden mehrere Prozessmodelle entwickelt. Ein in der Mathematikdidaktik viel zitiertes Modell stammt von Boero (1999). Dieses Modell, das ausgehend von Beobachtungen professioneller Mathematikerinnen und Mathematiker entwickelt wurde, umfasst sechs Schritte mathematischer Erkenntnisgewinnung, die nicht unbedingt in der angegebenen rein linearen Form durchlaufen werden. Die sechs Schritte sind (vereinfacht dargestellt) die Entwicklung einer Vermutung (1), die Formulierung einer Behauptung (2), die Exploration derselben und deren Umfelds (3), die Auswahl von Argumenten und deren Anordnung zu einem Argumentgefüge (4), das Aufschreiben des Beweises gemäß mathematischen Standards (5) und die Annäherung an einen formalen Beweis (6). Mit Blick auf Aufgaben, die einen Beweis einer vorgegebenen Aussage fordern, beschreibt Kirsten (2017, 2018) ein modifiziertes Phasenmodell. Insbesondere ersetzt sie die ersten beiden Phasen des Modells nach Boero (1999) durch die Exploration der vorgegebenen Aussage und die Entwicklung einer mentalen Repräsentation. Sie betont, dass explorative Tätigkeiten und der Aufbau einer individuellen mentalen Repräsentation der zu beweisenden Aussage trotz der Vorgabe dieser Aussage wichtig sind. „Einen sicheren Weg, der zu einem Beweis führt, gibt es [allerdings] nicht" (Davis & Hersh, 1985, S. 153). Gerade auch mit Fokus auf Beweisprozesse von Lernenden ist zu bedenken, dass die Phasen aus Prozessmodellen nicht unbedingt alle durchlaufen werden und tatsächliche Beweisprozesse eventuell mit abgebrochenen, rudimentären, fehlerhaften oder unvollständigen Beweisversuchen enden (u. a. Rott, 2014).

Reihe von Normen, um die Verständlichkeit von Beweisen zu fördern. Diese beeinflussten die Anerkennung eines Beweises jedoch nicht wesentlich. Beweise, die diesen Normen nicht entsprechen, könnten trotzdem als gültig eingestuft werden. Zu diesen Normen gehöre, dass Routineberechnungen und offensichtliche Begründungen weggelassen werden, damit die zentralen Beweisideen und Argumente gut erkennbar seien. Zudem sollten, in Einklang mit den Ausführungen von Wittmann (2009), keine unrelevanten Sachverhalte im Beweis vorkommen, Beweise möglichst klar gegliedert und wichtige Folgerungen bei der Setzung hervorgehoben werden. Bezüglich der Wahl von Symbolen würde darauf geachtet, dass diese mit den gängigen Notationen in Einklang stehen, z. B. x eine Variable bezeichnet, und unterschiedliche Objekte mit unterschiedlichen Symbolen belegt werden.

Diese von Dawkins und Weber (2017) publizierten Werte und Normen werden auch in den Ergebnissen der wissenschaftssoziologischen Studie von Bettina Heintz (2000) deutlich. Sie führt beispielsweise den Anspruch, lückenlos zu argumentieren und sich trotzdem auf Relevantes zu konzentrieren, sowie den Anspruch einer möglichst eindeutigen Darstellungsweise, aber auch die Notwendigkeit der Richtigkeit der Überlegungen mithilfe von Zitaten von Mathematikerinnen und Mathematikern aus:

> *«Das Aufschreiben ist oft mühsam, extrem mühsam. Denn dann muss man wirklich korrekt sein, und es muss hundert Prozent Hand und Fuss* [*sic*] *haben. Man darf auch keinen Satz zuviel* [*sic*] *schreiben, sondern genau, wie man es macht. Es muss kristallklar sein. Und das Resultat muss optimal dargestellt sein. Und das ist manchmal sehr schwierig. Ein Resultat so zu präsentieren, dass man genau das sagt, was man sagen will.»* (Heintz, 2000, S. 172, H. i. O.)
>
> *«In den publizierten Arbeiten sieht man ganz viele Sachen nicht mehr, die ich für mich durchgearbeitet und aufgeschrieben habe. Entweder kommen sie nicht mehr vor, oder ich stecke sie in den Anhang. Aber im Grunde genommen ist das nur technisches Zeug, das das Resultat irgendwie verfälschen würde. Es würde es verdecken. Es schreckt die Leute auch ab. Und darum lasse ich es entweder weg oder tue es in den Anhang.»* Zu diesen *«nicht erwähnenswerten»* Komponenten gehören vor allem auch die […] induktiven und experimentellen Methoden. (Heintz, 2000, S. 229, H. i. O.)

Insgesamt geht es also darum, präzise aufzuzeigen, dass und warum eine Aussage gilt. Damit dies gelingt, werden die Argumente korrekt, logisch stringent, gewissermaßen vollständig, klar strukturiert, präzise und fokussiert dargelegt. Dementsprechend lässt sich zusammenfassend mit Hanna et al. (2012) festhalten:

> Mathematical proof consists, of course, of explicit chains of inference following agreed rules of deduction, and is often characterised by the use of formal notation, syntax and rules of manipulation. Yet clearly, for mathematicians proof is much more than a sequence of correct steps; it is also, and perhaps most importantly, a sequence of ideas and insights with the goal of mathematical understanding – specifically, understanding why a claim is true. (Hanna et al., 2012, S. 444)

Die vom idealen Mathematiker als erster Definitionsversuch beschriebene Vorstellung eines Beweises als vollständige und rein formal-abstrakte deduktive Herleitung stellt sich folglich in der Praxis (zumindest etwas) anders dar. Dementsprechend gibt der ideale Mathematiker auf Nachfrage des Studenten zu: „Natürlich tut das keiner! Es würde ja ewig dauern“ (Davis & Hersh, 1985, S. 36). Bei einer Rückführung neuer Aussagen bis zu den Grundlagen der betreffenden Theorie und einer Explizierung jedes einzelnen Schrittes würden bereits Beweise ein-

facher Aussagen außerordentlich lang werden (u. a. Jahnke & Ufer, 2015).[28] Ein solcher Beweis ist folglich, wenn überhaupt, nur schwer möglich, sehr zeitintensiv, fehleranfällig, kaum nachprüfbar und wenig verständnisfördernd (u. a. Hanna, 1983; Heintz, 2000; Hersh, 1993).[29] Anstatt einer vollständigen Explizierung geht es in der Praxis vielmehr um die **prinzipielle Möglichkeit der Annäherung an einen formal-strengen Beweis**. „Man zeigt einfach, daß [*sic*] man es tun könnte, wenn man wollte. Das reicht!“ (Davis & Hersh, 1985, S. 36).

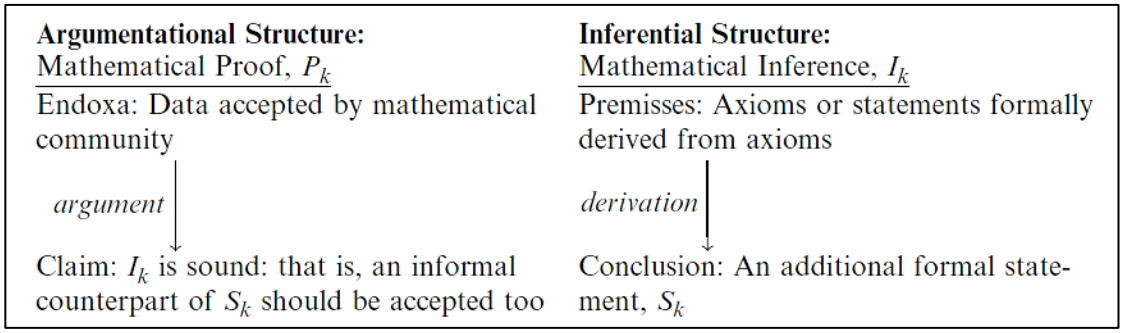


Abbildung 6: Modell der parallelen Struktur mathematischer Beweise von Aberdein (2013, S. 363)

Diese Verbindung zwischen dem Ideal eines formal-strengen Beweises und der gängigen Praxis stellt Andrew Aberdein (2013) in einem Modell mathematischer Beweise dar (Abb. 6). Demzufolge besteht jeder Beweis aus zwei parallelen Strukturen, einer inferential structure und einer argumentational structure. Erstere bezeichnet eine formal-strenge Kette von Deduktionen, in der ausgehend von Axiomen und bereits aus den Axiomen formal hergeleiteten Aussagen die gewünschte Konklusion Schritt für Schritt lückenlos abgeleitet wird. Die argumentational structure umfasst Argumente, mit denen Mathematikerinnen und Mathematiker versuchen, sich gegenseitig von der Existenz einer dazugehörigen inferential structure und damit von der Gültigkeit der informelleren argumentational structure als Beweis der infrage stehenden Aussage zu überzeugen. Die Grundlagen der argumentational structure sind weniger Axiome als von der mathematischen Gemeinschaft als gültig anerkannte Aussagen. Die Schlussweise ist insbesondere in Publikationen deduktiv, wobei aber unter anderem anschauliche Argumente und Rückgriffe auf Diagramme vorkommen können. Außerdem wird nicht jeder einzelne Beweisschritt ausgeführt. Es werden Argumente gebündelt dargestellt und Lücken offengelassen, die die Lesenden ergänzen können. In Einklang mit dieser Vorstellung einer parallelen Struktur mathematischer Beweise schreiben Ufer et al. (2009) kurz:

> Ein Beweis kann […] als ein Medium angesehen werden, das notwendige Informationen zur Rekonstruktion des zugehörigen idealen formalen Beweises in sich trägt und nur in Abhängigkeit vom sozialen Kontext als gültig akzeptiert wird. (Ufer et al., 2009, S. 31)

[28] Jahnke und Ufer (2015) merken dazu an, dass ein vollständig strenger Beweis der Gleichung $1000000 + 1000000 = 2000000$ allein eintausend Seiten benötigt.

[29] Beispielsweise erklärt Hersh (1993): „[…] for many mathematical investigations, full formalization and complete formal proof, even if possible in principle, may be impossible in practice. They require time, patience, and interest beyond the capacity of any human mathematician” (Hersh, 1993, S. 390). Hanna (1983) schreibt: „While a proof is considered a prerequisite for the publication of a theorem, it need be neither rigorous nor complete. Indeed the surveyability of a proof, the holistic conveyance of its ideas in a way that makes them intelligible and convincing, is of much more importance than its formal adequacy. Since fully adequate, step-by-step proof is in most cases impracticable, and since surveyability is lost when proofs become too long, proofs are conventionally elliptical and brief. Much is left to the reader’s judgement” (Hanna, 1983, S. 86).

In der wissenschaftlichen Praxis wird ein Beweis i. d. R. vor der Veröffentlichung insbesondere hinsichtlich seiner Korrektheit und Vollständigkeit geprüft (Heintz, 2000). Die Gruppe an Personen, die eine solche Überprüfung auf angemessene Weise vornehmen kann, ist je nach Themengebiet sowie der Komplexität der Aussage und des dazugehörigen Beweises unterschiedlich groß. Wegen der zunehmenden internen Spezialisierung der Mathematik werden heutzutage neue Beweise teilweise nur von wenigen Mathematikerinnen und Mathematikern verstanden (a. a. O.). Davis (1986) erklärt: „[In] order to evaluate a mathematical paper, you must be part of a certain mathematical subculture. Once you are there, then you will know whether a paper has proved what it claims to prove or it has not“ (Davis, 1986, S. 358). Dementsprechend kommt der ideale Mathematiker auf wiederholte Nachfrage des Studenten zu der Aussage: „Also, das [ein Beweis] ist ein Argument, das den überzeugt, der sich auf dem Gebiet auskennt“ (Davis & Hersh, 1985, S. 37). Und auch der „echte“ Mathematiker Hersh (1993) selbst schreibt:

> In mathematical practice, in the real life of living mathematicians, proof is *a convincing argument, as judged by qualified judges*. (Hersh, 1993, S. 389, H. i. O.)

Bei der Beurteilung eines Beweises durch qualifizierte Expertinnen oder Experten wird nicht immer jeder einzelne Beweisschritt akribisch kontrolliert, sondern zum Teil primär die Stimmigkeit und Kohärenz der zentralen Ideen überprüft (Heinze, 2010; Thurston, 1994). Wie detailliert geprüft wird, hängt u. a. von der Bedeutung des Beweises ab (Heintz, 2000). Bei dieser Prüfung können, so wie bei jeder menschlichen Tätigkeit, Fehler passieren. Die Kriterien, die dabei mehr oder weniger bewusst zugrunde gelegt werden, können sich als Ergebnisse eines Aushandlungs- und Enkulturationsprozesses im Laufe der Zeit verändern und zwischen verschiedenen Teilgebieten der Mathematik unterscheiden (z. B. Heintz, 2000; Reid & Knipping, 2010). Schließlich ist bei der Anwendung dieser Kriterien eine gewisse Subjektivität nicht ausschließbar. Auch nach einer Publikation wird mit einer neuen Aussage i. A. zunächst vorsichtig umgegangen, denn der Geltungsanspruch eines, wenn auch bereits publizierten, Beweises ist nicht unanfechtbar. Oftmals werden neue Aussagen erst nach mehreren Jahren (wirklich) unbedacht verwendet. Reid und Knipping (2010) führen diese Praxis näher aus:

> Once a proof is “accepted,” the results of the proof are regarded as true (with very high probability). It may take generations to detect an error in a proof. If a theorem is widely known and used, its proof frequently studied, if alternate proofs are invented, if it has known applications and generalisations and is analogous to known results in related areas, then it comes to be regarded as “rock bottom.” In this way, of course, all arithmetic and Euclidean geometry are rock bottom. (Reid & Knipping, 2010, S. 21)

Insgesamt betrachtet hängt die Frage, was als Beweis akzeptiert wird, also wesentlich von der mathematischen Gemeinschaft und dem konkreten Handeln der Forscherinnen und Forscher ab. Manin (1977) hebt diese **soziale Dimension** in seiner Beweisdefinition explizit hervor:

> A proof becomes a proof after the social act of accepting it as a proof. (Manin, 1977, S. 48)

Auch die Annahme, dass ein Beweis im Prinzip vollständig formalisiert und expliziert werden kann, wenn man nur lange genug daran arbeitet, wird von manchen Mathematikerinnen und Mathematikern infrage gestellt (z. B. Davis, 1986; Heintz, 2000; Manin, 1977). Grund dafür ist

u. a., dass sich das implizite Wissen, das in jedem Beweis vorausgesetzt wird, niemals vollständig explizieren lässt. Alles mathematische Wissen und damit auch die deduktive Methode an sich ist als sozial konstruiert anzusehen. Jahnke und Ufer (2015) erklären, dass ...

> [...] sich in den einzelnen mathematischen Teilgebieten stillschweigend ein je spezifischer Konsens herausgebildet [hat], wie ein für den Spezialisten akzeptabler und lesbarer Beweis aufgeschrieben sein sollte. Die Regeln dafür sind nirgends kodifiziert, es zeigt sich aber, dass die mathematische Gemeinschaft effizient darin ist, der nachwachsenden Generation zu vermitteln, wie ein Beweis in einem mathematischen Gebiet auszusehen hat. (Jahnke & Ufer, 2015, S. 333)

Heintz (2000) beschreibt, dass der Konsens in der Mathematik im Vergleich zu anderen Wissenschaften groß sei und sich sowohl die mathematische Gemeinschaft als auch der mathematische Korpus trotz aller zentrifugalen Kräfte, wie eine zunehmende Spezialisierung, durch Kohärenz auszeichne. Der ideale Mathematiker bestreitet trotz seiner Aussage, ein Beweis sei ein Argument, das den überzeuge, der sich auf dem Gebiet auskennt, letzten Endes die Personenabhängigkeit des Beweisbegriffs und motiviert den Studenten, sich auf den üblichen mathematischen Enkulturationsprozess einzulassen:

> Aber nein. Mit subjektiv hat das überhaupt nichts zu tun! Jeder weiß doch, was ein Beweis ist. Lesen Sie ein paar Bücher, hören Sie Vorlesungen bei einem tüchtigen Mathematiker, und dann wird Ihnen schon ein Licht aufgehen. (Davis & Hersh, 1985, S. 37)

3.2 Problematisierung des Beweisbegriffs innerhalb der Mathematik für die Arbeit mit Schülerinnen und Schülern

Der beschriebene Beweisbegriff innerhalb der Mathematik (Abschn. 3.1) bildet prinzipiell die Zielperspektive im Rahmen von Beweisaktivitäten mit Schülerinnen und Schülern. Allerdings erscheint es aus mathematikdidaktischer Sicht weder sinnvoll noch möglich, den Beweisbegriff eins zu eins bei der Arbeit mit Schülerinnen und Schülern zu übernehmen. Beispielsweise ist ein rein deduktives Schließen insbesondere von jungen Schülerinnen und Schülern nicht zu erwarten. Insbesondere muss die Praxis der Anerkennung eines Beweisversuchs als Beweis angepasst werden. Daher gibt es in der Mathematikdidaktik eine breite Auseinandersetzung mit dem Beweisbegriff, die sich unter anderem auf die axiomatische Fundierung, die zulässigen Schlussweisen, die Darstellungsmittel und die Anerkennung von Beweisversuchen als Beweis bezieht und noch nicht abgeschlossen ist. In der mathematikdidaktischen Literatur kommen dementsprechend unterschiedliche Definitionen des Beweisbegriffs vor (z.B. Jahnke, 2009; Knipping & Reid, 2015; Stylianides, 2007; Reiss & Ufer, 2009; Wittmann & Müller, 1988). Im Folgenden wird die Definition von Stylianides (2007) vorgestellt, in der verschiedene Aspekte des Beweisbegriffs innerhalb der Mathematik relativierend auf die Arbeit mit Schülerinnen und Schülern übertragen werden. Stylianides (2007) betont, dass die Definition für die Arbeit mit Schülerinnen und Schülern eines großen Spektrums an Klassenstufen geeignet ist, was zur Anlage der vorliegenden Arbeit passt (Abschn. 1.1, 6.1, 10.2). Mehrere der Anregungen von Stylianides (2007) sind zudem bei der nachfolgenden Entwicklung eines Modells für die Analyse von Beweisversuchen zu Allaussagen von Bedeutung (Abschn. 3.3). Auch wird der Beweisbegriff dieser Arbeit unter Rückbezug auf seine Definition eingegrenzt (Abschn. 3.4). Die Beweis-

definition von Stylianides (2007) lautet:

> *Proof* is a mathematical argument, a connected sequence of assertions for or against a mathematical claim, with the following characteristics:
>
> 1. It uses statements accepted by the classroom community (*set of accepted statements*) that are true and available without further justification;
> 2. It employs forms of reasoning (*modes of argumentation*) that are valid and known to, or within the conceptual reach of, the classroom community; and
> 3. It is communicated with forms of expression (*modes of argument representation*) that are appropriate and known to, or within the conceptual reach of, the classroom community.
>
> (Stylianides, 2007, S. 291, H. i. O.)

Einen Beweis definiert Stylianides (2007) als ein mathematisches Argument[30] mit drei besonderen Charakteristika. Das erste von ihm angeführte Charakteristikum betrifft die Aussagen, die in einem Beweis als gültig vorausgesetzt werden. Im Unterschied zur Wissenschaft Mathematik findet in der Schule kein axiomatischer Aufbau der Mathematik statt. Daher bezieht sich Stylianides auf das lokale Ordnen, also die deduktive Strukturierung eines Teilbereichs der Mathematik. Als Argumentationsgrundlage dienen dabei Aussagen, die im konkreten Unterrichtssetting als gültig angenommen werden. Dieses Vorgehen impliziert jedoch keine Beliebigkeit. Stylianides (2007) erklärt:

> The terms "true", "valid", and "appropriate" in the descriptions of the set of accepted statements, modes of argumentation, and modes of argument representation, respectively, should be understood in the context of what is typically agreed upon nowadays in the field of mathematics. (Stylianides, 2007, S. 292-293)

Auch beim Beweisen im schulischen Kontext sollten die zugrunde gelegten Aussagen aus fachlich-normativer Sicht gültig sein.

Das zweite angeführte Charakteristikum von Beweisen bezieht sich auf die zulässigen Beweisverfahren, wie direkter Beweis, Beweis durch Widerspruch und vollständige Induktion, sowie die erlaubten Schlussweisen. Während in der Fachwissenschaft nur deduktive Schlüsse erlaubt sind, beschreiben Mathematikdidaktikerinnen und -didaktiker Vorformen dieser Schlussweise, die auf einer anschaulicheren Erkenntnisebene operieren und trotzdem die Allgemeingültigkeit einer Aussage begründen können (Abschn. 3.3.2). Von solchen Schlussweisen abgrenzend betont Stylianides (2007), dass induktive Prüfungen auch beim schulischen Lernen nicht als Beweis angesehen werden sollten:

> The definition excludes the possibility of empirical arguments being considered as proofs at any level of schooling, where by *empirical arguments* I mean arguments based on the use of examples that offer confirming, yet incomplete, evidence that a mathematical claim is true. (Stylianides, 2007, S. 298, H. i. O.)

Induktionen und weitere aus fachlich-normativer Perspektive ungültige Schlussweisen sind trotzdem für die Deskription von Beweisversuchen von Schülerinnen und Schülern von Bedeutung, da es vorkommen kann, dass sie solche Schlussweisen benutzen (z. B. Abschn. 3.3.2).

[30] Die Begriffe „Argument" und „Argumentation" sowie deren Beziehungen zu Beweisen und Beweisprozessen werden in der Mathematikdidaktik diskutiert. Da diese Diskussion nicht zum inhaltlichen Verständnis der vorliegenden Arbeit beiträgt, wird an dieser Stelle auf einige Quellen verwiesen. Zusammenfassungen unterschiedlicher Sichtweisen sind u. a. bei Reid und Knipping (2010) und bei Durand-Guerrier et al. (2012) zu finden. Tiefere Einblicke in einzelne Standpunkte geben z. B. Balacheff (1991), Douek (1998), Duval (1995) und Pedemonte (2007).

Der dritte Punkt in der Definition von Stylianides (2007) bezieht sich auf die Darstellungsweise von Beweisen. Innerhalb der Mathematik werden Beweise vor der Veröffentlichung normalerweise zu einem gewissen Grad formal abgefasst (Abschn. 3.1.2). Allerdings sind junge Schülerinnen und Schüler mit den dabei erforderlichen Darstellungsmitteln noch nicht vertraut und auch viele fortgeschrittene Lernende scheinen Schwierigkeiten mit der formalsprachlichen Abfassung eigener Beweisversuche und dem Nachvollziehen formal dargestellter Beweise zu haben (Abschn. 3.3.11, 5.3). Nicht rein formalsprachlich dargestellte Beweise werden zudem als einsichtiger und einprägsamer angesehen, geht es beim Beweisen im Schulkontext doch vor allem darum, zu begründen, warum eine Aussage gilt, und das Lernen durch Verstehen und die Vernetzung von Wissen zu fördern (z. B. Hanna, 1989; Kautschitsch, 1989; Malle, 2002; Wittmann, 2009). Da zudem die Darstellungsweise eines Beweises dessen prinzipielle Gültigkeit allenfalls geringfügig beeinflusst, ist der Einbezug anderer Darstellungsmittel ein probater Ansatzpunkt zur Anpassung des Beweisbegriffs für die schulische Arbeit. Die von Stylianides (2007) angeführten „modes of argument representation“ (Stylianides, 2007, S. 291) umfassen u. a. ikonische und enaktive Darbietungen, präformale Ausdrucksweisen und die Verwendung von Hilfsmitteln, wie Tabellen oder konkrete Anschauungsobjekte.

Bei der Beschreibung aller drei Charakteristika von Beweisen bezieht sich Stylianides (2007) auf die Klassengemeinschaft. Während in der wissenschaftlichen Praxis die mathematische Community den Beweisbegriff prägt und über die Gültigkeit von Beweisversuchen entscheidet, übernimmt diese Rolle in der Schule je nach Unterrichtspraxis die Lehrkraft oder die Klassengemeinschaft. Auch Schülerinnen und Schüler durchlaufen in ihrer Schulzeit einen Enkulturationsprozess, bei dem sich die Beweiskultur in Abhängigkeit ihrer kognitiven Entwicklung, der curricularen Vorgaben sowie der Lehrkraft und Klassengemeinschaft verändert.

Darüber hinaus ist hervorzuheben, dass Styliandes (2007) in den Begriff des Beweises mathematische Argumentgefüge, die eine Aussage widerlegen, mit einbezieht. Dies erscheint in Anbetracht der Beziehung zwischen dem Beweisen und Widerlegen von Aussagen plausibel und wird deshalb auch in dieser Arbeit so gehandhabt. Insbesondere werden Widerlegungen von Allaussagen als Beweise von Existenzaussagen angesehen (Abschn. 3.1.1, 3.4).

3.3 Ein Modell zur Analyse von Beweisversuchen zu Allaussagen

Im vorhergehenden Abschnitt wurden Einblicke in die mathematikdidaktische Diskussion zum Beweisbegriff gegeben. Diese Überlegungen sowie bereits vorhandene Ansätze zur Beschreibung von Beweisversuchen aus der mathematikdidaktischen Literatur werden im Folgenden für die Analyse von Beweisversuchen aufgegriffen. Sie bilden eine wesentliche Grundlage eines Modells zur Analyse von Beweisversuchen von Allaussagen, das im Rahmen einer qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) entwickelt wurde (Kap. 7). Ziel der Entwicklung dieses Modells war es, eine zu den Forschungsfragen und den zu untersuchenden Beweisversuchen passende, präzise, trennscharfe und systematisch differenzierte Analyse zu ermöglichen. Im vorliegenden Abschnitt wird dieses Modell entfaltet.

Um die Leserinnen und Leser zu orientieren, wird in Abschnitt 3.3.1 zunächst die Grundstruktur des Modells skizziert. Danach werden die verschiedenen Elemente des Modells sukzessive erläutert (Abschn. 3.3.2 – 3.3.11). Dabei ist jeder der Abschnitte in drei Teile, die Entwicklung der Kategorien, die Vorstellung der Kategorien und Hinweise zu den Kategorien, gegliedert.

Es wird jeweils zuerst die Entwicklung des fokussierten Elements des Analysemodells thematisiert. Hier werden die bei der Entwicklung leitenden Überlegungen und die berücksichtigten mathematikdidaktischen Konzepte präsentiert. Von den mathematikdidaktischen Konzepten werden im Sinne einer kompakten und stringenten Darstellung stets nur die relevanten Aspekte vorgestellt. Zugleich können unterschiedliche Elemente des Modells ähnliche theoretische Grundlagen erfordern, so dass es zu Wiederholungen oder Querverweisen kommt.

Anknüpfend an die theoretischen Grundlagen und damit einhergehenden Überlegungen zur Entwicklung der Begrifflichkeiten des Modells werden in jedem Abschnitt 3.3.2 bis 3.3.11 die im Analysemodell unterschiedenen Kategorien definiert. Im Sinne einer guten Lesbarkeit werden bei den Definitionen nicht immer die kleinen Details ausgeführt, aber stets alle wesentlichen Informationen präsentiert.[31] Ebenfalls im Sinne der Verständlichkeit sowie um einen Einblick zu geben, wie die Begrifflichkeiten des Analysemodells dann wirksam werden, sind bei den Kategoriendefinitionen jeweils Beweisversuche von Schülerinnen und Schülern angeführt. Die Beweisversuche, die im Folgenden zur Veranschaulichung dienen, basieren auf der nachfolgenden Aufgabenstellung. Diese Aufgabe wurde in der im Rahmen der vorliegenden Arbeit durchgeführten Studie mit Teilnehmenden der Regionalrunde der Mathematik-Olympiade 2016/17 in der siebten und achten Klassenstufe gestellt (Abschn. 6.2).

> Ali, Björn, Claudia und Denise experimentieren mit natürlichen Zahlen einschließlich der Null. […]
>
> d) Denise berechnet die Summe von vier aufeinanderfolgenden ungeraden Zahlen (z. B. $5 + 7 + 9 + 11 = 32$). Sie behauptet: „Die Summe von **vier <u>aufeinanderfolgenden</u> ungeraden** Zahlen ist immer durch 8 teilbar.“ Hilf Denise zu beweisen, dass ihre Behauptung <u>richtig</u> ist.

Am Ende jedes Abschnitts 3.3.2 bis 3.3.11 sind Hinweise zu den Kategorien zu finden, die sich je nach inhaltlicher Relevanz auf die Einordnung der Kategorien im Forschungsfeld, auf die Nummerierung und Abgrenzung der Kategorien oder auf deren Interpretation beziehen.

Der vorliegende Abschnitt 3.3 verknüpft also theoretische Grundlagen, eigene Theoriebildung und Einblicke in die im Rahmen dieser Arbeit durchgeführte empirische Studie. Auf diese Weise ist es möglich, umfangreiche theoretische Grundlagen fokussiert darzustellen und die erfolgte Verbindung theoretischer Grundlagen und eigener Theoriebildung offenzulegen. Zudem ist das Analysemodell inklusive der eigenen Begrifflichkeiten erforderlich, um präzise zu beschreiben, was unter dem Begriff „Beweis“ in der vorliegenden Arbeit verstanden wird (Abschn. 3.4). Deshalb wird in diesem Abschnitt 3.3 ein Stück weit auf den praktischen Teil dieser Arbeit und dessen Ergebnisse vorgegriffen. Die Aufgabenbearbeitungen, die zur Veranschaulichung gebildeter Kategorien dienen, geben Einblicke in die Vielfalt der Beweisversuche von Teilnehmenden der Mathematik-Olympiade. Alle weiteren Ergebnisse sind in Kapitel 8 zu finden.

[31] Weitere Details können gerne angefragt werden.

3.3.1 Skizze des Modells

Um eine erste Orientierung zu geben, zeigt die Abbildung 7 die Grundstruktur des im Rahmen dieser Arbeit entwickelten Modells zur Analyse in schriftlicher Form vorliegender Beweisversuche zu Allaussagen. Demnach wird zwischen Beweisversuchen mit begründendem Ansatz, prüfendem Ansatz, weder begründendem noch prüfendem Ansatz und nicht eindeutig kategorisierbarem Ansatz sowie fehlenden Bearbeitungen[32] differenziert.

Danach werden bei Beweisversuchen mit begründendem Ansatz sowie bei denen mit prüfendem Ansatz unter anderem die Korrektheit, Beweisideen, Beispiele, die Schlussweise und die Darstellungsform analysiert. Bei Beweisversuchen mit weder begründendem noch prüfendem Ansatz werden insbesondere verschiedene Arten der Bearbeitung differenziert, wobei die Korrektheit der Beweisversuche, zugrundeliegende Ideen und die Schlussweise Berücksichtigung finden. Auch Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz werden, soweit möglich, näher beschrieben.

Auf diese Weise dient das Analysemodell als Rahmen für eine systematisch differenzierte Deskription von Beweisversuchen, der ermöglicht, die Besonderheiten von Beweisversuchen mit begründendem, prüfendem, weder begründendem noch prüfendem bzw. nicht eindeutig kategorisierbarem Ansatz zu berücksichtigen und zugleich Aussagen über alle zu analysierenden Beweisversuche zu treffen. Im Folgenden wird das Analysemodell sukzessive näher erläutert, ausdifferenziert und anhand von Beweisversuchen von Lernenden illustriert (Abschn. 3.3.2 – 3.3.11).

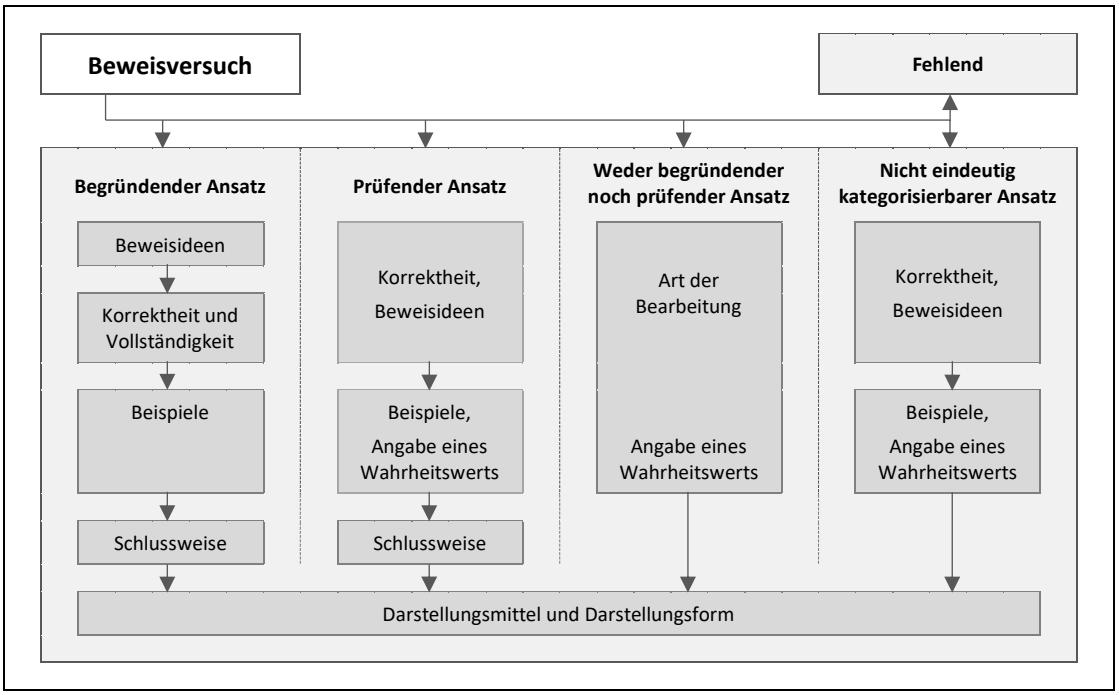


Abbildung 7: Grundstruktur des Modells zur Analyse von Beweisversuchen zu Allaussagen

[32] Im Folgenden werden die Begriffe „Beweisversuch“ und „Bearbeitung“ als Synonyme verwendet, um in schriftlicher Form vorliegende Bearbeitungen von Beweisaufgaben zu bezeichnen.

3.3.2 Begründender, prüfender oder sonstiger Ansatz

In dem skizzierten, für diese Arbeit entwickelten Analysemodell wird zunächst zwischen Beweisversuchen mit begründendem, prüfendem, weder begründendem noch prüfendem oder nicht eindeutig kategorisierbarem Ansatz sowie fehlenden Bearbeitungen differenziert (Abb. 8). Mit dieser Kategorisierung wird überblicksartig erfasst, inwiefern ein Beweisversuch zu der zu zeigenden Aussage abgegeben wurde sowie inwieweit Beispiele, ein Argumentgefüge und Überlegungen zur Allgemeingültigkeit der zu zeigenden Aussage erkennbar sind.

Im vorliegenden Abschnitt geht es darum, diese Kategorisierung einzuführen. Es wird dargelegt, wie die angegebene Unterteilung in die fünf Oberkategorien entstanden ist. Dafür werden der Entwicklungsprozess auf Basis der herangezogenen theoretischen Grundlagen nachgezeichnet und die bei der Entwicklung leitenden Überlegungen ausgeführt. Leserinnen und Leser, die sich mit den theoretischen Grundlagen und dem Entwicklungsprozess erst einmal nicht auseinandersetzen möchten, können direkt zur anschließenden Vorstellung der Kategorien übergehen. Hier werden die fünf Kategorien nacheinander definiert und anhand von Bearbeitungen von Schülerinnen und Schülern illustriert. Abschließend folgen ergänzende Hinweise zur Codierung von Beweisversuchen bezüglich dieser Kategorien und zur Interpretation vorgenommener Zuordnungen. Die Ergebnisse der Anwendung der Kategorien im Rahmen der eigenen Studie werden in Abschnitt 8.1 beschrieben.

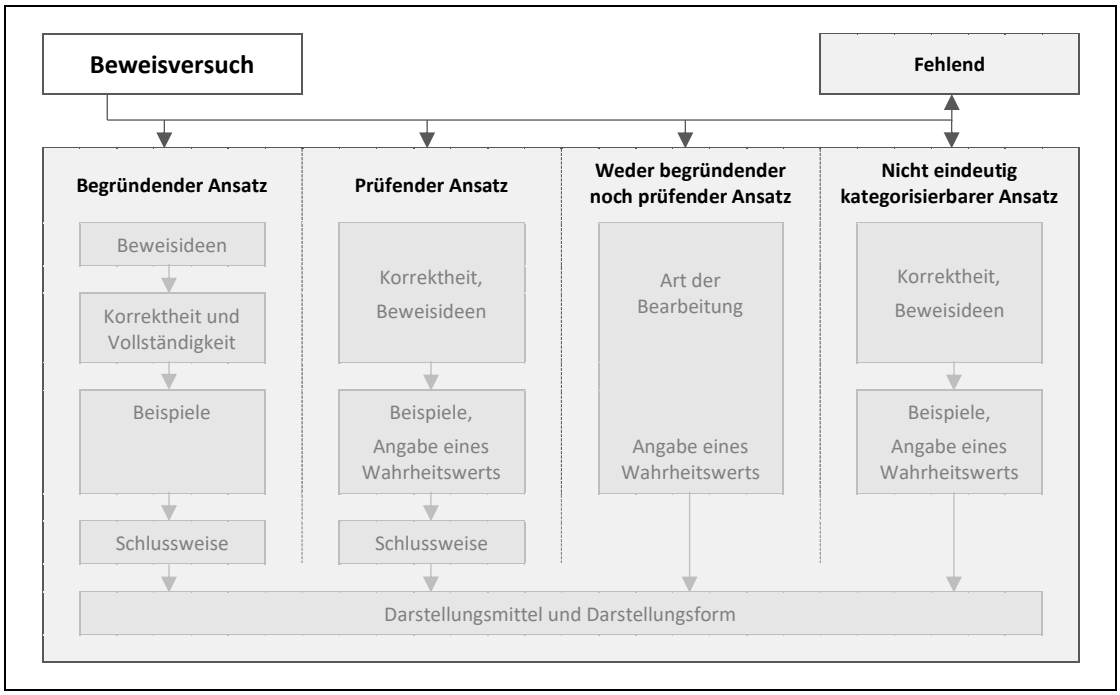


Abbildung 8: Analysemodell. Fokus: Begründender, prüfender oder sonstiges Ansatz

3.3.2.1 Entwicklung der Kategorien

Den Ausgangspunkt für die Entwicklung der Unterscheidung von Beweisversuchen mit begründendem, prüfendem und weder begründendem noch prüfendem Ansatz bildet das **Konzept der Argumentationsbasis** von Fischer und Malle (1985) sowie Malle (2002). Dieses wird als

erster mathematikdidaktischer Zugang ausgewählt, da es sich durch die Orientierung am Lernstand der Kinder besonders für die Analyse eines großen Spektrums an Beweisversuchen von Schülerinnen und Schülern eignet. Nach diesem Ansatz „bedarf [jeder Beweisversuch] einer Argumentationsbasis, d. h. eines Fundaments, auf das man sich bei seiner Argumentation stützt“ (Malle, 2002, S. 4). Die Autoren erklären: „Eine Menge von Aussagen, die als richtig angesehen werden, soll zusammen mit den Schlußweisen [*sic*], die als zulässig anerkannt werden, als Argumentationsbasis bezeichnet werden“ (Fischer & Malle, 1985, S. 180). Eine Argumentationsbasis besteht demnach aus einer Wissensbasis im Sinne einer „Menge an Aussagen, die als richtig angesehen werden“ (Fischer & Malle, 1985, S. 180) und einer Schlussweise. Zur Schlussweise schreiben Fischer und Malle (1985):

> Die Richtigkeit einer Aussage α, die etwa von einer Person A gemacht wird, kann einer Person B (wobei $\mathrm{A} = \mathrm{B}$ möglich ist) unmittelbar klar erscheinen […]. Sollte B die Aussage α bezweifeln, so kann A auf verschiedene Weisen Gründe für deren Gültigkeit anführen, z. B.:
>
> (a) Bestätigung der Richtigkeit durch einen glaubwürdigen Zeugen, etwa durch einen Text in einem entsprechenden Fachbuch (Berufung auf eine Autorität).
>
> (b) Anführen von Aussagen, die von B als richtig angesehen werden und deren Richtigkeit hinreichend für die Richtigkeit von [einer Aussage] α sind (deduktives Schließen).
>
> (c) Anführen von Folgerungen aus α, die von B als richtig angesehen werden, deren Richtigkeit aber nicht hinreichend für die Richtigkeit von α ist (reduktives Schließen). Ein Spezialfall liegt vor, wenn aus der Verifizierung von α an einzelnen Elementen einer Menge auf die Gültigkeit von α für alle Elemente dieser Menge geschlossen wird (induktives Schließen).
>
> (d) Anführen von Aussagen, die von B als richtig angesehen werden und deren Richtigkeit in einem gewissen, nicht deduktiven Zusammenhang mit der Richtigkeit von α steht (Analogieschlüsse, Wahrscheinlichkeitsaussagen usw.). (Fischer & Malle, 1985, S. 178-179, H. i. O.)

Beweisversuche können somit auf unterschiedlichen Schlussweisen beruhen. Während in der Wissenschaft Mathematik lediglich das deduktive Schließen als gültig erachtet wird, ist die Schlussweise in Beweisversuchen von Schülerinnen und Schülern Studien zufolge häufig induktiv (Abschn. 3.1.2, 5.1.1). Außerdem kann es vorkommen, dass Lernende weder deduktiv noch induktiv argumentieren, sich z. B. auf eine Autorität berufen.

Mit Blick auf eine erste grobe Kategorisierung von Beweisversuchen im Rahmen dieser Arbeit legen diese Überlegungen eine Unterscheidung zwischen Bearbeitungen mit deduktiver, induktiver und weder deduktiver noch induktiver Schlussweise nahe. Allerdings könnten bei dieser Dreiteilung Beweisversuche von Schülerinnen und Schülern nicht eindeutig eingeordnet werden, wenn die Schlussweise deduktive und induktive Anteile aufweist. Es sind folglich weiter gefasste Kategorienbezeichnungen als die Begriffe „deduktiv“ und „induktiv“ erforderlich, um solche Beweisversuche einordnen zu können. Im Rahmen dieser Arbeit werden für diese weiten Kategorien die Bezeichnungen „begründender Ansatz“ und „prüfender Ansatz“ gewählt. Dabei zielen Beweisversuche mit begründendem Ansatz auf die Allgemeingültigkeit einer zu zeigenden Aussage, während dies bei Beweisversuchen mit prüfendem Ansatz nicht expliziert ist. Für eine genaue begriffliche Klärung und Einordnung werden im Folgenden zentrale mathematikdidaktische Konzepte zu **Vorformen wissenschaftlicher Beweise** fokussiert (Abb. 9).

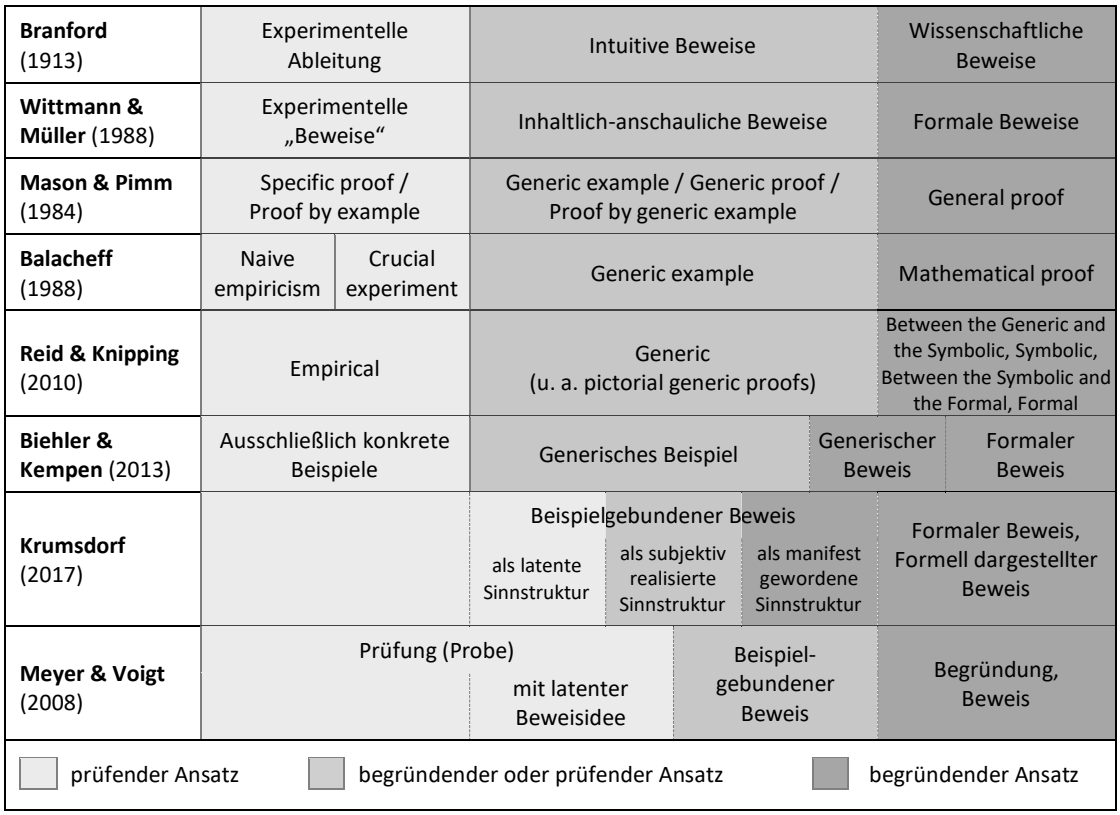


Abbildung 9: Vereinfachte Übersicht über relevante Aspekte berücksichtigter mathematikdidaktischer Konzepte sowie Zuordnung von Beweisversuchen mit begründendem bzw. prüfendem Ansatz zu den Begrifflichkeiten dieser Konzepte

Bereits Anfang des zwanzigsten Jahrhunderts folgerte Branford (1913) aus der Unerreichbarkeit des Ideals formal-strenger Beweise für die Pädagogik, „daß [*sic*] die Strenge oder der Grad der wissenschaftlichen Beweisstufe der Reife des Schülers angemessen sein muß [*sic*]“ (Branford, 1913, S. 107). Auf dieser Überlegung aufbauend hat er drei Stufen des Beweisverfahrens angeführt, mit dem die „in Rede stehende Wahrheit und die wachsende Erkenntnis derselben dem Schüler am besten übermittelt werden kann“ (Branford, 1913, S. 100): Die Stufe der experimentellen Ableitung, die Stufe der intuitiven Ableitung und der wissenschaftliche Beweis. Erstere bezeichnet die Überprüfung konkreter Fälle, die die Allgemeingültigkeit von Allaussagen nicht garantieren kann. Bei der Stufe der intuitiven Ableitung und beim wissenschaftlichen Beweisen würde hingegen der Frage nachgegangen, ob eine Aussage in allen Fällen richtig sei. Sowohl mit einem wissenschaftlichen Beweis als auch mit einem intuitiven Beweis könnten demnach universelle Wahrheiten aufgestellt werden. Als intuitiven Beweis beschreibt er exemplarisch, wie Lernende ausgehend von Erfahrungen mit zwei zusammengehefteten, sich kreuzenden Stäben Einsicht in die Allgemeingültigkeit des Scheitelwinkelsatzes gewinnen können. Solche intuitiven Beweise seien „eine mehr vorläufige und weniger strenge Art des idealen wissenschaftlichen Beweises“ (Branford, 1913, S. 108). Eine klare Grenzlinie zwischen intuitiven und wissenschaftlichen Beweisen gäbe es aber nicht (Branford, 1913).

In den vergangenen Jahrzehnten wurden insbesondere, aber nicht nur mit Blick auf junge Schülerinnen und Schüler weit über zwanzig solcher Vorformen wissenschaftlicher Beweise inner-

halb der Mathematik beschrieben (Krumsdorf, 2017).[33] Sie sollen eine schuladäquate Zwischenstufe zwischen empirischen Verifikationen und dem Beweisbegriff innerhalb der Mathematik bilden, die es Lernenden ermöglicht, über die Begrenztheit rein empirischer Argumente hinauszugehen und den allgemeingültigen mathematischen Beweis in einer für sie verständlichen Form zu erfahren (Biehler & Kempen, 2016). Die Grundzüge dieser Vorformen sind meistens ähnlich. So wie die beschriebenen intuitiven Beweise werden sie oftmals an Beispielen durchgeführt. Diese stehen jedoch nicht für sich selbst, sondern deuten darüber hinaus. Die an Beispielen durchgeführten Operationen können auf die gesamte betrachtete Grundmenge verallgemeinert werden. Je nach Ansatz werden die Überlegungen zusätzlich in allgemein gehaltener Sprache mehr oder weniger ausformuliert.

Wittmann und Müller (1988) beispielsweise greifen die von Branford (1913) beschriebene Dreistufigkeit sowie dessen Trennung von empirischen Verifikationen und allgemeingültigen oder verallgemeinerbaren Argumentgefügen explizit auf, indem sie zwischen experimentellen, inhaltlich-anschaulichen und formalen Beweisen unterscheiden. Dabei stützten sich inhaltlich-anschauliche Beweise „auf Konstruktionen und Operationen, von denen intuitiv erkennbar ist, dass sie sich auf eine ganze Klasse von Beispielen anwenden lassen und bestimmte Folgerungen nach sich ziehen“ (Wittmann & Müller, 1988, S. 249). Analog zu Branford (1913) grenzen sie diese von experimentellen Beweisen ab, die in der Verifikation einer endlichen Anzahl von Beispielen bestünden. Zwischen inhaltlich-anschaulichen und formalen Beweisen gäbe es hingegen keine genaue Grenzlinie, sie würden sich nur im Ausmaß der logischen Strenge unterscheiden (Wittmann und Müller, 1988).

Mason und Pimm (1984) verwenden die Begriffe „generic example“ (Mason & Pimm, 1984, S. 284) und „generic proof“ (Mason & Pimm, 1984, S. 284). Sie schreiben: „A generic example is an actual example, but one presented in such a way as to bring out its intended role as the carrier of the general. […] The generic proof, although given in terms of a particular number, nowhere relies on any specific properties of that number” (Mason & Pimm, 1984, S. 284). Eine explizite Unterscheidung der verwendeten Begriffe „generic example“ und „generic proof“ wird von den Autoren nicht vorgenommen. So wie Branford (1913) und Wittmann und Müller (1988) differenzieren sie vielmehr zwischen „specific, generic and general“ (Mason & Pimm, 1984, S. 284) sowie an späterer Stelle zwischen „proof by example“ (Mason & Pimm, 1984, S. 286) und „proof by generic example“ (Mason & Pimm, 1984, S. 286). Dabei weisen sie darauf hin, dass die Abgrenzung zwischen specific und generic examples in der Praxis nicht unbedingt so klar sei, wie sie aus theoretischer Perspektive erscheinen möge. Es sei insbesondere möglich, dass Lernende ein generic example ausschließlich als specific example ansehen. Sich bewusst zu machen, wofür ein Beispiel exemplarisch steht, die vom Autor bzw. von der Autorin

[33] Eine Erläuterung aller dieser Konzepte würde den Rahmen dieser Arbeit überschreiten und ist für die Entwicklung und Erläuterung des eigenen Analysemodells nicht erforderlich. Interessierte Leserinnen und Leser finden bei Krumsdorf (2017) kurze Beschreibungen einer Vielzahl an Ansätzen, die er unter dem Begriff „beispielgebundener Beweis“ subsummiert. Biehler und Kempen (2016) haben vor allem die Konzepte aus der deutschsprachigen Mathematikdidaktik sehr sorgfältig aufgearbeitet. Reid und Knipping (2010) vergleichen zahlreiche international bekannte Ansätze vor dem Hintergrund eines eigenen Begriffsnetzes. Darüber hinaus ist es sinnvoll, in den Originalquellen nachzulesen. Manche können dem Literaturverzeichnis dieser Arbeit entnommen werden (u. a. Bell, 1976; Wittmann & Müller, 1988).

intendierten verallgemeinerbaren Schlüsse zu abstrahieren und zu erkennen, ob diese ohne Einschränkung auf den allgemeinen Fall übertragen werden können, sei anspruchsvoll und von Lernenden nicht unbedingt zu erwarten.

Balacheff (1988) verwendet ebenfalls den Begriff „generic example“ (Balacheff, 1988, S. 219) und definiert ihn wie folgt:

> The generic example involves making explicit the reasons for the truth of an assertion by means of operations or transformations on an object that is not there in its own right, but as a characteristic representative of its class. The account involves the characteristic properties and structures of a class, while doing so in terms of the names and illustration of one of its representatives. (Balacheff, 1988, S. 219)

In Einklang mit der Definition von Mason und Pimm (1984) handelt es sich beim generic example laut Balacheff (1988) um ein Beispiel, das für etwas Allgemeines steht. Dabei betont er, dass dieses Beispiel alle relevanten Eigenschaften der Gesamtheit der Fälle, auf die sich die infrage stehende Aussage bezieht, aufweist. Die im Beispiel erkennbaren Eigenschaften und Überlegungen könnten auf alle Fälle verallgemeinert werden, ohne diesen strukturellen Kern zu verändern. Im Vergleich zu einer reinen induktiven Prüfung einer Aussage würde beim generic example die Verallgemeinerbarkeit der Überlegungen mitgedacht und von den spezifischen Besonderheiten des betrachteten Falls abstrahiert. Die Explikation der Überlegungen erfolge aber anhand der vergleichsweise anschaulichen konkreten Beispiele.

Reid und Knipping (2010) fassen unter dem Terminus „generic“ (Reid & Knipping, 2010, S. 130) bzw. „generic proofs“ (Reid & Knipping, 2010, S. 134) Beweisversuche zusammen, die auf Beispielen basieren, welche im Gegensatz zu rein empirischen Verifikationen nicht nur für sich selbst stehen, sondern eine größere Klasse von Fällen repräsentieren. In diese weite Definition können unter anderem die vorgestellten intuitiven Beweise von Branford (1913), die inhaltlich-anschaulichen Beweise von Wittmann und Müller (1988) sowie die generic examples nach Mason und Pimm (1984) und nach Balacheff (1988) integriert werden.

Bei allen diesen Vorformen wissenschaftlicher Beweise stellt sich die Frage, inwieweit diese als allgemeingültige Beweise anerkannt werden können und sollten. Wie bereits angedeutet, ist es Erfahrungsberichten und Studien zufolge keineswegs selbstverständlich, dass Lernende die von generischen Beispielen prinzipiell abstrahierbaren allgemeingültigen Argumentgefüge wahrnehmen und anerkennen (u. a. Balacheff, 1988; Krumsdorf, 2017; Mason & Pimm, 1984; Wittmann & Ziegenbalg, 2004; Kap. 5). Besonders plausibel erscheint diese Problematik bei pictorial generic examples, die Reid und Knipping (2010) als einen Untertyp von generic proofs klassifizieren.[34] Bei pictorial generic examples wird eine ikonische Darstellung (z. B. eine Zeichnung) als Prototyp für alle infrage stehenden Fälle verwendet. Das Hineinsehen des „Beweises“

[34] Pictorial generic examples werden unter verschiedenen, nicht ganz kongruenten Termini wie „‘Siehe‘-Beweise“ (Winter, 1991, S. 136) und „proofs without words“ (Nelsen, 2000, S. ix) unter anderem von Alsina und Nelsen (2013), Kautschitsch (1989), Nelsen (2000), Neubrand und Möller (1990), Winter (1991) und Wittmann und Ziegenbalg (2004) thematisiert. Im Schulunterricht etabliert und in dieser Arbeit relevant erscheinen insbesondere Beweisversuche mithilfe von figurierten Zahlen. Figurierte Zahlen sind natürliche Zahlen, die z. B. mit Punkt- oder Kästchenmustern ikonisch dargestellt werden können. Die Abbildung 70 in Abschnitt 3.3.11 zeigt einen Beweisversuch einer Schülerin, in dem figurierte Zahlen verwendet werden.

in eine solche Abbildung erfordert anspruchsvolle Denkprozesse. Es müssen die in der ikonischen Darstellung implizit enthaltenen relevanten strukturellen Merkmale, Beziehungen und sukzessiven Folgen von Operationen extrahiert und verallgemeinert werden. Dazu sind die relevanten Wissenselemente zu aktivieren und anzuwenden, Zusammenhänge zu erkennen und die Schlüsse zumindest gedanklich zu verbalisieren, was einen Wechsel zwischen der ikonischen und einer verbalen oder gar symbolischen Darstellung einschließt. Bei jedem Schluss sollte zudem reflektiert werden, ob und ggf. warum die Operation allgemein durchführbar ist (Kautschitsch 1989; Neubrand & Möller, 1990; Winter, 1991). In Anbetracht dieser beim Nachvollziehen eines pictorial generic example erforderlichen Denkprozesse schreibt Winter (1991): „Jedenfalls ist dieser scheinbar so durchsichtige anschauliche Beweis alles andere als eine bare Selbstverständlichkeit“ (Winter, 1991, S. 137). Bezüglich der Frage, inwieweit pictorial generic examples Beweise sind, gibt es leicht unterschiedliche Standpunkte. Der Mathematiker Nelsen (2000) schreibt in Bezug auf ikonisch dargestellte Beweise ohne Worte („proof without words“):

> Generally, proofs without words (PWWs) are pictures or diagrams that help the reader see *why* a particular mathematical statement may be true and also to see *how* one might begin to go about proving it true. (Nelsen, 2000, S. ix, H. i. O.).

Dabei fällt auf, dass Nelsen (2000) mit den Formulierungen „may be true“ und „might begin“ eine durchaus vorsichtige Ausdrucksweise wählt. Demzufolge sind proofs without words eine Keimzelle eines allgemeingültigen, wissenschaftlichen Beweises. Wittmann und Ziegenbalg (2004), die sich insbesondere mit figurierten Zahlen beschäftigen, betonen wiederholt, dass sie ikonisch dargestellte generische Beispiele alleine noch nicht als Beweis ansehen:

> Wir weisen noch einmal darauf hin, dass die Präsentierung eines suggestiven Musters als „Beweis ohne Worte“ nicht ausreicht. Es muss schon durch einen erklärenden Text sichergestellt werden, dass die zur Begründung von Beziehungen angewandten Operationen wirklich allgemein ausführbar sind. (Wittmann & Ziegenbalg, 2004, S. 42, H. i. O.)

Dieser bis heute nicht ganz klare Diskussionsstand bezieht sich nicht nur auf pictorial generic examples. Biehler und Kempen (2016) arbeiten heraus, dass zwischen den in der internationalen Literatur häufigen Termini „generic example“ und „generic proof“ oftmals nicht präzise differenziert wird. Insbesondere sei bis heute mehrdeutig, „[was] genau zu einem generic example an Erläuterungen hinzukommen muss, um es zu einem generic proof zu machen“ (Biehler & Kempen, 2016, S. 167). Biehler und Kempen (2013) fordern im Rahmen einer Einführungsveranstaltung für Lehramtsstudierende, dass das im generischen Beispiel prinzipiell erkennbare, allgemeingültige Argumentgefüge zusätzlich zur Angabe des generischen Beispiels ausformuliert wird. „Erst in dem Zusammenspiel von generischen Beispielen und explizierter Begründung wird dann von einem generischen Beweis gesprochen“ (Kempen, 2019, S. 36). Zur Illustration ihres Konzepts formulieren Biehler und Kempen (2013) u. a. den folgenden von ihnen als generischen Beweis bezeichneten Beweisversuch zu der Aussage, dass die Summe einer ungeraden natürlichen Zahl und ihres Doppelten ungerade ist.

> $1 + 2 \cdot 1 = 3 \cdot 1 = 3$ $\quad 3 + 2 \cdot 3 = 3 \cdot 3 = 9$ $\quad 5 + 2 \cdot 5 = 3 \cdot 5 = 15$
>
> Comparing the equations, one can recognize that the result must always be three times the initial number. Since three times an odd number is always odd, the result is an odd number. (Biehler & Kempen, 2013, S. 89)

Krumsdorf (2017) führt eine andere Abstufung ein. Er beschäftigt sich in seiner Arbeit mit „beispielgebundenem Beweisen“ (Krumsdorf, 2017, S. 139). Sobald ein einzelnes Argument eines Argumentgefüges beispielgebunden ist, bezeichnet er das Argumentgefüge im Ganzen als beispielgebunden. Die zuvor skizzierten generic examples, inhaltlich-anschaulichen Beweise etc. werden von Krumsdorf (2017) unter dem Begriff „beispielgebundener Beweis“ subsummiert. Auf seinem Verständnis beispielgebundener Beweise aufbauend unterscheidet er zwischen beispielgebundenen Beweisen mit latenter, subjektiv realisierter und manifest gewordener Sinnstruktur (Krumsdorf, 2017). Rein generische Beispiele (ohne ergänzende Erklärung) bezeichnet Krumsdorf (2017) als beispielgebundene Beweise mit latenter Sinnstruktur, wenn zwar ein Experte, aber nicht der betrachtende Lernende in diesem Beispiel eine Begründung für die Allgemeingültigkeit einer Aussage erkennt. Wenn sich der betrachtende Lernende der Verallgemeinerbarkeit der am Beispiel durchgeführten Operationen bewusst ist, spricht Krumsdorf (2017) von einem beispielgebundenen Beweis mit subjektiv realisierter Sinnstruktur. In dieser Definition eines beispielgebundenen Beweises mit subjektiv realisierter Sinnstruktur wird wohlgemerkt keine Explizierung gefordert. Ob es sich aus der Perspektive eines Lernenden nur um einen Beweis mit latenter oder mit subjektiv realisierter Sinnstruktur handelt, ist folglich schwierig zu beurteilen. Dementsprechend erklärt Krumsdorf (2009): „Manifest wird der beispielgebundene Beweis allerdings erst in seiner Entäußerung, etwa im Diskurs“ (Krumsdorf, 2009, S. 712). Ein Argumentgefüge bezeichnet er als beispielgebundenen Beweis mit manifest gewordener Sinnstruktur, wenn das generische Beispiel um Überlegungen ergänzt wird, die auf seine Allgemeingültigkeit hinweisen (Krumsdorf, 2017). Was als ergänzende Erläuterungen ausreicht, wird von ihm offengehalten. Die Allgemeingültigkeit müsse „in überzeugender Weise versprachlicht, entäußert, dargestellt [werden]“ (Krumsdorf, 2017, S. 140). Eine vollständige Explizierung des allgemeinen Argumentgefüges sei aber nicht erforderlich. In diesem Fall könnten die Ausführungen bereits als wirklicher mathematischer Beweis aufgefasst werden (Krumsdorf, 2017). Die Definition eines generischen Beweises nach Biehler und Kempen (2013) bewegt sich somit an der Grenze dessen, was Krumsdorf (2017) überhaupt noch als beispielgebunden bezeichnen würde. Andersherum bedeutet dies, dass Biehler und Kempen (2013) beispielgebundene Beweise mit latenter oder subjektiv realisierter Sinnstruktur sowie zumindest einen Teil der von Krumsdorf (2017) als beispielgebundene Beweise mit manifest gewordener Sinnstruktur bezeichneten Argumentgefüge als generische Beispiele, aber noch nicht als generische Beweise bezeichnen würden.

In den Begrifflichkeiten von Meyer und Voigt (2008) handelt es sich bei beispielgebundenen Beweisen mit latenter Sinnstruktur nur um eine Prüfung oder Probe mit latenter Beweisidee. Wenn diese latente Beweisidee dem Lernenden bewusst sei und die Person erkenne, dass die am Beispiel durchgeführten deduktiven Schlüsse von dessen konkreten Besonderheiten losgelöst werden können, sei ein beispielgebundener Beweis geführt (Meyer & Voigt, 2008).

Insgesamt betrachtet gibt es ein Kontinuum denkbarer Vorformen wissenschaftlicher Beweise, die sich zwischen Beweisversuchen mit deduktiver und solchen mit induktiver Schlussweise bewegen. Zu der Frage, inwieweit diese Vorformen als allgemeingültige Beweise anerkannt werden können und sollten, gibt es bis heute keinen klaren Diskussionsstand, sondern mehrere

parallel existierende Ansätze. Für die vergleichende Analyse von Beweisversuchen von Schülerinnen und Schülern **im Rahmen dieser Arbeit** gilt es also, eine inhaltlich sinnvolle und praktikable Abgrenzung zu treffen. Ein Beweisversuch wird der Oberkategorie *begründender Ansatz* zugeordnet, wenn ein Argumentgefüge vorkommt, das einen Bezug zu der infrage stehenden Aussage aufweist und eine Beweisidee enthält, deren Allgemeingültigkeit oder Verallgemeinerbarkeit in der Bearbeitung explizit angedeutet ist. Alle Beweisversuche, in denen kein begründender Ansatz, aber ein Beispiel erkennbar ist, das einen Bezug zu der infrage stehenden Aussage aufweist, werden der Oberkategorie *prüfender Ansatz* zugeordnet.

Bei dieser Abgrenzung ist zu bedenken, dass die Beweisversuche im Rahmen dieser Arbeit ausschließlich in schriftlicher Form vorliegen. Es ist beispielsweise nicht möglich, durch ein zusätzliches Interview herauszufinden, inwieweit die Verallgemeinerbarkeit einer im Schriftdokument erkennbaren Beweisidee von der Autorin bzw. dem Autor subjektiv realisiert wurde. Da es nach den vorhergehenden Ausführungen alles andere als selbstverständlich erscheint, dass Schülerinnen und Schüler eine von einem Beispiel abstrahierbare Beweisidee erkennen, werden im Rahmen dieser Arbeit nur solche Beweisversuche der Oberkategorie *begründender Ansatz*, das heißt derselben Oberkategorie wie Beweisversuche mit deduktiver Schlussweise, zugeordnet, in denen die Verallgemeinerbarkeit von an einem oder mehreren Beispielen durchgeführten Operationen explizit angedeutet ist. Diese Abgrenzung ist in der obigen, vereinfachten Übersicht über die berücksichtigten mathematikdidaktischen Konzepte durch farbliche Hinterlegungen verortet (Abb. 9). Beispielsweise werden die von Meyer und Voigt (2008) beschriebenen Prüfungen, und insbesondere auch die Prüfungen mit latenter Beweisidee, im Rahmen dieser Arbeit als Beweisversuche mit prüfendem Ansatz bezeichnet. Beispielgebundene Beweise im Sinne von Meyer und Voigt (2008) können sowohl Beweisversuche mit begründendem als auch solche mit prüfendem Ansatz sein. Während Meyer und Voigt (2008) es im Wesentlichen von der mentalen Realisierung der Verallgemeinerbarkeit an Beispielen durchgeführter Operationen abhängig machen, ob es sich um einen beispielgebundenen Beweis oder eine Prüfung mit latenter Beweisidee handelt, wird in der vorliegenden Arbeit die explizite Andeutung der Verallgemeinerbarkeit als anhand von Schriftdokumenten erkennbarem Kriterium zur Abgrenzung von Beweisversuchen mit begründendem und solchen mit prüfendem Ansatz verwendet. Da eine mentale Realisierung nicht unbedingt mit einer expliziten, schriftlichen Andeutung der Verallgemeinerbarkeit einhergehen muss, können beispielgebundene Beweise im Sinne von Meyer und Voigt (2008) sowohl in die Kategorie *begründender Ansatz* als auch in die Kategorie *prüfender Ansatz* fallen. Im Vergleich zu den Ausführungen von Meyer und Voigt (2008) besteht eine weitere Besonderheit im Analysemodell der vorliegenden Arbeit darin, dass es für die Bezeichnung eines Beweisversuchs als Beweisversuch mit begründendem Ansatz ausreicht, wenn Grundzüge einer Beweisidee erkennbar sind. Es müssen nicht unbedingt alle für einen Beweis erforderlichen Operationen aus der Bearbeitung hervorgehen. Außerdem müssen die durchgeführten Operationen, deren Verallgemeinerbarkeit explizit angedeutet ist, nicht tatsächlich verallgemeinerbar sein. Sie können Fehler enthalten, wie im Folgenden näher erläutert wird.

Beweisversuche, in denen an Beispielen eine Beweisidee erkennbar ist, deren Verallgemeinerbarkeit angedeutet wird, bis hin zu Beweisversuchen mit rein deduktiver Schlussweise werden

im Rahmen dieser Arbeit in der Oberkategorie ***begründender Ansatz*** zusammengefasst. Der Ausdruck „begründender Ansatz", der im Analysemodell als Kategorienname dient, beruht auf dem Begriff „Begründung". Dieser soll im Vergleich zum Beweisbegriff hervorheben, dass neben Beweisen im Sinne korrekter und vollständiger, deduktiver Argumentgefüge (Abschn. 3.4) auch weitere Begründungsformen mitgedacht sind. Der Begriff der Begründung wird allerdings in der mathematikdidaktischen Literatur unterschiedlich verwendet, was eine Auseinandersetzung mit verschiedenen Definitionen und eine klare Positionierung für den Rahmen dieser Arbeit erfordert.

Beispielsweise nutzen Fischer und Malle (1985) einen weiten Begründungsbegriff, der neben deduktivem Schließen u. a. auch Induktionen, Wahrscheinlichkeitsaussagen und Berufungen auf eine Autorität umfasst. Beweise werden dabei als besondere Formen von Begründungen angesehen. Auch von Brunner (2014) wird das Beweisen dem Begründen untergeordnet. Von ihr „wird Begründen als Oberbegriff verstanden und als Kontinuum mit den Stationen alltagsbezogenes Argumentieren, Argumentieren mit mathematischen Mitteln, logisches Argumentieren mit mathematischen Mitteln und formal-deduktiven Beweisen konzeptualisiert" (Brunner, 2014, S. 30). Reiss und Ufer (2009) sehen Beweise ebenfalls als einen Spezialfall von Begründungen an. Der Begründungsbegriff wird dabei jedoch enger gefasst. „Begründen [wird] als der Teilbereich von Argumentieren verstanden, der sich auf die primär deduktive Absicherung einer als plausibel angenommenen Behauptung bezieht" (Reiss & Ufer, 2009, S. 158). Beweise würden sich zudem durch „Spezifika wie den Bezug auf eine Rahmentheorie und einen gewissen Grad an formalsprachlicher Abfassung" (Reiss & Ufer, 2009, S. 158) auszeichnen. Eine ähnliche Festlegung haben Meyer und Voigt (2008) vorgenommen. Bei ihnen ist der Ausdruck „Begründung" ein „Oberbegriff für die Prozesse der deduktiven Erkenntnissicherung" (Meyer & Voigt, 2008, S. 129). Sie „sprechen von einem Beweis, wenn die Begründung eine stringente Kette von Deduktionen ist oder wenn unterstellt werden kann, dass der Beweisführende die Begründung so darstellen könnte und es nur aus praktischen Gründen unterlässt" (Meyer & Voigt, 2008, S. 129). Im Vergleich zu Reiss und Ufer (2009) betonen sie die Stringenz einer Begründung als Spezifikum von Beweisen, die Darstellungsweise wird von ihnen nicht explizit berücksichtigt. Von Bruder und Müller (1983), Müller (1991) sowie Bruder und Pinkernell (2011) werden Begründungen, anders als bei den bisher genannten Begriffsdefinitionen, als „einfachste, d. h. als ‚einschrittige' Beweise verstanden" (Müller 1991, S. 738). Eine solche Begründung könne unter anderem durch Bezug auf eine Definition, einen Satz oder ein Verfahren erfolgen (Bruder & Pinkernell, 2011).

Anknüpfend an mehrere der genannten Ansätze, wird der Begriff „Begründung" im Rahmen dieser Arbeit als ein Hyperonym für ein gewisses Spektrum an Möglichkeiten der Erkenntnissicherung, welches mathematische Beweise umfasst, verwendet. Ein Beweisversuch wird als Begründung bezeichnet, wenn ein Argumentgefüge[35] vorhanden ist, das zu einem Beweis der infrage stehenden Aussage weiterentwickelt werden kann. Dabei muss sich das Argumentgefüge

[35] Dieser Terminus wird in der vorliegenden Arbeit in Anlehnung an Schwarzkopf (2000) und Krumsdorf (2017) verwendet und ermöglicht, zwischen einzelnen Argumenten und einem Argumentgefüge zu unterscheiden. In dieser Arbeit wird als ein Argument bezeichnet, wenn ausgehend von mehr oder weniger explizierten Voraussetzungen auf eine Behauptung geschlossen wird. Bei einem Argument handelt es sich um einen einzelnen Schluss, bei dem

auf die infrage stehende Aussage beziehen, bis auf Ungenauigkeiten in der Formulierung korrekt sein und die Allgemeingültigkeit oder Verallgemeinerbarkeit der im Beweisversuch aufgezeigten Überlegungen explizit andeuten. Hinsichtlich der von der Autorin oder dem Autor verwendeten Darstellungsmittel wird der Begründungsbegriff nicht eingegrenzt. Gleiches gilt für die Komplexität des Argumentgefüges, eine Begründung kann sowohl ein- als auch mehrschrittig sein, wobei ein Argument als ein Begründungsschritt verstanden wird.[36] Außerdem muss die Schlussweise im Vergleich zu den Definitionen von Meyer und Voigt (2008) sowie Reiss und Ufer (2009) nicht unbedingt rein deduktiv sein. Beweisversuche, die ausschließlich auf einer unsicheren Schlussweise, wie einer Induktion oder einer Berufung auf eine Autorität beruhen, werden jedoch, anders als in den Definitionen von Fischer und Malle (1985) und von Brunner (2014), klar von Begründungen abgegrenzt. Mit dem beschriebenen Begründungsbegriff dieser Arbeit wird die mit Blick auf den Beweisbegriff innerhalb der Mathematik zentrale Idee der prinzipiellen Annäherbarkeit an einen formal-strengen Beweis auf die Analyse von Beweisversuchen von Schülerinnen und Schülern übertragen.

Die Oberkategorie *begründender Ansatz* umfasst im Rahmen dieser Arbeit alle Beweisversuche, in denen eine Begründung im beschriebenen Sinne vorkommt. Darüber hinaus ist es selbst bei Beweisversuchen mit deduktiver Schlussweise[37] möglich, dass sie keine weiterentwickelbare Beweisidee enthalten oder sich auf eine modifizierte Interpretation der infrage stehenden Aussage beziehen. Dann handelt es sich zwar um den Versuch einer Begründung, als Begründung der infrage stehenden Aussage kann der Beweisversuch aber nicht bezeichnet werden. Der weichere Ausdruck „begründender Ansatz“ soll verdeutlichen, dass auch solche Beweisversuche in diese Oberkategorie fallen. Positiv formuliert wird ein Beweisversuch der Kategorie *begründender Ansatz* zugeordnet, wenn das Argumentgefüge einen Bezug zu der infrage stehenden Aussage aufweist und eine Beweisidee enthält, deren Allgemeingültigkeit oder Verallgemeinerbarkeit in der Bearbeitung explizit angedeutet ist.

Von Beweisversuchen mit begründendem Ansatz werden in dieser Arbeit Beweisversuche mit prüfendem Ansatz unterschieden. Die Bezeichnung „**prüfender Ansatz**“ basiert auf dem Begriff „Prüfung“. Dieser wird von Meyer und Voigt (2009) verwendet. Nach den Autoren ist „die Schlussform der Induktion charakteristisch für das Prüfen eines Satzes“ (Meyer & Voigt, 2009, S. 51). Auf Grundlage der Theorie der Schlussformen des Philosophen C. S. Peirce erläutern sie, dass eine Allaussage mittels Induktion nur geprüft, weder entdeckt noch gesichert werden kann (Meyer & Voigt, 2009). Eine Induktion ist kein wahrheitsübertragender, sondern ein unsicherer Schluss. Sie kann die Plausibilität einer Allaussage erhöhen und z. B. zum Verständnis der infrage stehenden Aussage beitragen, jedoch nicht ihre Allgemeingültigkeit garantieren.

Auf den Überlegungen zur Schlussweise der Induktion aufbauend, unterscheiden Meyer und Voigt (2009) u. a. zwischen einer Prüfung an einem speziellen Fall, einer Prüfung an mehreren

die Schlussweise nicht nur deduktiv, sondern z. B. auch induktiv sein kann. Ein Argumentgefüge besteht aus einem oder mehreren, üblicherweise verknüpften Argumenten.

[36] Die u. a. von Bruder und Pinkernell (2011) als Begründung bezeichneten einschrittigen Beweise werden folglich als eine Art von Begründungen dem hier verwendeten Begründungsbegriff untergeordnet.

[37] Das Verständnis deduktiven Schließens wird für die Analyse von Beweisversuchen im Rahmen dieser Arbeit im Vergleich zum logischen Ideal einer Deduktion erweitert, da die einzelnen Deduktionen in Beweisversuchen von Lernenden fehlerhaft und unvollständig sein können (Abschn. 3.1.2, 3.3.6).

Fällen und einer Prüfung mit latenter Beweisidee. „Die Art der induktiven Prüfung beeinflusst die Höhe der Glaubwürdigkeit und kann sogar zu einer rein deduktiven Begründung überleiten“ (Meyer & Voigt, 2009, S. 51; Abschn. 3.3.7, 3.3.8). Mit Blick auf Beweisversuche von Schülerinnen und Schülern kann es dementsprechend vorkommen, dass Lernende eine Aussage mit genau einem oder mit mehreren Beispielen überprüfen. Auch kann an den Beispielen aus Expertensicht eine Beweisidee zu erkennen sein. Insofern die Verallgemeinerbarkeit dieser Beweisidee nicht explizit aus der Bearbeitung hervorgeht, wird der Beweisversuch im Rahmen dieser Arbeit nicht als Beweisversuch mit begründendem Ansatz bezeichnet, sondern in die Kategorie *prüfender Ansatz* eingeordnet.

In dieser Arbeit wird der Ausdruck „prüfender Ansatz“ anstatt des Begriffs „Prüfung“ für die Bezeichnung der Oberkategorie verwendet, weil die Beispiele in Beweisversuchen von Lernenden Fehler enthalten können. Außerdem ist es möglich, dass in Beweisversuchen zwar ein Beispiel, aber im Vergleich zu einer Induktion kein expliziter Schluss auf die Behauptung vorkommt. Unter dem Begriff „prüfender Ansatz“ werden Beweisversuche subsummiert, in denen mindestens ein Beispiel vorkommt, das einen klaren Bezug zu der infrage stehenden Aussage aufweist. Im Sinne einer kompetenzorientierten Analyse werden Beweisversuche nur der Kategorie *prüfender Ansatz* zugeordnet, wenn kein begründender Ansatz erkennbar ist.

Neben Beweisversuchen mit begründendem Ansatz und solchen mit prüfendem Ansatz, kann es entsprechend den vorgestellten Ausführungen von Fischer und Malle (1985) sein, dass sich Lernende auf eine Autorität berufen, Analogieschlüsse durchführen oder Wahrscheinlichkeitsaussagen tätigen. Solche Beweisversuche, die auf eine Beschäftigung mit der zu zeigenden Aussage hindeuten, jedoch weder einen begründenden noch einen prüfenden Ansatz enthalten, werden unter der Oberkategorie ***weder begründender noch prüfender Ansatz*** erfasst.

Weiterhin ist es möglich, dass die Zuordnung eines Beweisversuchs zu einer der Kategorien *begründender Ansatz*, *prüfender Ansatz* und *weder begründender noch prüfender Ansatz* misslingt, etwa bei sehr vage formulierten Bearbeitungen oder solchen, die auf einer falschen Interpretation der zu zeigenden Aussage beruhen. Um trotzdem beliebige Beweisversuche codieren zu können, wird die Kategorie ***nicht eindeutig kategorisierbarer Ansatz*** ergänzt.

Schließlich kann es passieren, dass Lernende zu einem Aufgabenteil keinen Beweisversuch abgeben. Gerade mit Blick auf die in dieser Arbeit fokussierten Beweisversuche von Teilnehmenden der Mathematik-Olympiade kann dies unterschiedliche Gründe haben. Es ist z. B. denkbar, dass eine Schülerin bzw. ein Schüler versucht hat, den Aufgabenteil zu bearbeiten, jedoch keine Beweisidee gefunden und deshalb nichts abgegeben hat. Zudem ist es möglich, dass der Aufgabenteil als schwierig oder uninteressant wahrgenommen und aus diesem Grund nicht beachtet wurde. Auch kann es vorkommen, dass eine Person keine Zeit oder keine Motivation mehr hatte, um sich mit dem Aufgabenteil zu beschäftigen und nicht einmal die Aufgabenstellung gelesen wurde. Ein fehlender Beweisversuch lässt folglich keine Aussage bezüglich Beweiskompetenzen des Lernenden zu. Fehlende Beweisversuche werden deshalb nicht in die Kategorie *weder begründender noch prüfender Ansatz* eingeordnet, sondern in einer separaten Oberkategorie, der Kategorie ***fehlend***, erfasst.

3.3.2.2 Vorstellung der Kategorien

Im Folgenden werden die fünf Oberkategorien *begründender Ansatz*, *prüfender Ansatz*, *weder begründender noch prüfender Ansatz*, *nicht eindeutig kategorisierbarer Ansatz* und *fehlend* nacheinander definiert und anhand einzelner Beweisversuche von Lernenden aus der siebten oder achten Klassenstufe illustriert. Diese Beweisversuche beziehen sich auf folgende Aussage: Die Summe von vier aufeinanderfolgenden ungeraden Zahlen ist immer durch 8 teilbar.

Ein Beweisversuch wird der Oberkategorie ***begründender Ansatz*** (Ans3) zugeordnet, wenn das Argumentgefüge einen Bezug zu der infrage stehenden Aussage aufweist und eine Beweisidee enthält, deren Allgemeingültigkeit oder Verallgemeinerbarkeit in der Bearbeitung explizit angedeutet ist. Als Beweisidee wird eine inhaltsbezogene Überlegung bezeichnet, die dem Argumentgefüge zugrunde liegt und auf Voraussetzungen der infrage stehenden Aussage oder anderen mathematischen Sätzen und Definitionen beruht. In einem Beweisversuch mit begründendem Ansatz muss die Schlussweise nicht unbedingt rein deduktiv sein. Zudem kann das Argumentgefüge Fehler oder Lücken enthalten.

Beispielsweise basiert die Bearbeitung 3.3.2a in Abbildung 10 auf der Beweisidee der Teilbarkeit der Summe zweier Zahlenpaare, wobei der letzte Satz in der Bearbeitung dessen Verallgemeinerbarkeit andeutet. Das Argumentgefüge ist korrekt, aber unvollständig.

Bearbeitung 3.3.2a: Klassenstufe 7, weiblich (Code: 0721706)

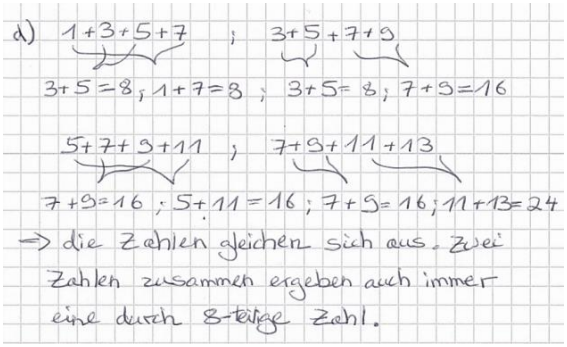


Abbildung 10: Bearbeitung 3.3.2a (Ans3 *begründender Ansatz*)

Unter der Oberkategorie ***prüfender Ansatz*** (Ans2) werden Beweisversuche subsummiert, in denen kein begründender Ansatz erkennbar ist, aber ein Beispiel vorkommt, das einen klaren Bezug zu der infrage stehenden Aussage aufweist. Dieses Beispiel kann Fehler enthalten und sich auch auf eine leicht modifizierte Interpretation der infrage stehenden Aussage beziehen.

Bei der Bearbeitung 3.3.2b in Abbildung 11 passt das angegebene Beispiel nicht ganz zu der hier fokussierten Aussage, weil die ungeraden Zahlen nicht aufeinanderfolgen. Dadurch, dass es sich um vier ungerade Zahlen handelt und deren Summe durch acht dividiert wird, ist trotzdem ein klarer Bezug zu dieser Aussage zu erkennen. Der zweite Satz „Deswegen ist die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer durch 8 teilbar“ verdeutlicht, dass der Siebtklässler das Beispiel zumindest als Bestätigung, wenn nicht sogar als „Beweis“ für die Allgemeingültigkeit der infrage stehenden Aussage ansieht. Inwieweit in einer Bearbeitung ausgehend von den Beispielen auf die Gültigkeit der infrage stehenden Aussage geschlossen wird, hat für die Zuordnung zur Kategorie prüfender Ansatz jedoch keine Relevanz.

Bearbeitung 3.3.2b: Klassenstufe 7, männlich (Code: 0711314)

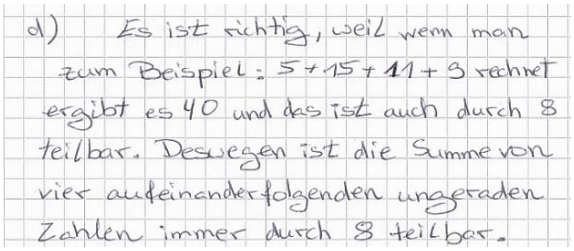


Abbildung 11: Bearbeitung 3.3.2b (Ans2 *prüfender Ansatz*)

Ein Beweisversuch gehört in die Oberkategorie ***weder begründendem noch prüfendem Ansatz*** (Ans1), wenn zum infrage stehenden Aufgabenteil eine Bearbeitung abgegeben wurde[38], die jedoch weder der Kategorie *begründender Ansatz* noch der Kategorie *prüfender Ansatz* zugeordnet werden kann.

In der Bearbeitung 3.3.2c in Abbildung 12 ist als Grund für die Gültigkeit der infrage stehenden Aussage angegeben, dass es kein Gegenbeispiel gebe. Die Ausführungen deuten somit eine gewisse Bewusstheit für die Problematik der Allgemeingültigkeit an, beruhen allerdings weder klar auf Voraussetzungen der zu beweisenden Aussage noch auf anderen mathematischen Sätzen oder Definitionen. Es ist folglich keine Beweisidee erkennbar, die Bearbeitung gehört nicht in die Kategorie *begründender Ansatz*. Da in der Bearbeitung kein Beispiel vorkommt, handelt es sich auch nicht um einen Beweisversuch mit prüfendem Ansatz. Deshalb wird sie in die Kategorie *weder begründender noch prüfender Ansatz* eingeordnet.

Bearbeitung 3.3.2c: Klassenstufe 8, männlich (Code: 0811901)

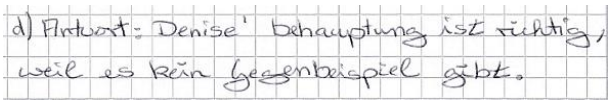


Abbildung 12: Bearbeitung 3.3.2c (Ans1 *weder begründender noch prüfender Ansatz*)

Weiterhin kann es vorkommen, dass Ausführungen zu dem infrage stehenden Aufgabenteil abgegeben wurden, die jedoch keiner der Kategorien *begründender Ansatz*, *prüfender Ansatz* und *weder begründender noch prüfender Ansatz* eindeutig zugeordnet werden können. Dann wird der Beweisversuch in die Oberkategorie ***nicht eindeutig kategorisierbarer Ansatz*** (Ans55) eingeordnet.

Die Abbildung 13 zeigt exemplarisch eine Bearbeitung, bei der es unklar ist, ob es sich um einen begründenden oder einen prüfenden Ansatz handelt. Der Schüler hat die vorgegebene Aussage für die zwei kleinstmöglichen Fälle, d. h. für die vier aufeinanderfolgenden ungeraden Zahlen $1, 3, 5, 7$ und $3, 5, 7, 9$ korrekt überprüft. Die Prüfung der Summe der vier nächstgrößeren aufeinanderfolgenden ungeraden Zahlen $5, 7, 9, 11$ hat er jedoch abgebrochen. Der Kommentar „wäre fünfstellig!“ und die Unterstreichung der Zahl 11 deuten an, dass der Schüler die infrage stehende Aussage so verstanden hat, dass die vier aufeinanderfolgenden ungeraden Zahlen

[38] Im Rahmen dieser Arbeit gilt eine Bearbeitung im Grenzfall auch als abgegeben, falls alle Ausführungen durchgestrichen sind.

insgesamt nur aus vier Ziffern bestehen dürfen. Diese Interpretation passt auch zu dem darunter stehenden Satz: „Es gibt nur zwei vierstelligen Zahlen dieser Sorte und bei Beiden trifft die Aussage zu.“ Durch die Uminterpretation ist aus der zu zeigenden Allaussage mit unendlichem Gültigkeitsbereich eine Allaussage mit sehr kleinem Gültigkeitsbereich geworden. Dem Schüler ist es mit der Überprüfung der genannten Fälle gelungen, seine Version der zu zeigenden Aussage zu beweisen. Allerdings lässt die Bearbeitung offen, ob er im Falle einer korrekten Interpretation, d. h. bei der Aussage mit unendlichem Gültigkeitsbereich, einen begründenden oder ausschließlich einen prüfenden Ansatz gewählt hätte. Während die Aufgabenstellung auf das Beweisen einer Allaussage mit unendlichem Gültigkeitsbereich fokussiert, lässt sich nur festhalten, dass der Schüler Beispiele betrachtet, dabei mit dem kleinstmöglichen Fall begonnen und systematisch probiert hat sowie dass ihm bei der Bearbeitung ein grundlegender Fehler passiert ist. Die exemplarisch vorgestellte Bearbeitung ist also hinsichtlich des gewählten Ansatzes nicht eindeutig kategorisierbar, weil die zu zeigende Aussage so sehr anders interpretiert wurde, dass unklar ist, wie der Schüler die Aufgabe im Falle einer korrekten Interpretation bearbeitet hätte. Beweisversuche in der Kategorie *nicht eindeutig kategorisierbarer Ansatz* sind weder unbedingt besser noch schlechter als solche in eine der Kategorien *begründender*, *prüfender* oder *weder begründender noch prüfender Ansatz*. Sie lassen sich nicht eindeutig einordnen und werden deshalb separat erfasst.

Bearbeitung 3.3.2d: Klassenstufe 7, männlich (Code: 0711225)

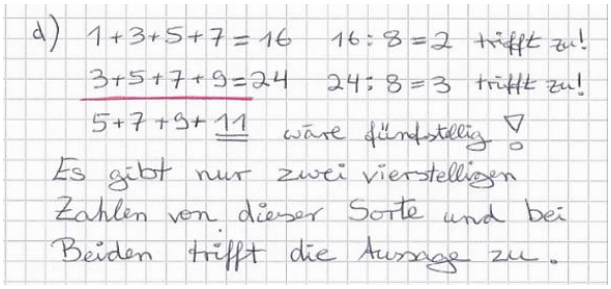


Abbildung 13: Bearbeitung 3.3.2d (Ans55 *nicht eindeutig kategorisierbarer Ansatz*)

Ein Beweisversuch gilt als **fehlend** (Ans99), wenn zu dem infrage stehenden Aufgabenteil kein Beweisversuch, d. h. insbesondere auch keine durchgestrichenen Ausführungen, abgegeben wurden.

3.3.2.3 Hinweise zu den Kategorien

Jeder Beweisversuch ist bei der Anwendung des Analysemodells genau einer der fünf Oberkategorien *begründender Ansatz*, *prüfender Ansatz*, *weder begründender noch prüfender Ansatz*, *nicht eindeutig kategorisierbarer Ansatz* und *fehlend* zuzuordnen. Dabei ist zu bedenken, dass es sich bei zu analysierenden Bearbeitungen prinzipiell um ein Kontinuum an denkbaren Beweisversuchen handelt, vor dessen Hintergrund möglichst sinnvolle Abgrenzungen getroffen wurden. Wie auch die gestrichelten Linien im Analysemodell (Abb. 8) andeuten sollen, können z. B. in der Kategorie *begründender Ansatz* sowohl Bearbeitungen vorkommen, die eindeutig Beweisversuche mit begründendem Ansatz sind, als auch solche, die an der Grenze zu Beweisversuchen mit prüfendem oder weder begründendem noch prüfendem Ansatz liegen.

Die Zuordnung eines Beweisversuchs zu der Kategorie *begründender Ansatz* sagt entsprechend obiger Definition aus, dass das Argumentgefüge einen Bezug zu der infrage stehenden Aussage aufweist und eine Beweisidee enthält, deren Allgemeingültigkeit oder Verallgemeinerbarkeit in der Bearbeitung explizit angedeutet ist. Sie lässt vermuten, dass der Autorin bzw. dem Autor die Allgemeingültigkeit der infrage stehenden Aussage sowie die Notwendigkeit der Berücksichtigung aller möglicher Fälle bewusst ist. Die Einordnung in die Kategorie *begründender Ansatz* bedeutet allerdings weder, dass das Argumentgefüge in allgemeingültiger Form formuliert wurde, noch dass es direkt zu einem Beweis verallgemeinerbar ist. Mit Blick auf die Weiterentwicklung von Beweisversuchen zu einem Beweis sind Bearbeitungen in der Kategorie *begründender Ansatz* nicht zwingend besser als solche in der Kategorie *prüfender Ansatz*.

Die Zuordnung eines Beweisversuchs zur Kategorie *prüfender Ansatz* drückt aus, dass die Bearbeitung ein Beispiel enthält, das einen Bezug zu der zu zeigenden Aussage aufweist. Es wurde folglich zumindest versucht, einen Spezialfall der Aussage zu betrachten. Dabei ist es möglich, dass eine Schülerin oder ein Schüler ausgehend von Beispielen eine korrekte und vollständige Begründung verbalisieren könnte, dies jedoch im vorliegenden Beweisversuch nicht getan hat. Zugleich kann es sein, dass Lernende vergeblich versucht haben, ausgehend von Beispielen eine Beweisidee zu generieren, dass sie Beispiele als Beweis einer Allaussage ansehen oder Schwierigkeiten beim Verständnis der zu zeigenden Aussage bestehen.

Die Zuordnung eines Beweisversuchs zur Kategorie *weder begründender noch prüfender Ansatz* weist darauf hin, dass eine, wenn auch eventuell nur kurze und oberflächliche, Auseinandersetzung mit der Aufgabenstellung erfolgt ist. Beweisversuche in dieser Kategorie können durchaus aufzeigen, dass Lernende um die Allgemeingültigkeit einer infrage stehenden Aussage wissen. Sie können ebenso andeuten, dass ihnen unklar war, was zu tun ist.

Die unterschiedenen Oberkategorien umfassen folglich jeweils ein großes Spektrum möglicher Beweisversuche. Um genauere Einblicke in die Beweisversuche in diesen Kategorien zu erhalten, werden sie in den folgenden Abschnitten weiter ausdifferenziert.

3.3.3 Begründender Ansatz: Beweisideen

Dieser und die nachfolgenden Abschnitte 3.3.4 bis 3.3.6 und 3.3.11 fokussieren auf Beweisversuche mit begründendem Ansatz. Im Rahmen einer detaillierten Analyse dieser Beweisversuche sollen nacheinander die Beweisideen, die Korrektheit und Vollständigkeit, die Schlussweisen, die Verwendung von Beispielen und die vorkommenden Darstellungsmittel beschrieben werden (Abschn. 3.3.1).

Die Beweisideen bilden das erste Merkmal von Beweisversuchen mit begründendem Ansatz, das im Rahmen der Analyse solcher Beweisversuche erfasst wird (Abb. 14). In diesem Abschnitt werden zuerst die Überlegungen erläutert, die zur Berücksichtigung der Beweisideen und zur Kategorienbildung beigetragen haben. Danach folgen die Kategoriendefinitionen und Hinweise zu den Kategorien. Die Ergebnisse der Analyse der Beweisideen bei Beweisversuchen mit begründendem Ansatz werden in Abschnitt 8.2.2 vorgestellt.

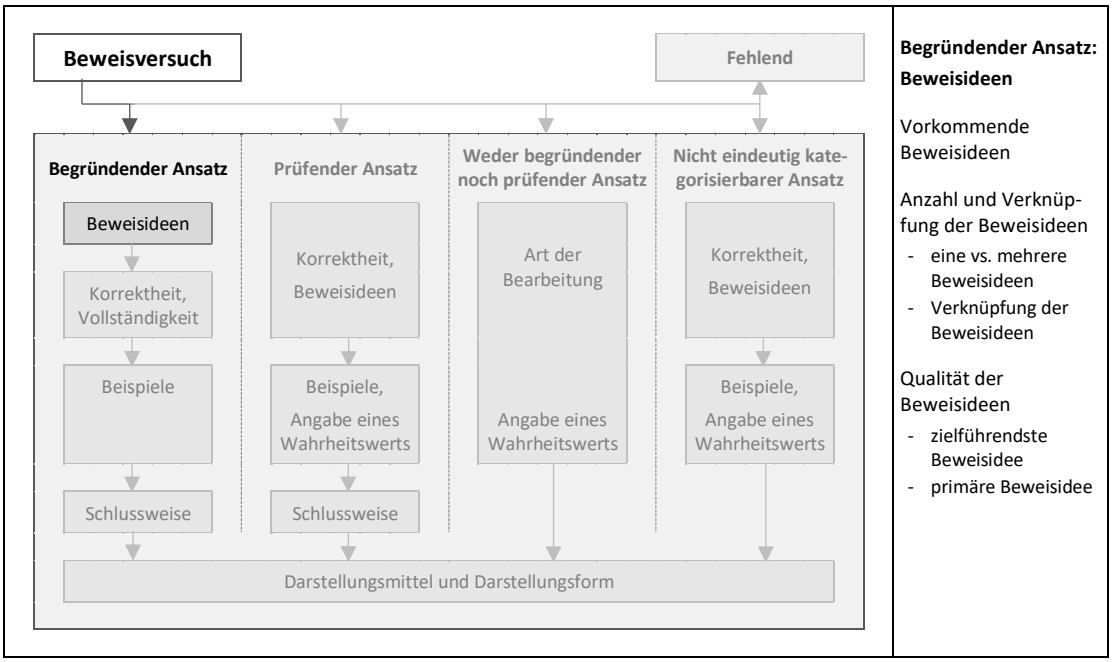


Abbildung 14: Analysemodell. Fokus: Begründender Ansatz, Beweisideen

3.3.3.1 Entwicklung der Kategorien

Nach dem vorgestellten Konzept der Argumentationsbasis von Fischer und Malle (1985), Malle (2002) basiert jeder Beweisversuch mit begründendem Ansatz auf verschiedenen Wissenselementen (Abschn. 3.3.2.1). Die Erfassung dieser Wissenselemente kann im Rahmen einer Analyse einzelner Beweisversuche detaillierte Einblicke in das Wissen und die Denkweise von Lernenden geben. Diese Vorgehensweise birgt jedoch neben dem hohen Zeitaufwand und den umfangreichen Ergebnissen die Schwierigkeit, dass in der Regel unklar ist, welche Wissenselemente und damit verbundenen Vorstellungen tatsächlich den explizierten Argumenten zugrunde lagen. Die in einem Beweisversuch genutzten Wissenselemente verbleiben teils implizit. Beispielsweise geht aus den nachfolgenden Bearbeitungen 3.3.3a-b (Abb. 15, 16) nicht klar hervor, welche Wissenselemente bei den dort zu sehenden Termumformungen mehr oder weniger bewusst mitgedacht wurden und welche Begriffsverständnisse den verwendeten Termini, wie „teilbar“ und „ungerade“, zugrunde lagen. Darüber hinaus können Beweisversuche, die sich in Bezug auf die Beweisidee unterscheiden, die Verwendung einer Vielzahl an gleichen oder ähnlichen Wissenselementen implizieren, so dass im Rahmen einer detaillierten Analyse die wesentlichen Unterschiede zwischen verschiedenen Beweisversuchen verwischen können. Zum Beispiel basieren beide Bearbeitungen 3.3.3a, 3.3.3b in den Abbildungen 15, 16 auf einem gewissen korrekten Verständnis des Ausdrucks „aufeinanderfolgende ungerade Zahlen“ sowie des Teilbarkeitsbegriffs. In beiden Bearbeitungen wird ausgenutzt, dass die Differenz zwischen zwei aufeinanderfolgenden ungeraden Zahlen zwei beträgt. Die Schülerin und der Schüler haben zudem gewisse Termumformungsregeln und Rechenverfahren korrekt angewendet. Dabei beruht die Bearbeitung 3.3.3b auf der Darstellung der infrage stehenden Summe in Abhängigkeit des kleinsten Summanden, während der Bearbeitung 3.3.3a vielmehr die Idee der stetigen Vergrößerung der Summe um acht und das Verfahren der vollständigen Induktion zugrunde liegen.

Um vor allem Einblicke in solche Kernideen der Lernenden zu erhalten, werden nicht en détail alle Wissenselemente, sondern Beweisideen erfasst. Eine Beweisidee wird als inhaltsbezogene Überlegung definiert, die dem Argumentgefüge zugrunde liegt und auf Voraussetzungen der infrage stehenden Aussage oder anderen mathematischen Sätze und Definitionen beruht. Die Zuordnung eines Beweisversuchs mit begründendem Ansatz zu einer Beweisidee impliziert im Allgemeinen zugleich die Nutzung bestimmter Wissenselemente, die im Rahmen einer anschließenden detaillierten Analyse genauer untersucht werden könnte.

Bei der Analyse von Beweisversuchen mit begründendem Ansatz werden alle vorkommenden Beweisideen erfasst. Dabei kann jedoch die Möglichkeit der Verwendung mehrerer Beweisideen innerhalb eines Beweisversuchs dazu führen, dass einzelne Beweisideen in einem sehr großen Anteil vorliegender Bearbeitungen auftauchen, obwohl sie oftmals ausschließlich von unter- oder nachgeordneter Bedeutung sind. Um diesen Einflussfaktor zu relativieren, wird ergänzend codiert, ob in den einzelnen Beweisversuchen mehrere Beweisideen zu finden und ggf. als ein einziges zusammenhängendes Argumentgefüge ausgeführt sind. Damit zudem die Qualität der verwendeten Beweisideen mehr Gewicht erhält, wird die zielführendsten Beweisidee, die in einem vorliegenden Beweisversuch vorkommt, kategorisiert und die Beweisidee, die den wesentlichen Ausgangspunkt oder den Rahmen des vorliegenden Argumentgefüges bildet, unter der Bezeichnung „primäre Beweisidee“ codiert.

3.3.3.2 Vorstellung der Kategorien

Unter einer Beweisidee wird in dieser Arbeit eine inhaltsbezogene Überlegung erfasst, die dem Argumentgefüge zugrunde liegt und auf Voraussetzungen der infrage stehenden Aussage oder anderen mathematischen Sätzen und Definitionen beruht. Welche Beweisideen sich unterscheiden lassen, ist von der zu zeigenden Aussage abhängig. Mit Blick auf die in dieser Arbeit fokussierte Teilbarkeit der Summe von vier aufeinanderfolgenden ungeraden Zahlen wurden vierzehn Beweisideen herausgearbeitet, die nachfolgend vorgestellt werden. Auf dieser Grundlage kann dann anhand der Daten analysiert werden, welche dieser Beweisideen in einem vorliegenden Beweisversuch mit begründendem Ansatz vorkommen. Darauf aufbauend lassen sich die Anzahl an verwendeten Beweisideen, deren Verknüpfung und Qualität codieren.

Unterschiedene Beweisideen

Ein Beweisversuch enthält die Beweisidee **BI1** ***sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante***, insbesondere wenn über die Invariante argumentiert wird, dass die Differenz zwischen der Summe von vier aufeinanderfolgenden ungeraden Zahlen und der Summe der nächstgrößeren (bzw. nächstkleineren) vier aufeinanderfolgenden ungeraden Zahlen acht bzw. ein Vielfaches von vier beträgt (Abb. 15). Ein Beweisversuch enthält ebenfalls die Beweisidee 1, wenn über die Invariante argumentiert wird, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer um ein Vielfaches von acht bzw. vier größer ist als die Summe der kleinsten vier aufeinanderfolgenden ungeraden natürlichen Zahlen. Außerdem kommt die Beweisidee 1 vor, wenn z. B. für die Teilbarkeit der Summe zweier aufeinanderfolgender ungerader Zahlen durch vier oder für die Teilbarkeit der Summe von vier gleichen ungeraden Zahlen durch vier analog argumentiert wird.

Bearbeitung 3.3.3a: Klassenstufe 8, männlich (Code: 0821301)

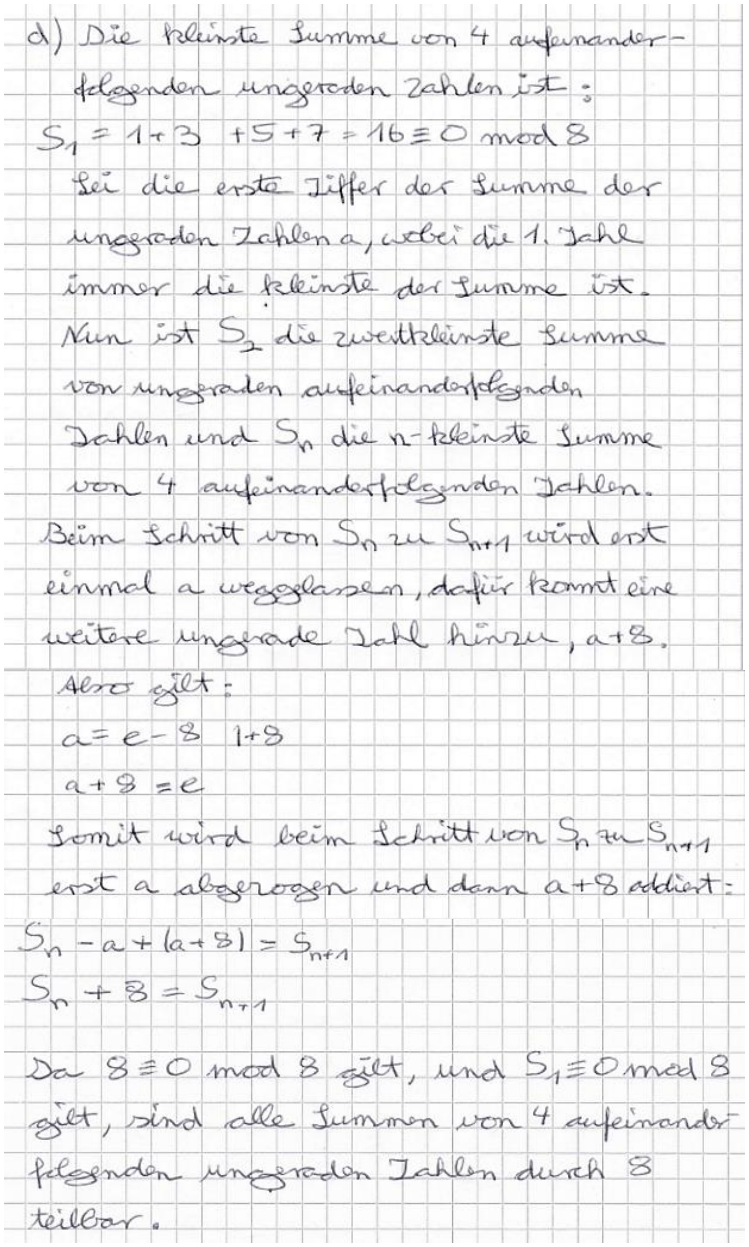


Abbildung 15: Bearbeitung 3.3.3b (BI1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*)

Ein Beweisversuch enthält die Beweisidee **BI2** ***Darstellung einer Summe in Abhängigkeit des kleinsten Summanden***, insbesondere wenn die Summe von vier aufeinanderfolgenden ungeraden Zahlen in Abhängigkeit des kleinsten Summanden dargestellt ist (Abb. 16). Es ist aber auch möglich, dass die Summe z. B. in Abhängigkeit des größten Summanden dargestellt ist. Zudem kommt die Beweisidee 2 vor, wenn etwa die Summe zweier aufeinanderfolgender ungerader Zahlen in Abhängigkeit einer dieser Zahlen dargestellt ist.

Bearbeitung 3.3.3b: Klassenstufe 8, weiblich (Code: 0811303)

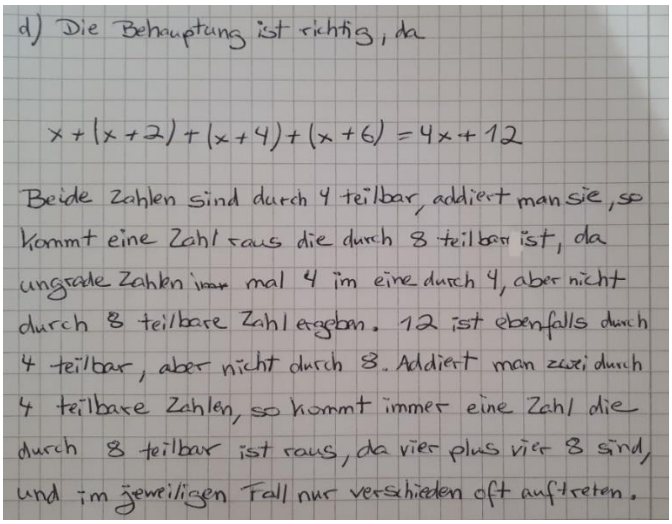


Abbildung 16: Bearbeitung 3.3.3b (BI2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*)

Ein Beweisversuch enthält die Beweisidee **BI3** ***Vielfaches einer mittleren Zahl oder Durchschnittsbildung***, insbesondere wenn die Ausführungen auf der Idee basieren, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen gleich dem Vierfachen der ganzen Zahl zwischen der zweitkleinsten und zweitgrößten der vier Zahlen bzw. gleich dem Vierfachen des arithmetischen Mittels dieser Zahlen ist (Abb. 17). Zudem enthält ein Beweisversuch die Beweisidee 3, wenn z. B. für die Teilbarkeit der Summe zweier aufeinanderfolgender ungerader Zahlen durch vier analog argumentiert wird.

Bearbeitung 3.3.3c: Klassenstufe 8, männlich (Code: 0810506)

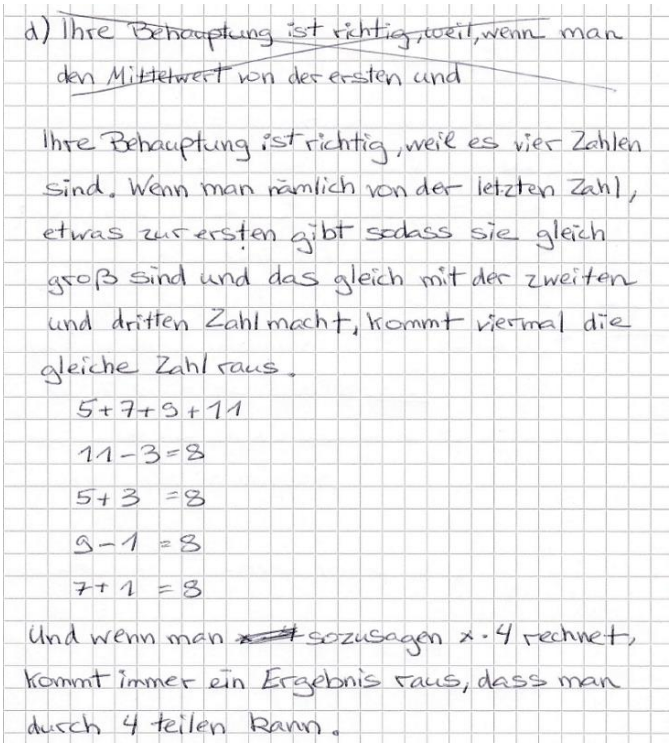


Abbildung 17: Bearbeitung 3.3.3c (BI3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*)

Ein Beweisversuch enthält die Beweisidee **BI4** ***Zahldarstellung im Dezimalsystem***, insbesondere wenn die Summe von vier aufeinanderfolgenden ungeraden Zahlen als Vielfaches von 40 plus eine der Summen $1 + 3 + 5 + 7$, $3 + 5 + 7 + 9$, $5 + 7 + 9 + 11$, $7 + 9 + 11 + 13$, $9 + 11 + 13 + 15$ dargestellt wird.[39] Allgemein enthält ein Beweisversuch die Beweisidee 4, wenn Eigenschaften des Dezimalsystems genutzt werden, d. h. beispielsweise wenn nur die Einerziffern betrachtet werden, sei es von einzelnen Summanden oder der Summe von vier aufeinanderfolgenden ungeraden Zahlen (Abb. 18).

Bearbeitung 3.3.3d: Klassenstufe 8, weiblich (Code: 0811224)

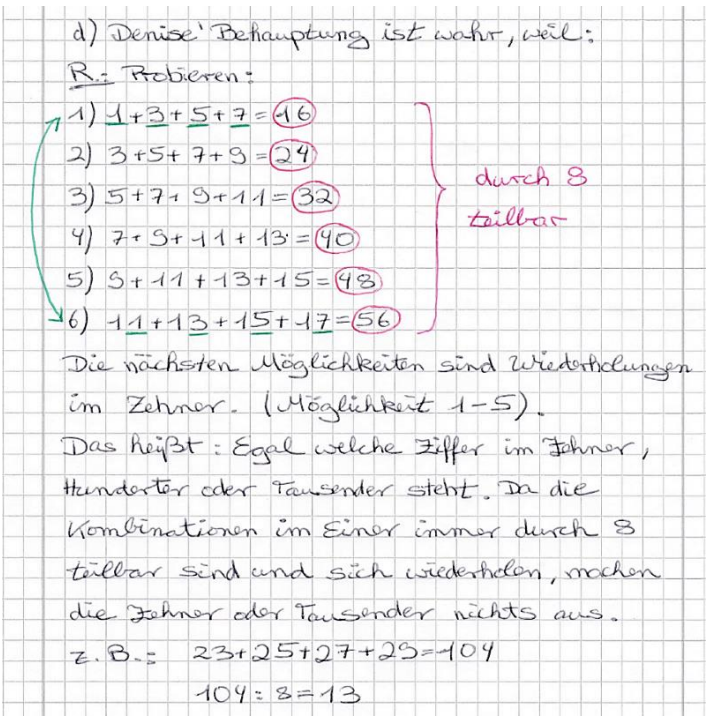


Abbildung 18: Bearbeitung 3.3.3d (BI4 *Zahldarstellung im Dezimalsystem*)

Ein Beweisversuch enthält die Beweisidee **BI5** ***Reste bei Division einzelner Summanden durch vier oder acht***, insbesondere wenn Ausführungen auf Überlegungen zur Teilbarkeit der einzelnen vier aufeinanderfolgenden ungeraden Zahlen durch vier oder acht beruhen. Dabei ist es sowohl möglich, dass mithilfe modularer Arithmetik argumentiert wird, als auch, dass vier aufeinanderfolgende ungerade Zahlen jeweils als Vielfache von vier oder acht plus eine der Ziffern 1, 3, 5 oder 7 dargestellt werden (Abb. 19). Zudem enthält ein Beweisversuch die Beweisidee 5, wenn z. B. für die Teilbarkeit von vier geraden aufeinanderfolgenden ungeraden Zahlen durch vier oder acht analog argumentiert wird. Eine Überlegung, inwieweit eine einzige gerade Zahl durch vier oder acht teilbar ist, sowie die Aussage, dass vier Zahlen addiert werden, die nicht in der Achter- oder Viererreihe vorkommen, reichen hingegen nicht aus, um einem Beweisversuch die Beweisidee 5 zuzuordnen.

[39] Konkret ist es beispielsweise möglich, die kleinste von vier aufeinanderfolgenden ungeraden natürlichen Zahlen als $10a + e$ mit $a \in \mathbb{N}$ und $e \in \{1; 3; 5; 7; 9\}$ darzustellen. Mithilfe einer vollständigen Fallunterscheidung und unter Berücksichtigung, dass die Differenz zwischen zwei aufeinanderfolgenden ungeraden Zahlen gleich zwei ist, folgt, dass die Summe von vier aufeinanderfolgenden ungeraden natürlichen Zahlen $40a + z$ mit $z \in \{16; 24; 32; 40; 48\}$ beträgt. Da beide Summanden in allen Fällen durch acht teilbar sind, ist deren Summe stets durch acht teilbar.

Bearbeitung 3.3.3e: Klassenstufe 8, männlich (Code: 0811808)

Abbildung 19: Bearbeitung 3.3.3e (BI5 *Reste bei Division einzelner Summanden durch vier oder acht*)

Ein Beweisversuch enthält die Beweisidee **BI6** ***Teilbarkeit der Summen zweier Zahlenpaare***, insbesondere wenn er auf der Idee beruht, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen gleich der Summe zweier dieser Zahlen plus der Summe der anderen beiden Zahlen ist, und die Teilbarkeit durch zwei, vier oder acht der beiden Summen betrachtet wird (Abb. 20). Zudem enthält ein Beweisversuch die Beweisidee 6, wenn z. B. für die Teilbarkeit der Summe von vier aufeinanderfolgenden geraden Zahlen durch vier analog argumentiert wird. Im Grenzfall wird einem Beweisversuch auch die Beweisidee 6 zugeordnet, wenn geschrieben steht, dass die Summe zweier aufeinanderfolgender ungerader Zahlen durch vier teilbar ist bzw. bei Division durch acht Rest null oder vier lässt.

Bearbeitung 3.3.3f: Klassenstufe 8, männlich (Code: 0812202)

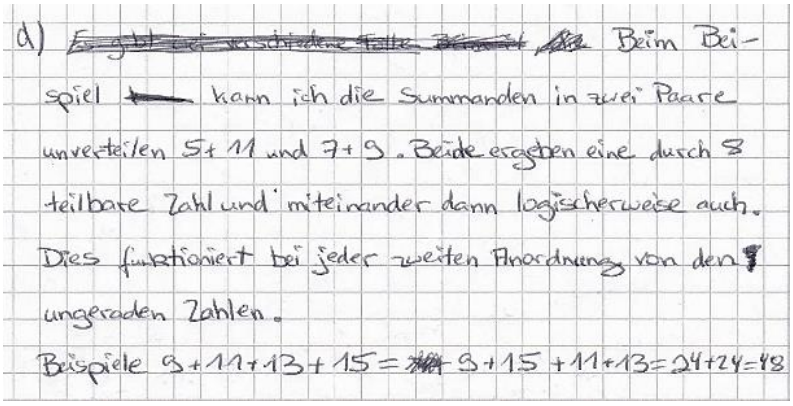


Abbildung 20: Bearbeitung 3.3.3f (BI6 *Teilbarkeit der Summen zweier Zahlenpaare*)

Ein Beweisversuch enthält die Beweisidee **BI7** ***Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen***, insbesondere wenn die Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen durch vier oder acht thematisiert wird (Abb. 20).[40] Zudem enthält ein Beweisversuch die Beweisidee 7, wenn die Teilbarkeit von vier geraden, nicht unbedingt aufeinanderfolgenden Zahlen durch vier oder acht thematisiert wird. Es reicht hingegen nicht aus, wenn nur darauf eingegangen wird, inwieweit eine einzige gerade Zahl durch vier oder acht teilbar ist.

Bearbeitung 3.3.3g: Klassenstufe 8, weiblich (Code: 0820303)

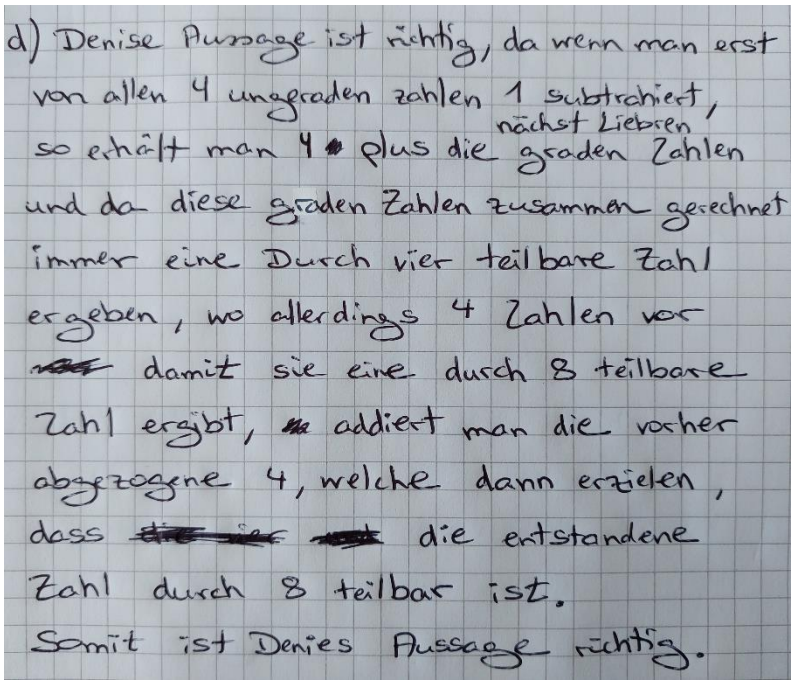


Abbildung 21: Bearbeitung 3.3.3g (BI7 *Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen*)

[40] Zum Beispiel über die Betrachtung der Teilbarkeit der einzelnen vier aufeinanderfolgenden geraden Zahlen (BI5) oder mithilfe der Darstellung der Summe in Abhängigkeit der kleinsten dieser Zahlen (BI2) kann bewiesen werden, dass die Summe von vier aufeinanderfolgenden geraden Zahlen gleich einem Vielfachen der Zahl acht plus vier ist. Da jede Summe von vier aufeinanderfolgenden ungeraden Zahlen um vier größer bzw. kleiner ist als eine Summe von vier aufeinanderfolgenden geraden Zahlen, folgt die gewünschte Behauptung.

Ein Beweisversuch enthält die Beweisidee **BI8** ***Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen***, wenn verwendet wird, dass eine ungerade Zahl gleich einer geraden Zahl plus bzw. minus eins ist oder dass eine ungerade Zahl gleich dem Doppelten einer ganzen Zahl plus bzw. minus eins ist (Abb. 22). Für die Zuordnung der Beweisidee 8 ist es dementsprechend nicht erforderlich, dass die Summe von mehreren ungeraden Zahlen betrachtet wird. Die ausschließliche Überlegung, dass zwischen zwei ungeraden Zahlen eine gerade Zahl liegt, reicht für die Zuordnung jedoch nicht aus.

Bearbeitung 3.3.3h: Klasse 8, männlich (Code: 0820904)

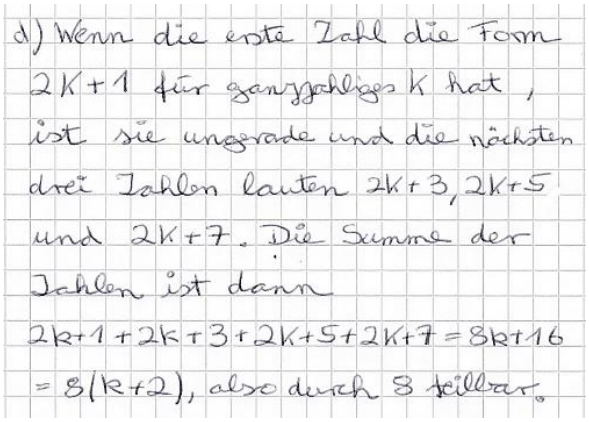


Abbildung 22: Bearbeitung 3.3.3h (BI8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*)

Ein Beweisversuch enthält die Beweisidee **BI9** ***Parität von Summen ungerader oder gerader Zahlen***, wenn Kenntnisse über die Parität von Summen gerader oder ungerader Zahlen als wesentliche inhaltsbezogene Überlegungen[41] verwendet werden (Abb. 23). Im Grenzfall wird Beweisversuchen auch die Beweisidee 9 zugeordnet, wenn Kenntnisse über die Parität von Produkten ungerader oder gerader Zahlen als wesentliche inhaltsbezogene Überlegungen verwendet werden.

Bearbeitung 3.3.3i: Klassenstufe 7, männlich (Code: 0712204)

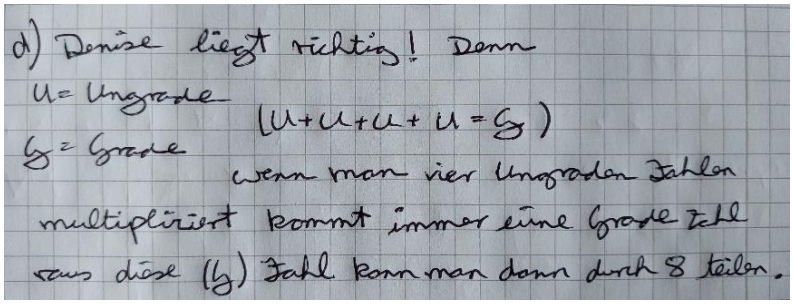


Abbildung 23: Bearbeitung 3.3.3i (BI9 *Parität von Summen ungerader oder gerader Zahlen*)

[41] Der in den Definitionen der Beweisideen 9 bis 13 vorkommende Ausdruck „wesentliche inhaltsbezogene Überlegung“ meint, dass die Wissenselemente nicht nur als Mittel zum Zweck beiläufig Verwendung finden, sondern als ein entscheidender Grund angeführt sind, warum eine Aussage gilt. Bei der Aussage kann es sich um die zu zeigende Aussage, um eine nicht ganz korrekte Interpretation der zu zeigenden Aussage oder um eine im Beweisversuch verwendete Aussage, wie z. B. dass die Summe zweier aufeinanderfolgender ungerader Zahlen durch vier teilbar ist, handeln. Die als wesentliche inhaltsbezogene Überlegung verwendeten Wissenselemente können falsch oder fehlerhaft angewendet sein.

Ein Beweisversuch enthält die Beweisidee **BI10** ***Abstände ungerader Zahlen***, wenn als eine wesentliche inhaltsbezogene Überlegung auf den Abstand zwischen ungeraden Zahlen Bezug genommen wird (Abb. 24). Dabei ist es egal, ob auf den Abstand zwischen zwei aufeinanderfolgenden ungeraden Zahlen oder z. B. auf den Abstand zwischen der kleinsten und der größten von vier aufeinanderfolgenden ungeraden Zahlen eingegangen wird. Im Grenzfall wird einem Beweisversuch auch die Beweisidee 10 zugeordnet, wenn als eine wesentliche inhaltsbezogene Überlegung auf den Abstand zwischen zwei geraden Zahlen Bezug genommen wird.

Bearbeitung 3.3.3j: Klassenstufe 7, männlich (Code: 0720520)

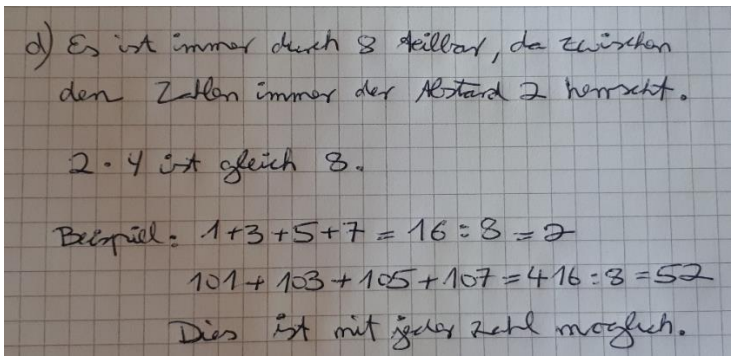


Abbildung 24: Bearbeitung 3.3.3j (BI10 *Abstände ungerader Zahlen*)

Ein Beweisversuch enthält die Beweisidee **BI11** ***vier Zahlen***, insbesondere wenn als eine wesentliche inhaltsbezogene Überlegung verwendet wird, dass es sich um vier Summanden handelt (Abb. 25). Allgemein enthält ein Beweisversuch die Beweisidee 11, wenn der Anzahl an Zahlen, die addiert werden, eine entscheidende Bedeutung zugewiesen wird. Einem Beweisversuch wird demzufolge u. a. auch im Falle von Verallgemeinerungen der zu zeigenden Aussage, wie etwa „Die Summe von vier ungeraden Zahlen ist immer durch vier teilbar“ oder „Die Summe von n ($n \in \mathbb{N}$) aufeinanderfolgenden ungeraden Zahlen ist durch n teilbar“, die Beweisidee 11 zugeordnet.

Bearbeitung 3.3.3k: Klassenstufe 7, männlich (Code: 0721101)

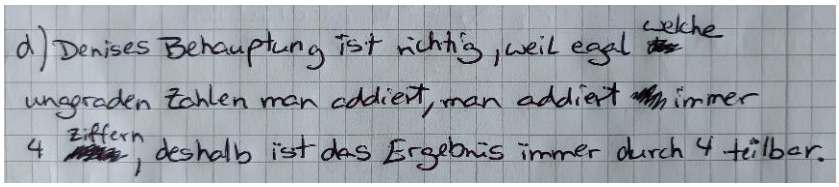


Abbildung 25: Bearbeitung 3.3.3k (BI11 *vier Zahlen*)

Ein Beweisversuch enthält die Beweisidee **BI12** ***Vierer- oder Achterreihe***, wenn als eine wesentliche inhaltsbezogene Überlegung angeführt ist, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer eine Zahl der Achter- bzw. Viererreihe ist oder dass die einzelnen Summanden nicht in der Achter- bzw. Viererreihe sind (Abb. 26). Wenn ungerade Zahlen als Vielfache von vier oder acht plus einem Rest dargestellt werden oder deren Reste bei Division durch vier bzw. acht betrachtet werden, wird hingegen als Beweisidee 5 Reste bei Division einzelner Summanden durch vier oder acht angesehen.

Bearbeitung 3.3.3l: Klassenstufe 7, weiblich (Code: 0711810)

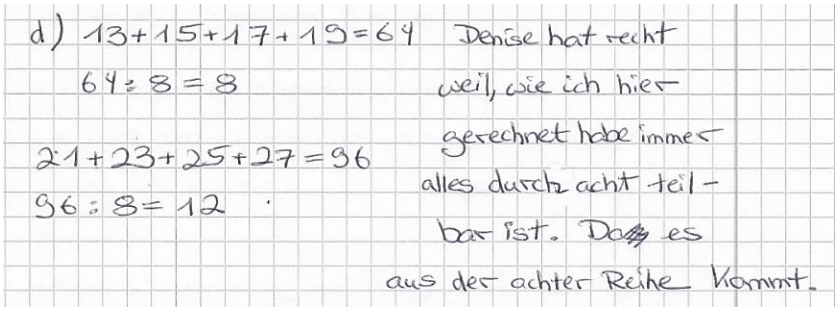


Abbildung 26: Bearbeitung 3.3.3l (BI12 *Vierer- oder Achterreihe*)

Ein Beweisversuch enthält die Beweisidee **BI13** ***Regeln zur Teilbarkeit von Produkten oder Summen durch vier oder acht***, wenn als eine wesentliche inhaltsbezogene Überlegung Kenntnisse zur Teilbarkeit von Produkten oder Summen durch vier oder acht expliziert sind. Dies trifft z. B. zu, wenn angeführt ist, dass das Vierfache einer geraden Zahl durch acht teilbar ist oder dass das Doppelte einer durch vier teilbaren Zahl durch acht teilbar ist (Abb. 27).

Bearbeitung 3.3.3m: Klassenstufe 7, weiblich (Code: 0711207)

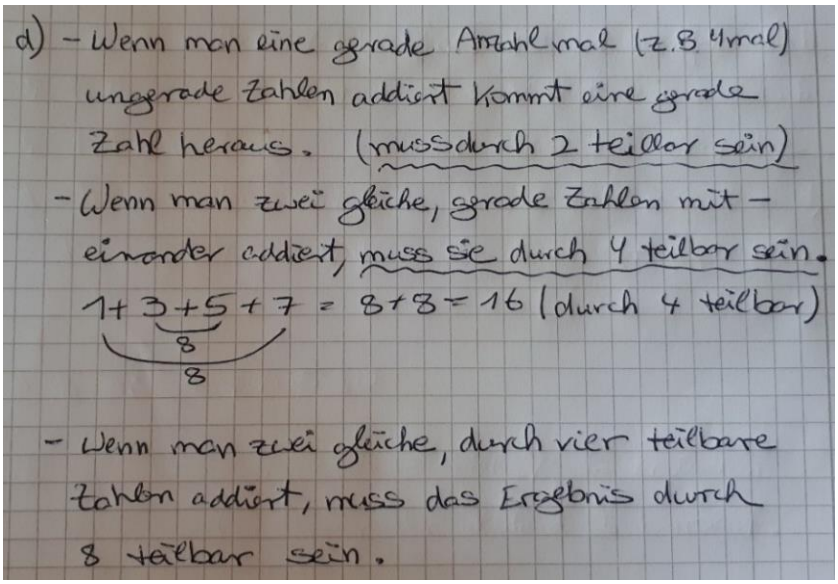


Abbildung 27: Bearbeitung 3.3.3m (BI13 *Regeln zur Teilbarkeit von Produkten oder Summen durch vier oder acht*)

Ein Beweisversuch enthält eine ***nicht eindeutig kategorisierbare Beweisidee*** (**BI55**), wenn eine Beweisidee vorkommt, die keiner der beschriebenen Beweisideen 1 bis 13 klar zugeordnet werden kann. Dies kann insbesondere vorkommen, wenn eine zu zeigende Aussage nicht ganz korrekt interpretiert wird und wenn die Ausführungen sehr viel Interpretationsspielraum lassen. Zudem ist es möglich, dass eine andere als die angeführten Beweisideen 1 bis 13 genutzt wird, wie z. B. dass die Summe von vier aufeinanderfolgenden ungeraden natürlichen Zahlen immer größer als acht ist oder dass eine Zahl durch acht teilbar sei, wenn sie durch zwei und durch vier teilbar ist (Abb. 28). Dabei gilt es zu bedenken, dass es sich nicht um eine Beweisidee (d. h. insbesondere auch nicht um eine nicht eindeutig kategorisierbare Beweisidee)

handelt, wenn im Wesentlichen die zu zeigende Aussage selbst oder ihre zu beweisende Allgemeingültigkeit als ein Grund für die Gültigkeit der Aussage angeführt ist (z. B. „weil die Summe immer durch acht teilbar ist“, „weil es kein Gegenbeispiel gibt“).

Bearbeitung 3.3.3n: Klassenstufe 8, weiblich (Code: 0821306)

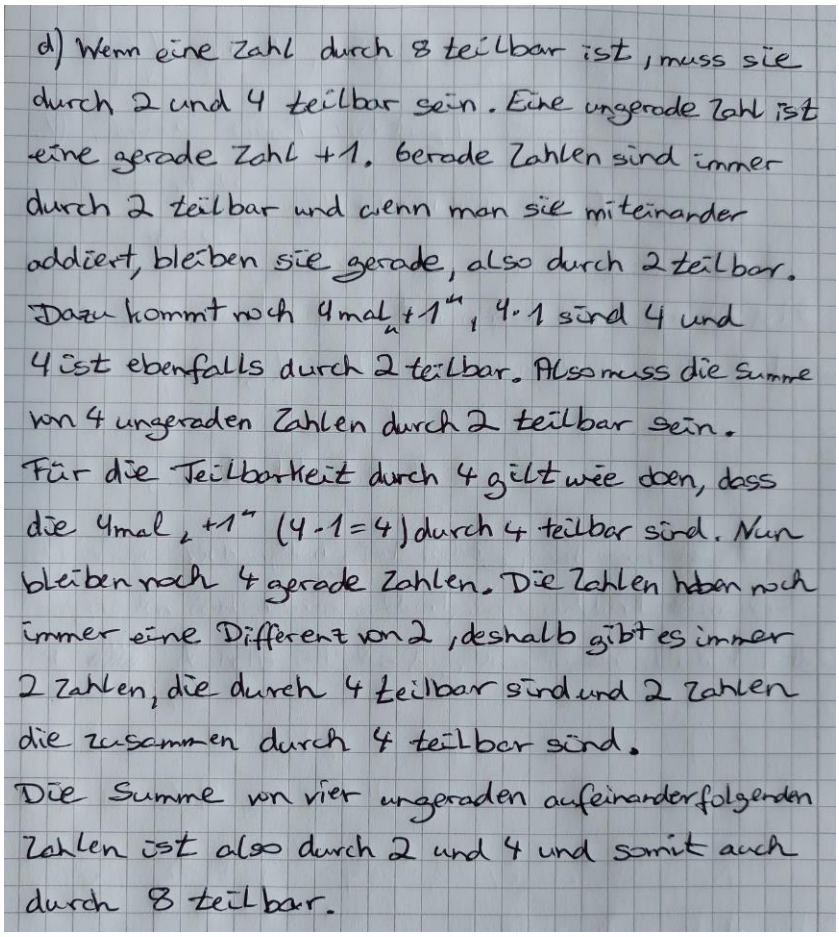
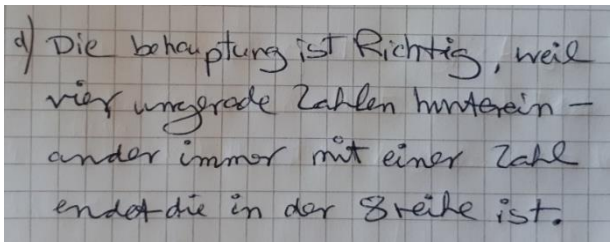


Abbildung 28: Bearbeitung 3.3.3n (BI55 *nicht eindeutig kategorisierbar*)

Vorkommende Beweisideen

In der Analyse der Beweisideen bei einem vorliegenden Beweisversuch mit begründendem Ansatz wird zuerst untersucht, ob die einzelnen unterschiedenen Beweisideen 1 bis 13, 55 vorkommen. Beispielsweise ist in der nachfolgenden Bearbeitung 3.3.3o in Abbildung 29 die Beweisidee 12 *Vierer- oder Achterreihe* angeführt, alle anderen Beweisideen sind nicht zu erkennen. In der Bearbeitung 3.3.3p kommen genau die Beweisideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, 10 *Abstände ungerader Zahlen* und 11 *vier Zahlen* vor. Die Bearbeitung 3.3.3q enthält genau die Beweisideen 6 *Teilbarkeit der Summen zweier Zahlenpaare* und 9 *Parität von Summen gerader oder ungerader Zahlen*.

Bearbeitung 3.3.3o: Klassenstufe 7, männlich (Code: 0711816)



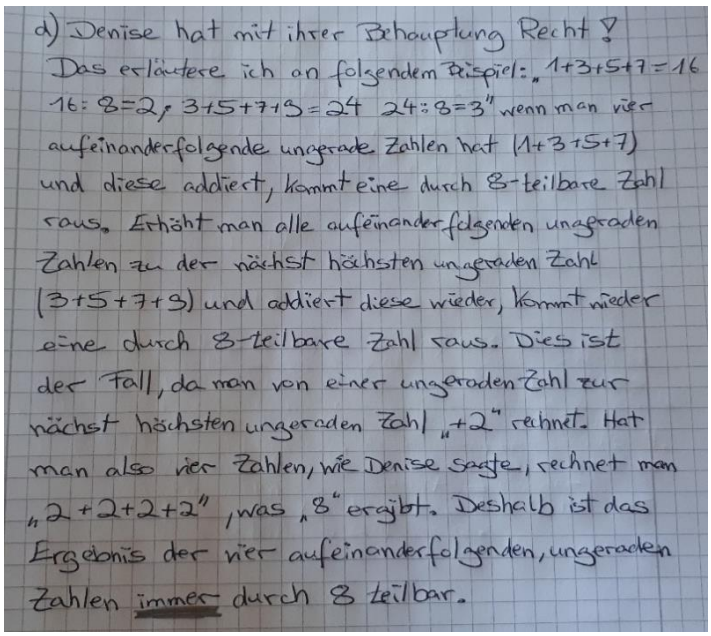
Begründender Ansatz: Beweisideen

Vorkommende Beweisideen (BI)
- BI12: 1 *kommt vor*
- Bei allen anderen BI jeweils: 0 *kommt nicht vor*

Anzahl, Verknüpfung der Beweisideen
- BIAnz1 *genau eine Beweisidee*
- BIVerkn88 *nicht relevant*

Qualität der Beweisideen
- BIziel1 *keine ausbaubare Beweisidee*
- BIp12 *Vierer- oder Achterreihe*

Bearbeitung 3.3.3p: Klassenstufe 8, weiblich (Code: 0811302)



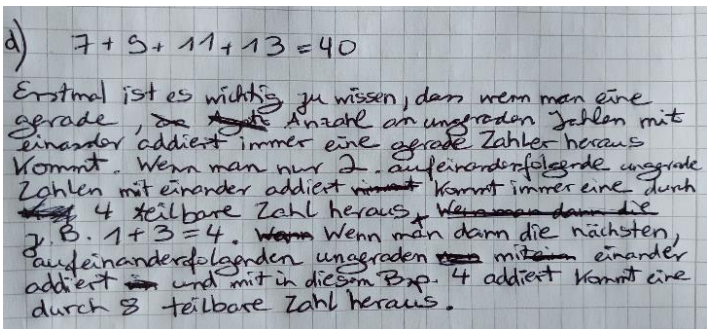
Begründender Ansatz: Beweisideen

Vorkommende Beweisideen (BI)
- BI1: 1 *kommt vor*
- BI10: 1 *kommt vor*
- BI11: 1 *kommt vor*
- Bei allen anderen BI jeweils: 0 *kommt nicht vor*

Anzahl, Verknüpfung der Beweisideen
- BIAnz2 *mehrere Beweisideen*
- BIVerkn1 *ein einziges zusammenhängendes Argumentgefüge*

Qualität der Beweisideen
- BIziel3 *zentrale Beweisidee*
- BIp1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*

Bearbeitung 3.3.3q: Klassenstufe 7, männlich (Code: 0721211)



Begründender Ansatz: Beweisideen

Vorkommende Beweisideen (BI)
- BI6: 1 *kommt vor*
- BI9: 1 *kommt vor*
- Bei allen anderen BI jeweils: 0 *kommt nicht vor*

Anzahl, Verknüpfung der Beweisideen
- BIAnz2 *mehrere Beweisideen*
- BIVerkn0 *eventuell mehrere Argumentgefüge*

Qualität der Beweisideen
- BIziel2 *ausbaubare Beweisidee*
- BIp6 *Teilbarkeit der Summen zweier Zahlenpaare*

Abbildung 29: Bearbeitungen 3.3.3p-q (Beweisversuche mit begründendem Ansatz: Analyse der Beweisideen)

Anzahl und Verknüpfung der Beweisideen

Auf der Analyse der vorkommenden Beweisideen aufbauend wird erfasst, ob mehrere Beweisideen zu erkennen sind. Wenn ein Beweisversuch mehrere Beweisideen enthält, wird zudem erfasst, ob alle Beweisideen als ein einziges zusammenhängendes Argumentgefüge ausgeführt sind. Sobald nicht ganz klar ist, ob es sich um ein einziges zusammenhängendes Argumentgefüge handelt, wird dabei die Kategorie BIVerkn0 *eventuell mehrere Argumentgefüge* gewählt. Falls ausschließlich genau eine Beweisidee vorkommt, wird 88 *nicht relevant* notiert. Beispielsweise ist in der Bearbeitung 3.3.3o in obiger Abbildung 29 ausschließlich eine der unterschiedenen Beweisideen zu erkennen, die Frage nach der Verknüpfung der Beweisideen ist also nicht relevant. In den Bearbeitungen 3.3.3p-q kommen mehrere Beweisideen vor. Dabei sind in der Bearbeitung 3.3.3p alle Beweisideen als ein einziges zusammenhängendes Argumentgefüge ausgeführt, während es sich bei den Ausführungen in der Bearbeitung 3.3.3q eher um zwei Argumentgefüge handelt.

Qualität der Beweisideen

Abschließend werden als zwei Anhaltspunkte für die Qualität der vorkommenden Beweisideen die zielführendste Beweisidee kategorisiert und die primäre Beweisidee codiert.

Zur Beschreibung der **zielführendsten Beweisidee** werden die Kategorien BIziel3 *zentrale Beweisidee*, BIziel2 *ausbaubare, aber keine zentrale Beweisidee* und BIziel1 *keine ausbaubare Beweisidee* differenziert (Bearbeitungen 3.3.3o-q in Abb. 29). Als zentrale Beweisideen werden die Beweisideen eins bis fünf bezeichnet, weil für ein korrektes und vollständiges Argumentgefüge in der Regel irgendeine dieser Beweisideen verwendet werden muss.[42] Ein Beweisversuch gehört dementsprechend in die Kategorie BIziel3 *zentrale Beweisidee*, wenn eine der Beweisideen eins bis fünf vorkommt. Ein Beweisversuch gehört in die Kategorie BIziel2 *ausbaubare, aber keine zentrale Beweisidee*, wenn keine der Beweisideen eins bis fünf, aber eine der Beweisideen sechs bis acht vorkommt. Jede dieser Beweisideen geht deutlich über die zu zeigende Aussage hinaus und kann zudem den Rahmen oder den wesentlichen Ausgangspunkt eines korrekten und vollständigen Argumentgefüges bilden, wenngleich in der Regel zusätzlich eine der zentralen Beweisideen erforderlich ist. Wenn ein Beweisversuch keine der Beweisideen eins bis acht enthält, gehört er in die Kategorie BIziel1 *keine ausbaubare Beweisidee*.

Von den in einem vorliegenden Beweisversuch vorkommenden Beweisideen wird diejenige, die mit Blick auf den Nachweis der zu zeigenden Aussage am grundlegendsten ist, d. h. den wesentlichen Ausgangspunkt oder den Rahmen des vorliegenden Argumentgefüges bildet, als **primäre Beweisidee** bezeichnet (→ Codes BIp1 bis BIp13, BIp55 entsprechend den Definiti-

[42] Wenn nicht die Teilbarkeit der Summe von vier aufeinanderfolgenden ungeraden Zahlen durch acht, sondern nur deren Teilbarkeit durch vier gezeigt werden muss, besteht die Möglichkeit, einen Beweis basierend auf der Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* ohne eine der Beweisideen 1 bis 5 zu formulieren. Ansonsten ist aber auch für den Beweis der Teilbarkeit der Summe von vier aufeinanderfolgenden ungeraden Zahlen durch vier stets eine der Beweisideen 1 bis 5 erforderlich. Daher erscheint die Bezeichnung der Beweisideen 1 bis 5 als zentrale Beweisideen trotzdem sinnvoll. Zur Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz siehe ergänzend Abschnitt 3.3.4.

onen der Beweisideen). Wenn in einem Beweisversuch mit begründendem Ansatz ausschließlich eine Beweisidee vorkommt, handelt es sich dabei zugleich um die primäre Beweisidee. Wenn mehrere Beweisideen miteinander verknüpft oder ineinander verschachtelt sind, wird die übergeordnete Beweisidee als primäre Beweisidee angesehen, d. h. diejenige, die den Rahmen des Argumentgefüges oder zumindest dessen wesentlichen Ausgangspunkt bildet. Wenn mehrere Beweisideen nicht klar als ein zusammenhängendes Argumentgefüge ausgeführt sind, wird das Argumentgefüge fokussiert, das mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage am besten ausgeführt ist.[43]

Beispielsweise ist in der Bearbeitung 3.3.3o in Abbildung 29 ausschließlich die Beweisidee 12 *Vierer- bzw. Achterreihe* zu erkennen, die folglich zugleich die primäre Beweisidee darstellt. In der Bearbeitung 3.3.3p ist die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* die übergeordnete und somit die primäre Beweisidee. Bei der Bearbeitung 3.3.3q ist das Argumentgefüge basierend auf der Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* am besten, die Beweisidee 6 ist also die primäre Beweisidee.

3.3.3.3 Hinweise zu den Kategorien

Bei der Analyse der Beweisideen bei Beweisversuchen mit begründendem Ansatz werden nacheinander alle vorkommenden Beweisideen, deren Anzahl und Verknüpfung sowie die zielführendste Beweisidee und die primäre Beweisidee erfasst. Entsprechend den Ausführungen zur zielführendsten Beweisidee unterscheiden sich die Beweisideen u. a. hinsichtlich ihrer Ausbaubarkeit zu einem korrekten und vollständigen Argumentgefüge.

Nur unter Verwendung einer der Beweisideen eins bis fünf (1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*, 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*, 4 *Zahldarstellung im Dezimalsystem*, 5 *Reste bei Division einzelner Summanden durch vier oder acht*) ist ein korrektes und vollständiges Argumentgefüge möglich. Dabei sind Beweisideen, die mithilfe elementarer Heurismen und inhaltlich-anschaulicher Überlegungen gut zugänglich sind, oder unter Verwendung von Algebra als offensichtlich erscheinen, zu Beginn genannt. Andere Beweisideen, die vom erforderlichen Argumentgefüge her komplexer sind, aus Expertensicht weniger intuitiv erscheinen oder die Verwendung in der Schule nicht unbedingt thematisierter Wissenselemente und Methoden, wie modulare Arithmetik, nahelegen, sind eher an späterer Stelle angeführt. Jede der Beweisideen eins bis fünf geht deutlich über die zu zeigende Behauptung, ihre Voraussetzungen und darin enthaltenen Begrifflichkeiten hinaus.

Die Beweisideen sechs bis acht (6 *Teilbarkeit der Summe zweier Zahlenpaare*, 7 *Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen*, 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*) können ebenfalls den wesentlichen Ausgangspunkt oder Rahmen eines korrekten und vollständigen Argumentgefüges bilden. Zudem gehen auch sie über die zu zeigende Aussage hinaus. Ihre Entwicklung kann etwa durch das Zerlegungsprinzip oder die Rückführung

[43] Es handelt sich um das Argumentgefüge, das bei der Analyse der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges der Einordnung als korrektes und vollständiges Argumentgefüge am nächsten ist. Zur Analyse der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges siehe Abschnitt 3.3.4.

von Unbekanntem auf Bekanntes unterstützt werden. Um auf Grundlage einer der Beweisideen sechs bis acht ein korrektes und vollständiges Argumentgefüge zu formulieren, ist jedoch zusätzlich mindestens eine der zentralen Beweisideen eins bis fünf erforderlich.

Die Beweisideen neun bis zwölf (9 *Parität von Summen ungerader oder gerader Zahlen*, 10 *Abstände ungerader Zahlen*, 11 *vier Zahlen*, 12 *Vierer- oder Achterreihe*) gehen nicht deutlich über die in der zeigenden Aussage enthaltenen Begrifflichkeiten und deren aus dem Schulunterricht i. d. R. gut bekannten Eigenschaften hinaus. Alleine mit diesen Beweisideen kann allenfalls ein Ansatz für ein weiterentwickelbares Argumentgefüge formuliert werden.[44] Trotzdem sind Beweisversuche mit einer der ausbaubaren Beweisideen eins bis acht (BIziel2-3) oder im Speziellen einer der zentralen Beweisideen eins bis fünf (BIziel3) hinsichtlich der Korrektheit und Vollständigkeit der Argumentgefüge nicht automatisch besser als solche, in denen ausschließlich eine oder mehrere der übrigen Beweisideen neun bis dreizehn oder nicht eindeutig kategorisierbare Beweisideen vorkommen (BIziel1).

Die Analyse der Beweisideen gibt Einblicke in die vorhandenen Kompetenzen und Denkweisen von Schülerinnen und Schülern. Vor allem wenn das Argumentgefüge weitgehend vollständig ist, impliziert das Vorkommen bestimmter Beweisideen die Verwendung bestimmter Wissenselemente. Zudem sind die Beweisideen in Zusammenhang mit dem genutzten Beweisverfahren zu sehen. Ein korrektes und vollständiges Argumentgefüge mit der primären Beweisidee der Vergrößerung der Summe um acht basiert auf dem Verfahren der vollständigen Induktion, auch wenn dieses Lernenden bei der Bearbeitung eventuell nicht bekannt ist. Fast alle anderen primären Beweisideen legen die Methode des direkten Beweises nahe. Weiterhin impliziert insbesondere die Nutzung einer der Beweisideen eins bis sechs die Anwendung eines heuristischen Prinzips, wie z. B. des Invarianzprinzips oder des Zerlegungsprinzips, oder eine Reformulierung in der Darstellung, und weist somit auf gewisse heuristische Kompetenzen hin. Darüber hinaus ist die Erfassung der Beweisideen, insbesondere der primären Beweisideen, bei der Beschreibung der Vollständigkeit des Argumentgefüges unerlässlich (Abschn. 3.3.4).

3.3.4 Begründender Ansatz: Korrektheit und Vollständigkeit

Beweisversuche mit begründendem Ansatz werden in einem zweiten Schritt hinsichtlich ihrer Korrektheit und Vollständigkeit[45] untersucht (Abb. 30). Dafür sind einige theoretische Grundlagen erforderlich, die im Rahmen einer Beschreibung des Vorgehens bei der Entwicklung der Kategorien zur Erfassung der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz sukzessive eingeführt werden. Anschließend folgen Kategoriendefinitionen zusammen mit exemplarischen Beweisversuchen von Lernenden und ergänzende Hinweise zu den Kategorien. Die Ergebnisse der Erfassung der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz werden in Abschnitt 8.2.1 präsentiert.

[44] Diesbezüglich eine Ausnahme bildet die Beweisidee 11 *vier Zahlen*, weil davon ausgehend ein korrektes und vollständiges Argumentgefüge möglich, wenn auch von Lernenden in den Klassenstufen 5 bis 8 nicht zu erwarten ist (Abschn. 6.2.2.2).

[45] Die Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz werden gemeinsam betrachtet, insbesondere da es im Falle gravierender Fehler, wie beispielsweise eines Zirkelschlusses, weder möglich noch sinnvoll ist, die Vollständigkeit des Argumentgefüges differenziert zu analysieren.

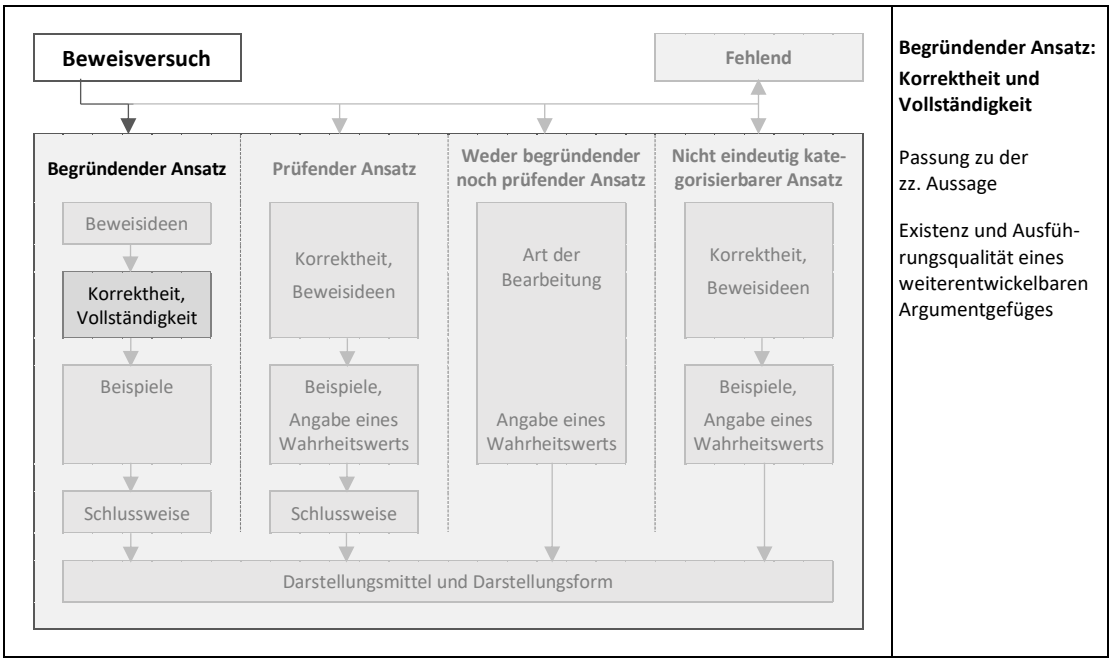


Abbildung 30: Analysemodell. Fokus: Begründender Ansatz, Korrektheit und Vollständigkeit

3.3.4.1 Entwicklung der Kategorien

Um Beweisversuche mit begründendem Ansatz hinsichtlich ihrer Korrektheit und Vollständigkeit zu beschreiben, wird zunächst das von **Heinze und Reiss (2003) sowie Ufer et al. (2009)** beschriebene Konzept herangezogen, weil in diesem mehrere Aspekte der Korrektheit und Vollständigkeit von Beweisversuchen von zentraler Bedeutung sind. Das Konzept bezieht sich primär auf geometrische Beweiskompetenzen von Gymnasiastinnen und Gymnasiasten ab der Klassenstufe 7. Ihre Sichtweise auf den Beweisbegriff fassen Ufer et al. (2009) zusammen:

> Mit den Einschränkungen bezüglich der axiomatischen Absicherung sowie der formal-sprachlichen Formulierung werden […] an die von Schülern im Rahmen des Unterrichts erbrachten Argumentationen letztlich ähnliche Maßstäbe angelegt, wie sie auch in der mathematischen Fachcommunity allgemein akzeptiert und angewendet werden. (Ufer et al., 2009, S. 35)

Dies ergänzend formulieren sie Kriterien für die Akzeptanz von Beweisversuchen als Beweis:

1. **Beweisschema:** Dieser Aspekt bezieht sich auf die Argumente, die in jedem einzelnen Schritt eines Beweises zugelassen sind. Im Fall des mathematischen Beweises sind dies allgemeingültige deduktive Schlüsse. […] Andere nicht zulässige Argumentationstypen wären empirisch-induktive Argumente, Berufung auf eine höhere Autorität oder auf die Anschauung. […] Die Fixierung des Arguments in formaler Sprache ist kein relevanter Teil des Aspekts Beweisschema. Sie ist für einen korrekten Beweis nicht zwingend notwendig. Ausschlaggebend ist letztlich die Transparenz des mathematischen Arguments, nicht seine korrekte Form.
2. **Beweisstruktur:** Die deduktiven Schlüsse müssen vom logischen Aufbau so angeordnet sein, dass sie mit den Voraussetzungen beginnen und bei der Behauptung enden. Weiterhin müssen die zur Stützung von Argumenten nötigen zusätzlichen Aussagen an der Stelle des entsprechenden Beweisschrittes bereits gesichert sein. Kurz gesagt soll der Beweis beweisen, dass die Behauptung aus den Voraussetzungen folgt. Insbesondere zirkuläre Argumentationen sind nicht erlaubt.
3. **Beweiskette:** Aufeinanderfolgende Beweisschritte müssen sich zu einer logischen Kette formieren in dem Sinne, dass jede neu gezeigte Aussage – nötigenfalls unter Hinzuziehung gesicherten Wissens – aus der im vorherigen Schritt gezeigten Aussage folgt. (Ufer et al., 2009, S. 35-36, H. i. O.)

Nach dem ersten Kriterium „Beweisschema“ sind ausschließlich deduktive Argumente zugelassen. Eine formalsprachliche Formulierung der Argumente ist aber nicht zwingend notwendig. Zudem müssen die einzelnen deduktiven Argumente nicht unbedingt fachlich korrekt sein, um dieses Kriterium zu erfüllen (Heinze & Reiss, 2003). Das zweite Kriterium „Beweisstruktur“ bezieht sich auf die Gesamtstruktur eines Beweises. Es besagt, dass das Argumentgefüge ausgehend von den Voraussetzungen zur Behauptung führen muss. Zirkelschlüsse sind demzufolge nicht erlaubt. Zudem wäre es ein Verstoß gegen dieses Kriterium, wenn das Argumentgefüge, z. B. wegen einer fehlerhaften Interpretation der zu beweisenden Aussage, nicht bei den tatsächlich gegebenen Voraussetzungen startet. Auch darf das Argumentgefüge nicht abbrechen oder zu einer anderen als der gewünschten Behauptung führen. Des Weiteren explizieren Heinze und Reiss (2003), dass der Beweisversuch keine Lücken enthalten darf, die die Struktur des Argumentgefüges unterbrechen. Das dritte Kriterium „Beweiskette“ bezieht sich auf die Verknüpfung der Argumente. Diese sollte korrekt sein. Beispielsweise sind fehlerhafte Termumformungen und falsche Verallgemeinerungen nicht erlaubt.

Während sich also das Kriterium „Beweisschema“ vor allem auf die Schlussweise bezieht, sind die Kriterien „Beweisstruktur“ und „Beweiskette“ für die Deskription der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz relevant. Dabei umfassen die zwei Kriterien jeweils ein großes Spektrum an möglichen Schwächen von Beweisversuchen. Das Kriterium „Beweisstruktur“ wird von Beweisversuchen mit grundlegenden logischen oder sachinhaltlichen Fehlern bis hin zu Beweisversuchen mit kleineren Unzulänglichkeiten in der Explizierung nicht erfüllt. Dem Kriterium „Beweiskette“ genügen weder Beweisversuche, die – mit Blick auf deren Weiterentwicklung zu einem Beweis – einen Wechsel zugrundeliegender Beweisideen erfordern, noch solche, bei denen nur ein Argument zu korrigieren ist. Um möglichst konkrete Einblicke in die Korrektheit und Vollständigkeit von Beweisversuchen zu erhalten, werden diese Kriterien in feinere Kategorien ausdifferenziert und neu strukturiert. Dabei werden mögliche Interpretationen der Kategorien und damit verbundene Anknüpfungspunkte für die Weiterentwicklung von Beweisversuchen zu einem Beweis bedacht. Aus dieser Perspektive erscheint es beispielsweise angemessen, Beweisversuche mit begründendem Ansatz, in denen etwa wegen einer fehlerhaften Interpretation der zu zeigenden Aussage nicht genau die vorgegebenen Voraussetzungen zugrunde gelegt werden, in einer Kategorie zu erfassen, weil solche Bearbeitungen es nahelegen, grundlegend zu klären, was zu tun ist. Zugleich ist es möglich, dass in Beweisversuchen, bei denen die Ausführungen auf eine nicht ganz korrekte Interpretation der zu zeigenden Aussage hinweisen, trotzdem ein weiterentwickelbares Argumentgefüge vorkommt. Daher wird bei der Beschreibung der Korrektheit und Vollständigkeit von Beweisversuchen im Rahmen dieser Arbeit zwischen der Passung der Ausführungen zu der zu zeigenden Aussage und der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges differenziert.

Zusätzlich zu dieser Umsortierung muss für eine reliable Codierung von Beweisversuchen der Interpretationsspielraum, den die Definitionen der Kriterien „Beweisstruktur“ und „Beweiskette“ lassen, reduziert werden. Auf Grundlage der vorgestellten Definitionen ist es beispielsweise nicht vollständig geklärt, wie Bearbeitungen, die ausschließlich kleine Ungenauigkeiten in der Formulierung enthalten, wie z. B. eine fehlerhafte Verwendung des mathematischen

Symbols für die Teilbarkeitsrelation, einzuordnen sind. Da ein Beweis nie vollständig aufgeschrieben ist, haben Heinze und Reiss (2003) zudem nicht genau festgelegt, ab wann eine Lücke die Struktur des Argumentgefüges unterbricht oder dieses nicht ausreichend explizit von den Voraussetzungen bis zur Behauptung führt, das heißt ab wann ein Argumentgefüge unvollständig ist.

Ein Ansatz zum Umgang mit diesem Interpretationsspielraum lässt sich in der Auseinandersetzung mit dem Kategoriensystem von **Kempen und Biehler (2014)** entwickeln. Zur Analyse von Beweisversuchen von Studierenden verwenden sie unter anderem die Kategorien „connected arguments with unrecoverable gap“ (Kempen & Biehler, 2014, S. 429), „connected argument, but incomplete“ (ebd.), „complete explanation […] with minor (formal) inaccuracies“ (ebd.) und „complete explanation“ (ebd.). Diese Unterscheidung regt dazu an, auch im Analysemodell dieser Arbeit zwischen unterschiedlich gravierenden Fehlern und Unvollständigkeiten zu differenzieren.

Dabei bleibt die Problematik, dass ein Beweis nie vollständig aufgeschrieben und die Entscheidung, ob ein Argumentgefüge eine Lücke enthält, stets mit Subjektivität verbunden ist (Abschn. 3.1.2). Zudem weist **Malle (2002)** darauf hin, dass neben der grundlegenden Problematik der Unmöglichkeit einer lückenlosen Explikation die Vollständigkeit eines Beweisversuchs stets in Abhängigkeit von der verwendeten Argumentationsbasis gesehen werden muss. Um eine große Anzahl an Beweisversuchen mit begründendem Ansatz hinsichtlich ihrer Vollständigkeit möglichst eindeutig einordnen und somit vergleichend beschreiben zu können, werden deshalb die in Abschnitt 3.3.3 unterschiedenen Beweisideen aufgegriffen und zu allen Beweisideen[46] Erwartungshorizonte formuliert, vor deren Hintergrund Beweisversuche mit begründendem Ansatz hinsichtlich ihrer Vollständigkeit beschrieben werden. Für die Entwicklung und Formulierung solcher Erwartungshorizonte wird das Schema von Toulmin (1958) verwendet.

Toulmin (1958) entwickelt ein Modell, das den Grundaufbau eines Arguments sowohl für mathematische als auch für außermathematische Kontexte beschreibt. Jahnke und Ufer (2015) fassen dieses prägnant zusammen:

> Toulmins Schema eines Schlusses besteht aus sechs Komponenten. Die Konklusion (K) ist die Aussage, von deren Gültigkeit ein Sprecher sein Gegenüber zu überzeugen versucht. Der Sprecher gründet seine Behauptung auf gewissen Daten (D). Um aus D die Konklusion K abzuleiten, beruft sich der Sprecher auf eine Schlussregel (SR). Diese Schlussregel mag ihrerseits durch eine Stützung (S) untermauert sein. Im Allgemeinen werden Schlussfolgerungen mit modalen Operatoren (O) versehen, die den subjektiven Überzeugungsgrad zum Ausdruck bringen, also Qualifikationen wie „vermutlich“, „höchst wahrscheinlich“ etc. Außerdem werden häufig Ausnahmebedingungen (AB) formuliert, unter denen der Schluss nicht gilt. Toulmins Schema […] sagt also, dass die Konklusion K aus den Daten D gemäß der Schlussregel SR gilt, die sich ihrerseits auf die Aussage S stützt. K hat einen Überzeugungsgrad O und gilt eventuell nicht, falls die Ausnahmebedingungen AB vorliegen. (Jahnke & Ufer, 2015, S. 339)

Dieses Schema wird in weiterentwickelter Form in mehreren mathematikdidaktischen Untersuchungen von Argumentationsprozessen verwendet (z. B. Knipping, 2003; Meyer, 2007; Pe-

[46] „Alle Beweisideen“ meint hierbei alle primären Beweisideen, die sich für eine korrekte und vollständige Begründung der infrage stehenden Aussage eignen.

demonte, 2007; Schwarzkopf, 2000) und kann auch für die Beschreibung rein schriftlich vorliegender Beweisversuche mit begründendem Ansatz herangezogen werden. Da in korrekten und vollständigen mathematischen Begründungen modale Operatoren und Ausnahmebedingungen nicht vorkommen, wird Toulmins Schema im Rahmen dieser Arbeit auf die drei Komponenten Datum (D), Schlussregel (SR) und Konklusion (K) reduziert. Um Missverständnisse wegen der Ähnlichkeit der Termini „Schlussregel" und „Schlussweise" zu vermeiden, wird entsprechend dem Vorgehen von Meyer (2007) der Begriff „Regel" verwendet. In einem einzelnen Argument wird demnach ausgehend von Daten (D) mithilfe einer allgemeinen Regel (R) auf eine Konklusion (K) geschlossen. Diese vereinfachte Form des Schemas von Toulmin entspricht strukturell dem logischen Ideal der Deduktion, nach dem ausgehend von einem Fall mithilfe eines allgemeinen Gesetzes auf ein Resultat geschlossen wird (Abschn. 3.1.2). Im Vergleich zum rein formal-logischen Schema der Deduktion nach Peirce steht beim Schema von Toulmin die inhaltliche Qualität der einzelnen Elemente mit im Fokus. Zudem müssen die einzelnen Elemente (insbesondere das Datum und die Regel) laut Toulmin (1958) nicht unbedingt sachlich korrekt sein, und auch wenn diese Prämissen korrekt sind, muss die angeführte Konklusion nicht zwingend daraus folgen. Deshalb erscheint das Schema von Toulmin für die Rekonstruktion von Beweisversuchen mit begründendem Ansatz von Schülerinnen und Schülern und die Festlegung, welche Daten, Regeln und Konklusionen für eine korrekte und vollständige Begründung erwartet werden, gut geeignet. Da mathematische Begründungen allerdings in der Regel nicht nur aus einem Argument bestehen, wird das Schema von Toulmin für die Analysen im Rahmen dieser Arbeit erweitert. Meyer (2007) redet in Anlehnung an Krummheuer (2003) von einem „mehrgliedrigen Argument" (Meyer, 2007, S. 88), wenn in einer Begründung mehrere Argumente aufeinander aufbauen. „Die Konklusion eines ersten Schlusses fungiert dann als Datum eines zweiten (usw.)" (Meyer, 2007, S. 88). Wenn die Regel eines Arguments mithilfe eines weiteren Arguments hergeleitet wird, d. h. die Konklusion eines zweiten Arguments darstellt, spricht Meyer (2007) von einem „mehrschichtigen Argument" (Meyer, 2007, S. 90). Darüber hinaus sind Kombinationen von mehrgliedrigen und mehrschichtigen Argumenten sowie die Verwendung mehrerer Daten oder Regeln innerhalb eines Arguments möglich. In dieser Arbeit werden Argumente, mehrgliedrige Argumente, mehrschichtige Argumente und sämtliche Kombinationen derselben in Anlehnung an Krumsdorf (2017) als Argumentgefüge bezeichnet. Dieser Ausdruck schließt auch einzelne Argumente und Gefüge aus mehreren Argumenten, die nicht verknüpft sind, ein. Auf diese Weise kann begrifflich zwischen Argumentgefügen und einzelnen Argumenten eines Argumentgefüges differenziert werden. Dabei können die einzelnen Argumente unterschiedlich komplex sein. Beispielsweise veranschaulicht Krumsdorf (2017), dass ein mehrgliedriges Argumentgefüge durch „Zusammenfassung aller Regeln zu einem Regelverbund" (Krumsdorf, 2017, S. 115) zu einem einzelnen Argument verkürzt werden kann. Andersherum lässt sich ein einzelnes Argument manchmal durch die Separation eines Regelverbunds oder eine Herleitung der Regel erweitern. Wie komplex ein Argument ist, hängt also mit Blick auf Beweisversuche von Lernenden von den Formulierungen der Schülerinnen und Schüler ab und mit Blick auf die Erwartungen an korrekte und vollständige Argumentgefüge von der Detailtiefe der konkreten, festgelegten Argumentgefüge. In die Entwicklung der Erwartungshorizonte für den Rahmen dieser Arbeit werden Aufgabenbearbeitungen von Lernenden einbezogen (Abb. 38; Abschn. 3.3.4.2).

Zusammenfassend werden für die Analyse der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz die von Heinze und Reiss (2003) und Ufer et al. (2009) formulierten Kriterien „Beweisstruktur“ und „Beweiskette“ ausdifferenziert und neu strukturiert. Die Frage, inwieweit die Ausführungen zu der zu zeigenden Aussage passen, und die Frage nach der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges werden separiert. Zudem werden verschieden gravierende Fehler und Unvollständigkeiten unterschieden und Erwartungshorizonte zu korrekten und vollständigen Argumentgefügen formuliert. Das daraus resultierende Kategoriensystem für die Analyse der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz wird im Folgenden präsentiert.

3.3.4.2 Vorstellung der Kategorien

Unter der Korrektheit und Vollständigkeit von Beweisversuchen werden im Rahmen dieser Arbeit die Passung der Ausführungen zu der zu zeigenden Aussage sowie die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges erfasst. Die Analyse der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz erfolgt dementsprechend in zwei Schritten, die nachfolgend genau beschrieben sind.

Passung zu der zu zeigenden Aussage

Zur Beschreibung der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz wird als erstes analysiert, inwieweit die Ausführungen zu der zu zeigenden Aussage passen. Dabei werden folgende Kategorien differenziert: PzzA1 *Hinweis auf einen grundlegenden Fehler*, PzzA2 *Bezug auf die zz. Aussage unklar*, PzzA3 *Passung zu der zz. Aussage*.

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **PzzA1** ***Hinweis auf einen grundlegenden Fehler***, wenn es einen Hinweis darauf gibt, dass die zu zeigende Aussage nicht ganz korrekt interpretiert wurde. In diesem Fall weisen die Ausführungen zwar einen klaren Bezug zu der Aussage auf, passen aber nicht ganz dazu. Beispielsweise kann sich eine Bearbeitung anstatt auf Summen von vier aufeinanderfolgenden ungeraden Zahlen auf das Produkt von vier gleichen ungeraden Zahlen beziehen (Bearbeitung 3.3.4a in Abb. 31). Zudem ist es z. B. möglich, dass die Umkehrung einer Aussage betrachtet oder Voraussetzungen, die nicht vorgegeben sind, als gültig angenommen werden. Bei Beweisversuchen in der Kategorie PzzA1 *Hinweis auf einen grundlegenden Fehler* gibt es also einen Hinweis darauf, dass die Interpretation der zu zeigenden Aussage anzupassen ist, damit ein korrektes und vollständiges Argumentgefüge entwickelt werden kann. Ein Beweisversuch gehört hingegen nicht in diese Kategorie, wenn es sich aus Expertensicht eher um sprachliche Ungenauigkeiten handelt als um einen grundlegenden Fehler sowie wenn die Ausführungen offenlassen, ob die zu zeigende Aussage vollständig korrekt interpretiert wurde.

Bearbeitung 3.3.4a: Klassenstufe 8, weiblich (Code: 0810601)

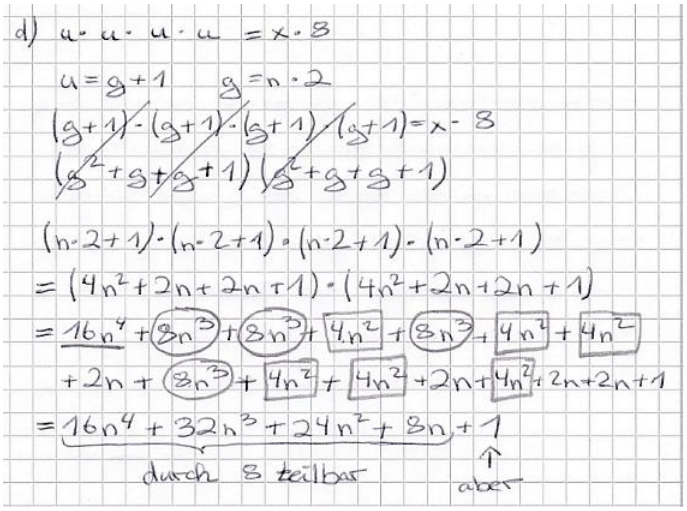


Abbildung 31: Bearbeitung 3.3.4a (PzzA1 *Hinweis auf einen grundlegenden Fehler*)

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **PzzA2** ***Bezug auf die zu zeigende Aussage unklar***, wenn in den Ausführungen kein Hinweis auf einen grundlegenden Fehler vorkommt, sie jedoch aus Expertensicht offenlassen, inwieweit sich der Beweisversuch genau auf die zu zeigende Aussage bezieht. Dies trifft u. a. zu, wenn ausschließlich die Summe zweier aufeinanderfolgender ungerader Zahlen oder die Parität der Summe von vier aufeinanderfolgenden ungeraden Zahlen betrachtet wird (Bearbeitung 3.3.4b in Abb. 32).

Bearbeitung 3.3.4b: Klassenstufe 7, weiblich (Code: 0710408)

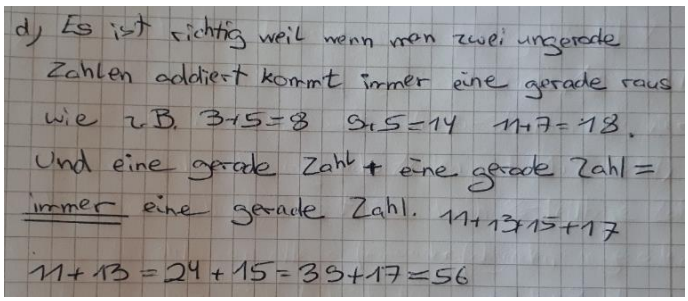


Abbildung 32: Bearbeitung 3.3.4b (PzzA2 *Bezug auf die zz. Aussage unklar*)

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **PzzA3** ***Passung zu der zu zeigenden Aussage***, wenn sich die Ausführungen aus Expertensicht sehr wahrscheinlich genau auf die zu zeigende Aussage beziehen. „Sehr wahrscheinlich“ meint, dass diese Passung eventuell noch klarer expliziert werden könnte. Ein Beweisversuch wird auch in die Kategorie PzzA3 eingeordnet, wenn das Argumentgefüge relativ vage ist, aber ein vollständig ausgeführtes Beispiel vorkommt, das genau zu der zu zeigenden Aussage passt und zusammen mit dem (übrigen) Argumentgefüge aus Expertensicht nahelegt, dass die Ausführungen sich genau auf die zu zeigende Aussage beziehen. In Beweisversuchen in dieser Kategorie gibt es keinen Anhaltspunkt dafür, dass die korrekte Interpretation der zu zeigenden Aussage für dessen Weiterentwicklung zu einem Beweis zu hinterfragen ist (Bearbeitung 3.3.4c in Abb. 33).

Bearbeitung 3.3.4c: Klassenstufe 7, männlich (Code: 0720807)

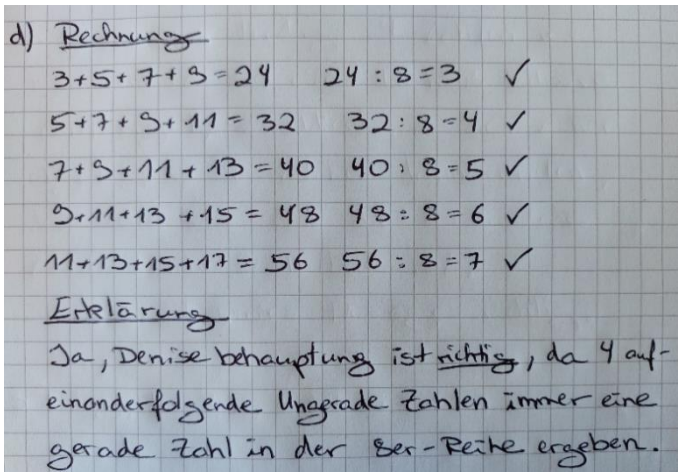


Abbildung 33: Bearbeitung 3.3.4c (PzzA3 *Passung zu der zz. Aussage*)

Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges

Bei der Analyse der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges wird jeder Beweisversuch mit *begründendem Ansatz* genau einer der folgenden sechs Kategorien zugeordnet: $KV1$ *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge*, $KV2$ *Ansatz für ein weiterentwickelbares Argumentgefüge*, $KV3$ *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler*, $KV4$ *weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke*, $KV5$ *weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung*, $KV6$ *korrektes und vollständiges Argumentgefüge*. Diese Kategorien werden nun nacheinander vorgestellt.

Ein Beweisversuch mit *begründendem Ansatz* gehört in die Kategorie **KV1** ***Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge***, wenn aus Expertensicht keine ausbaubare Beweisidee[47] erkennbar ist und zudem keine inhaltliche Überlegung vorkommt, die als Teil eines korrekten und vollständigen Argumentgefüges verwendet werden kann und deutlich über die zu zeigende Aussage und ihre Voraussetzungen hinausgeht (Bearbeitung 3.3.4d in Abb. 34).

Bearbeitung 3.3.4d: Klassenstufe 7, männlich (Code: 0711508)

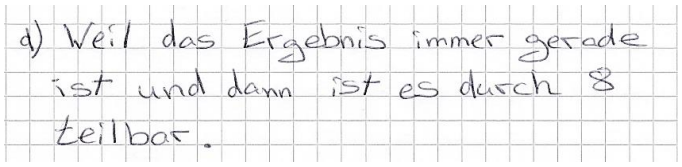


Abbildung 34: Bearbeitung 3.3.4d (KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge*)

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **KV2** ***Ansatz für ein weiterentwickelbares Argumentgefüge***, wenn einer der folgenden drei Punkte erfüllt ist.

[47] Als ausbaubare Beweisidee wird im Rahmen dieser Arbeit jede Beweisidee verstanden, die den Rahmen oder den wesentlichen Ausgangspunkt eines korrekten und vollständigen Argumentgefüges zu der zu zeigenden Aussage bilden kann (Abschn. 3.3.3).

(I) Aus Expertensicht ist eine inhaltliche Überlegung erkennbar, die als Teil eines korrekten und vollständigen Argumentgefüges zu der zu zeigenden Aussage verwendet werden kann.

(II) Aus Expertensicht ist ansatzweise eine Beweisidee erkennbar, von der ausgehend ein korrektes und vollständiges Argumentgefüge zu der zu zeigenden Aussage entwickelt werden kann.

(III) Eine ausbaubare Beweisidee ist explizit angedeutet, jedoch noch vage, nicht als mehrschrittiges Argumentgefüge oder nur ansatzweise korrekt dargelegt.

Beispielsweise gehört die Bearbeitung 3.3.4e aus Abbildung 35 in die Kategorie KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge*, weil die ausbaubare Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* aus Expertensicht erkennbar ist, wenngleich die Betrachtung der Teilbarkeit durch zwei für den Beweis der Teilbarkeit durch acht nicht ausreicht.

Bearbeitung 3.3.4e: Klassenstufe 7, weiblich (Code: 0711907)

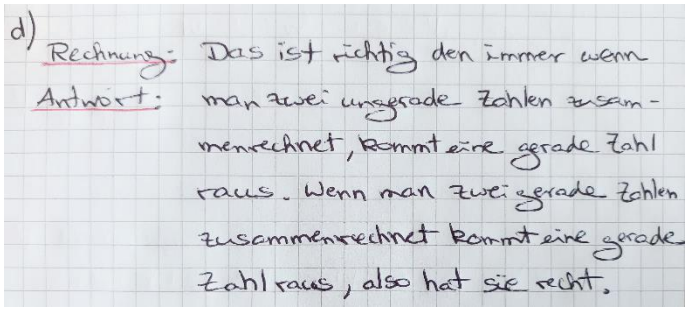


Abbildung 35: Bearbeitung 3.3.4e (KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge*)

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **KV3** ***weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler***, wenn eine Beweisidee in Form eines mehrschrittigen Argumentgefüges expliziert ist, das zu einem korrekten und vollständigen Argumentgefüge weiterentwickelt werden kann, die Ausführungen allerdings einen inhaltlichen oder grundlegenden Fehler enthalten. Als grundlegender Fehler gilt, wenn sich die Ausführungen auf eine nicht ganz korrekte Interpretation der zu zeigenden Aussage beziehen. Als inhaltliche Fehler gelten alle anderen Fehler, die über kleine Schwächen in der Darstellung der Überlegungen hinausgehen. Dazu gehören u. a. unvollständige Fallunterscheidungen, fehlerhafte Termumformungen und die Nutzung ungültiger Wissenselemente. Als inhaltlicher Fehler gilt auch, wenn eine notwendige, jedoch keine hinreichende Bedingung für die Gültigkeit der infrage stehenden Aussage gezeigt ist.

Die exemplarische Bearbeitung 3.3.4f in Abbildung 36 enthält ein weiterentwickelbares Argumentgefüge basierend auf der Idee, dass sich die Summe von vier aufeinanderfolgenden Zahlen stets um acht vergrößert. Der Achtklässler hat dies korrekt und in allgemein gehaltener Sprache erläutert. Allerdings fehlt der Induktionsanfang vollständig. Deshalb ist sein Schluss auf die Behauptung, dass die Summe immer durch acht teilbar ist, nicht hinreichend legitimiert. Dieser Schluss ist also streng genommen falsch, weshalb die Bearbeitung in die Kategorie KV3 einzuordnen ist.

Bearbeitung 3.3.4f: Klassenstufe 8, männlich (Code: 0821403)

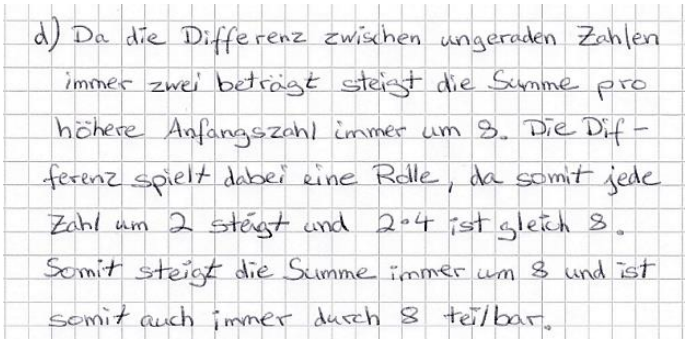


Abbildung 36: Bearbeitung 3.3.4f (KV3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler*)

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **KV4** ***weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke***, wenn eine Beweisidee in Form eines mehrschrittigen Argumentgefüges expliziert ist, das zu einem korrekten und vollständigen Argumentgefüge weiterentwickelt werden kann, und zudem das angegebene Argumentgefüge weitgehend korrekt ist, aber eine inhaltliche Lücke enthält. Ein Argumentgefüge wird als weitgehend korrekt bezeichnet, wenn es weder einen grundlegenden noch einen inhaltlichen Fehler enthält. Es enthält eine inhaltliche Lücke, wenn eine inhaltliche Überlegung fehlt, d. h. die ausgeführten Beweisideen für ein korrektes und vollständiges Argumentgefüge noch ergänzt werden müssten. Bei inhaltlichen Lücken liegt die Vermutung nahe, dass sie vom Lernenden auf Nachfrage eventuell nicht sofort geschlossen werden könnten. Dabei hängt die Entscheidung, ob ein weiterentwickelbares Argumentgefüge eine inhaltliche Lücke enthält, von den festgelegten Erwartungen an ein korrektes und vollständiges Argumentgefüge ab.

Die Abbildung 37 zeigt exemplarisch den Erwartungshorizont an Beweisversuche, die auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* (Variante 2.1.1 *Betrachtung der jeweils nächstgrößeren Summe, Teilbarkeit durch acht*) beruhen.[48] Die dort verwendeten Abkürzungen D, R und K stehen nacheinander für die Begriffe „Datum“, „Regel“ und „Konklusion“. Für ein korrektes und vollständiges Argumentgefüge müssen alle Daten, Regeln und Konklusionen sowie deren Verknüpfungen, die mit durchgehenden Linien gekennzeichnet sind, explizit ausgeführt sein. Die mit gestrichelten Linien gekennzeichneten Elemente dürfen auch nur implizit in einer Bearbeitung enthalten sein. Die Explizierung dieser mit gestrichelten Linien umrandeten Elemente erscheint wünschenswert, ist jedoch für die Zuordnung zu der Kategorie *korrektes und vollständiges Argumentgefüge* nicht unbedingt notwendig. Wenn bei einem Datum, einer Regel oder einer Konklusion mehrere Unterpunkte angegeben oder mit einem Schrägstrich getrennt sind, muss einer davon aus der Bearbeitung hervorgehen. Die Schlussweise und die verwendeten Darstellungsmittel dürfen von denen im Schema abweichen. Die im Schema kursiv gedruckte Abkürzung „Beh.“ steht für die gewünschte Behauptung, im vorliegenden Fall also für die Aussage, dass die betrachtete Summe immer durch acht teilbar ist. Neben einer solchen Wiederholung der Behauptung aus der Aufgabenstellung werden auch Formulierungen der Art „also ist die Aussage korrekt“ oder „Sie hat recht“ bis hin zu „Ja, weil …“ zu Beginn einer Bearbeitung als synonyme Äußerungen akzeptiert.

[48] Der Codierleitfaden mit allen Erwartungshorizonten kann gerne per Mail angefordert werden.

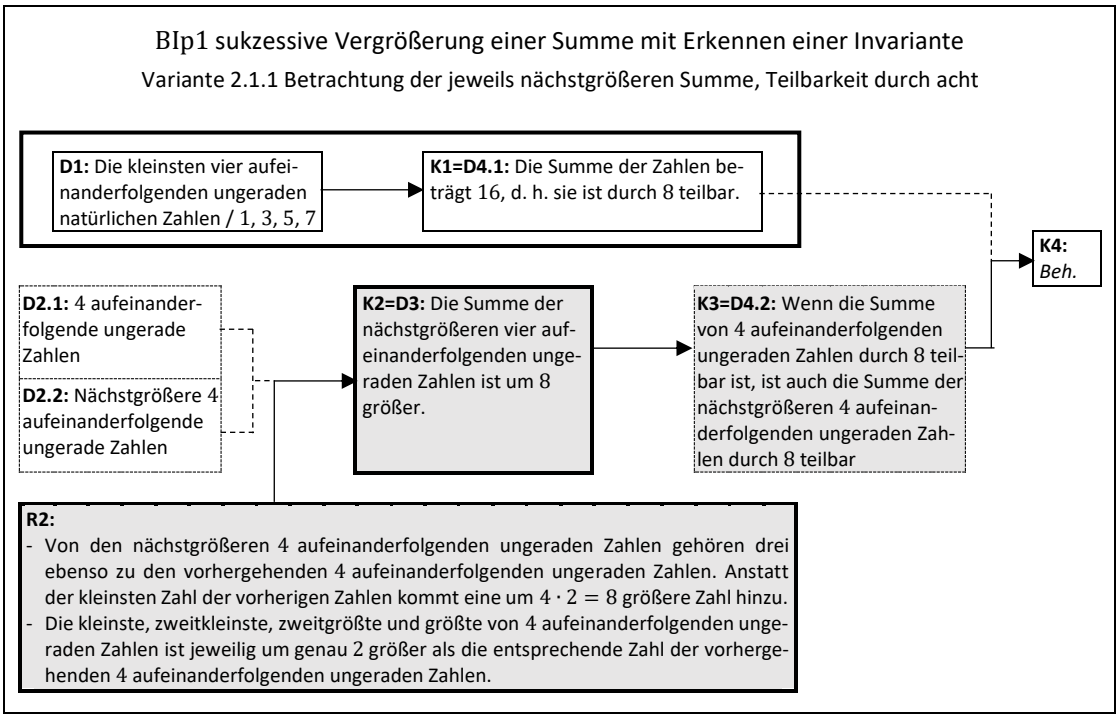


Abbildung 37: Exemplarischer Erwartungshorizont (BIp1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, Variante 2.1.1 *Betrachtung der jeweils nächstgrößeren Summe, Teilbarkeit durch acht*)

Wenn in einem Argumentgefüge im Vergleich zum Erwartungshorizont für die gewählte primäre Beweisidee ein breit umrandetes Element fehlt, enthält das Argumentgefüge eine inhaltliche Lücke. Dies trifft z. B. auf die Bearbeitung 3.3.4g in Abbildung 38 zu. Die Schülerin beschreibt, dass die Summe der nächstgrößeren vier aufeinanderfolgenden ungeraden natürlichen Zahlen immer um acht größer ist (K2). Die Bearbeitung liefert jedoch keinen Anhaltspunkt dafür, warum aus Sicht der Schülerin die Summe immer um acht größer wird. Das heißt, das breit umrandete Element R2 fehlt. Zudem hat die Achtklässlerin als Induktionsanfang die vier aufeinanderfolgenden ungeraden Zahlen $5, 7, 9, 11$ gewählt. Die beiden kleinsten Fälle, d. h. die vier aufeinanderfolgenden ungeraden Zahlen $1, 3, 5, 7$ und $3, 5, 7, 9$ wurden nicht berücksichtigt. Da die Berücksichtigung dieser beiden Fälle für einen Beweis der infrage stehenden Aussage notwendig ist, wird auch diese Auslassung nicht nur als Ungenauigkeit in der Formulierung, sondern als inhaltliche Lücke bezeichnet.

Bearbeitung 3.3.4g: Klassenstufe 8, weiblich (Code: 0811813)

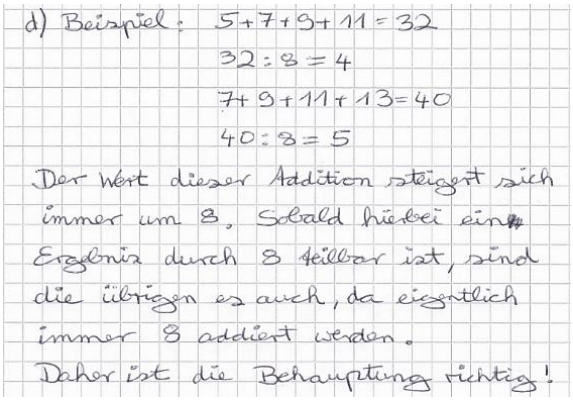


Abbildung 38: Bearbeitung 3.3.4g (KV4 *weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke*)

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **KV5** ***weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung***, wenn eine Beweisidee in Form eines mehrschrittigen Argumentgefüges, das zu einem korrekten und vollständigen Argumentgefüge weiterentwickelt werden kann, expliziert wurde, und zudem das angegebene Argumentgefüge weitgehend korrekt und vollständig ist, allerdings einen Formulierungsfehler oder eine Formulierungslücke enthält. Dabei gilt ein Argumentgefüge als weitgehend korrekt und vollständig, wenn dieses keinen grundlegenden oder inhaltlichen Fehler sowie keine inhaltliche Lücke enthält. Als Formulierungsfehler werden kleine Mängel in der numerischen, ikonischen, narrativen oder symbolischen Darstellung angesehen, die von einer Autorin bzw. einem Autor mit großer Wahrscheinlichkeit nach einem kurzen Hinweis sofort behoben werden könnten und die prinzipielle Korrektheit des Argumentgefüges nicht beeinflussen.[49] Als Formulierungslücke wird bezeichnet, wenn in einem Beweisversuch mit *begründendem Ansatz* im Vergleich zum zu den gewählten Beweisideen passenden Erwartungshorizont ein mit einer dünnen, durchgezogenen Linie umrandetes Element fehlt oder ein mit einer breiten, durchgezogenen Linie umrandetes Element unvollständig ausgeführt ist.

Beispielsweise enthält die Bearbeitung 3.3.4h in Abbildung 39 ein weitgehend korrektes und vollständiges Argumentgefüge mit Formulierungslücken und einem Formulierungsfehler, weshalb sie der Kategorie KV5 zuzuordnen ist. Die Bearbeitung 3.3.4h beruht, so wie die vorhergehende Bearbeitung 3.3.4g in Abbildung 38, auf der primären Beweisidee 1 *sukzessive Ver-*

Bearbeitung 3.3.4h: Klassenstufe 7, weiblich (Code: 0722305)

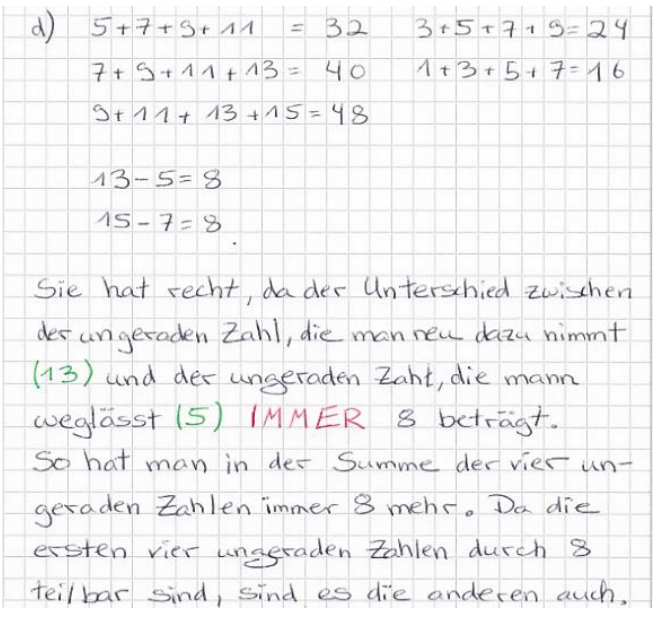


Abbildung 39: Bearbeitung 3.3.4h (KV5 *weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung*)

[49] An dieser Stelle ist mit zu erwähnen, dass bei der Beschreibung der Korrektheit und Vollständigkeit eines Beweisversuchs mit begründendem Ansatz in dieser Arbeit außen vor bleibt, inwieweit gegebenenfalls verwendete Variablen definiert wurden. Die korrekte Definition von Variablen kann von Lernenden am Anfang der Klassenstufen 5 bis 7, in deren Schulunterricht in der Regel nicht einmal Variablen eingeführt wurden, nicht erwartet werden. Inwieweit Variablen in einem Beweisversuch definiert sind, wird deshalb im Rahmen der Beschreibung der verwendeten Darstellungsmittel ohne Wertung miterfasst, wie in Abschnitt 3.3.11 beschrieben wird.

größerung einer Summe mit Erkennen einer Invariante. Im Vergleich zum geforderten Argumentgefüge (Abb. 37) ist die unter dem ersten Spiegelstrich angeführte Regel R2 erwähnt, aber unvollständig ausgeführt. Am Beispiel der aufeinanderfolgenden ungeraden Zahlen $5, 7, 9, 11$ und $7, 9, 11, 13$ erklärt die Schülerin, dass der Unterschied zwischen der kleinsten von vier aufeinanderfolgenden ungeraden Zahlen, die bei der Berechnung der nächstgrößeren Summe wegfällt (hier: 5), und der ungeraden Zahl, die bei dieser Berechnung neu hinzukommt (hier: 13), acht beträgt. Dabei deutet die Verwendung, Großschreibung und farbliche Hervorhebung des Ausdrucks „IMMER“ an, dass sie sich der Verallgemeinerbarkeit dieser Überlegung sicher ist. Im Vergleich zu den festgelegten Anforderungen lässt die Bearbeitung jedoch offen, warum der Unterschied zwischen dem wegfallenden und dem hinzukommenden Summanden immer acht beträgt. Außerdem könnte die Verknüpfung der Elemente noch verbessert werden. Dies ist insbesondere am letzten Satz ihrer Bearbeitung erkennbar: „Da die ersten vier ungeraden Zahlen durch 8 teilbar sind, sind es die anderen auch.“ Die Aussage, dass die kleinstmögliche Summe durch acht teilbar ist, steht hier ohne Bezug zur oben auf dem Blatt durchgeführten Berechnung dieser Summe. Zudem wird in diesem Satz explizit nur vom Induktionsanfang bzw. der Konklusion K1 auf die Behauptung geschlossen. Die zuvor angeführte stetige Vergrößerung der Summe um acht wurde wahrscheinlich mitgedacht, aber nicht explizit aufgegriffen. Auch diese fehlende Verbindung von K2 und K4 lässt sich als Formulierungslücke oder Formulierungsfehler bezeichnen. Weiterhin ist die Überlegung, dass „die ersten vier ungeraden Zahlen durch 8 teilbar sind“, wortwörtlich genommen nicht korrekt. Da aber offenbar die von der Schülerin oberhalb angeführte Summe der kleinsten vier aufeinanderfolgenden ungeraden Zahlen („$1 + 3 + 5 + 7 = 16$“) gemeint ist, wird die inkorrekte Aussage nicht als inhaltlicher Fehler, sondern nur als Formulierungsfehler bewertet. Insgesamt ist die Bearbeitung 3.3.4g somit der Kategorie $\mathrm{KV5}$ *weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung* zuzuordnen.

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **KV6** ***korrektes und vollständiges Argumentgefüge***, wenn ein Argumentgefüge basierend auf einer ausbaubaren Beweisidee korrekt[50] ausgeführt ist und zudem alle im zu den gewählten Beweisideen passenden Erwartungshorizont mit einer durchgehenden Linie umrandeten Elemente in angemessen verknüpfter Form vorkommen.

Beispielsweise gehört die Bearbeitung 3.3.4i aus Abbildung 40 in die Kategorie $\mathrm{KV6}$ *korrektes und vollständiges Argumentgefüge*. Alle laut dem relevanten, in Abbildung 37 dargestellten Erwartungshorizont notwendigen Elemente D1, K1=D4.1, D2.1, D2.2, R2, K2=D3 und K4 sowie die geforderten Verknüpfungen sind erkennbar und korrekt ausgeführt. Zuerst weist die Schülerin nach, dass die Summe der kleinsten vier aufeinanderfolgenden ungeraden Zahlen durch acht teilbar ist (D1, K1). Am Beispiel dieser und der darauffolgenden vier aufeinanderfolgenden ungeraden Zahlen wird anschließend begründet, warum die nächstgrößere Summe stets um acht größer ist. Als Regel (R2) verwendet sie dabei, dass bei der nächstgrößeren Summe von vier aufeinanderfolgenden ungeraden Zahlen jeder der Summanden um genau zwei größer ist als

[50] Wie bereits erwähnt, wird es dabei nicht als Fehler angesehen, wenn Variablen ohne Definition vorkommen. Außerdem werden präformale Darstellungen akzeptiert und Rechtschreib-, Grammatik- und Zeichensetzungsfehler hingenommen.

Bearbeitung 3.3.4i: Klassenstufe 8, weiblich (Code: 0811104)

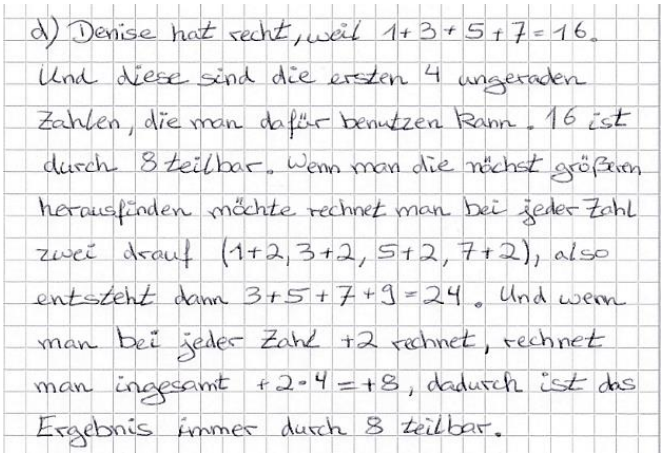


Abbildung 40: Bearbeitung 3.3.4i (KV6 *korrektes und vollständiges Argumentgefüge*)

die entsprechenden Summanden der vorhergehenden Summe. Das Datum (D2.1, D.2.2) und die Regel (R2) sind beispielgebunden formuliert, was aber, wie oben angegeben, bei der Beurteilung der Korrektheit und Vollständigkeit eines Argumentgefüges nicht berücksichtigt wird. Den Schluss auf die Konklusion (K2), dass die Summe der nächstgrößeren vier aufeinanderfolgenden ungeraden Zahlen um acht größer ist, hat die Schülerin dabei bereits in allgemeiner Sprache ausgedrückt („Und wenn man bei jeder Zahl $+2$ rechnet, rechnet man insgesamt $+2 \cdot 4 = +8$“). Von dieser Überlegung ausgehend schließt sie auf die Behauptung (K4, „dadurch ist das Ergebnis immer durch 8 teilbar“). Hier scheint implizit mitgedacht worden zu sein, dass wenn eine Summe von vier aufeinanderfolgenden ungeraden Zahlen durch acht teilbar ist, auch die Summe der nächstgrößeren vier aufeinanderfolgenden Zahlen durch acht teilbar ist (K3) und dass die kleinstmögliche Summe tatsächlich durch acht teilbar ist (K1). Da K3 und ein expliziter Einbezug des Induktionsanfangs beim Schluss auf die Behauptung nicht unbedingt gefordert sind, ist das Argumentgefüge der Schülerin vollständig.

3.3.4.3 Hinweise zu den Kategorien

Zur Beschreibung der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz werden die Passung der Ausführungen zu der zu zeigenden Aussage und die Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges analysiert.

Die Ergebnisse der Analyse der **Passung der Ausführungen zu der zu zeigenden Aussage** sollen einen Anhaltspunkt geben, inwieweit die zu zeigende Aussage korrekt interpretiert wurde. Dass die Ausführungen in Beweisversuchen von Lernenden genau zu der zu zeigenden Aussage passen, ist Forschungsergebnissen zufolge keine Selbstverständlichkeit (Abschn. 5.1.1). Zugleich erscheint eine korrekte Interpretation der zu zeigenden Aussage als eine Voraussetzung für die Entwicklung, Formulierung eines Beweises zu dieser Aussage. Insbesondere Beweisversuche in der Kategorie PzzA1 *Hinweis auf einen grundlegenden Fehler* haben in den Begrifflichkeiten von Heinze und Reiss (2003) ein Defizit im Bereich der Beweisstruktur. Bei der Weiterentwicklung solcher Beweisversuche wäre die Interpretation der zu zeigenden Aussage zu hinterfragen und ggf. zu klären.

Beweisversuche in den verschiedenen Kategorien zur Beschreibung der **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** können ebenfalls bezüglich des von Heinze und Reiss (2003) sowie Ufer et al. (2009) beschriebenen Konzepts eingeordnet und hinsichtlich ihrer Weiterentwickelbarkeit zu einem korrekten und vollständigen Argumentgefüge interpretiert werden. Korrekte und vollständige Argumentgefüge (KV6) werden den von Heinze und Reiss (2003), Ufer et al. (2009) formulierten Kriterien „Beweisstruktur" und „Beweiskette" tendenziell gerecht. Bei Beweisversuchen in den Kategorien KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge*, KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge* hingegen ist es möglich, dass keines dieser beiden Kriterien erfüllt ist. Um ausgehend von Beweisversuchen in diesen Kategorien KV1-2 einen Beweis zu entwickeln, sind die Ausführungen gründlich zu überdenken, gegebenenfalls fachliche Unsicherheiten zu klären und eventuell eine neue ausbaubare Beweisidee zu generieren. Von Beweisversuchen in der Kategorie KV3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler* wird auch mindestens eines der Kriterien „Beweiskette" und „Beweisstruktur" nicht erfüllt. Im Falle eines inhaltlichen oder grundlegenden Fehlers ist das Argumentgefüge zu korrigieren, eventuell sind zudem größere Ergänzungen und damit verbunden eine Weiterentwicklung der erkennbaren Beweisideen erforderlich.[51] Beweisversuche in der Kategorie KV4 *weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke* enthalten Unvollständigkeiten, die in den Begrifflichkeiten von Heinze und Reiss (2003) die Struktur des Argumentgefüges unterbrechen. Bearbeitungen mit inhaltlicher Lücke werden also dem Kriterium „Beweisstruktur" nicht vollständig gerecht. Das Schließen einer inhaltlichen Lücke kann neben einer genaueren Explizierung auch eine sachinhaltliche Weiterentwicklung der Beweisideen erfordern. Im Vergleich dazu sind bei Beweisversuchen in der Kategorie KV5 *weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung* ausschließlich kleine Änderungen oder Ergänzungen erforderlich. Wegen der Abhängigkeit von den Erwartungen an ein korrektes und vollständiges Argumentgefüge und der Unterschiedlichkeit der Beweisversuche in der Kategorie KV5 lässt sich nicht pauschal sagen, ob sie die von Heinze und Reiss (2003), Ufer et al. (2009) formulierten Anforderungen erfüllen.

Bei diesen Überlegungen zur Analyse der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges ist zu bedenken, dass Beweisversuche sämtlicher Kategorien KV1 bis KV4 zusätzlich die Unzulänglichkeiten von Beweisversuchen in einer der Kategorien mit einer höheren Nummer aufweisen können. Zum Beispiel können Beweisversuche in der Kategorie KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge* prinzipiell inhaltliche Lücken oder Formulierungsfehler aufweisen. Einerseits gehen dadurch Informationen verloren. Andererseits erscheint die beschriebene Kategorisierung plausibel, da es mit Blick auf die Weiterentwicklung eines Beweisversuchs mit begründendem Ansatz zu einem Beweis z. B. wenig zielführend wäre, einzelne Formulierungen zu optimieren, wenn er allenfalls einen Ansatz für ein weiterentwickelbares Argumentgefüge enthält.

[51] Dabei kann es sich lohnen, zuerst die korrekte Interpretation der zu zeigenden Aussage sicher zu stellen, insbesondere falls eine Bearbeitung in eine der Kategorien PzzA1-2 gehört.

3.3.5 Begründender Ansatz: Beispiele

Einen Ansatzpunkt für die Weiterentwicklung von Beweisversuchen mit begründendem Ansatz können angegebene Beispiele darstellen. Auch in Beweisversuchen mit begründendem Ansatz können Beispiele vorkommen und verschiedene, wichtige Funktionen haben. Daher geht es in einem weiteren Analyseschritt darum, die Verwendung von Beispielen detailliert zu beschreiben, sofern dies anhand der verschriftlichten Beweisversuche möglich ist (Abb. 41). Im Folgenden werden die Entwicklung der Kategorien zur Beschreibung der Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz, deren Definitionen und mögliche Interpretationen präsentiert. Die Ergebnisse der Analyse der Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz werden in Abschnitt 8.2.3 vorgestellt.

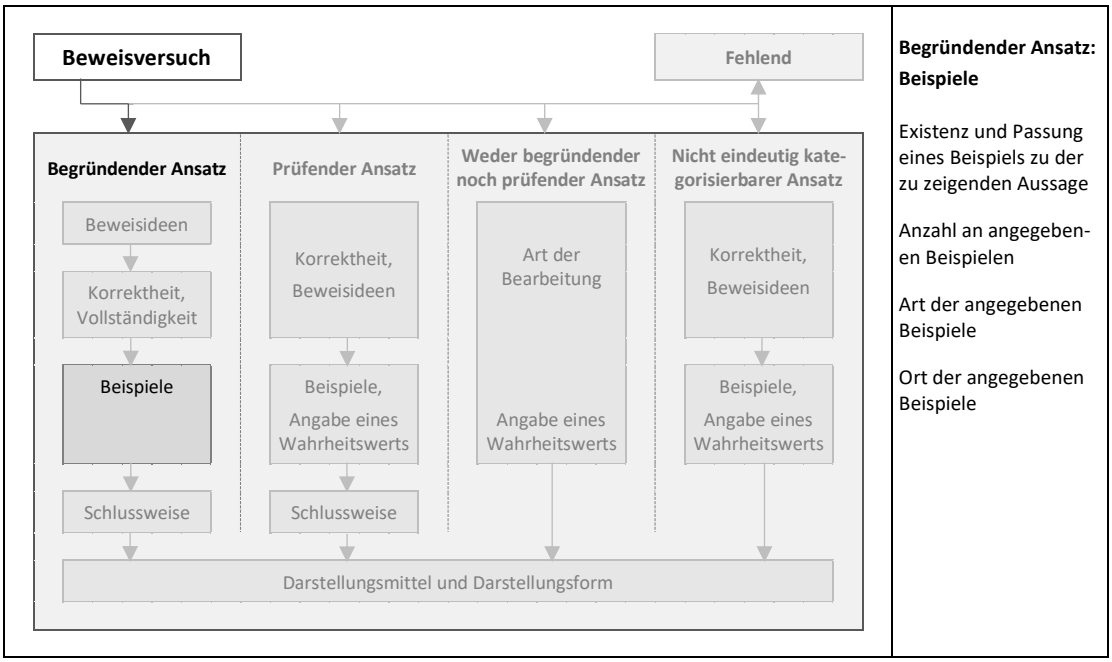


Abbildung 41: Analysemodell. Fokus: Begründender Ansatz, Beispiele

3.3.5.1 Entwicklung der Kategorien

Für die Entwicklung der Kategorien zur Beschreibung der Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz wird zunächst die Literatur gesichtet. Demnach können Beispiele bei der Bearbeitung von Beweisaufgaben verschiedene **Funktionen** haben.[52] Sie können eine zu zeigende Aussage veranschaulichen und helfen, diese zu verstehen (z. B. Beutelspacher, 2009; Lynch & Lockwood, 2019). Beispiele können auch berechnet werden, um die Gültigkeit einer Aussage zu überprüfen, den Gültigkeitsbereich auszutesten und Vertrauen in die Gültigkeit der Aussage zu gewinnen (z. B. Beutelspacher, 2009; Lockwood et al., 2016). Weiterhin können Beispiele genutzt werden, um eine Beweisidee zu generieren (z. B. Ellis et al., 2012; Lynch & Lockwood, 2019). Im Rahmen der wissenschaftssoziologischen Studie von Heintz (2000) beschreibt ein Mathematiker sein Tun wie folgt:

[52] Die folgenden Ausführungen fokussieren primär auf Beweisaufgaben, in denen eine Allaussage mit unendlichem Gültigkeitsbereich zu zeigen ist.

> Natürlich versucht man, allgemeine Sätze zu beweisen. Man hat aber manchmal überhaupt keine Ahnung, wie man das angreifen muss. Dann ist die Regel: man [*sic*] betrachte den ersten Spezialfall, den man nicht versteht. Dann untersucht man den. Das ist ganz experimentell. Und dann betrachtet man andere Spezialfälle. Also entweder sind die Spezialfälle selbst schon wieder interessant, und dann hat man bereits etwas gemacht. Oder sie sind nicht so interessant, und dann hofft man immer, dass man schlussendlich die allgemeine Idee haben wird. Also in diesem Sinne – es ist wirklich experimentell (Heintz, 2000, S. 151).

In der Wissenschaft Mathematik ist es eine übliche Vorgehensweise, Eigenschaften von Einzelfällen zu analysieren oder etwa systematisch zu probieren, in der Hoffnung an den Beispielen Verallgemeinerbares zu sehen. Es gibt empirische Evidenz dafür, dass es auch manchen Lernenden durch die Arbeit mit Beispielen gelingt, zu erkennen, warum eine Aussage gilt (z. B. Ellis et al., 2012; Metzer & Zaslavsky, 2019). Zudem können Beispiele verwendet werden, um verallgemeinerbare Überlegungen zu erläutern oder zu veranschaulichen (z. B. Balacheff, 1988; Mason & Pimm, 1984). Ebenfalls gibt es Lernende, die Beispiele anstelle von Fachvokabularien nutzen, um allgemein gedachte Überlegungen auszudrücken (Balacheff, 1988). Außerdem ist es möglich, dass Lernende innerhalb eines Argumentgefüges Beispiele anführen, um einzelne Argumente zu stützen. Weiterhin kommt es vor, dass Schülerinnen und Schüler nach der Formulierung eines Beweisversuchs mit begründendem Ansatz noch ein Beispiel betrachten, beispielsweise um ihre allgemein gehaltenen Überlegungen zu prüfen oder um die Gültigkeit der Aussage zusätzlich mit einem Beispiel zu bekräftigen (Fischbein 1982; Krumsdorf 2017; Porteous, 1990; Abschn. 5.2). Darüber hinaus kann es sein, dass Lernende in einem Beweisversuch mit begründendem Ansatz ein Beispiel angeben, weil sie dieses als notwendigen Bestandteil eines Beweises ansehen (z. B. Fischbein, 1982) oder sie es gewöhnt sind, in Aufgabenbearbeitungen ein Beispiel dazuzuschreiben.

Die Erkenntnisse zu möglichen Gründen für die Angabe von Beispielen ergänzend, geben die Ergebnisse des vierjährigen amerikanischen **Forschungsprojekts „The Role and Use of Examples in learning to Prove“** (Zaslavsky & Knuth, 2019) Einblick in die Art und Weise der Verwendung sowie in den Nutzen von Beispielen bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit. In halbstrukturierten einstündigen Einzelinterviews werden 38 Lernende ab der siebten Klassenstufe sowie 13 Mathematikerinnen und Mathematiker aufgefordert, Aufgaben mit Beweisnotwendigkeit zu bearbeiten. In der Untersuchung hilft die Betrachtung von Beispielen allen Lernenden, die ein Beispiel nutzen, wenn auch in unterschiedlichem Ausmaß (Ozgur et al., 2019). Bei einer Aufgabe zur Teilbarkeit der Summe aufeinanderfolgender ganzer Zahlen betrachten in der Untersuchung erfolgreiche[53] Lernende häufig Beispiele, mit dem Ziel zu verstehen, warum eine Aussage gilt, während die nicht erfolgreichen besonders häufig ein Beispiel nutzen, um eine Aussage zu verstehen (Ozgur et al., 2019). In den Interviews gelingt rund zwei Drittel der Lernenden, die bei dieser Aufgabe Beispiele nutzen, mit dem Ziel zu verstehen, warum eine Aussage gilt, eine gangbare, wenn auch nicht unbedingt vollständige Begründung zu entwickeln (Lynch & Lockwood, 2019). Die Analyse der Beispielnutzung der teilnehmenden Schülerinnen und Schüler zeigt weiterhin, dass sie Beispiele in mehr als der Hälfte

[53] „[…] by successful case we mean the ability to develop a viable justification that accounts for why the conjecture must be true” (Ozgur et al., 2019, S. 284).

der Fälle generisch[54] nutzen (Aricha-Metzer & Zaslavsky, 2019). Die Ergebnisse deuten an, dass eine generische Beispielnutzung eine notwendige, aber keine hinreichende Bedingung für eine produktive Beispielnutzung darstellt (Aricha-Metzer & Zaslavsky, 2019). Dabei scheint die Fähigkeit, in einem oder mehreren Beispielen das Allgemeine zu sehen, bei Lernenden sehr unterschiedlich ausgeprägt zu sein (Balacheff, 1988; Krumsdorf, 2017; Metzer & Zaslavsky, 2019).

Bezüglich der Art der ausgewählten Beispiele verdeutlichen die Ergebnisse des Forschungsprojekts, dass sowohl Lernende als auch Mathematikerinnen, Mathematiker häufig das kleinste Beispiel oder ein Beispiel, das sich der eigenen Einschätzung nach leicht berechnen lässt, wählen, insbesondere für anfängliche Explorationen vorgegebener Aussagen (Lynch & Lockwood, 2019). Zudem verdeutlichen die Ergebnisse, dass bei den teilnehmenden Lernenden aus middle schools und high schools die Beispielnutzung häufiger produktiv[60] ist, wenn der Interviewer sorgfältig ausgewählte Beispiele vorschlägt, als wenn die Lernenden selbst Beispiele auswählen (Ozgur et al., 2019). Bei der selbstständigen Auswahl von Beispielen gehen sowohl einige erfolgreiche[59] als auch einige nicht erfolgreiche Lernende systematisch vor (Ozgur et al., 2019). Nicht erfolgreiche Lernende betrachten zudem häufig ein möglichst vielfältiges Set an Beispielen, während dies bei erfolgreichen Lernenden wesentlich seltener vorkommt, weil sie elaboriertere Strategien, wie etwa die Suche nach Invarianten, anwenden (Ozgur et al., 2019).

Krumsdorf (2017) folgert aus seinen Einzelfallstudien, dass „die Wahl und die Anzahl an Beispielen wesentlich den beispielgebundenen Beweisverlauf eines Schülers beeinflusst“ (Krumsdorf, 2017, S. 353) und kommt zu dem Schluss, dass die Betrachtung weniger und insbesondere auch schwieriger oder nicht gut vorstellbarer Beispiele, wie etwa großer Beispiele, die Entwicklung eines beispielgebundenen Beweises eher begünstigt als die Betrachtung möglichst vieler Beispiele.

Dabei basieren die Ergebnisse des Forschungsprojekts „The Role and Use of Examples in learning to Prove“ (Zaslavsky & Knuth, 2019) und die von Krumsdorf (2017) auf Prozessanalysen und audiovisuellen Daten, während im Rahmen der vorliegenden Arbeit rein in schriftlicher Form vorliegende Produkte von Beweisprozessen (konkret: Beweisversuche) zu analysieren sind. Bei Beweisversuchen lässt sich herausarbeiten, ob Beispiele vorkommen, wie viele Beispiele angegeben sind und welche Eigenschaften diese Beispiele aufweisen. Auch ist es möglich, Einblicke zu geben, welche Funktionen die Beispiele in den Beweisversuchen allerdings aus der Perspektive von Lesenden haben und welche Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis der zu zeigenden Aussage durch die angegebenen Beispiele zu erkennen sind. Bei der Analyse der Verwendung von Beispielen in Beweisversuchen ist es möglich und sinnvoll, die beschriebenen Forschungsergebnisse zu Arten an Bei-

[54] „In our analysis we distinguish between *empirical example-use* and *generic example-use*, and examine whether the example-uses that we identified were productive for proving, in terms of developing a proof, a deductive argument, or a sound justification that may lead to a proof” (Aricha-Metzer & Zaslavsky, 2019, S. 304, H. i. O.). „In the context of this study, we use empirical example-use to mean the use of specific examples to make sense of, check out, or verify conjectures, focusing on the specifics of an example without looking at the specifics in a general way. […] In contrast to an empirical example-use, in a generic example-use one sees through the specifics some general case. It does not necessarily mean that one sees the full structure or the idea of a proof through the generic example“ (Aricha-Metzer & Zaslavsky, 2019, S. 305).

spielen und zu Funktionen der Beispiele zu berücksichtigen. Zugleich ist es notwendig, die Begrifflichkeiten und Kategorien aus den Prozessanalysen an die Andersartigkeit der verschriftlichten Beweisversuche anzupassen. Die Begrifflichkeiten sind teils zu ersetzen und zu ergänzen sowie für die **vorliegende Studie** zu konkretisieren. Dabei wird insbesondere auch der Nutzen der Kategorien zur Beantwortung der Forschungsfragen dieser Arbeit im Blick behalten. Diese Entwicklung erfolgt im Sinne eines kohärenten Analysemodells parallel zu der Ausarbeitung der Kategorien für die Analyse von Beweisversuchen mit prüfendem Ansatz, die in den Abschnitten 3.3.7, 3.3.8 beschrieben wird. Dementsprechend sind Teile des folgenden Kategoriensystems zur Analyse von Beispielen in Beweisversuchen mit begründendem Ansatz ebenfalls bei der Analyse von Beweisversuchen mit prüfendem Ansatz relevant.

3.3.5.2 Vorstellung der Kategorien

Die Analyse der Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz erfolgt in mehreren Schritten (Abb. 42). Im ersten Schritt werden die Existenz und ggf. Passung eines Beispiels zu der zu zeigenden Aussage untersucht. Im zweiten Schritt stehen die Anzahl und die Art aller angegebenen Beispiele im Fokus. Drittens wird der Ort der angegebenen Beispiele codiert. Davon ausgehend, dass ein Beweisversuch entsprechend der Anordnung der Ausführungen in der Bearbeitung von oben nach unten verfasst wurde, macht es mit Blick auf die Interpretation von Charakteristika angegebener Beispiele einen Unterschied, wo sich die Beispiele im Verhältnis zum (übrigen) Argumentgefüge befinden (Abschn. 3.3.5.3). Deshalb werden in einem vierten Schritt die Beispiele oberhalb, danach die Beispiele unterhalb und abschließend die Beispiele innerhalb des (übrigen) Argumentgefüges separat fokussiert und hinsichtlich ihrer Eigenschaften analysiert. Im Folgenden werden diese Schritte und die dazugehörigen Kategorien nach einer Einführung der vorkommenden grundlegenden Termini sukzessive vorgestellt.

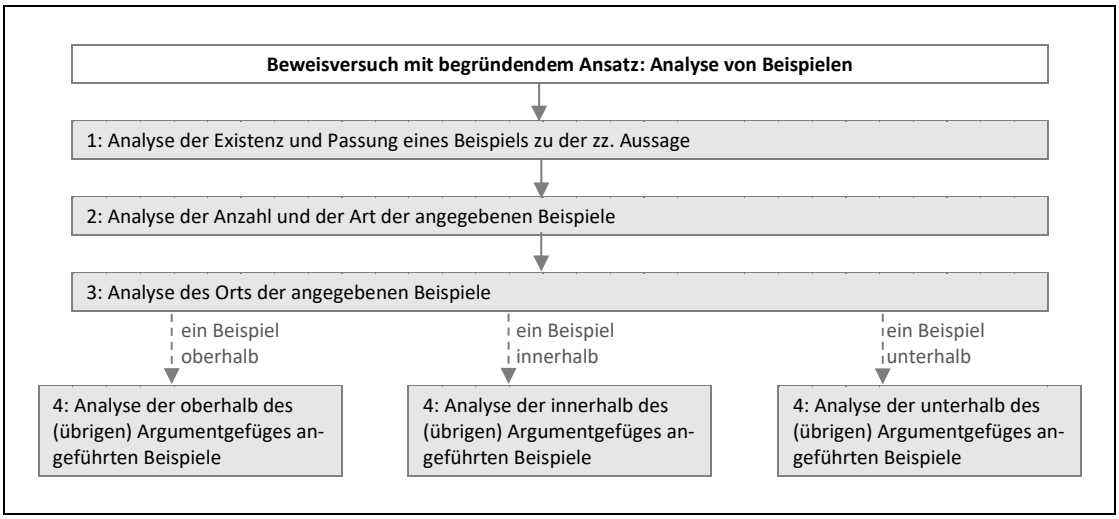


Abbildung 42: Beweisversuche mit begründendem Ansatz: Vorgehensweise bei der Untersuchung der Verwendung von Beispielen

Grundlegende Termini[55]

Als **Beispiel** wird in dieser Arbeit jede (numerische) Konkretisierung oder Exemplifizierung einer Idee bezeichnet (z. B. „$3 + 5 + 7 + 9 = 24$“, „$1 + 1 + 1 + 1 = 4$“, „$7 + 3 = 10$ gerade“). Dementsprechend handelt es sich auch um ein Beispiel, wenn die Ausführungen aus Expertensicht zugleich als Induktionsanfang, als Fall einer vollständigen Fallunterscheidung oder etwa als eine Bestimmung der Summe der Reste im Rahmen einer Modulo-Rechnung fungieren. Falls die Betrachtung eines Beispiels in mehreren Schritten erfolgt oder an einem Beispiel ergänzende Überlegungen durchgeführt werden, die z. B. eine Beweisidee veranschaulichen, werden die gesamten auf das konkrete Beispiel bezogenen Ausführungen als ein einziges Beispiel angesehen (z. B. „$1 + 3 = 4 \quad 5 + 7 = 12 \quad 4 + 12 = 16 \quad 16 : 8 = 2$).

Bezogen auf die in dieser Arbeit fokussierte Beweisaufgabe gilt ein Beispiel als **Beispiel zu der zu zeigenden Aussage**, wenn vier aufeinanderfolgende ungerade Zahlen addiert werden oder eine Summe von vier aufeinanderfolgenden ungeraden Zahlen durch vier (Kl. 5-6) bzw. acht (Kl. 7-8) dividiert wird (z. B. „$7 + 9 + 11 + 13 = 40$“, „$40 : 8 = 5$“). Alle anderen Beispiele werden **sonstiges Beispiel** genannt. Es kann zum Beispiel exemplifiziert werden, dass die Summe zweier ungerader Zahlen immer gerade ist, etwa anhand des Beispiels $1 + 3 = 4$. Zudem ist es möglich, dass ein sonstiges Beispiel, wie z. B. $1 + 13 + 27 + 39 = 80$, von einem Lernenden als ein Beispiel zu der zu zeigenden Aussage gedacht war, es sich aus Expertensicht jedoch nicht um ein solches handelt.

Ein Beispiel zu der zu zeigenden Aussage ist ein **eigenes Beispiel**, wenn es sich nicht um den in der Aufgabenstellung vorgegebenen Einzelfall der vier aufeinanderfolgenden ungeraden Zahlen $5, 7, 9, 11$ handelt (z. B. „$1 + 3 + 5 + 7 = 16 \quad 16 : 8 = 2$“).

Ein Beispiel zu der zu zeigenden Aussage gilt als **vollständig**, wenn vier aufeinanderfolgende ungerade Zahlen addiert und das Ergebnis durch vier (Kl. 5-6) bzw. acht (Kl. 7-8) dividiert wird. Dabei wird eine Aussage der Art „24 ist durch acht teilbar“ mit einer klaren Fokussierung auf die Teilbarkeit durch vier (Kl. 5-6) bzw. acht (Kl. 7-8) ebenso akzeptiert wie eine rechnerische Division (z. B. „$1 + 3 + 5 + 7 = 16 : 8 = 2$“, „$1 + 3 + 5 + 7 = 16$ ist durch acht teilbar“). Ein Beispiel zu der zu zeigenden Aussage gilt als **fast vollständig**, wenn entweder die Addition oder die Division nicht vollständig expliziert, aber angedeutet wurde (z. B. „ $1 + 3 + 5 + 7 = 16$, hier geht‘s“, „Die kleinste Summe beträgt 16 und $16 : 4 = 4$“). Ein Beispiel zu der zu zeigenden Aussage gilt als **unvollständig**, wenn es weder vollständig noch fast vollständig ist (z. B. „$1 + 3 + 5 + 7$“, „$56 : 8 = 7$“).

[55] Im Folgenden werden alle im Rahmen der vorliegenden Arbeit festgelegten Begrifflichkeiten eingeführt, die sowohl bei der Analyse der Verwendung von Beispielen bei Beweisversuchen mit begründendem Ansatz als auch bei der Analyse der Beweisversuche mit prüfendem Ansatz (Abschn. 3.3.7, 3.3.8) oder bei der Analyse der in Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz angegebenen Beispiele (Abschn. 3.3.10) genutzt werden.

Ein Beispiel zu der zu zeigenden Aussage gilt als **korrekt**, wenn weder ein grundlegender oder ein inhaltlicher Fehler noch ein Formulierungsfehler zu erkennen ist (z. B. „$1 + 3 + 5 + 7 = 16 \quad 16 : 8 = 2$“, „$11 + 13 + 15 + 17 = 56$“).[56] Ein Beispiel zu der zu zeigenden Aussage enthält einen **Formulierungsfehler**, wenn das Beispiel kleine Mängel in der Darstellung aufweist, die über eine rein präformale Darstellung hinausgehen, jedoch das Finden einer Beweisidee wohl kaum behindern würden (z. B. „$1 + 3 + 5 + 7 = 16 : 8 = 2$“). Ein Beispiel zu der zu zeigenden Aussage wird als weitgehend korrekt bezeichnet, wenn es korrekt ist oder ausschließlich Formulierungsfehler enthält. Ein Beispiel zu der zu zeigenden Aussage ist nicht weitgehend korrekt, sondern enthält einen **grundlegenden oder inhaltlichen Fehler**, wenn es nicht ganz zu der zu zeigenden Aussage passt oder eine Unzulänglichkeit, wie einen Rechenfehler, enthält, die das Finden einer Beweisidee behindern kann (z. B. „$3 + 5 + 7 + 9 = 22$“).

Beispiele zu der zu zeigenden Aussage, die direkt aufeinanderfolgen, gelten im Rahmen dieser Arbeit als **Reihenfolgebeispiele** (z. B. „$1 + 3 + 5 + 7 = 16, 3 + 5 + 7 + 9 = 24$“). Ergänzend gelten als **andere systematische Beispiele** zwei oder mehr angegebene Beispiele, bei denen aus Expertensicht eine Systematik zu erkennen ist (z. B. „$1 + 3 + 5 + 7 = 16, 9 + 11 + 13 + 15 = 48$“, „$1 + 1 + 1 + 1 = 4, 1 + 2 + 4 + 1 = 8, 1 + 3 + 4 + 4 = 12$“).

Ein Beispiel wird als **besonderes Beispiel** bezeichnet, wenn es aus Expertensicht bei der Erarbeitung einer Beweisidee eine besondere Stellung einnehmen kann. Zu den besonderen Beispielen gehört das **kleinste Beispiel**. Bei einem Beispiel zu der zu zeigenden Aussage handelt es sich um das kleinste Beispiel, wenn der Fall der vier aufeinanderfolgenden ungeraden Zahlen $1, 3, 5, 7$ ausgeführt ist. Weiterhin gibt es andere besondere Beispiele, die zur Entwicklung einer ausbaubaren Beweisidee wesentlich beitragen können. Mit Blick auf die in dieser Arbeit fokussierte Allaussage handelt es sich um ein anderes besonderes Beispiel, insbesondere wenn eine Summe von vier gleichen ungeraden Zahlen (z. B. „$1 + 1 + 1 + 1 = 4$“), eine Summe von vier aufeinanderfolgenden geraden Zahlen (z. B. „$2 + 4 + 6 + 8 = 20$“) oder eine Summe von zwei aufeinanderfolgenden ungeraden Zahlen angegeben ist (z. B. „$1 + 3 = 4$“).

Große Beispiele werden im Rahmen dieser Arbeit nicht als besondere Beispiele angesehen, weil die Angabe eines großen Beispiels vielfältige Gründe haben kann und große Beispiele nicht unbedingt eine besondere Stellung bei der Entwicklung einer Beweisidee einnehmen. Mit Blick auf die in dieser Arbeit fokussierte Aussage wird ein Beispiel zu der zu zeigenden Aussage als großes Beispiel bezeichnet, wenn alle Summanden zumindest zweistellig sind und zwischen diesem und den anderen angegebenen Beispielen zu der zu zeigenden Aussage mindestens vier weitere Beispiele dazwischenliegen.

Ein Beispiel wird im Rahmen dieser Arbeit als ein **notwendiges Beispiel** bezeichnet, wenn es aus innerfachlicher Sicht ein notwendiger Bestandteil des Argumentgefüges ist. Insbesondere sind Beispiele, die zugleich einen Induktionsanfang oder einen Fall einer vollständigen Fallunterscheidung darstellen, notwendige Beispiele.

[56] Wie der Ausdruck „zu erkennen ist“ andeutet, wird mit Blick auf die Korrektheit der Beispiele zu der zeigenden Aussage deren Vollständigkeit nicht berücksichtigt. Ein Beispiel zu der zu zeigenden Aussage kann folglich korrekt sein, auch wenn es unvollständig ausgeführt wurde.

Existenz und Passung eines Beispiels zu der zu zeigenden Aussage

Zur Beschreibung, inwieweit in Beweisversuchen mit begründendem Ansatz ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage angegeben und somit die Aussage für einen Spezialfall bestätigt ist, wird jeder Beweisversuch mit begründendem Ansatz genau einer der in Tabelle 1 angegeben Kategorien zugeordnet. Beispielweise gehört die Bearbeitung 3.3.5a in die Kategorie PBsp5 und die Bearbeitung 3.3.5b in die Kategorie PBsp2 (Abb. 43).

PBsp5 ***vollständig und korrekt ausgeführtes eigenes Beispiel zu der zz. Aussage*****:** Es kommt ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zz. Aussage vor (z. B. „$1 + 3 + 5 + 7 = 16, 16 : 8 = 2$“).
PBsp4 ***unvollständig oder fehlerhaft ausgeführtes eigenes Beispiel zu der zz. Aussage*****:** Es kommt ein eigenes Beispiel zu der zu zeigenden Aussage vor, jedoch keines, das vollständig und korrekt ausgeführt wurde (z. B. „$1 + 3 + 5 + 7 = 16 : 8 = 3$“, „$11 + 13 + 15 + 17 = 56$“).
PBsp3 ***vorgegebenes Beispiel zu der zz. Aussage*****:** Es kommt kein eigenes Beispiel zu der zu zeigenden Aussage vor, aber das in der Aufgabenstellung vorgegebene Beispiel der aufeinanderfolgenden ungeraden Zahlen $5, 7, 9, 11$ (z. B. „$5 + 7 + 9 + 11 = 32 : 8 = 4$“).
PBsp2 ***sonstiges Beispiel, Addition von vier ungeraden Zahlen*****:** Es kommt kein Beispiel zu der zu zeigenden Aussage vor. Es wurde aber eine Summe von vier ungeraden Zahlen berechnet (z. B. „$1 + 5 + 11 + 23 = 40$“).
PBsp1 ***sonstiges Beispiel, keine Addition von vier ungeraden Zahlen*****:** Es kommt kein Beispiel zu der zu zeigenden Aussage, aber ein sonstiges Beispiel vor. Es wurde keine Summe von vier ungeraden Zahlen berechnet (z. B. „$7 + 13 = 20$“).
PBsp0 ***kein Beispiel*****:** Es kommt kein Beispiel vor.

Tabelle 1: Kategorien zur Beschreibung der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage insbesondere bei Beweisversuchen mit begründendem Ansatz

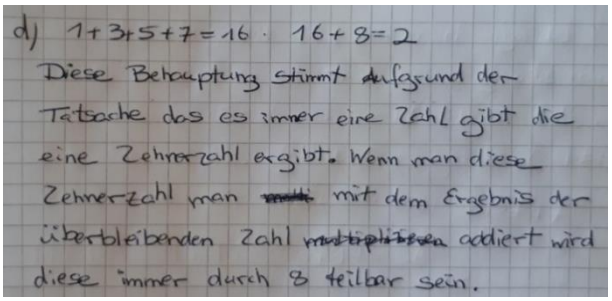
Anzahl an angegebenen Beispielen

Bei jedem Beweisversuch mit begründendem Ansatz wird festgehalten, wie viele Beispiele insgesamt, wie viele Beispiele zu der zu zeigenden Aussage und wie viele sonstige Beispiele in dem Beweisversuch vorkommen. Dazu wird jeweils zwischen der Angabe von keinem, genau einem, zwei bis fünf und mehr als fünf Beispielen unterschieden, wie es exemplarisch an den Bearbeitungen 3.3.5a-b in Abbildung 43 veranschaulicht ist.

Art der angegebenen Beispiele

Bei jedem Beweisversuch mit begründendem Ansatz und Beispiel wird nach der Analyse der Anzahl an Beispielen untersucht, inwieweit die Beispiele eine **Systematik** aufweisen. Dabei wird jeder der Beweisversuche genau einer der Kategorien BspSyst3 *mindestens drei Reihenfolgebeispiele*, BspSyst2 *genau zwei Reihenfolgebeispiele*, BspSyst1 *andere systematische Beispiele* und BspSyst0 *keine systematischen Beispiele* zugeordnet. Zudem wird erfasst, ob das **kleinste Beispiel** der vier aufeinanderfolgenden ungeraden Zahlen $1, 3, 5, 7$ oder ein anderes besonderes Beispiel angegeben ist. Außerdem wird festgehalten, ob ein **großes Beispiel** vorkommt. Die Analyse der Art der angegebenen Beispiele ist an den Bearbeitungen 3.3.5a-b in Abbildung 43 veranschaulicht.

Bearbeitung 3.3.5a: Klassenstufe 8, männlich (Code: 0811802)



Ort der angegebenen Beispiele
- BspOrtoben1 *ein Bsp. oberhalb des (übrigen) Argumentgefüges*
- BspOrtinnen0 *kein Bsp. innerhalb des (übrigen) Argumentgefüges*
- BspOrtunten0 *kein Bsp. unterhalb des (übrigen) Argumentgefüges*
- BspOrtnek0 *kein Bsp. an einem nicht eindeutig kategorisierbaren Ort*
- BspOrtges1 *oberhalb*

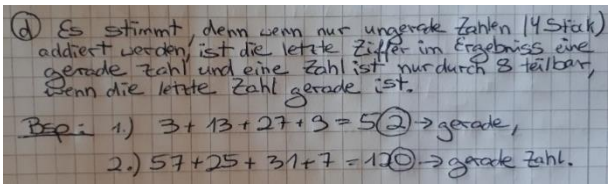
Begründender Ansatz: Beispiele

Existenz, Passung eines Beispiels zu der zz. Aussage
- PBsp5 *vollständig und korrekt ausgeführtes eigenes Beispiel zu der zz. Aussage*

Anzahl an angegebenen Beispielen
- BspN1 *genau ein Beispiel*
- BspNzzA1 *genau ein Beispiel zu der zz. Aussage*
- BspNsonst0 *kein sonstiges Beispiel*

Art der angegebenen Beispiele
- BspSyst0 *keine syst. Beispiele*
- Bspbes2 *kleinstes Beispiel*
- Bspgr0 *kein großes Beispiel*

Bearbeitung 3.3.5b: Klassenstufe 8, weiblich (Code: 0810414)



Ort der angegebenen Beispiele
- BspOrtoben0 *kein Bsp. oberhalb des (übrigen) Argumentgefüges*
- BspOrtinnen0 *kein Bsp. innerhalb des (übrigen) Argumentgefüges*
- BspOrtunten1 *ein Bsp. unterhalb des (übrigen) Argumentgefüges*
- BspOrtnek0 *kein Bsp. an einem nicht eindeutig kategorisierbaren Ort*
- BspOrtges3 *unterhalb*

Begründender Ansatz: Beispiele

Existenz, Passung eines Beispiels zu der zz. Aussage
- PBsp2 *sonstiges Beispiel, Addition von vier ungeraden Zahlen*

Anzahl an angegebenen Beispielen
- BspN2 *zwei bis fünf Beispiele*
- BspNzzA0 *kein Beispiel zu der zz. Aussage*
- BspNsonst2 *zwei bis fünf sonstige Beispiele*

Art der angegebenen Beispiele
- BspSyst0 *keine syst. Beispiele*
- Bspbes0 *kein besonderes Beispiel*
- Bspgr0 *kein großes Beispiel*

Abbildung 43: Bearbeitungen 3.3.5a-b (Beweisversuche mit begründendem Ansatz: Analyse der Verwendung von Beispielen)

Ort der angegebenen Beispiele

Weiterhin wird bei jedem Beweisversuch mit begründendem Ansatz und Beispiel der Ort der angegebenen Beispiele beschrieben. Für die Beschreibung des Orts der angegebenen Beispiele werden die Begriffe „oberhalb“, „innerhalb“, „unterhalb“ und „nicht eindeutig kategorisierbar“ verwendet.

Ein Beispiel gilt als **oberhalb**, wenn es sich räumlich oberhalb des (übrigen) Argumentgefüges befindet oder vom Textfluss her dem (übrigen) Argumentgefüge klar vorausgeht. Ein Beispiel befindet sich also auch oberhalb, wenn dieses vom Textfluss her dem (übrigen) Argumentgefüge klar vorausgeht, aber räumlich in der Zeile angegeben ist, in der das (übrige) Argumentgefüge beginnt. Ein Beispiel wird ebenfalls als oberhalb bezeichnet, wenn ihm ausschließlich eine Angabe eines Wahrheitswerts vorausgeht (z. B. „Denises Aussage ist richtig. Zum Beispiel $9 + 11 + 13 + 15 = 48, 48 : 8 = 6$. Das ist so, weil …“). Ein Beispiel gilt als **unterhalb**, wenn es sich räumlich unterhalb des (übrigen) Argumentgefüges befindet oder vom Textfluss her dem (übrigen) Argumentgefüge klar nachfolgt. Ein Beispiel gilt als **innerhalb**, wenn es im Verhältnis zum (übrigen) Argumentgefüge mittig angeführt ist. Dabei ist ein reiner Verweis auf ein an einem anderen Ort angeführtes Beispiel, etwa mit einer Formulierung der Art „wie man an

dem Beispiel sieht“, nicht ausreichend. Das Beispiel muss zahlenmäßig aufgegriffen sein. In Abgrenzung zu Beispielen, die sich unterhalb befinden, wird der Ort eines Beispiels auch als innerhalb bezeichnet, wenn ein am Ende eines Argumentgefüges angeführtes Beispiel etwa durch eine Formulierung der Art „…, weil z. B. …“ als fester Bestandteil des Argumentgefüges dargestellt ist. Insbesondere gelten Beispiele als innerhalb, die am Ende eines letzten Satzes in Klammern angeführt und somit Teil dieses letzten Satzes sind. Der Ort eines Beispiels gilt als **nicht eindeutig kategorisierbar**, wenn er nicht eindeutig als ober-, inner- oder unterhalb bezeichnet werden kann. Dies gilt u. a. für Beispiele, die sich räumlich gesehen neben dem (übrigen) Argumentgefüge befinden.

Bei der Analyse des Orts der Beispiele wird in zwei Schritten vorgegangen. Zuerst wird nacheinander erfasst, ob ein Beispiel oberhalb des (übrigen) Argumentgefüges angegeben ist, ob eines innerhalb des (übrigen) Argumentgefüges vorkommt, ob sich eines unterhalb des (übrigen) Argumentgefüges befindet und ob ein Beispiel an einem nicht eindeutig kategorisierbaren Ort steht. Danach wird unter Berücksichtigung aller angegebenen Beispiele beschrieben, ob sich diese 1 *oberhalb*, 12 *ober- und innerhalb*, 2 *innerhalb*, 23 *inner- und unterhalb*, 3 *unterhalb*, 13 *ober- und unterhalb*, 123 *ober-, inner- und unterhalb* oder 55 *an einem nicht eindeutig kategorisierbaren Ort* befinden. Beispielsweise ist in der Bearbeitung 3.3.5a in Abbildung 43 ein Beispiel oberhalb des (übrigen) Argumentgefüges angegeben, aber kein Beispiel innerhalb, unterhalb oder an einem nicht eindeutig kategorisierbaren Ort. Kurz gesagt befinden sich also alle Beispiele oberhalb des (übrigen) Argumentgefüges (→ $\mathrm{BspOrtges1}$). In der Bearbeitung 3.3.5b in Abbildung 43 ist weder oberhalb noch innerhalb des (übrigen) Argumentgefüges ein Beispiel angegeben, aber unterhalb des (übrigen) Argumentgefüges sind zwei Beispiele zu finden. Ein Beispiel, das sich bezüglich des Orts nicht eindeutig kategorisieren lässt, kommt nicht vor. Insgesamt befinden sich also alle Beispiele unterhalb des (übrigen) Argumentgefüges (→ $\mathrm{BspOrtges3}$).

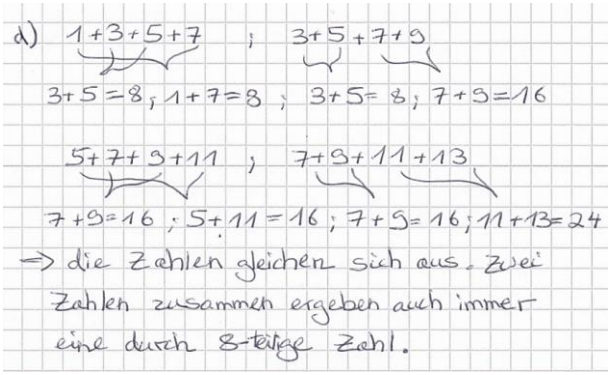
Beispiele oberhalb / unterhalb / innerhalb des (übrigen) Argumentgefüges

Nach der Erfassung des Orts der angegebenen Beispiele werden nacheinander die oberhalb, die unterhalb und die innerhalb des (übrigen) Argumentgefüges angegebenen Beispiele fokussiert und jeweils als Einheit separat analysiert.

Zuerst werden die Beispiele **oberhalb** des (übrigen) Argumentgefüges fokussiert und als Einheit analysiert. Analog zu der Untersuchung der Anzahl und Art aller angegebenen Beispiele werden die Anzahl und Art der oberhalb des (übrigen) Argumentgefüges angeführten Beispiele codiert. Zusätzlich wird analysiert, ob an einem der Beispiele oberhalb des (übrigen) Argumentgefüges aus Expertensicht eine Beweisidee hervorgehoben ist.[57] Eine Beweisidee kann z. B. durch nicht unbedingt notwendige Zwischenschritte oder die Verwendung ikonischer Darstellungsmittel aus Expertensicht hervorgehoben sein (z. B. „$5 + 7 + 9 + 11 = 12 + 20 = 12 + 12 + 8 = 8 + 4 + 8 + 4 + 8 = 4 \cdot 8$“). Demgegenüber wird eine reine Angabe von Beispielen zu der zu

[57] Es ist möglich, dass ein Beispiel von einer Autorin bzw. einem Autor zur Veranschaulichung einer Beweisidee angeführt wurde oder an einem Beispiel eine Beweisidee erkannt wurde, die Beweisidee jedoch aus Expertensicht nicht hervorgehoben ist. Ebenfalls ist es möglich, dass aus Expertensicht an einem Beispiel eine Beweisidee hervorgehoben ist, ohne dass dies von der Autorin bzw. dem Autor beabsichtigt war.

Bearbeitung 3.3.5c: Klassenstufe 7, weiblich (Code: 0721706)



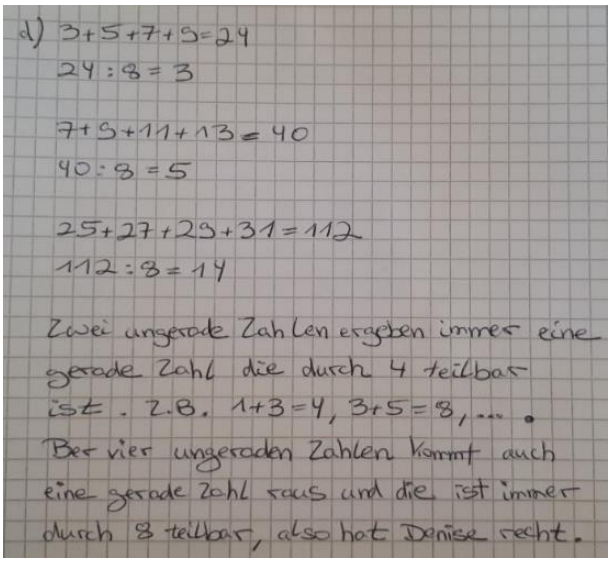
Begründender Ansatz: Beispiele oberhalb des (übrigen) Argumentgefüges

Anzahl an angegebenen Beispielen
- BspobenN2 *zwei bis fünf Beispiele*
- BspobenNzzA2 *zwei bis fünf Beispiele zu der zu zeigenden Aussage*
- BspobenNsonst0 *kein sonstiges Beispiel*

Art der angegebenen Beispiele
- BspobenSyst3 *mindestens drei Reihenfolgebeispiele*
- Bspobenbes2 *kleinstes Beispiel*
- Bspobengr0 *kein großes Beispiel*

Hervorhebung einer Beweisidee
- BspobenBI1 *eine Beweisidee hervorgehoben*

Bearbeitung 3.3.5d: Klassenstufe 7, weiblich (Code: 0721801)



Integration der Beispiele in das Argumentgefüge
- Bspinnennotw0 *kein notwendiges Beispiel*
- BspinnenTeilAG0 *kein Bsp. fester Bestandteil des Argumentgefüges*
- Bspinnenerg1 *ein ergänzendes Beispiel*
- Bspinnenweg0 *kein Beispiel weglassbar*

Begründender Ansatz: Beispiele oberhalb des (übrigen) Argumentgefüges

Anzahl an angegebenen Beispielen:
- BspobenN2 *zwei bis fünf Beispiele*
- BspobenNzzA2 *zwei bis fünf Beispiele zu der zu zeigenden Aussage*
- BspobenNsonst0 *kein sonstiges Beispiel*

Art der angegebenen Beispiele
- BspobenSyst0 *keine Systematik erkennbar*
- Bspobenbes0 *kein besonderes Beispiel*
- Bspobengr1 *ein großes Beispiel*

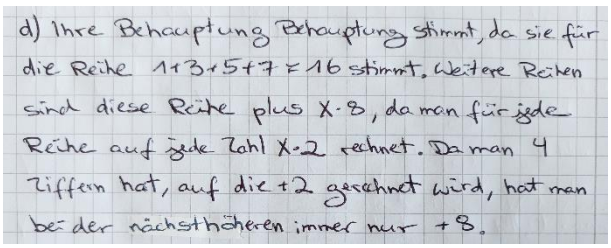
Hervorhebung einer Beweisidee
- BspobenBI0 *keine Beweisidee hervorgehoben*

Begründender Ansatz: Beispiele innerhalb des (übrigen) Argumentgefüges

Anzahl an angegebenen Beispielen:
- BspinnenN2 *zwei bis fünf Beispiele*
- BspinnenNzzA0 *kein Beispiel zu der zz. Aussage*
- BspinnenNsonst2 *zwei bis fünf sonstige Beispiele*

Art der angegebenen Beispiele
- BspinnenSyst1 *andere systematische Beispiele*
- Bspinnenbes1 *anderes besonderes Beispiel*
- Bspinnengr0 *kein großes Beispiel*

Bearbeitung 3.3.5e: Klassenstufe 8, männlich (Code: 0811002)



Integration der Beispiele in das Argumentgefüge
- Bspinnennotw1 *ein notwendiges Beispiel*
- BspinnenTeilAG1 *ein Bsp. fester Bestandteil des Argumentgefüges*
- Bspinnenerg0 *kein ergänzendes Beispiel*
- Bspinnenweg0 *kein Beispiel weglassbar*

Begründender Ansatz: Beispiele innerhalb des (übrigen) Argumentgefüges

Anzahl an angegebenen Beispielen
- BspobenN1 *genau ein Beispiel*
- BspobenNzzA1 *genau ein Beispiel zu der zu zeigenden Aussage*
- BspobenNsonst0 *kein sonstiges Beispiel*

Art der angegebenen Beispiele
- BspobenSyst0 *keine systematischen Beispiele*
- Bspobenbes2 *kleinstes Beispiel*
- Bspobengr0 *kein großes Beispiel*

Abbildung 44: Bearbeitungen 3.3.5c-e (Beweisversuche mit begründendem Ansatz: Analyse der Beispiele oberhalb des (übrigen) Argumentgefüges und Analyse der Beispiele innerhalb des (übrigen) Argumentgefüges)

zeigenden Aussage, d. h. auch eine reine Angabe von Reihenfolgebeispielen, nicht als ausreichend angesehen. Die gesamte Analyse der Beispiele oberhalb des (übrigen) Argumentgefüges ist in Abbildung 44 exemplarisch an zwei Bearbeitungen ausgeführt.

Im Anschluss an die Analyse der oberhalb des (übrigen) Argumentgefüges befindlichen Beispiele werden vollständig analog die **unterhalb** des (übrigen) Argumentgefüges angegebenen Beispiele fokussiert und als Einheit analysiert.

Abschließend werden die Beispiele **innerhalb** des (übrigen) Argumentgefüges fokussiert und ebenfalls als Einheit analysiert. Die Analyse der Beispiele innerhalb des (übrigen) Argumentgefüges weicht ausschließlich in einem Punkt von der Analyse der Beispiele oberhalb bzw. unterhalb des (übrigen) Argumentgefüges ab. Anstatt der Hervorhebung einer Beweisidee wird die Integration der Beispiele in das Argumentgefüge analysiert. Erstens wird erfasst, ob ein notwendiges Beispiel[58] vorkommt. Zweitens wird festgehalten, ob eines der innerhalb des (übrigen) Argumentgefüges angeführten Beispiele insofern ein fester Bestandteil des Argumentgefüges ist, als dass das Weglassen dieses Beispiels zu unvollständigen Sätzen oder Unterbrechungen im Argumentgefüge führen würde. Drittens wird angegeben, ob ein Beispiel innerhalb des (übrigen) Argumentgefüges vorkommt, das die Ausführungen inhaltlich ergänzt, aber kein notwendiges Beispiel ist. Viertens wird codiert, ob eines der Beispiele innerhalb des (übrigen) Argumentgefüges weggelassen werden könnte, ohne die Struktur des Argumentgefüges oder die im Argumentgefüge dargelegten inhaltlichen Überlegungen zu beeinflussen. Die gesamte Analyse der Beispiele innerhalb des (übrigen) Argumentgefüges ist in Abbildung 44 anhand zweier Bearbeitungen veranschaulicht.

3.3.5.3 Hinweise zu den Kategorien

Der erste Schritt der Analyse der Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz, d. h. die Analyse der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage, legt offen, ob ein Beispiel vorkommt, und insbesondere eines, das die zu zeigende Aussage bestätigt. Dies ist mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage vor allem von Bedeutung, wenn ein Beweisversuch zwar einen begründenden Ansatz, aber kein weiterentwickelbares Argumentgefüge enthält. Die Analyse zeigt, inwieweit es gelungen ist, die Aussage für einen Spezialfall zu bestätigen (Abschn. 8.2.3.1, 8.2.5.1 für die Ergebnisse).

In allen weiteren Analyseschritten werden die Beweisversuche mit begründendem Ansatz, die ein Beispiel enthalten, fokussiert und Eigenschaften der angegebenen Beispiele analysiert. Dabei wird, wie beschrieben, zwischen Beispielen ober-, inner- und unterhalb des (übrigen) Argumentgefüges differenziert. Davon ausgehend, dass ein Beweisversuch entsprechend der Anordnung der Ausführungen in der Bearbeitung von oben nach unten verfasst wurde, macht es einen Unterschied, wo sich die Beispiele im Verhältnis zum (übrigen) Argumentgefüge befinden. Während oberhalb angeführte Beispiele tendenziell dem Verständnis einer infrage ste-

[58] Ein Beispiel wird im Rahmen dieser Arbeit als notwendiges Beispiel bezeichnet, wenn es aus innerfachlicher Sicht ein notwendiger Bestandteil des Argumentgefüges ist. Bei notwendigen Beispielen handelt es sich weniger um Beispiele im Sinne einer (numerischen) Konkretisierung und Exemplifizierung einer Idee als etwa um einen Induktionsanfang oder einen Fall einer vollständigen Fallunterscheidung.

henden Aussage, deren Plausibilisierung oder der Entwicklung einer Beweisidee dienen, können unterhalb angeführte Beispiele eher zur Veranschaulichung des Argumentgefüges für Lesende, als Kontrolle der zuvor ausgeführten Überlegungen oder als eine Art abschließende Bestätigung der Gültigkeit der zu zeigenden Aussage gedacht sein.[59] Damit einhergehend liegt die Vermutung nahe, dass ober-, inner- und unterhalb angeführte Beispiele unterschiedliche Charakteristika aufweisen. Während als ein mögliches Charakteristikum bei den Beispielen oberhalb und denen unterhalb des (übrigen) Argumentgefüges die Hervorhebung einer Beweisidee analysiert wird, wird bei den Beispielen innerhalb des (übrigen) Argumentgefüges die Integration in das Argumentgefüge betrachtet, weil Beispiele innerhalb des (übrigen) Argumentgefüges i. d. R. per se in Zusammenhang mit einer oder mehreren Beweisideen zu sehen sind. Die Ergebnisse der Erfassung der Eigenschaften der angegebenen Beispiele sollen einen Einblick geben, in welcher Form Beispiele in Beweisversuchen mit begründendem Ansatz häufig oder eher selten vorkommen, und somit zur Beantwortung der Frage nach Charakteristika der Beweisversuche beitragen. Nebenbei können die Ergebnisse Anhaltspunkte geben, warum die Beispiele angeführt wurden (Abschn. 8.2.3.2, 8.2.3.3 für die Ergebnisse).

Welche Funktion den Beispielen nicht aus der Perspektive der Lernenden, sondern aus der Perspektive von Lesenden zukommt und welche Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis der zu zeigenden Aussage durch die angegebenen Beispiele zu erkennen sind, lässt sich in einer ganzheitlichen Betrachtung einzelner Beweisversuche im Gesamten herausarbeiten. Um Einblicke in verschiedene Anknüpfungspunkte zu geben, die sich aus den angegebenen Beispielen ergeben können, werden deshalb im Rahmen der weiteren Auswertung und Ergebnispräsentation Einzelfallanalysen durchgeführt (Abschn. 7.2.1.6, 7.3.8 zur Methodik, 8.2.3.3 für die Ergebnisse).

3.3.6 Begründender Ansatz: Schlussweisen

Insbesondere die beschriebene Untersuchung der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz erfolgt unabhängig von der verwendeten Schlussweise, die in einem separaten Analyseschritt untersucht werden soll (Abschn. 3.3.4; Abb. 45). Dafür wird im Folgenden eine Ausdifferenzierung von Schlussweisen in Beweisversuchen mit begründendem Ansatz eingeführt. Der erste Unterabschnitt beschäftigt sich mit den Überlegungen, die bei der Entwicklung dieser Ausdifferenzierung leitend waren. Daran anknüpfend werden die einzelnen Kategorien definiert, und abschließend kommentiert. Die Ergebnisse der Analyse der Schlussweise bei Beweisversuchen mit begründendem Ansatz werden in Abschnitt 8.2.1 präsentiert.

[59] Die Ergebnisse der Beschreibung der Verwendung von Beispielen in Beweisversuchen sind diesbezüglich ausschließlich als Anhaltspunkte zu verstehen. Sichere Aussagen sind entsprechend den anfänglichen Ausführungen zu Funktionen von Beispielen in Abschnitt 3.3.5.1 nicht möglich.

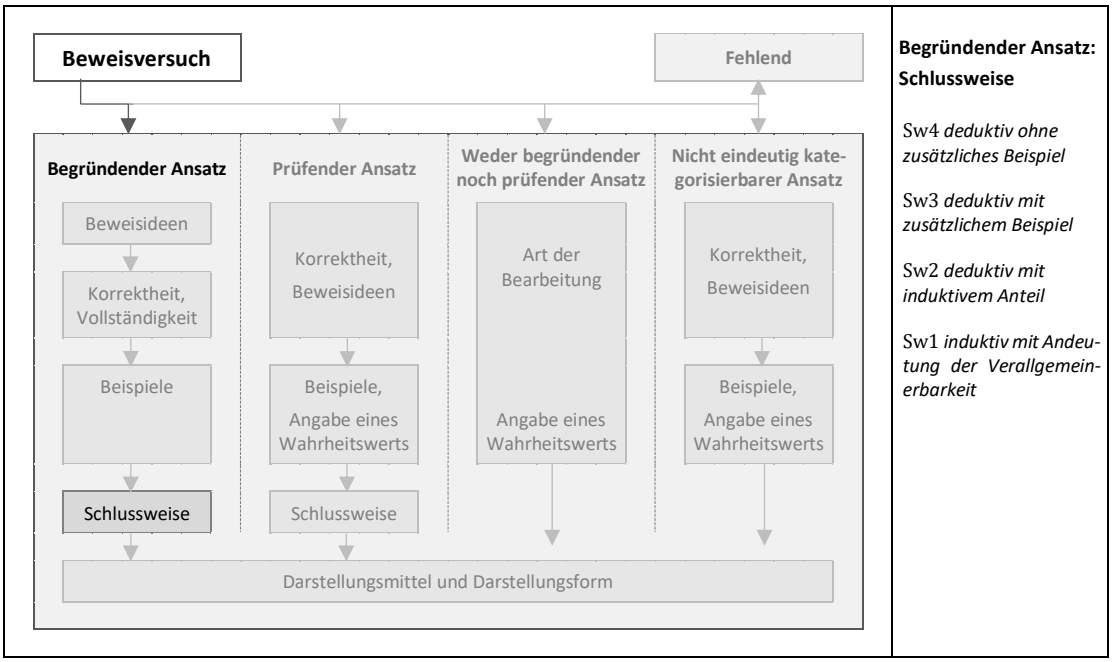


Abbildung 45: Analysemodell. Fokus: Begründender Ansatz, Schlussweise

3.3.6.1 Entwicklung der Kategorien

Zu Beweisversuchen mit begründendem Ansatz gehören Beweisversuche, in denen an Beispielen eine Beweisidee erkennbar ist, deren Verallgemeinerbarkeit angedeutet wird, bis hin zu Beweisversuchen mit rein deduktiver Schlussweise (Abschn. 3.3.2). Bei der Ausdifferenzierung von Beweisversuchen mit begründendem Ansatz sollen Beweisversuche mit deduktiver Schlussweise und solche, in denen die Beweisideen primär aus Beispielen hervorgehen, zwei verschiedene Unterkategorien bilden.

Als Grundlage dafür ist der **Begriff „Schlussweise“** zu konkretisieren. Welche Schlussweise einem in rein schriftlicher Form vorliegenden Beweisversuch mit begründendem Ansatz tatsächlich zugrunde liegt, lässt sich i. d. R. nicht mit Sicherheit sagen. Zum Beispiel können Aussagen, wie etwa „Die Summe zweier aufeinanderfolgender ungerader Zahlen ist immer durch vier teilbar“, auf einem induktiven Schluss basieren, deduktiv hergeleitet worden sein oder zur Wissensbasis des Lernenden gehören. Wegen dieser Unsicherheit wird der Begriff „Schlussweise“ im Rahmen dieser Arbeit explizit auf das eingeschränkt, was bei den Beweisversuchen mit begründendem Ansatz erkennbar ist, nämlich das Verhältnis angegebener Beispiele zu allgemein gehaltenen Ausführungen. Trotzdem werden in den Kategorienbezeichnungen möglichst übliche Termini, wie „deduktiv“ und „induktiv“, verwendet.

Eine Deduktion ist vereinfacht gesagt ein „Schluss von einem Fall und einem Gesetz auf ein Resultat“ (Meyer & Voigt, 2008, S. 133). Bei mathematischen Beweisen handelt es sich um ein Gefüge von Deduktionen. Die einzelnen Deduktionen sind üblicherweise miteinander verknüpft, sodass beispielsweise das Resultat einer Deduktion eine Prämisse der nächsten Deduktion bildet. In Beweisversuchen von Schülerinnen und Schülern ist es aber ebenfalls möglich, dass Deduktionen ohne explizite Verbindung aufeinander folgen oder einzelne Elemente von Deduktionen fehlen. Zudem ist die Qualität der einzelnen Elemente einer Deduktion für die

Analyse von Beweisversuchen von Lernenden im Rahmen dieser Arbeit zu erweitern. Es wird auch als Deduktion angesehen, wenn verwendete Prämissen Fehler enthalten oder die angeführten Konklusionen nicht zwingend daraus folgen. Die Schlussweise eines Beweisversuchs mit begründendem Ansatz wird in dieser Arbeit als **deduktiv** bezeichnet, wenn alle Argumente des Argumentgefüges im beschriebenen Sinne deduktiv sind.

Unter einer Induktion versteht man den Schluss von einem Fall und einem Resultat auf ein Gesetz (Meyer & Voigt, 2008). Sie lässt sich mit dem Motto „Was einmal gilt, gilt immer“ oder etwa „Was dreimal gilt, gilt immer“ zusammenfassen. Es wird also von Konkretem, wie der Gültigkeit einer Aussage für einen speziellen Fall, auf Allgemeines geschlossen. Bei Beweisversuchen, in denen an Beispielen eine Beweisidee erkennbar ist, deren Verallgemeinerbarkeit angedeutet wird, ist die Schlussweise vordergründig induktiv. Die Bearbeitungen weisen jedoch zwei Besonderheiten auf, die Existenz einer Beweisidee und die explizite Andeutung in Bezug auf deren Verallgemeinerbarkeit, weshalb die Schlussweise im Folgenden mit dem Ausdruck „**induktiv mit Andeutung der Verallgemeinerbarkeit**“ beschrieben wird.

Weiterhin ist es möglich, dass ein Beweisversuch mit begründendem Ansatz hinsichtlich der Schlussweise **zwischen deduktiv und induktiv mit Andeutung der Verallgemeinerbarkeit** liegt. Um auch bei solche Beweisversuchen die Schlussweise zu beschreiben, werden die Ansätze von Biehler und Kempen (2013) sowie Krumsdorf (2017) herangezogen, da sie eine Abgrenzung verschiedener Vorformen wissenschaftlicher Beweise vornehmen. Biehler und Kempen (2013) unterscheiden zwischen generischen Beispielen und generischen Beweisen, wobei sich generische Beweise durch eine zusätzliche Ausformulierung des im generischen Beispiel prinzipiell erkennbaren, allgemeinen Argumentgefüges auszeichnen. In den Begrifflichkeiten von Krumsdorf (2017) sind generische Beweise i. d. R. nicht beispielgebunden, zumindest wenn das Argumentgefüge vollständig auch in allgemein gehaltener Sprache formuliert ist. Krumsdorf (2017) bezeichnet ein Argumentgefüge als beispielgebunden, sobald ein einzelnes Argument eines Argumentgefüges beispielgebunden ist (Abschn. 3.3.2.1). Die in beiden Ansätzen zentrale Frage, inwieweit ein vorliegendes Argumentgefüge in allgemein gehaltener Sprache formuliert ist, wird im Rahmen dieser Arbeit genutzt, um Zwischenkategorien zur Beschreibung der Schlussweise in Beweisversuchen mit begründendem Ansatz zu bilden. Zudem wird bei der Kategorienbildung berücksichtigt, dass in einem Beweisversuch mit deduktiver Schlussweise aus Expertensicht überflüssige Beispiele angegeben sein können, die jedoch aus Sicht der Autorin bzw. des Autors einen wesentlichen Bestandteil des Argumentgefüges darstellen, weil sie z. B. als Stützung oder als „Beweis“ einer verwendeten Aussage gedacht sind. Im Folgenden werden die Kategorien zur Beschreibung der Schlussweise in Beweisversuchen mit begründendem Ansatz zunächst vorgestellt und anschließend kommentiert.

3.3.6.2 Vorstellung der Kategorien

Der Begriff „Schlussweise“ bezeichnet im Rahmen dieser Arbeit bezogen auf Beweisversuche mit begründendem Ansatz das Verhältnis angegebener Beispiele zu allgemein gehaltenen inhaltlichen Überlegungen. Bei der Analyse der Schlussweise wird jeder Beweisversuch mit begründendem Ansatz genau einer der folgenden vier Kategorien zugeordnet: Sw1 *induktiv mit*

Andeutung der Verallgemeinerbarkeit, Sw2 *deduktiv mit induktivem Anteil*, Sw3 *deduktiv mit zusätzlichem Beispiel*, Sw4 *deduktiv ohne zusätzliches Beispiel*.

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **Sw1** ***induktiv mit Andeutung der Verallgemeinerbarkeit*****,** wenn das Argumentgefüge primär auf Schlüssen beruht, die an konkreten Beispielen durchgeführt werden, und die Verallgemeinerbarkeit dieser Schlüsse postuliert wird. Ob die erkennbaren Überlegungen tatsächlich auf alle relevanten Fälle übertragbar sind und inwieweit deren Verallgemeinerung ein korrektes und vollständiges Argumentgefüge ergeben, wird nicht berücksichtigt. In Abgrenzung zu der Kategorie Sw2 *deduktiv mit induktivem Anteil* wird die Verallgemeinerbarkeit der Überlegungen nur durch Ausdrücke wie „z. B.“, „und so weiter“, „immer“ oder eine im Wesentlichen auf einen Satz beschränkte, allgemein gehaltene Skizze einer Beweisidee angedeutet. Alternativ gehen aus Expertensicht die an den angegebenen Beispielen dargestellten oder aus den Beispielen abstrahierbaren inhaltlichen Überlegungen deutlich über die allgemein gehaltenen Ausführungen hinaus und ergänzen die allgemein gehaltenen Ausführungen wesentlich. Beispielsweise ist in der Bearbeitung 3.3.6a in Abbildung 46 ein Argumentgefüge überwiegend an Beispielen dargestellt, wobei der Satz „Denise hat Recht, denn die Zahl wird immer um 2, 4, 6, 8 usw größer“ auf dessen Verallgemeinerbarkeit hinweist.

Bearbeitung 3.3.6a: Klassenstufe 7, weiblich (Code: 0710112)

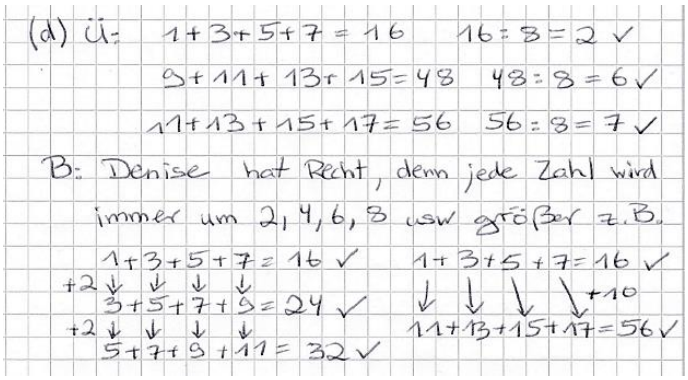


Abbildung 46: Bearbeitung 3.3.6a (Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit*)

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **Sw2** ***deduktiv mit induktivem Anteil***, wenn die Schlussweise überwiegend deduktiv, aber ein Teil der Schlüsse an konkreten Beispielen ausgeführt oder ein Beispiel, das die Ausführungen inhaltlich ergänzt, angegeben ist. Im Vergleich zu der Kategorie Sw3 *deduktiv mit zusätzlichem Beispiel* können auch die Beispiele, die kein notwendiger Bestandteil des Argumentgefüges sind, nicht alle weggelassen werden, ohne die Struktur des Argumentgefüges oder die im Argumentgefüge dargelegten inhaltlichen Überlegungen zu beeinträchtigen. In der Bearbeitung 3.3.6b in Abbildung 47 zum Beispiel wird die Idee der Vergrößerung der Summe um acht anhand der aufeinanderfolgenden ungeraden Zahlen $1, 3, 5, 7$ und $3, 5, 7, 9$ sowie $5, 7, 9, 11$ erläutert. Ein Streichen der mit „(2)“ gekennzeichneten beispielgebundenen Ausführungen würde die Vollständigkeit des Argumentgefüges beeinflussen.

Bearbeitung 3.3.6b: Klassenstufe 8, männlich (Code: 0821910)

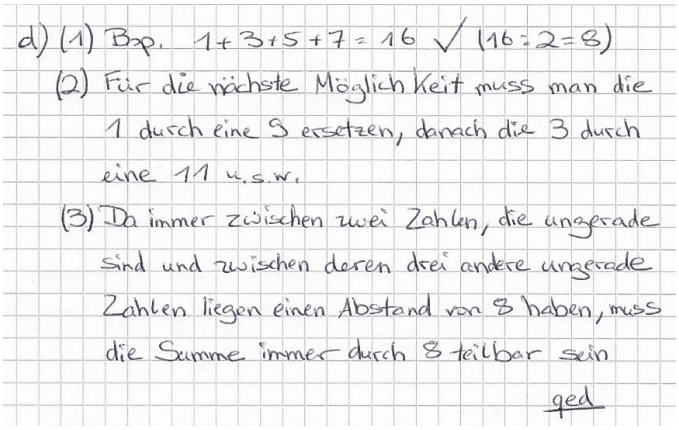


Abbildung 47: Bearbeitung 3.3.6b (Sw2 *deduktiv mit induktivem Anteil*)

Ein Beweisversuch mit *begründendem Ansatz* gehört in die Kategorie **Sw3 *deduktiv mit zusätzlichem Beispiel***, wenn das Argumentgefüge nur deduktive Schlüsse enthält, aber ein Beispiel vorkommt, das kein notwendiger Bestandteil des Argumentgefüges ist. Die angegebenen Beispiele können von außen betrachtet sowohl für sich selbst stehen als auch die Beweisidee veranschaulichen. Sie können darauf hinweisen, wie die zu zeigenden Aussage interpretiert wurde. Alle Beispiele, die kein notwendiger Bestandteil des Argumentgefüges sind, könnten aber weggelassen werden, ohne die deduktive Struktur des Argumentgefüges zu unterbrechen oder die im Argumentgefüge dargelegten inhaltlichen Überlegungen zu beeinträchtigen. Dies trifft zum Beispiel bei der Bearbeitung 3.3.6c in Abbildung 48 zu.

Bearbeitung 3.3.6c: Klassenstufe 7, männlich (Code: 0710804)

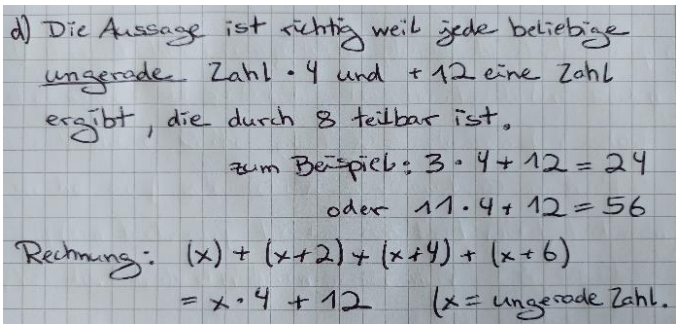


Abbildung 48: Bearbeitung 3.3.6c (Sw3 *deduktiv mit Angabe eines Beispiels*)

Ein Beweisversuch mit begründendem Ansatz gehört in die Kategorie **Sw4 *deduktiv ohne zusätzliches Beispiel***, wenn jeder Schluss im Argumentgefüge deduktiv ist. Dabei müssen die einzelnen Elemente der Deduktionen nicht unbedingt sachlich korrekt sein, und auch wenn die Prämissen korrekt sind, müssen die angeführten Konklusionen nicht zwingend daraus folgen. Auch ist es möglich, dass einzelne Elemente deduktiver Schlüsse oder die Verknüpfungen dieser Elemente nur implizit vorkommen oder fehlen. In Abgrenzung zu der Kategorie Sw3 *deduktiv mit zusätzlichem Beispiel* kommen in Beweisversuchen mit deduktiver Schlussweise allenfalls Beispiele vor, die notwendige Bestandteile des Argumentgefüges sind. Beispiele sind not-

wendige Bestandteile des Argumentgefüges, insbesondere wenn sie im Rahmen einer vollständigen Fallunterscheidung als Induktionsanfang fungieren oder Fälle einer vollständigen Fallunterscheidung bilden. Beispielsweise fällt die Bearbeitung 3.3.6d in Abbildung 49 in die Kategorie Sw4 *deduktiv ohne zusätzliches Beispiel*.

Bearbeitung 3.3.6d: Klassenstufe 8, weiblich (Code: 0811311)

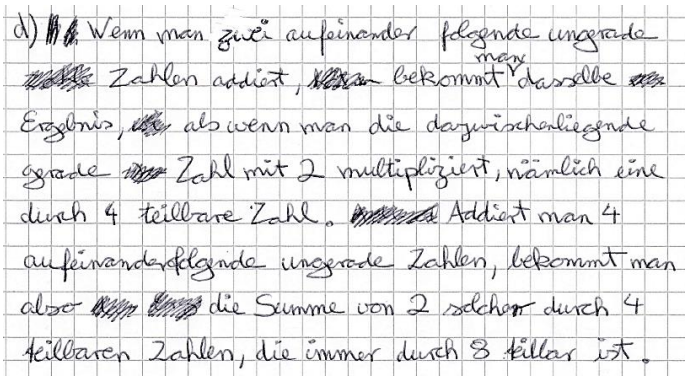


Abbildung 49: Bearbeitung 3.3.6d (Sw4 *deduktiv ohne zusätzliches Beispiel*)

3.3.6.3 Hinweise zu den Kategorien

Zur Beschreibung der Schlussweise bei Beweisversuchen mit begründendem Ansatz werden vier Kategorien differenziert: Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit*, Sw2 *deduktiv mit induktivem Anteil*, Sw3 *deduktiv mit zusätzlichem Beispiel*, Sw4 *deduktiv ohne zusätzliches Beispiel*. Dabei sind die Kategorien Sw4 *deduktiv ohne zusätzliches Beispiel* und Sw3 *deduktiv mit zusätzlichem Beispiel* ähnlich, da das Argumentgefüge jeweils nur deduktive Schlüsse enthält, die Schlussweise also deduktiv ist und Beweisversuche mit beiden Schlussweisen Sw3-4 prinzipiell als Beweis gelten können (Abschn. 3.4). Trotzdem werden diese Kategorien nicht zusammengefasst, weil die Kategorie Sw3 *deduktiv mit zusätzlichem Beispiel* eine Zwischenstufe zwischen den Kategorien Sw2 und Sw4 bildet. Die Angabe eines oder mehrerer Beispiele in Beweisversuchen mit der Schlussweise Sw3 kann u. A. damit zusammenhängen, dass die Überlegungen des Lernenden eventuell noch nicht vollständig von Beispielen abgelöst sind. Zugleich bedeutet die Zuordnung eines Beweisversuchs zu der Kategorie Sw3 *deduktiv mit zusätzlichem Beispiel*, dass die Schülerin bzw. der Schüler ein deduktives, wenn auch nicht unbedingt korrektes und vollständiges Argumentgefüge formuliert hat. Dies trifft bei Beweisversuchen in den Kategorien Sw2 *deduktiv mit induktivem Anteil* und Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit* nicht zu. Ähnlich wie bei beispielgebundenen Beweisen im Sinne von Krumsdorf (2017) sowie generischen Beispielen nach Biehler und Kempen (2013) ist bei Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise (Sw1-2) das erkennbare Argumentgefüge nicht in Gänze allgemeingültig ausgeführt. Dabei beschreiben die in dieser Arbeit verwendeten Termini, wie z. B. „deduktiv mit induktivem Anteil“, anders als bei den Ansätzen von Krumsdorf (2017) und Kempen (2013), ausschließlich die Schlussweise. Inwieweit die an Beispielen erkennbaren oder allgemein formulierten Überlegungen ein korrektes und vollständiges Argumentgefüge darstellen, ist bei der Beschreibung der Schlussweise unrelevant.

Zusammen betrachtet unterscheiden sich die Kategorien Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit*, Sw2 *deduktiv mit induktivem Anteil*, Sw3 *deduktiv mit zusätzlichem Beispiel* und Sw4 *deduktiv ohne zusätzliches Beispiel* graduell hinsichtlich der Abstraktheit der Argumentgefüge sowie der Explikation der Allgemeingültigkeit. Die Weiterentwicklung eines Argumentgefüges, bei dem die Schlussweise als induktiv mit Andeutung der Verallgemeinerbarkeit bezeichnet wird, hin zu einem deduktiven Argumentgefüge erfordert eine zunehmende Dekontextualisierung, d. h. eine Ablösung der Überlegungen von den spezifischen Besonderheiten eines Beispiels und eine Verallgemeinerung derselben (Balacheff, 1988). Mit dieser zunehmenden Abstraktion sind auch höhere Anforderungen an die Versprachlichung verbunden, weshalb ein deduktives Argumentgefüge aus sprachlicher Sicht anspruchsvoller erscheint. Ein Beweisversuch mit deduktiver Schlussweise ist aber nicht grundsätzlich besser als eine Bearbeitung, bei der die Schlussweise etwa als deduktiv mit induktivem Anteil zu bezeichnen ist, da die Einordnung der Schlussweise unabhängig von der Analyse der Beweisideen und der Korrektheit und Vollständigkeit von Beweisversuchen erfolgt. Schließlich ist zu beachten, dass es bei Beweisversuchen mit begründendem Ansatz ein Kontinuum denkbarer Schlussweisen gibt, in dem lediglich mit dem Ziel der Kategorienbildung möglichst inhaltlich aussagekräftige und anwendbare Abgrenzungen getroffen wurden. Zwei Beweisversuche innerhalb einer Kategorie sind auch bezüglich der Schlussweise nicht unbedingt gleich und zwei Beweisversuche aus verschiedenen Kategorien nicht unbedingt deutlich verschieden.

3.3.7 Prüfender Ansatz: Beispiele und Angabe eines Wahrheitswerts

Dieser und der nachfolgende Abschnitt fokussieren auf Beweisversuche mit prüfendem Ansatz. Im Vergleich zu Beweisversuchen mit begründendem Ansatz ist in Beweisversuchen mit prüfendem Ansatz nicht unbedingt eine Beweisidee zu erkennen oder deren Verallgemeinerbarkeit nicht explizit angedeutet. Beweisversuche mit prüfendem Ansatz charakterisieren sich insbesondere durch die Existenz eines Beispiels, das einen klaren Bezug zu der infrage stehenden Aussage aufweist (Abschn. 3.3.2). Offen ist dabei, ob ein Beispiel die infrage stehende Aussage bestätigt. Zudem können sich die Beweisversuche mit prüfendem Ansatz hinsichtlich diverser Eigenschaften, wie der Anzahl und Systematik angegebener Beispiele, sowie daraus resultierender Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis unterscheiden.

Dieser und der nachfolgende Abschnitt beschreiben das Kategoriensystem für die detaillierte Analyse der Beweisversuche mit prüfendem Ansatz (Abb. 50; Abschn. 3.3.8). Die Analyse soll Einblicke geben, inwieweit die zu zeigende Aussage in den Beweisversuchen mit prüfendem Ansatz für einen Spezialfall bestätigt wird und welche Eigenschaften die Beweisversuche mit besonderem Fokus auf Anknüpfungspunkte für die Entwicklung eines Beweises aufweisen. Die Ergebnisse werden in Abschnitt 8.3 vorgestellt.

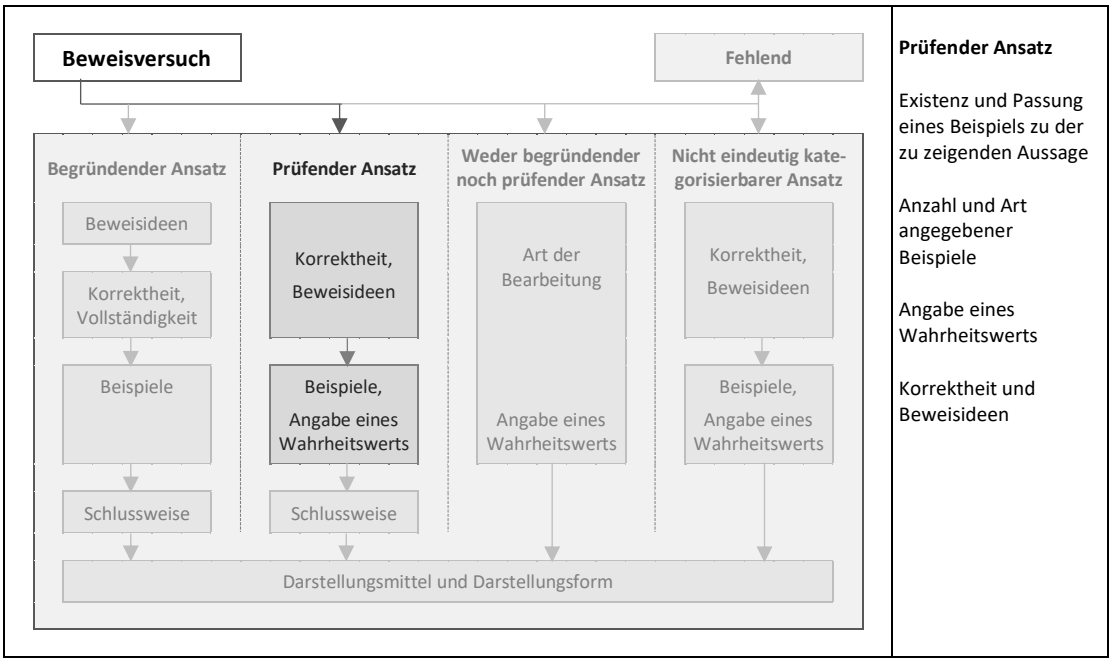


Abbildung 50: Analysemodell. Fokus: Prüfender Ansatz, Korrektheit, Beweisideen, Beispiele und Angabe eines Wahrheitswerts

3.3.7.1 Entwicklung der Kategorien

Für die Analyse von Eigenschaften der Beweisversuche mit prüfendem Ansatz wird zuerst die **mathematikdidaktische Literatur** hinsichtlich möglicher, relevanter Facetten solcher Beweisversuche gesichtet.

Beispielsweise berücksichtigt Bell (1976) in seinem Kategoriensystem, dass die angegebenen Beispiele Fehler enthalten können: „*Empirical O*: Failure to generate correct examples or to comply with given conditions“ (Bell, 1976, S. 28, H. i. O.). Weiterhin beschreibt Bell (1976) mit den Kategorien „non-systematic“ (Bell, 1976, S. 28), „partially systematic“ (ebd.) und „systematic“ (ebd.), inwieweit angegebene Beispiele alle möglichen Fälle einer Menge abdecken. Chazan (1993) bezeichnet als „kinds or types“ (Chazan, 1993, S. 369), wenn Lernende versuchen, bei der Überprüfung einer Allaussage jede Art von Fällen, wie z. B. jede Art von Dreiecken, zu berücksichtigen. Dabei werden im Vergleich zu den Kategorien von Bell (1976) nicht mehrere Fälle einer Teilmenge, sondern Einzelfälle aus verschiedenen Teilmengen betrachtet. Die Kategorie „extending a pattern“ nach Smith und Henderson (1959) charakterisiert sich insbesondere durch die Betrachtung aufeinanderfolgender Beispiele aus der infrage stehenden Grundgesamtheit (Reid & Knipping, 2010).[60]

Balacheff (1988) unterscheidet zwischen den Kategorien „naive empiricism“ (Balacheff, 1988, S. 218) und „crucial experiment“ (ebd.). Diese differieren hinsichtlich der Anzahl an betrachteten Beispielen und der Explizierung der Frage nach der Allgemeingültigkeit:

> Naive empiricism consists of asserting the truth of a result after verifying several cases. […] We use the expression [‘crucial experiment’] for a slightly different process, one of verifying a proposition on an instance which ‘doesn’t come for free’, asserting that ‘if it works here, it will always

[60] Die Originalquelle ist für die Autorin der vorliegenden Arbeit nicht zugänglich, weshalb auf die Ausführungen von Reid und Knipping (2010) verwiesen wird.

> work'. […] This type of validation is distinguishable from naive empiricism in that the pupil poses explicitly the problem of generality and resolves it by staking all on the outcome of a particular case that she recognises to be not too special. (Balacheff, 1988, S. 218-219)

Beide Eigenschaften, die Anzahl an Beispielen und der Grad der Berücksichtigung der Frage nach der Allgemeingültigkeit, sind auch im Begriffsnetz von Meyer und Voigt (2009) relevant. Sie unterscheiden u. a. zwischen einer Prüfung an einem speziellen Fall, an einem typisch anderen Fall, an mehreren Fällen, an einer Klasse von Fällen und mit latenter Beweisidee. Bei einer Prüfung an einem speziellen Fall wird eine infrage stehende Allaussage an genau einem Fall überprüft. Als Prüfung an einem typisch anderen Fall bezeichnen sie, wenn die Verifizierung an einem Fall erfolgt, der möglichst verschieden von allen anderen zuvor betrachteten Fällen ist. Auf diese Weise könne der Geltungsbereich einer Aussage ausgetestet und die Glaubwürdigkeit derselben deutlich erhöht werden. Weiterhin könne eine Prüfung an mehreren Fällen erfolgen. Nach einer Prüfung an mehreren Fällen erscheine ein Satz in der Regel glaubwürdiger als nach der Prüfung an einem einzigen Fall. Außerdem könne vor allem die Prüfung an mehreren, möglichst verschiedenen Fällen zum Verständnis einer infrage stehenden Aussage beitragen. Bei manchen Aussagen biete es sich weiterhin an, die Aussage, anstatt an einem oder mehreren Einzelfällen, direkt an einer Klasse von Fällen zu überprüfen. Dies sei im Vergleich zu den zuvor genannten Optionen anspruchsvoller, weil die Überlegungen bereits auf allgemeinerer Ebene erfolgen und somit sowohl abstrakter sind als auch höhere Anforderungen an die Versprachlichung implizieren. Dadurch könne eine Prüfung an einer Klasse von Fällen aber zugleich den Beweis der infrage stehenden Aussage besonders gut vorbereiten. Letzteres gelte auch für Prüfungen mit latenter Beweisidee. Bei einer Prüfung mit latenter Beweisidee würde das Resultat eines konkreten Falls deduktiv durch die Anwendung mathematischer Definitionen und Sätze, wie etwa algebraischer Regeln der Termumformung, bestimmt, sodass die „Bestimmung des konkreten Resultates schon strukturell die Begründung des Satzes vorwegnimmt" (Meyer & Voigt, 2009, S. 51). Folglich können an konkreten Beispielen verallgemeinerbare Überlegungen erkennbar sein, wie es auch zahlreiche weitere Autorinnen und Autoren betonen (z. B. Biehler & Kempen, 2016). Dabei ist es möglich, jedoch keine Selbstverständlichkeit, dass Lernende die von den Beispielen prinzipiell abstrahierbaren, verallgemeinerbaren Argumentgefüge wahrnehmen und anerkennen (Abschn. 3.3.2, 5.2.2).

Zusammenfassend betrachtet unterscheiden sich die Kategorien der genannten Autorinnen und Autoren hinsichtlich der Anzahl und Art der betrachteten Beispiele, der Korrektheit der Ausführungen, der Erkennbarkeit einer Beweisidee sowie der Berücksichtigung der Frage nach der Allgemeingültigkeit. Diese Aspekte werden im **Analysemodell der vorliegenden Arbeit** aufgegriffen und um einen Punkt ergänzt. Vor dem Hintergrund der weiten Definition der Oberkategorie *prüfender Ansatz* kann es bei Beweisversuchen mit prüfendem Ansatz nicht nur sein, dass die zu zeigende Aussage als gültig dargestellt ist, sondern auch, dass Lernende die Gültigkeit der Aussage negieren oder einschränken oder dass sie keinen Wahrheitswert explizieren. Deshalb wird die Existenz und Art eines angegebenen Wahrheitswerts ebenfalls im Analysemodell berücksichtigt.

Welche Kategorien genau für eine Beschreibung von Eigenschaften von Beweisversuchen mit prüfendem Ansatz hinsichtlich der genannten Aspekte geeignet sind, ist von der zu zeigenden Aussage und der Forschungsfrage abhängig. Bezüglich der Systematik der Beispiele beispielsweise ist es bei der hier fokussierten Aussage über die Teilbarkeit der Summe von vier aufeinanderfolgenden ungeraden Zahlen erfassbar und ggf. mit Blick auf die Weiterentwicklung eines Beweisversuchs mit prüfendem Ansatz zu einem Beweis relevant, wenn aufeinanderfolgende Beispiele angegeben sind. Demgegenüber erscheint eine Untersuchung, inwieweit im Sinne von Bell (1976) bzw. Chazan (1993) alle Beispiele einer gewissen Teilmenge oder einzelne Beispiele verschiedener Teilmengen in Beweisversuchen vorkommen, bei der hier fokussierten Aussage kaum möglich und auch nicht lohnend.

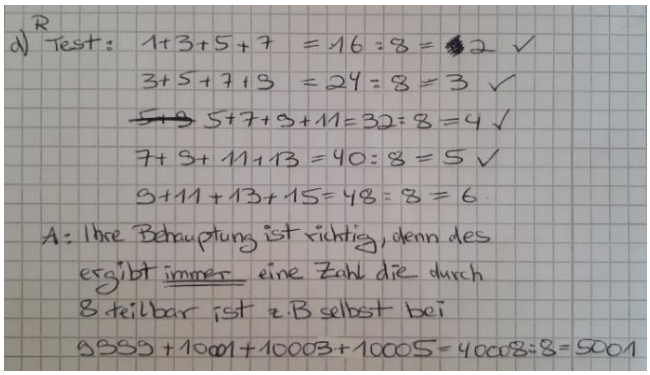
Zusätzlich zur Beschreibung diverser Eigenschaften von Beweisversuchen mit prüfendem Ansatz soll mithilfe des Kategoriensystems grundlegend herausgearbeitet werden können, inwieweit es den Teilnehmenden der Studie gelingt, die vorgegebene Allaussage zu beweisen. Da die Oberkategorie *prüfender Ansatz* ausschließlich über die Existenz eines Beispiels, das einen Bezug zu der infrage stehenden Aussage aufweist, definiert ist, wird – ähnlich wie bei der Analyse der Verwendung von Beispielen bei Beweisversuchen mit begründendem Ansatz (Abschn. 3.3.5)[61] – in einem ersten Schritt erfasst, inwieweit die Schülerinnen und Schüler ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage angegeben und somit die zu zeigende Aussage selbstständig für einen Spezialfall bestätigt haben, bevor weitere Eigenschaften der Beweisversuche mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung untersucht werden.

3.3.7.2 Vorstellung der Kategorien

Bei Beweisversuchen mit prüfendem Ansatz werden entsprechend obiger Abbildung 50 die Existenz und Passung eines Beispiels zu der zu zeigenden Aussage sowie weitere Eigenschaften der angegebenen Beispiele und der Angabe eines Wahrheitswerts untersucht. Die dabei vorkommenden grundlegenden Termini, wie „Beispiel zu der zu zeigenden Aussage“ und „sonstiges Beispiel“, wurden für die Analyse der Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz eingeführt (Abschn. 3.3.5) und werden bei der Analyse der Beweisversuche mit prüfendem Ansatz genauso genutzt. In Abbildung 51 sind exemplarisch mehrere Bearbeitungen mit prüfendem Ansatz und deren Codierungen zu sehen, die die nachfolgend beschriebene Analyse der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage sowie weiterer Eigenschaften der angegebenen Beispiele und der Angabe eines Wahrheitswerts veranschaulichen.

[61] Im Sinne eines kohärenten Kategoriensystems wird die Analyse der Beweisversuche mit prüfendem Ansatz parallel zur in Abschnitt 3.3.5 beschriebenen Analyse der Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz entwickelt.

Bearbeitung 3.3.7a: Klassenstufe 7, weiblich (Code: 0711002)



Angabe eines Wahrheitswerts
- Wwert1 *Wahrheitswert angegeben, Affirmation der Aussage*
- UnzBsp0 *kein Hinweis auf die Unzulänglichkeit von Beispielen*
- AllgzzA1 *Hinweis auf die Allgemeingültigkeit der Aussage*
- Gegenbsp0 *keine Verwendung des Prinzips des Gegenbeispiels*

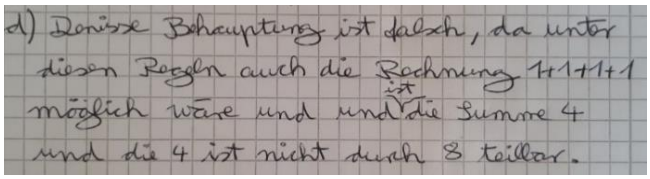
Prüfender Ansatz

Existenz und Passung eines Beispiels zu der zz. Aussage: PBsp4 *unvollständig oder fehlerhaft ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage*

Anzahl und Art angegebener Beispiele
- BspN3 *mehr als fünf Beispiele*
- BspNzzA3 *mehr als fünf* Beispiele zu der zu zeigenden Aussage
- BspNsonst0 *kein sonstiges Beispiel*
- BspSyst3 *mindestens drei Reihenfolgebeispiele*
- Bspbes2 *kleinstes Beispiel*
- Bspgr1 *ein großes Beispiel*
- Bspkv4 *mehrere Beispiele vollständig, weitgehend korrekt*

Korrektheit und Beweisideen
- PzzA3 *Passung zu der zz. Aussage*
- KV0 *keine Beweisidee*
- BIp88 *nicht relevant*

Bearbeitung 3.3.7b: Klassenstufe 7, männlich (Code: 0711910)



Angabe eines Wahrheitswerts
- Wwert1 *Wahrheitswert angegeben, Affirmation der Aussage*
- UnzBsp0 *kein Hinweis auf die Unzulänglichkeit von Beispielen*
- AllgzzA0 *kein Hinweis auf die Allgemeingültigkeit der Aussage*
- Gegenbsp1 *Verwendung des Prinzips des Gegenbeispiels*

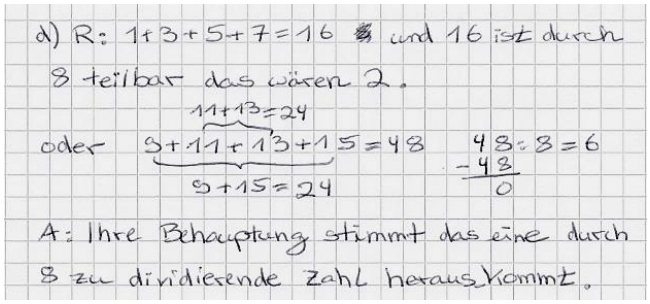
Prüfender Ansatz

Existenz und Passung eines Beispiels zu der zz. Aussage: PBsp2 *sonstiges Beispiel, Addition von vier ungeraden Zahlen*

Anzahl und Art angegebener Beispiele
- BspN1 *genau ein Beispiel*
- BspNzzA0 *kein* Beispiel zu der zu zeigenden Aussage
- BspNsonst1 *genau ein sonstiges Beispiel*
- BspSyst0 *keine Systematik erkennbar*
- Bspbes1 *sonstiges besonderes Beispiel*
- Bspgr0 *kein großes Beispiel*
- Bspkv1 *kein Beispiel zu der zz. Aussage*

Korrektheit und Beweisideen
- PzzA1 *Hinweis auf einen grundlegenden Fehler*
- KV0 *keine Beweisidee*
- BIp88 *nicht relevant*

Bearbeitung 3.3.7c: Klassenstufe 8, weiblich (Code: 0810212)



Angabe eines Wahrheitswerts
- Wwert1 *Wahrheitswert angegeben, Affirmation der Aussage*
- UnzBsp0 *kein Hinweis auf die Unzulänglichkeit von Beispielen*
- AllgzzA0 *kein Hinweis auf die Allgemeingültigkeit der Aussage*
- Gegenbsp0 *keine Verwendung des Prinzips des Gegenbeispiels*

Prüfender Ansatz

Existenz und Passung eines Beispiels zu der zz. Aussage: PBsp5 *vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage*

Anzahl und Art angegebener Beispiele
- BspN2 *zwei bis fünf Beispiele*
- BspNzzA2 *zwei bis fünf* Beispiele zu der zu zeigenden Aussage
- BspNsonst0 *kein sonstiges Beispiel*
- BspSyst1 *andere syst. Beispiele*
- Bspbes2 *kleinstes Beispiel*
- Bspgr0 *kein großes Beispiel*
- Bspkv4 *mehrere Beispiele vollständig, weitgehend korrekt*

Korrektheit und Beweisideen
- PzzA3 *Passung zu der zz. Aussage*
- KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge*
- BIp6 *Teilbarkeit der Summen zweier Zahlenpaare*

Abbildung 51: Bearbeitungen 3.3.7a-c (Beweisversuche mit prüfendem Ansatz: Analyse der Existenz, Passung, Anzahl und Art von Beispielen, der Angabe eines Wahrheitswerts, der Korrektheit und der Beweisideen)

Existenz und Passung eines Beispiels zu der zu zeigenden Aussage

Aufbauend auf den eingeführten Begrifflichkeiten wird bei jedem Beweisversuch mit prüfendem Ansatz zuerst untersucht, inwieweit die Schülerinnen und Schüler ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage angegeben und somit die zu zeigende Aussage selbstständig für einen Spezialfall bestätigt haben. Dazu wird jeder Beweisversuch mit prüfendem Ansatz in Analogie zu der Untersuchung der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage bei Beweisversuchen mit begründendem Ansatz genau einer von sechs Kategorien zugeordnet (Abschn. 3.3.5). Im Vergleich zu der Analyse von Beweisversuchen mit begründendem Ansatz kommt die Kategorie PBsp0 *kein Beispiel* bei Beweisversuchen mit prüfendem Ansatz nicht vor, weil in allen Beweisversuchen mit prüfendem Ansatz per Definition ein Beispiel angegeben ist. In Abbildung 51 sind exemplarisch mehrere Bearbeitungen mit prüfendem Ansatz und deren Codierungen zu sehen.

Anzahl und Art angegebener Beispiele

Nach der Analyse der Passung eines Beispiels zu der zu zeigenden Aussage wird, analog zu der Analyse der Beispiele bei Beweisversuche mit begründendem Ansatz, die Anzahl an angegebenen Beispielen codiert (Abschn. 3.3.5; Bearbeitungen 3.3.7a-c in Abb. 51). Danach wird, ebenfalls analog zur Analyse von Beweisversuchen mit begründendem Ansatz, die Art der angegebenen Beispiele fokussiert. Es wird nacheinander analysiert, inwieweit die angegebenen Beispiele eine Systematik aufweisen, ob ein besonderes Beispiel vorkommt und ob ein großes Beispiel vorkommt (Abschn. 3.3.5; Bearbeitungen 3.3.7a-c in Abb. 51). In Ergänzung dazu wird beschrieben, inwieweit in den Beweisversuchen mit prüfendem Ansatz vollständig und weitgehend korrekt ausgeführte Beispiele zu der zu zeigenden Aussage vorkommen.[62] Dazu wird jeder Beweisversuch mit prüfendem Ansatz genau einer der in Tabelle 2 angegebenen Kategorien zugewiesen. Beispielsweise fällt die Bearbeitung 3.3.7a in die Kategorie Bspkv3 *mehrere Beispiele vollständig, weitgehend korrekt*, die Bearbeitung 3.3.7b hingegen in die Kategorie Bspkv1 *kein Beispiel vollständig, weitgehend korrekt* (Abb. 51).

Bspkv3 m*ehrere Beispiele vollständig, weitgehend korrekt*: In dem Beweisversuch kommen mindestens zwei vollständig und weitgehend korrekt ausgeführte Beispiele zu der zz. Aussage vor. Hinweis: Ein Beispiel der Art „$3 + 5 + 7 + 9 = 24$. 24 ist durch acht teilbar.“ wird an dieser Stelle nicht als ausreichend angesehen, es muss der Quotient angegeben sein, weil ansonsten Zusammenhänge zwischen dem Quotienten und den Summanden weniger ersichtlich sind.
Bspkv2 *genau ein Beispiel vollständig, weitgehend korrekt*: In dem Beweisversuch kommt genau ein vollständig und weitgehend korrekt ausgeführtes Beispiel zu der zz. Aussage vor. Hinweis: Eine Aussage der Art „24 ist durch acht teilbar“ wird an dieser Stelle nicht als ausreichend angesehen, es muss der Quotient angegeben sein.
Bspkv1 *kein Beispiel vollständig, weitgehend korrekt*: In dem Beweisversuch kommt kein vollständig und weitgehend korrekt ausgeführtes Beispiel zu der zz. Aussage vor. Der Beweisversuch kann weder der Kategorie Bspkv3 noch der Kategorie Bspkv2 zugeordnet werden.

Tabelle 2: Kategorien zur Beschreibung, inwieweit in Beweisversuchen mit prüfendem Ansatz vollständig und weitgehend korrekt ausgeführte Beispiele zu der zu zeigenden Aussage vorkommen

[62] Vollständig und weitgehend korrekt ausgeführte Beispiele zu der zu zeigenden Aussage können z. B. das Finden der Beweisidee 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung* und damit die Entwicklung eines korrekten und vollständigen Argumentgefüges unterstützen.

Angabe eines Wahrheitswerts

Während in Beweisversuchen mit begründendem Ansatz i. d. R. die Allgemeingültigkeit der zu zeigenden Aussage angegeben ist, können Beweisversuche mit prüfendem Ansatz z. B. eine Negation der Aussage enthalten (Abschn. 3.3.2). Zur Beschreibung der **Existenz und Art eines angegebenen Wahrheitswerts** bei Beweisversuchen mit prüfendem Ansatz wird jeder der Beweisversuche genau einer der Kategorien Wwert1 *Wahrheitswert angegeben, Affirmation der Aussage*, Wwert0 *Wahrheitswert angegeben, Negation der Aussage*, Wwert55 *Wahrheitswert angegeben, nicht eindeutig kategorisierbarer Wahrheitswert* und Wwert99 *kein Wahrheitswert angegeben* zugeordnet (Tab. 3; Abb. 51).

Wwert1 ***Wahrheitswert angegeben, Affirmation der Aussage*****:** Die zu zeigende Aussage (oder eine nicht ganz korrekte Interpretation dieser Aussage) ist explizit als gültig bezeichnet. Neben Formulierungen mit Begriffen wie „gültig“ und „wahr“ werden auch Aussagen wie „Die Behauptung ist korrekt“ und „Es stimmt!“ sowie Reformulierungen der infrage stehenden Aussage als Angabe ihrer Gültigkeit angesehen. Hinweis: Da die Gültigkeit der zu zeigenden Aussage in der Aufgabenstellung vorgegeben ist, wird im Grenzfall auch als Affirmation der Aussage angesehen, wenn ein Beweisversuch mit „Weil […]“ oder z. B. „Denn […]“ beginnt und passend zur Gültigkeit der Aussage eine Beweisidee an Beispielen erläutert ist.
Wwert0 ***Wahrheitswert angegeben, Negation der Aussage*****:** Die zu zeigende Aussage (oder eine nicht ganz korrekte Interpretation dieser Aussage) ist explizit als ungültig bezeichnet. Insbesondere werden Formulierungen mit Begriffen wie „ungültig“ und „falsch“ sowie Aussagen wie „Ihre Behauptung stimmt nicht“ als Negation der Aussage angesehen.
Wwert55 ***Wahrheitswert angegeben, nicht eindeutig kategorisierbar*****:** Es ist ein Wahrheitswert angegeben, der weder als Affirmation der Aussage noch als Negation der Aussage bezeichnet werden kann. Insbesondere sind Bearbeitungen, in denen eine in der Aufgabenstellung nicht vorgegebene Bedingung für die Gültigkeit der Aussage angegeben ist, in diese Kategorie einzuordnen. Hinweis: Ein Beweisversuch gehört auch in diese Kategorie, wenn explizit ein Wahrheitswert angegeben ist, der sich jedoch nicht auf die Aussage an sich, sondern auf angegebene Beispiele bezieht, so dass unklar ist, inwieweit die Aussage (und nicht nur die angegebenen Spezialfälle) als gültig angesehen werden (z. B. „Bei diesen Beispielen stimmt die Aussage von Denise“).
Wwert99 ***kein Wahrheitswert angegeben*****:** In der Bearbeitung ist kein Wahrheitswert angegeben. Hinweis: Insbesondere wird die alleinige Angabe eines Häkchens oder eines Ausdrucks wie „geht“ am Ende einer Addition nicht als Angabe eines Wahrheitswerts angesehen, weil diese evtl. ausschließlich die Division durch vier oder acht ersetzen.

Tabelle 3: Kategorien zur Beschreibung, ob einer und ggf. welcher Wahrheitswert in Beweisversuchen angegeben ist

Um bei Beweisversuchen mit prüfendem Ansatz, in denen kein Wahrheitswert angegeben ist (Wwert99), eine Tendenz auszumachen, ob die Aussage eher als gültig oder als ungültig angesehen wird, werden bei diesen Beweisversuchen die Anzahl und die Reihenfolge von Positiv- und Negativbeispielen analysiert und die drei Kategorien PosNeg1 *ausschließlich Positivbeispiele*, PosNeg0 *genau ein Negativbeispiel, das zudem das letzte Beispiel ist* und PosNeg55 *nicht eindeutig kategorisierbar* differenziert. Ein Beispiel wird als Positivbeispiel bezeichnet, wenn es die Gültigkeit der zu zeigenden Aussage zumindest scheinbar bestätigt. Mit Blick auf die im Rahmen dieser Arbeit fokussierte Aussage trifft dies insbesondere zu, wenn eine angegebene Summe von vier Zahlen durch vier (Kl. 5, 6) bzw. acht (Kl. 7, 8) teilbar ist (z. B. „$1 + 7 + 9 + 15 = 32$“). Ein Beispiel wird als Negativbeispiel bezeichnet, wenn es die Gültigkeit der zu zeigenden Aussage scheinbar widerlegt. Dies trifft insbesondere zu, wenn eine Summe von vier Zahlen nicht durch vier (Kl. 5, 6) bzw. acht (Kl. 7, 8) teilbar ist (z. B. „$1 + 5 + 7 + 9 = 22$“). Falls ein Beispiel nicht eindeutig entweder ein Positivbeispiel oder ein Negativbeispiel ist, wird es als nicht eindeutig kategorisierbar angesehen (z. B. sonstige Beispiele wie „$9 - 7 = 2$“).

In Ergänzung zu der Analyse der Existenz und Art eines angegebenen Wahrheitswerts wird untersucht, ob in den Beweisversuchen mit prüfendem Ansatz ein **Hinweis auf die Unzulänglichkeit von Beispielen, die Allgemeingültigkeit der Aussage oder die Verwendung des Prinzips des Gegenbeispiels** vorkommt. Unter dem Unterpunkt „Hinweis auf die Unzulänglichkeit von Beispielen“ wird codiert, ob explizit angedeutet ist, dass Beispiele als Beweis der Aussage nicht ausreichen. Unter dem Unterpunkt „Hinweis auf die Allgemeingültigkeit der Aussage“ wird codiert, ob die Allgemeingültigkeit der zu zeigenden Aussage (oder einer nicht ganz korrekten Interpretation derselben) hervorgehoben ist. Die Allgemeingültigkeit kann beispielsweise mit einem Ausdruck, wie „immer“ oder „egal welche Zahlen man wählt“, oder mit einem Hinweis auf die Fortsetzbarkeit von Beispielen, wie etwa „usw.“ unterhalb der Beispiele, hervorgehoben sein. Weiterhin kann im Sonderfall einer Negation der Aussage die Verwendung des Prinzips des Gegenbeispiels andeuten, dass die Frage nach der Allgemeingültigkeit zumindest unterbewusst bedacht wurde. Von der Verwendung des Prinzips des Gegenbeispiels wird ausgegangen, wenn in einem Beweisversuch genau ein Negativbeispiel vorkommt, welches zudem als Grund für die Ungültigkeit der Aussage dargestellt ist. Dies wird unter dem Unterpunkt „Verwendung des Prinzips des Gegenbeispiels“ festgehalten. Die gesamte Analyse der Angabe eines Wahrheitswerts ist in Abbildung 51 exemplifiziert.

Korrektheit und Beweisideen

In einem weiteren Schritt werden die Korrektheit von Beweisversuchen mit prüfendem Ansatz und eventuell vorkommende Beweisideen analysiert. Unter der Korrektheit eines Beweisversuchs werden, entsprechend den Ausführungen in Abschnitt 3.3.4, der Grad dessen Passung zu der zu zeigenden Aussage und die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges verstanden. Anknüpfend an die Analyse von Beweisversuchen mit begründendem Ansatz werden bei Beweisversuchen mit prüfendem Ansatz zur Beschreibung der **Passung der Ausführungen zu der zu zeigenden Aussage** die Kategorien PzzA1 *Hinweis auf einen grundlegenden Fehler*, PzzA2 *Bezug auf die zu zeigende Aussage unklar* und PzzA3 *Passung zu der zu zeigenden Aussage* verwendet. Beweisversuche mit prüfendem Ansatz gehören in die Kategorie PzzA1 *Hinweis auf einen grundlegenden* Fehler, wenn sie einen Hinweis darauf enthalten, dass die zu zeigende Aussage nicht ganz korrekt interpretiert wurde. Dies ist insbesondere bei allen Beweisversuchen mit prüfendem Ansatz der Fall, in denen kein Beispiel zu der zu zeigenden Aussage oder eine Negation der Aussage vorkommt. Beispielsweise gehört die Bearbeitung 3.3.7b in Abbildung 51 in die Kategorie PzzA1 *Hinweis auf einen grundlegenden Fehler*. Die Bearbeitungen 3.3.7a,c hingegen können beide der Kategorie PzzA3 *Passung zu der zu zeigenden Aussage* zugeordnet werden. Ein Beweisversuch mit prüfendem Ansatz wird der Kategorie PzzA3 *Passung zu der zu zeigenden Aussage* zugeordnet, wenn sich die Ausführungen aus Expertensicht sehr wahrscheinlich genau auf die zu zeigende Aussage beziehen. Dies ist insbesondere der Fall, wenn alle angegebenen Beispiele weitgehend korrekt ausgeführte Beispiele zu der zu zeigenden Aussage sind und zudem ein eigenes Beispiel sowie ein vollständig ausgeführtes Beispiel vorkommen. Dabei kann der Beweisversuch eine Affirmation

der Aussage oder auch keinen Wahrheitswert[63] enthalten. Für die Zuordnung zur Kategorie PzzA3 *Passung zu der zu zeigende Aussage* muss nicht unbedingt ein Hinweis auf die Allgemeingültigkeit dieser Aussage erkennbar sein. Bei allen Beweisversuchen mit prüfendem Ansatz, die weder in die Kategorie PzzA1 *Hinweis auf einen grundlegenden Fehler* noch in die Kategorie PzzA3 *Passung zu der zu zeigenden Aussage* gehören, wird der Bezug auf die zu zeigende Aussage als unklar angesehen (PzzA2).

Zur Beschreibung der **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** bei Beweisversuchen mit prüfendem Ansatz können ebenfalls die bei Beweisversuchen mit begründendem Ansatz verwendete Kategorien aufgegriffen werden. Bei Beweisversuchen mit prüfendem Ansatz ist es allerdings möglich, dass kein Argumentgefüge und insbesondere auch keine Beweisidee erkennbar ist. Deshalb wird im Vergleich zur Analyse von Beweisversuchen mit begründendem Ansatz die Kategorie KV0 *keine Beweisidee* ergänzt.[64] Außerdem werden im Vergleich zur Analyse von Beweisversuchen mit begründendem Ansatz die Kategorien KV4, KV5, KV6 zu der Kategorie KV4-6 *weitgehend korrektes Argumentgefüge* zusammengefasst.[65] Ansonsten lassen sich die in Abschnitt 3.3.4 beschriebenen Kategorien aus der Analyse von Beweisversuchen mit begründendem Ansatz im Sinne eines kohärenten Kategoriensystems übernehmen. Falls ein Beweisversuch mit prüfendem Ansatz in eine der Kategorien KV1-6 gehört, d. h. zumindest eine Beweisidee erkennbar ist, wird ergänzend die **primäre Beweisidee** so wie bei Beweisversuchen mit begründendem Ansatz erfasst (Abschn. 3.3.3 zu den Kategorien). Beispielsweise gehört die Bearbeitung 3.3.7c in Abbildung 51 in die Kategorie KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge*. Es ist ansatzweise die ausbaubare Beweisidee 6 *Teilbarkeit der Summe zweier Zahlenpaare* zu erkennen, die zugleich die primäre Beweisidee darstellt.[66] Die Bearbeitungen 3.3.7a-b hingegen gehören beide in die Kategorie KV0 *keine Beweisidee*, die primären Beweisidee wird also nicht untersucht.

[63] Es werden auch Beweisversuche mit prüfendem Ansatz, in denen kein Wahrheitswert angegeben ist, in die Kategorie PzzA3 *Passung zu der zu zeigenden Aussage* eingeordnet, weil die Allgemeingültigkeit der zu zeigenden Aussage in der Aufgabenstellung vorgegeben ist.

[64] Ein Beweisversuch gehört auch in diese Kategorie, wenn Reihenfolgebeispiele vorkommen, an denen nichts hervorgehoben ist. Gleiches gilt für Beweisversuche mit einem besonderen oder einem großen Beispiel.

[65] Dadurch, dass die Argumentgefüge in Beweisversuchen mit prüfendem Ansatz vollständig an einem oder mehreren Beispielen ausgeführt sind, kann eine Differenzierung etwa zwischen weitgehend korrekten und vollständigen Argumentgefügen mit einer Ungenauigkeit in der Formulierung und korrekten und vollständigen Argumentgefügen schwierig sein. Auf eine präzise Beschreibung der Vollständigkeit der Argumentgefüge wird daher bei Beweisversuchen mit prüfendem Ansatz verzichtet. Bei Interesse kann die Vollständigkeit in Einzelfallanalysen im Rahmen der weiteren Auswertung der Codierungen näher betrachtet werden.

[66] Bei der Betrachtung der aufeinanderfolgenden ungeraden Zahlen $9, 11, 13, 15$ werden die zweitkleinste und die zweitgrößte sowie die kleinste und die größte dieser Zahlen separat addiert. Beide Summen betragen 24. Sie sind gleich groß, genau doppelt so groß wie die gerade Zahl zwischen dem zweitkleinsten und zweitgrößten Summanden und durch vier teilbar. Diese Auffälligkeiten könnten mit dem Abstand der einzelnen Summanden zu der geraden Zahl zwischen dem zweitkleinsten und zweitgrößten Summanden begründet werden. Da das Doppelte einer durch vier teilbaren Zahl durch acht teilbar ist, folgt aus der Gleichheit und der Teilbarkeit durch vier der betrachteten zwei Summen, dass die Summe von vier aufeinanderfolgenden ungeraden durch acht teilbar sein muss. Die erkennbaren Auffälligkeiten könnten folglich zu einer Begründung der infrage stehenden Aussage ausgebaut werden.

3.3.7.3 Hinweise zu den Kategorien

In der detaillierten Analyse der Beweisversuche mit prüfendem Ansatz werden zahlreiche Eigenschaften der Beispiele und der Angabe eines Wahrheitswerts kategorisiert. Damit wird versucht, der Frage nach dem Nachweis der Gültigkeit der zu zeigenden Aussage, den Charakteristika der Beweisversuche und möglichen Anknüpfungspunkten für deren Weiterentwicklung zu einem Beweis nachzugehen.

Die anfängliche Analyse der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage fokussiert auf die Frage nach dem Nachweis der Gültigkeit der zu zeigenden Aussage. Sie verdeutlicht, inwieweit die zu zeigende Aussage in den Beweisversuchen für einen in der Aufgabenstellung nicht vorgegebenen Spezialfall bestätigt wird (Abschn. 8.3.1 für die Ergebnisse). Die Analyse der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage lässt allerdings keinen direkten Rückschluss darauf zu, inwieweit die zu zeigende Aussage korrekt interpretiert wurde, da ausschließlich eines von eventuell mehreren Beispielen betrachtet wird. Dazu sind weitere Analysen notwendig, in die alle angegebenen Beispiele einbezogen werden.

Alle weiteren Analysen dienen vor allem dazu, in deskriptiver Form darzulegen, welche Eigenschaften in den Beweisversuchen mit prüfendem Ansatz besonders häufig oder eher selten vorkommen. Die beschriebenen Eigenschaften lassen sich zudem hinsichtlich möglicher Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche mit prüfendem Ansatz zu einem Beweis interpretieren. Die Analysen der Passung der Ausführungen zu der zu zeigenden Aussage und der Angabe eines Wahrheitswerts liefern Anhaltspunkte, inwieweit die zu zeigende Aussage korrekt interpretiert wurde (Abschn. 8.3.2.1 für die Ergebnisse). Die Analysen der Erkennbarkeit und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Anzahl und Art angegebener Beispiele geben einen Einblick, inwieweit in den Beweisversuchen sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges zu finden sind (Abschn. 8.3.2.2 für die Ergebnisse). Zusätzlich zu der Interpretation der zu zeigenden Aussage und Anknüpfungspunkten für die Entwicklung eines korrekten und vollständigen Argumentgefüges ist die verwendete Schlussweise mit Blick auf die Weiterentwicklung eines Beweisversuchs mit prüfendem Ansatz zu einem Beweis von Bedeutung. Die Analyse der Schlussweise bei Beweisversuchen mit prüfendem Ansatz wird im nachfolgenden Abschnitt eingeführt.

3.3.8 Prüfender Ansatz: Schlussweise

Bei Beweisversuchen mit prüfendem Ansatz soll die Schlussweise untersucht werden, weil sie mit Blick auf die Weiterentwicklung eines Beweisversuchs mit prüfendem Ansatz zu einem Beweis von Bedeutung ist (Abb. 52; Abschn. 8.3.2.3).

Die Analyse der Schlussweise bei Beweisversuchen mit prüfendem Ansatz wird allerdings auf Bearbeitungen mit einer Affirmation der Aussage fokussiert, d. h. auf die Bearbeitungen mit prüfendem Ansatz, in denen die zu zeigende Aussage (oder eine nicht ganz korrekte Interpretation derselben) explizit als gültig bezeichnet ist (Abschn. 3.3.7). Im Falle einer Negation der Aussage wird im Rahmen der Analyse der Angabe des Wahrheitswerts erfasst, ob die Ausführungen die Verwendung des Prinzips des Gegenbeispiels nahelegen (Abschn. 3.3.7). Wenn ein

nicht eindeutig kategorisierbarer Wahrheitswert oder kein Wahrheitswert angegeben ist, ist eine Analyse der Schlussweise wegen der großen Vielfalt der möglichen Denkweisen nicht praktikabel.

Weiterhin lassen sich anhand der in rein schriftlicher Form vorliegenden Beweisversuche mit prüfendem Ansatz auch im Falle einer Affirmation der Aussage ausschließlich Hinweise auf die gewählte Schlussweise beschreiben und anschließend deuten, wie im Folgenden näher ausgeführt wird. Unter dem Begriff „Schlussweise“ wird deshalb in dieser Arbeit bei Beweisversuchen mit prüfendem Ansatz, anknüpfend an die Analyse der Schlussweise bei Beweisversuchen mit begründendem Ansatz (Abschn. 3.3.6), das Verhältnis angegebener Beispiele zur Angabe der Gültigkeit einer Aussage verstanden und mithilfe verschiedener Kategorien beschrieben. Die folgenden Ausführungen beginnen mit den Überlegungen, die bei der Entwicklung der Kategorien leitend waren. Danach werden die entwickelten Kategorien definiert und abschließend kommentiert. Die Ergebnisse der Analyse der Schlussweise werden in Abschnitt 8.3.2.3 präsentiert.

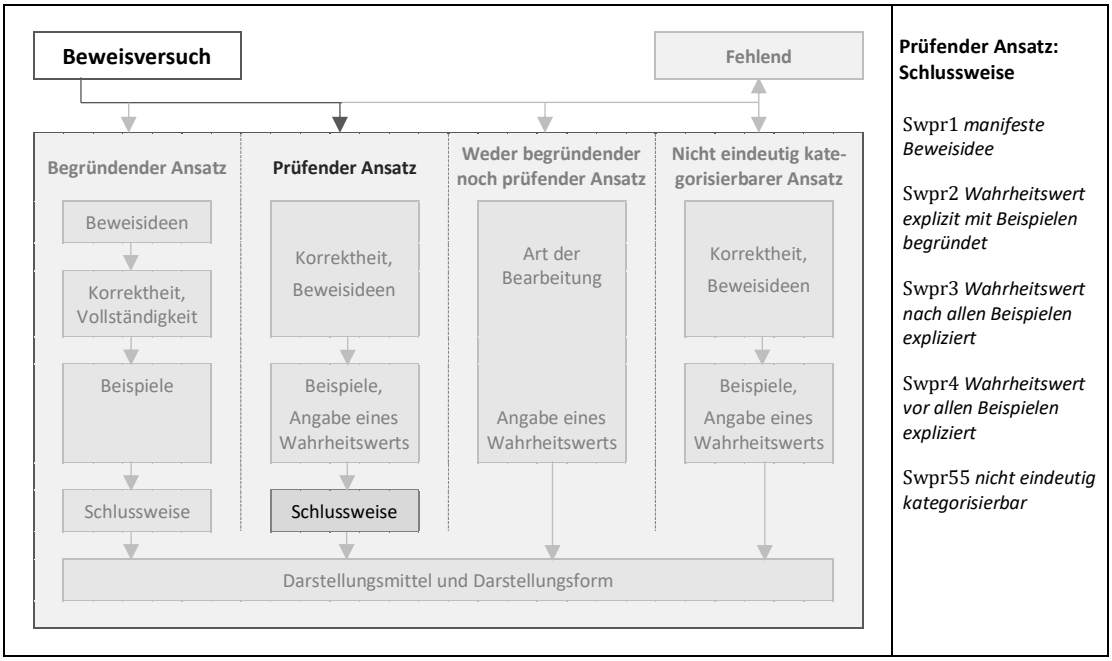


Abbildung 52: Analysemodell. Fokus: Prüfender Ansatz, Schlussweise

3.3.8.1 Entwicklung der Kategorien

Bei der Entwicklung des Kategoriensystems zur Beschreibung der Schlussweise bei Beweisversuchen mit prüfendem Ansatz und einer Affirmation der Aussage ist die Frage leitend, inwieweit die Lernenden einen induktiven Schluss durchgeführt haben. In einer Induktion wird von einem Fall und einem Resultat auf ein (allgemeingültiges) Gesetz geschlossen. Ein Beweisversuch mit prüfendem Ansatz kann auf induktivem Schließen basieren. Es ist jedoch auch möglich, dass Lernende nicht von Beispielen auf die Gültigkeit dieser Aussage schließen, sondern vielmehr die Gültigkeit der Aussage durch Beispiele veranschaulichen wollen (Kempen & Biehler, 2014; Recio & Godino, 2001). Im Vergleich zur Definition eines induktiven Schlusses ist es weiterhin denkbar, dass Lernende ausgehend von Beispielen auf eine gewisse Gültigkeit der

Aussage, jedoch nicht unbedingt bewusst auf deren Allgemeingültigkeit schließen. Die Frage nach der Allgemeingültigkeit kann Lernenden unterschiedlich präsent sein, wie u. a. die Unterscheidung zwischen „naive empiricism“ (Balacheff, 2018, S. 218) und „crucial experiments“ (ebd.) von Balacheff (1988) andeutet. Außerdem kann es sein, dass nicht alleine wegen eines oder mehrerer Beispiele, sondern vielmehr Dank einer an Beispielen erkannten Beweisidee auf die Gültigkeit der Aussage geschlossen wird. Diese Beweisidee kann aus Sicht von Lesenden manifest sein, es ist aber auch möglich, dass sie nicht klar zu erkennen ist. Zum Beispiel weist Krumsdorf (2017) mit der Definition eines „subjektiv realisierten Beweises“ (Krumsdorf, 2017, S. 140) auf die Möglichkeit hin, dass Lernenden die Verallgemeinerbarkeit von am Beispiel durchgeführten Operationen bewusst sein kann, ohne dass sie diese explizieren. Reid und Knipping (2010) beschreiben in Anlehnung an Smith und Henderson (1959) mit der Kategorie „extending a pattern“ (Reid & Knipping, 2010, S. 132), dass ausgehend von Reihenfolgebeispielen eine gewisse Invariante erkannt und somit auf die Allgemeingültigkeit einer Aussage geschlossen werden kann. Coe und Ruthven (1994) berichten von einem Schüler, der (laut einer Interviewaussage) von der Gültigkeit einer Aussage überzeugt ist, weil er anhand seiner Beispiele eine Regelmäßigkeit erkannt hat (Abschn. 5.2).

Da die Beweisversuche im Rahmen dieser Arbeit ausschließlich in schriftlicher Form vorliegen, beschreibt der Terminus „Schlussweise“ nicht unbedingt die tatsächliche Denkweise der Lernenden, sondern, mit Fokus auf Beweisversuche mit prüfendem Ansatz, ausschließlich das Verhältnis angegebener Beispiele zu der Angabe der Gültigkeit einer Aussage. Ausgehend von den Schriftdokumenten lassen sich nur Eigenschaften der Beweisversuche beschreiben, die Anhaltspunkte für die tatsächliche Denkweise der Lernenden bieten. Anknüpfend an die vorhergehenden Überlegungen werden als solche Eigenschaften der Ort der Angabe der Gültigkeit und deren Verknüpfung mit den Beispielen, die Existenz eines Hinweises auf die Frage nach der Allgemeingültigkeit und die Erkennbarkeit einer manifesten oder einer eventuell subjektiv realisierten Beweisidee berücksichtigt und miteinander kombiniert.

3.3.8.2 Vorstellung der Kategorien

Zur Beschreibung der Schlussweise bei Beweisversuchen mit prüfendem Ansatz werden entsprechend Abbildung 53 die Kategorien $\mathrm{Swpr1}$ bis $\mathrm{Swpr4.4}$ und $\mathrm{Swpr55}$ unterschieden. Dabei bezeichnen $\mathrm{Swpr2.1}$ bis $\mathrm{Swpr2.4}$ Unterkategorien der Kategorie $\mathrm{Swpr2}$ *Gültigkeit explizit mit Beispiel verknüpft*, wie im Folgenden näher ausgeführt wird.[67]

Ein Beweisversuch mit prüfendem Ansatz gehört in die Kategorie **Swpr1** ***manifeste Beweisidee***, wenn eine aus Expertensicht erkennbare Beweisidee z. B. mithilfe verbaler Darstellungsmittel so klar ausgedrückt ist, dass von einer subjektiven Realisierung derselben ausgegangen werden kann (Bearbeitung 3.3.8a in Abb. 54). Im Vergleich zu Beweisversuchen mit begründendem Ansatz ist die Verallgemeinerbarkeit der Beweisidee nicht explizit angedeutet.

[67] Um die Kategorien möglichst gut veranschaulichen zu können, werden im Folgenden nicht nur Beweisversuche aus den Klassenstufen 7 und 8, sondern auch solche aus den Klassenstufen 5 und 6 angeführt. Die Beweisversuche von Lernenden in den Klassenstufen 5 und 6 beziehen sich auf die Aussage, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer durch vier teilbar ist (Abschn. 6.2.2).

Ort der Angabe der Gültigkeit und dessen Verknüpfung mit Beispielen	Erkennbarkeit einer manifesten oder einer evtl. subjektiv realisierten Beweisidee, Existenz eines Hinweises auf die Allgemeingültigkeit der Aussage					
	manifeste Beweisidee	*evtl. subj. real. BI, Hinweis auf die Allg.gültigkeit*	*evtl. subjektiv realisierte BI*	*Hinweis auf die Allg.gültigkeit*	*keine BI, kein Hinweis auf Allg.gültigkeit*	*nicht eindeutig kategorisierbar*
Gültigkeit explizit mit Beispielen verknüpft (Swpr2)	Swpr1	Swpr2.1	Swpr2.2	Swpr2.3	Swpr2.4	
Gültigkeit nach allen Beispielen expliziert (Swpr3)		Swpr3.1	Swpr3.2	Swpr3.3	Swpr3.4	
Gültigkeit vor allen Beispielen expliziert (Swpr4)		Swpr4.1	Swpr4.2	Swpr4.3	Swpr4.4	
nicht eindeutig kategorisierbar						Swpr55

subj.: subjektiv real.: realisierte BI: Beweisidee Allg.gültigkeit: Allgemeingültigkeit

Abbildung 53: Kategorien zur Beschreibung der Schlussweise bei Beweisversuchen mit prüfendem Ansatz und einer Affirmation der Aussage

Bearbeitung 3.3.8a: Klassenstufe 5, männlich (Code: 0520416)

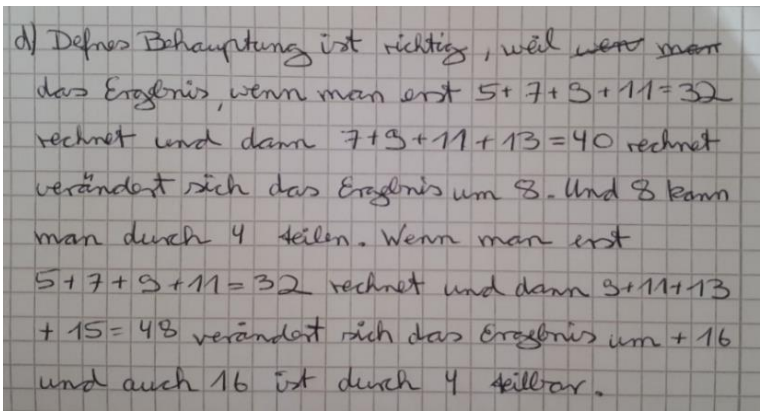
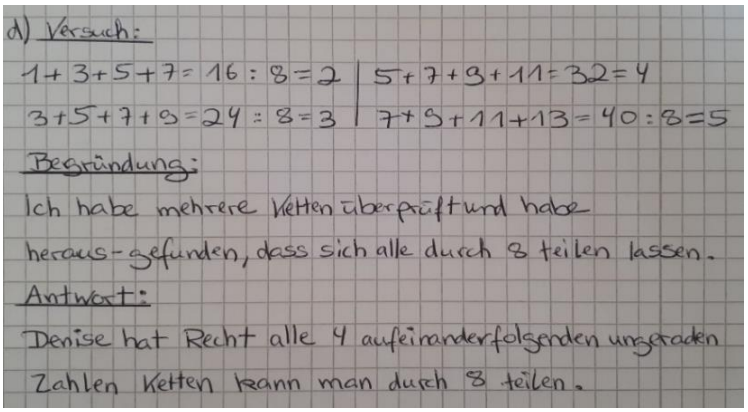


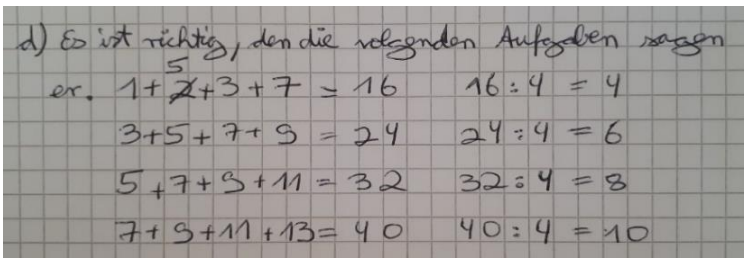
Abbildung 54: Bearbeitung 3.3.8a (Swpr1 *manifeste Beweisidee*)

Beweisversuche in der Kategorie **Swpr2** ***Gültigkeit explizit mit Beispielen verknüpft*** enthalten keine manifeste Beweisidee, sie charakterisieren sich dadurch, dass Beispiele explizit als Grund für die Affirmation der Aussage dargestellt sind. Dabei ist in Beweisversuchen in den Unterkategorien 2.1 und 2.2 ein Hinweis auf eine eventuell subjektiv realisierte Beweisidee zu erkennen. Das heißt, dass aus Expertensicht eine latente Beweisidee oder Beispiele mit einer klar erkennbaren Systematik vorkommen. Beweisversuche in den Unterkategorien 2.1 und 2.3 enthalten einen Hinweis auf die Allgemeingültigkeit der Aussage. In Beweisversuchen in der Unterkategorie 2.4 ist weder eine eventuell subjektiv realisierte Beweisidee noch ein Hinweis auf die Allgemeingültigkeit der Aussage zu erkennen (Bearbeitungen 3.3.8b-e in Abb. 55).

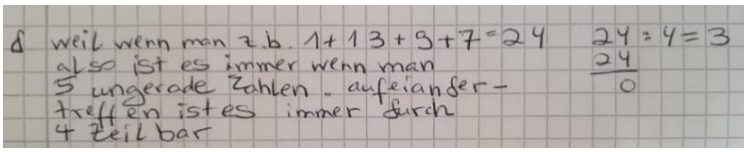
Bearbeitung 3.3.8b: Klassenstufe 7, weiblich (Code: 0710408)



Bearbeitung 3.3.8c: Klassenstufe 5, männlich (Code: 0520106)



Bearbeitung 3.3.8d: Klassenstufe 5, männlich (Code: 0522317)



Bearbeitung 3.3.8e: Klassenstufe 5, männlich (Code: 0510212)

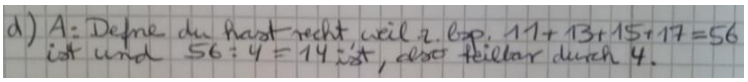
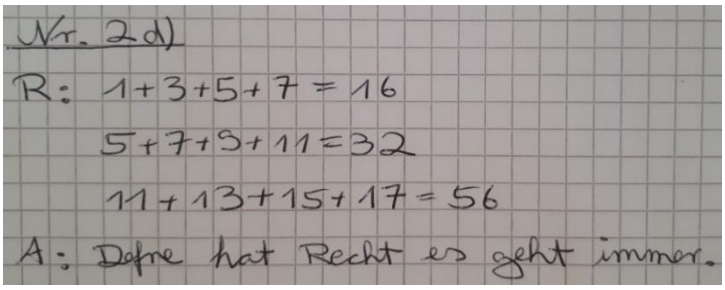


Abbildung 55: Bearbeitung 3.3.8b (Swpr2.1 *Wahrheitswert explizit mit Beispielen begründet, evtl. subjektiv realisierte Beweisidee, Hinweis auf Allgemeingültigkeit*), Bearbeitung 3.3.8c (Swpr2.2 *Wahrheitswert explizit mit Beispielen begründet, evtl. subjektiv realisierte Beweisidee*), Bearbeitung 3.3.8d (Swpr2.3 *Wahrheitswert explizit mit Beispielen begründet, Hinweis auf Allgemeingültigkeit*), Bearbeitung 3.3.8e (Swpr2.4 *Wahrheitswert explizit mit Beispielen begründet*)

Im Vergleich zu den Beweisversuchen in der Kategorie Swpr2 ist der angegebene Wahrheitswert bei Beweisversuchen in den Kategorien Swpr3-4 nicht explizit mit den Beispielen verknüpft. Bei Beweisversuchen in der Kategorie **Swpr3** ***Gültigkeit nach allen Beispielen expliziert*** folgt die Angabe der Gültigkeit vom Textfluss her nach allen Beispielen (Bearbeitung 3.3.8f in Abb. 57).[68] Bei Beweisversuchen in der Kategorie **Swpr4** ***Gültigkeit vor allen Beispielen expliziert*** befindet sich die Angabe der Gültigkeit vom Textfluss her vor allen Beispielen (Bearbeitung 3.3.8g in Abb. 56).

[68] Die Differenzierung zwischen den Unterkategorien 3.1 bis 3.4 (sowie den Unterkategorien 4.1 bis 4.4) erfolgt analog zu der beschriebenen Unterscheidung zwischen den Kategorien 2.1 bis 2.4.

Bearbeitung 3.3.8f: Klassenstufe 6, weiblich (Code: 0621006)



Bearbeitung 3.3.8g: Klassenstufe 5, männlich (Code: 0510835)

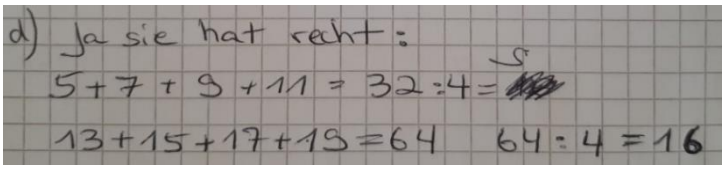


Abbildung 56: Bearbeitung 3.3.8f (Swpr3.3 *Wahrheitswert nach allen Beispielen expliziert, Hinweis auf Allgemeingültigkeit*), Bearbeitung 3.3.38g (Swpr4.4 *Wahrheitswert vor allen Beispielen expliziert*)

Schließlich ist es möglich, dass sich Beweisversuche nicht eindeutig einer der genannten Kategorien zuordnen lassen (**Swpr55** ***nicht eindeutig kategorisierbar***). Beispielsweise sind in der Bearbeitung 3.3.8h in Abbildung 57 die Beispiele nicht explizit als Grund für die Gültigkeit der Aussage angeführt, weshalb die Bearbeitung nicht in die Kategorie Swpr2 gehört. Außerdem gehört die Bearbeitung mangels einer manifesten Beweisidee nicht in die Kategorie Swpr1. In der Bearbeitung 3.3.8h ist die Gültigkeit der Aussage sowohl vor als auch nach den Beispielen angeführt, sodass sie ebenfalls nicht klar in eine der Kategorien Swpr3, Swpr4 gehört.

Bearbeitung 3.3.8h: Klassenstufe 6, männlich (Code: 0611828)

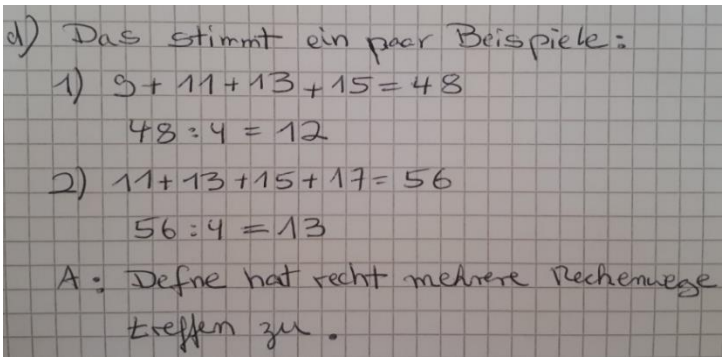


Abbildung 57: Bearbeitung 3.3.8h (Swpr55 *nicht eindeutig kategorisierbar*)

3.3.8.3 Hinweise zu den Kategorien

Mit Blick auf die Frage nach einem induktiven Schluss sind besonders Beweisversuche in der Kategorie Swpr2.3 *Wahrheitswert explizit mit Beispielen begründet, Hinweis auf Allgemeingültigkeit* hervorzuheben, da die Beispiele explizit als Grund für die Allgemeingültigkeit der Aussage angeführt sind, ohne dass ein Anhaltspunkt für eine eventuell subjektiv realisierte Beweisidee vorkommt. Bei Beweisversuchen in dieser Kategorie deuten die Ausführungen am deutlichsten auf einen induktiven Schluss hin.

Beweisversuche in der Kategorie $\mathrm{Swpr1}$ *manifeste Beweisidee* bewegen sich, Dank dem an Beispielen ausgeführten Argumentgefüge, an der Grenze der Beweisversuche mit prüfendem Ansatz zu solchen mit begründendem Ansatz. Sie sind beispielgebundenen Beweisen mit subjektiv realisierter Sinnstruktur im Sinne von Krumsdorf (2017) und beispielgebundenen Beweisen im Sinne von Meyer und Voigt (2008) prinzipiell am nächsten, wobei in der vorliegenden Arbeit die an Beispielen durchgeführten Überlegungen nicht unbedingt ein korrektes und vollständiges Argumentgefüge bilden. Hinsichtlich der Qualität der Überlegungen gibt die Analyse der Erkennbarkeit und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges Aufschluss (Abschn. 3.3.7).

Bei Beweisversuchen in den Kategorien $\mathrm{Swpr2.1,2.2,3.1,3.2,4.1,4.2}$ gibt es in der Bearbeitung einen Hinweis darauf, dass eventuell eine erkannte Beweisidee zur Angabe der Gültigkeit der Aussage beigetragen hat. Die Bezeichnung „eventuell subjektiv realisierte Beweisidee“ knüpft an die Begrifflichkeiten von Krumsdorf (2017) an, wobei die Bearbeitungen offenlassen, ob die Lernenden tatsächlich eine Beweisidee erkannt haben.

Bei Beweisversuchen in den Kategorien $\mathrm{Swpr4.1}$ bis $\mathrm{Swpr4.4}$ befindet sich der Wahrheitswert vom Textfluss her vor allen Beispielen, sodass die Beispiele die Gültigkeit der Aussage aus Expertensicht illustrieren. Inwieweit die Beispiele aus Sicht der Lernenden z. B. dazu angeführt wurden, um die Gültigkeit der Aussage zu veranschaulichen oder um die Aussage zu verstehen, lässt sich alleine anhand der schriftlichen Ausführungen nicht sicher sagen.

3.3.9 Weder begründender noch prüfender Ansatz: Angabe eines Wahrheitswerts und Art der Bearbeitung

Dieser Abschnitt fokussiert auf Beweisversuche mit weder begründendem noch prüfendem Ansatz. Bei diesen Beweisversuchen deutet es sich an, dass sich die Autorin bzw. der Autor mit der Aufgabenstellung beschäftigt hat, aber dennoch enthalten sie weder einen begründenden noch einen prüfenden Ansatz (Abschn. 3.3.2). Mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage ist folglich weder eine Beweisidee noch ein bestätigendes Beispiel formuliert. Die Analyse der Beweisversuche mit weder begründendem noch prüfendem Ansatz soll verdeutlichen, inwieweit in diesen Bearbeitungen die Allgemeingültigkeit der zu zeigenden Aussage expliziert ist. Zudem soll sie ergänzende Einblicke in die Eigenschaften der Beweisversuche mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis liefern. Der dafür relevante Teil des Analysemodells (Abb. 58) wird im vorliegenden Abschnitt eingeführt und kommentiert. Die Ergebnisse der Analyse der Beweisversuche mit weder begründendem noch prüfendem Ansatz werden in Abschnitt 8.4 präsentiert.

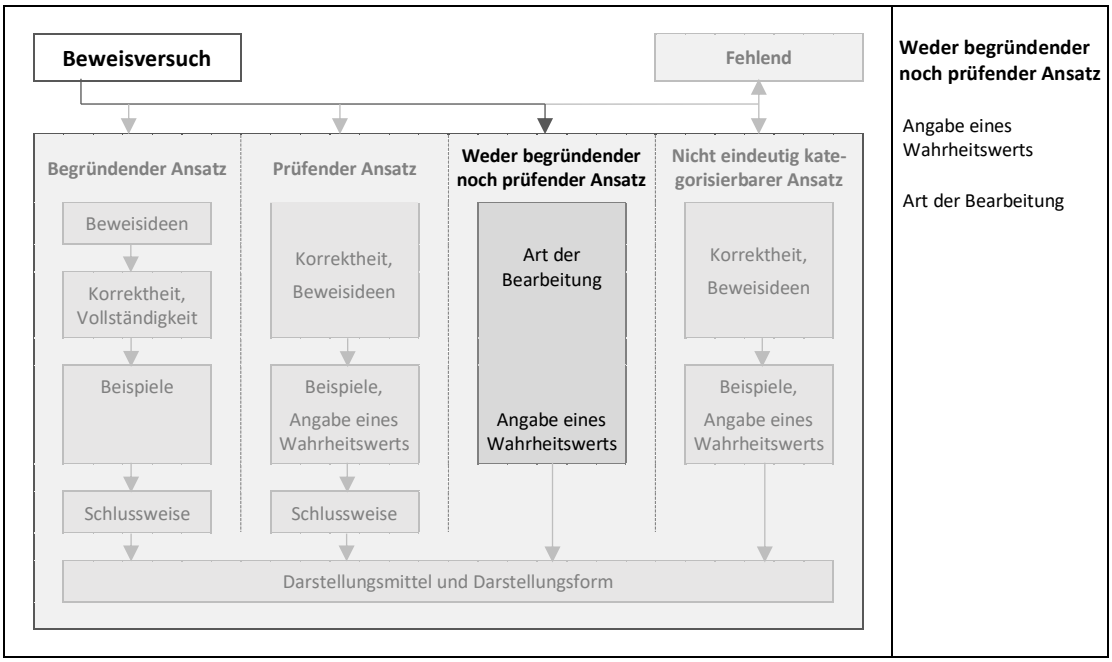


Abbildung 58: Analysemodell. Fokus: Weder begründender noch prüfender Ansatz, Angabe eines Wahrheitswerts und Art der Bearbeitung

3.3.9.1 Entwicklung der Kategorien

Entsprechend der doppelten Fragestellung nach der Angabe der Allgemeingültigkeit der zu zeigenden Aussage und weiteren Eigenschaften der Beweisversuche wird bei der Entwicklung der Kategorien in zwei Schritten vorgegangen. Bezüglich der **Angabe der Allgemeingültigkeit** lässt sich an die Analyse der Beweisversuche mit prüfendem Ansatz anknüpfen, in der die Existenz und Art eines angegebenen Wahrheitswerts sowie die Existenz eines Hinweises auf die Frage nach der Allgemeingültigkeit untersucht werden (Abschn. 3.3.7). Die dafür eingeführten Begrifflichkeiten und Abgrenzungen lassen sich für die Analyse der Beweisversuche mit weder begründendem noch prüfendem Ansatz übernehmen.

Als Grundlage für die Untersuchung von **Eigenschaften der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für die Entwicklung eines Beweises** wird zunächst mathematikdidaktische Literatur gesichtet, in der Beweisversuche mit weder begründendem noch prüfendem Ansatz – wenn auch unter anderen Begrifflichkeiten – angesprochen werden. Beispielsweise berücksichtigen Fischer und Malle (1985) im Rahmen ihres Konzepts der Argumentationsbasis als mögliche Schlussweisen Berufungen auf eine Autorität, d. h. „Bestätigung[en] der Richtigkeit [einer Aussage] durch einen glaubwürdigen Zeugen, etwa durch einen Text in einem entsprechenden Fachbuch“ (Fischer & Malle, 1985, S. 180) sowie Analogieschlüsse und Wahrscheinlichkeitsaussagen. Die Ergebnisse der Untersuchungen von Galbraith (1981) weisen außerdem darauf hin, dass Lernende manchmal ausschließlich den Wahrheitswert einer infrage stehenden Aussage angeben. Weiterhin nutzt Kempen (2017) bei der Auswertung von Beweisversuchen zu Allaussagen die Kategorie „Nennung des Satzes“ (Kempen, 2017, S. 252). Demnach wird als Grund für die Gültigkeit einer Aussage diese Aussage genannt oder paraphrasiert. Die Ausführungen von Tebaartz und Lengnink (2015) machen ergänzend darauf aufmerksam, dass Lernende eventu-

ell die zu zeigende Aussage wiedergeben, ohne diese als Grund für die Gültigkeit derselben darzustellen. Zudem kommt es in der Untersuchung von Tebaartz (2013) vor, dass die Ausführungen in Beweisversuchen keinen Zusammenhang zu der zu zeigenden Aussage erkennen lassen. Ausgehend von diesen in der mathematikdidaktischen Literatur zu findenden Anregungen werden im Rahmen dieser Arbeit als Kategorien zur Beschreibung von Beweisversuchen mit weder begründendem noch prüfendem Ansatz verschiedene Arten der Bearbeitung differenziert, die sich sowohl sachinhaltlich als auch hinsichtlich möglicher Anknüpfungspunkte für die Entwicklung eines Beweises zu der vorgegebenen Aussage unterscheiden.

3.3.9.2 Vorstellung der Kategorien

Bei jedem Beweisversuch mit weder begründendem noch prüfendem Ansatz werden die Angabe eines Wahrheitswerts und die Art der Bearbeitung analysiert.

Angabe eines Wahrheitswerts

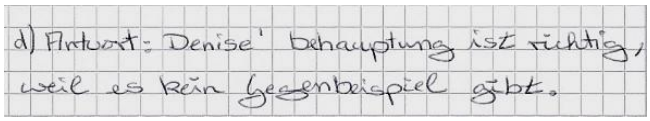
Um zu beschreiben, inwieweit in den Beweisversuchen mit weder begründendem noch prüfendem Ansatz die Allgemeingültigkeit der zu zeigenden Aussage (oder einer nicht ganz korrekten Interpretation derselben) expliziert ist, werden die Existenz und Art eines angegebenen Wahrheitswerts sowie die Existenz eines Hinweises auf die Frage nach der Allgemeingültigkeit analysiert. In der Analyse der **Existenz und Art eines angegebenen Wahrheitswerts** werden die Kategorien Wwert1 *Wahrheitswert angegeben, Affirmation der Aussage*, Wwert0 *Wahrheitswert angegeben, Negation der Aussage*, Wwert55 *Wahrheitswert angegeben, nicht eindeutig kategorisierbarer Wahrheitswert* und Wwert99 *kein Wahrheitswert angegeben* differenziert. Mit Blick auf die Frage nach der Allgemeingültigkeit der Aussage wird zudem erfasst, ob in den Beweisversuchen ein **Hinweis auf die Frage nach der Allgemeingültigkeit der Aussage** vorkommt. Alle Analyseschritte erfolgen vollständig analog zu denen bei Beweisversuchen mit prüfendem Ansatz (Abschn. 3.3.7.2 zu Beweisversuchen mit prüfendem Ansatz).

Arten der Bearbeitung

Bei der Analyse von Beweisversuchen mit weder begründendem noch prüfendem Ansatz werden zur Beschreibung der Art der Bearbeitung vier Oberkategorien und mehrere Unterkategorien unterschieden (Abb. 63).[69] Ein Beweisversuch mit weder begründendem noch prüfendem Ansatz gehört in die Oberkategorie **WNart4** ***Berufung auf die Aussage***, wenn die Aussage selbst als Grund für ihre Gültigkeit angegeben ist. In Beweisversuchen in der Kategorie WNart4.1 *Allgemeingültigkeit der Aussage als Grund* steht als Grund für die Gültigkeit der zu zeigenden Aussage, dass diese für alle Fälle gilt bzw. es kein Gegenbeispiel gibt (Bearbeitung 3.3.9b in Abb. 59). Im Vergleich dazu ist in Beweisversuchen in der WNart4.2 *Gültigkeit der Aussage als Grund* die Allgemeingültigkeit der Aussage nicht hervorgehoben (Bearbeitung 3.3.9b in Abb. 59).

[69] Um viele der Kategorien anhand einzelner Bearbeitungen veranschaulichen zu können, werden in diesem Abschnitt auch Beweisversuche von Lernenden in den Klassenstufen 5, 6 abgebildet. Die Beweisversuche aus den Klassenstufen 5, 6 beziehen sich auf die Aussage, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen durch vier teilbar ist (Abschn. 6.2.2).

Bearbeitung 3.3.9a: Klassenstufe 8, männlich (Code: 0811901)



Bearbeitung 3.3.9b: Klassenstufe 5, weiblich (Code: 0521855)

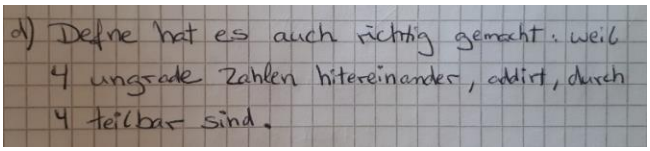
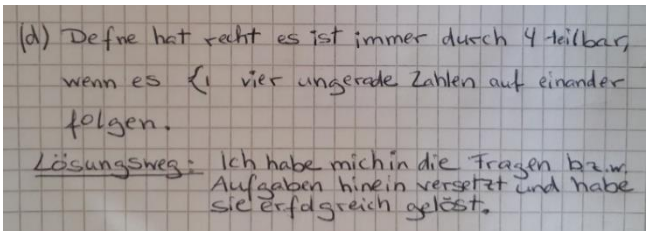
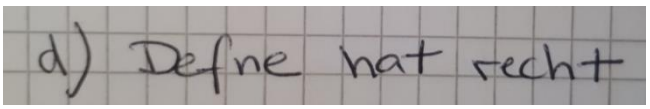


Abbildung 59: Bearbeitung 3.3.9a (WNart4.1 *Berufung auf die Aussage, Allgemeingültigkeit der Aussage als Grund*), Bearbeitung 3.3.9b (WNart4.2 *Berufung auf die Aussage, Gültigkeit der Aussage als Grund*)

Bearbeitung 3.3.9c: Klassenstufe 6, männlich (Code: 0610442)



Bearbeitung 3.3.9d: Klassenstufe 6, weiblich (Code: 0611912)



Bearbeitung 3.3.9e: Klassenstufe 7, männlich (Code: 0722308)

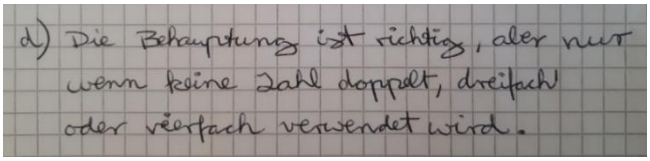


Abbildung 60: Bearbeitung 3.3.9c (WNart3.1 *Angabe der Aussage oder eines Wahrheitswerts, Affirmation mit Betonung der Allgemeingültigkeit*), Bearbeitung 3.3.9d (WNart3.2 *Angabe der Aussage oder eines Wahrheitswerts, Affirmation ohne Betonung der Allgemeingültigkeit*), Bearbeitung 3.3.9e (WNart3.3 *Angabe der Aussage oder eines Wahrheitswerts, keine klare Affirmation*)

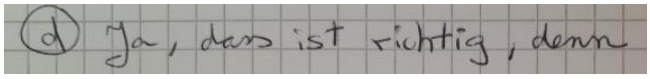
Beweisversuche in der Oberkategorie **WNart3** ***Angabe der Aussage oder eines Wahrheitswerts*** können nicht der Oberkategorie 4 *Berufung auf die Aussage* zugeordnet werden. In Beweisversuchen in der Oberkategorie 3 *Angabe der Aussage oder eines Wahrheitswerts* ist ausschließlich die zu zeigende Aussage wiedergegeben oder angegeben, inwieweit diese Aussage wahr ist. Dabei kann die Aussage wortwörtlich, in paraphrasierter oder nicht ganz korrekter Form wiedergegeben sein. Mit Blick auf die Angabe eines Wahrheitswerts ist es möglich, dass dieser nicht als absolut, sondern als Wahrscheinlichkeitsaussage oder als persönliche Überzeugung dargestellt ist.

In Beweisversuchen in der Kategorie WNart3.1 *Affirmation mit Betonung der Allgemeingültigkeit* weisen die Ausführungen auf die Allgemeingültigkeit der Aussage hin (Bearbeitung 3.3.9c in Abb. 60). In Beweisversuchen in der Kategorie WNart3.2 *Affirmation ohne Betonung der Allgemeingültigkeit* deuten die Ausführungen auf die Gültigkeit der Aussage hin, ohne dass deren

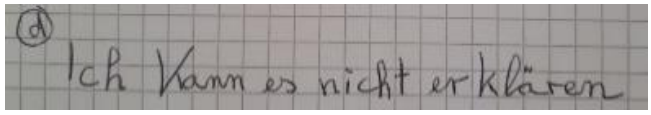
Allgemeingültigkeit betont ist (Bearbeitung 3.3.9d in Abb. 60). Beweisversuche in der Kategorie WNart3.3 *keine klare Affirmation der Aussage* sind der Oberkategorie 3, aber keiner der Kategorien 3.1, 3.2 zuzuordnen, beispielsweise weil eine Bedingung angeführt wurde, wann die Aussage gilt und wann nicht (Bearbeitungen 3.3.9e in Abb. 60).

Beweisversuche in der Oberkategorie **WNart2** ***abgebrochene Bearbeitung*** können keiner der beschriebenen Oberkategorien 3, 4 zugeordnet werden. Die Kategorie WNart2 *abgebrochene Bearbeitung* subsummiert vier Arten an Beweisversuchen mit weder begründendem noch prüfendem Ansatz. Ein Beweisversuch gehört in die Kategorie WNart2.1 *Abbruch nach Angabe der Allgemeingültigkeit*, wenn auf die Allgemeingültigkeit der Aussage hingewiesen und eine angestrebte Fortsetzung der Ausführungen beispielsweise mit einer Kausalkonjunktion wie „weil" angedeutet ist, die Bearbeitung jedoch endet.[70] Bei Beweisversuchen in der Kategorie WNart2.2 *Abbruch nach Angabe der Gültigkeit* wird im Vergleich dazu ausschließlich auf die Gültigkeit der Aussage, nicht explizit auf deren Allgemeingültigkeit hingewiesen (Bearbeitung 3.3.9f in Abb. 61). Beweisversuche in der Kategorie WNart2.3 *Ausführungen ohne Angabe der Gültigkeit* deuten ebenfalls eine angestrebte Fortsetzung der Ausführungen an, enthalten jedoch keinen Hinweis auf die Gültigkeit der Aussage. Dies trifft unter anderem zu, wenn ausschließlich ein strukturierender Begriff, wie „Beispiel:", oder eine Anmerkung, wie „Ich weiß nicht, wie man das begründen kann", aufgeschrieben wurde (Bearbeitung 3.3.9g in Abb. 61). In die Kategorie WNart2.4 *durchgestrichene Ausführungen* gehört ein Beweisversuch, wenn fast alle[71] Ausführungen durchgestrichen oder etwa mit einem Tintenkiller gelöscht wurden (Bearbeitungen 3.3.9h in Abb. 61).

Bearbeitung 3.3.9f: Klassenstufe 7, männlich (Code: 0711913)



Bearbeitung 3.3.9g: Klassenstufe 5, männlich (Code: 0521619)



Bearbeitung 3.3.9h: Klassenstufe 8, männlich (Code: 0811218)

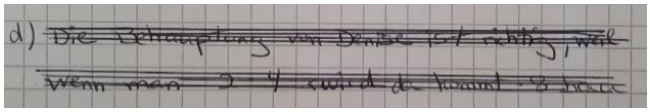
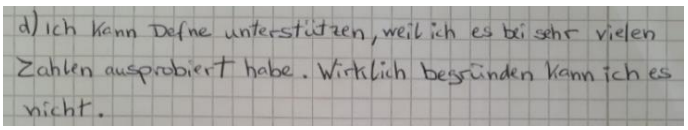


Abbildung 61: Bearbeitung 3.3.9f (WNart2.2 *abgebrochene Bearbeitung, Abbruch nach Angabe der Gültigkeit*), Bearbeitung 3.3.9g (WNart2.3 *abgebrochene Bearbeitung, durchgestrichene Ausführungen*), Bearbeitung 3.3.9h (WNart2.4 *abgebrochene Bearbeitung, durchgestrichene Ausführungen*)

[70] Bei der Entwicklung und Pilotierung des Kategoriensystems kam kein Beweisversuch vor, der in die Kategorie WNart2.1 gehört. Im Sinne der Konsistenz des Kategoriensystem wurde die Kategorie trotzdem mit aufgenommen.
[71] Der Ausdruck „fast alle" meint, dass allenfalls kleine Teile der Ausführungen, wie etwa „Beispiel:", nicht durchgestrichen sind. Bei Beweisversuchen in der Kategorie WNart2.4 ist zumindest ansatzweise zu erkennen, dass eventuell ein Beispiel oder eine Beweisidee angegeben wurde. Die Ausführungen sind jedoch durchgestrichen, sodass die Bearbeitung keiner der Oberkategorien *prüfender Ansatz* und *begründender Ansatz* zugeordnet werden kann.

Jeder Beweisversuch mit weder begründendem noch prüfendem Ansatz, der in keine der beschriebenen Kategorien fällt, wird der Restkategorie **WNart1** ***sonstiges*** zugeordnet. Beispielsweise gibt es Lernende, die behaupten, ein oder mehrere Beispiele betrachtet zu haben, ohne diese anzugeben (Bearbeitung 3.3.9i in Abb. 62). Auch ist es möglich, dass Ausführungen keinen klaren Bezug zur gegebenen Aussage aufweisen (Bearbeitung 3.3.9j in Abb. 62).[72] Darüber hinaus können z. B. Berufungen auf eine Autorität vorkommen.

Bearbeitung 3.3.9i: Klassenstufe 6, männlich (Code: 0620604)



Bearbeitung 3.3.9j: Klassenstufe 7, männlich (Code: 0710416)

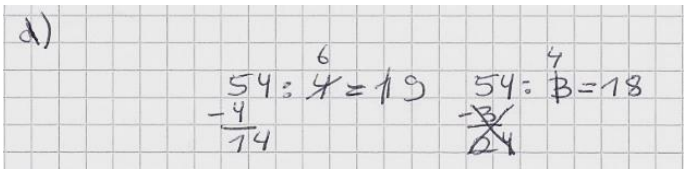


Abbildung 62: Bearbeitung 3.3.9i-j (WNart1 *sonstiges*)

3.3.9.3 Hinweise zu den Kategorien

Bei Beweisversuchen mit weder begründendem noch prüfendem Ansatz wird erstens untersucht, inwieweit die Allgemeingültigkeit der zu zeigenden Aussage (oder einer nicht ganz korrekten Interpretation dieser Aussage) angegeben und somit deren Allgemeingültigkeit entsprechend der Aufgabenstellung bejaht ist. Dies ist mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage, d. h. insbesondere auch aus innermathematischer Sicht, von Bedeutung (Abschn. 8.4.1, 8.4.3 zu den Ergebnissen).

Zweitens wird bei Beweisversuchen mit weder begründendem noch prüfendem Ansatz aus einer primär mathematikdidaktischen Sicht die Art der Bearbeitung beschrieben (Abschn. 8.4.2, 8.4.3.2 zu den Ergebnissen). Die Unterscheidung verschiedener Arten der Bearbeitung knüpft an vorhandene mathematikdidaktische Literatur zu Beweisversuchen mit weder begründendem noch prüfendem Ansatz an und ermöglicht es, mehrere Eigenschaften derselben gebündelt in einem Analyseschritt zu erfassen. Zudem wurden bei der Differenzierung der Arten der Bearbeitung mögliche Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche mitgedacht. In Beweisversuchen mit weder begründendem noch prüfendem Ansatz können die Ausführungen darauf hinweisen, dass die zu zeigende Aussage als gültig angesehen wird.[73] Außerdem können Sie andeuten, dass die Frage nach der Allgemeingültigkeit bedacht wurde. Des Weiteren können Ausführungen in Beweisversuchen mit weder begründendem noch prüfendem Ansatz darauf hindeuten, dass die Lernenden versucht haben, einen Grund für die Gültigkeit der Aussage oder ein Beispiel anzugeben. Inwieweit diese Anknüpfungspunkte je nach Art

[72] Die in der Bearbeitung betrachtete Zahl 54 ist keine Summe von vier aufeinanderfolgenden ungeraden Zahlen. Sie wird in der Bearbeitung zudem nicht durch acht, sondern durch sechs bzw. vier dividiert.

[73] Dies bedeutet nicht unbedingt, dass die Begrifflichkeiten der zu zeigenden Aussage korrekt interpretiert wurden. Inwieweit dies der Fall ist, lassen Beweisversuche mit weder begründendem noch prüfendem Ansatz i. d. R. offen.

Kategorien	Anknüpfungspunkte für die Entwicklung eines Beweises			
	Angabe der Gültigkeit	Hinweis auf die Frage nach der Allgemeingültigkeit	ansatzweise Angabe eines Grundes	ansatzweise Angabe eines Beispiels
4.1 *Berufung auf die Aussage, Allgemeingültigkeit der Aussage als Grund*	x	x	x	--
4.2 *Berufung auf die Aussage, Gültigkeit der Aussage als Grund*	x	--	x	--
3.1 *Angabe der Aussage oder eines Wahrheitswerts, Affirmation mit Betonung der Allgemeingültigkeit*	x	x	--	--
3.2 *Angabe der Aussage oder eines Wahrheitswerts, Affirmation ohne Betonung der Allgemeingültigkeit*	x	--	--	--
3.3 *Angabe der Aussage oder eines Wahrheitswerts, keine klare Affirmation*	--	x--	--	--
2.1 *abgebrochene Bearbeitung, Abbruch nach Angabe der Allgemeingültigkeit*	x	x	x	
2.2 *abgebrochene Bearbeitung, Abbruch nach Angabe der Gültigkeit*	x	--	x	
2.3 *abgebrochene Bearbeitung, Ausführungen ohne Angabe der Gültigkeit*	--	x--	x	
2.4 *abgebrochene Bearbeitung, durchgestrichene Ausführungen*	--	--	x	
1 *sonstiges*	x--	x--	x--	x--
	x: trifft zu	x--: trifft vielleicht zu	--: trifft nicht zu	

Abbildung 63: Beweisversuche mit weder begründendem noch prüfendem Ansatz, Arten der Bearbeitung und Anknüpfungspunkte für die Entwicklung eines Beweises

der Bearbeitung in den Beweisversuchen mit weder begründendem noch prüfendem Ansatz zu erkennen sind, ist in Abbildung 63 zu sehen.

In Beweisversuchen in den Kategorien WNart4.1,4.2 ist vom Ansatz her ein Grund für die Gültigkeit der Aussage expliziert. Die Beweisversuche legen nahe, dass die Notwendigkeit der Angabe eines Grundes erkannt wurde. Mit Blick auf deren Weiterentwicklung stellt sich i. d. R. die Frage, welche „Gründe“ in der Mathematik als Beweis anerkannt werden.

Beweisversuche in den Kategorien WNart4.1,4.2,3.1,3.2,2.1,2.2 weisen darauf hin, dass die zu zeigende Aussage (oder eine nicht ganz korrekte Interpretation derselben) als gültig angesehen wird. Dabei unterscheiden sich Beweisversuche in den Kategorien 4.1 und 4.2, 3.1 und 3.2 sowie 3.1 und 3.2 jeweils hinsichtlich des Hinweises auf die Frage nach der Allgemeingültigkeit, der nur bei Beweisversuchen in den Kategorien 4.1, 3.1, 2.1 vorkommt.

Bei Beweisversuchen in den Kategorien WNart3.3,2.3 ist nicht angegeben, dass die Aussage gültig ist, wobei bei denen in der Kategorie 3.3 ein anderer Wahrheitswert, bei denen in der Kategorie 2.3 kein Wahrheitswert angedeutet ist. Für eine Weiterentwicklung der Beweisversuche zu einem Beweis wäre eventuell eine Klärung der Interpretation der vorgegebenen Aussage und dessen Wahrheitswerts erforderlich.

Beweisversuche in der Kategorie WNart2.4 *abgebrochene Bearbeitung, durchgestrichene Ausführungen* sowie solche in der Kategorie WNart1 *sonstiges* können hinsichtlich möglicher Anknüpfungspunkte für die Entwicklung eines Beweises sehr unterschiedlich sein. Beweisversuche in der Kategorie WNart2.4 lassen sich, so wie alle Beweisversuche in den Kategorien WNart2.1-2.3, als abgebrochene Bearbeitung bezeichnen. Im Unterschied zu allen anderen Beweisversuchen mit weder begründendem noch prüfendem Ansatz wurde von der Autorin bzw. dem Autor eventuell eine Beweisidee oder ein Beispiel aufgeschrieben, im Vergleich zu Beweisversuchen in einer der Oberkategorien *begründender Ansatz* und *prüfender Ansatz* nur durchgestrichen. Beweisversuche in der Kategorie WNart2.4 stellen somit innerhalb der Beweisversuche mit weder begründendem noch prüfendem Ansatz einen Sonderfall dar, weil sie mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung über die anderen Beweisversuche mit weder begründendem noch prüfendem Ansatz deutlich hinausgehen können. Die Kategorie WNart1 *sonstiges* ist weit gefasst. Beispielsweise deutet die Bearbeitung 3.3.9i in Abbildung 62 an, dass der Schüler die Idee hatte, Beispiele zu betrachten sowie eine Begründung zu formulieren. Zudem legt insbesondere die treffende Verwendung des Verbs „unterstützen" eine Bewusstheit für die Unzulänglichkeit von Beispielen und die Notwendigkeit einer allgemeingültigen Begründung nahe. In der Bearbeitung 3.3.9j in Abbildung 62 hingegen ist allenfalls die Idee der Betrachtung eines Beispiels zu erkennen.

In die Kategorie WNart1 *sonstiges* werden unter anderem auch Berufungen auf eine Autorität aufgenommen. Berufungen auf eine Autorität werden nicht in eine eigene Kategorie eingeordnet, weil sie bei der Entwicklung des Kategoriensystems nicht vorkamen. Dies kann sowohl mit der Vorgabe einer den Lernenden tendenziell unbekannten Aussage als auch mit dem Setting der Wettbewerbsklausur, in dem die Beweisversuche entstanden sind, zusammenhängen (Abschn. 6.2 zum Design). Insbesondere mit Blick auf die Restkategorie WNart1 kann es sinnvoll sein, die vorgenommene Differenzierung von Arten der Bearbeitung bei der Analyse von Beweisversuchen zu anderen Aussagen oder aus anderen Settings anzupassen.

3.3.10 Nicht eindeutig kategorisierbarer Ansatz

Dieser Abschnitt fokussiert auf Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz. In Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz sind Ausführungen zu dem infrage stehenden Aufgabenteil vorhanden, können jedoch keiner der Kategorien *begründender Ansatz*, *prüfender Ansatz* und *weder begründender noch prüfender Ansatz* eindeutig zugeordnet werden (Abschn. 3.3.2). Um die Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz bei der Beschreibung von Charakteristika aller Beweisversuche so gut es geht miteinbeziehen zu können, werden bei diesen Beweisversuchen, soweit möglich, die gleichen Eigenschaften analysiert wie bei denen, die einem bestimmten Ansatz zugeordnet werden können. Beispielsweise lassen sich auch bei Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz die Passung der Ausführungen zu der zu zeigenden Aussage sowie die Anzahl und Art an Beispielen beschreiben (Abb. 64).

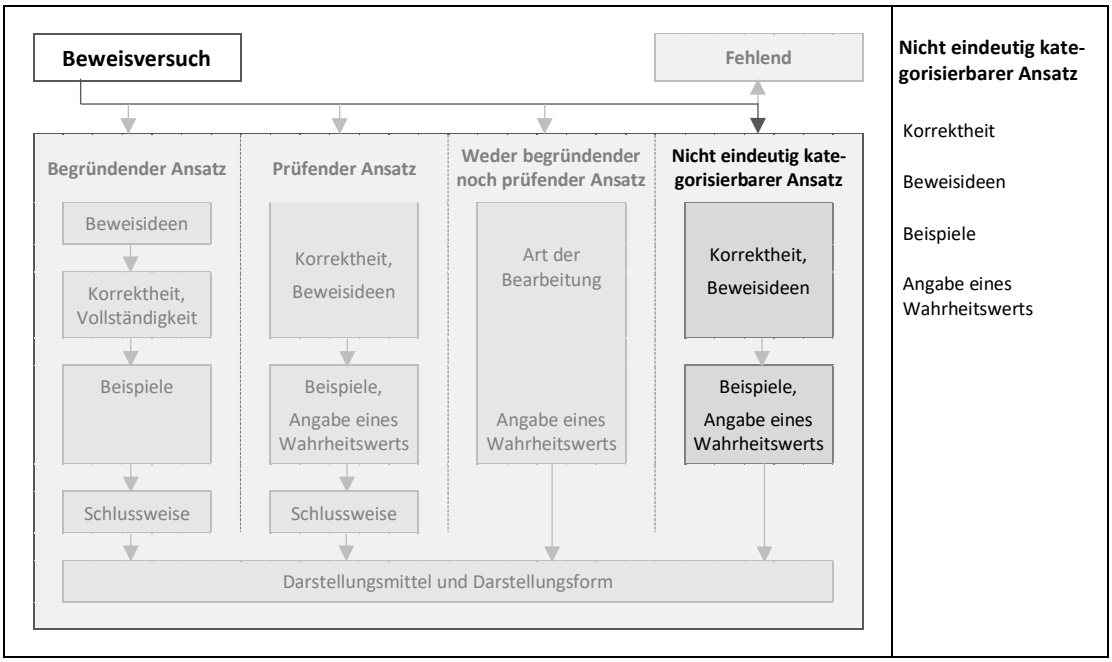


Abbildung 64: Analysemodell. Fokus: Nicht eindeutig kategorisierbarer Ansatz, Korrektheit, Beweisideen, Beispiele und Angabe eines Wahrheitswerts

3.3.10.1 Entwicklung der Kategorien

Um die Ergebnisse der Analyse der Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz möglichst gut mit denen der detaillierten Analysen der Beweisversuche mit begründendem, prüfendem und weder begründendem noch prüfendem Ansatz zusammenführen zu können, werden für die Analyse der Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz keine neuen Kategorien generiert. Es werden systematisch aus den Analysen der anderen Beweisversuche alle Eigenschaften und dazugehörigen Kategorien ausgewählt, die bei Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz erfasst werden können. Dabei werden die Kategoriendefinitionen, soweit notwendig, leicht an die Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz adaptiert. Die daraus resultierende Analyse von Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz wird im Folgenden präsentiert.

3.3.10.2 Vorstellung der Kategorien

Bei Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz werden nacheinander die Passung der Ausführungen zu der zu zeigenden Aussage, die Erkennbarkeit einer Beweisidee und ggf. die primäre Beweisidee, die Verwendung von Beispielen sowie die Angabe eines Wahrheitswerts analysiert. Im Folgenden werden diese Analysen kurz allgemein beschrieben. Zudem sind in Abbildung 65 exemplarisch die Codierungen dreier Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz zu sehen.[74]

Korrektheit: Passung zu der zu zeigenden Aussage

Analog zur Analyse von Beweisversuchen mit begründendem oder prüfendem Ansatz (Abschn. 3.3.4.2, 3.3.7.2) wird mithilfe der Kategorien PzzA1 *Hinweis auf einen grundlegenden*

[74] Im Sinne einer guten Veranschaulichung ist dabei auch ein Beweisversuch aus der Klassenstufe 6 angegeben.

Fehler, PzzA2 *Bezug auf die zz. Aussage unklar* und PzzA3 *Passung zu der zz. Aussage* beschrieben, inwieweit die Ausführungen in Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz zu der zu zeigenden Aussage passen.

Beweisideen

Im Vergleich zu der Analyse von Beweisversuchen mit begründendem Ansatz ist es bei Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz erforderlich, zuerst die Erkennbarkeit einer Beweisidee zu erfassen. Unter einer Beweisidee wird in dieser Arbeit eine inhaltsbezogene Überlegung verstanden, die dem Argumentgefüge zugrunde liegt und auf Voraussetzungen der infrage stehenden Aussage oder anderen mathematischen Sätzen und Definitionen beruht (Abschn. 3.3.3). Dabei ist es nicht notwendig, dass die Ausführungen genau zu der zu zeigenden Aussage passen. Zudem reicht es aus, wenn die Beweisidee in einer von mehreren möglichen Interpretationen eines Beweisversuchs mit nicht eindeutig kategorisierbarem Ansatz vorkommt. Wenn in einem Beweisversuch mit nicht eindeutig kategorisierbarem Ansatz eine Beweisidee erkennbar ist, wird ergänzend die primäre Beweisidee codiert. Die Codierung der primären Beweisidee erfolgt analog zur Analyse von Beweisversuchen mit begründendem oder prüfendem Ansatz (Abschn. 3.3.3.2, 3.3.7.2).

Beispiele

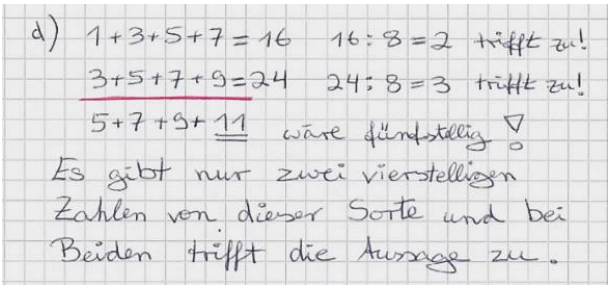
Ebenfalls anknüpfend an die Analysen von Beispielen in Beweisversuchen mit begründendem oder prüfendem Ansatz werden bei Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz nacheinander die Existenz und Passung eines Beispiels zu der zu zeigenden Aussage, die Anzahl an Beispielen und die Art der Beispiele untersucht.

Zur Beschreibung, inwieweit die zu zeigende Aussage in Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz für einen Spezialfall bestätigt ist, wird jeder Beweisversuch mit nicht eindeutig kategorisierbarem Ansatz, analog zur Analyse der Existenz und Passung eines Beispiels bei Beweisversuchen mit begründendem Ansatz, einer von sechs Kategorien zugeordnet (Abschn. 3.3.5.2 zu den Kategorien). Anschließend wird, so wie bei Beweisversuchen mit begründendem Ansatz und solchen mit prüfendem Ansatz, die Anzahl an vorkommenden Beispielen erfasst. Zudem wird, analog zur Analyse von Beweisversuchen mit begründendem oder prüfendem Ansatz, beschrieben, inwieweit systematische Beispiele, ein besonderes Beispiel und ein großes Beispiel vorkommen (Abschn. 3.3.5.2, 3.3.7.2 zu den Kategorien).

Angabe eines Wahrheitswerts

Nach den Beispielen steht die Angabe eines Wahrheitswerts im Fokus. So wie bei Beweisversuchen mit prüfendem Ansatz und denen mit weder begründendem noch prüfendem Ansatz werden mit den Kategorien Wwert1 *Wahrheitswert angegeben, Affirmation der Aussage*, Wwert0 *Wahrheitswert angegeben, Negation der Aussage*, Wwert55 *Wahrheitswert angegeben, nicht eindeutig kategorisierbarer Wahrheitswert* und Wwert99 *kein Wahrheitswert angegeben* die Existenz und ggf. die Art eines Wahrheitswerts untersucht (Abschn. 3.3.7.2 zu den Kategoriendefinitionen).

Bearbeitung 3.3.10a: Klassenstufe 7, männlich (Code: 0711225)



Passung zu der zu zeigenden Aussage
PzzA1 *Hinweis auf einen grundlegenden Fehler*

Beweisideen
- ExBI1 *Beweisidee erkennbar*
- BIp55 *nicht eindeutig kategorisierbar*

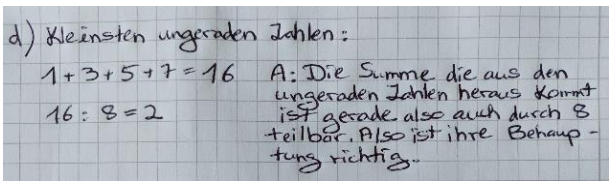
Nicht eindeutig kategorisierbarer Ansatz

Existenz und Passung eines Beispiels zu der zz. Aussage: PBsp5 *vollständig und korrekt ausgeführtes eigenes Beispiel zu der zz. Aussage*

Anzahl und Art angegebener Beispiele
- BspN2 *zwei bis fünf Beispiele*
- BspNzzA2 *zwei bis fünf Beispiele zu der zu zeigenden Aussage*
- BspNsonst0 *kein sonstiges Beispiel*
- BspSyst0 *keine systematischen Beispiele*
- Bspbes2 *kleinstes Beispiel*
- Bspgr0 *kein großes Beispiel*

Angabe eines Wahrheitswerts: Wwert1 Wahrheitswert angegeben, Affirmation der Aussage

Bearbeitung 3.3.10b: Klassenstufe 8, weiblich (Code: 0821604)



Passung zu der zu zeigenden Aussage
PzzA3 *Passung zu der zz. Aussage*

Beweisideen
- ExBI1 *Beweisidee erkennbar*
- BIp9 *Parität von Summen ungerader oder gerader Zahlen*

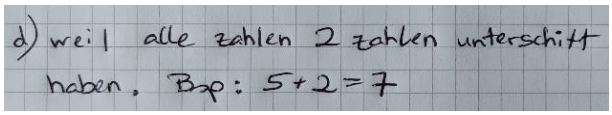
Nicht eindeutig kategorisierbarer Ansatz

Existenz und Passung eines Beispiels zu der zz. Aussage: PBsp5 *vollständig und korrekt ausgeführtes eigenes Beispiel zu der zz. Aussage*

Anzahl und Art angegebener Beispiele
- BspN1 *genau ein Beispiel*
- BspNzzA2 *genau ein Beispiel zu der zu zeigenden Aussage*
- BspNsonst0 *kein sonstiges Beispiel*
- BspSyst0 *keine systematischen Beispiele*
- Bspbes2 *kleinstes Beispiel*
- Bspgr0 *kein großes Beispiel*

Angabe eines Wahrheitswerts: Wwert1 *Wahrheitswert angegeben, Affirmation der Aussage*

Bearbeitung 3.3.10c: Klassenstufe 6, männlich (Code: 0612108)



Passung zu der zu zeigenden Aussage
PzzA2 *Bezug auf die zz. Aussage unklar*

Beweisideen
- ExBI1 *Beweisidee erkennbar*
- BIp10 *Abstände ungerader Zahlen*

Nicht eindeutig kategorisierbarer Ansatz

Existenz und Passung eines Beispiels zu der zz. Aussage: PBsp1 *sonstiges Beispiel, keine Addition von vier ungeraden Zahlen*

Anzahl und Art angegebener Beispiele
- BspN1 *genau ein Beispiel*
- BspNzzA0 k*ein Beispiel zu der zz. Aussage*
- BspNsonst1 *genau ein sonstiges Beispiel*
- BspSyst0 *keine systematischen Beispiele*
- Bspbes1 *anderes besonderes Beispiel*
- Bspgr0 *kein großes Beispiel*

Angabe eines Wahrheitswerts: Wwert1 *Wahrheitswert angegeben, Affirmation der Aussage*

Abbildung 65: Bearbeitungen 3.3.10a-c (Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz: Analyse der Passung der Ausführungen zu der zu zeigenden Aussage, der Erkennbarkeit einer Beweisidee und ggf. der primären Beweisidee, der Existenz, Passung, Anzahl und Art von Beispielen sowie der Existenz und Art eines Wahrheitswerts)

3.3.10.3 Hinweise zu den Kategorien

Bei Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz werden in der beschriebenen Form, möglichst analog zur Analyse von Beweisversuchen mit begründendem, prüfendem oder weder begründendem noch prüfendem Ansatz mehrere Eigenschaften en détail untersucht. Wenn dabei an einzelnen Stellen keine klar begründbare Zuordnung zu einer Kategorie möglich ist, werden die Beweisversuche diesbezüglich als nicht eindeutig kategorisierbar be-

zeichnet. Auf diese Weise lassen sich auch die Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz bei der Untersuchung von Charakteristika von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade in möglichst aussagekräftiger Form berücksichtigen.

3.3.11 Darstellungsmittel und Darstellungsform

Unabhängig von dem Ansatz werden die Beweisversuche in einem letzten Analyseschritt hinsichtlich der verwendeten Darstellungsmittel und ihrer Darstellungsform analysiert (Abb. 66). Dafür werden wichtige Aspekte zum Darstellen, die auch in der Literatur zum Beweisen aufgegriffen werden, wie z. B. die Verwendung symbolischer und ikonischer Darstellungsmittel, fokussiert. Auf eine umfassende Rezeption der Theorie zum Darstellen wird aufgrund des anderen Fokus der Arbeit verzichtet. Ausgehend von wenigen theoretischen Grundlagen wird im Folgenden das dafür entwickelte Kategoriensystem erklärt.[75]

Die Analyse der Darstellungsmittel und der Darstellungsform soll Einblicke geben, welche Darstellungsmittel in Beweisversuchen von Teilnehmenden der Mathematik-Olympiade vorkommen und wie sie im Wesentlichen dargestellt sind. Die diesbezüglichen Ergebnisse werden im achten Kapitel präsentiert (insb. Abschn. 8.2.4, 8.2.6.5, 8.3.3, 8.4.2, 8.5.2.4).

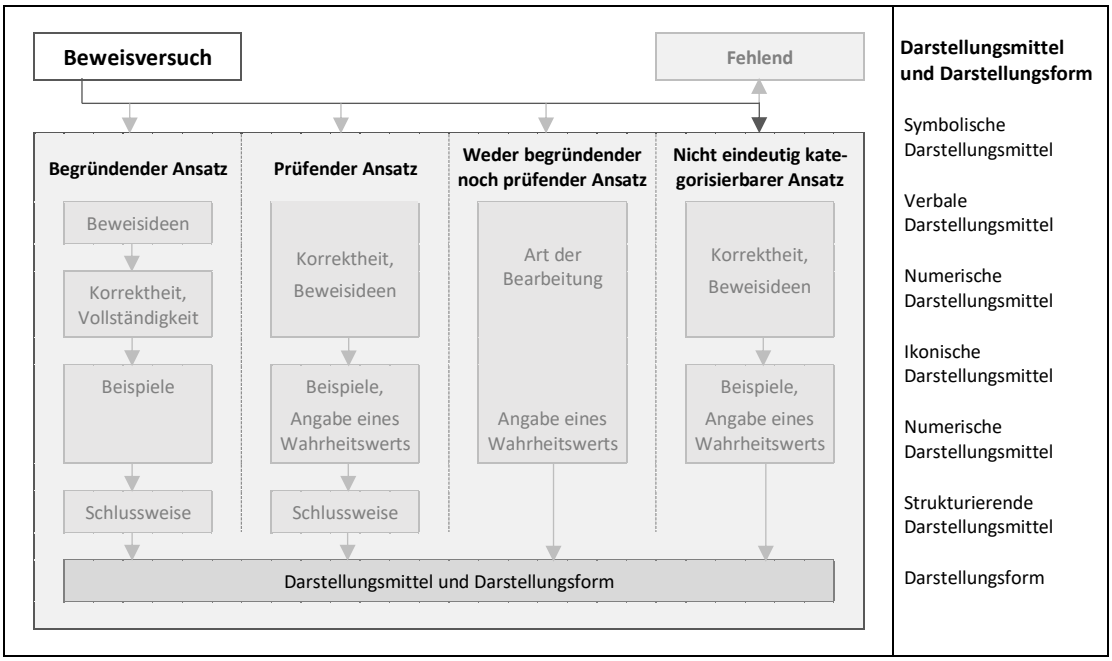


Abbildung 66: Analysemodell. Fokus: Darstellungsmittel und Darstellungsform

[75] Der Autorin der vorliegenden Arbeit ist es ein Anliegen, die vorkommenden Darstellungsmittel und die Darstellungsform als diskutierte, gerade mit Blick auf junge Lernende interessante Eigenschaften von Beweisversuchen in das Analysemodell mit aufzunehmen. Eine fundierte und detaillierte Analyse der Verwendung von Darstellungsmitteln lässt sich als ein kleiner Baustein dieses Dissertationsprojekts allerdings nicht leisten. Wer es genauer wissen und erforschen möchte, darf sich selbstverständlich melden. Gerne werden die im Rahmen dieser Arbeit erhobenen Beweisversuche für etwaige Folgeprojekte zur Verfügung gestellt.

3.3.11.1 Entwicklung der Kategorien

Entsprechend den einleitenden Überlegungen werden bei der Entwicklung der Kategorien nacheinander symbolische, verbale, ikonische, numerische und strukturierende Darstellungsmittel fokussiert.

In der Wissenschaft Mathematik werden Beweise üblicherweise möglichst präzise dargestellt. Dafür wird in der Regel eine mathematische Fachsprache verwendet, was insbesondere den Gebrauch **symbolischer** Darstellungsmittel einschließt (Abschn. 3.1.2). In der mathematikdidaktischen Literatur zum Beweisen im Schulunterricht hingegen wird häufig betont, dass die symbolische Darstellung nicht ins Zentrum gestellt werden sollte (u. a. Hanna, 1989; Ufer et al., 2009; Wittmann & Müller, 1988). Maier (1999) erklärt:

> Mit [einer Festlegung auf einen extensiven Symbolgebrauch] ist die Gefahr verbunden, dass mindestens ein Teil der Schüler mit der Aufgabe des Beweisens einfach deshalb Schwierigkeiten bekommt, weil ihm diese Idealform der Kodierung von Beweisgedanken nicht gelingt und er umgekehrt weitgehend mit Symbolen dargestellte Beweise nicht zu dekodieren und damit nicht zu verstehen vermag. (Maier, 1999, S. 25)

Empirische Untersuchungen zeigen, dass es manchen Lernenden gelingt, eine Allaussage in primär symbolisch dargestellter Form zu beweisen, und deuten zugleich darauf hin, dass vielen Lernenden die Verwendung symbolischer Darstellungsmittel bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit Probleme bereitet (Abschn. 5.3). In der Studie von Healy und Hoyles (2000) ist die Qualität der primär symbolisch dargestellten Beweisversuche schlechter als die der primär **verbal** dargestellten Bearbeitungen. Um die Form von Beweisversuchen von Schülerinnen und Schülern zu erfassen, unterscheiden Healy und Hoyles (2000) unter anderem die Kategorien „formal (algebraic)“ (Healy & Hoyles, 2000, S. 408) und „narrative“ (ebd.). Dazu erläutern Reid und Knipping (2010):

> Some proofs […] use only or mainly words, and have been called “narrative” (Healy & Hoyles, 2000) for this reason. Others use very few words and many symbols specific to mathematics […]. These have been called “symbolic” (Healy & Hoyles, 2000). These are not two distinct subcategories, however, as proof might be sited anywhere on a continuum from highly narrative to almost completely symbolic. (Reid & Knipping, 2010, S. 140)

In der Praxis gibt es ein breites Spektrum zwischen rein symbolisch und rein verbal dargestellten Beweisversuchen. Zudem ist es möglich, dass in Beweisversuchen auch ikonische oder numerische Darstellungsmittel vorkommen.

In der mathematikdidaktischen Literatur liegt neben den symbolischen und verbalen Darstellungsmitteln ein besonderes Augenmerk auf ikonischen Darstellungsmitteln. Einerseits bestehen Vorbehalte hinsichtlich der Gültigkeit **ikonisch** dargestellter Beweise. Zudem kann das Hineinsehen des „Beweises“ in eine ikonische Darstellung anspruchsvoll sein (z. B. Dreyfus et al., 2012; Winter, 1991; Abschn. 3.3.2). Andererseits können mithilfe ikonischer Darstellungen Ideen sichtbar gemacht werden, die aus formalen Definitionen oder reinem algebraischen Kalkül nicht direkt hervorgehen (Dreyfus et al., 2012). Zudem erfordern ikonische Darstellungen von Beweisen keine Algebra, sodass sie auch für junge Schülerinnen und Schüler zugänglich sind (z. B. Neubrand & Möller, 1990). Viele Möglichkeiten zum Argumentieren und Beweisen mit Lernenden in der Grundschule und zu Beginn der Sekundarstufe I bieten figurierte Zahlen,

die inzwischen „ein konsentierter Inhalt der Mathematiklehrerbildung [sind]“ (Jahnke & Ufer, 2015, S. 338). Zudem machen Leiß und Blum (2006) deutlich, dass auch mit ikonischen Darstellungsmitteln wie Klammern und Pfeilen Beweisideen verdeutlicht werden können.

Die Nutzung **numerischer** Darstellungsmittel, etwa bei der Angabe von Beispielen, ist zumindest mit Blick auf Beweisversuche von Lernenden ebenfalls relevant. Numerische Darstellungsmittel werden in der Untersuchung von Healy und Hoyles (2000) nicht klar separiert. Um die Form von Beweisversuchen von Schülerinnen und Schülern zu erfassen, haben Healy und Hoyles (2000) zwischen den Kategorien „formal (algebraic)“ (Healy & Hoyles, 2000, S. 408), „narrative“ (ebd.), „empirical“ (ebd.) und „other (e. g. visual proofs)“ (ebd.) differenziert. Dabei können in Beweisversuchen, die nach deren Kategorisierung als empirical bezeichnet werden, durchaus unterschiedliche Darstellungsmittel (z. B. numerische, ikonische, verbale Darstellungsmittel) vorkommen. In der Untersuchung von Healy und Hoyles (2000) sowie auch in anderen Studien und theoretischen Klassifikationen von Beweisversuchen sind die verwendeten Darstellungsmittel in Verbindung mit anderen Merkmalen von Beweisversuchen berücksichtigt (z. B. Grundey, 2015; Kempen, 2019; Leiß & Blum, 2006; Leuders, 2010).

Strukturierende Darstellungsmittel werden in den der Autorin bekannten Studien mit Schülerinnen und Schülern nicht gesondert aufgeführt. In der Wissenschaft Mathematik werden Beweise jedoch meistens klar strukturiert (Abschn. 3.1.2). Als einen Anhaltspunkt für Studienanfängerinnen und -anfänger im Bereich der Mathematik schreibt Beutelspacher (2009), dass ein Beweis so formuliert sein sollte, dass die Leserinnen und Leser möglichst wenig Mühe und Zeit aufwenden müssen, um diesen zu verstehen. Dazu empfiehlt er beispielsweise, anfangs präzise zu sagen, welche Aussage bewiesen werden soll und lange Beweise klar zu gliedern. Es sei für Lesende zudem hilfreich, das Ende eines Beweises z. B. mit dem Beweisabschlusszeichen, einer Abkürzung wie q. e. d. (quod erat demonstrandum) oder einem Satz wie ‚Damit ist die Behauptung bewiesen‘ zu kennzeichnen (Beutelspacher, 2009). Für Teilnehmende der Mathematik-Olympiade in den Klassenstufen sieben und acht hat die Gruppe des Aufgabenausschusses, die für diese Klassenstufen verantwortlich ist, in den Schuljahren 2010/2011 und 2011/2012 auf den Aufgabenblättern der Schulrunde einige mathematische Grundlagen und Anforderungen, die mit den Wettbewerbsaufgaben verbunden sind, erläutert.[76] Dabei empfehlen sie den Lernenden, Bedingungen und Schlussfolgerungen, auf die im weiteren Verlauf einer Aufgabenbearbeitung zurückgegriffen wird, durchzunummerieren:

> Um angeben zu können, aus welcher gegebenen Bedingung eine Schlussfolgerung gezogen wurde, ist es günstig, die gegebenen Bedingungen zu bezeichnen, im Fall von Aufgabe 500713 mit (1), (2) und (3). Dies trifft auch auf *abgeleitete Feststellungen* zu, auf die später zurückgegriffen wird. Diese können bei Aufgabe 500713 mit (4), (5) usw. bezeichnet werden. (Aufgabenausschuss des Mathematik-Olympiaden e. V., 2010, S. 2, H. i. O.)

Beweisabschlusszeichen, Nummerierungen und weitere strukturierende Darstellungsmittel werden in die Analyse der Darstellungsmittel im Rahmen der vorliegenden Arbeit ergänzend

[76] Diese Erläuterungen sind ausschließlich auf den genannten Aufgabenblättern aus den Schuljahren 2010/2011, 2011/2012 zu finden. Es ist folglich nicht davon auszugehen, dass diese Erläuterungen Teilnehmenden der im Rahmen dieser Arbeit, im Schuljahr 2016/2017 durchgeführten Studie bekannt sind.

zu symbolischen, verbalen, ikonischen und numerischen Darstellungsmitteln mit aufgenommen. Auf diese Weise lässt sich ein Einblick geben, inwieweit Teilnehmende der Mathematik-Olympiade strukturierende Darstellungsmittel verwenden. Zudem ermöglicht die Berücksichtigung strukturierender Darstellungsmittel, alle in einem Beweisversuch vorkommenden Ausführungen[77] entweder als symbolisch, als verbal, als ikonisch, als numerisch oder als strukturierend zu bezeichnen.

Da die verwendeten Darstellungsmittel nicht unbedingt mit weiteren Merkmalen von Beweisversuchen, wie etwa den ausgeführten Beweisideen, zusammenhängen, werden die verwendeten Darstellungsmittel **in der vorliegenden Arbeit** separat betrachtet. Da prinzipiell sowohl in Beweisversuchen mit begründendem Ansatz als auch in solchen mit prüfendem oder nicht eindeutig kategorisierbarem Ansatz alle unterschiedenen Darstellungsmittel vorkommen können, werden die verwendeten Darstellungsmittel unabhängig vom Ansatz mit dem gleichen Kategoriensystem analysiert. Mit dem Kategoriensystem wird erstens erfasst, welche Darstellungsmittel in den Beweisversuchen der Lernenden vorkommen, um aufzuzeigen, ob beispielsweise symbolische[78] oder ikonische Darstellungsmittel verwendet werden, selbst wenn diese von marginaler Bedeutung sind. In einem zweiten Schritt soll erfasst werden, was die Darstellungsform ist, um Randständiges zu relativieren und die Ergebnisse der im Rahmen dieser Arbeit durchgeführten Studie in den derzeitigen Forschungsstand einordnen zu können.

3.3.11.2 Vorstellung der Kategorien

Mit dem vorliegenden Analysemodell soll erfasst werden, welche Darstellungsmittel in den Beweisversuchen der Lernenden vorkommen und wie die Beweisversuche im Wesentlichen dargestellt sind. Dazu werden symbolische, ikonische, numerische, verbale und strukturierende Darstellungsmittel unterschieden sowie die Darstellungsform beschrieben.

Symbolische Darstellungsmittel

In dieser Arbeit werden unter der Verwendung von symbolischen Darstellungsmitteln in Anlehnung an Maier (1999) der Gebrauch von Symbolen mit fester Bedeutung, wie etwa dem Symbol für natürliche Zahlen und dem Zeichen für Teilbarkeit, die den Lernenden nicht unbedingt aus der Schule bekannt sind, sowie der Gebrauch von Buchstabenvariablen und damit einhergehend von Termen, Funktionen etc. zusammengefasst.

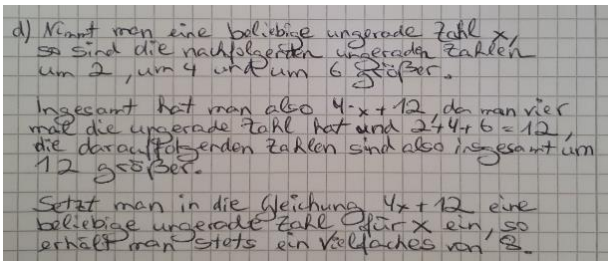
Bei der Analyse wird als grundlegende Information erfasst, ob im vorliegenden Beweisversuch ein symbolisches Darstellungsmittel vorkommt. Beispielsweise kommen in den Bearbeitungen 3.3.11a-b in Abbildung 67 symbolische Darstellungsmittel vor, in den Bearbeitung 3.3.11c-e in Abbildung 68 nicht.

---

[77] Ausschließlich die Nummerierung des Aufgabenteils (z. B. „d“, „Aufgabe 2d“, „Denise:“) und eventuelle Bemerkungen der Art „Auf der nächsten Seite geht's weiter“ bleiben bei der Analyse der Darstellungsmittel, so wie bei der gesamten Analyse eines vorliegenden Beweisversuchs, unberücksichtigt.

[78] Falls in einem Beweisversuch Buchstabenvariablen vorkommen, soll nebenbei ergänzend erfasst werden, ob diese angemessen definiert sind. Die Variablendefinitionen sollen an dieser Stelle betrachtet werden, weil sie vor dem Hintergrund des schulischen Vorwissens junger Lernender bei der Analyse der Korrektheit und Vollständigkeit von Beweisversuchen nicht berücksichtigt werden (Abschn. 3.3.4).

Bearbeitung 3.3.11a: Klassenstufe 7, männlich (Code: 0711201)



Darstellungsmittel und Darstellungsform

Symbolische Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Symbol mit fester Bed.: 0 *kommt nicht vor*
- Buchstabenvariable: 1 *kommt vor*
  - Anzahl: 1 *genau eine*
  - Definition: 2 *korrekte Def. aller Var.*

Ikonische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

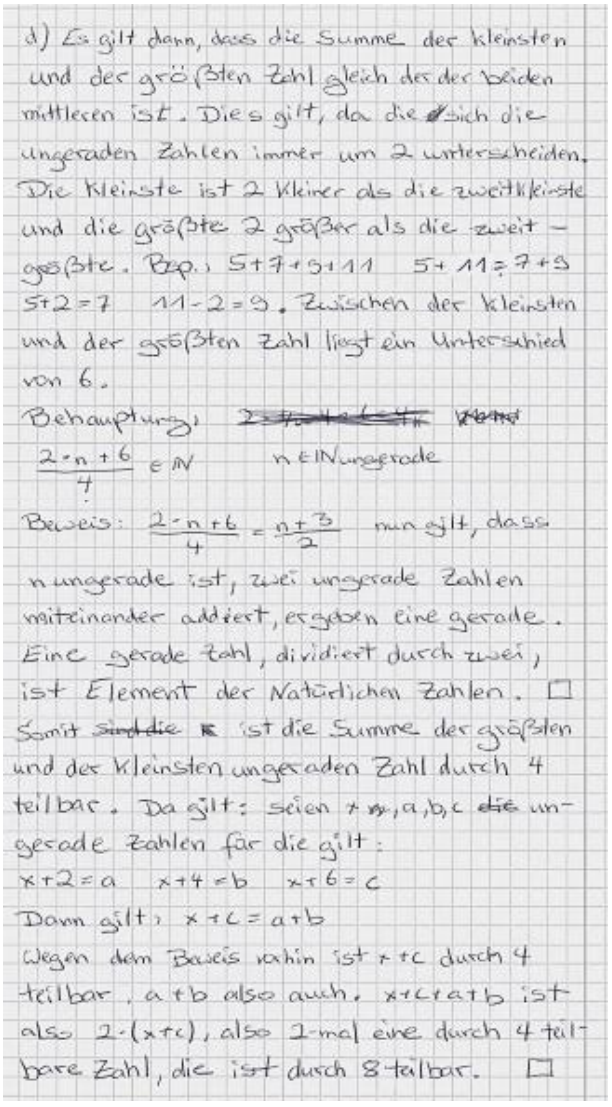
Numerische Darstellungsmittel
- Verwendung: 1 *trifft zu*

Verbale Darstellungsmittel
- Verwendung: 1 *trifft zu*

Strukturierende Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Darstellungsform
- Dstf14 *symbolisch-verbal*

Bearbeitung 3.3.11b: Klassenstufe 7, männlich (Code: 0711004)



Darstellungsmittel und Darstellungsform

Symbolische Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Symbol mit fester Bed.: 1 *kommt vor*
  - Art: ∈, ℕ
- Buchstabenvariable: 1 *kommt vor*
  - Anzahl: 2 *mehrere*
  - Definition: 2 *korrekte Def. aller Var.*

Ikonische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Numerische Darstellungsmittel
- Verwendung: 1 *trifft zu*

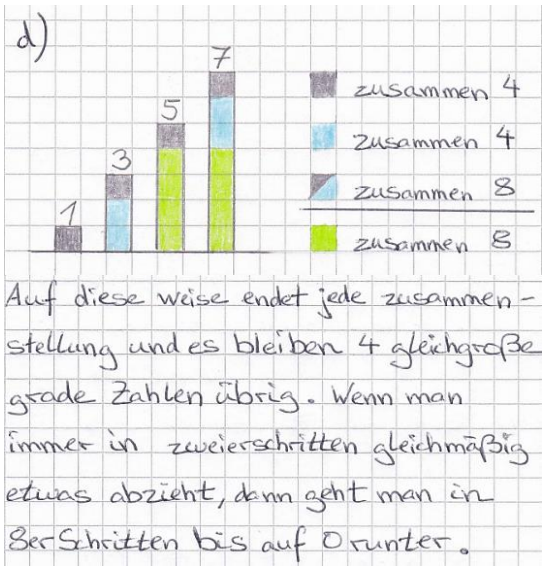
Verbale Darstellungsmittel
- Verwendung: 1 *trifft zu*

Strukturierende Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Math. Strukturierungselement: 1 *kommt vor*
- Strukturierender Begriff: 0 *kommt nicht vor*
- Nummerierung: 0 *kommt nicht vor*

Darstellungsform
- Dstf134 *symbolisch-numerisch-verbal*

Abbildung 67: Bearbeitungen 3.3.11a-b (Beweisversuche mit symbolischen Darstellungsmitteln: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform)

Bearbeitung 3.3.11c: Klassenstufe 7, weiblich (Code: 0711004)



Darstellungsmittel und Darstellungsform

Symbolische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Ikonische Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Figurierte Zahlen: 1 *kommen vor*
- Klammern, Pfeile, Linien: 1 *kommen vor*
- Einkreisungen, Unterstreichungen o. Ä.: 0 *kommen nicht vor*
- Häkchen, Kreuzchen: 0 *kommen nicht vor*
- Tabelle: 0 *kommt nicht vor*
- Sonstiges: 1 *kommt vor*

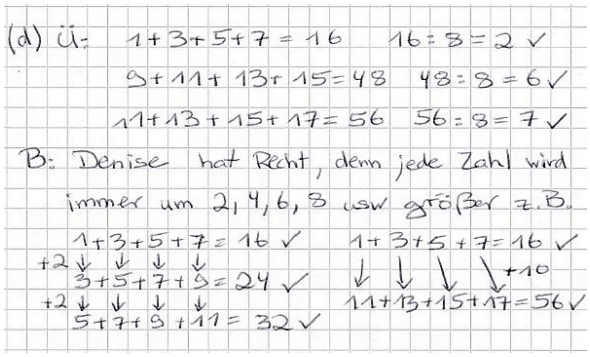
Numerische Darstellungsmittel
- Verwendung: 1 *trifft zu*

Verbale Darstellungsmittel
- Verwendung: 1 *trifft zu*

Strukturierende Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Darstellungsform
- Dstf24 *ikonisch-verbal*

Bearbeitung 3.3.11d: Klassenstufe 7, weiblich (Code: 0710112)



Strukturierende Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Mathematisches Strukturierungselement: 0 *kommt nicht vor*
- Strukturierender Begriff: 1 *kommt vor*
- Nummerierung: 0 *kommt nicht vor*

Darstellungsmittel und Darstellungsform

Symbolische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

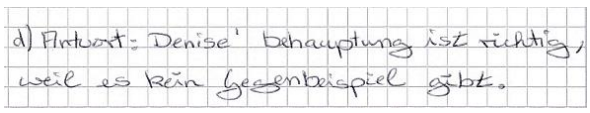
Ikonische Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Figurierte Zahlen: 0 *kommen nicht vor*
- Klammern, Pfeile, Linien: 1 *kommen vor*
- Einkreisungen, Unterstreichungen o. Ä.: 0 *kommen nicht vor*
- Häkchen, Kreuzchen: 1 *kommen vor*
- Tabellen: 0 *kommen nicht vor*
- Sonstiges: 0 *kommt vor*

Numerische Darstellungsmittel
- Verwendung: 1 *trifft zu*

Verbale Darstellungsmittel
- Verwendung: 1 *trifft zu*

Darstellungsform
- Dstf234 *ikonisch-numerisch-verbal*

Bearbeitung 3.3.11e: Klassenstufe 8, männlich (Code: 0811901)



Strukturierende Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Mathematisches Strukturierungselement: 0 *kommt nicht vor*
- Strukturierender Begriff: 1 *kommt vor*
- Nummerierung: 0 *kommt nicht vor*

Darstellungsmittel und Darstellungsform

Symbolische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Ikonische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Numerische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Verbale Darstellungsmittel
- Verwendung: 1 *trifft zu*

Darstellungsform
- Dstf4 *verbal*

Abbildung 68: Bearbeitungen 3.3.11c-d (Beweisversuche mit ikonischen Darstellungsmitteln: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform), Bearbeitung 3.3.11e (Beweisversuch ohne numerische Darstellungsmittel: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform)

Wenn in einem Beweisversuch ein symbolisches Darstellungsmittel vorkommt, werden ergänzend die Art der symbolischen Darstellungsmittel, die Korrektheit der Verwendung von Variablen und die Variablendefinitionen untersucht. Es wird erfasst, ob in dem Beweisversuch Symbole mit fester Bedeutung, wie etwa Quantoren, das zahlentheoretische Symbol für Kongruenz und das Zeichen für Teilbarkeit, vorkommen, die den Lernenden nicht unbedingt aus der Schule bekannt sind. Gegebenenfalls wird auch festgehalten, welche Symbole vorkommen. Danach liegt der Fokus auf Buchstabenvariablen (im Folgenden kurz „Variablen“). Außerdem wird codiert, ob der vorliegende Beweisversuch eine Variable enthält. Bei den Beweisversuchen mit einer Variablen wird ergänzend erfasst, ob genau eine oder mehrere verschiedene Variablen verwendet werden. Außerdem wird erfasst, inwieweit alle Variablen korrekt definiert sind. Ein Beweisversuch gehört diesbezüglich in die Kategorie DsymVarDef2 *korrekte Definition aller Variablen*, wenn der Gültigkeitsbereich aller Variablen in angemessener Form angegeben ist. Er gehört in die Kategorie DsymVarDef1 *Ansatz einer Variablendefinition*, wenn eine, aber nicht alle verwendeten Variablen definiert sind oder mindestens eine Definition angegeben, aber zu vage oder falsch ist. Er gehört in die Kategorie DsymVarDef0 *ohne Variablendefinition*, wenn keine der verwendeten Variablen definiert ist. Beispielsweise kommt in der Bearbeitung 3.3.11a in Abbildung 68 genau eine Variable vor, die über den Ausdruck „beliebige ungerade Zahl“ angemessen definiert ist. Symbole mit fester Bedeutung sind nicht zu sehen. In der Bearbeitung 3.3.11b in Abbildung 68 sind mehrere Variablen verwendet und alle korrekt definiert. Außerdem kommen in der Bearbeitung 3.3.11b das Elementzeichen und das Symbol für die Menge der natürlichen Zahlen vor.

Ikonische Darstellungsmittel

Als ikonische Darstellungsmittel gelten in dieser Arbeit figurierte Zahlen, Klammern, Pfeile, Linien, Einkreisungen, farbliche Markierungen, Tabellen, Diagramme etc., die aus Expertensicht im vorliegenden Beweisversuch eine inhaltliche Überlegung ausdrücken oder hervorheben.

Nicht als ikonische Darstellungsmittel gelten insbesondere …

- Darstellungsmittel, die im vorliegenden Beweisversuch einer Korrektur dienen (z. B. Durchstreichungen sowie Sternchen, die auf ergänzende Ausführungen verweisen),
- Darstellungsmittel wie Klammern, die im vorliegenden Beweisversuch Bestandteile eines symbolsprachlichen oder numerischen Ausdrucks sind,
- Darstellungsmittel wie Klammern, Linien und Beweisabschlusszeichen, die den vorliegenden Beweisversuch im Wesentlichen die formale Struktur der Ausführungen verdeutlichen.

Bezüglich ikonischer Darstellungsmittel wird zuerst erfasst, ob im vorliegenden Beweisversuch ein ikonisches Darstellungsmittel vorkommt. Beispielsweise kommen in den Bearbeitungen 3.3.11c-d ikonische Darstellungsmittel vor, in den Bearbeitung 3.3.11a-b nicht (Abb. 67-68). Wenn in einem vorliegenden Beweisversuch ein ikonisches Darstellungsmittel vorkommt, wird ergänzend dessen Art beschrieben. Dazu werden verschiedene Arten ikonischer Darstellungsmittel, wie z. B. figurierte Zahlen und Klammern, Pfeile, Linien, unterschieden. Bezüglich jeder der Arten wird eingetragen, ob sie im vorliegenden Beweisversuch zu finden sind (Bearbeitungen 3.3.11c-d in Abb. 68).

Numerische Darstellungsmittel

In dieser Arbeit gelten als numerische Darstellungsmittel sämtliche Zahlzeichen und Rechenzeichen, die im vorliegenden Beweisversuch nicht Teil eines symbolsprachlichen Ausdrucks sind.[79] Bei der Analyse wird ausschließlich erfasst, ob im vorliegenden Beweisversuch numerische Darstellungsmittel vorkommen oder nicht. Beispielsweise kommen in allen Bearbeitungen 3.3.11a-d numerische Darstellungsmittel vor, in der Bearbeitung 3.3.11e hingegen nicht (Abb. 67-68).

Verbale Darstellungsmittel

In dieser Arbeit gelten als verbale Darstellungsmittel alle Buchstaben, Wörter etc., die aus Expertensicht im vorliegenden Beweisversuch nicht ausschließlich als strukturierende Begriffe genutzt oder fester Bestandteil eines symbolsprachlichen oder numerischen Ausdrucks sind. Bei der Analyse wird ausschließlich erfasst, ob im vorliegenden Beweisversuch verbale Darstellungsmittel vorkommen oder nicht. Beispielsweise enthalten die Bearbeitungen 3.3.11a-e verbale Darstellungsmittel, die Bearbeitung 3.3.11f hingegen nicht (Abb. 67-69).

Strukturierende Darstellungsmittel

Im Rahmen dieser Arbeit gelten als strukturierende Darstellungsmittel alle Darstellungsmittel, die aus Expertensicht im vorliegenden Beweisversuch primär der Strukturierung der Ausführungen dienen. Das sind insbesondere

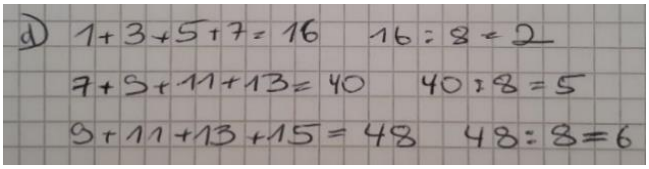
- in dieser Arbeit als mathematische Strukturierungselemente bezeichnete Angaben einer zu zeigenden Aussage oberhalb des eigentlichen Beweisversuchs sowie Kennzeichnungen des Beginns oder des Abschlusses eines Beweisversuchs,
- in dieser Arbeit als strukturierende Begriffe bezeichnete Ausdrücke, wie z. B. „Antwort:“, „Beispiel:“ und „Begründung:“, und entsprechende Abkürzungen, mit denen ein Abschnitt einer Bearbeitung gekennzeichnet ist,
- in dieser Arbeit als Nummerierungen bezeichnete mindestens zweischrittige Nummerierungen von Schlussfolgerungen, Lösungsschritten, Beispielen oder Ähnlichem,
- sonstige Auflistungen, etwa mithilfe von Aufzählungszeichen, beispielsweise von Fällen einer Fallunterscheidung,
- Klammern, Pfeile, Linien, Sternchen und farbliche Markierungen, die aus Expertensicht primär die Anordnung verschiedener Ausführungen in einer Bearbeitung verdeutlichen.

Bei der Analyse von Beweisversuchen wird als grundlegende Information erfasst, ob strukturierende Darstellungsmittel vorkommen. Beispielsweise kommen in der Bearbeitung 3.3.11g strukturierende Darstellungsmittel vor, in der Bearbeitung 3.3.11f hingegen nicht (Abb. 69).

[79] Hinweise: Wenn Zeichen wie das Additions- und das Gleichheitszeichen, die als Teil eines numerischen Ausdrucks vorkommen können, aus Expertensicht eher als Stenografien verwendet sind, werden diese dem Verbalen zugeordnet. Gleiches gilt bei Ausdrücken wie „4mal“ und „4er-Reihe“. Des Weiteren werden Nummerierungen nicht als numerische, sondern als strukturierende Darstellungsmittel erfasst.

Wenn in einem vorliegenden Beweisversuch strukturierende Darstellungsmittel vorkommen, wird ergänzend festgehalten, ob mathematische Strukturierungselemente zu erkennen sind, ob strukturierende Begriffe verwendet werden und ob eine Nummerierung erfolgt ist. In der Bearbeitung 3.3.11g in Abbildung 69 beispielsweise sind Lösungsschritte nummeriert. Das Beispiel ist mit der Abkürzung „Bsp." gekennzeichnet und der Abschluss des Beweisversuchs ist mit der Abkürzung „q. e. d." markiert. Folglich sind ein mathematisches Strukturierungselement, ein strukturierender Begriff und eine Nummerierung zu erkennen.

Bearbeitung 3.3.11f: Klassenstufe 8, weiblich (Code: 0810601)



Strukturierende Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Darstellungsmittel und Darstellungsform

Symbolische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

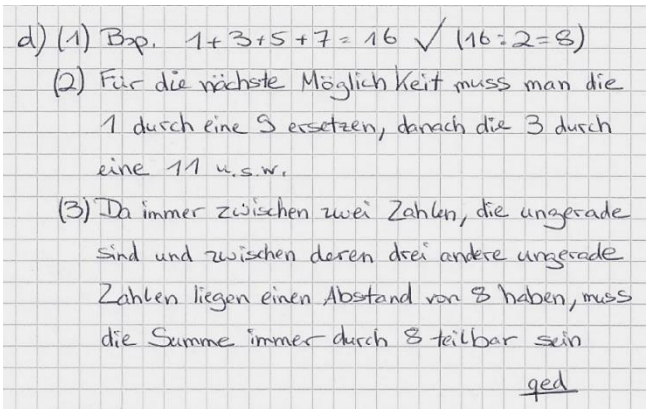
Ikonische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Numerische Darstellungsmittel
- Verwendung: 1 *trifft zu*

Verbale Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Darstellungsform
- Dstf3 *numerisch*

Bearbeitung 3.3.11g: Klassenstufe 8, männlich (Code: 0821910)



Strukturierende Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Mathematisches Strukturierungselement: 1 *kommt vor*
- Strukturierender Begriff: 1 *kommt vor*
- Nummerierung: 1 *kommt vor*

Darstellungsmittel und Darstellungsform

Symbolische Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Ikonische Darstellungsmittel
- Verwendung: 1 *trifft zu*
- Figurierte Zahlen: 0 *kommen nicht vor*
- Klammern, Pfeile, Linien: 0 *kommen nicht vor*
- Einkreisungen, Unterstreichungen o. Ä.: 1 *kommen vor*
- Häkchen, Kreuzchen: 1 *kommen vor*
- Tabellen: 0 *kommen nicht vor*
- Sonstiges: 0 *kommt vor*

Numerische Darstellungsmittel
- Verwendung: 1 *trifft zu*

Verbale Darstellungsmittel
- Verwendung: 0 *trifft nicht zu*

Darstellungsform
- Dstf34 *numerisch-verbal*

Abbildung 69: Bearbeitung 3.3.11f (Beweisversuch ohne verbale Darstellungsmittel: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform), Bearbeitung 3.3.11g (Beweisversuch mit verschiedenartigen strukturierenden Darstellungsmitteln: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform)

Darstellungsform

Aufbauend auf den separaten Analysen der unterschiedenen Darstellungsmittel wird die Darstellungsform, im Sinne der wesentlichen Form der Darstellung sachinhaltlicher Ausführungen, untersucht. Dabei sind die Kategorien Dstf1 *symbolisch*, Dstf2 *ikonisch*, Dstf3 *numerisch*, Dstf4 *verbal* und sämtliche Kombinationen derselben, wie z. B. Dstf14 *symbolisch-verbal*, so-

wie Beweisversuche mit einer nicht eindeutig kategorisierbaren Darstellungsform (Dstf55) möglich.[80]

In Beweisversuchen können verschiedene Arten an Darstellungsmitteln vorkommen. Um nun zu entscheiden, was die Darstellungsform ist, d. h. wie die sachinhaltlichen Überlegungen im Wesentlichen dargestellt sind, wird zunächst geprüft, mit welchen Arten an Darstellungsmitteln eine sinnbezogene Einheit formuliert wurde. Als sinnbezogene Einheiten werden dabei insbesondere Argumente, stützende oder konkretisierende Beispiele und Angaben eines Wahrheitswerts angesehen. Wenn alle sinnbezogenen Einheiten in der gleichen Form vorliegen, so handelt es sich um die Reinform der Darstellung (z. B. Dstf1 *symbolisch*). Kommen sinnbezogene Einheiten in unterschiedlichen Darstellungsformen vor, so handelt es sich um eine Mischform der Darstellung (z. B. Dstf14 *symbolisch-verbal*), die allerdings ein ganzes Spektrum an Darstellungsformen, z. B. von fast rein symbolisch bis hin zu verbal mit symbolischen Anteilen, abbildet.

Folglich gehört ein Beweisversuch beispielsweise in die Kategorie Dstf3 *numerisch*, wenn alle sinnbezogenen Einheiten in numerischer Darstellungsform vorliegen (Bearbeitung 3.3.11f in Abb. 69). Er gehört in die Kategorie Dstf24 *ikonisch-verbal*, wenn sowohl eine ikonisch als auch eine verbal dargestellte sinnbezogene Einheit vorkommen, aber keine symbolisch oder numerisch dargestellte Sinneinheit (Bearbeitung 3.3.11c in Abb. 68). Ein Beweisversuch gehört in die Kategorie Dstf134 *symbolisch-numerisch-verbal*, wenn in der Bearbeitung in der symbolischen, numerischen und verbalen Darstellungsform sinnbezogene Einheiten ausgedrückt sind (Bearbeitung 3.3.11b in Abb. 67). Ein Beweisversuch wird in die Kategorie Dstf55 *nicht eindeutig kategorisierbar* eingeordnet, wenn er keiner der anderen Kategorien eindeutig zugewiesen werden kann oder er durchgestrichen ist.

3.3.11.3 Hinweise zu den Kategorien

Um zu beschreiben, welche Darstellungsmittel in den Beweisversuchen der Lernenden vorkommen, werden – wie dargelegt – symbolische, ikonische, numerische, verbale und strukturierende Darstellungsmittel unterschieden. Diese sind so definiert, dass alle in einer Aufgabenbearbeitung vorkommenden Ausführungen genau einer dieser unterschiedenen Arten an Darstellungsmitteln zugeordnet werden können.

Bei der anschließenden Analyse der Darstellungsform und der Interpretation derer Ergebnisse ist zu bedenken, dass sich beispielweise hinter der Unterscheidung der Kategorien Dstf3 *numerisch*, Dstf34 *numerisch-verbal* und Dstf4 *verbal* ein Kontinuum von rein numerisch bis hin zu rein verbal dargestellten Beweisversuchen verbirgt. Etwa bei Beweisversuchen in den Kategorien Dstf3 *numerisch* und Dstf4 *verbal* liegen alle sinnbezogenen Einheiten in der gleichen Darstellungsform vor, während die Zwischenkategorien, wie z. B. Dstf34 *numerisch-verbal*, ein ganzes Spektrum an Beweisversuchen umfassen können. Für die Erfassung der Darstellungsform wird, wie ausgeführt, der Ausdruck „sinnbezogene Einheit“ verwendet, wohl wissend, dass dieser anders interpretiert werden kann.

[80] Da die sachinhaltlichen Überlegungen im Fokus stehen, werden die in einem vorliegenden Beweisversuch zu findenden symbolischen, ikonischen, numerischen und verbalen Darstellungsmittel, d. h. nicht die strukturierenden Darstellungsmittel berücksichtigt.

3.4 Der Beweisbegriff in der vorliegenden Arbeit

Nachdem die Begrifflichkeiten des Analysemodells eingeführt wurden (Abschn. 3.3), ist es möglich, präzise zu beschreiben, was unter dem Begriff „Beweis“ in der vorliegenden Arbeit verstanden wird. Im Folgenden wird der in dieser Arbeit verwendete Beweisbegriff ausgehend vom in der Mathematik geteilten Beweisbegriff, der Problematisierung des Beweisbegriffs aus mathematikdidaktischer Sicht, der Terminologie des Analysemodells und den in Kapitel 2 beschriebenen Charakteristika der Mathematik-Olympiade eingegrenzt.

Der in der Mathematik geteilte Beweisbegriff bildet den Ausgangspunkt für die Eingrenzung des Beweisbegriffs der vorliegenden Arbeit. Ein wichtiges Charakteristikum mathematischer Beweise ist demnach das deduktive Schließen. Deshalb werden in dieser Arbeit nur Beweisversuche mit begründendem Ansatz in einer der Kategorien *deduktiv ohne zusätzliches Beispiel* und *deduktiv mit zusätzlichem Beispiel* als Beweis angesehen.[81] Zusätzlich zum deduktiven Schließen muss das Argumentgefüge im Rahmen dieser Arbeit, so wie in der Wissenschaft Mathematik, korrekt und vollständig sein, damit es als Beweis bezeichnet wird. Dabei wird es jedoch nicht als Fehler angesehen, wenn Variablen ohne Definition verwendet werden. Zudem werden präformale Darstellungen akzeptiert sowie Rechtschreib-, Grammatik- und Zeichensetzungsfehler hingenommen (Abschn. 3.3.4). Die Vollständigkeit von Argumentgefügen wird im Verhältnis zu vorab festgelegten Erwartungshorizonten beurteilt (Abschn. 3.3.4). Ob ein Beweisversuch in dieser Arbeit als Beweis gilt, ist folglich von den formulierten Erwartungen und der konkreten Codierung des Beweisversuchs abhängig. So wie in der wissenschaftlichen Praxis erscheint auch hier eine gewisse Subjektivität als unumgänglich.

Die vorgenommene Eingrenzung des Beweisbegriffs auf korrekte und vollständige, deduktive Argumentgefüge mag gerade mit Blick auf junge Schülerinnen und Schüler anspruchsvoll wirken. Vor dem Hintergrund der Mathematik-Olympiade erscheint sie trotzdem als sinnvoll. Die Mathematik-Olympiade, die sich im Spannungsfeld zwischen schulischer und wissenschaftlicher Mathematik bewegt, hat zum Ziel, mathematisch besonders leistungsfähige Lernende zu finden und zu fördern und zur Gewinnung des akademischen Nachwuchses beizutragen (Kap. 2). Die in dieser Arbeit vorgenommene Einschränkung bezüglich der Definition von Variablen (Abschn. 3.3.4) wird in den fortgeschrittenen Klassenstufen und Wettbewerbsrunden, wie z. B. bei der Bundesrunde in der achten Klassenstufe, normalerweise nicht mehr toleriert. Zugleich bedeutet die getroffene Eingrenzung auf korrekte und vollständige, deduktive Beweisversuche keineswegs, dass sämtliche Beweisversuche, die nicht als Beweis bezeichnet werden, schlecht sind. Beweisversuche, bei denen die Schlussweise deduktiv mit induktivem Anteil oder induktiv mit Andeutung der Verallgemeinerbarkeit ist, und z. B. Beweisversuche, die ein weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke enthalten, können als Vorformen von Beweisen angesehen werden. Zweck des Analysemodells ist eine vergleichende Deskription von Beweisversuchen und dafür sind Abgrenzungen erforderlich.

[81] Da bei Beweisversuchen in der Kategorie *deduktiv mit zusätzlichem Beispiel* alle Beispiele weggelassen werden können, ohne die Struktur des Argumentgefüges oder die im Argumentgefüge dargelegten inhaltlichen Überlegungen zu beeinflussen (Abschn. 3.3.6), können Beweisversuche beider Kategorien als Beweis gelten.

Im Vergleich zu der in Abschnitt 3.2 vorgestellten Beweisdefinition von Stylianides (2007), die für die Arbeit mit Schülerinnen und Schülern gedacht ist, wird der Beweisbegriff im Rahmen dieser Arbeit enger gefasst. Anstelle der Klassengemeinschaft oder der Lehrkraft als entscheidender Instanz werden für die vergleichende Analyse von Beweisversuchen Erwartungshorizonte gesetzt, die sich am innermathematischen Beweisverständnis orientieren. Zugleich werden, passend zu der Definition von Stylianides (2007), Vorformen von Beweisen, die etwa hinsichtlich der Schlussweise abweichen, durch die Begrifflichkeiten des Analysemodells trotzdem wertgeschätzt. Zudem lässt die Definition, in Einklang mit der Definition von Stylianides (2007), hinsichtlich der Darstellungsmittel und auch hinsichtlich der Beweisideen Spielraum. Prinzipiell ist es nicht entscheidend, welche Beweisideen verwendet werden, deren Korrektheit und Tragfähigkeit vorausgesetzt. Außerdem werden im Rahmen dieser Arbeit keine Vorgaben hinsichtlich der verwendeten Darstellungsmittel getroffen. Ein z. B. rein verbal dargestellter Beweisversuch kann, anknüpfend an die Ausführungen in den Abschnitten 3.1.2 und 3.2, genauso ein Beweis sein, wie einer, in dem symbolische Darstellungsmittel vorkommen. Zudem wird die u. a. von Heintz (2000) und Wittmann (2009) angesprochene minimale Struktur von Beweisen nicht berücksichtigt, da diese weder entscheidend für die Gültigkeit eines Beweisversuchs als Beweis erscheint noch von jungen Teilnehmenden eines Klausurwettbewerbs zu erwarten ist (Abschn. 3.1.2, 3.3).

Darüber hinaus ist zu betonen, dass sich der Begriff „Beweis“ nicht nur auf Allaussagen, sondern auf Beweise beliebiger Arten mathematischer Aussagen bezieht. Er umfasst auch das Widerlegen, d. h. das Beweisen der Negation von Aussagen (Abschn. 3.1.1, 3.2). Des Weiteren kann eine Aussage, die bewiesen werden soll, sowohl einen unendlichen als auch einen endlichen Gültigkeitsbereich haben, und sowohl allgemeingültige Schlüsse als auch Überlegungen anhand konkreter Objekte erfordern. Das Argumentgefüge, das zum Beweisen einer Aussage notwendig ist, kann beliebig komplex sein, insbesondere können also auch einschrittige Argumentgefüge (bzw. Argumente) als Beweis gelten (Abschn. 3.3.2). Auch wenn in einer Aufgabe eine Lösung bestimmt und die Richtigkeit dieser Lösung gezeigt werden soll, wird dies als Beweisen angesehen (Abschn. 2.3, 3.3.2). Das, was gemeinhin mit Ausdrücken wie „Erklären von Lösungswegen“ beschrieben wird, kann demensprechend durchaus als Beweisen bezeichnet werden. Dazu ist es wichtig, dass nicht nur eine grobe Idee, sondern ein korrektes und gewissermaßen vollständiges, deduktives Argumentgefüge erwartet wird. Diese Sichtweise öffnet die Möglichkeit, den Beweisbegriff auch bei Aufgabenbearbeitungen sehr junger Lernender zu verwenden und diese als erste, elementare Formen von Beweisen wertzuschätzen.

Mit dem Beweisbegriff in dieser Arbeit sind also einerseits gewisse Ansprüche verbunden, die an den Beweisbegriff innerhalb der Mathematik anknüpfen. Andererseits werden das kleinere Repertoire an Darstellungsmitteln junger Schülerinnen und Schüler und das Setting der zeitlich begrenzten Wettbewerbsklausur bedacht sowie elementare Formen von Beweisen, wie einschrittige Argumentgefüge und rein zahlengebundene Überlegungen, in den Beweisbegriff dieser Arbeit eingeschlossen.

3.5 Zusammenfassung des dritten Kapitels

Annäherung an den Beweisbegriff

Mit dem Ausdruck „Beweisbegriff“ wird in dieser Arbeit das geteilte begriffliche Verständnis des Terminus „Beweis“ bezeichnet. Der **Beweisbegriff innerhalb der Mathematik** ist nicht so ganz klar und strikt, wie er vielleicht aus der Außensicht erscheinen mag. Trotz eines weitgehenden Konsenses innerhalb der mathematischen Community gibt es keine allgemeingültigen Kriterien dafür, was als Beweis akzeptiert wird. Beim Formulieren eines Beweises versuchen Mathematikerinnen und Mathematiker aufzuzeigen, dass und warum eine Aussage gilt. Dabei wird darauf geachtet, dass das Argumentgefüge deduktiv, korrekt, logisch stringent, gewissermaßen vollständig und präzise ist. Um eine möglichst große Exaktheit und Eindeutigkeit zu erreichen, werden üblicherweise eine präzise Fachsprache, Algebra und strukturierende Darstellungsmittel verwendet. Eine solche formale Abfassung ist aber nicht unbedingt notwendig. Entscheidend für die Akzeptanz eines Beweisversuchs ist die prinzipielle Möglichkeit zur Annäherung an einen formal-strengen Beweis. Ob ein Beweisversuch diesem Anspruch gerecht wird, ist von der Beurteilung desselben durch andere Expertinnen und Experten der mathematischen Community abhängig.

Der Beweisbegriff innerhalb der Mathematik bildet prinzipiell die Zielperspektive im Rahmen von **Beweisaktivitäten mit Schülerinnen und Schülern**. Allerdings erscheint es aus mathematikdidaktischer Sicht weder sinnvoll noch möglich, den Beweisbegriff eins zu eins bei der Arbeit mit Schülerinnen und Schülern zu übernehmen. Daher gibt es in der Mathematikdidaktik eine breite Auseinandersetzung mit dem Beweisbegriff, die sich unter anderem auf die axiomatische Fundierung, die zulässigen Schlussweisen, die Darstellungsmittel und die Anerkennung von Beweisversuchen als Beweis bezieht und noch nicht abgeschlossen ist. In der mathematikdidaktischen Literatur kommen dementsprechend unterschiedliche Definitionen des Beweisbegriffs vor.

Die verschiedenen mathematikdidaktischen Sichtweisen auf den Beweisbegriff sowie vorhandene Konzepte zur Beschreibung von Beweisversuchen aus der mathematikdidaktischen Literatur werden in dieser Arbeit für die Analyse von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade aufgegriffen. Als Beweisversuche werden dabei in schriftlicher Form vorliegende Bearbeitungen von Beweisaufgaben verstanden. Um ein großes Spektrum an **Beweisversuchen zu Allaussagen** aussagekräftig, präzise und systematisch differenziert beschreiben zu können, wurde ausgehend von der Fragestellung dieser Arbeit, mathematikdidaktischer Theorie, Ergebnissen empirischer Studien und Beweisversuchen von Lernenden ein **Analysemodell** entwickelt.

Nach diesem Analysemodell (Abb. 70) wird zwischen Beweisversuchen mit begründendem, prüfendem, weder begründendem noch prüfendem oder nicht eindeutig kategorisierbarem Ansatz sowie fehlenden Bearbeitungen unterschieden. Beweisversuche mit begründendem Ansatz werden bezüglich der Korrektheit und Vollständigkeit des Argumentgefüges, der verwendeten Beweisideen, der Schlussweise, der Darstellungsmittel und der Darstellungsform analysiert. Bei Beweisversuchen mit prüfendem Ansatz werden u. a. die Korrektheit und

Beweisideen sowie die Beispiele möglichst analog untersucht. Bei Beweisversuchen mit weder begründendem noch prüfendem Ansatz wird insbesondere zwischen verschiedenen Arten der Bearbeitung differenziert, wobei die Korrektheit der Beweisversuche, zugrundeliegende Ideen und die Schlussweise Berücksichtigung finden. Bei Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz werden, soweit möglich, die gleichen Eigenschaften analysiert wie bei denen, die einem bestimmten Ansatz zugeordnet werden können. Das Analysemodell ermöglicht auf diese Weise eine systematisch differenzierte Deskription von Beweisversuchen, die die Besonderheiten von Beweisversuchen mit begründendem, prüfendem, weder begründendem noch prüfendem sowie nicht eindeutig kategorisierbarem Ansatz berücksichtigt und zugleich Aussagen über alle zu analysierenden Beweisversuche ermöglicht.

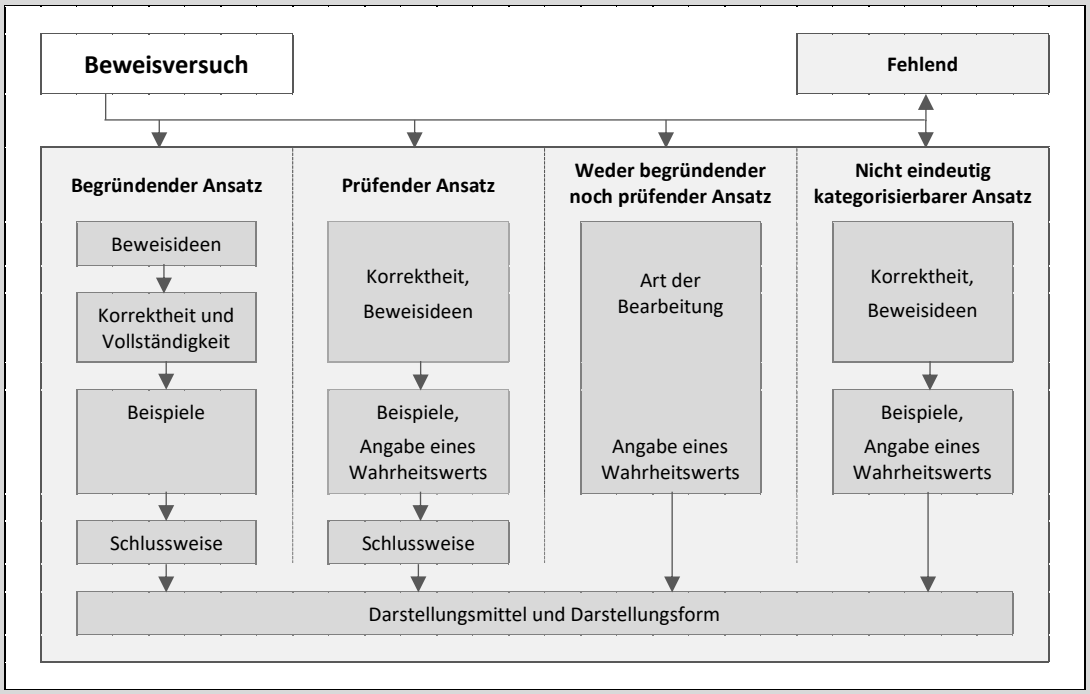


Abbildung 70: Grundstruktur des Modells zur Analyse von Beweisversuchen zu Allaussagen

Beweisversuche, in denen ein korrektes und vollständiges, deduktives Argumentgefüge vorkommt, werden in dieser Arbeit als Beweis einer infrage stehenden Allaussage bezeichnet. Diese Aussage kann sowohl einen unendlichen als auch einen endlichen Gültigkeitsbereich haben und sowohl allgemeingültige Schlüsse als auch Überlegungen anhand konkreter Objekte erfordern. Das Argumentgefüge, das zum Beweisen dieser Aussage notwendig ist, kann beliebig komplex sein, es können also auch einschrittige Argumentgefüge als Beweis gelten. Der **Beweisbegriff dieser Arbeit** erstreckt sich auf alle Arten mathematischer Aussagen und umfasst das Widerlegen von Aussagen. Auch wenn in einer Aufgabe eine Lösung bestimmt und die Richtigkeit dieser Lösung gezeigt werden soll, wird dies als Beweisen angesehen, insofern ein korrektes und gewissermaßen vollständiges, deduktives Argumentgefüge erwartet wird.

4 Beweiskompetenz: Begriffsannäherung und curriculare Vorgaben

In diesem Kapitel geht es um den Begriff „Beweiskompetenz“ und Kompetenzerwartungen an Schülerinnen und Schüler. Kompetenzen und der Kompetenzbegriff sind heutzutage in aller Munde. Im Mathematikunterricht sollen Schülerinnen und Schüler entsprechend den 2003 eingeführten Bildungsstandards allgemeine und inhaltsbezogene mathematische Kompetenzen aufbauen. Eine allgemeine mathematische Kompetenz lautet „Mathematisch argumentieren“ (KMK, 2003, S. 8). Unter diesem Oberbegriff wurden Aspekte, wie etwa das Führen von Beweisen und das Bewerten von Beweisversuchen, in die Bildungsstandards aufgenommen.

In Anlehnung an den Kompetenzbegriff wird in der vorliegenden Arbeit der Begriff „Beweiskompetenz“ verwendet. Was genau unter dem Begriff „Beweiskompetenz“ zu verstehen ist, ist jedoch nicht in den Bildungsstandards ausgeführt. Daher haben verschiedene Autorinnen und Autoren versucht, diesen Begriff präziser zu fassen. Daran anknüpfend wird auch für diese Arbeit zuerst eine Annäherung an den Begriff „Beweiskompetenz“ vorgenommen (Abschn. 4.1). Auf diesem Begriffsverständnis aufbauend werden curriculare Vorgaben zu der Teilkompetenz des Führens von Beweisen, die in dieser Arbeit im Fokus steht, vorgestellt und Erwartungen an Teilnehmende der Mathematik-Olympiade in den Klassenstufen fünf bis acht aufgezeigt (Abschn. 4.2). Diese Vorgaben und Erwartungen werden unter anderem bei der Interpretation und Diskussion der Ergebnisse der Analyse von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade berücksichtigt (Abschn. 9.3).

4.1 Begriff „Beweiskompetenz“

Ziel dieses Abschnitts ist es, den Begriff „Beweiskompetenz“ für den Rahmen dieser Arbeit zu beschreiben. Die Grundlage dafür bildet der **Kompetenzbegriff** an sich. Zu diesem Begriff wurden unterschiedliche Definitionen publiziert (z. B. Klieme & Leutner, 2006; Rychen & Salganik, 2003; Weinert, 2001). In der vorliegenden Arbeit wird die im deutschsprachigen Raum weit verbreitete Kompetenzdefinition von Franz E. Weinert (2001) verwendet. Weinert (2001, S. 27-28) definiert Kompetenzen als …

> […] die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.

Kompetenz meint in diesem Sinne stets Handlungskompetenz, die sich in herausfordernden Situationen zeigt. Für ein kompetentes Handeln ist es notwendig, bestimmte Probleme potentiell lösen zu können. Dies reicht allerdings nicht aus. Die kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, müssen dann auch eingesetzt werden. Diese Nutzung setzt neben den kognitiven Fähigkeiten und Fertigkeiten auch entsprechende motivationale,

volitionale und soziale Bereitschaften und Fähigkeiten voraus. Eine Person ist nach der Definition von Weinert (2001) kompetent, wenn die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll genutzt werden. Der Ausdruck „erfolgreich" verweist dabei auf die Qualität des Handlungsergebnisses und legt eine Differenzierung zwischen mehr oder weniger kompetentem Handeln nahe. Der Begriff „verantwortungsvoll" impliziert u. a. über die Folgen einer Handlung nachzudenken.[82]

Aufbauend auf diesem Kompetenzbegriff nach Weinert (2001) wird im Folgenden der Begriff „**Beweiskompetenz**" konkretisiert. Dabei lässt sich an Definitionen anderer Autorinnen und Autoren anknüpfen, wenngleich sich diese nicht alle auf Weinert (2001) beziehen. Beispielsweise schreibt Ufer (2008):

> In Anlehnung an den Kompetenzbegriff von Klieme und Leutner (2006) wird *geometrische Beweiskompetenz* verstanden als individuelle kontextspezifische Leistungsdisposition, die sich auf die Anforderungen von Beweisproblemen in der elementaren Geometrie bezieht. Explorative Arbeitsweisen, wie die Untersuchung von Vermutungen, sind hier nur insofern enthalten, wie sie für die Konstruktion eines Beweises von Bedeutung sind. (Ufer, 2008, S. 771, H. i. O.)

Eine ähnliche Eingrenzung trifft Reiss (o. J.): „Beweiskompetenzen sollen alle Fähigkeiten umfassen, die selbstständige Argumentationen und Beweise von Schülerinnen und Schülern verlangen" (Reiss, o. J., S. 13). In diesen beiden Definitionen wird das Führen von Beweisen ins Zentrum gestellt und der Begriff „Beweiskompetenz" auf die dafür erforderlichen kognitiven Kompetenzen begrenzt.[83]

Brunner (2013, 2014) rückt ebenfalls das Führen von Beweisen in den Mittelpunkt ihrer Begriffsbeschreibung. Das Anforderungsniveau könne hierbei entsprechend dem Ansatz von Reiss et al. (2006) vom Anwenden mathematischer Begriffe und Sätze bei einfachen Berechnungsproblemen, über die Formulierung von Argumenten bis hin zur Entwicklung mehrschrittiger Argumentgefüge reichen. Vor allem Letzteres verlange die Anwendung von Heurismen und metakognitiver Fähigkeiten. Zudem betont Brunner (2014), dass die Bewältigung der Anforderungssituationen egal welchen Niveaus eine Bereitschaft zur kognitiven Auseinandersetzung voraussetzt. Im Vergleich zu den Definitionen von Ufer (2008) und Reiss (o. J.) werden also die von Weinert (2001) angeführten „motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten" (Weinert, 2001, S. 28) berücksichtigt.

Grundey (2015) knüpft an die Kompetenzdefinition von Weinert (2001), die Ausführungen von Brunner (2014) und verschiedene curriculare Vorgaben, wie die Bildungsstandards der Kultusministerkonferenz (2012), an. Darauf aufbauend unterscheidet sie zwischen verschiedenen Teilkompetenzen des Konstrukts „Beweiskompetenz". Als Teilkompetenzen erwähnt sie das Führen von Beweisen, das Nachvollziehen von Beweisen sowie das Validieren und Bewerten von Beweisversuchen.

---

[82] Diese in der von Weinert (2001) für eine Vielzahl von Kontexten konzipierten Kompetenzdefinition enthaltene Forderung eines verantwortungsvollen Handelns, die beispielsweise bezogen auf Genforschung oder Kernphysik sehr plausibel und wichtig erscheint, ist mit Blick auf mathematische Kompetenzen von Lernenden kaum relevant und wird deshalb nicht erneut aufgegriffen.

[83] Dies passt zum Kompetenzbegriff von Klieme und Leutner (2006), der sich auf „auf *kognitive Dispositionen*" (Klieme & Leutner, 2006, S. 880, H. i. O.) beschränkt.

Diese Überlegungen von Ufer (2008), Reiss (o. J.), Brunner (2013, 2014) und Grundey (2015) können genutzt werden, um den Begriff „Beweiskompetenz“ aufbauend auf der Kompetenzdefinition nach Weinert (2001) zu spezifizieren und für den Rahmen dieser Arbeit präzise zu fassen.

Nach der Kompetenzdefinition von Weinert (2001) geht es darum, „bestimmte Probleme zu lösen“ (Weinert, 2001, S. 27). Fokussiert auf Beweiskompetenz handelt es sich dabei um Tätigkeiten, die in einem direkten Zusammenhang zu Beweisen stehen. Der Beweisbegriff wird dabei weit gefasst. Er bezieht sich auf beliebige mathematische Aussagen mit endlichem oder unendlichem Gültigkeitsbereich und umfasst sowohl ein- als auch mehrschrittige Argumentgefüge (Abschn. 3.4). Zu Tätigkeiten in einem direkten Zusammenhang zu Beweisen gehört es, solche Argumentgefüge zu entwickeln, d. h. Beweise zu führen. Dies kann sowohl eigenständig als auch in Gruppen erfolgen. Auch muss dabei nicht unbedingt der gesamte Prozess vom Verstehen einer infrage stehenden Aussage, über das Entwickeln einer Beweisidee bis hin zur Formulierung eines korrekten und vollständigen, deduktiven Argumentgefüges durchlaufen werden. Oftmals als Vorübungen zum Führen von Beweisen angesehene Aktivitäten, wie die Anordnung vorgegebener Argumente zu einem Beweis, die Vervollständigung mit Lücken versehener Beweisversuche und die Formulierung eines Beweises ausgehend von einem generischen Beispiel, werden ebenfalls als Tätigkeiten, die in einem direkten Zusammenhang zu Beweisen stehen, bezeichnet. Hinzu kommen Aktivitäten wie das Nachvollziehen, Validieren und Bewerten von Beweisversuchen. Auch das Wiedergeben und Erklären von Beweisen sowie die Formulierung von Fragen, die für Beweisaktivitäten charakteristisch sind, sind Tätigkeiten in einem direkten Zusammenhang zu Beweisen. Nicht dazu zählt hingegen ein reines Angeben von Lösungen zu mehr oder weniger problemhaltigen Aufgaben, da hier kein solcher Zusammenhang z. B. in Form der Explizierung logischer Schlüsse oder einer Auseinandersetzung mit Beweisversuchen vorhanden ist. Für das von Brunner (2013, 2014) als erstes Niveau von Beweiskompetenz bezeichnete „Anwenden mathematischer Begriffe und Sätze bei einfachen Berechnungsproblemen“ (Brunner, 2014, S. 80) ist demnach entsprechend der Terminologie dieser Arbeit keine Beweiskompetenz erforderlich. Gleiches gilt für das Prüfen und Anwenden einer wahren Allaussage, auch wenn solche Tätigkeiten durchaus die Entwicklung eines Beweises vorbereiten können. Zudem sind explorative Arbeitsweisen, wie auch das Formulieren von Vermutungen, entsprechend der Festlegung von Ufer (2008) nur insofern enthalten, wie sie für die Konstruktion eines Beweises oder z. B. die Anordnung vorgegebener Argumente zu einem Beweis Relevanz haben. An diese Eingrenzung der Tätigkeiten, die in einem direkten Zusammenhang zu Beweisen stehen, und den Ansatz von Grundey (2015) anknüpfend, werden im Rahmen dieser Arbeit das Führen von Beweisen, das Vervollständigen lückenhafter Beweisversuche, das Validieren und Bewerten von Beweisversuchen u. s. w. als Teilkompetenzen von Beweiskompetenz angesehen.

Um Tätigkeiten, die in einem direkten Zusammenhang zu Beweisen stehen, durchzuführen, sind nach Weinert (2001) gewisse „kognitive Fähigkeiten und Fertigkeiten“ (Weinert, 2001, S. 27) erforderlich. Dazu gehören auf jeden Fall elementare Fähigkeiten im Bereich des formalen logischen Denkens, ein passendes Bereichswissen und eine tragfähige Vorstellung davon, was einen mathematischen Beweis ausmacht (u. a. Brunner, 2013; Epp, 2003; Heinze & Reiss,

2003; Reiss et al., 2002). Weiterhin erscheinen angemessene Vorstellungen zur Entwicklung von Beweisen, eine gewisse Vertrautheit mit verschiedenen Beweisverfahren, ein sicherer Umgang mit Heurismen und metakognitive Kompetenzen als hilfreich (u. a. Bruder 1985; Brunner 2013; Kuntze, 2009). Zudem bedarf es einer gewissen Darstellungs- und Kommunikationskompetenz. Argumente sind nachzuvollziehen, zu hinterfragen, zu korrigieren und zu ergänzen, zu versprachlichen und zu verknüpfen. Dabei können verschiedene Darstellungsmittel zum Einsatz kommen (u. a. Bruder 1985; Brunner, 2013; Hoyles & Healy, 2007; Abschn. 3.3.11). Zusätzlich zu den kognitiven Dispositionen sind nach Weinert (2001) entsprechende „motivationale, volitionale und soziale Bereitschaften und Fähigkeiten" (Weinert, 2001, S. 28) erforderlich, um Tätigkeiten, die in einem direkten Zusammenhang zu Beweisen stehen, erfolgreich durchzuführen. Die genannten und weitere Voraussetzungen können je nach Anforderungssituation unterschiedlich wichtig sein. Beispielsweise sind angemessene Vorstellungen zum Beweisbegriff beim Bewerten von Beweisversuchen unbedingt notwendig und müssen je nach Aufgabenstellung auch versprachlicht werden können, während es beim Führen von Beweisen ausreichen kann, unterbewusst entsprechend solcher Vorstellungen zu handeln.

In der **vorliegenden Arbeit** steht die Teilkompetenz des Führens von Beweisen im Fokus. Anhand von Aufgabenbearbeitungen von Teilnehmenden der Mathematik-Olympiade werden nicht deren Beweiskompetenzen[84] in Gänze erfasst, sondern Ausprägungen der Teilkompetenz des Führens von Beweisen untersucht (Abschn. 1.1, 6.1). Bei der Mathematik-Olympiade wird in jeder Wettbewerbsrunde dazu aufgefordert, in einem beschränkten zeitlichen Rahmen mehrere Problemaufgaben eigenständig und schriftlich zu bearbeiten. Dabei sind vorgegebene Aussagen und selbstständig erarbeitete Lösungen zu beweisen (Abschn. 2.3, 10.3). Die Wettbewerbsteilnahme kann folglich als herausfordernde Situation angesehen werden, die insbesondere Einblicke in die Teilkompetenz des Führens von Beweisen von Schülerinnen und Schülern ermöglicht. Die Vielzahl an kognitiven Voraussetzungen, die beim Führen von Beweisen erforderlich sind, sowie entsprechende motivationale und volitionale Dispositionen fließen dabei mit ein, lassen sich aber nicht direkt analysieren.

Zusammenfassend betrachtet ist Beweiskompetenz folglich ein komplexes Konstrukt. Aufbauend auf der Kompetenzdefinition von Weinert (2001) werden im Rahmen dieser Arbeit unter dem Begriff „Beweiskompetenz" ein Bündel von bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten verstanden, um Tätigkeiten, die in einem direkten Zusammenhang zu Beweisen stehen, durchzuführen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die kognitiven Fähigkeiten und Fertigkeiten in variablen Situationen erfolgreich nutzen zu können. Dabei werden die verschiedenen Tätigkeiten, die in einem direkten Zusammenhang zu Beweisen stehen, wie z. B. das Führen von Beweisen, das Vervollständigen lückenhafter Beweisversuche sowie das Validieren und Bewerten von Beweisversuchen, als Teilkompetenzen von Beweiskompetenz angesehen. In der vorliegenden Arbeit wird die Teilkompetenz des Führens von Beweisen

[84] Mit dem Ausdruck „Beweiskompetenz" wird in dieser Arbeit das vollständige theoretische Konstrukt bezeichnet, das alle Teilkompetenzen umfasst. Die Mehrzahl „Beweiskompetenzen" wird für die Ausprägungen von Beweiskompetenz bei verschiedenen Lernenden verwendet.

fokussiert. Um Beweise zu führen sind eine Vielzahl kognitiver Voraussetzungen sowie entsprechende motivationale und volitionale Dispositionen erforderlich. Diese fließen folglich in die vorliegende Arbeit mit ein, werden aber nicht direkt untersucht.

4.2 Curriculare Vorgaben zu Beweiskompetenzen

Um herauszuarbeiten, was von Teilnehmenden der Klassenstufen 5 bis 8 der Mathematik-Olympiade in Nordrhein-Westfalen (NRW)[85] in Bezug auf das Führen von Beweisen erwartet werden kann, werden in diesem Abschnitt die diesbezüglichen curricularen Vorgaben in den Bildungsstandards und den Lehrplänen NRW analysiert.[86]

4.2.1 Vorgaben in den Bildungsstandards

In diesem Abschnitt werden zunächst die Bildungsstandards im Fach Mathematik für den Primarbereich hinsichtlich Beweiskompetenzen beleuchtet. Die hier ausgewiesenen Teilkompetenzen können von Lernenden aller Klassenstufen fünf bis acht erwartet werden. Danach werden Vorgaben aus den Bildungsstandards für den Mittleren Schulabschluss beschrieben, weil diese Hinweise auf mögliche Unterschiede zwischen Lernenden in den Klassenstufen 5 bis 8 geben.

4.2.1.1 Bildungsstandards im Fach Mathematik für den Primarbereich

In den Bildungsstandards im Fach Mathematik für den Primarbereich sind Kompetenzerwartungen an Lernende Ende der vierten Klassenstufe angegeben. Hier ist insbesondere die **allgemeine mathematische Kompetenz „Argumentieren“** (KMK, 2004, S. 8) relevant, die für Lernende Ende der vierten Klassenstufe mit folgenden Stichpunkten konkretisiert wird:

- mathematische Aussagen hinterfragen und auf Korrektheit prüfen,
- mathematische Zusammenhänge erkennen und Vermutungen entwickeln,
- Begründungen suchen und nachvollziehen. (KMK, 2004, S. 8)

Der erste Stichpunkt bezieht sich auf das Prüfen von Aussagen. Er deutet an, dass es Lernenden Ende der vierten Klassenstufe durchaus gelingen sollte, Allaussagen durch ein Gegenbeispiel zu widerlegen und in der Auseinandersetzung mit wahren Aussagen Beispiele anzugeben. Darüber hinaus kann der dritte Stichpunkt, das Suchen von Begründungen, der Teilkompetenz des

[85] Die Beweisversuche, die im Rahmen dieser Arbeit analysiert werden, wurden im Rahmen einer Mathematik-Olympiade in NRW erhoben (Kap. 6).

[86] In dieser Arbeit werden die Bildungsstandards und die Lehrpläne NRW fokussiert, da sie die mit Blick auf Teilnehmende der Mathematik-Olympiade in NRW relevantesten curricularen Vorgaben darstellen. In der internationalen mathematikdidaktischen Literatur werden stattdessen primär die Principles and Standards for School Mathematics, die vom National Council of Teachers of Mathematics (NCTM), der Vertretung von Mathematiklehrkräften in Schulen und Mathematikdidaktikerinnen und -didaktikern an amerikanischen Universitäten, nach einer breit geführten Diskussion im Jahr 2000 veröffentlicht wurden, berücksichtigt. Nach den Principles and Standards for School Mathematics sollen Lernende am Ende der Schulzeit u. a. mathematische Beweise entwickeln können. Dabei wird ein kontinuierlicher Aufbau einer mathematischen Argumentations- und Beweiskompetenz ab dem Anfang der Schulzeit empfohlen (NCTM, 2000). Diese Idee einer kontinuierlichen Förderung spiegelt sich in den von der Kultusministerkonferenz beschlossenen Bildungsstandards für die verschiedenen Klassenstufen und Schulformen, wie es die nachfolgenden Ausführungen auch andeuten.

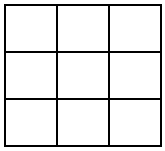
Führens von Beweisen zugeordnet werden. Was dabei mit dem Ausdruck „Begründungen“ gemeint ist, wird in den Bildungsstandards nicht explizit erläutert (KMK, 2004). Stattdessen umfassen die Standards mehrere Aufgabenbeispiele, um das Spektrum an Anforderungen, welches Schülerinnen und Schüler am Ende der vierten Klassenstufe bewältigen sollen, zu konkretisieren und zu veranschaulichen. Das Suchen von Begründungen findet in zwei Aufgabenbeispielen Berücksichtigung, die im Folgenden kurz vorgestellt werden.

8. Beispiel: Muster aus Streifen

[…]

1. Aufgabe:

Dieses Muster beginnt mit einem Dreierstreifen. Zeichne das 4. Bild. *(AB I)*

1. Bild	2. Bild	3. Bild	4. Bild
			

[…]

3. Aufgabe:

Jens zeichnet dieses Muster immer weiter. Kann es sein, dass er auf diese Weise ein fertiges Bild mit genau 125 Kästchen zeichnet? *(AB II)*

Ja ☐ Nein ☐

Begründung: _____

_____

(KMK, 2004, S. 23-24, H. i. O.)

Im dritten Aufgabenteil des Aufgabenbeispiels „Muster aus Streifen“ (KMK, 2004, S. 23) sollen die Lernenden angegeben und begründen, dass ein Bild des vorgegebenen Musters nicht aus genau 125 Kästchen bestehen kann. Bei der infrage stehenden Aussage handelt es sich um eine Allaussage mit unendlichem Gültigkeitsbereich. Diese lässt sich über die Invariante, dass die Anzahl an Kästchen eines jeden Bildes durch drei teilbar ist, beweisen. Außerdem ist es möglich, durch systematisches Probieren herauszufinden, dass das 41. Bild aus 123 und das 42. Bild aus 126 Kästchen besteht, und davon ausgehend die Ungültigkeit der Existenzaussage zu zeigen. Ein Beweisversuch mit prüfendem Ansatz und ohne latente Beweisidee würde dementsprechend von den Verantwortlichen vermutlich nicht als Begründung angesehen. Ob etwa Beweisversuche mit begründendem Ansatz der Art „125 ist nicht durch 3 teilbar. Also geht es nicht.“ ausreichen oder vollständigere Argumentgefüge erwartet werden, ist hingegen fraglich. Die zwei Zeilen, die in der Aufgabenstellung für die Antwort vorgegeben sind, suggerieren, dass solche Beweisversuche durchaus als „Begründung“ ausreichend sein können, die Formulierung eines vollständigeren Argumentgefüges nicht verlangt wird.

9. Beispiel: Hunderter-Tafel

[…]

5. Aufgabe:

Helene behauptet: „Die Summe von zwei Zahlen nebeneinander kann nie eine gerade Zahl sein." *(AB III)*

Stimmt das? Ja ☐ Nein ☐

Begründung: _____
_____
_____
_____
_____

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

6. Aufgabe:

Berechne jeweils die Summe der drei aufeinander folgenden Zahlen. *(AB III)*

4	5	6	Summe	
7	8	9	Summe	
14	15	16	Summe	

Vergleiche die Summe mit der mittleren Zahl. Was fällt dir auf? Begründe!

_____

(KMK, 2004, S. 25-26, H. i. O.)

Beim Aufgabenbeispiel „Hunderter-Tafel" (KMK, 2004, S. 25) werden die Lernenden nach mehreren Bestimmungsaufgaben, bei denen die Angabe der Lösung ausreicht, in den Aufgabenteilen fünf und sechs auch zum Begründen aufgefordert. Der fünfte Aufgabenteil fokussiert auf Zahlen, die in der Hunderter-Tafel nebeneinander stehen. Dabei ist nicht ganz klar, was mit dem Ausdruck „zwei Zahlen nebeneinander" (KMK, 2004, S. 25) gemeint ist. Wenn dieser Ausdruck Zahlen einschließt, die sich in der Hunderter-Tafel direkt untereinander befinden, dann ist die vorgegebene Allaussage falsch und ein Gegenbeispiel als Begründung ausreichend. Wenn der Ausdruck „zwei Zahlen nebeneinander" (KMK, 2004, S. 25) meint, dass sich zwei Zahlen in der gleichen Zeile der Hunderter-Tafel und direkt nebeneinander befinden, dann ist die vorgegebene Allaussage wahr und über die Invariante, dass stets eine ungerade und eine gerade Zahl addiert werden, begründbar.

Im sechsten Aufgabenteil geht es darum, dass die Summe dreier aufeinanderfolgender Zahlen dem Dreifachen der mittleren Zahl entspricht. Ausgehend vom Vergleich dreier Beispiele sollen die Lernenden diese Vermutung entwickeln und anschließend begründen. Hier lässt sich beispielsweise darüber argumentieren, dass die größte von drei aufeinanderfolgenden Zahlen immer um eins größer und die kleinste immer um eins kleiner ist als die mittlere dieser Zahlen. Ähnlich wie beim Aufgabenteil 3 des Aufgabenbeispiels „Muster aus Streifen" erscheint somit auch beim Aufgabenteil 6 des Aufgabenbeispiels „Hunderter-Tafel" insbesondere das Invarianzprinzip als zielführend. Vermutlich würde von den Verantwortlichen bei diesen Aufgabenteilen zumindest ein Beweisversuch mit prüfendem Ansatz und latenter Beweisidee oder ein Beweisversuch mit begründendem Ansatz, aber nicht unbedingt deduktiver Schlussweise ge-

fordert. Das notwendige Argumentgefüge hat bei beiden Aufgabenteilen geringe Komplexität. Inwieweit dieses vollständig ausformuliert werden soll, bleibt in den Bildungsstandards offen.

4.2.1.2 Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss

In den Bildungsstandards für den Mittleren Schulabschluss sind Beweiskompetenzen unter dem Oberbegriff **„Mathematisch argumentieren“** (KMK, 2003, S. 8) berücksichtigt. Im Vergleich zu den Bildungsstandards für den Primarbereich wurde die Kompetenzbezeichnung „Argumentieren“ (KMK, 2004, S. 8) durch den Zusatz „mathematisch“ ergänzt. Dieser verdeutlicht, „dass es sich beim mathematischen Argumentieren offensichtlich nicht um alltagsnahes Argumentieren handelt, sondern um mathematisches bzw. um spezifisch in der Mathematik lokalisiertes und durchgeführtes Argumentieren“ (Brunner, 2014, S. 31). In den Bildungsstandards für den Mittleren Schulabschluss wird folglich hervorgehoben, dass es bei der Kompetenz „Mathematisch argumentieren“ um eine besondere Form von Argumentationen und Argumentgefügen geht.

Die allgemeine mathematische Kompetenz „Mathematisch argumentieren“ wird durch folgende drei Stichpunkte konkretisiert:

- Fragen stellen, die für die Mathematik charakteristisch sind („Gibt es…?“, „Wie verändert sich …?“, „Ist das immer so …?“) und Vermutungen begründet äußern,
- Mathematische Argumentationen entwickeln (wie Erläuterungen, Begründungen, Beweise),
- Lösungswege beschreiben und begründen. (KMK, 2003, S. 8)

Diese allgemeine Beschreibung der Kompetenz „Mathematisch argumentieren“ wird in den Bildungsstandards für den Mittleren Schulabschluss zusätzlich in die drei Anforderungsbereiche (I) Reproduzieren, (II) Zusammenhänge herstellen und (III) Verallgemeinern und Reflektieren ausdifferenziert (Abb. 71). „Im Allgemeinen nehmen Anspruch und kognitive Komplexität vo[m] Anforderungsbereich [(I) bis zum Anforderungsbereich (III)] zu“ (KMK, 2003, S. 13).

Reproduzieren	Zusammenhänge herstellen	Verallgemeinern und Reflektieren
Mathematisch argumentieren Dazu gehört:		
- Routineargumentationen wiedergeben (wie Rechnungen, Verfahren, Herleitungen, Sätze, die aus dem Unterricht vertraut sind) - mit Alltagswissen argumentieren	- überschaubare mehrschrittige Argumentationen entwickeln - Lösungswege beschreiben und begründen - Ergebnisse bzgl. ihres Anwendungskontextes bewerten - Zusammenhänge, Ordnungen und Strukturen erläutern	- Komplexe Argumentationen erläutern oder entwickeln - Verschiedene Argumentationen bewerten - Fragen stellen, die für die Mathematik charakteristisch sind und Vermutungen begründet äußern

Abbildung 71: Ausdifferenzierung der Kompetenz „Mathematisch argumentieren“ in drei Anforderungsbereiche (KMK, 2003, S. 13-14, H. i. O.)

Das **Führen von Beweisen** wurde vor allem durch den zweiten Unterpunkt „mathematische Argumentationen entwickeln“ (KMK, 2003, S. 8) in die Bildungsstandards aufgenommen. Dabei deutet die Auflistung „Erläuterungen, Begründungen, Beweise“ (KMK, 2003, S. 8) an, dass das Spektrum an „Argumentationen“ (KMK, 2003, S. 8) weit gefasst ist. Es kann sich um Beweise, aber auch um gewisse Vorformen derselben handeln (Leiß & Blum, 2006; Reiss & Ufer, 2009). Sie können den Lernenden aus dem Unterricht mehr oder weniger bekannt sein und unter-

schiedliche Komplexität aufweisen (KMK, 2003; Abb. 71). Zur Veranschaulichung der Anforderungen an Lernende am Ende der Sekundarstufe I sind, so wie in den Bildungsstandards im Fach Mathematik für den Primarbereich, mehrere Aufgabenbeispiele angegeben.

> **(10) Fakultät**
>
> **Aufgabenstellung**
>
> Das Produkt aller natürlicher Zahlen von 1 bis n kann verkürzt aufgeschrieben werden:
>
> Beispiel: $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 5!$ (gesprochen: „fünf Fakultät“)
>
> Allgemein: $1 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots \cdot n = n!$ (gesprochen: „n Fakultät“)
>
> a) Berechnen Sie ohne Taschenrechner: $6!$
>
> b) Berechnen Sie mit Taschenrechner: $20!$
> Wie viele Stellen besitzt die errechnete Zahl?
>
> c) Auf wie viele Nullen endet die Zahl $20!$?
> Begründen und vergleichen Sie mit dem Taschenrechnerergebnis. (KMK, 2003, S. 29, H. i. O.)

Bei dem Aufgabenbeispiel „Fakultät“ (KMK, 2003, S. 29) wird der Aufgabenteil c) dem höchstens Anforderungsbereich der Kompetenz „Mathematisch argumentieren“ zugeordnet (KMK, 2003). Hier ist im Anschluss an einfache Bestimmungsaufgaben eine „komplexe Argumentation“ (KMK, 2003, S. 14) zu entwickeln. Die angegebene Lösungsskizze lautet:

> Die Nullen am Ende der Zahl ergeben sich durch die Zehnerfaktoren und durch die Faktorenpaare, die jeweils 10 ergeben: 10, 20, $2 \cdot 5$, $4 \cdot 15$. Die Zahl endet also mit vier Nullen, nach Taschenrechneranzeige mit acht. Die Anzeige ist also nicht genau. (KMK, 2003, S. 29)

So wie in diesem Aufgabenbeispiel werden auch in den anderen Aufgabenbeispielen „inhaltlich-anschauliche oder rechnerische Argumentationen, jedoch keine formalen Beweise“ (Wittmann, 2009, S. 48) gefordert. Die zu beweisenden Aussagen beziehen sich teilweise, so wie im obigen Beispiel, ausschließlich auf einzelne Zahlen, manchmal haben sie einen unendlichen Gültigkeitsbereich. Die Komplexität des erforderlichen Argumentgefüges ist dabei in der Regel gering. Außerdem befinden sich die Aufgabenteile, die „überschaubare mehrschrittige“ (KMK, 2003, S. 13) oder „komplexe“ (ebd.) Argumentgefüge erfordern, meistens am Ende der Aufgaben und bauen auf den vorhergehenden Aufgabenstellungen auf. Dies und die primäre Zuordnung des Entwickelns von Argumentgefügen zu den Anforderungsbereichen II bis III deutet an, dass es sich dabei um eine anspruchsvolle Aktivität handelt, die trotz der Bezeichnung als Regelstandard besonders von leistungsstarken Lernenden erwartet werden kann.

4.2.2 Vorgaben in Nordrhein-Westfalen

Die Vorgaben in den Bildungsstandards im Fach Mathematik werden in Nordrhein-Westfalen (NRW) durch den Lehrplan Mathematik für die Grundschule und die Kernlehrpläne für die Sekundarstufe I aufgegriffen und umgesetzt. In Abgrenzung und in Beziehung zu den Vorgaben in den Bildungsstandards werden im Folgenden diese curricularen Vorgaben für NRW beleuchtet. Dabei beschäftigt sich der erste Abschnitt mit den Kompetenzerwartungen, die an Lernende am Ende der Grundschulzeit gestellt werden. Der zweite Abschnitt fokussiert die Kompetenzerwartungen an Lernende in der Sekundarstufe I an Gymnasien.[87]

[87] Da die meisten Teilnehmenden der in dieser Arbeit vorgestellten Studie ein Gymnasium besuchen und die Vorgaben bezüglich des Führens von Beweisen in den Kernlehrplänen für Real- und für Gesamtschulen mit denen am

4.2.1.1 Lehrplan Mathematik für die Grundschulen des Landes NRW

Zu der prozessbezogenen Kompetenz „Argumentieren“ (MSW NRW, 2008b, S. 11) sind im Lehrplan Mathematik für die Grundschulen des Landes NRW folgende verbindlichen Erwartungen für alle Schülerinnen und Schüler am Ende der vierten Klassenstufe angegeben:

> Die Schülerinnen und Schüler
> - stellen Vermutungen über mathematische Zusammenhänge oder Auffälligkeiten an (vermuten)
> - testen Vermutungen anhand von Beispielen und hinterfragen, ob ihre Vermutungen, Lösungen, Aussagen etc. zutreffend sind (überprüfen)
> - bestätigen oder widerlegen ihre Vermutungen anhand von Beispielen und entwickeln – ausgehend von Beispielen – ansatzweise allgemeine Überlegungen oder vollziehen diese nach (folgern)
> - erklären Beziehungen und Gesetzmäßigkeiten an Beispielen und vollziehen Begründungen anderer nach (begründen) (MSW NRW, 2008b, S. 11)

Ein Blick in den Lehrplan Mathematik für die Grundschulen des Landes NRW zeigt ein relativ ähnliches Bild wie die Beschreibung der Bildungsstandards für den Primarbereich (Abschn. 4.2.1.1). Im Vergleich zu den Vorgaben in den Bildungsstandards für den Primarbereich wird im Lehrplan für NRW das Arbeiten mit Beispielen und Gegenbeispielen noch klarer hervorgehoben. Die Lernenden sollen Vermutungen anhand von Beispielen bestätigen und widerlegen. Zudem sollen sie ausgehend von Beispielen ansatzweise allgemeine Überlegungen entwickeln und Beziehungen und Gesetzmäßigkeiten an Beispielen erklären. Demzufolge sollen die Lernenden beim Beweisen wahrer Allaussagen ansatzweise über ein reines Prüfen oder die Angabe von Beispielen hinausgehen können, was zu den Aufgabenbeispielen in den Bildungsstandards passt (KMK, 2004; Abschn. 4.2.1.1).

4.2.1.2 Kernlehrplan für die Sekundarstufe I des Gymnasiums (G8) in NRW

Im Kernlehrplan für die Sekundarstufe I des Gymnasiums (G8) in Nordrhein-Westfalen sind Kompetenzerwartungen an Lernende am Ende der Klassenstufen sechs, acht und neun festgelegt. Erwartungen an Beweiskompetenzen werden dabei unter der prozessbezogenen Kompetenz „Argumentieren/Kommunizieren“ (MSW NRW, 2007, S. 18) aufgeführt. Mit Fokus auf Lernende am Ende der sechsten Klassenstufe heißt es kurz: „Schülerinnen und Schüler […] nutzen intuitiv verschiedene Arten des Begründens (Beschreiben von Beobachtungen, Plausibilitätsüberlegungen, Angeben von Beispielen oder Gegenbeispielen)“ (MSW NRW, 2007, S. 18). So wie im Lehrplan für die Grundschule des Landes NRW wird von Lernenden am Ende der sechsten Klassenstufe gefordert, dass sie Beispiele und Gegenbeispiele angeben können. Dabei fällt auf, dass der Begriff „Gegenbeispiel“ explizit vorkommt. Die Anforderung, ausgehend von Beispielen ansatzweise allgemeine Überlegungen zu entwickeln, ist hingegen nicht klar in den Vorgaben enthalten.

Gymnasium weitgehend übereinstimmen, wird nur auf den Kernlehrplan für Gymnasien Bezug genommen (MSW NRW, 2004a, 2004b, 2007).

Fokussiert auf Schülerinnen und Schüler am Ende der achten Klassenstufe steht im Kernlehrplan: „Schülerinnen und Schüler […] nutzen mathematisches Wissen für Begründungen, auch in mehrschrittigen Argumentationen“ (MSW NRW, 2007, S. 24). Am Ende der achten Klassenstufe sollen die Schülerinnen und Schüler also unter Verwendung ihres mathematischen Wissens mehrschrittige Argumentgefüge entwickeln und formulieren können (MSW NRW, 2007). Am Ende der Klassenstufe neun wird darüber hinausgehend erwartet, dass sie „mathematisches Wissen und mathematische Symbole für Begründungen und Argumentationsketten [nutzen]“ (MSW NRW, 2007, S. 29). Im Vergleich zu den Bildungsstandards wird folglich der Beweisbegriff nicht verwendet, aber von Lernenden am Ende der neunten Klassenstufe eine gewisse symbolische Darstellung von Argumentgefügen gefordert.

Was die Verwendung elementarer Algebra betrifft, lässt sich ergänzend festhalten, dass diese in den Kompetenzerwartungen an Schülerinnen und Schüler an Ende der sechsten Klassenstufe nicht vorkommt. Schülerinnen und Schüler am Ende der achten Klassenstufe sollen unter anderem Terme zusammenfassen und ausmultiplizieren sowie lineare Gleichungen lösen können.

4.2.3 Erwartungen an Teilnehmende der Mathematik-Olympiade

Ausgehend von den beschriebenen curricularen Vorgaben lassen sich Tendenzen formulieren, welche Kompetenzen im Führen von Beweisen von Teilnehmenden der Mathematik-Olympiade aus NRW am Anfang der Klassenstufen fünf bis acht zu erwarten sind.[88] Dabei wird angenommen, dass die Teilnehmenden der Mathematik-Olympiade in Mathematik zu den leistungsstärksten Schülerinnen und Schülern gehören.

Nach den Bildungsstandards für die Primarstufe und dem Lehrplan für Grundschulen des Landes NRW können leistungsstarke Schülerinnen und Schüler aller Klassenstufen fünf bis acht zu mathematischen Aussagen Beispiele oder Gegenbeispiele angeben. Es sollte ihnen gelingen, einfache spezielle Existenzaussagen zu beweisen und Allaussagen zu widerlegen. Weiterhin ist zu erwarten, dass sie in der Auseinandersetzung mit elementaren wahren Allaussagen Beispiele betrachten, Invarianten identifizieren und allgemeingültige Beweisideen skizzieren.

Mit Blick auf Lernende in den verschiedenen Klassenstufen fünf bis acht und insbesondere auf diejenigen ab der Klassenstufe sieben ist zu erwarten, dass sie ihre Beweisideen zunehmend deduktiv, detailliert und unter explizitem Bezug auf ihre mathematischen Kenntnisse formulieren. Besonders leistungsstarke Schülerinnen und Schüler in der Klassenstufe acht können unter Verwendung mathematischen Wissens mehrschrittige Argumentgefüge entwickeln und formulieren, die primär auf inhaltlich-anschaulichen oder rechnerischen Überlegungen basieren. Zum Teil verwenden sie dabei symbolische Darstellungsmittel, so wie es offiziell von Lernenden am Ende der neunten Klassenstufe erwartet wird.

[88] Die in dieser Arbeit vorgestellte Studie wurde im November 2016 mit Teilnehmenden der Mathematik-Olympiade aus den Klassenstufen fünf bis acht durchgeführt (Abschn. 6.2).

Einschränkend ist festzuhalten, dass die Vorgaben in den Bildungsstandards und auch die Vorgaben in den Lehrplänen des Landes NRW Interpretationsspielraum lassen, sodass auch mit Fokus auf leistungsstarke Lernende in der gleichen Klassenstufe heterogene Leistungen zu erwarten sind. Diese Überlegung stützend lassen es die Fallstudien von Girnat (2009) fraglich erscheinen, inwieweit die gemeinsamen Bildungsstandards in Bezug auf geometrisches Beweisen zu einem vergleichbaren Unterricht führen. Reiss und Ufer (2009) explizieren, dass die tatsächlichen Kompetenzen von Lernenden nicht mit den normativen Vorgaben der Curricula identisch sind. „Dennoch ist es sinnvoll, sich mit dieser normativen Ebene auseinanderzusetzen, denn nur so wird die Diskrepanz zwischen Anforderungen und dem Erfolg von Unterricht konkret" (Reiss & Ufer, 2009, S. 158-159). Im folgenden fünften Kapitel werden die normativen curricularen Vorgaben durch eher deskriptive Ergebnisse empirischer Studien zu Beweiskompetenzen von Schülerinnen und Schülern ergänzt. Die curricularen Vorgaben und die Forschungsergebnisse werden eine wichtige Hintergrundinformation bei der Konstruktion der Beweisaufgabe für die eigene Studie und der Einordnung der Ergebnisse dieser Studie bilden (Abschn. 6.2.2, 9.3).

4.3 Zusammenfassung des vierten Kapitels

Beweiskompetenz: Begriffsannäherung und curriculare Vorgaben

In der vorliegenden Arbeit wird der **Begriff „Beweiskompetenz"** verwendet. Die Grundlage dafür bildet die Kompetenzdefinition von Weinert (2001). Beweiskompetenz wird aufbauend auf dieser Kompetenzdefinition als spezielle Form von Kompetenz angesehen, die sich auf Tätigkeiten in direktem Zusammenhang zu Beweisen bezieht. Um solche Tätigkeiten durchzuführen, sind zahlreiche kognitive Dispositionen sowie „motivationale, volitionale und soziale Bereitschaften und Fähigkeiten" (Weinert, 2001, S. 28) notwendig. In dieser Arbeit steht das Führen von Beweisen als Teilkompetenz von Beweiskompetenz im Fokus.

Nach den **Vorgaben in den Bildungsstandards** und den **curricularen Vorgaben in NRW** lässt sich davon ausgehen, dass leistungsstarke Lernende in der fünften Klassenstufe das Angeben von Beispielen und Gegenbeispielen sowie erste Entwicklungen von Beweisideen bereits kennengelernt haben. Von Lernenden in höheren Klassenstufen sind zunehmend deduktive und zunehmend komplexe Argumentgefüge zu erwarten, wenngleich die tatsächlichen Beweiskompetenzen von Schülerinnen und Schülern nicht unbedingt mit den Vorgaben in den Bildungsstandards identisch sind.

5 Ergebnisse empirischer Studien zu Beweiskompetenzen von Schülerinnen und Schülern

Seit vielen Jahren werden Beweiskompetenzen von Schülerinnen und Schülern beforscht. Beispielsweise untersucht Walsch (1975) Beweiskompetenzen von Lernenden in der siebten Klassenstufe und Bell (1976) Beweiskompetenzen von vierzehn- bis fünfzehnjährigen britischen Lernenden. Zugleich sind Studien zu Beweiskompetenzen von Schülerinnen und Schülern weiterhin aktuell (u. a. Gerloff, 2021; Moya et al., 2015).

In diesem Kapitel werden Ergebnisse empirischer Studien zu Beweiskompetenzen von Schülerinnen und Schülern vorgestellt, um das Forschungsanliegen dieser Arbeit und die Ergebnisse der im Rahmen dieser Arbeit durchgeführten Studie einordnen zu können. Der Fokus liegt dabei auf Untersuchungen zur Teilkompetenz des Führens von Beweisen. Es werden insbesondere Ergebnisse aus Studien vorgestellt, in denen Lernende aufgefordert werden, alleine oder in Kleingruppen mathematische Allaussagen zu beweisen. Aus der Vielzahl solcher Untersuchungen werden primär diejenigen ausgewählt, die sich auf Schülerinnen und Schüler am Ende der Grundschulzeit oder in der Sekundarstufe I und auf schriftliche Aufgabenbearbeitungen beziehen, weil die Ergebnisse dieser Studien mit Blick auf die Fragestellung der vorliegenden Arbeit besonders relevant sind.

In den vorhandenen Studien werden teils unterschiedliche Begrifflichkeiten oder ähnliche Begriffe mit verschiedenen Konnotationen verwendet, die zudem nicht mit dem für die empirische Studie in dieser Arbeit entwickelten Begriffsnetz übereinstimmen. Um die Ergebnisse der empirischen Studien nicht zu verfälschen, werden die Termini der verschiedenen Forscherinnen und Forscher, wie etwa „generisches Beispiel“, „Beweis“ und „formal“, so weit wie möglich übernommen.[89]

Die nun folgende Präsentation von Ergebnissen empirischer Studien wird in mehrere Abschnitte gegliedert. Zuerst steht das Gelingen von Bearbeitungen von Aufgaben mit Beweisnotwendigkeit im Fokus (Abschn. 5.1). Anschließend geht es um Ergebnisse von Studien, die sich mit der Verwendung von Beispielen in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit beschäftigen (Abschn. 5.2). Danach werden Forschungsergebnisse vorgestellt, die Einblicke bezüglich der Verwendung von Darstellungsmitteln in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit geben (Abschn. 5.3). Die Ergebnisse werden innerhalb jedes Abschnitts 5.1 bis 5.3 resümiert und abschließend diskutiert (Abschn. 5.4).

[89] In Abschnitt 3.3 wurden die Ansätze und insbesondere Kategorisierungen mehrerer Forscherinnen und Forscher vorgestellt, sodass dort die Forschungsergebnisse ergänzende Informationen zu finden sind.

5.1 Gelingen von Beweisversuchen

Dieser Abschnitt beschäftigt sich mit der Frage, inwieweit es Schülerinnen und Schülern gelingt, Allaussagen mit unendlichem Gültigkeitsbereich zu beweisen. Es werden Ergebnisse empirischer Studien beschrieben, die Einblicke in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit geben, und in Ergänzung dazu Erkenntnisse vorgestellt, mit welchen Faktoren die Qualität solcher Aufgabenbearbeitungen zusammenhängt.

5.1.1 Bearbeitungen von Aufgaben mit Beweisnotwendigkeit

Die Sichtung von Studien zur Teilkompetenz des Führens von Beweisen verdeutlicht, dass die meisten Schülerinnen und Schüler dabei Schwierigkeiten haben. Insbesondere scheinen auch einem Teil der mathematisch leistungsstarken Lernenden Beweise von Allaussagen mit unendlichem Gültigkeitsbereich nicht zu gelingen. Dies wird im Folgenden anhand von Studien aufgezeigt. Diese Studien werden zuerst einzeln vorgestellt, weil die Anlage und verwendeten Begrifflichkeiten der Untersuchungen sehr unterschiedlich sind. Leserinnen und Leser, die ausschließlich die Kernpunkte interessieren, können direkt zum zweiten Teil dieses Abschnitts übergehen, in dem Erkenntnisse, die sich aus den Studien ziehen lassen, zusammengefasst sind.

5.1.1.1 Vorstellung einzelner Studien

Anfang der 60er Jahre lässt **Walsch (1975)** die Lernenden aus sechs Klassen zu Beginn und am Ende der Klassenstufe 7 eine geometrische Allaussage beweisen. Das notwendige Bereichswissen und die erforderliche Beweisfigur wurden jeweils vorab im Unterricht erarbeitet. Die Aufgabe der Lernenden besteht also primär darin, mehrere gegebene Aussagen zu einem korrekten Argumentgefüge zu verknüpfen. Dies schaffen sowohl zu Beginn als auch am Ende des Schuljahres etwa 12% aller Lernenden. Rund 15 Jahre später wiederholen **Flade und Pruzina (1990)** die Studie mit Lernenden am Ende der siebten Klassenstufe. Der Beweis gelingt rund 10% der Teilnehmenden. Dies zeigt, dass die Kompetenz des Führens von Beweisen vor über 45 Jahren und auch vor etwa 30 Jahren nicht so hoch entwickelt war, wenn es um das Finden und Verschriftlichen logischer Verknüpfungen geht.

Bell (1976) fordert 32 britische Schülerinnen und Schüler im Alter von vierzehn bis fünfzehn Jahren auf, Aufgaben mit Beweisnotwendigkeit eigenständig schriftlich zu bearbeiten. Bei einer Aufgabe geht es um eine zahlentheoretische Allaussage mit einem kleinen endlichen Gültigkeitsbereich, bei einer anderen um eine geometrische Allaussage mit unendlichem Gültigkeitsbereich. Bei beiden Aufgaben passen zwischen 35% und 50% der Bearbeitungen nicht ganz zu der infrage stehenden Aussage, was hervorhebt, dass Schülerinnen und Schüler vorgegebene Aufgaben nicht unbedingt korrekt interpretieren. Bei der Aufgabe, die sich auf einen endlichen Gültigkeitsbereich bezieht, sind in 19% der Bearbeitungen alle möglichen Einzelfälle korrekt überprüft, während keine Bearbeitung einen Beweis der Allaussage mit unendlichem Gültigkeitsbereich enthält. Dies legt die Vermutung nahe, dass es mehr Lernenden gelingt, eine Allaussage mit einem kleinen endlichen Gültigkeitsbereich zu beweisen als eine Allaussage mit unendlichem Gültigkeitsbereich zu zeigen.

Senk (1982) sowie **Senk und Usiskin (1983)** berichten von einem Forschungsprojekt im Bereich der Elementargeometrie. Im Rahmen dieses Projekts werden 1520 US-amerikanische Schülerinnen und Schüler zur eigenständigen, schriftlichen Bearbeitung mehrerer Beweisaufgaben unterschiedlichen Schwierigkeitsgrads aufgefordert. Die Teilnehmenden sind in den Klassenstufen 7 bis 12, mehrheitlich in der Klassenstufe 10 und haben im Unterricht das Beweisen elementargeometrischer Aussagen thematisiert. Senk (1982) hält fest:

> About 70% of the students can do simple proofs requiring only one deduction beyond those made from the given. […] Achievement is considerably lower on proofs requiring auxiliary lines or longer chains of reasoning. […] After a full year of a geometry course with proof, only about half the students can do any more than simple proofs. (Senk, 1982, S. 9)

Die Ergebnisse von Senk (1982) zeigen also, dass rund 30% der Schülerinnen und Schüler das Beweisen einfacher elementargeometrischer Aussagen nicht gelingt, auch wenn dieses im Schulunterricht thematisiert wurde. Zugleich scheinen nach der Behandlung des Beweisens im Schulunterricht etwa die Hälfte der Lernenden mehrschrittige Beweise führen können. Außerdem deuten die Ergebnisse darauf hin, dass der Anteil an Schülerinnen und Schülern, denen ein Beweis gelingt, u. a. vom Schwierigkeitsgrad der Aufgabe abhängt.

In einer Studie von **Balacheff (1988)** sollen 14 Paare französischer dreizehn- bis vierzehnjähriger Schülerinnen und Schüler eine Formel für die Anzahl an Diagonalen in einem n-Eck entwickeln, wobei die Aufgabenstellung nicht explizit zum Beweisen auffordert. Der Hälfte der Schülerpaare gelingt es, eine angemessene Formel zu entwickeln. Acht der Paare (ca. 57%) gehen im gesamten Bearbeitungsprozess nicht über die Betrachtung von Einzelfällen hinaus. Bei drei Paaren (ca. 21%) nutzt einer der Lernenden ein generisches Beispiel. Die übrigen drei Paare (ca. 21%) formulieren deduktive Argumente. Dabei verwendet ein Paar einen Spezialfall nicht im Sinne eines generischen Beispiels, sondern nur als Hilfe zur Versprachlichung der allgemeinen Überlegungen. Dies deutet an, dass viele Lernende ausschließlich empirisch-induktiv[90] argumentieren und manche Lernende Beispiele nutzen, um eine Beweisidee zu kommunizieren oder allgemeingültige Überlegungen zu formulieren.

Porteous (1990) untersucht Beweiskompetenzen von 50 britischen Lernenden im Alter von elf bis sechszehn Jahren. In einer schriftlichen Befragung sollen die Teilnehmenden angeben, inwieweit sie denken, dass die Summe dreier aufeinanderfolgender Zahlen immer durch drei teilbar ist und ihre Antwort möglichst vollständig erklären. Porteous (1990) schreibt: „The proportion of respondents who [...] provided logical support of their own for the statement was 10%. [...] there was a significant growth in the figure with age ($p < 0.01$)“ (Porteous, 1990, S. 591-592). Zusätzlich zu der schriftlichen Befragung sollen die Lernenden in Interviews entscheiden, ob verschiedene vorgegebene Allaussagen richtig sind. Porteous (1990) berichtet: „The true generalisations were usually accepted on empirical grounds, the proportion of subjects providing a proof being about 15%“ (Porteous, 1990, S. 595). Die Ergebnisse von Porteous (1990) stützen die Tendenz, dass viele Lernende ausschließlich empirisch-induktiv argumentieren. Zudem deuten die Ergebnisse an, dass Schülerinnen und Schüler höherer Klassen-

[90] Mit dem Terminus „empirisch-induktiv“ werden in der Regel Bearbeitungen bezeichnet, in denen Beispiele angeführt und eventuell von den Beispielen auf die Allgemeingültigkeit einer Allaussage geschlossen wird. Üblicherweise wird davon ausgegangen, dass die Beispiele zu der zu zeigenden Aussage passen und diese somit bestätigen.

stufen zu einem größeren Anteil versuchen, eine Begründung zu formulieren, als diejenigen zu Beginn der Sekundarstufe I.

In der Untersuchung von **Coe und Ruthven (1994)** sollen sechszehn- bis siebzehnjährige britische Schülerinnen und Schüler eines Colleges zu verschiedenen Sachverhalten selbstständig Vermutungen entwickeln. Bei einem Teil der Aufgaben werden sie zusätzlich explizit zum Beweisen aufgefordert. Bei beiden Aufgabenformaten wird in der Mehrheit der analysierten Bearbeitungen allenfalls empirisch-induktiv argumentiert.

Knuth et al. (2002) fordern ungefähr 350 US-amerikanische Schülerinnen und Schüler aus den Klassenstufen 6 bis 8 in einem schriftlichen Test auf, zu zeigen, dass die Summe zweier beliebiger aufeinanderfolgender Zahlen immer eine ungerade Zahl ist und dass die Summe zweier beliebiger gerader Zahlen immer gerade ist. Die Ergebnisse bestätigen die z. B. von Balacheff (1988), Porteous (1990) und Coe und Ruthven (1994) festgestellte Tendenz, dass viele Lernende empirisch-induktiv argumentieren, für die fokussierte Zielgruppe. Knuth et al. (2002) schreiben, dass 70% der Beweisversuche zu den genannten Allaussagen auf der Nutzung von Beispielen basieren, wobei eventuell generische Beispiele miteinbezogen sind.

Krumsdorf (2009a, 2009b, 2011, 2017) beschäftigt sich mit dem beispielgebundenen Beweisen (Abschn. 3.3.2). Die Ergebnisse heben hervor, dass es einigen Lernenden aus der vierten Klassenstufe in Einzelinterviews gelingt, einen beispielgebundenen Beweis zu führen (Krumsdorf, 2011, 2017). Zugleich zeigen die Ergebnisse, dass manche Gymnasiastinnen und Gymnasiasten in der fortgeschrittenen Sekundarstufe I ausschließlich empirisch-induktiv argumentieren oder die zu zeigende Aussage als Argument für deren Gültigkeit anführen (Krumsdorf, 2009a, 2017).

Moya et al. (2015) fordern in Spanien dreizehn durchschnittlich leistungsstarke und drei mathematisch talentierte Lernende[91] in der siebten Klassenstufe sowie vier mathematisch talentierte Lernende in der zehnten Klassenstufe auf, mehrere geometrische Aufgaben mit Beweisnotwendigkeit und unendlichem Gültigkeitsbereich eigenständig schriftlich zu bearbeiten.

Für die Teilnehmenden in der siebten Klassenstufe stellen Moya et al. (2015) heraus: Rund 58% der Aufgabenbearbeitungen der durchschnittlich leistungsstarken Lernenden gehören in eine der Kategorien[92] „no answer“ (Moya et al., 2015, S. 175) und „empirical failed“ (ebd.), weitere 30% in eine der Kategorien „naive empiricism“ (ebd.) und „crucial experiment“ (ebd.). Von den Aufgabenbearbeitungen der drei mathematisch talentierten Lernenden fällt keine in eine der Kategorien „no answer“ (Moya et al., 2015, S. 175) und „empirical failed“ (ebd.), rund 58% gehören in eine der Kategorien „naive empiricism“ (ebd.) und „crucial experiment“ (ebd.). Ein „generic example“ (Moya et al., 2015, S. 175) kommt sowohl bei den durchschnittlich leistungsstarken als auch bei den mathematisch talentierten Lernenden in einzelnen Aufgabenbearbeitungen vor. Eine korrekte „deductive justification“ (Marrades & Gutiérrez, 2000, S. 7) ist hingegen ausschließlich in einzelnen Aufgabenbearbeitungen der mathematisch talentierten Lernenden zu finden (Moya et al., 2015). Diese Ergebnisse lassen vermuten, dass mathematisch

[91] Als mathematisch talentierte Lernende bezeichnen die Verantwortlichen u. a. mathematisch begabte Lernende. Genaue Informationen diesbezüglich und zur Auswahl der Teilnehmenden kommen in der Publikation nicht vor.
[92] Die Kategorien werden von Moya et al. (2015) im Sinne von Marrades und Gutiérrez (2000) und teils angelehnt an Balacheff (1988) verwendet.

talentierten Lernenden Bearbeitungen von Aufgaben mit Beweisnotwendigkeit im Mittel ähnlich gut oder besser gelingen als anderen Lernenden aus der gleichen Klassenstufe.

Weiterhin sind nach Moya et al. (2015) die Aufgabenbearbeitungen der mathematisch talentierten Lernenden aus der zehnten Klassenstufe im Mittel besser als die der mathematisch talentierten Lernenden aus der siebten Klassenstufe. Beispielsweise enthalten bei einer Aufgabe zwei der vier Bearbeitungen aus der zehnten Klassenstufe ein generic example oder eine korrekte deductive justification, jedoch keine der Bearbeitungen aus der siebten Klassenstufe. Dabei sind allerdings die geringen Fallzahlen zu bedenken.

Im **britischen Forschungsprojekt „Justifying and Proving in School Mathematics"** (1995 – 1998) werden Beweiskompetenzen von rund 2500 vierzehn- bis fünfzehnjährigen, leistungsstarken Zehntklässlerinnen und Zehntklässlern untersucht (Healy & Hoyles, 2000; Hoyles & Healy, 2007).[93] In einem schriftlichen Test am Ende der zehnten Klassenstufe sollen sie eine aus dem Schulunterricht bekannte und eine unbekannte Allaussage beweisen:

> Prove that **when you add any 2 odd numbers, your answer is always even.** (Write down your answer in the way that would get you the best mark you can.) […]
>
> Prove that **if p and q are any two odd numbers, $(\mathbf{p}+\mathbf{q})\times(\mathbf{p}-\mathbf{q})$ is always a multiple of 4.** (Write your answer in the way that would get you the best mark you can.)
> (Healy & Hoyles, 2000, S. 404, H. i. O.)[94]

Von den Beweisversuchen zu der aus dem Schulunterricht bekannten Allaussage können rund 22% als Beweis angesehen werden, weitere 18% enthalten eine unvollständige deduktive Begründung (Kategorie „Partial proof, including all information needed but some reasoning omitted" (Healy & Hoyles, 2000, S. 408)). Bei der unbekannten Allaussage können rund 3% der Beweisversuche als Beweis angesehen werden, rund 6% der Beweisversuche enthalten eine unvollständige deduktive Begründung.[95] In einer landesweiten **Studie in Taiwan** werden rund 1100 Lernende der neunten Klassenstufe in einer schriftlichen Befragung aufgefordert, die beiden Allaussagen aus dem Forschungsprojekt von Healy und Hoyles (2000) zu beweisen (Chin & Lin, 2009). Die Teilnehmenden sind so wie die Zehntklässlerinnen und Zehntklässler im britischen Projekt vierzehn bis fünfzehn Jahre alt. In der Studie in Taiwan beträgt der Anteil an Beweisen bei beiden Aufgaben 6%. Bei der bekannten Allaussage werden weitere 31% als „partial proof" (Chin & Lin, 2009, S. 62) kategorisiert, bei der unbekannten Allaussage rund 14%. Die Verantwortlichen schreiben: „Compared with the UK, Taiwanese students performed much better in A4 [unfamiliar conjecture], and about the same in A3 [familiar conjecture]" (Chin & Lin, 2009, S. 61). In einer **Studie im Iran** werden 338 Lernende (high school second graders, ca. 15 bis 18 Jahre), die einen mathematisch-naturwissenschaftlichen Schwerpunkt gewählt haben, aufgefordert, die bekannte algebraische Allaussage aus dem britischen For-

[93] Die Teilnehmenden wurden anhand der Ergebnisse eines nationalen Mathematiktests, in dem mathematisches Beweisen nicht vorkam, ausgewählt und gehörten zu den besten 20% bis 25%.

[94] Zusätzlich zum Bekanntheitsgrad unterscheiden sich die zwei Aussagen u. a. auch hinsichtlich der vorkommenden Darstellungsmittel und der Komplexität des notwendigen Argumentgefüges. Die nachfolgend beschriebenen, unterschiedlichen Ergebnisse bei der Bearbeitung dieser Aufgaben können folglich nicht nur am Bekanntheitsgrad der Aussagen, sondern auch an anderen Faktoren liegen.

[95] Die erkennbaren Tendenzen zeigen sich im Rahmen der Untersuchungen in Großbritannien auch für geometrische Allaussage bei leicht anderen Erfolgszahlen (Healy & Hoyles, 2007).

schungsprojekt zu beweisen (Reyhani et al., 2011). Die Ergebnisse sind schwächer als in den anderen beiden Ländern. Insbesondere liegt der Anteil der Aufgabenbearbeitungen, die als Beweis bezeichnet werden können, bei der bekannten algebraischen Aussage unter 5%. Insgesamt betrachtet deuten die Ergebnisse der drei beschriebenen Studien an, dass die Teilkompetenz des Führens von Beweisen bei Lernenden aus verschiedenen Ländern unterschiedlich ausgeprägt ist. Dazu passend weisen auch die Ergebnisse von Bearbeitungen der geometrischen Beweisaufgabe K18 im Rahmen der **Third International Mathematics and Science Study** (TIMSS/III), die sich auf Schülerinnen und Schülern am Ende der Sekundarstufe II (secondary schooling) bezieht, auf Unterschiede zwischen verschiedenen Ländern hin (Baumert et al., 1999; Reid & Knipping, 2010).

Mit Fokus auf Deutschland und die Sekundarstufe I liegen recht aktuelle **Studien mit Gymnasiastinnen und Gymnasiasten ab der Klassenstufe 7** im Bereich der Geometrie vor (u. a. Reiss et al., 2002; Ufer, 2008). Die Teilnehmenden der Studien werden insbesondere aufgefordert, schriftliche Leistungstests zu bearbeiten. Die Tests umfassen Aufgaben, in denen die Anwendung von Begriffen und Regeln abgefragt wird, sowie Beweisaufgaben, die einschrittige oder mehrschrittige Argumentgefüge erfordern.

Reiss (o. J.) zeigt einzelne prototypische Bearbeitungen mehrschrittiger Beweisaufgaben. Sie veranschaulichen, dass manchen Lernenden Bearbeitungen solcher Aufgaben gelingen, andere z. B. Zirkelschlüsse formulieren, Prämissen aus der Anschauung entnehmen oder durchaus korrekte Überlegungen nicht angemessen ordnen und verknüpfen. Diese und weitere Arten der Bearbeitung, wie empirisch-induktive Ansätze, werden sowohl bei Untersuchungen mit Lernenden in den Klassenstufen 7 und 8 als auch in Studien mit älteren Schülerinnen und Schülern beobachtet (Reiss, o. J.; Reiss et al., 2006). Die Kompetenz des Führens von Beweisen scheint somit bei Schülerinnen und Schülern, insbesondere auch bei solchen in der gleichen Klassenstufe, unterschiedlich ausgeprägt zu sein.

Reiss et al. (2002) berichten über eine Studie, in der von 659 Lernenden am Ende der siebten Klassenstufe in einem schriftlichen Leistungstest im Bereich der Geometrie Daten erhoben werden. Die erhaltene Lösungswahrscheinlichkeit beträgt bezogen auf alle Items, die mehrschrittige Argumentgefüge erfordern, rund 24%. Dabei zeigen die Ergebnisse klare Unterschiede auf individueller Ebene. Während die Schülerinnen und Schüler des schwächsten Leistungsdrittels zum Teil allenfalls ein paar Aufgaben zum Basiswissen korrekt bearbeiten, können die leistungsstärksten Lernenden die meisten Aufgaben aller drei Niveaustufen lösen. Andersherum betrachtet weisen die Ergebnisse darauf hin, dass auch viele Lernende des leistungsstärksten Drittels Schwierigkeiten beim Führen von Beweisen haben. Die Lösungsrate über alle Items, die mehrschrittige Argumentgefüge erfordern, beträgt bei ihnen 50%.

Wenige Jahre nach der von Reiss et al. (2002) beschriebenen Untersuchung wird erneut ein Leistungstest mit etwa 700 Gymnasiastinnen und Gymnasiasten am Ende der Klassenstufe 7 durchgeführt (Reiss et al., 2006). Die Ergebnisse zeigen ein ähnliches Bild wie die der Untersuchung von Reiss et al. (2002). In Fortführung der Studie von Reiss et al. (2006) wird mittels schriftlicher Tests die Entwicklung geometrischer Beweiskompetenzen von Lernenden in sieb-

ten bis neunten Klassen untersucht (Ufer, 2008). Die Daten zeigen die Tendenz, dass der relative Anteil an Schülerinnen und Schülern, die mehrschrittige Beweisaufgaben lösen können, im Verlauf der Sekundarstufe I steigt. Dabei deutet sich aber an, dass dieser Zuwachs „mehr auf eine Entwicklung des geometrischen Basiswissens zurückzuführen ist als auf eine prinzipiell bessere Fähigkeit, mehrschrittige Beweise erfolgreich zu bearbeiten“ (Ufer, 2008, S. 774).

Küchemann und Hoyles (2002) berichten von einer Längsschnittstudie mit leistungsstarken Lernenden in den Klassenstufen 8 bis 10 in England. Im Rahmen eines schriftlichen Tests am Ende der achten, neunten und zehnten Klassenstufe sollen die Probanden entscheiden, ob die Aussage „If the PRODUCT of two whole numbers is ODD, their SUM is EVEN“ (Küchemann & Hoyles, 2002, S. 199, H. i. O.) wahr ist und ihre Antwort erklären. Der Vergleich der Daten von 1100 Schülerinnen und Schülern aus den Klassenstufen 8 und 9 lässt kleine, aber statistisch signifikante Verbesserungen erkennen. Beispielsweise ist in der neunten Klassenstufen in 14% im Vergleich zu 10% der Bearbeitung die wahre Implikation als wahr beurteilt und angemessen deduktiv begründet. Der ergänzende intraindividuelle Vergleich der Aufgabenbearbeitungen in den beiden Schuljahren zeigt, dass diese in 22% der Fälle in der Klassenstufe 9 besser sind als im Jahr zuvor, in 15% der Fälle schlechter. Ein Vergleich der Ergebnisse aus der Klassenstufe 9 mit denen des Tests in der zehnten Klassenstufe zeigt ähnliche Tendenzen. Insgesamt kommen die Verantwortlichen, vor dem Hintergrund der Bearbeitungen des genannten und weiterer Aufgabenteile, zu folgendem Schluss: „Thus our study provides some but rather limited support to the assumption that students move from empirical to more deductive approaches with increasing maturity“ (Küchemann & Hoyles, 2002, S. 217).

Die Studie von Küchemann und Hoyles (2002) sowie die meisten anderen bisher beschriebenen Untersuchungen fokussieren auf Schülerinnen und Schüler frühestens ab der siebten Klassenstufe. Demgegenüber beschäftigen sich insbesondere Neumann et al. (2014), Brunner et al. (2022), Gerloff (2021) und Vogt (2021) gezielt mit jüngeren Schülerinnen und Schülern.[96]

Neumann et al. (2014) fordern 243 Schülerinnen und Schüler aus den Klassenstufen 3, 4, 6 auf, mehrere arithmetische Aufgaben eigenständig schriftlich zu bearbeiten. In jeder der Aufgaben sollen die Teilnehmenden Aufgaben, wie z. B. Additionspäckchen, berechnen, daran Regelmäßigkeiten erkennen, eigene dazu passende Aufgaben angeben, die erkannte Regelmäßigkeit beschreiben und abschließend begründen. Die Analyse der Aufgabenbearbeitungen zeigte keine statistisch signifikanten Unterschiede zwischen den Klassenstufen. Vielmehr deuten die Ergebnisse auf eine „Homogenisierung der gezeigten Leistungen im Verlauf der Schulalter“ (Neumann et al., 2014) hin. Dies ergänzend schreiben die Autorinnen:

> Im unteren Leistungsbereich ist eine Abnahme der Anzahl der SchülerInnen bei gleichzeitiger Zunahme der Begründungskompetenz über die Klassenstufen hinweg festzustellen. […] Erstaunlich ist, dass im Hochleistungsbereich keine entsprechende Zunahme über die Jahrgangsstufen erkennbar ist. (Neumann et al., 2014, S. 122)

[96] Den genannten Untersuchungen von Neumann et al. (2014), Brunner et al. (2022), Gerloff (2021) und Vogt (2021) ist gemein, dass sie auf Schülerinnen und Schüler aus den Klassenstufen 3 bis 6 und auf eigenständige schriftliche Bearbeitungen von Aufgaben mit Beweisnotwendigkeit fokussierten. Daher sind sie mit Blick auf die Fragestellung der vorliegenden Arbeit relevant und werden im Folgenden näher beschrieben. Zusätzlich gibt es zahlreiche Untersuchungen mit Lernenden in den Klassenstufen 3 bis 6, in denen es primär um mündliches Argumentieren, Begründen und Beweisen geht (z. B. Fetzer, 2011; Hartkens, 2018; Krummheuer, 2001; Schwarzkopf, 2000; Steinbring, 2009). Für nähere Informationen zu diesen Studien wird auf die angegebene Literatur verwiesen.

Im Vergleich zu der Untersuchung von Neumann et al. (2014) kommen **Brunner et al. (2022)** in ihrer Quasi-Längsschnittstudie mit 866 Schweizer Schülerinnen und Schülern aus den Klassenstufen 4 bis 6 zu etwas anderen Ergebnissen. Die Teilnehmenden dieser Studie sollen eine Aufgabe aus dem Bereich „Arithmetik“ und eine aus dem Bereich „Geometrie“ bearbeiten. Die Aufgaben sind jeweils in vier Aufgabenteile gegliedert, die nacheinander die vier von Brunner et al. (2022) unterschiedenen Teilkompetenzen mathematischen Argumentierens, „Erkennen, Beschreiben, Begründen, Verallgemeinern“ (Brunner et al., 2002, S. 474), fokussieren. In der Aufgabe aus dem Bereich „Arithmetik“ beispielsweise geht es um die Summe von drei aufeinanderfolgenden (natürlichen) Zahlen. Es sind mehrere Beispiele („$1 + 2 + 3 =$ ____, $2 + 3 + 4 =$ ____, $3 + 4 + 5 =$ ____ […]“ (Brunner et al., 2022, S. 475)) vorgegeben, die im ersten Aufgabenteil berechnet und durch weitere Beispiele ergänzt werden sollen. Im zweiten Aufgabenteil soll ein anhand der Beispiele erkennbares Muster beschrieben werden. Der dritte Aufgabenteil zielt dann auf eine Begründung („Warum entsteht aus den drei Pluszahlen *genau dieses* Ergebnis? Erkläre, zeichne oder beschreibe!“ (Brunner et al., 2022, S. 475, H. i. O.)) und der vierte auf eine Verallgemeinerung („Wie muss deine Erklärung heißen, damit sie *immer* funktioniert, also für irgendein Beispiel“ (Brunner et al., 2022, S. 475, H. i. O.)).

Bei der Bearbeitung der Aufgabenteile ist es möglich, unterschiedliche inhaltliche Ideen zu explizieren. Brunner et al. (2022) unterscheiden u. a. die Ideen „Anhängen“ (Brunner et al., 2022, S. 479) und „Umstrukturieren“ (ebd.). Bezogen auf die Aufgabe aus dem Bereich „Arithmetik“ wird beim Anhängen „jeweils der erste Summand weggelassen und durch die nächste Zahl in der Zahlenreihe ersetzt“ (Brunner et al., 2022, S. 479) und beim Umstrukturieren erkannt, dass „es sich […] um eine Verdreifachung des mittleren Summanden handelt“ (ebd.). Die Analyseergebnisse zeigen, dass alle von Brunner et al. (2022) unterschiedenen Ideen in den Aufgabenbearbeitungen vorkommen. In allen Klassenstufen ist die Idee „Anhängen“ eher häufig, die Idee „Umstrukturierung“ hingegen selten. Zugleich zeigen sich hinsichtlich der vorkommenden inhaltlichen Ideen statistisch signifikante Unterschiede zwischen den einzelnen Klassenstufen vier bis sechs. Außerdem sind mit Blick auf die vorkommenden inhaltlichen Ideen Unterschiede zwischen der Aufgabe aus dem Bereich „Arithmetik“ und der Aufgabe aus dem Bereich „Geometrie“ zu erkennen.

Zusätzlich zu den inhaltlichen Ideen untersuchen Brunner et al. (2022) die Existenz einer inhaltlichen, mathematischen Bearbeitung der einzelnen Aufgabenteile und ggf. deren Ausführungsqualität. Bei beiden Aufgaben ist in ca. der Hälfte der vorhandenen Bearbeitungen des dritten Aufgabenteils zum Begründen eine zumindest teilweise korrekte Begründung formuliert. Dabei deuten die Ergebnisse zu diesem Aufgabenteil eine Leistungszunahme von der vierten zur sechsten Klassenstufe an, wenngleich die Unterschiede zur fünften Klassenstufe nicht alle statistisch signifikant sind. Im Vergleich zum Aufgabenteil zum Begründen sind zum abschließenden Aufgabenteil zum Verallgemeinern insbesondere im Fall der Geometrieaufgabe wenige Bearbeitungen vorhanden. Zudem sind die vorhandenen Bearbeitungen seltener korrekt und vollständig als die zum Aufgabenteil zum Begründen.

Gerloff (2021) untersucht die schriftlichen Bearbeitungen eines Sets geometrischer Aufgaben von 117 Lernenden am Ende der dritten Klassenstufe und von 121 Lernenden am Ende der vierten Klassenstufe. Die Aufgaben sind jeweils hälftig den Bereichen „Muster und Strukturen“ sowie „Raumvorstellungen“ zuzuordnen und unterscheiden sich hinsichtlich der Explizität der Begründungsaufforderung. Beispielsweise sind in einem Aufgabenteil aus dem Bereich „Muster und Strukturen“ die ersten drei Glieder der Folge der Dreieckszahlen ikonisch dargestellt. In der impliziten Variante wird ausschließlich gefragt, wie die Folge weiter geht, während in der expliziten Variante der Imperativ „Begründe:“ (Gerloff, 2021, S. 208) explizit zum Begründen auffordert. Jede Aufgabenbearbeitung ordnet Gerloff (2021) genau einer der folgenden Niveaustufen zu:

> Stufe 0) kein passender Zugang […]. Stufe 1) Lösung (Aussagenfindung) […]. Stufe 2) Beschreibung als deklarierte Begründung […]. Stufe 3) Konkretes Begründen (Einzelfallebene) […]. Stufe 4) Abstraktes Begründen (verallgemeinerte Ebene) […]. Stufe 5) Konkretes und abstraktes Begründen in logischer Schlussfolge (beide Ebenen) […]. (Gerloff, 2021, S. 251-259)

Die Ergebnisse zeigen, dass alle unterschiedenen Niveaustufen vorkommen und zwar in Bezug auf beide Klassenstufen drei und vier. Dabei weisen die relativen Häufigkeitsverteilungen auf Unterschiede zwischen den Klassenstufen hin. Von den Bearbeitungen aus der dritten Klassenstufe gehören rund 47%, von denen aus der vierten Klassenstufe rund 32% zur Stufe 0. Diese „umfasst fehlende, vollständig fehlerhafte und unpassende Antworten […], die inhaltlich nicht falsch sind, aber keinen passenden Bezug zu der Fragestellung aufweisen“ (Gerloff, 2021, S. 251). Zudem ist der Anteil der Bearbeitungen auf den Stufen 3 bis 5, die von Gerloff (2021) alle als „Begründungen“ (Gerloff, 2021, S. 250) bezeichnet werden, in der vierten Klassenstufe statistisch hoch signifikant größer als in der dritten Klassenstufe bei geringer Effektstärke.

Die Begründungen für sich betrachtet, fällt auf, dass in beiden Klassenstufen 3, 4 die Mehrheit der Bearbeitungen zu der Stufe 4 gehören. „Die Kinder begründen also, trotz des anspruchsvollen Abstrahierens und der dafür notwendigen Herstellung von Zusammenhängen zwischen mehreren Fällen, vorzugsweise bereits in einer für mehrere oder alle Fälle verallgemeinerten Form“ (Gerloff, 2021, S. 268). Bearbeitungen auf der Stufe 5 hingegen sind in beiden Klassenstufen mit einem Anteil von unter 5% an allen Bearbeitungen selten.

Beim Vergleich der Bearbeitungen zu den impliziten und den expliziten Aufgabenvarianten fällt auf, dass die relativen Anteile der Bearbeitungen auf Stufe 0 fast identisch sind. Der Anteil der Begründungen hingegen ist mit 55% im Vergleich zu 13,5% bei den expliziten Aufgabenvarianten statistisch hoch signifikant größer mit mittlerer Effektstärke. In fast der Hälfte der Fälle sind die impliziten Aufgabenvarianten „mit einer formulierten, passfähigen Aussage ohne Begründung beantwortet“ (Gerloff, 2021, S. 271).

Des Weiteren stellt Gerloff (2021) fest, dass es sich bei den Bearbeitungen der Aufgabenteile aus dem Bereich „Muster und Strukturen“ statistisch hoch signifikant häufiger um Begründungen handelt als bei den Bearbeitungen der Aufgabenteile aus dem Bereich „Raumvorstellungen“. Allerdings scheint die Klassenzugehörigkeit für die Begründungshäufigkeit entscheidender zu sein als der Inhaltsbereich (Gerloff, 2021; Abschn. 5.1.2.3).

Während sich die Studie von Gerloff (2021) auf die Geometrie bezieht, verwendet **Vogt (2021)** zwei stochastische Aufgaben. In der ersten Aufgabe sind drei Paare an Urnen mit roten und schwarzen Kugeln vorgegeben. Mit Blick auf jedes Urnenpaar ist zu entscheiden, bei welcher Urne man die höchste Gewinnchance hat, danach sind die Entscheidungen (gebündelt) zu begründen. Die zweite Aufgabe ist weitgehend identisch, ausschließlich das Verhältnis der Kugeln in den Urnen ist etwas anspruchsvoller gewählt. Vogt (2021) fordert 184 Viertklässlerinnen und Viertklässler auf, die beiden stochastischen Aufgaben eigenständig schriftlich zu bearbeiten. Alle Aufgabenbearbeitungen, außer fehlende Bearbeitungen und Bearbeitungen, wie z. B. „Ich kann das nicht", ordnet sie genau einer der folgenden fünf ordinal skalierten Kategorien zu: „Zusammenhänge (z. T.) beschrieben; ansatzweise begründet; beispielbezogen begründet; z.T. verallgemeinernd begründet; verallgemeinernd/formal begründet" (Vogt, 2021, S. 112). Der Anteil der Bearbeitungen mit einer beispielbezogenen oder verallgemeinernden Begründung (Kategorien 3-5) liegt bei beiden Aufgaben unter 10%. Bei der ersten Aufgabe fallen 64%, bei der zweiten 37% in die zweite Kategorie „ansatzweise begründet". Für eine Begründung müssen bei jeder der Aufgaben im Wesentlichen zwei inhaltliche Überlegungen ausgeführt werden. Die genannten Anteile der Kategorie „ansatzweise begründet" verdeutlichen, dass oftmals eine dieser Überlegungen, aber nicht beide in den Ausführungen der Lernenden vorkommen. Weiterhin ist, wie angegeben, der Anteil der Kategorie „ansatzweise begründet" bei der zweiten Aufgabe kleiner als bei der ersten und stattdessen der Anteil der Basiskategorie „Zusammenhänge (z. T.) beschrieben" größer. Dies deutet darauf hin, „dass die stochastische Aufgabenkomplexität Einfluss auf die Begründungsqualität der Schüler_innen hatte" (Vogt, 2021, S. 215). Der Aspekt der Komplexität von Aufgaben wird in Abschnitt 5.1.2 aufgegriffen.

5.1.1.2 Studienübergreifende Betrachtung

Zusammenfassend betrachtet zeigen die vorgestellten Studien, dass manche Lernende in den Sekundarstufen mathematische Allaussagen mit unendlichem Gültigkeitsbereich selbstständig beweisen können (z. B. Chin et al., 2009; Healy & Hoyles, 2000). Sie deuten jedoch zugleich an, dass dies der großen Mehrheit der Lernenden und insbesondere auch vielen im sozialen Vergleich mathematisch leistungsstarken Schülerinnen und Schülern nicht gelingt (u. a. Bell, 1976; Flade & Pruzina, 1990; Healy & Hoyles, 2000; Reyhani et al., 2011; Walsch, 1975). Dabei finden sich in manchen, aber nicht in allen Querschnittstudien Hinweise darauf, dass Schülerinnen und Schüler höherer Klassenstufen Aufgaben mit Beweisnotwendigkeit im Durchschnitt ähnlich gut oder besser bearbeiten (Brunner et al., 2022; Gerloff, 2021; Moya et al., 2015; Neumann et al., 2014; Porteous, 1990). Längsschnittstudien mit Lernenden siebter bis zehnter Klassen zeigen keine klare positive Tendenz (Küchemann & Hoyles, 2002; Reiss et al., 2002; Ufer, 2008).

Sowohl mit Blick auf Schülerinnen und Schüler in verschiedenen Klassenstufen als auch mit Fokus auf Lernende in der gleichen Klassenstufe weisen die Studien auf deutliche Leistungsunterschiede hin (u. a. Gerloff, 2021; Reiss et al., 2002; Healy & Hoyles, 2000). Neben manchen Schülerinnen und Schülern, denen Beweise von Allaussagen mit unendlichem Gültigkeitsbereich gelingen, gibt es Lernende, die bei Aufgaben mit Beweisnotwendigkeit keine Bearbeitung abgeben, sich nicht genau auf die zu zeigende Aussage beziehen, die Aussage selbst als Grund

für ihre Gültigkeit anführen oder einen rein empirisch-induktiven Ansatz wählen (u. a. Balacheff, 1988; Bell, 1976; Coe & Ruthven, 1994; Hoyles & Healy, 2007; Krumsdorf, 2009a, 2011). Zudem gibt es Lernende, die ein generisches Beispiel formulieren oder ein Beispiel nutzen, um ihre allgemeingültigen Überlegungen zu versprachlichen (Balacheff, 1988; Krumsdorf, 2017; Moya et al., 2015). Weiterhin kommt es vor, dass Schülerinnen und Schüler nicht angemessen zwischen Gesuchtem und Gegebenem unterscheiden oder für eine Begründung durchaus geeignete Überlegungen nicht zu einem stringenten Argumentgefüge verknüpfen (Reiss, o. J.; Reiss et al., 2006; Hoyles & Healy, 2007). Begründungen von Lernenden sind teils unvollständig, wobei große Teile des erforderlichen Argumentgefüges oder auch nur Kleinigkeiten fehlen können (u. a. Healy & Hoyles, 2000; Lin, 2005; Vogt, 2021).

Die Häufigkeiten solcher Arten an Bearbeitungen sind je nach Studie verschieden und nur schwierig zu vergleichen, weil sehr unterschiedliche Aufgabenstellungen verwendet und deren Bearbeitungen mit verschiedenen Foki und Kategoriensystemen ausgewertet werden. In mehreren Studien, die auf Aufgaben mit Beweisnotwendigkeit und einem unendlichen Gültigkeitsbereich fokussieren, wird ein großer Anteil an Bearbeitungen mit empirisch-induktivem Ansatz, ohne deduktive Begründung hervorgehoben (u. a. Balacheff, 1988; Coe & Ruthven, 1994; Knuth et al., 2002; Küchemann & Hoyles, 2002; Porteous, 1990). Außerdem erscheint es auf Grundlage der Studien recht klar, dass Lernenden Beweisversuche zu aus dem Schulunterricht bekannten Allaussagen mit unendlichem Gültigkeitsbereich im Mittel besser gelingen als dies für unbekannte Aussagen der Fall ist (Chin & Lin, 2009; Healys & Hoyles, 2000; Hoyles & Healy, 2007). Mathematisch leistungsstarke Schülerinnen und Schüler scheinen Aufgaben mit Beweisnotwendigkeit im Mittel gleich gut oder besser zu bearbeiten als andere Lernende in der gleichen Klassenstufe (Moya et al., 2015; Reiss et al., 2002). Es scheint hinsichtlich der Qualität von Beweisversuchen von Schülerinnen und Schülern Unterschiede zwischen verschiedenen Ländern zu geben (u. a. Chin & Lin, 2009; Healy & Hoyles, 2000).

5.1.2 Zusammenhänge zwischen Bearbeitungen von Aufgaben mit Beweisnotwendigkeit und weiteren Faktoren

Dieser Abschnitt fokussiert auf Zusammenhänge zwischen Bearbeitungen von Aufgaben mit Beweisnotwendigkeit und weiteren Faktoren. Die Beschreibung von Studien in Abschnitt 5.1.1 hat bereits Einblicke gegeben, inwieweit Bearbeitungen von Aufgaben mit Beweisnotwendigkeit mit der Klassenstufe der Lernenden zusammenhängen. Außerdem deuten die in Abschnitt 5.1.1 beschriebenen Ergebnisse an, dass Bearbeitungen von Aufgaben mit Beweisnotwendigkeit mit Merkmalen dieser Aufgaben, wie z. B. dem Gültigkeitsbereich, dem Komplexitätsgrad und der Explizität der Beweisanforderung, zusammenhängen. Im Folgenden werden gezielt weitere Forschungsergebnisse dazu vorgestellt, inwieweit Bearbeitungen von Aufgaben mit Beweisnotwendigkeit mit Faktoren, wie z. B. dem Geschlecht von Schülerinnen und Schülern, deren Performanz beim Validieren und Bewerten fiktiver Beweisversuche und dem erlebten Mathematikunterricht, zusammenhängen. Leserinnen und Leser, die ausschließlich die Kernpunkte interessieren, wird vorgeschlagen, direkt zum Abschnitt 5.1.2.5, der Zusammenfassung, überzugehen.

5.1.2.1 Cognitive Development and Achievement in Secondary School Geometry (CDASSG) project

Im Rahmen des in Abschnitt 5.1.1.1 erwähnten, von Senk (1982) sowie Senk und Usiskin (1983) präsentierten Projekts mit über 1000 US-amerikanischen Schülerinnen und Schülern aus den Klassenstufen 7 bis 12 sollen die Teilnehmenden am Ende des Schuljahres 1980/81 zusätzlich zu elementargeometrischen Beweisaufgaben auch einen „Van Hiele Geometry Test“ (Senk, 1982, S. 6) und einen „standardized geometry achievement test“ (ebd.) bearbeiten. Außerdem sind die Teilnehmenden bereits am Anfang des Schuljahres 1980/81 zur Bearbeitung eines „Entering Geometry Student Test“ (Senk, 1982, S. 5) und eines „Van Hiele Geometry Test“ (ebd.) aufgefordert worden. Die Ergebnisse beider Van Hiele Geometry Tests und die Ergebnisse des geometry achievement tests korrelieren hoch signifikant mit den Leistungen bei den Beweisaufgaben (Senk, 1982). Außerdem hält Senk (1982) fest, dass hinsichtlich des Gelingens der Beweisversuche deutliche Unterschiede zwischen Schülerinnen und Schülern aus verschiedenen Schulen zu erkennen sind. Statistisch signifikante Geschlechterunterschiede sind hinsichtlich des Gelingens der Beweisversuche nicht zu sehen (Senk & Usiskin, 1983).

5.1.2.2 Das Forschungsprojekt „Justifying and Proving in School Mathematics“

Im Rahmen des britischen Forschungsprojekts „Justifying and Proving in School Mathematics“ (Healy & Hoyles, 2000; Hoyles & Healy, 2007; Abschn. 5.1.1.1) mit mathematisch leistungsstarken Zehntklässlerinnen und Zehntklässlern werden Zusammenhänge zwischen den Beweisversuchen der Teilnehmenden und weiteren Faktoren systematisch untersucht.

Die Analyse zeigt insbesondere signifikante Zusammenhänge zwischen den **Ergebnissen im allgemeinen Mathematiktest**, mit dem die Teilnehmenden ausgewählt wurden, und der Qualität der Beweisversuche zu den verschiedenen Allaussagen, obwohl mathematisches Beweisen im allgemeinen Mathematiktest nicht geprüft wurde (Healy & Hoyles, 2000; Hoyles & Healy, 2007).

Die Qualität der Beweisversuche korreliert darüber hinaus mit den Bearbeitungen weiterer Testaufgaben, wenn auch mit geringerer Effektstärke. Zum Beispiel sollen die Teilnehmenden zusätzlich zu der Bearbeitung der Beweisaufgaben entscheiden, inwieweit – unter der Annahme, dass eine bestimmte Allaussage bewiesen wurde – ein vorgegebener Spezialfall dieser Aussage gültig ist oder es eines zusätzlichen Nachweises bedürfe. Die Analysen zeigen signifikante Zusammenhänge zwischen den Bearbeitungen dieser Aufgabe und der Qualität der Beweisversuche zu drei von vier Allaussagen (Healy & Hoyles, 2000; Hoyles & Healy, 2007). Sie deuten an, dass eine gewisse **Einsicht in die Allgemeingültigkeit bewiesener Aussagen** mit der Qualität eigener Beweisversuche zusammenhängt.

Des Weiteren werden mit Blick auf die Beweisaufgaben aus dem Bereich Arithmetik und Algebra signifikante **Geschlechterunterschiede** zugunsten der Mädchen deutlich (Healy & Hoyles, 2000). Zwischen den Bearbeitungen der geometrischen Beweisaufgaben und dem Geschlecht der Lernenden sind jedoch keine statistisch signifikanten Zusammenhänge zu erkennen (Hoyles & Healy, 2007).

Zudem werden **Merkmale des Schulunterrichts und der Lehrkraft** in die Analyse miteinbezogen. Die Ergebnisse deuten darauf hin, dass die Performanz in Hinblick auf Beweisaufgaben vom besuchten Mathematikunterricht der Teilnehmenden abhängt. Hoyles und Healy (2007) erläutern mit Fokus auf die Beweisversuche zu den geometrischen Allaussagen:

> It is also of interest, that students in classes with a larger proportion taking the higher- rather than the middle-tier GCSE paper[97] were better at construction proofs for the unfamiliar conjecture and the experience of writing proofs improved scores for the proofs of the familiar conjecture. Both factors point to the influence of teaching. The latter points to the importance of familiarity in setting down an argument in geometry. The former suggests that teachers may have a rather different approach to teaching about proving if there is a majority of students in the class who are highly motivated to take a challenging test at the end of the year. The point is that this influence can be detected among all students in the class regardless of prior mathematics attainment and regardless of whether (or not) the student was actually due to sit the challenging test. Maybe it is only in these classes that students have gained enough familiarity with simple geometrical content and language to be able to begin to engage in argumentation? Maybe it is only in these classes that students are given some instruction in organising an argument in Euclidean geometry? (Hoyles & Healy, 2007, S. 105)

5.1.2.3 Studien mit Gymnasiastinnen und Gymnasiasten in den Klassenstufen 7, 8

Die u. a. von Reiss et al. (2002) vorgestellten Ergebnisse von Leistungstests mit Gymnasiastinnen und Gymnasiasten aus siebten und achten Klassen deuten an, dass gute Leistungen beim **Anwenden geometrischer Sätze und Begriffe bei einfachen Berechnungsproblemen** eine notwendige, jedoch keine hinreichende Voraussetzung für eine erfolgreiche Bearbeitung mehrschrittiger Beweisaufgaben darstellen (Abschn. 5.1.1.1). Zusätzlich zu den Leistungstests werden in der von Reiss et al. (2002) präsentierten Studie **mathematische Beliefs, Kompetenzen im wissenschaftlichen Denken, das mathematikspezifische Interesse und die mathematikbezogene Motivation** der Lernenden erhoben (Reiss et al., 2002; Heinze & Reiss, 2004). Die Ergebnisse zeigen bei den Schülerinnen und Schülern aller Klassen weitgehend unabhängig von ihrem Abschneiden im Leistungstest „ein recht einheitliches Bild von Mathematik, das wesentlich durch Formalismus und Exaktheit geprägt ist“ (Reiss et al., 2002, S. 60). Es scheint keinen deutlichen, direkten Zusammenhang zwischen mathematischen Beliefs von Schülerinnen und Schülern verschiedener Klassen und deren Leistungen bei Beweisaufgaben zu geben. Kompetenzen im wissenschaftlichen Denken scheinen hingegen durchaus eine Rolle zu spielen (Reiss et al., 2002). Mit Blick auf das mathematikspezifische Interesse und die mathematikbezogene Motivation der Lernenden zeigen sich signifikante Korrelationen zwischen den Faktoren „Interesse/Freude an Mathematik“ (Heinze & Reiss, 2004, S. 244) sowie „Angst vor Leistungsversagen“ (ebd.) und den Ergebnissen im Leistungstest.

Ufer et al. (2009) untersuchen Daten aus zwei Teilerhebungen mit insgesamt rund 700 Gymnasiastinnen und Gymnasiasten in der achten Klassenstufe. Die Teilnehmenden werden zu Leistungstests der beschriebenen Art (Abschn. 5.1.1.1) und zusätzlich zum **Validieren und Bewerten fiktiver Beweisversuche** aufgefordert. Ufer et al. (2009) stellen fest, dass die Leistungen beim Validieren und Bewerten fiktiver Beweisversuche signifikant mit den Ergebnissen der

[97] GCSE (General Certificate of Secondary Education) is the public examination taken by students in England and Wales at the end of their compulsory schooling (age 16 years). Students are entered to one of three levels in the examination […].

Bearbeitungen der einschrittigen Beweisaufgaben (Niveau II) und noch deutlicher mit den Leistungen bei den mehrschrittigen Beweisaufgaben (Niveau III) korrelieren. Eine Regressionsanalyse ergibt, „dass etwa 10% der Varianz der Punktzahl der Beweisitems (Niveau II und III) durch das Methodenwissen erklärt werden können“ (Ufer et al., 2009, S. 45). Dies sei sicher kein großer Effekt, lege jedoch nahe, dass das Verständnis für Kriterien zur Akzeptanz von Beweisen durchaus mit der Fähigkeit zur Konstruktion von Beweisen zusammenhängt (Ufer et al., 2009). Kuntze (2009) beschäftigt sich in einem weiteren Sinne mit **Metawissen zum mathematischen Beweisen**. Dieses umfasst „verschiedene Aspekte wie etwa die drei Bereiche *beweisbezogenes Methodenwissen*, *Wissen über die Entwicklung von Beweisen* und *Wissen über Funktionen des Beweisens*“ (Kuntze, 2009, S. 220, H. i. O). Die Ergebnisse einer Studie mit Gymnasiastinnen und Gymnasiasten der achten Klassenstufe deuten an, dass Metawissen zum Beweisen einen relevanten Faktor bei der Entwicklung von Kompetenzen im Führen von Beweisen darstellt (Kuntze, 2009).

Heinze et al. (2007) analysieren die Daten von Leistungstests aus vier empirischen Studien mit insgesamt über 2500 Schülerinnen und Schülern der Klassenstufen 7, 8 hinsichtlich **Geschlechterunterschiede**, wobei sie keine nennenswerten Unterschiede feststellen.

Unter anderem Reiss et al. (2002) werten die Ergebnisse der Leistungstests auf Klassenebene aus. Dabei werden bezogen auf beide Messzeitpunkte, Ende der siebten und Mitte der achten Klassenstufe, hoch signifikante **Unterschiede zwischen den getesteten Klassen** deutlich, obwohl das im Test erforderliche Bereichswissen zuvor in allen Klassen im Unterricht behandelt wurde. Die Unterschiede zeigen sich vor allem bei den einschrittigen und den mehrschrittigen Beweisaufgaben. In den leistungsschwächeren Klassen gibt es bei diesen Aufgaben kaum korrekte Lösungen, wohingegen in den leistungsstärksten Klassen ein breites Spektrum an Leistungen vorkommt. Rund die Hälfte der Schülerinnen und Schüler, die beim zweiten Test Mitte der achten Klasse deutlich besser abschneiden, sind in nur drei der 27 Klassen. Rund die Hälfte der Lernenden, die sich deutlich verschlechtert haben, kommen ausschließlich aus fünf verschiedenen Klassen (Heinze & Reiss, 2004). Zudem stellen Reiss et al. (2002) fest, dass einige Schülerinnen und Schüler ihre Rechnungen in den einfachen Anwendungsaufgaben begründen, auch wenn eine Herleitung nicht verlangt ist. Dies kommt in leistungsstärkeren Schulklassen häufiger vor als in leistungsschwächeren.

5.1.2.4 Weitere Ergebnisse zur Relevanz des erlebten Mathematikunterrichts

Wie in den Abschnitten 5.1.2.2, 5.1.2.3 beschrieben deuten Ergebnisse mehrerer Untersuchungen mit Schülerinnen und Schülern in der Sekundarstufe I darauf hin, dass Leistungen bei der Bearbeitung von Beweisaufgaben mit Merkmalen des erlebten Mathematikunterrichts zusammenhängen. Diese Ergebnisse ergänzend stellt Gerloff (2021), die **Aufgabenbearbeitungen von Schülerinnen und Schülern in den Klassenstufen 3, 4** untersucht, deutliche Klassenunterschiede fest. Grundlegend ist zu sehen, dass der Anteil fehlender und deshalb auf Stufe 0 eingeordneter Aufgabenbearbeitungen je nach Klasse zwischen 0% und rund 20% beträgt, wobei die Werte auch innerhalb der beiden Klassenstufen 3, 4 deutlich variieren. Des Weiteren schreibt Gerloff (2021) mit Blick auf die Bearbeitungen der Aufgaben mit impliziter Begründungsaufforderung:

> Die Verschriftlichung einer Begründungsantwort bei lediglich impliziter Begründungsaufforderung steht, auch bei grundsätzlichem Zugang zur Aufgabenstellung, in einem hoch signifikanten Zusammenhang zur Klassenzugehörigkeit. Dabei zeigt sich, dass die Klassenzugehörigkeit einen noch stärkeren Effekt auf die Begründungshäufigkeit bei impliziten Aufforderungen besitzt als die Jahrgangsstufe. (Gerloff, 2021, S. 275)

Außerdem erscheint die Klassenzugehörigkeit in der Untersuchung von Gerloff (2021) für die Begründungshäufigkeit entscheidender zu sein als der Inhaltsbereich der Aufgaben.

Während sich u. a. bei der Untersuchung von Gerloff (2021) nicht klar sagen lässt, inwieweit der erlebte Mathematikunterricht oder andere Faktoren für die Klassenunterschiede verantwortlich sind, weisen verschiedene Interventionsstudien darauf hin, dass Schülerinnen und Schüler im Rahmen geeigneter Lernumgebungen Fortschritte im Führen von Beweisen machen können. Dazu gehören die Interventionsstudie von Bezold (2009) mit Drittklässlerinnen und Drittklässlern, die Interventionsstudie von Vogt (2021) mit Viertklässlerinnen und Viertklässlern, sowie die Interventionsstudie von Kuntze (2006) mit Schülerinnen und Schülern achter Klassen und die von Reiss et al. (2006) mit Schülerinnen und Schülern in den Klassenstufen 7 bis 8. Auch die Ergebnisse der Entwicklungsforschungsstudie von Hein (2021) mit Schülerinnen Schülern achter bis zehnter Klassen sowie Ergebnisse der Untersuchungen von Lin (2005) mit Lernenden aus neunten Klassen deuten darauf hin, dass Schülerinnen und Schüler durch eine **gezielte Förderung** beim Führen von Beweisen diesbezüglich Fortschritte machen können. Somit legen zahlreiche Forschungsergebnisse die Vermutung nahe, dass der erlebte Mathematikunterricht für die Leistungen von Schülerinnen und Schülern bei Aufgaben mit Beweisnotwendigkeit von großer Bedeutung ist.

5.1.2.5 Zusammenfassung

Zusammenfassend deuten die Ergebnisse an, dass die Qualität von Beweisversuchen von Schülerinnen und Schülern mit deren Leistungen bei mathematischen Aufgaben, die keine Beweiskompetenzen erfordern, zusammenhängt. Auch scheinen Leistungen beim Validieren und Bewerten fiktiver Beweisversuche, das vorhandene Metawissen zum mathematischen Beweisen und im Speziellen eine gewisse Einsicht in die Allgemeingültigkeit bewiesener Allaussagen durchaus mit der Qualität von Beweisversuchen zu korrelieren. Zudem scheinen Zusammenhänge zu einzelnen Merkmalen des fachspezifischen Interesses und der Motivation von Schülerinnen und Schülern zu bestehen. Bei Beweisversuchen zu algebraischen Allaussagen zeigen sich in der Studie von Healy und Hoyles (2000) Geschlechterunterschiede zugunsten der Mädchen, während in mehreren Analysen mit Blick auf Beweisversuche im Bereich der Geometrie keine nennenswerten Geschlechterunterschiede auftreten. Mehrere Studien deuten auf Zusammenhänge zwischen der Qualität von Beweisversuchen und dem erlebten Mathematikunterricht hin. Außerdem lässt sich festhalten, dass die Klassenstufe von Schülerinnen und Schülern sowie Merkmale der zu bearbeitenden Aufgabe mit Blick auf die Qualität von Beweisversuchen zumindest von gewisser Bedeutung sind.

5.2 Verwendung von Beispielen

Nachdem im Abschnitt 5.1 das Gelingen von Bearbeitungen von Aufgaben mit Beweisnotwendigkeit im Fokus stand, werden in diesem Abschnitt Ergebnisse empirischer Studien gebündelt, die sich mit der Verwendung von Beispielen in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit beschäftigen. Beispiele können in Beweisversuchen von Schülerinnen und Schülern eine zentrale Rolle spielen (Abschn. 3.3, 5.1). Es kann sein, dass Schülerinnen und Schüler ausschließlich empirisch-induktiv argumentieren. Zudem ist es möglich, dass sie z. B. ein generisches Beispiel oder einen beispielgebundenen Beweis formulieren. Auch können Beispiele in Beweisversuchen mit begründendem Ansatz vorkommen, in denen nicht unbedingt beispielgebunden argumentiert wird. Im Folgenden wird auf wesentliche Forschungsergebnisse zur Art und zum Umfang der Verwendung von Beispielen in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit eingegangen. Zudem werden Gründe für diese Beispielnutzung näher beschrieben sowie Einblicke in Zusammenhänge zwischen der Verwendung von Beispielen und der dem Gelingen von Beweisversuchen gegeben (Abschn. 5.2.1 – 5.2.3). Wesentliche Ergebnisse werden am Ende dieses Abschnitts kurz zusammengefasst (Abschn. 5.2.4).

5.2.1 Empirisch-induktive Ansätze

Im vorhergehenden Abschnitt 5.1 ist bereits deutlich geworden, dass Lernende bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit oftmals einen empirisch-induktiven Ansatz wählen. Mit dem Terminus „empirisch-induktiv“ werden in diesem Abschnitt Bearbeitungen bezeichnet, in denen Beispiele angeführt und eventuell von den Beispielen auf die Allgemeingültigkeit einer Allaussage geschlossen wird. Generische Beispiele und beispielgebundene Beweise werden dabei nicht mitgedacht.

Publikationen mehrerer Forscherinnen und Forscher machen darauf aufmerksam, dass Bearbeitungen mit empirisch-induktivem Ansatz verschiedene **Ausprägungen** aufweisen können und geben Einblicke, wie häufig diese sind.[98] In der britischen Studie von Küchemann und Hoyles (2002) mit mathematisch leistungsstarken Achtklässlerinnen und Achtklässlern geben viele der Teilnehmenden genau ein Beispiel an, wenige mehrere Beispiele. Einzelne Lernende verdeutlichen in ihrer Bearbeitung, dass die Angabe von Beispielen für einen Beweis nicht ausreicht. In der Untersuchung von Moya et al. (2015) mit mathematisch talentierten und durchschnittlich leistungsstarken Lernenden kommen Aufgabenbearbeitungen vor, die laut den Verantwortlichen auf ein crucial experiment im Sinne Balacheffs (1988) hindeuten. Zudem kommen Aufgabenbearbeitungen vor, die von ihnen als naive empiricism oder als empirical failed bezeichnet werden (Abschn. 5.1.1). In der Untersuchung von Knuth et al. (2002) sind in über 50% der Beweisversuche aus den Klassenstufen 6 bis 8 ein oder mehrere unsystematische Beispiele angeführt. Ein kleinerer Anteil der Beweisversuche enthält Beispiele verschiedener Art, wie etwa kleine und sehr große Beispiele.

[98] Im Folgenden wird das Design der Studien, die bereits in Abschnitt 5.1 ausschnittweise vorgestellt wurden, allenfalls skizziert. Ergänzende Informationen sind in Abschnitt 5.1 und den angegebenen Publikationen zu finden.

Dass eine Schülerin oder ein Schüler bei einer Aufgabe mit Beweisnotwendigkeit einen empirisch-induktiven Ansatz wählt, kann unterschiedliche **Gründe** haben. Beispiele scheinen aus Sicht einiger Lernender durchaus überzeugende Kraft zu besitzen und ihnen auszureichen, um sehr sicher zu gehen. Im britischen Forschungsprojekt von Healy und Hoyles (2000) sollen die mathematisch leistungsstarken Zehntklässlerinnen und Zehntklässler auch fiktive Beweisversuche mit empirisch-induktivem Ansatz bewerten. Healy und Hoyles (2000) schreiben:

> These data indicates that fewer than one third of the students […] felt that empirical argument had no explanatory value at all. […] Examples provide an immediate entry into what the conjecture is about and help to convince oneself or another of its truth. (Healy & Hoyles, 2000, S. 412)

Dabei ist es möglich, dass Lernenden die prinzipiell fehlende Allgemeingültigkeit einzelner Beispiele bewusst ist und sie ihnen trotzdem viel Aussagekraft zusprechen. Chazan (1993) befragt etwa 15 Jahre alte US-amerikanische Lernende in Interviews zu fiktiven Beweisversuchen. Die Lernenden, die Beweisversuche mit empirisch-induktivem Ansatz als Beweis ansehen, haben Strategien bei der Verwendung von Beispielen als Stützung einer allgemeinen Gültigkeit entwickelt, welche die Grenzen der Verwendung von Beispielen aus ihrer Sicht abmildern. Sie erklären z. B., dass man einen möglichst allgemeinen Fall überprüfen müsse, dass es wichtig sei, beim Überprüfen alle verschiedenen Arten von Fällen zu berücksichtigen, dass sie sich sämtliche Möglichkeiten im Kopf vorgestellt hätten, oder dass sie sich auf der Grundlage der Beispiele zwar nicht ganz sicher seien, dass die infrage stehende Aussage gelte, aber ziemlich sicher. Coe und Ruthven (1994) führen mit einigen Lernenden ein follow-up Interview durch. Ein Schüler erklärt, dass er sich ausgehend von seinen sechs Beispielen der Gültigkeit der Vermutung zu fast 100% sicher sei. Ein anderer Schüler zieht seine Überzeugung aus einer Regelmäßigkeit, die er in seinen Beispielen erkannt hat: „No [I have no doubt], because all this progresses in the same sort of way. It's not going to suddenly radically change. There isn't suddenly going to be a huge gap. It's always progressing in the same sort of way" (Coe & Ruthven, 1994, S. 50). Im Projekt von Healy und Hoyles (2000) erklären manche Teilnehmende, die einen empirisch-induktiven Beweisversuch zu einer bekannten Allaussage als allgemeingültig bewertet haben, im Rahmen der follow-up Interviews, „that they had made this response as a result of having 'looked through' the particular cases to the generality because the truth of the conjecture was so obvious" (Healy & Hoyles, 2000, S. 411).

Außerdem ist es möglich, dass Lernende bei Aufgaben mit Beweisnotwendigkeit einen rein empirisch-induktiven Ansatz formulieren, weil ihnen die Entwicklung einer Begründung nicht gelingt oder sie sich nicht mehr zutrauen. In den follow-up Interviews von Coe und Ruthven (1994) sagen einzelne Lernende, dass sie sich die Begründungsfrage im Rahmen der Aufgabe schon gestellt hätten, aber es schwierig fänden, diese zu beantworten. In den follow-up Interviews von Healy und Hoyles (2000) erläutern einige Teilnehmende, dass sie einen empirisch-induktiven Beweisversuch in der Aufgabe zum Validieren und Bewerten nur als Beweisversuch ausgewählt hätten, der dem eigenen Ansatz am nächsten käme, weil ihnen bewusst gewesen sei, dass ihnen kein besserer Beweisversuch gelingen würde.

Weiterhin kann es vor allem bei Aufgaben, in denen nicht explizit zum Beweisen aufgefordert wird, vorkommen, dass Lernende diese Anforderung nicht berücksichtigen. Die Ergebnisse von Gerloff (2021) und von Coe und Ruthven (1994) deuten an, dass dies durchaus häufig zutrifft. In der Untersuchung von Gerloff (2021) ist, wie in Abschnitt 5.1.1 beschrieben, in etwa der Hälfte der Bearbeitungen von impliziten Aufgabenvarianten, zwar eine zu der Aufgabenstellung passende Aussage, aber keine Begründung dieser Aussage formuliert. Der Anteil der Bearbeitungen, in denen eine Begründung vorkommt, ist bei den expliziten Aufgabenvarianten statistisch hoch signifikant größer mit mittlerer Effektstärke als bei den impliziten. In den follow-up Interviews von Coe und Ruthven (1994) erklären mehrere Teilnehmende, dass ihnen die Begründungsfrage nicht relevant oder wichtig erscheine, wie die Autoren u. a. mit folgendem Zitat veranschaulichen: „[...] when we found it [a conjecture] we didn't actually think why, we just put it down and went on to the next thing, I suppose we didn't think it was relevant" (Coe & Ruthven, 1994, S. 51).

Zusätzlich deuten Studien an, dass eine gewisse Unsicherheit oder Skepsis bezüglich der Korrektheit und Allgemeingültigkeit mathematischer Beweise zur Beschränkung auf einen empirisch-induktiven Ansatz beitragen können. Im Projekt „Justifying and Proving in School Mathematics" (Healy & Hoyles, 2000; Hoyles & Healy, 2007) geben manche Teilnehmende an, dass ein Spezialfall einer bewiesenen Allaussage eines weiteren Nachweises bedürfe. In der Interviewstudie von Chazan (1993) sehen manche Teilnehmende mathematische Beweise nur als Evidenz an. Diese Lernenden sagen u. a., dass es außerhalb der Fälle, die durch den Beweis abgedeckt seien, ein Gegenbeispiel geben könne und dass die Annahmen, auf denen ein Beweis beruhe, falsch sein können.

Wie häufig es vorkommt, dass Schülerinnen und Schüler bei einer Aufgabe mit Beweisnotwendigkeit einen empirisch-induktiven Ansatz abgeben, weil sie die betrachteten Beispiele von der Gültigkeit der Aussage überzeugen oder weil sie sich z. B. nicht mehr zutrauen, lässt sich auf Grundlage der vorhandenen Forschungsergebnisse nicht genau sagen. Sowohl im britischen Forschungsprojekt von Healy und Hoyles (2000) als auch in den ähnlichen Studien in Taiwan und im Iran (Chin & Lin, 2009; Reyhani et al., 2011) wählen beim Validieren und Bewerten fiktiver Beweisversuche zu algebraischen Allaussagen je unter 10% der Teilnehmenden einen Beweisversuch mit empirisch-induktivem Ansatz als Bearbeitung, die von ihrer Lehrkraft die beste Note erhalten würde. Zudem ordnen je mehr als die Hälfte der Teilnehmenden, teils über 75%, die Beweisversuche mit empirisch-induktivem Ansatz als nicht allgemeingültig ein (Chin & Lin, 2009; Healy & Hoyles, 2000; Reyhani et al., 2011). Ufer et al. (2009) analysieren Daten von knapp 700 Gymnasiastinnen und Gymnasiasten in der Klassenstufe acht, die zum Validieren und Bewerten fiktiver Beweisversuche aufgefordert werden. In rund zwei Drittel der Bearbeitungen ist angekreuzt, dass es sich bei dem vorgegebenen empirisch-induktiven Beweisversuch nicht um einen korrekten mathematischen Beweis handelt, in rund der Hälfte der Fälle ist die Ablehnung adäquat begründet. Insgesamt betrachtet deuten die Forschungsergebnisse an, dass vielen Lernenden zumindest intuitiv bewusst ist, dass empirisch-induktive Beweisversuche nicht ausreichen, um eine Allaussage zu beweisen, und dass mehr von ihnen erwartet wird.

5.2.2 Generische Beispiele und beispielgebundene Beweise

Studien belegen, dass manche Schülerinnen und Schüler bei Aufgaben mit Beweisnotwendigkeit ein **generisches Beispiel** formulieren (Balacheff, 1988; Moya et al., 2015). Dabei verdeutlichen die Beobachtungen von Balacheff (1988), dass es keine Selbstverständlichkeit ist, dass Schülerinnen und Schüler die Verallgemeinerbarkeit von an einem Beispiel durchgeführten Überlegungen erkennen und von den Besonderheiten der Beispiele abstrahieren. Manche Lernende scheint ein crucial experiment im Sinne Balacheffs (1988) eher von der Gültigkeit einer Allaussage zu überzeugen als ein generisches Beispiel.

Krumsdorf (2011, 2017) betont die mit dem **beispielgebundenen Beweisen** verbundene doppelte Herausforderung, das Allgemeingültige von den Besonderheiten der Beispiele in angemessener Form zu abstrahieren und die allgemeingültigen Überlegungen zu manifestieren. In seinen Einzelfallstudien nehmen manche Lernende beim Abstrahieren von Beispielen zu weitreichende Verallgemeinerungen vor (Krumsdorf, 2017). Weiterhin gibt es nach Krumsdorf (2009a, 2017) Lernende, die Teile eines erforderlichen Argumentgefüges sehr klar manifestieren, während andere latent bleiben und eventuell auch nicht kognitiv realisiert werden. Manchen Schülerinnen und Schülern gelingt es sehr gut, allgemeingültige Überlegungen mittels selbstgeprägter Wortbildungen und variablenhaltiger Umgangssprache zu manifestieren, einigen Lernenden fällt es schwer, „kreativ und versiert die Umgangssprache beim beispielgebundenen Beweisen zu nutzen“ (Krumsdorf, 2009a, S. 713; Krumsdorf, 2017). Grundsätzlich folgert Krumsdorf (2017) aus seinen Beobachtungen, dass bei der Interpretation sowohl konkreter als auch allgemein formulierter Aussagen Vorsicht geboten ist. Konkret Gesagtes könne allgemeiner gedacht werden. Eine allgemein formulierte Aussage müsse nicht mit einer subjektiven Realisierung des Allgemeingültigen einhergehen, sondern sei eventuell nur eine Sprechweise, um gefühlte Erwartungen zu erfüllen.

Bezüglich der Art und Anzahl an Beispielen folgert Krumsdorf (2017) aus den Einzelfallstudien, dass „die Wahl und die Anzahl an Beispielen wesentlich den beispielgebundenen Beweisverlauf eines Schülers beeinflusst“ (Krumsdorf, 2017, S. 353). Er kommt zu dem Schluss, dass die Betrachtung möglichst vieler Beispiele zu einer Verfestigung eines rein empirisch-induktiven Vorgehens führen kann. Die Betrachtung weniger und insbesondere auch schwieriger oder nicht gut vorstellbarer Beispiele, wie etwa großer Beispiele, würde die Entwicklung eines beispielgebundenen Beweises eher begünstigen (Krumsdorf, 2017). Aus dieser Beobachtung lässt sich insbesondere die Option abstrahieren, dass Lernende in einer Aufgabenbearbeitung ein großes oder in anderer Hinsicht besonderes Beispiel anführen, weniger um die Gültigkeit der Aussage an sich zu prüfen, als um eine Beweisidee auszutesten und auszuschärfen.

5.2.3 Beispiele in Aufgabenbearbeitungen mit begründendem Ansatz

Weiterhin weisen die Beobachtungen von Krumsdorf (2009a, 2017) und weiterer Forscherinnen und Forscher darauf hin, dass Beispiele auch in Aufgabenbearbeitungen mit begründendem Ansatz vorkommen können, in denen nicht unbedingt beispielgebunden argumentiert wird. **Krumsdorf (2009a, 2017)** kommt zu dem Ergebnis, dass manche „SchülerInnen während des beispielgebundenen Beweisens oder im Anschluss daran noch zu induktiven Prüfungen im

Sinne einer zusätzlichen Bestätigung greifen oder die Grenzen ihrer Verallgemeinerungen erproben“ (Krumsdorf, 2009a, S. 684). Zudem schreibt er: „Auch ein Bedürfnis nach der Konkretisierung im induktiven Prüfen mag Schüler dazu bewegen, die nach einem formal gehaltenen Beweis unfraglich gewordene Behauptung induktiv zu bekräftigen“ (Krumsdorf, 2017, S. 343). Diese Beobachtungen, dass und warum manche Lernende nach dem Beweisen noch ein Beispiel betrachten, kann durch die Forschungsergebnisse von **Fischbein (1982) und Porteous (1990)** ergänzt werden. In einer Studie von Fischbein (1982) mit israelischen Schülerinnen und Schülern in den Klassenstufen 10 bis 12 geben viele Teilnehmende trotz der Annahme, dass ein vorliegender Beweis korrekt ist, an, dass weitere Prüfungen notwendig seien, um das eigene Vertrauen in die infrage stehende Allaussage zu erhöhen. Sie erklären z. B., dass es immer Ausnahmen geben könne, dass weitere Prüfungen das Ergebnis präziser machen würden und dass man eine Aussage an Zahlen überprüfen müsse, um sie zu beweisen. Porteous (1990) fordert die Teilnehmenden in seiner Interviewstudie auf, die Gültigkeit eines Spezialfalls einer zuvor betrachteten Allaussage zu beurteilen. Mehrere Lernende, die die Allaussage zuvor bewiesen haben, wiederholen den Beweis am Spezialfall. Porteous (1990) erläutert:

> The response reflects a desire to ensure that the proof is valid. The data of the “particular” are being employed to check the proof itself, to make sure that its argument is correct, not to check a particular statement concerning those data. (Porteous, 1990, S. 593-594)

Die Untersuchungen von Krumsdorf (2009a, 2017), Porteous (1990) und Fischbein (1982) zusammen betrachtet liefern somit empirische Anhaltspunkte dafür, dass es sehr verschiedene Gründe gibt, warum manche Lernende trotz eines vorliegenden Beweises oder eines eigenen allgemein gehaltenen oder verallgemeinerbaren Argumentgefüges noch ein Beispiel betrachten und dass diese Betrachtung in unterschiedlicher Form, d. h. insbesondere als induktive Prüfung oder als eine erneute Durchführung allgemeingültiger Überlegungen erfolgen kann.

Prinzipiell sind vielfältige Verwendungen von Beispielen in Bearbeitungen mit begründendem Ansatz, in denen nicht unbedingt beispielgebunden argumentiert wird, denkbar und scheinen durchaus auch vorzukommen. **Küchemann (2008)** berichtet über Unterrichtsstunden mit Lernenden achter bis zehnter Klassen, in denen die Schülerinnen und Schüler Aussagen über aufeinanderfolgende Zahlen beweisen sollen, und zeigt einige Aufgabenbearbeitungen. In manchen Bearbeitungen befinden sich oberhalb, innerhalb oder unterhalb des (übrigen) Argumentgefüges Beispiele. Mal ist an ihnen eine Beweisidee hervorgehoben, mal nicht. Teils handelt es sich um genau ein Beispiel, teils um mehrere Beispiele und manchmal weisen die Beispiele besondere Eigenschaften, wie etwa Regelmäßigkeiten, auf.

Lynch und Lockwood (2019) sowie Ozgur et al. (2019) berichten vom amerikanischen **Forschungsprojekt „The Role and Use of Examples in learning to Prove“**, in dem Lernende ab der siebten Klassenstufe in Einzelinterviews Aufgaben mit Beweisnotwendigkeit bearbeiten sollen.[99] Sie berichten, dass ein großer Anteil der Schülerinnen und Schülern kleine oder einfach zu berechnende Beispiele und teils das kleinste Beispiel wählen, insbesondere im Rahmen einer anfänglichen Exploration einer Aussage (Lynch & Lockwood, 2019). Sowohl einige Schülerinnen und Schüler, die im Interview eine gangbare, wenn auch nicht unbedingt vollständige

[99] Ergänzende Informationen zu diesem Forschungsprojekt sind auch in Abschnitt 3.3.5.1 zu finden.

Begründung entwickeln, als auch einige derjenigen, denen dies nicht gelingt, gehen bei der selbstständigen Auswahl von Beispielen systematisch vor (Ozgur et al., 2019). Schülerinnen und Schüler, die eine gangbare Begründung entwickeln, betrachten seltener ein vielfältiges Set an Beispielen und nutzen Beispiele häufiger, um z. B. nach Invariante zu suchen (Ozgur et al., 2019).

5.2.4 Zusammenfassung

Insgesamt betrachtet belegen vorhandene Forschungsergebnisse trotz aller inhaltlicher und begrifflicher Diversität, dass Schülerinnen und Schüler Beispiele in empirisch-induktiven Bearbeitungen, als generische Beispiele, in beispielgebundenen Argumentgefügen und auch in ansonsten eher deduktiv gehaltenen Bearbeitungen mit begründendem Ansatz anführen. Sie verdeutlichen zudem, dass die Angabe von Beispielen aus sehr verschiedenen Gründen erfolgen kann. Insbesondere ist es möglich und nicht unwahrscheinlich, dass Lernenden die Unzulänglichkeit einer selbst erstellten empirisch-induktiven Bearbeitung bewusst ist. Außerdem gibt es empirische Hinweise darauf, dass die Art der ausgewählten Beispiele und der Umgang mit diesen Beispielen damit zusammenhängen, inwieweit Lernenden Bearbeitungen von Aufgaben mit Beweisnotwendigkeit gelingen.

5.3 Verwendung von Darstellungsmitteln

Im Rahmen der Entwicklung eines Kategoriensystems für die Analyse von Beweisversuchen zu einer Allaussage ist bereits deutlich geworden, dass in Beweisversuchen von Schülerinnen und Schülern verschiedene Darstellungsmittel vorkommen können (Abschn. 3.3.11). Im Folgenden werden Forschungsergebnisse vorgestellt, die Einblicke bezüglich der tatsächlichen Verwendung von Darstellungsmitteln in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit geben.

5.3.1 Symbolische und verbale Darstellungsmittel

Es zeigt sich in verschiedenen Studien, dass die **Verwendung von Algebra** nicht so stark ausgeprägt ist. Dabei scheint der konkrete Anteil an Lernenden, die in Beweisversuchen Variablen nutzen, eventuell länderspezifisch zu sein und mit Faktoren, wie etwa dem Bekanntheitsgrad oder dem Komplexitätsgrad einer zu beweisenden Aussage, zusammenzuhängen. Bell (1976) betont, dass in keiner Bearbeitung der Aufgabe aus dem Bereich Arithmetik/Algebra durch die vierzehn bis fünfzehn Jahre alten Teilnehmenden Algebra vorkommt. In der Interviewstudie von Porteous (1990) verwendet ausschließlich einer von fünfzig Lernenden, ein zwölfjähriger Schüler, in seinem Beweisversuch zu einer Allaussage Algebra. Im britischen Forschungsprojekt „Justifying and Proving in School mathematics“ (Healy & Hoyles, 2000; Hoyles & Healy, 2007) mit mathematisch leistungsstarken Zehntklässlerinnen und Zehntklässlern werden die Beweisversuche der Lernenden nach ihrer hauptsächlichen Darstellungsform („major forms in which their arguments were presented” (Healy & Hoyles, 2000, S. 407)) klassifiziert. Von den Beweisversuchen zu der aus dem Schulunterricht bekannten algebraischen Aussage gehören 11% zu der Kategorie „formal (algebraic)“ (Healy & Hoyles, 2000, S. 408), von den Beweisversuchen zu

der unbekannten algebraischen Aussage 3%. In der ähnlichen Untersuchung in Taiwan trifft dies auf etwa 33% der Beweisversuche zu der bekannten algebraischen Allaussage und auf rund 11% der Beweisversuche zu der unbekannten algebraischen Allaussage zu (Chin & Lin, 2009; Abschn. 5.1.1.1). In der ähnlichen Untersuchung im Iran werden 44% der Beweisversuche zu der aus dem Schulunterricht bekannten algebraischen Allaussage der Kategorie „formal" (Reyhani et al., 2011, S. 239) zugeordnet.

Wenngleich Forschungsergebnisse zeigen, dass es manchen Lernenden gelingt, eine Allaussage in symbolisch dargestellter Form zu beweisen (u. a. Healy & Hoyles, 2000), gibt es Hinweise darauf, dass einigen Lernenden die Verwendung symbolischer Darstellungsmittel bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit Schwierigkeiten bereitet und ihnen dabei Fehler passieren. Im Projekt „Justifying and Proving in School mathematics" (Healy & Hoyles, 2000; Hoyles & Healy, 2007) ist die Qualität der primär formal (algebraisch) dargestellten Beweisversuche schlechter als die der primär verbal dargestellten Bearbeitungen. Beispielsweise sind 17% der primär formal dargestellten Bearbeitungen zu der aus dem Schulunterricht bekannten Allaussage als gültige Beweise anzusehen und 42% der primär verbal dargestellten Bearbeitungen. Küchemann (2008) beobachtet in Unterrichtsstunden zu Eigenschaften aufeinanderfolgender Zahlen, dass die Lernenden aus den Klassenstufen acht bis zehn bei der Verwendung von Algebra oft Unsicherheiten zeigen und Fehler machen, wenn sie versuchen, Algebra zu verwenden. Beispielsweise verwenden einige Lernende die gleiche Variable für verschiedene Zahlen.

Neben der Frage nach der Verwendung von Algebra geben manche Untersuchungen auch Einblicke in die **Verwendung verbaler Darstellungsmittel** in Bearbeitungen von Beweisaufgaben. In der Studie von Reyhani et al. (2011) sind unter 2% der Beweisversuche rein verbal dargestellt und unter 5% der Beweisversuche gehören zu der Kategorie „Combinatorial response (formal and narrative)" (Reyhani et al., 2011, S. 238). Im britischen Projekt „Justifying and Proving in School Mathematics" (Healy & Hoyles, 2000; Hoyles & Healy, 2007) hingegen machen primär verbal dargestellte Beweisversuche bei den algebraischen Beweisaufgaben große Anteile aus. Beispielsweise sind bei der aus dem Schulunterricht nicht bekannten algebraischen Aussage 32% aller Beweisversuche bzw. ca. 83% derjenigen, die über einen empirisch-induktiven Ansatz hinaus gehen, primär verbal dargestellt. Im Themenbereich Arithmetik und Algebra deuten die Bearbeitungen der Aufgabe zum Validieren und Bewerten an, dass verbal dargestellte Beweisversuche für die Lernenden am ehesten zugänglich sind. Bei den narrativ dargestellten Beweisversuchen ist der Anteil der Lernenden, die deren Korrektheit und Allgemeingültigkeit richtig bewerten, mit über 60% am höchsten. Außerdem wird den verbal dargestellten Beweisversuchen am häufigsten erklärende und kommunizierende Funktion zugewiesen (Healy & Hoyles, 2000).

5.3.2 Ikonische Darstellungsmittel

Weiterhin zeigen die Ergebnisse des Projekts „Justifying and Proving in School Mathematics" (Healy & Hoyles, 2000; Hoyles & Healy, 2007), dass es Lernende gibt, die einen Beweisversuch primär **ikonisch** darstellen. Der Anteil primär ikonisch dargestellter Beweisversuche liegt bei

der bekannten algebraischen Allaussage bei 8%. Bei der unbekannten algebraischen Allaussage kommt hingegen kein primär ikonisch dargestellter Beweisversuch vor (Healy & Hoyles, 2000). Es lässt sich vermuten, dass die primäre oder ausschließliche Nutzung ikonischer Darstellungsmittel, wie figurierter Zahlen, unter anderem von Eigenschaften der zu zeigenden Aussage abhängt. In den von Küchemann (2008) abgebildeten Bearbeitungen von Aufgaben mit Beweisnotwendigkeit, die sich auf aufeinanderfolgende Zahlen beziehen und von Lernenden achter bis zehnter Klassen verfasst wurden, kommen verschiedene ikonische Darstellungsmittel, wie z. B. figurierte Zahlen, Einkreisungen, Klammern und Pfeile vor. Fokussiert auf figurierte Zahlen berichtet Küchemann (2008), dass diese eher selten und nicht unbedingt effektiv verwendet werden, wenngleich einzelne Lernende auch erfolgreich im Wesentlichen mithilfe figurierter Zahlen argumentieren. In der Untersuchung von Gerloff (2021) mit Schülerinnen und Schülern aus den Klassenstufen 3, 4 kommen Aufgabenbearbeitungen vor, die „*schriftsprachlich mit Zeichnung/Markierung*“ (Gerloff, 2021, S. 310, H. i. O.) dargestellt sind, aber keine mit einer ausschließlichen „*Verdeutlichung über Zeichnung/Markierung*“ (Gerloff, 2021, S. 311, H. i. O.). Dabei sind in der Studie von Gerloff (2021) bei allen Aufgaben Linien für die Aufgabenbearbeitung vorgegeben, was eventuell zum Fehlen rein ikonisch dargestellter Aufgabenbearbeitungen beiträgt.

5.3.3 Zusammenfassung

Zusammenfassend deuten manche Forschungsergebnisse an, dass Lernende in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit kaum Algebra verwenden, sondern ihre Bearbeitungen, wenn sie über einen empirisch-induktiven Ansatz hinausgehen, primär verbal darstellen. Die Verwendung von Algebra oder primär verbaler Darstellungsmittel scheint jedoch länderspezifisch ausgeprägt zu sein und von weiteren Faktoren, wie etwa der Bekanntheit oder dem Komplexitätsgrad der infrage stehenden Aussage, abzuhängen. Weiterhin deuten Forschungsergebnisse an, dass primär ikonisch dargestellte Beweisversuche manchmal vorkommen.

5.4 Diskussion

Die Ausführungen in den vorhergehenden Abschnitten 5.1 bis 5.3 verdeutlichen, dass empirische Studien zum Führen von Beweisen eine lange Tradition haben und zugleich weiterhin durchgeführt werden. In vielen der vorhandenen Studien liegt ein Fokus auf der Frage, inwieweit Schülerinnen und Schülern Beweise mathematischer Aussagen gelingen. Dabei bezieht sich ein Teil der Untersuchungen nicht speziell auf Beweisaufgaben, sondern auf Aufgaben mit Beweisnotwendigkeit, teils ohne explizite Aufforderung, einen Beweis zu formulieren, so dass vielfältige Faktoren die Aufgabenbearbeitungen beeinflusst haben können. In den Analysen von Aufgabenbearbeitungen in den verschiedenen Studien werden teils sehr unterschiedliche Begrifflichkeiten verwendet. Diese Begrifflichkeiten sind in den Publikationen oftmals nicht alle klar definiert, was zusammengenommen die Vergleichbarkeit verschiedener Untersuchungen einschränkt.

Die Ergebnisse der Studien zum Führen von Beweisen weisen darauf hin, dass Beweise den meisten Schülerinnen und Schülern, und insbesondere auch vielen der mathematisch vergleichsweise leistungsstarken Lernenden, nicht gelingen. Zudem geben sie Einblicke in die sehr unterschiedlichen Ausprägungen von Bearbeitungen von Aufgaben mit Beweisnotwendigkeit, wie etwa empirisch-induktive Ansätze und unvollständige Begründungen, und deren relative Häufigkeiten. Des Weiteren liefern einige Studien Erkenntnisse bezüglich der Verwendung von Beispielen in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit sowie hinsichtlich der Verwendung von Darstellungsmitteln und der Darstellungsform solcher Bearbeitungen. Systematische Analysen, in denen mehrere Eigenschaften von Beweisversuchen von Schülerinnen und Schülern klar analytisch getrennt untersucht werden, gibt es aber bisher kaum. Zudem stehen in den vorhandenen Analysen von Bearbeitungen von Aufgaben mit Beweisnotwendigkeit in der Regel eher die Defizite und Schwierigkeiten von Lernenden im Fokus als mögliche Anknüpfungspunkte für eine Weiterentwicklung der Aufgabenbearbeitungen. Systematische Analysen von Eigenschaften von Beweisversuchen von Lernenden in Verbindung mit einer primär kompetenzorientierten Sichtweise könnten mit Blick auf Förderung mögliche Handlungsperspektiven aufzeigen.

Die meisten empirischen Untersuchungen schriftlicher Bearbeitungen von Schülerinnen und Schülern zu Aufgaben mit Beweisnotwendigkeit beziehen sich auf Lernende ab (oftmals Ende) der Klassenstufe 7. Einzelne Studien, die auf Schülerinnen und Schülern in den Klassenstufen 3 bis 6 fokussieren, beschäftigen sich in einem weiten Sinne mit schriftlichen Bearbeitungen von Aufgaben mit Beweisnotwendigkeit, jedoch allenfalls ansatzweise mit Beweisversuchen zu Allaussagen mit unendlichem Gültigkeitsbereich. In allen diesen Studien mit Lernenden in den Klassenstufen 3 bis 6 ist es, zumindest in vorbereitenden Aufgabenteilen, möglich und hinreichend rein beispielgebunden zu argumentieren. Von Lernenden zu Beginn der Sekundarstufe I wurden bisher kaum Beweisversuche zu Allaussagen mit unendlichem Gültigkeitsbereich untersucht.

Ergebnisse empirischer Studien mit Schülerinnen und Schülern ab der Klassenstufe 7, in denen Allaussagen mit unendlichem Gültigkeitsbereich im Fokus stehen, deuten darauf hin, dass die Verwendung von Algebra bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit eine Hürde und Fehlerquelle sein kann. Es stellt sich die Frage, inwieweit jungen Lernenden vor der Einführung in die Algebra und auch vor der Behandlung des Beweisens von Allaussagen am Beispiel der Elementargeometrie Beweise solcher Aussagen gelingen, und welche Eigenschaften deren Beweisversuche aufweisen. Ist es möglich und sinnvoll, insbesondere mathematisch leistungsstarke Schülerinnen und Schüler bereits in der Klassenstufe fünf oder sechs in mathematisches Beweisen einzuführen, um präalgebraische, inhaltlich-anschauliche Denkweisen zu bewahren und elementare Herangehensweisen zu fördern

In den vorhandenen Studien zum Führen von Beweisen werden im Speziellen auch nicht die Kompetenzen von Teilnehmenden der Mathematik-Olympiade untersucht.[100] Die Teilnehmenden der Mathematik-Olympiade stellen eine sehr geeignete, interessante Untersuchungspopulation und der Wettbewerb einen guten, authentischen Rahmen für eine Erhebung von Beweisversuchen insbesondere auch junger, mathematisch vergleichsweise leistungsstarker Lernender dar (Kap. 2 zur Mathematik-Olympiade; Abschn. 6.1 zum Forschungsdesiderat).

5.5 Zusammenfassung des fünften Kapitels

Ergebnisse empirischer Studien zu Beweiskompetenzen von Schülerinnen und Schülern

Ergebnisse zahlreicher empirischer Untersuchungen weisen darauf hin, dass die meisten Schülerinnen und Schüler, und insbesondere auch Teile der mathematisch leistungsstarken Lernenden, beim Führen von Beweisen Schwierigkeiten haben und ihnen Beweise von Allaussagen mit unendlichem Gültigkeitsbereich nicht gelingen. Während einzelne Querschnittsstudien andeuten, dass Lernende höherer Klassenstufen Aufgaben mit Beweisnotwendigkeit im Durchschnitt ähnlich gut oder besser bearbeiten, zeigen Längsschnittstudien keine klare positive Tendenz. Weitgehend unabhängig von der Klassenstufe scheinen Bearbeitungen von Lernenden zu Aufgaben mit Beweisnotwendigkeit sehr vielfältige Ausprägungen aufzuweisen, wobei solche mit ausschließlich empirisch-induktivem Ansatz in mehreren Studien besonders häufig sind. Des Weiteren deuten Forschungsergebnisse an, dass die **Qualität von Beweisversuchen von Schülerinnen und Schülern** unter anderem mit deren Leistungen bei Aufgaben, die keine Beweiskompetenzen erfordern, deren Leistungen beim Validieren und Bewerten fiktiver Beweisversuche, einzelnen Merkmalen des fachspezifischen Interesses und der Motivation von Lernenden, dem erlebten Mathematikunterricht sowie Merkmalen der zu zeigenden Aussage zusammenhängen.

Studien belegen, dass Schülerinnen und Schüler **Beispiele** in vielfältig ausgeprägten Bearbeitungen von Aufgaben mit Beweisnotwendigkeit, unter anderem auch in ansonsten eher deduktiv gehaltenen Bearbeitungen, anführen. Sie verdeutlichen zudem, dass die Angabe von Beispielen aus sehr verschiedenen Gründen erfolgen kann. Insbesondere ist es möglich und Studien zufolge nicht unwahrscheinlich, dass Lernenden die Unzulänglichkeit einer selbst erstellten empirisch-induktiven Bearbeitung bewusst ist.

Manche Studien deuten an, dass Lernende ihre Bearbeitungen, wenn sie über einen empirisch-induktiven Ansatz hinausgehen, primär verbal darstellen. Die Verwendung von Algebra scheint länderspezifisch zu sein und eventuell von Eigenschaften der zu zeigenden Aussage abzuhängen. Die primäre Nutzung ikonischer **Darstellungsmittel** scheint eher selten vorzukommen und eventuell mit der zu zeigenden Aussage zusammenzuhängen.

[100] Eine Ausnahme bildet eine erste eigene Untersuchung, die im Rahmen der wissenschaftlichen Hausarbeit im Lehramtsstudium durchgeführt wurde. Bei einer mathematischen Akademie insbesondere für erfolgreich Teilnehmende der Mathematik-Olympiade ab der Klassenstufe 5 werden alle Schülerinnen und Schüler zur eigenständigen schriftlichen Bearbeitung von Aufgaben mit Beweisnotwendigkeit aufgefordert, um erste Einblicke in deren Kompetenzen im Führen von Beweisen zu erhalten (Tebaartz, 2013; Tebaartz, 2015; Tebaartz & Lengnink, 2015).

Insgesamt gibt es viele Studien zum Führen von Beweisen und zugleich mehrere offene Fragen, etwa bezüglich Eigenschaften von Beweisversuchen junger Lernender zu Beginn der Sekundarstufe I.

6 Forschungsdesiderat, Fragestellungen und Design

In diesem Kapitel geht es um die konkrete Planung und Durchführung der Studie zu Beweisversuchen von Teilnehmenden der Mathematik-Olympiade. Es werden die Fragestellungen dieser Studie erläutert (Abschn. 6.1) sowie die Anlage der Studie und die in ihr verwendeten Untersuchungsinstrumente eingeführt (Abschn. 6.2).

6.1 Forschungsdesiderat und Fragestellungen

Die vorliegende Arbeit beschreibt, entsprechend den Ausführungen in Abschnitt 1.1, einen Ausschnitt eines Forschungsprojekts, das aus einer ehrenamtlichen Tätigkeit im Kontext der Mathematik-Olympiade heraus motiviert wurde. Für die Dissertationsschrift wurde aus diesem Forschungsprojekt ein Teilaspekt begründet ausgewählt, der an vorhandene empirische Studien anknüpft und sich zugleich als Forschungsdesiderat beschreiben lässt. Konkret handelt es sich bei der ausgewählten (Teil)Studie aus zwei Perspektiven um ein Forschungsdesiderat, zum einen mit Blick auf die Mathematik-Olympiade, zum zweiten vor dem Hintergrund der mathematikdidaktischen Forschung zum Führen von Beweisen. Im Folgenden werden diese beiden Forschungsdesiderate zunächst getrennt erläutert und, davon ausgehend, zur Fragestellung der vorliegenden Arbeit kombiniert. Dabei sind die nachfolgenden Ausführungen nicht neu. Sie fassen die vorhergehenden Kapitel 1 bis 5 aus den genannten zwei Perspektiven zusammen, um die Forschungsdesiderate, die diese Dissertation bearbeitet, und die daraus resultierenden Fragestellungen darzustellen.

Die **Mathematik-Olympiade** in Deutschland ist ein traditionsreicher Stufenwettbewerb für mathematisch interessierte Schülerinnen und Schüler. Sie ist strukturell und institutionell fest verankert, und stellt mit bis zu 200 000 Teilnehmerinnen und Teilnehmern pro Schuljahr einen großen Mathematik-Wettbewerb für Schülerinnen und Schüler in Deutschland dar. Ein wesentliches Ziel der Mathematik-Olympiade ist es, leistungsstarke und potenziell besonders leistungsfähige Kinder und Jugendliche zu finden, zu fördern und zu einem mathematiknahen Studium zu motivieren (Bauermann, 2018; Mathematik-Olympiaden e. V., o. J.; Kap. 2). Dies ergänzend hebt die Kultusministerkonferenz in ihrer Förderstrategie für leistungsstarke und leistungsfähige Schülerinnen und Schüler die Bedeutung von Wettbewerben wie der Mathematik-Olympiade folgendermaßen hervor:

> Schülerwettbewerbe […] tragen dazu bei, frühzeitig das Interesse an bestimmten Fachrichtungen zu wecken, Talente zu finden und von der Breitenmotivation zur Förderung besonderer Fähigkeiten zu kommen. Wettbewerbe leisten einen wichtigen Beitrag zur Gewinnung von Fachkräften und akademischem Nachwuchs, zum Beispiel in den MINT-Fächern und in den Ingenieurwissenschaften. (KMK, 2015, S. 6)

Insgesamt betrachtet ist die Mathematik-Olympiade in Deutschland nicht mehr wegzudenken. Trotzdem steht die mathematikdidaktische Forschung zur Mathematik-Olympiade noch am

Anfang. Der Autorin der vorliegenden Arbeit sind diesbezüglich ausschließlich forschungsbasierte Arbeiten aus Lettland bekannt. Veilande et al. (2018) analysieren insbesondere schriftliche Bearbeitungen von Problemlöseaufgaben der Open Mathematical Olympiads of Latvia aus der Klassenstufe 9. Sie kommen zu dem grundlegenden Ergebnis, dass den meisten Teilnehmenden keine vollständigen Lösungen gelingen. Unvollständige Lösungen können vielfältige Ursachen haben. Besonders häufig haben die Teilnehmenden aus der Klassenstufe 9 nach Veilande et al. (2018) Schwierigkeiten im Umgang mit Algebra. Zudem werden die Teilnehmenden sowohl bei der Mathematik-Olympiade in Lettland als auch bei der in Deutschland aufgefordert, vorgegebene Aussagen und selbstständig erarbeitete Lösungen zu beweisen. Die Ergebnisse von Veilande et al. (2018) lassen offen, inwieweit die unvollständigen Lösungen auch auf Schwierigkeiten beim Führen von Beweisen zurückzuführen sind. Kompetenzen im Führen von Beweisen sind eine zentrale Voraussetzung für Erfolg im Wettbewerb (Abschn. 2.3.2; 10.3).

Die Mathematik-Olympiade ist somit ein interessantes, bisher jedoch kaum beachtetes Forschungsfeld. Da die Anforderung, mathematische Aussagen zu beweisen, ein Charakteristikum dieses Wettbewerbs darstellt, erscheint es als sinnvoll, in einer Untersuchung zur Mathematik-Olympiade das Führen von Beweisen zu fokussieren. Zudem ist das Führen von Beweisen über die Mathematik-Olympiade hinaus von besonderer Relevanz.

Das **Führen von Beweisen** ist für die Sicherung, Systematisierung, Erklärung, Entdeckung und Kommunikation mathematischen Wissens von großer Bedeutung und unterscheidet die Wissenschaft Mathematik von vielen anderen Wissenschaften (u. a. Davis & Hersh, 1986; de Villiers, 1990; Heintz, 2000; Abschn. 1.1, 3.1). Außerdem sind Beweise und die Tätigkeit des Beweisens beim Mathematiklernen wichtig. Sie können das Lernen durch Verstehen fördern, das Ordnen, Vernetzen und Vertiefen mathematischen Wissens unterstützen und z. B. zu einem authentischen Bild der Wissenschaft Mathematik beitragen (z. B. Hanna et al., 2012; Malle, 2002; Reiss, 2009; Abschn. 1.1).

In den Bildungsstandards wurde das Beweisen unter dem Oberbegriff „Argumentieren“ (KMK, 2004, S. 8) bzw. „Mathematisch argumentieren“ (KMK, 2003, S. 8; KMK, 2012, S. 15) als allgemeine mathematische Kompetenz aufgenommen. Lernende sollen am Ende der vierten Klassenstufe Beispiele und Gegenbeispiele angeben können (KMK, 2004; Abschn. 4.2). Das Beweisen von Allaussagen wird üblicherweise am Beispiel der Elementargeometrie im Schulunterricht der Klassenstufe 7 thematisiert. Schülerinnen und Schüler sollen am Ende der Sekundarstufe I mehrschrittige Argumentgefüge und am Ende der Sekundarstufe II durchaus auch formale Beweise entwickeln und formulieren können (KMK, 2003, 2012; Abschn. 4.2).

In zahlreichen mathematikdidaktischen Studien werden Bearbeitungen von Aufgaben mit Beweisnotwendigkeit untersucht (Kap. 5; z. B. Bell, 1976; Brunner et al., 2022; Chin & Lin, 2009; Knuth et al., 2002; Healy & Hoyles, 2000; Moya et al., 2015; Porteous, 1990; Reiss et al., 2002). Die Ergebnisse der Studien weisen darauf hin, dass Beweise den meisten Schülerinnen und Schülern, und insbesondere auch vielen der mathematisch vergleichsweise leistungsstarken Lernenden, nicht gelingen. Zusätzlich geben sie Einblicke in vorkommende Ausprägungen und Charakteristika von Bearbeitungen von Aufgaben mit Beweisnotwendigkeit. Systematische

Analysen, in denen mehrere Eigenschaften von Beweisversuchen von Schülerinnen und Schülern, wie deren Korrektheit und Vollständigkeit, vorkommende Beweisideen und die Darstellungsform, analytisch getrennt untersucht werden, gibt es jedoch kaum.

In einigen der vorhandenen Untersuchungen werden andere Aufgabenformate mit Beweisnotwendigkeit, teils ohne explizite Aufforderung, einen Beweis zu formulieren, verwendet, so dass vielfältige Faktoren die Aufgabenbearbeitungen beeinflussen können. Außerdem beziehen sich viele der Untersuchungen frühestens auf Schülerinnen und Schüler ab der Klassenstufe 7. Beweisversuche von Schülerinnen und Schülern in den Klassenstufen 5, 6 wurden bisher kaum untersucht. Erkenntnisse diesbezüglich würden somit den bisherigen Forschungsstand ergänzen. Sie erscheinen zudem im Vergleich zu Beweisversuchen von Lernenden in höheren Klassenstufen interessant. Während vorhandene Forschungsergebnisse darauf hindeuten, dass die Verwendung von Algebra bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit durch Schülerinnen und Schüler eine Hürde und Fehlerquelle sein kann, sind Lernende in den Klassenstufen 5 und 6 i. d. R. noch nicht mit der Verwendung von Algebra vertraut. Während zumindest Lernenden ab der Klassenstufe 8 einzelne Beweise elementargeometrischer Allaussagen bekannt sein sollten, wird im Schulunterricht in den Klassenstufen 5, 6 i. d. R. das Beweisen von Allaussagen nicht explizit thematisiert. Eine Analyse von Beweisversuchen von Lernenden etwa in den Klassenstufen 5 bis 8 würde es ermöglichen, Eigenschaften von Beweisversuchen dieser Lernenden mit unterschiedlichen Voraussetzungen und Vorkenntnissen zu vergleichen. Diesbezügliche Ergebnisse wären zusätzlich zur mathematikdidaktischen Theoriebildung auch mit Blick auf die Förderung der Teilkompetenz des Führens von Beweisen interessant. Die Forschung würde so eine Basis schaffen, auf der ein Förderkonzept entwickelt werden kann, das bereits in den Klassenstufen 5, 6 in mathematisches Beweisen einführt und präalgebraische, inhaltlich-anschauliche Denkweisen sowie elementare Herangehensweisen unterstützt.

In der vorliegenden Arbeit werden das Desiderat einer Untersuchung im Forschungsfeld der **Mathematik-Olympiade** und das Desiderat einer vergleichenden Analyse von **Beweisversuchen** von Lernenden in den Klassenstufen 5 bis 8 kombiniert. Diese Kombination erscheint aus beiden Perspektiven als sinnvoll, zum einen da die Anforderung, mathematische Aussagen zu beweisen, ein Charakteristikum der Mathematik-Olympiade darstellt und zum anderen da dieser Wettbewerb einen sehr geeigneten, authentischen Rahmen für eine Erhebung von Beweisversuchen aus allen Klassenstufen 5 bis 8 darstellt.

In der vorliegenden Arbeit geht es also darum, Beweisversuche von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 systematisch vergleichend zu untersuchen. Dabei stellt sich grundlegend die Frage, inwieweit diesen Lernenden ein Beweis gelingt. Zudem ist es von Interesse, welche Charakteristika deren Beweisversuche, in Bezug auf die Beispielnutzung, die Beweisideen und die vorkommenden Darstellungsmittel, aufweisen. Außerdem soll explizit ein Augenmerk darauf liegen, inwieweit Anknüpfungspunkte für eine Weiterentwicklung derer Beweisversuche zu einem Beweis in den Beweisversuchen der Schülerinnen und Schüler zu erkennen sind, um Einblicke in Ansatzpunkte für eine Förderung im Führen von Beweisen zu gewinnen.

Für diese systematische vergleichende Untersuchung von Beweisversuchen werden, da Allaussagen mit unendlichem Gültigkeitsbereich für die Mathematik besonders charakteristisch sind, Beweisversuche zu einer Allaussage mit unendlichem Gültigkeitsbereich fokussiert. Gerade auch mit Blick auf junge Schülerinnen und Schüler zu Beginn der Sekundarstufe I wird eine Implikation, d. h. die einfachste Form einer Allaussage, als zu beweisende Aussage ausgewählt. Nebenbei stehen auch in vielen anderen Studien zum Führen von Beweisen Allaussagen mit unendlichem Gültigkeitsbereich, meistens Implikationen, im Fokus (Kap. 5; z. B. Healy & Hoyles, 2000; Knuth et al., 2002; Portous, 1990), was für eine Einordnung der Ergebnisse der vorliegenden Arbeit günstig ist.

Fachinhaltlich erscheint für die vergleichende Untersuchung der Themenbereich der Teilbarkeit ganzer Zahlen geeignet. In diesem Themenbereich besteht die Möglichkeit, auf elementarem Niveau Beweise zu führen. Wissenselemente aus diesem Themenbereich werden zum Teil in der Grundschule eingeführt und danach kaum vertieft. Zudem kommen bei der Mathematik-Olympiade Aufgaben aus dem Themenbereich der Teilbarkeit ganzer Zahlen regelmäßig in allen Klassenstufen 5 bis 8 vor (Abschn. 1.1, 10.3).

Diese beschriebenen Überlegungen zusammengenommen, lauten die **Forschungsfragen** der vorliegenden Arbeit, entsprechend den Angaben in Abschnitt 1.1:

1. Inwieweit gelingt es Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 eine mathematische Allaussage mit unendlichem Gültigkeitsbereich aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen, welche Charakteristika zeichnen deren Beweisversuche aus und inwieweit sind Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen?
2. Welche Gemeinsamkeiten und Unterschiede bestehen diesbezüglich zwischen Beweisversuchen von Teilnehmenden der Mathematik-Olympiade mit Blick auf verschiedene Klassenstufen?

6.2 Design der Studie

Um die Forschungsfragen dieser Arbeit zu untersuchen, wurde im Rahmen der Regionalrunde der Mathematik-Olympiade in Nordrhein-Westfalen im November 2016 eine Studie durchgeführt. Sie umfasst eine Beweisaufgabe als Teil der regulären Wettbewerbsklausur und eine schriftliche Befragung direkt im Anschluss an die Klausur. Im Folgenden wird die Anlage der Studie näher beschrieben und es werden die in ihr verwendeten Untersuchungsinstrumente begründet. Ein besonderer Fokus liegt dabei auf der Aufgabe, die zur Erhebung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade verwendet wurde. Die Konstruktion dieser Aufgabe wird im Anschluss an den Überblick über die Anlage und die Untersuchungsinstrumente der Studie erläutert.

6.2.1 Anlage und Untersuchungsinstrumente der Studie

Grundlegendes Charakteristikum der Studie ist deren Durchführung im Rahmen einer regulär stattfindenden Mathematik-Olympiade. Auf diese Weise war es möglich, authentisches Datenmaterial von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade zu erhalten.

Die Mathematik-Olympiade umfasst in allen Klassenstufen 5 bis 8 eine Schul-, eine Regional- und eine Landesrunde, wovon die Regionalrunde, d. h. die zweite Wettbewerbsstufe, für die Datenerhebung gewählt wurde. Die Regionalrunde ist am besten geeignet, um Durchführungsobjektivität zu gewährleisten und zugleich ein großes Spektrum unterschiedlich leistungsstarker Teilnehmerinnen und Teilnehmer einzubeziehen. Sie ist im Vergleich zur Schulrunde als Klausurwettbewerb konzipiert und ermöglicht dadurch sicherzustellen, dass alle Schülerinnen und Schüler die Beweisaufgabe eigenständig und in einem begrenzten Zeitfenster bearbeiten. Zugleich ist das Spektrum an Teilnehmerinnen und Teilnehmern im Rahmen der Regionalrunde im Vergleich zur darauffolgenden Landesrunde sehr groß. Es reicht von Lernenden, die durchaus mathematisch interessiert sind, aber noch keine Förderung in Mathematik erhalten haben, bis hin zu besonders leistungsstarken Lernenden, die zudem an Spitzenförderprogrammen, wie Jugend trainiert Mathematik, teilnehmen (Abschn. 2.2).

In der Regionalrunde der Mathematik-Olympiade werden in allen Klassenstufen 5 bis 8 regulär Aufgaben mit Beweisnotwendigkeit gestellt. Trotzdem wurde für die Erhebung von Beweisversuchen im Rahmen dieser Dissertation eine spezifisch fokussierte Beweisaufgabe konstruiert, die den Teilnehmerinnen und Teilnehmern der Regionalrunde der Mathematik-Olympiade zur Bearbeitung vorgelegt wurde. Dies erwies sich aus folgenden Gründen als nötig: Da sich die in der Regionalrunde regulär gestellten Aufgaben deutlich zwischen den Klassenstufen unterscheiden, ist ein Vergleich von Beweisversuchen von Lernenden in den Klassenstufen 5 bis 8 anhand des üblichen Aufgabenmaterials schwierig. Zudem umfassen viele Wettbewerbsaufgaben zusätzlich zum Beweisen einen Entdeckungsauftrag, der die Analyse und Interpretation von Aufgabenbearbeitungen mit Fokus auf die Bewältigung der Beweisanforderungen verzerren könnte. Darüber hinaus bietet die Konstruktion einer spezifisch fokussierten Beweisaufgabe die Chance, eine für den Themenbereich der Teilbarkeit ganzer Zahlen charakteristische Aufgabe zu formulieren sowie gezielt bestimmte Facetten mathematischen Beweisens zu untersuchen und einige Schwierigkeitsparameter zu kontrollieren, die die Bearbeitung einer Aufgabe beeinflussen können (Abschn. 6.2.2).

Zusätzlich zu dieser spezifisch fokussierten Beweisaufgabe wurde ein kurzer Fragebogen erstellt. Er umfasst personenbezogene Daten, wie etwa Fragen zur Teilnahme an mathematischen Förderangeboten und an der Mathematik-Olympiade, um eine differenzierte Betrachtung von Auffälligkeiten in den Beweisversuchen der Lernenden zu ermöglichen (Abschn. 10.4). Mit dem Ziel, eine angemessene Interpretation der Beweisversuche der Lernenden zu unterstützen, umfasst der Fragebogen außerdem Aufgaben zum Validieren und Bewerten fiktiver Beweisversuche. Von diesen Aufgaben wurden zwei Versionen erstellt. Die eine Version fokussiert auf Beweisversuche zu Allaussagen, die andere auf Beweisversuche zu Existenzaussagen (Abschn. 10.2). Um die Bearbeitung des Fragebogens mit ca. 15 Minuten Bearbeitungszeit in einem zumutbaren Rahmen zu halten, wurde zu gleichen Teilen in jeden Fragebogen nur eine

der beiden Versionen aufgenommen. Es wurde festgelegt, dass die Teilnehmenden den Fragebogen direkt im Anschluss an die Klausur erhalten, auch wenn vor allem bei den jungen Lernenden nach der Wettbewerbsklausur mit gewissen Ermüdungserscheinungen gerechnet werden muss. Das Austeilen des Fragebogens direkt im Anschluss an die Klausur hilft zu gewährleisten, dass alle Teilnehmenden einen Fragebogen erhalten und eigenständig bearbeiten. Die Bearbeitung des Fragebogens wurde nicht vor, sondern nach der Wettbewerbsklausur angesetzt, damit die fiktiven Beweisversuche, die im Fragebogen abgebildet sind, die Bearbeitung der Beweisaufgabe innerhalb der Wettbewerbsklausur nicht beeinflussen.

Neben der Konstruktion der spezifisch fokussierten Beweisaufgabe und des Fragebogens waren einige organisatorische Vorbereitungen erforderlich. Bereits ein Jahr im Voraus wurde die Durchführung der Datenerhebung mit den Vorsitzenden des Landesverbands Mathematikwettbewerbe NRW e. V. abgesprochen. Für die Datenerhebung wurde das Bundesland Nordrhein-Westfalen (NRW) gewählt, da die Autorin dieser Arbeit durch ihre ehrenamtlichen Tätigkeiten mit den dortigen Strukturen gut vertraut ist und sich die vorhandenen Strukturen für die Durchführung einer solchen Erhebung gut eignen. Zugleich sind die Teilnehmerzahlen mit etwa 5000 bis 7000 Schülerinnen und Schülern der Klassenstufen 5 bis 13 in der Regionalrunde einer jeden Mathematik-Olympiade in NRW so hoch, dass auf den Einbezug weiterer Bundesländer verzichtet werden konnte (Abschn. 2.2.2 zur Mathematik-Olympiade in NRW).

Für die Durchführung der Studie in verschiedenen Regionen Nordrhein-Westfalens war es notwendig, in jeder Klassenstufe eine der regulären Wettbewerbsaufgaben, die vom Aufgabenausschuss des Mathematik-Olympiaden e. V. herausgegeben werden, durch die für diese Studie spezifisch fokussierte Beweisaufgabe zu ersetzen. Dabei wurden die regionalen Besonderheiten, wie z. B. die Klausurlänge, berücksichtigt und darauf geachtet, die für diese Studie konzipierte Beweisaufgabe möglichst als zweite Aufgabe in die Klausur einzufügen, damit die Bearbeitung tendenziell nicht aus zeitlichen oder motivationalen Gründen beendet wird. Es sollte sichergestellt werden, dass sich die Lernenden tendenziell frühzeitig mit der Beweisaufgabe auseinandersetzen, aber vorab Erfolgserlebnisse bei einer leichteren Aufgabe sammeln können. Um eine möglichst objektive Durchführung zu gewährleisten, wurden die Regionalkoordinatorinnen und -koordinatoren weiterhin darum gebeten, in der Klausur keine Fragen zu beantworten, die sich auf die Bearbeitung der Beweisaufgabe beziehen, und Lösungen frühestens am Samstagabend, d. h. nach Abschluss der Regionalrunde in allen Regionen Nordrhein-Westfalens herauszugeben. In den zwei teilnehmerreichsten Regionen wurde die Durchführung der Erhebung vor Ort durch die Autorin dieser Arbeit unterstützt. In den meisten anderen Regionen wurden die Fragebögen den Regionalkoordinatorinnen und -koordinatoren per Post übermittelt und der Rücklauf persönlich abgesprochen.

Insgesamt haben 44 der 55 Regionen Nordrhein-Westfalens an der Studie teilgenommen, dies entspricht 80%. Bezogen auf die Regionen, in denen regelmäßig über 100 Schülerinnen und Schüler an der Regionalrunde teilnehmen, war die Quote noch höher. Dies stellte sicher, dass ein großes Spektrum an Teilnehmenden der Mathematik-Olympiade und insbesondere auch einzelne besonders erfolgreiche Schülerinnen und Schüler in der Studie berücksichtigt werden.

6.2.2 Aufgabenkonstruktion

Für die Datenerhebung wurde eine spezifisch fokussierte Beweisaufgabe in die Wettbewerbsklausur der Regionalrunde 2016 der Mathematik-Olympiade in NRW implementiert (Abschn. 6.2.1). In diesem Abschnitt geht es darum, die Konstruktion der Aufgabe vorzustellen.

Um bei der Konstruktion dieser Beweisaufgabe einige Schwierigkeitsparameter gezielt kontrollieren zu können, wurde vorab herausgearbeitet, welche schwierigkeitsgenerierenden Merkmale die Anforderungen, die strukturell in Aufgabenstellungen angelegt sind, beeinflussen können. Zudem wurden anhand ausgewählter Aufgaben der Mathematik-Olympiade die regulären Anforderungen, die im Bereich des Führens von Beweisen gestellt werden, untersucht, mit dem Ziel, die eigene Beweisaufgabe in Einklang mit den Anforderungen in den vorhergehenden Wettbewerben zu konstruieren.[101]

Aufbauend auf der Analyse schwierigkeitsgenerierender Merkmale und der regulären Wettbewerbsanforderungen wurden für die Erhebung von Beweisversuchen zunächst zwei Versionen einer Beweisaufgabe entwickelt. Diese wurden anschließend von Juni bis Juli 2016 in vier hessischen Schulen, im MatheClub Wetzlar und im Rahmen eines Förderangebots an der Justus-Liebig-Universität Gießen mit rund achtzig mathematisch interessierten, unterschiedlich leistungsstarken Lernenden aus den Klassenstufen 5 bis 8 pilotiert. Die Ergebnisse aus der Pilotierung wurden genutzt, um die finale Aufgabe auszuwählen und für den Einsatz im Wettbewerb zu optimieren. Im Folgenden werden, anschließend an einen Überblick über die konzipierte Beweisaufgabe, objektive Anforderungsstrukturen[102] und mögliche Bearbeitungen des Aufgabenteils d) en détail präsentiert, denn genau die Beweisversuche zum Aufgabenteil d) werden im Rahmen dieser Arbeit analysiert (Abschn. 1.1, 6.1 zu den Forschungsfragen und Fokussierungen).

6.2.2.1 Überblick über die Aufgabe

In Abbildung 74 ist die konzipierte Beweisaufgabe für Teilnehmende der Mathematik-Olympiade in den Klassenstufen 5, 6 bzw. 7, 8 zu sehen. Die Aufgabe besteht aus vier Aufgabenteilen a) bis d), in denen unterschiedliche Facetten mathematischen Beweisens fokussiert werden. So sind in den Aufgabenteilen a), b) Existenzaussagen und in den Aufgabenteilen c), d) Allaussagen zu zeigen.

Zusätzlich unterscheiden sich die Aufgabenteile hinsichtlich des Gültigkeitsbereichs und der Abstraktheit der zu zeigenden Aussage, besonders hilfreicher Heurismen und der Komplexität des zu formulierenden Argumentgefüges. Diesbezüglich wurden gezielt Schwierigkeiten implementiert, um die Kompetenzen von Lernenden hinsichtlich dieser für Beweisaufgaben charakteristischen Anforderungen zu erfassen und untersuchen zu können. Zugleich wurden dafür die Anforderungen in anderen Bereichen niedrig gehalten, wie nachfolgend beschrieben wird.

[101] Die Anforderungen, die bei der Mathematik-Olympiade im Bereich des Führens von Beweisen gestellt werden, und die schwierigkeitsgenerierenden Merkmale wurden im Rahmen des dieser Arbeit übergeordneten Forschungsprojekts untersucht (Abschn. 1.1, 10.2, 10.3).

[102] Der Ausdruck „objektiv“ soll in Anlehnung an Bruder (1981, 1985) verdeutlichen, dass es um Anforderungen geht, die in der Aufgabe angelegt sind. Der tatsächliche, subjektive Schwierigkeitsgrad der Aufgabe hängt u. a. vom Ausgangsniveau sowie den Vorerfahrungen der Lernenden ab und kann sich individuell unterscheiden.

Für die Studie konzipierte Beweisaufgabe

Klassenstufen 5, 6

André, Björn, Claudia und Defne experimentieren mit natürlichen Zahlen einschließlich der Null, also mit Zahlen wie $0, 1, 2, 3, 4, 5, 6, 7,$ usw.

a) André behauptet: „Es gibt eine **dreistellige** Zahl, die beide Bedingungen (1) und (2) erfüllt.
   (1) Wenn man die drei Ziffern der Zahl multipliziert, dann ist das Ergebnis 18.
   (2) Wenn man die drei Ziffern der Zahl addiert, dann ist das Ergebnis 12."
   Hilf André zu zeigen, dass seine Behauptung richtig ist.
   Hinweis: Dreistellige Zahlen (z. B. die Zahl 702) bestehen aus drei Ziffern (z. B. hier aus der Hunderterziffer 7, Zehnerziffer 0, Einerziffer 2). Bei einer dreistelligen Zahl ist die Hunderterziffer nie Null.

b) Björn interessiert sich für Quadratzahlen. Er behauptet: „Wenn man von einer **Quadratzahl** eins subtrahiert, dann erhält man immer eine ungerade Zahl." Aber André ist skeptisch. Hilf ihm zu zeigen, dass Björns Behauptung falsch ist.
   Hinweis: „Quadratzahl" heißt jede Zahl, die sich als Produkt zweier gleicher natürlicher Zahlen darstellen lässt. Zum Beispiel gilt $16 = 4 \cdot 4$. Also ist 16 eine Quadratzahl.

c) Claudia sagt: „2016 ist die **kleinste vierstellige** Zahl, die alle folgenden Bedingungen erfüllt:
   (1) Alle Ziffern sind verschieden.
   (2) Die Einerziffer ist dreimal so groß wie die Tausenderziffer.
   (3) Wenn man die Einer- und Tausenderziffer multipliziert, dann ist das Ergebnis durch 4 teilbar."
   Hilf Claudia zu begründen, warum ihre Behauptung richtig ist.

d) Defne berechnet die Summe von vier aufeinanderfolgenden ungeraden Zahlen (z. B. $5 + 7 + 9 + 11 = 32$). Sie behauptet: „Die Summe von **vier aufeinanderfolgenden ungeraden** Zahlen ist immer durch 4 teilbar." Hilf Defne zu begründen, warum ihre Behauptung richtig ist.

Klassenstufen 7, 8

Ali, Björn, Claudia und Denise experimentieren mit natürlichen Zahlen einschließlich der Null.

a) Ali behauptet: „Es gibt eine **vierstellige** Zahl, die beide Bedingungen (1) und (2) erfüllt.
   (1) Wenn man die vier Ziffern der Zahl multipliziert, dann ist das Ergebnis 108.
   (2) Wenn man die vier Ziffern der Zahl addiert, dann ist das Ergebnis 18."
   Hilf Ali zu zeigen, dass seine Behauptung richtig ist.
   Hinweis: Vierstellige Zahlen (z. B. die Zahl 3027) bestehen aus vier Ziffern (z. B. hier aus der Tausenderziffer 3, Hunderterziffer 0, Zehnerziffer 2, Einerziffer 7). Bei einer vierstelligen Zahl ist die Tausenderziffer nie Null.

b) Björn interessiert sich für Quadratzahlen. Er behauptet: „Wenn man von einer **Quadratzahl** eins subtrahiert, dann erhält man immer eine ungerade Zahl." Aber Ali ist skeptisch. Hilf ihm zu zeigen, dass Björns Behauptung falsch ist.
   Hinweis: „Quadratzahl" heißt jede Zahl, die sich als Produkt zweier gleicher natürlicher Zahlen darstellen lässt. Zum Beispiel gilt $16 = 4 \cdot 4$. Also ist 16 eine Quadratzahl.

c) Claudia sagt: „2016 ist die **kleinste vierstellige** Zahl, die alle folgenden Bedingungen erfüllt:
   (1) Alle Ziffern sind verschieden.
   (2) Die Einerziffer ist dreimal so groß wie die Tausenderziffer.
   (3) Wenn man die Einer- und Tausenderziffer multipliziert, dann ist das Ergebnis durch 4 teilbar."
   Hilf Claudia zu begründen, warum ihre Behauptung richtig ist.

d) Denise berechnet die Summe von vier aufeinanderfolgenden ungeraden Zahlen (z. B. $5 + 7 + 9 + 11 = 32$). Sie behauptet: „Die Summe von **vier aufeinanderfolgenden ungeraden** Zahlen ist immer durch 8 teilbar." Hilf Denise zu beweisen, dass ihre Behauptung richtig ist.

Abbildung 72: Für die Studie konzipierte Beweisaufgabe für die Klassenstufen 5 bis 8

Entsprechend den Forschungsfragen dieser Arbeit und des übergeordneten Forschungsprojekts gehört die gesamte Aufgabe zum Themenbereich der Teilbarkeit ganzer Zahlen. Dabei wurde der Gültigkeitsbereich explizit auf natürliche Zahlen eingegrenzt, weil nur diese den Lernenden ab der Klassenstufen 5 sicher bekannt sind. Um Unklarheiten vorzubeugen, wurde für Lernende in den Klassenstufen 5, 6 der Bereich der natürlichen Zahlen in der Aufgabenstellung kurz definiert, während dies in den Klassenstufen 7, 8 nach den Erfahrungen aus der Pilotierung nicht notwendig ist. Als natürliche Zahl wird entsprechend der gültigen DIN-Norm 5473 auch die Zahl Null angesehen.

Um die Anforderungen bei der Decodierung der Aufgabenstellung niedrig zu halten, wurden auch weitere Fachausdrücke wie „Quadratzahl“ und „vier aufeinanderfolgende ungerade Zahlen“, die sich in der Pilotierung häufig als schwierig erwiesen, in der Aufgabenstellung erklärt oder exemplifiziert. Ansonsten wurden primär Fachbegriffe wie „multiplizieren“ und „Ziffer“ verwendet, die den Lernenden aller Klassenstufen 5 bis 8 aus dem Schulunterricht bekannt sein sollten. Zusätzlich wurde versucht, mögliche lexikalisch-morphologische, syntaktische und textlinguistische Hürden zu vermeiden (Prediger, 2013 zu sprachlich bedingten Hürden). Inhaltlich wichtige Begriffe wurden mittels optischer Markierungen hervorgehoben. In der Aufgabenstellung kommen keine symbolischen Darstellungsmittel vor. Als Operatoren wurden die Begriffe „zeigen“ und „begründen“ (Kl. 5, 6) bzw. „zeigen“, „begründen“ und „beweisen“ (Kl. 7, 8) gewählt, weitgehend analog zu den vorhergehenden Mathematik-Olympiaden.

Damit die einzelnen Aufgabenteile a) bis d) unabhängig voneinander bearbeitet werden können und die Beschäftigung mit einzelnen Aufgabenteilen möglichst keine Vor-/Nachteile für die Bearbeitung eines anderen Aufgabenteils mit sich bringt, beziehen sich die einzelnen Teile a) bis d) auf unterschiedliche Teilmengen und Besonderheiten natürlicher Zahlen. Auch die vorkommenden Fachtermini und die für die Bearbeitung notwendigen Wissenselemente wurden möglichst inkongruent gewählt, wenngleich etwa Wissenselemente zur Parität von Summen ungerader und gerader Zahlen insbesondere bei der der Bearbeitung des Aufgabenteils b), aber auch bei den anderen Aufgabenteilen helfen können.

Um trotz dieser Unterschiedlichkeit und Unabhängigkeit der Aufgabenteile die Aufgabe kohärent und ansprechend zu gestalten sowie die Arbeitsaufträge zu motivieren, wurde die Aufgabe leicht eingekleidet. Bei der Einkleidung wurde darauf geachtet, dass sie die Aufgabenbearbeitung nicht erschwert. Insbesondere ist durch die Einkleidung weder der Einbezug von Wissenselementen aus der außermathematischen Erfahrungswelt noch eine Verknüpfung von Wissenselementen aus dem Bereich der Teilbarkeit ganzer Zahlen mit solchen aus anderen mathematischen Themenbereichen erforderlich. Insgesamt wurden somit die Anforderungen, die aus der Formulierung der Aufgabenstellung resultieren, niedrig und für Lernende aller Klassenstufen 5 bis 8 möglichst gleich gehalten.

6.2.2.2 Der Aufgabenteil d)

Dies gilt insbesondere auch für den Aufgabenteil d). Da in der vorliegenden Arbeit Beweisversuche zum Aufgabenteil d) untersucht werden, werden im Folgenden objektive Anforderungsstrukturen und mögliche Bearbeitungen dieses Aufgabenteils en détail ausgeführt.

Im Aufgabenteil d) ist zu beweisen, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen durch vier (Kl. 5, 6) bzw. durch acht (Kl. 7, 8) teilbar ist (Abb. 74).[103] Die Teilnehmenden in den Klassenstufen 7, 8 werden zum anspruchsvolleren Beweis der Teilbarkeit durch acht aufgefordert, u. a. weil die Bearbeitung der Aufgabe ansonsten für Lernende in der Klassenstufe 8 durch die Kenntnis von Algebra prinzipiell leicht wäre, wie aus den nachfolgenden Ausführungen hervorgeht. Andersherum die Teilnehmenden in den Klassenstufen 5, 6 zum Beweis der Teilbarkeit durch acht aufzufordern, ließe sich im Vergleich zu den curricularen Vorgaben und den regulären Wettbewerbsanforderungen nicht vertreten. Daher unterscheidet sich der Aufgabenteil d) in den Klassenstufen 5, 6 und 7, 8 in der dargelegten Form. Bei der nachfolgenden Beschreibung objektiver Anforderungsstrukturen und möglicher Bearbeitungen des Aufgabenteils d) werden die beiden Varianten dieses Aufgabenteils gemeinsam betrachtet und relevante Unterschiede dargelegt.

Aussagetyp und Gültigkeitsbereich

Im Aufgabenteil d) ist in allen Klassenstufen 5 bis 8 eine wahre Allaussage mit unendlichem Gültigkeitsbereich zu beweisen. Im Vergleich zum Aufgabenteil c), in dem es um eine wahre Allaussage mit endlichem Gültigkeitsbereich geht, ist der Aufgabenteil d) hinsichtlich des Gültigkeitsbereichs und damit der Abstraktheit der zu zeigenden Aussage anspruchsvoller. Die Anforderung, eine wahre Allaussage mit unendlichem Gültigkeitsbereich zu beweisen, geht in den Klassenstufen 5, 6 und zu Beginn der Klassenstufe 7 über die in den curricularen Vorgaben für NRW formulierten Kompetenzerwartungen hinaus, wenngleich Lernende mit dieser Anforderung in der Mathematik-Olympiade ab der fünften Klassenstufe konfrontiert werden (Abschn. 2.3.2, 4.2.2).

Decodierung der Aufgabenstellung

Damit es für alle Wettbewerbsteilnehmenden klar erkennbar ist, dass es sich um eine Allaussage handelt, wurde der Begriff „immer“ in der Aufgabenstellung verwendet. Um eine korrekte Interpretation der Aufgabenstellung zusätzlich zu unterstützen, wurde der zentrale Ausdruck „vier aufeinanderfolgende ungerade Zahlen“ mittels optischer Markierungen hervorgehoben und an einem Beispiel illustriert. Als Beispiel wurde ein einfach zu berechnendes Beispiel, aber nicht das kleinste Beispiel gewählt, um keinen wesentlichen Bestandteil eines möglichen Beweises vorzugeben. Das Beispiel wurde nicht vollständig ausgeführt, um keine Beziehung zwischen den Summanden und dem Ergebnis der Division der Summe dieser Zahlen hervorzuheben und dadurch die Beweisfindung zu erleichtern.

Mögliche Beweisideen und die Entwicklung, Formulierung von Argumentgefügen

Für die Entwicklung eines Beweises zu der zu zeigenden Aussage ist die Verwendung eines heuristischen Prinzips, wie z. B. des Invarianzprinzips, oder eine Reformulierung in der Darstellung

[103] Die in den Klassenstufen 7, 8 im Aufgabenteil d) zu zeigende Aussage wird auch von Esther Brunner (2013) verwendet. Brunner (2013) befasst sich insbesondere mit der Unterrichtspraxis von Lehrkräften im Bereich des Beweisens. Individuelle Beweiskompetenzen von Schülerinnen und Schülern stehen dabei weniger im Fokus.

erforderlich. Dabei sind vielfältige primäre Beweisideen als Grundlage eines korrekten und vollständigen Argumentgefüges möglich (Abschn. 3.3.3). Ausgehend von jeder der **primären Beweisideen 1 bis 8 und 11** lässt sich die vorgegebene Aussage beweisen. Im Folgenden werden bezogen auf jede dieser primären Beweisideen mögliche Herangehensweisen und Schwierigkeiten beschrieben. Die anderen im Analysemodell berücksichtigten Beweisideen, die sich jedoch nicht als Grundlage für einen Beweis der zu zeigenden Aussage eignen, werden im Folgenden nicht beschrieben. Alle Beweisideen sind in Abschnitt 3.3.3.2 definiert.

Die **primäre Beweisidee 1** *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* lässt sich besonders gut durch die Betrachtung von Reihenfolgebeispielen, d. h. durch systematisches Probieren und das Invarianzprinzip, entwickeln. Sie erscheint Dank der geringen Anforderungen im Bereich des Bereichswissens und der curricularen Vorgaben, die die Arbeit mit Beispielen betonen (Abschn. 4.2.2), für Lernende aller Klassenstufen 5 bis 8 besonders gut zugänglich. Die Summe der nächstgrößeren vier aufeinanderfolgenden ungeraden Zahlen ist immer um acht größer als die Summe der vorhergehenden vier aufeinanderfolgenden ungeraden Zahlen. Daher ist es denkbar, dass die Beziehung zwischen der sukzessiven Vergrößerung dieser Summe und der zu zeigenden Behauptung für Teilnehmende in den Klassenstufen 7, 8, die die Teilbarkeit durch acht beweisen sollen, offensichtlicher erscheint als für Teilnehmende in den Klassenstufen 5, 6, die zum Beweis der Teilbarkeit durch vier aufgefordert werden. Es ist folglich möglich, dass der kleine Unterschied in der Aufgabenstellung dazu beiträgt, dass die Beweisidee 1 von Teilnehmenden in den Klassenstufen 7, 8 häufiger genutzt wird als von denjenigen in den Klassenstufen 5, 6.

Ein Beweis der zu zeigenden Aussage über die sukzessive Vergrößerung der Summe um acht basiert auf dem Prinzip der vollständigen Induktion, welches im Schulunterricht in NRW normalerweise nicht thematisiert wird. Es ist insbesondere möglich, dass Lernenden die Relevanz des Induktionsanfangs nicht bewusst ist, sodass dieser fehlt oder nicht das kleinste Beispiel angegeben ist. Außerdem wird eventuell nicht begründet, warum sich die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer um acht vergrößert, was eine inhaltliche Lücke darstellt. Eine korrekte und vollständige Begründung der sukzessiven Vergrößerung der Summe um acht impliziert, dass diese hinterfragt wird und zusätzlich zur Idee 1 die Beweisideen 10 *Abstände ungerader Zahlen* und 11 *vier Zahlen* genutzt werden. Bezüglich der Schlussweise erscheint es fraglich, inwieweit Lernende bei Verwendung der Beweisidee 1 die Notwendigkeit sehen, ihre tendenziell an Beispielen durchgeführten Überlegungen zusätzlich allgemeingültig zu formulieren. Es liegt die Vermutung nahe, dass bei Beweisversuchen basierend auf der Beweisidee 1 die Schlussweise häufig induktiv mit Andeutung der Verallgemeinerbarkeit oder deduktiv mit induktivem Anteil ist (Abschn. 3.3.6 zur Schlussweise in Beweisversuchen mit begründendem Ansatz).

Die **primäre Beweisidee 2** *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* charakterisiert eine Reformulierung in der Darstellung. Diese Reformulierung (z. B. $S \coloneqq \mathrm{x} + (\mathrm{x}+2) + (\mathrm{x}+4) + (\mathrm{x}+6), \mathrm{x} \in \mathbb{N}$) basierend auf der Beweisidee 2 und unter impliziter Nutzung der Beweisideen 10 *Abstände ungerader Zahlen*, 11 *vier Zahlen* stellt mit Blick auf die in den Klassenstufen 5, 6 zu zeigende Teilbarkeit durch vier die Hauptschwierigkeit dar, die zu

zeigende Aussage lässt sich in diesem Fall mit einem zwei- bis dreischrittigen Argumentgefüge beweisen. Wenn Teilnehmenden aus den Klassenstufen 5, 6 eine solche Reformulierung gelingt, erscheint es gut möglich, dass sie es schaffen, ein korrektes und vollständiges Argumentgefüge zu formulieren.

Um ausgehend von der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* die in den Klassenstufen 7, 8 geforderte Teilbarkeit durch acht zu beweisen, ist mindestens eine weitere der Beweisideen 1 bis 8 erforderlich, die deutlich über die zu zeigende Aussage und ihre Voraussetzungen hinausgehen. Um zu zeigen, dass die Summe $S := x + (x + 2) + (x + 4) + (x + 6) = 4x + 12$ durch acht teilbar ist, muss berücksichtigt werden, dass x laut den Voraussetzungen eine ungerade Zahl ist. Dann lässt sich z. B. über die sukzessive Vergrößerung des Terms $4x + 12$ (Beweisidee 1) oder die Reste der zwei Summanden bei Division durch acht (Beweisidee 5) argumentieren. Das bedeutet, dass ein weitgehend korrektes und vollständiges Argumentgefüge ausgehend von der primären Beweisidee 2 in den Klassenstufen 7, 8 anspruchsvoller ist als in den Klassenstufen 5, 6. Es ist denkbar, dass manche Lernende in den Klassenstufen 7, 8 einen Beweisversuch mit dem Ansatz $S := x + (x + 2) + (x + 4) + (x + 6) = 4x + 12$ beginnen und sie anschließend zunächst Beispiele betrachten, um eine weitere Beweisidee zu finden, oder ihnen kein korrektes und vollständiges Argumentgefüge gelingt.

Die **primäre Beweisidee 3** *Vielfaches einer mittleren Zahl oder Durchschnittsbildung* lässt sich mit Blick auf die in den Klassenstufen 5, 6 zu zeigende Teilbarkeit durch vier über die Beobachtung finden, dass der Quotient stets der Zahl zwischen dem zweikleinsten und zweitgrößten Summanden entspricht. Diese Invariante ist im Falle der Betrachtung mehrerer vollständig und korrekt ausgeführter Beispiele zu der zu zeigenden Aussage besonders gut zu erkennen. Zusätzlich zur Beobachtung dieser Invariante kann etwa die Veränderung der vier Summanden zu vier gleichen Zahlen basierend auf dem Wissenselement, dass der Abstand zwischen zwei aufeinanderfolgenden ungeraden Zahlen immer zwei beträgt, wesentlich zur Entwicklung der Beweisidee 3 beitragen. Ein korrektes und vollständiges Argumentgefüge basierend auf der primären Beweisidee 3 erfordert, mit Fokus auf den Nachweis der Teilbarkeit durch vier, keine andere Beweisidee, die deutlich über die zu zeigende Aussage hinausgeht.

Im Vergleich dazu ist der in den Klassenstufen 7, 8 geforderte Beweis der Teilbarkeit durch acht basierend auf der Beweisidee 3 anspruchsvoller, weil die Invarianten weniger offensichtlich erscheinen und zudem für den Beweis mehr Beweisideen sowie ein etwas komplexeres Argumentgefüge erforderlich sind. Im Falle der Division durch acht entspricht der Quotient nicht der Zahl zwischen dem zweitkleinsten und zweitgrößten Summanden, sondern dessen Hälfte. Es ist mit zu berücksichtigen, dass die Zahl zwischen dem zweitkleinsten und zweitgrößten Summanden stets gerade und das Vierfache einer geraden Zahl durch acht teilbar ist.

Die Verwendung der **primären Beweisidee 4** *Zahldarstellung im Dezimalsystem* kann besonders nahe liegen, wenn Lernende mit Betrachtungen modulo zehn oder konkret z. B. mit Beweisen von Endstellenregeln für Teilbarkeit gut vertraut sind. Dann erscheint es eventuell als tragfähiger Ansatz, die Summanden oder deren Summe modulo zehn zu betrachten. Eine reine Betrachtung modulo zehn ist jedoch für einen Beweis der zu zeigenden Aussage nicht direkt

zielführend, weil zehn weder durch acht noch durch vier teilbar ist, so dass diese Beweisversuche allenfalls einen Ansatz für ein weiterentwickelbares Argumentgefüge darstellen können. Ein Beweis über die Betrachtung der Summanden modulo 100 ist prinzipiell möglich, aber mühsam, weil die Gültigkeit der Aussage für fünfzig Einzelfälle nachzuweisen wäre.

Es ist denkbar, dass Lernende in den Klassenstufen 5 bis 8 die primäre Beweisidee 4 *Zahldarstellung im Dezimalsystem* finden und ihnen darauf basierend ein Beweis gelingt, wenn sie sich etwa mit dem kleinsten Beispiel und dem Beispiel der aufeinanderfolgenden ungeraden Zahlen $11, 13, 15, 17$ beschäftigen, oder wenn sie mindestens sechs bis sieben Reihenfolgebeispiele beginnend mit dem kleinsten Beispiel betrachten. Dann kann auffallen, dass die Summe im sechstkleinsten Fall der vier aufeinanderfolgenden ungeraden Zahlen $11, 13, 15, 17$ um vierzig größer ist als beim kleinsten Beispiel der Zahlen $1, 3, 5, 7$ und im siebtkleinsten Fall der Zahlen $13, 15, 17, 19$ um vierzig größer als beim zweitkleinsten Beispiel der Zahlen $3, 5, 7, 9$. Diese Beobachtung ist plausibel, weil jeder der Summanden um $(5 \cdot 2 =)\ 10$ größer ist als beim kleinsten bzw. zweitkleinsten Beispiel. Immer wenn man fünf Reihenfolgebeispiele „weiter geht“, vergrößert sich die Summe um vierzig, was durch vier (Kl. 5, 6) bzw. acht (Kl. 7, 8) teilbar ist. Wenn man zeigt, dass die Summe in den fünf kleinsten Fällen jeweils durch vier bzw. acht teilbar ist, lässt sich somit auf die Gültigkeit der zu zeigenden Aussage schließen. Ein solches Vorgehen und entsprechendes Argumentgefüge ist wegen der Vielzahl notwendiger Reihenfolgebeispiele inklusive des kleinsten Beispiels und da sich alleine anhand von zwei bis drei Reihenfolgebeispielen bereits die Invariante der Vergrößerung der Summe um acht (Beweisidee 1) erkennen lässt, d. h. das Finden der Beweisidee 1 wahrscheinlicher erscheint, nicht unbedingt zu erwarten.

Weniger inhaltlich-anschaulich lässt sich die zu zeigende Aussage basierend auf der primären Beweisidee 4 beweisen, indem man z. B. den kleinsten Summanden in der Form $10\mathrm{a} + \mathrm{e}$ mit $\mathrm{a} \in \mathbb{N}$ und $\mathrm{e} \in \{1; 3; 5; 7; 9\}$ darstellt. Davon ausgehend kann unter Nutzung der Beweisideen $2, 10, 11$ und einer vollständigen Fallunterscheidung die Gültigkeit der vorgegebenen Aussage hergeleitet werden (Abschn. 3.3.3.2). Dabei sind mit den Ideen $4, 2$ zumindest zwei klar über die zu zeigende Aussage hinausgehende Beweisideen erforderlich und symbolische Darstellungsmittel hilfreich. Zudem sind vollständige Fallunterscheidungen im Kernlehrplan für das Gymnasium in NRW nicht erwähnt (MSW NRW, 2007, 2008a). Insgesamt erscheint es folglich unwahrscheinlich, dass Lernende in den Klassenstufen 5 bis 8 einen Beweis der zu zeigenden Aussage basierend auf der primären Beweisidee 4 formulieren.

Bei der **primären Beweisidee 5** *Reste bei Division einzelner Summanden durch vier oder acht* ist es zum einen möglich, dass explizit mithilfe modularer Arithmetik argumentiert wird, was Lernenden normalerweise nicht aus dem Schulunterricht, aber eventuell aus mathematischen Förderangeboten bekannt ist. Mit dem Wissen, dass bei Division durch acht je genau einer der Summanden den Rest $1, 3, 5, 7$ lässt bzw. bei Division durch vier genau zwei der Summanden den Rest 1 und die anderen beiden den Rest 3 lassen, kann die zu zeigende Aussage in einem kurzen Argumentgefüge bewiesen werden, und zwar ohne dass eine weitere der unterschiedenen Beweisideen 1 bis $13, 55$ explizit ausgeführt werden muss. Daher ist zu erwarten, dass

vielen der Lernenden ggf. zumindest ein weitgehend korrektes und vollständiges Argumentgefüge gelingt. Alternativ zur expliziten Nutzung modularer Arithmetik ist es möglich, dass Lernende – etwa basierend auf ihren Erfahrungen mit dem Teilen mit Rest aus der Grundschulzeit – die einzelnen Summanden als Vielfache der Zahl vier bzw. acht plus den verschiedenen Resten betrachten. Sowohl im Falle der expliziten Nutzung modularer Arithmetik als auch im Falle der Betrachtung der Summanden als Vielfache der Zahl vier bzw. acht plus einem Rest, kann man direkt über die Summe der Reste argumentieren oder eine vollständige Fallunterscheidung, etwa in Abhängigkeit der Teilbarkeit des kleinsten Summanden, durchführen. Vor allem bei der Durchführung einer vollständigen Fallunterscheidung ist zu beachten, dass z. B. in dem Fall, in dem der kleinste Summand gleich $4x + 3$ ($x \in \mathbb{N}$) ist, der zweitkleinste Summand nicht $4x + 1$, sondern $4(x + 1) + 1$ entspricht. Dies kann eine Fehlerquelle bei der Verschriftlichung vielleicht sachinhaltlich eigentlich korrekter Überlegungen darstellen.

Im Fall der **primären Beweisidee 6** *Teilbarkeit der Summen zweier Zahlenpaare* werden anstatt einer Summe von vier aufeinanderfolgenden ungeraden Zahlen in der Regel zwei Summen von je zwei Zahlen betrachtet. Dabei kann über die Summen der kleinsten und der größten beiden Summanden, über die Summen der „äußeren“ und der „mittleren“ Summanden oder darüber, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen stets der Summe zweier durch acht (bzw. durch vier) teilbarer Summen entspricht, argumentiert werden. Unabhängig davon lässt sich die Teilbarkeit der beiden Summen durch zwei, vier oder acht betrachten.

Wie sich etwa in der Auseinandersetzung mit Beispielen beobachten lässt, ist die Summe zweier aufeinanderfolgender ungerader Zahlen immer durch vier teilbar. Über diese Aussage kann zu der in den Klassenstufen 5, 6 zu zeigenden Teilbarkeit durch vier ein weitgehend korrektes Argumentgefüge entwickelt werden.[104] Es ist folglich gut denkbar, dass Lernende in den Klassenstufen 5, 6 basierend auf der primären Beweisidee 6 ein weitgehend korrektes Argumentgefüge formulieren, wenn auch eventuell eines mit einer inhaltlichen Lücke. Für ein weitgehend korrektes und vollständiges Argumentgefüge ist, vor dem Hintergrund der festgelegten Erwartungshorizonte (Abschn. 3.3.4), auch zu begründen, warum die Summe zweier aufeinanderfolgender ungerader Zahlen durch vier teilbar ist. Dies ist mithilfe jeder der zentralen Beweisideen 1 bis 5 möglich. Damit zusammenhängend erfordert ein weitgehend korrektes und vollständiges Argumentgefüge beruhend auf der primären Beweisidee 6 eine größere Anzahl an Deduktionen als etwa eines beruhend auf der primären Beweisidee 1. Ein weitgehend korrektes und vollständiges Argumentgefüge basierend auf der primären Beweisidee 6 ist also vergleichsweise anspruchsvoll, was dazu beitragen kann, dass Lernenden zwar ein weiterentwickelbares Argumentgefüge, aber kein korrektes und vollständiges Argumentgefüge gelingt.

Die in den Klassenstufen 5, 6 geforderte Teilbarkeit durch vier lässt sich anstatt über die Teilbarkeit durch vier der Summen zweier Zahlenpaare auch über deren Teilbarkeit durch zwei und etwa den Abstand dieser Summen beweisen. Da das Wissenselement, dass die Summe von

[104] Zudem sind die Summe der kleinsten und der größten sowie die Summe der zweitkleinsten und der zweitgrößten von vier aufeinanderfolgenden ungeraden Zahlen durch vier teilbar. Darüber kann analog argumentiert werden. Die nachfolgenden Ausführungen zur Begründung dieser Regelmäßigkeit gelten ebenfalls analog.

zwei ungeraden Zahlen gerade ist, Lernenden aller Klassenstufen 5 bis 8 aus dem Schulunterricht gut bekannt sein sollte, liegt die Vermutung nahe, dass zumindest einzelne junge Lernende dies versuchen.

Für einen Beweis der in den Klassenstufen 7, 8 geforderten Teilbarkeit durch acht ist die Betrachtung der Teilbarkeit durch zwei der Summen je zweier von vier aufeinanderfolgenden ungeraden Zahlen nicht ausreichend. Es muss über die Teilbarkeit durch vier oder über die Teilbarkeit durch acht der Summen zweier Zahlenpaare argumentiert werden. Zudem ist das erforderliche Argumentgefüge komplexer. Wenn man beispielsweise über die Teilbarkeit durch vier der Summe der kleinsten und der Summe der größten beiden von vier aufeinanderfolgenden ungeraden Zahlen argumentiert, reicht es mit Blick auf die in den Klassenstufen 7, 8 vorgegebene Aussage im Vergleich zu der in den Klassenstufen 5, 6 nicht aus auszuführen, dass und warum diese beiden Summen durch vier teilbar sind. Für einen hinreichend legitimierten Schluss auf die in den Klassenstufen 7, 8 zu zeigende Teilbarkeit durch acht muss zusätzlich begründet werden, warum aus der Teilbarkeit durch der genannten zwei Summen die Teilbarkeit durch acht der Gesamtsumme von vier aufeinanderfolgenden ungeraden Zahlen resultiert. Ein Beweis basierend auf der primären Beweisidee 6 ist mit Blick auf die in den Klassenstufen 7, 8 zu zeigenden Teilbarkeit durch acht anspruchsvoller als mit Blick auf die in den Klassenstufen 5, 6 zu zeigende Teilbarkeit durch vier.

Sowohl mit Blick auf den Beweis der Teilbarkeit durch acht als auch mit Blick auf den Beweis der Teilbarkeit durch vier erscheint die Verwendung der primären Beweisidee 6 als recht fehleranfällig. Insbesondere ist es denkbar, dass Lernende ausgehend von einem einzelnen Beispiel oder ausgehend von mehreren ungünstig ausgewählten Beispielen falsche Verallgemeinerungen vornehmen, wie etwa dass die Summe der „äußeren“ und die Summe der „mittleren“ beiden Summanden immer durch acht teilbar sind. Zudem ist es möglich, dass Lernende fehlerhafte Aussagen, wie dass die Summe zweier durch zwei teilbarer Zahlen immer durch vier teilbar ist, nutzen, um auf die zu zeigende Aussage zu schließen. Auch kann es passieren, dass Lernende ausschließlich die Teilbarkeit durch vier anstatt der Teilbarkeit durch acht bzw. ausschließlich die Teilbarkeit durch zwei anstatt der Teilbarkeit durch vier von vier aufeinanderfolgenden ungeraden Zahlen begründen, sodass der Schluss auf die zu zeigende Aussage nicht hinreichend legitimiert ist.

Im Falle der **primären Beweisidee 7** *Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen* wird die zu zeigende Aussage reformuliert. Anstelle von vier aufeinanderfolgenden ungeraden Zahlen werden im Kern vier aufeinanderfolgende gerade Zahlen betrachtet. Dies kann insbesondere mit Blick auf die in den Klassenstufen 5, 6 zu zeigende Teilbarkeit durch vier nahe liegen, weil die Summe von vier aufeinanderfolgenden geraden Zahlen stets um vier kleiner bzw. größer ist als die Summe der vier benachbarten aufeinanderfolgenden ungeraden Zahlen. Wird dabei die Teilbarkeit der Summe von vier aufeinanderfolgenden geraden Zahlen durch vier als gültig vorausgesetzt, enthält das Argumentgefüge allerdings im Vergleich zu den formulierten Erwartungshorizonten zumindest eine inhaltliche Lücke (Abschn. 3.3.4). Zu beweisen, dass die Summe von vier aufeinanderfolgenden geraden Zahlen durch vier teilbar ist, entspricht von den objektiven Anforderungsstrukturen her dem Beweis der eigentlich vorgegebenen Aussage

über die Summe von vier aufeinanderfolgenden ungeraden Zahlen. Vor allem die Überlegung, dass gerade Zahlen entweder durch vier teilbar sind oder durch zwei, aber nicht durch vier teilbar sind, ist Schülerinnen und Schülern aber eventuell besser vertraut, als dass aufeinanderfolgende ungerade Zahlen stets abwechselnd um eins größer bzw. kleiner als ein Vielfaches von vier sind, was die subjektiv wahrgenommene Schwierigkeit des Problems reduzieren kann.

Im Falle der **primären Beweisidee 8** *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* wird auf Grundlage des Wissenselements argumentiert, dass eine ungerade Zahl gleich einer geraden Zahl plus bzw. minus eins ist, oder ausgehend davon, dass eine ungerade Zahl gleich dem Doppelten einer ganzen Zahl plus bzw. minus eins ist. Zumindest dass eine ungerade Zahl gleich einer geraden Zahl plus bzw. minus eins ist, sollte allen Teilnehmenden ab der Klassenstufe 5 aus der Grundschulzeit bekannt sein (MSW NRW, 2008). Die Verwendung der primären Beweisidee 8 ist in den Klassenstufen 5, 6 insbesondere in Verbindung mit der Beweisidee 7 Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen zu erwarten, wie in den Ausführungen zur Beweisidee 7 angedeutet ist. Mit Blick auf die in den Klassenstufen 7, 8 geforderte Teilbarkeit durch acht ist ein Beweis mit primärer Beweisidee 8 in Verbindung mit einer der Beweisideen 2 Darstellung einer Summe in Abhängigkeit des kleinsten Summanden (z. B. über den Term $S \coloneqq (2n+1)+(2n+3)+(2n+5)+(2n+7), n \in \mathbb{N}$) und 3 Vielfaches einer mittleren Zahl oder Durchschnittsbildung (z. B. über den Term $S \coloneqq (g-3)+(g-1)+(g+1)+(g+3), g \in 2\mathbb{N}, g \geq 4$) gut möglich.

Weiterhin kann die zu zeigende Aussage auch ausgehend von der **primären Beweisidee 11** *vier Zahlen* bewiesen werden, die sich durch die Berücksichtigung der Anzahl an Summanden als wesentliche inhaltliche Überlegung charakterisiert. Basierend auf der primären Beweisidee 11 ist ein korrektes und vollständiges Argumentgefüge möglich, wenn eine korrekte Verallgemeinerung der zu zeigenden Ausage betrachtet wird. Dazu gehört z. B. die Aussage, dass die Summe von n ($n \in \mathbb{N}$) aufeinanderfolgenden ungeraden Zahlen durch $2n$ teilbar ist. Der Beweis einer solchen Verallgemeinerung ist schwieriger als der Beweis der in der Aufgabenstellung vorgegebenen Aussage. Deshalb liegt die Vermutung nahe, dass allenfalls einzelne Lernende aus den Klassenstufen 5 bis 8 ein weitgehend korrektes Argumentgefüge basierend auf der primären Beweisidee 11 formulieren.[105]

Zusammenfassend lässt sich bezüglich der **primären Beweisideen 1 bis 8 und 11** festhalten, dass prinzipiell auf Grundlage jeder dieser Beweisideen ein Beweis der zu zeigenden Aussage möglich, jedoch das Gelingen eines Beweises ausgehend von einer der Ideen 4 *Zahldarstellung im Dezimalsystem* und 11 *vier Zahlen* nicht zu erwarten ist. Mit Blick auf die in den Klassenstufen 5, 6 geforderte Teilbarkeit durch vier erscheint ein weitgehend korrektes und vollständiges

[105] Trotzdem ist zu erwarten, dass vor allem in den Klassenstufen 5, 6, in denen die Teilbarkeit durch vier bewiesen werden soll, die primäre Beweisidee 11 *vier Zahlen* vorkommt, etwa in der Form „Die Summe von vier aufeinanderfolgenden ungeraden Zahlen ist immer durch vier teilbar, weil immer genau vier Zahlen addiert werden“. Dass es sich um vier Summanden handelt, ist in der Aufgabenstellung vorgegeben und erscheint eventuell einigen Lernenden durch die Übereinstimmung der Anzahl an Summanden und des Teilers als entscheidend.

Argumentgefüge auf Grundlage jeder der primären Beweisideen 1, 2, 3, 5 gut möglich.[106] Zudem ist es gut denkbar, dass Lernende in diesen Klassenstufen eine der Beweisideen 6, 7 nutzen und ein weitgehend korrektes Argumentgefüge mit einer inhaltlichen Lücke formulieren. Im Vergleich zur Teilbarkeit durch vier ist ein Beweis der in den Klassenstufen 7, 8 geforderten Teilbarkeit durch acht basierend auf den primären Beweisideen 1, 5 ähnlich anspruchsvoll, im Falle jeder der primären Beweisideen 2, 3, 6, 7 jedoch klar anspruchsvoller. Es ist zu erwarten, dass Lernende in den Klassenstufen 7, 8 ausgehend von den Beweisideen 2, 3, 6, 7 durchaus weiterentwickelbare Argumentgefüge formulieren, ihnen aber eher selten korrekte und vollständige Argumentgefüge gelingen.

Schlussweise

Bei sämtlichen primären Beweisideen 1 bis 13, 55 sind prinzipiell die Schlussweisen Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit*, Sw2 *deduktiv mit induktivem Anteil*, Sw3 *deduktiv mit zusätzlichem Beispiel* und Sw4 *deduktiv ohne zusätzliches Beispiel* möglich. Zudem können Beweisideen an Beispielen ohne Andeutung der Verallgemeinerbarkeit ausgeführt werden, sodass es sich um einen Beweisversuch mit prüfendem Ansatz handelt. Sowohl Beweisversuche mit zwischen induktiver und deduktiver Schlussweise (Sw1-2) als auch solche mit prüfendem Ansatz erscheinen bei der primären Beweisideen 1 *sukzessiven Vergrößerung einer Summe mit Erkennen einer Invariante* als wahrscheinlich, wie im Rahmen der Überlegungen zur primären Beweisidee 1 erklärt wurde. Demgegenüber ist es denkbar, dass gerade im Falle der primären Beweisidee 9 *Summen ungerader oder gerader Zahlen* deduktiv (Sw3-4) argumentiert wird, weil Wissenselemente zur Parität von Summen ungerader und gerader Zahlen den Lernenden aller Klassenstufen 5 bis 8 gut bekannt sein sollten und daher vielleicht besonders häufig ohne Überlegungen an Beispielen sofort angewendet werden.

Darstellungsmittel

Der Aufgabenteil d) kann prinzipiell mithilfe verbaler, numerischer, ikonischer oder symbolischer Darstellungsmittel bearbeitet werden. Dabei ist ein Beweis der zu zeigenden Aussage ohne **verbale Darstellungsmittel** kaum vorstellbar, wenngleich es nicht notwendig ist, eine ganze sinnbezogene Einheit überwiegend verbal darzustellen. **Numerische Darstellungsmittel** sind insbesondere im Falle der Angabe eines Beispiels zu erwarten. Beispiele können aber auch ikonisch oder verbal dargestellt werden.

An **ikonischen Darstellungsmitteln** können u. a. figurierte Zahlen, Klammern, Pfeile, Linien und Häkchen verwendet werden, wobei nicht unbedingt zu erwarten ist, dass die Schülerinnen und Schüler alle sinnbezogenen Einheiten ikonisch darstellen. Es ist z. B. möglich, dass die ikonischen Darstellungsmittel in Verbindung mit numerischen oder verbalen Darstellungsmitteln verwendet werden. Ikonische Darstellungsmittel können auch zusammen mit symbolischen Darstellungsmitteln vorkommen, insbesondere falls Schülerinnen und Schüler versuchen, die zu zeigende Aussage deduktiv zu beweisen.

[106] Dabei lassen sich die Beweisideen 1, 3 gut über die Arbeit mit Beispielen und die Verwendung des Invarianzprinzips entwickeln, während für die erfolgreiche Nutzung der primären Beweisideen 2, 5 bestimmtes Basiswissen hilfreich, jedoch den Lernenden nicht unbedingt bekannt ist.

Die Verwendung figurierter Zahlen im Speziellen ist im Kontext der primären Beweisidee 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*, aber auch bei anderen primären Beweisideen möglich (Abb. 73). Ikonische Darstellungsmittel wie Klammern, Pfeile, Linien, Häkchen und farbliche Markierungen können bei der Entwicklung und Formulierung von Beweisversuchen basierend auf vielen der unterschiedenen primären Beweisideen helfen (Abb. 74). Da ikonisch dargestellte Überlegungen vielleicht häufig ausschließlich auf Spezialfälle bezogen werden und die Lernenden vielleicht nicht die Notwendigkeit sehen, ihre Überlegungen zusätzlich allgemeingültig zu formulieren oder nicht wissen, wie dies möglich ist, liegt die Vermutung nahe, dass die Schlussweise in Beweisversuchen mit einer Darstellungsform mit ikonischem Anteil meistens zwischen induktiv und deduktiv liegt oder es sich um Beweisversuche mit prüfendem Ansatz handelt.

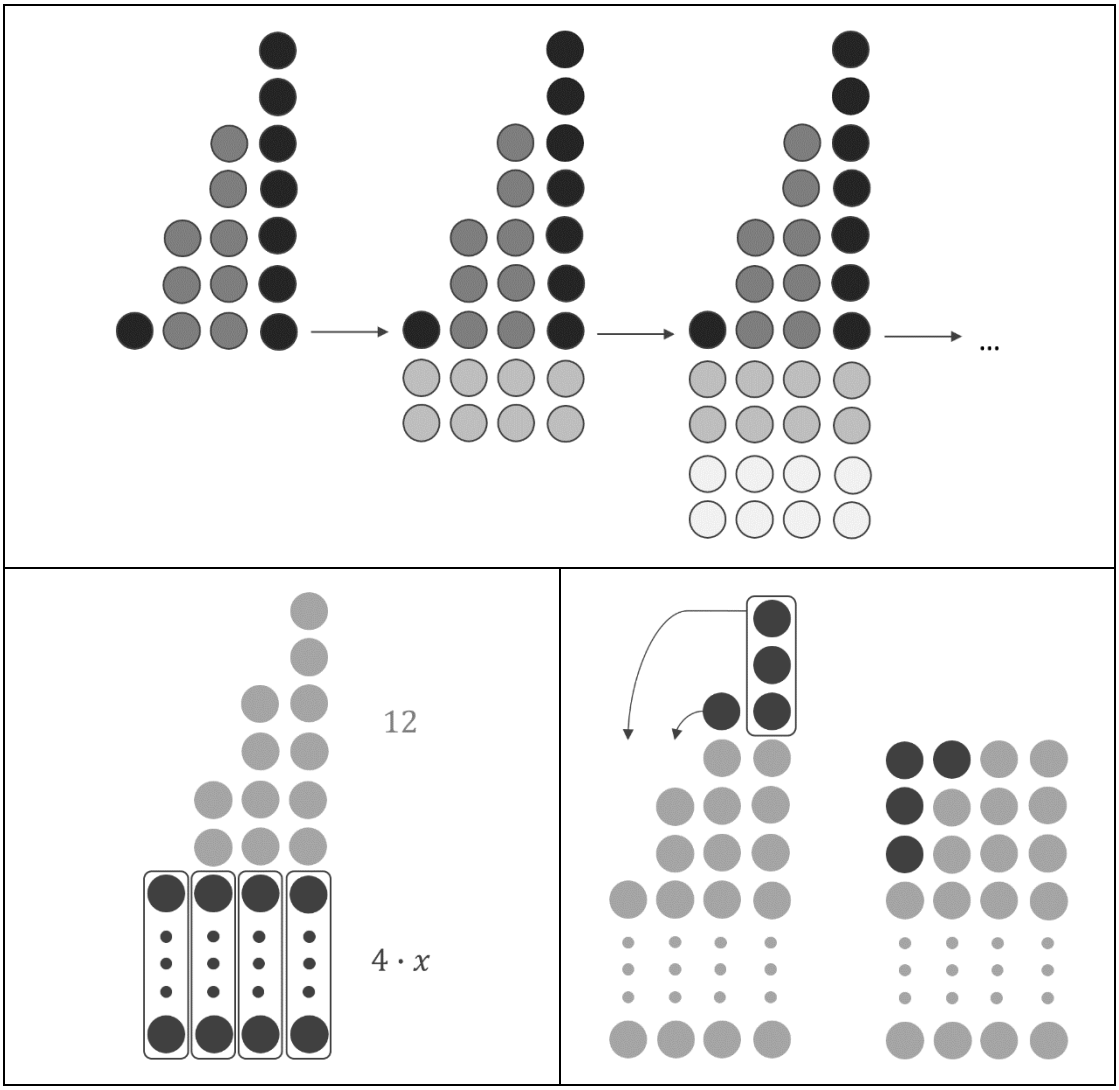


Abbildung 73: Darstellung der Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* (oben), der Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* (unten links) und der Beweisidee 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung* (unten rechts) insbesondere mithilfe figurierter Zahlen

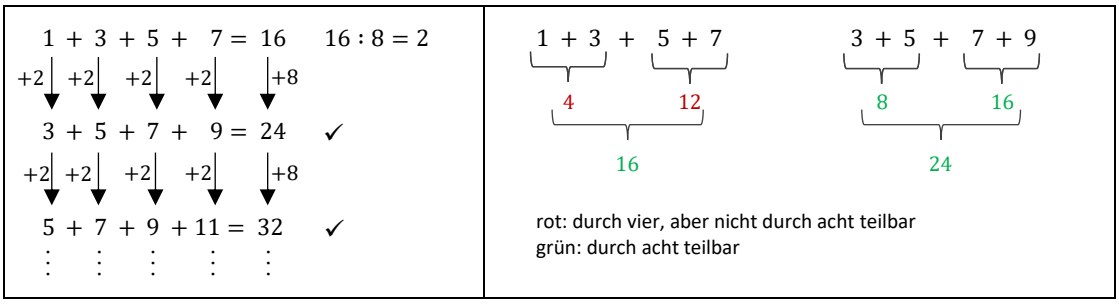


Abbildung 74: Darstellung der Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* (links) und der Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* (rechts) mithilfe verschiedener ikonischer Darstellungsmittel

Symbolische Darstellungsmittel sind für die Entwicklung und Formulierung eines Beweises zu der zu zeigenden Aussage nicht notwendig. Schülerinnen und Schüler in den Klassenstufen 5, 6 sowie zu Beginn der Klassenstufe 7 sind in der Regel mit der Verwendung elementarer Algebra auch nicht vertraut. Nach dem Kernlehrplan für Gymnasien in NRW sollten Schülerinnen und Schüler aber spätestens am Ende der Klassenstufe 8 mit Termen mit einer Variablen umgehen können (Abschn. 4.2.2). Möchte man die im Rahmen dieser Arbeit fokussierte Allaussage mithilfe symbolischer Darstellungsmittel beweisen, so sind, was die Verwendung symbolischer Darstellungsmittel betrifft, im Wesentlichen das Aufstellen eines Terms mit einer Variablen und elementare Termumformungen ausreichend. Wie in der Abbildung 75 veranschaulicht, lässt sich die zu zeigende Aussage unter Verwendung einer Variablen sehr gut ausgehend von der Beweisidee 2 (Kl. 5, 6) bzw. durch die Kombination der Beweisideen 2, 8 (Kl. 7, 8) beweisen. Außerdem sind z. B. die Terme $(g-3)+(g-1)+(g+1)+(g+3)$, $g \in 2\mathbb{N}$, $g \geq 4$ (Beweisidee 3) oder $10x+e$ mit $x \in \mathbb{N}$, $e \in \{1,3,5,7,9\}$ (Beweisidee 4) als wesentliche Bestandteile eines Beweises der zu zeigenden Aussage vorstellbar. Variablen können darüber hinaus z. B. bei einer vollständigen Induktion basierend auf der primären Beweisidee 1 oder der Ausführung einer vollständigen Fallunterscheidung nach der Teilbarkeit des kleinsten Summanden durch acht (Beweisidee 5) helfen.

Beweis der zu zeigenden Aussage in symbolischer Darstellungsform

Klassenstufen 5, 6: Beweis der Teilbarkeit durch vier

Sei $n \in \mathbb{N}$, $2 \nmid n$.
$n+n+2+n+4+n+6 = 4n+12$
Da $4|4n$ und $4|12$, gilt $4|(4n+12)$. ■

Klassenstufen 7, 8: Beweis der Teilbarkeit durch acht

Sei $x \in \mathbb{N}$.
$2x+1+(2x+3)+(2x+5)+(2x+7) = 8x+16$
Da $8|8x$ und $8|16$, gilt $8|(8x+16)$. ■

Abbildung 75: Beweis der zu zeigenden Aussage in symbolischer Darstellungsform

Charakteristika des Aufgabenteils d) im Kontext der Teilbarkeit ganzer Zahlen

Zusammenfassend lässt sich ausgehend von den vorhergehenden Überlegungen herausstellen, dass die im Aufgabenteil d) zu beweisende Aussage über die Teilbarkeit der Summe von vier aufeinanderfolgenden ungeraden Zahlen durchaus für den Themenbereich der Teilbarkeit ganzer Zahlen charakteristisch ist.

Grundlegend kennzeichnet den Aufgabenteil d) die Vorgabe einer Allaussage mit unendlichem Gültigkeitsbereich. Beweise von Allaussagen mit unendlichem Gültigkeitsbereich sind eine Besonderheit der Mathematik im Vergleich zu anderen Wissenschaften (Abschn. 1.1, 3.1). Für den Beweis der im Aufgabenteil d) vorgegebenen Allaussage ist elementares Bereichswissen ausreichend. Die Möglichkeit der Formulierung mathematisch voraussetzungsarmer Beweisaufgaben ist ein Charakteristikum des Themenbereichs der Teilbarkeit ganzer Zahlen und nebenbei ein Grund, weshalb dieser Themenbereich gewählt wurde (Abschn. 1.1).

Zugleich kennzeichnet den Aufgabenteil d) die Notwendigkeit einer Reformulierung in der Darstellung oder der Anwendung eines heuristischen Prinzips sowie das Erfordernis eines mehrschrittigen Argumentgefüges geringer bis mittlerer Komplexität. Dabei kann das Argumentgefüge auf vielfältigen primären Beweisideen basieren. Mehrere der Beweisideen und damit einhergehende Wissenselemente und Heurismen können auch bei zahlreichen anderen Aussagen aus dem Themenbereich der Teilbarkeit ganzer Zahlen verwendet werden. Dazu gehören unter anderem die Betrachtung der Reste einzelner Summanden bei einer Division, die Darstellung von Summen oder Produkten in Abhängigkeit einer der betrachteten Zahlen, Argumentationen über eine sukzessive Vergrößerung oder Verkleinerung und Überlegungen zur Parität von Summen oder Produkten ganzer Zahlen.

Typisch für den Themenbereich der Teilbarkeit ganzer Zahlen sind beim Aufgabenteil d) elementare Zugänge zu der zu zeigenden Aussage sowie Vorformen mathematischer Beweise und Beweise ohne Variablennutzung möglich. Beispielsweise lassen sich die ausbaubaren Beweisideen 1, 3, 4, 6 sehr gut ausgehend von der Beschäftigung mit Beispielen entwickeln. Als Vorformen mathematischer Beweise sind z. B. Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge bei zwischen induktiver und deduktiver Schlussweise sowie Beweisversuche mit prüfendem Ansatz und einem weiterentwickelbaren Argumentgefüge denkbar. Hinsichtlich der Verwendung von Darstellungsmitteln besteht viel Spielraum. Insbesondere können verschiedene ikonische Darstellungsmittel zur Entwicklung und Formulierung von Begründungen genutzt werden. Die Teilbarkeit der Summe von vier aufeinanderfolgenden ungeraden Zahlen durch acht lässt sich nicht direkt durch eine symbolische Darstellung der Summe in Abhängigkeit des kleinsten Summanden beweisen, sondern erfordert im Falle dieser primären Beweisidee die Verwendung einer weiteren klar über die zu zeigende Aussage hinausgehenden Beweisidee. Insbesondere ein Beweis der in den Klassenstufen 7, 8 zu zeigenden Teilbarkeit durch acht erscheint ohne Variablennutzung, basierend auf inhaltlich-anschaulichen Überlegungen einfacher als eine rein symbolische Herangehensweise. Dies trifft auch bei anderen Aussagen aus dem Themenbereich der Teilbarkeit ganzer Zahlen zu.

6.3 Zusammenfassung des sechsten Kapitels

Forschungsdesiderat, Fragestellungen und Design

Die vorliegende Arbeit knüpft an vorhandene Forschungsergebnisse an und beschreibt zugleich ein **Forschungsdesiderat** aus zwei Perspektiven: Zum einen mit Blick auf die Mathematik-Olympiade und zum zweiten vor dem Hintergrund anderer mathematikdidaktischer Studien zum Führen von Beweisen.

Die Mathematik-Olympiade ist ein Stufenwettbewerb für mathematisch interessierte Schülerinnen und Schüler. Sie erscheint als interessantes, jedoch bisher kaum beachtetes Forschungsfeld. Da die Anforderung, mathematische Aussagen zu beweisen, ein Charakteristikum der Mathematik-Olympiade ist, erscheint es als eine sinnvolle Option, eine Untersuchung im Bereich des Führens von Beweisen anzusetzen.

Das Führen von Beweisen, insbesondere das Beweisen von Allaussagen mit unendlichem Gültigkeitsbereich, ist ein Charakteristikum der Mathematik und sowohl aus mathematikdidaktischer Sicht als auch nach den Bildungsstandards für das Lernen von Schülerinnen und Schülern von Bedeutung. Das Beweisen von Allaussagen wird üblicherweise im Schulunterricht der Klassenstufe 7 explizit eingeführt. Studien mit Schülerinnen und Schülern, i. d. R. frühestens ab der Klassenstufe 7, weisen darauf hin, dass den meisten und insbesondere auch vielen der mathematisch vergleichsweise leistungsstarken Lernenden, das Beweisen solcher Aussagen nicht gelingt. Beweisversuche von Schülerinnen und Schülern in den Klassenstufen 5, 6 wurden bisher kaum untersucht. Erkenntnisse zu Beweisversuchen dieser jungen Lernenden und deren Unterschiede zu denen von Schülerinnen und Schülern in den Klassenstufen 7, 8 würden den bisherigen Forschungsstand ergänzen und Einblicke in Ansatzpunkte für eine Förderung im Führen von Beweisen ermöglichen.

In der vorliegenden Arbeit werden das Desiderat einer Untersuchung im Forschungsfeld der Mathematik-Olympiade und das Desiderat einer vergleichenden Analyse von Beweisversuchen von Lernenden in den Klassenstufen 5 bis 8 miteinander kombiniert und auf Beweisversuche zu einer Allaussage mit unendlichem Gültigkeitsbereich aus dem Themenbereich der Teilbarkeit ganzer Zahlen fokussiert. Die **Forschungsfragen** dieser Arbeit lauten:

1. Inwieweit gelingt es Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 eine mathematische Allaussage aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen, welche Charakteristika zeichnen deren Beweisversuche aus und inwieweit sind Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen?
2. Welche Gemeinsamkeiten und Unterschiede bestehen diesbezüglich zwischen Beweisversuchen von Teilnehmenden der Mathematik-Olympiade mit Blick auf verschiedene Klassenstufen?

Um diese Fragen zu untersuchen, wurde bei der **Regionalrunde der Mathematik-Olympiade** in Nordrhein-Westfalen im November 2016 eine **spezifisch fokussierte Beweisaufgabe** als Teil der regulären Wettbewerbsklausur gestellt, von der die Bearbeitungen des Aufgabenteils d) im Rahmen dieser Arbeit ausgewertet werden. Im Aufgabenteil d) der Beweisaufgabe sollten die Teilnehmenden zeigen, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer durch vier (Kl. 5, 6) bzw. acht (Kl. 7, 8) teilbar ist. Diese Aussage ist eine Allaussage mit unendlichem Gültigkeitsbereich, deren Beweis eine Reformulierung in der Darstellung oder die Anwendung eines heuristischen Prinzips sowie die Formulierung eines mehrschrittigen Argumentgefüges erfordert. Während diesbezüglich gezielt Schwierigkeiten gesetzt wurden, wurden die Anforderungen in anderen Bereichen, wie der Decodierung der Aufgabenstellung, niedrig gehalten. Nebenbei lässt die vorgegebene Aussage bezüglich möglicher Beweisideen, Schlussweisen, Darstellungsmittel und der Nutzung von Beispielen große Freiheiten. Die fokussierte Aussage erscheint Dank dieser vielfältigen Möglichkeiten und dadurch, dass sie voraussetzungsarm ist, für den Themenbereich der Teilbarkeit ganzer Zahlen charakteristisch.

7 Datenerhebung und Auswertung

Anknüpfend an das Design der eigenen Studie (Abschn. 6.2) wird in diesem Kapitel der tatsächliche Verlauf der Datenerhebung skizziert. Dabei werden die Anzahlen an erhaltenen und für die Auswertung zur Verfügung stehenden Daten präsentiert (Abschn. 7.1). Anschließend werden die methodischen Grundlagen, die bei der Auswertung der erhobenen Beweisversuche verwendet werden, zunächst allgemein und danach mit Fokus auf die durchgeführte Studie beschrieben (Abschn. 7.2, 7.3).

7.1 Datenerhebung

An der Regionalrunde der Mathematik-Olympiade in Nordrhein-Westfalen haben insgesamt 5775 Schülerinnen und Schüler ab der Klassenstufe 5 bis zum Abitur teilgenommen. Entsprechend dem beschriebenen Design der Studie haben in 44 der 55 Regionen Nordrhein-Westfalens Teilnehmende der Regionalrunde 2016 die für diese Studie eigens konzipierte Beweisaufgabe als Teil der Wettbewerbsklausur und den kurzen Fragebogen direkt nach der Klausur erhalten (Abschn. 6.2). Die Klausuren und ausgefüllten Fragebögen der Schülerinnen und Schüler trafen bis Mitte Januar 2017 sukzessive in der Universität ein. Die Dokumente wurden bei Bedarf gescannt und zurückgeschickt. Außerdem wurden die Vollständigkeit und Lesbarkeit aller Datensätze sowie die Angaben aus den Fragebögen erfasst. Mit Blick auf die Fragestellungen dieser Arbeit liegen 3400 weitgehend vollständige Datensätze vor.[107] Von diesen 3400 Datensätzen stammen 1309 aus der fünften, 919 aus der sechsten, 765 aus der siebten und 407 aus der achten Klassenstufe.[108]

7.2 Auswertungsmethoden

Um die Forschungsfragen dieser Arbeit auf Grundlage der vorhandenen 3400 Datensätze zu beantworten, wird eine qualitativ-strukturierende Inhaltsanalyse nach Schreier (2012, 2014) durchgeführt. Mit dem Verfahren der qualitativ-strukturierenden Inhaltsanalyse ist es möglich, große Datenmengen in systematischer und zugleich flexibler Form inhaltlich aussagekräftig zu beschreiben und zu vergleichen. Dadurch ist dieses Verfahren für die Untersuchung von Be-

[107] Die Gesamtanzahl an Schülerinnen und Schülern aus den Klassenstufen 5 bis 8, die an der Regionalrunde 2016 in einer der 44 Regionen teilgenommen haben, liegt etwas über 3400. Einzelne Datensätze sind unvollständig in der Universität angekommen und manche Dokumente sind durch das Scannen unleserlich geworden. Die für die Auswertung zur Verfügung stehenden Datensätze umfassen jeweils die Bearbeitung des relevanten Aufgabenteils d) und können eindeutig einer der Klassenstufen 5 bis 8 zugeordnet werden. Die Angaben zum Wettbewerbserfolg und zur Teilnahme an außerunterrichtlichen mathematischen Förderangeboten sind eventuell nicht ganz vollständig. Da diese Angaben für die Beantwortung der Forschungsfragen dieser Arbeit nicht relevant sind, werden die Datensätze trotzdem in die Auswertung einbezogen.

[108] Die Anzahlen an Datensätzen aus den verschiedenen Klassenstufen unterscheiden sich deutlich, da an der Regionalrunde in Nordrhein-Westfalen üblicherweise sehr viele Fünftklässlerinnen und Fünftklässler, etwas weniger Lernende aus der sechsten und siebten Klassenstufe und deutlich weniger Lernende aus der achten Klassenstufe teilnehmen (Abschn. 2.2.2).

weisversuchen im Rahmen dieser Arbeit optimal geeignet. Zusätzlich zum Verfahren der qualitativ-strukturierenden Inhaltsanalyse sind einige statistische Grundlagen zur Stichprobenziehung und zur Datenanalyse erforderlich. Diese Auswertungsmethoden werden im vorliegenden Abschnitt 7.2 allgemein beschrieben.

7.2.1 Qualitative Inhaltsanalyse

Die Beweisversuche der Schülerinnen und Schüler werden mit dem Verfahren der qualitativen Inhaltsanalyse ausgewertet. Anders als der Ausdruck „Verfahren“ suggerieren mag, handelt es sich bei der Methode der qualitativen Inhaltsanalyse allerdings nicht um einen eindeutigen Algorithmus, der angewendet wird, sondern vielmehr um einen Oberbegriff für verschiedene, wenn auch ähnliche Vorgehensweisen bei der Auswertung empirischer Daten. So werden in der Literatur zur qualitativen Inhaltsanalyse verschiedene Varianten dieses Verfahrens, wie etwa die inhaltlich-strukturierende, die evaluative, die zusammenfassende, die explikative und die typenbildende qualitative Inhaltsanalyse, unterschieden (Gläser & Laudel, 2009; Kuckartz, 2012; Mayring, 2010; Rustemeyer, 1992). Schreier (2014) vergleicht diese und weitere Varianten qualitativer Inhaltsanalyse. Sie kommt zu dem Schluss, dass im Wesentlichen zwei Basisformen qualitativer Inhaltsanalyse existieren, qualitativ-strukturierende Inhaltsanalyse und Inhaltsanalyse durch Extraktion. Ausschließlich diese zwei Basisformen werden von ihr als eigene Verfahren angesehen.

> Für alle übrigen Varianten, die in der Literatur genannt werden, gilt jedoch, dass sie nicht notwendig als eigenständige Verfahren gelten können, sondern an einer bestimmten Stelle von der generischen Variante qualitativ-strukturierender Inhaltsanalyse abweichen bzw. diese in einer bestimmten Weise spezifizieren. (Schreier, 2014, S. 22)

Schreier (2014) arbeitet beispielsweise heraus, dass die inhaltliche und die evaluative strukturierende Inhaltsanalyse eine bestimmte Art von Kategorien fokussieren, und sich die typenbildende Inhaltsanalyse durch die Verwendung eines bestimmten Auswertungsverfahrens im letzten Schritt der qualitativen Inhaltsanalyse auszeichnet. Die in der Literatur vorkommenden Varianten ließen sich zudem durch weitere Varianten systematisch ergänzen und überdies miteinander kombinieren (Schreier, 2014). Auf diesen Überlegungen aufbauend entwirft Schreier (2014) ein Werkzeugkasten-Modell der qualitativ-strukturierenden Inhaltsanalyse, das die verschiedenen in der Literatur vorkommenden Varianten in impliziter, systematischer Form integriert.

Diese übergeordnete Sichtweise von Schreier (2014) auf qualitativ-strukturierende Inhaltsanalyse ermöglicht, in verschiedenen Untersuchungen, wie der Untersuchung von Beweisversuchen im Rahmen dieser Arbeit und der Analyse der Anforderungen zum Beweisen bei der Mathematik-Olympiade im Rahmen des Forschungsprojekts (Abschn. 1.1, 10.3), das gleiche Auswertungsverfahren zu verwenden und an die unterschiedlichen Forschungsfragen, theoretischen Grundlagen und zu untersuchenden Materialien anzupassen, um möglichst valide Ergebnisse zu erhalten. Deshalb wird in dieser Arbeit das Verfahren der qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) verwendet, das im Folgenden näher vorgestellt wird. Dabei basieren die Ausführungen vollständig auf dieser Literatur. Wenn andere Quellen verwendet werden, sind diese explizit angegeben.

7.2.1.1 Grundzüge qualitativ-strukturierender Inhaltsanalyse

> [Qualitative content analysis] is a method for systematically describing the meaning of qualitative material. It is done by classifying material as instances of the categories of a coding frame. (Schreier, 2012, S. 1)

Kern der qualitativen Inhaltsanalyse ist die Anwendung eines Kategoriensystems auf das zu untersuchende Material. Das Material wird somit auf die Informationen reduziert, die mithilfe des Kategoriensystems erfasst werden. Dieses einheitliche und datenreduzierende Vorgehen ermöglicht, Eigenschaften aus den Materialien herauszuarbeiten und verschiedene Daten diesbezüglich systematisch zu beschreiben und zu vergleichen.

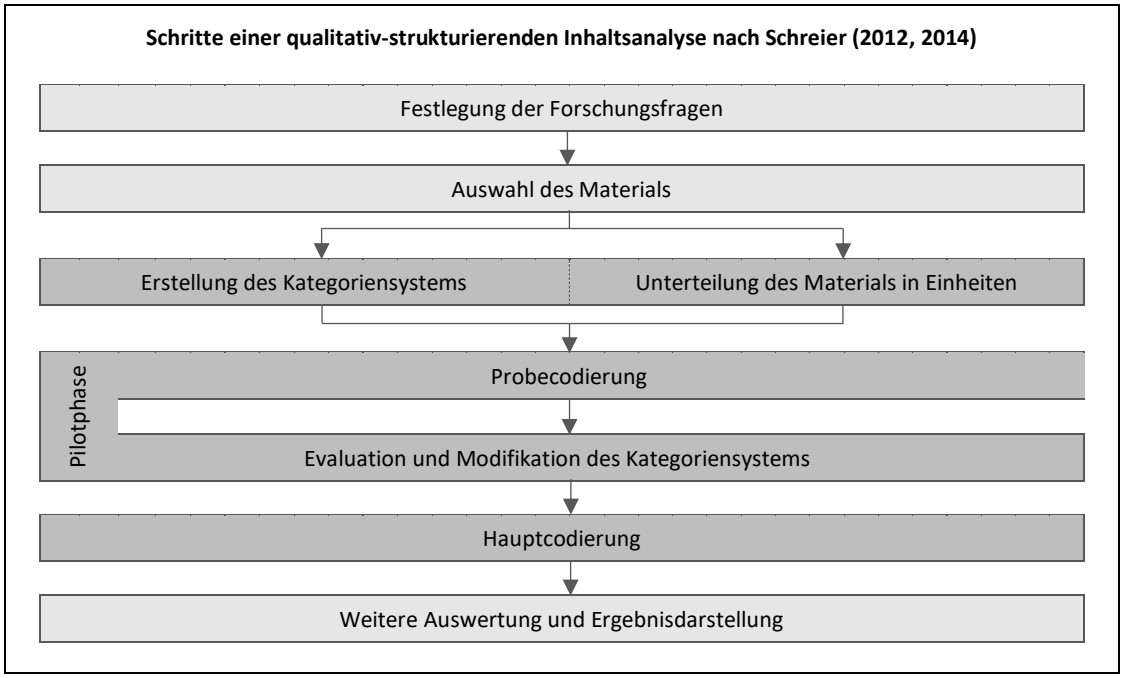


Abbildung 76: Schritte einer qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014)

Die Grundstruktur des Ablaufs einer qualitativ-strukturierenden Inhaltsanalyse ist immer gleich (Abb. 76). Den Anfang bildet die Festlegung der Forschungsfrage. Danach ist das zu untersuchende Material auszuwählen. Es ist also zu entscheiden, welches Material der qualitativen Inhaltsanalyse zugrunde liegen soll. Anschließend erfolgt die Entwicklung des Kategoriensystems. Außerdem ist das zu untersuchende Material in Codiereinheiten zu unterteilen, auf die das Kategoriensystem einzeln angewendet werden soll. Sind die Unterteilung des Materials in Einheiten und die Erstellung des Kategoriensystems abgeschlossen, wird das Kategoriensystem pilotiert. Diese Pilotphase umfasst eine Probecodierung mit anschließender Evaluation und ggf. Modifikation des Kategoriensystems. Nach einer erfolgreichen Pilotierung erfolgt die Hauptcodierung. Das Kategoriensystem wird nun sukzessive auf das gesamte zu analysierende Material angewendet. Die dabei erhaltenen Codierungen werden weiter ausgewertet und aufbereitet und abschließend als Ergebnisse präsentiert.

In einer qualitativ-strukturierenden Inhaltsanalyse wird somit theoretisch immer die gleiche Abfolge von Schritten durchlaufen. Dieser Ablauf erfolgt im Wesentlichen linear, wobei es erforderlich sein kann, von der Pilotphase noch einmal zu der Erstellung des Kategoriensystems

zurückzukehren, und danach eine erneute Pilotierung durchzuführen. Zudem kann die Unterteilung des Materials in Einheiten je nach Art des Materials und Art der zu bildenden Codiereinheiten zu leicht unterschiedlichen Zeitpunkten erfolgen. Da die Unterteilung des Materials in Einheiten stets passend zum Kategoriensystem erfolgen sollte, sind diese beiden Schritte im Ablaufmodell auf einer Ebene dargestellt (Abb. 76). Die Erstellung des Kategoriensystems, die Unterteilung des Materials in Einheiten, die Probecodierung, die Evaluation und Modifikation des Kategoriensystems und die Hauptcodierung sind charakteristisch für qualitativ-strukturierende Inhaltsanalyse. Die Festlegung der Forschungsfragen und die Auswahl des Materials hingegen erfolgen in der Regel bereits im Rahmen des Designs der Studie, auch wenn es erforderlich sein kann, Forschungsfragen zu Beginn der Auswertung zu konkretisieren oder einen Teil der erhobenen Daten für die Analyse zu selektieren. Die abschließende Aufbereitung und Präsentation der Ergebnisse ist ebenfalls ein wichtiger Schritt in einer jeden Studie.

Bei der Realisierung der einzelnen Schritte besteht jeweils Spielraum, was die Gegenstandsangemessenheit des Verfahrens sicherstellt. Beispielsweise können bei der Entwicklung des Kategoriensystems verschiedene Strategien genutzt und unterschiedliche Arten an Kategorien gebildet werden. Auch bei der Probecodierung und der Evaluation des Kategoriensystems sind unterschiedliche Vorgehensweisen möglich. In der weiteren Auswertung der erhaltenen Codierungen können verschiedene qualitative und quantitative Methoden verwendet werden. Schreier (2014) schreibt zusammenfassend:

> Charakteristisch für qualitative Inhaltsanalyse ist somit gerade die Kombination vorab spezifizierter Schritte einerseits mit deren vielfältigen Realisationsmöglichkeiten andererseits: Das Durchlaufen einer festgelegten Abfolge von Schritten gewährleistet die Systematik, während die unterschiedlichen Möglichkeiten, die Schritte konkret zu realisieren, die Gegenstandsangemessenheit des Verfahrens sichern. Genau diese Kombination von Systematik und Gegenstandsangemessenheit macht das Verfahren der qualitativen Inhaltsanalyse aus. (Schreier, 2014, S. 25)

Im Folgenden werden einzelne Schritte des Basisablaufs einer qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) und verschiedene Realisationsmöglichkeiten dieser Schritte näher erläutert. Dabei werden die im Rahmen dieser Arbeit besonders relevanten Aspekte fokussiert.

7.2.1.2 Unterteilung des Materials in Einheiten

In einer qualitativen Inhaltsanalyse wird das gesamte zu untersuchende Material in Analyse-, Codier- und Kontexteinheiten unterteilt.

> Each case on which [qualitative content analysis] is carried out constitutes a unit of analysis. (Schreier, 2012, S. 130)

Eine Analyseeinheit bildet jeder Teil des Materials, der mit der qualitativen Inhaltsanalyse ausgewertet und im Rahmen der Ergebnispräsentation separat beschrieben werden soll. Eine Analyseeinheit kann beispielsweise ein Interview oder eine Aufgabenbearbeitung sein. Jede Analyseeinheit umfasst eine oder mehrere Codiereinheiten.

> Units of coding are those parts of the units of analysis that can be interpreted in a meaningful way with respect to your categories and that fit within one subcategory of your coding frame. (Schreier, 2012, S. 131).

Eine Codiereinheit ist jeder Teil einer Analyseeinheit, der im Rahmen der Inhaltsanalyse einzeln codiert wird. Die Codiereinheiten sollten dementsprechend so gewählt werden, dass sie bezüglich der verschiedenen Merkmale des Kategoriensystems eindeutig einordbar sind.[109]

Um eine Codiereinheit angemessen zu interpretieren und entsprechend zu codieren, können weitere Teile des Materials, meistens weitere Teile der Analyseeinheit, erforderlich sein. Die Teile des Materials, die bei der Interpretation einer Codiereinheit berücksichtigt werden dürfen, bilden die Kontexteinheit.

> The context unit is that portion of the surrounding material that you need to understand the meaning of a given unit of coding. (Schreier, 2012, S. 133)

In einer qualitativen Inhaltsanalyse wird jede Codiereinheit unter eventuellem Einbezug der dazugehörigen Kontexteinheit einzeln anhand eines Kategoriensystem analysiert.

7.2.1.3 Erstellung des Kategoriensystems

Bei der Entwicklung eines Kategoriensystems ist es möglich, deduktiv, induktiv oder deduktiv-induktiv vorzugehen. Als deduktive Strategie bezeichnet man eine theoriegeleitete Kategorienbildung. Das Kategoriensystem wird in diesem Fall ausgehend von vorhandenen Informationen, wie theoretischen Konzepten, anderen Studien, Alltagserfahrungen oder einem Interviewleitfaden, ohne Einsicht in die eigenen zu analysierenden Materialien gebildet. Beispielsweise ist es möglich, ein Kategoriensystem von einer ähnlichen Studie, die auf eine etwas andere Zielgruppe fokussiert, zu übernehmen. Ein solches deduktives Vorgehen ist erforderlich, wenn mit den vorliegenden Daten Hypothesen getestet werden sollen. Im Rahmen einer qualitativen Inhaltsanalyse, bei der es meistens um eine Beschreibung des Datenmaterials geht, empfiehlt es sich, einen Teil des eigenen Materials in die Kategorienbildung einzubeziehen, um die Passung des Kategoriensystems sicherzustellen (Schreier, 2012). Bei einer induktiven Kategorienbildung wird das gesamte Kategoriensystem aus dem eigenen Material heraus gebildet. Eine mögliche Vorgehensweise ist es, sukzessive Codiereinheiten zu sichten und damit z. B. neue Kategorien aufzumachen. Im Verlauf dieses Prozesses werden kontinuierlich neue Kategorien gebildet, kleine Kategorien zusammengefasst, vorhandene Kategorien erweitert und präzisiert sowie die verschiedenen Kategorien voneinander abgegrenzt. Eine ausschließlich datenbasierte Kategorienbildung ist allerdings im Rahmen einer qualitativen Inhaltsanalyse ebenfalls untypisch, da die Forschungsfrage und vorhandene Konzepte und Studien normalerweise bereits bestimmte Oberkategorien nahelegen. Daher werden die Kategorien im Rahmen einer qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012) meistens deduktiv-induktiv gebildet. Eine typische Kombination deduktiver und induktiver Strategien besteht darin, deduktiv zu beginnen, d. h. ausgehend von der Forschungsfrage und vorhandenen Konzepten oder Kategoriensystemen, Oberkategorien festzulegen, und danach die Unterkategorien aus dem Material heraus zu entwickeln. Es sind aber auch vielfältige weitere, beliebig komplexe Kombinationen möglich (Schreier, 2012).

[109] Zugleich sollte andersherum auch das Kategoriensystem zu den Codiereinheiten passen. Wegen dieser wechselseitigen Beziehung ist die Unterteilung des Materials in Einheiten in einer qualitativen Inhaltsanalyse immer im Zusammenhang zur Erstellung des Kategoriensystems zu sehen (Schreier, 2012).

Bei jeder Kombination deduktiver und induktiver Strategien stellt sich die Frage, wie viel Material in die Entwicklung des Kategoriensystems miteinbezogen werden sollte. Nach Schreier (2012) hängt die Antwort davon ab, wie unterschiedlich die zu untersuchenden Daten sind. Es sollte so lange neues Material miteinbezogen werden, wie bei der Sichtung neue Aspekte und Kategorien zum Vorschein kommen, bis also eine Sättigung eingetreten ist. Das einbezogene Material ist im Rahmen der Kategorienbildung zudem so lange immer wieder erneut zu durchlaufen, bis alle Codiereinheiten eindeutig eingeordnet werden können. Die Unterkategorien, die zu einem Merkmal gehören, müssen disjunkt und erschöpfend sein, um eine eindeutige Zuordnung aller Codiereinheiten zu ermöglichen. Um sicherzustellen, dass alle einbezogenen und auch alle anderen zu untersuchenden Codiereinheiten klar eingeordnet werden können, ist es sinnvoll, bei jedem Merkmal eine Restkategorie zu ergänzen. Außerdem ist bei der Entwicklung der Kategorien immer wieder die Sachangemessenheit und Aussagekraft der getroffenen Unterscheidungen zu hinterfragen. Deutet sich z. B. an, dass viele Codiereinheiten in Restkategorien fallen, sind weitere Überarbeitungen erforderlich.

Im Laufe der Kategorienbildung sind alle Unterkategorien, in die Codiereinheiten eingeordnet werden, präzise zu definieren. Eine Kategoriendefinition besteht aus drei bis vier Teilen: Einem Namen, einer Beschreibung, Ankerbeispielen und, soweit erforderlich, Entscheidungsregeln (Schreier, 2012). Der Name einer Kategorie sollte möglichst den Kern einer Kategorie widerspiegeln. Die Beschreibung sollte Charakteristika der Kategorie oder Indikatoren umfassen, die darauf hinweisen, dass eine Codiereinheit in diese Kategorie gehört. Auch ist es sinnvoll zu erwähnen, was bei der Zuordnung zu einer Kategorie keine Rolle spielt, um Unsicherheiten bei der Zuordnung zu vermeiden. Die Ankerbeispiele dienen der Veranschaulichung der Kategorienbeschreibung. Sie sollten, abgesehen von einer rein deduktiven Kategorienbildung, aus dem Datensatz stammen, der in die Entwicklung des Kategoriensystems miteinbezogen wurde. Die Auswahl der Ankerbeispiele ist auf möglichst wenige, typische Fälle zu beschränken, so dass die Ankerbeispiele die Verständlichkeit der Kategorienbeschreibung unterstützen und die Handhabbarkeit des in der Regel per se umfangreichen Codierleitfadens nicht gefährden. Name, Beschreibung und Ankerbeispiele sind viertens bei Bedarf durch Entscheidungsregeln zu ergänzen, die den Umgang mit Grenzfällen vorgeben. Der Aufbau des Kategoriensystems und alle Kategoriendefinitionen werden üblicherweise in einem Codierleitfaden festgehalten, der dann die Grundlage – sozusagen die Anleitung – für die Codierung des zu untersuchenden Materials bildet.

7.2.1.4 Die Pilotphase

Wenn das Kategoriensystem und der Codierleitfaden scheinbar fertig sind, erfolgt zunächst eine Pilotphase, um die Qualität des Kategoriensystems zu überprüfen und zu optimieren. In der Pilotphase wird das Kategoriensystem zuerst auf einen Teil des zu untersuchenden Datenmaterials angewendet.[110] Dafür sind, soweit möglich, Codiereinheiten auszuwählen, die nicht in die Entwicklung des Kategoriensystems einbezogen wurden. Außerdem sollten möglichst

[110] Eine Ausnahme bilden Untersuchungen, die auf eine Hypothesentestung zielen (Schreier, 2012). Diese Möglichkeit ist in der vorliegenden Arbeit jedoch unrelevant und bleibt deshalb an dieser Stelle und auch bei allen weiteren Ausführungen unberücksichtigt.

unterschiedliche Codiereinheiten berücksichtigt werden, damit viele Unterkategorien des Kategoriensystems in der **Probecodierung** ausprobiert werden. Über die genaue Anzahl an zu berücksichtigenden Codiereinheiten ist in Abwägung der Variabilität des Materials und der praktischen Umsetzbarkeit zu entscheiden. Schreier (2012) empfiehlt, etwa zehn bis zwanzig Prozent des gesamten zu untersuchenden Materials für die Probecodierung auszuwählen.

Die Probecodierung der ausgewählten Daten kann durch die forschende Person alleine oder durch mehrere Personen erfolgen. Wenn mehrere Personen mitwirken, ist sicherzustellen, dass alle Beteiligten vor der Probecodierung mit dem Forschungsanliegen und dem Codierleitfaden vertraut werden. Das bedeutet, dass allen Beteiligten Zeit einzuräumen ist, den Codierleitfaden zu lesen und dessen Anwendung zunächst gemeinsam mit der forschenden Person und anschließend selbstständig auszuprobieren. Nach dieser Eingewöhnungsphase sind alle für die Probecodierung ausgewählten Daten durch zwei (oder mehr) Personen unabhängig voneinander zu analysieren. Schwierigkeiten bei der Durchführung der Codierungen werden von den einzelnen Personen notiert, jedoch nicht diskutiert.

Nach der Probecodierung wird das **Kategoriensystem evaluiert und ggf. modifiziert**. Üblicherweise wird zu jedem Merkmal des Kategoriensystems zuerst ein Maß für die Übereinstimmung der Codierungen berechnet. Wenn zwei (oder mehr) Personen das Material unabhängig voneinander codiert haben, handelt es sich um ein Maß für Intercoderreliabilität. Reliabilität ist ein Gütekriterium aus der quantitativen Forschungstradition.

> Unter Reliabilität beziehungsweise Zuverlässigkeit (auch die Begriffe Genauigkeit und Präzision werden benutzt) versteht man ein Maß für die Reproduzierbarkeit der Messergebnisse. Dasselbe Ergebnis soll bei wiederholter Anwendung des Instruments [Erhebungs- oder Messinstruments] stets wieder erzielt werden. (Häder, 2015, S. 104)

Intercoderreliabilität beschreibt, inwieweit die Ergebnisse der Codierung nicht nur subjektive, sondern intersubjektive Gültigkeit haben. Es handelt sich also primär um ein Maß für Objektivität im Sinne der Unabhängigkeit der Ergebnisse von der forschenden Person (Mayring, 2010; Schreier, 2013). Eine hohe Übereinstimmung der Codierungen zwischen verschiedenen Personen trägt aber zugleich zu einer niedrigen Fehleranfälligkeit und somit zur Reproduzierbarkeit der Messergebnisse bei, wie sie mit dem Kriterium der Reliabilität angestrebt wird (Schreier, 2013).

Die Ergebnisse der Berechnung[111] von Werten für die Intercoderreliabilität als zahlenmäßige, zusammenfassende Angaben weisen darauf hin, welche Kategorien inkonsistent verwendet werden und überarbeitet werden sollten. Zusätzlich zu diesen Rechenergebnissen ist für deren angemessene Interpretation und eine effektive Überarbeitung zu berücksichtigen, wo genau Abweichungen sind und warum sich die Codierungen unterscheiden.[112] Auch sind die während der Probecodierung notierten Schwierigkeiten zu diskutieren. Darüber hinaus sollten Anhalts-

[111] Zur rechnerischen Bestimmung der Intercoderreliabilität siehe Abschnitt 7.2.2.

[112] Beispielsweise können Abweichungen systematisch auftreten, da Kategorien nicht trennscharf definiert oder Kategoriendefinitionen unterschiedlich interpretiert wurden. Auch ist es möglich, dass nicht die Kategoriendefinitionen, sondern die zu analysierenden Codiereinheiten verschieden interpretiert und deshalb unterschiedlich codiert wurden. Zudem können Flüchtigkeitsfehler passiert sein (Schreier, 2012).

punkte für die Validität des Kategoriensystems in die Evaluation des Kategoriensystems miteinbezogen werden. Validität bezeichnet in der quantitativen Forschung den „Grad der Genauigkeit, mit dem ein Test das misst, was er messen soll“ (Häder, 2015, S. 109). In einer qualitativ-strukturierenden Inhaltsanalyse stellt sich vor allem die Frage, inwieweit das Kategoriensystem die zu untersuchenden Charakteristika der Materialien adäquat erfasst (Schreier, 2012). Eine konsistente Anwendbarkeit des Kategoriensystems ist eine Voraussetzung für Validität, reicht jedoch nicht aus. Weitere Anhaltspunkte für die Validität des Kategoriensystems geben Beurteilungen des Kategoriensystems durch Expertinnen und Experten, die sich mit den zugrundeliegenden Theorien auskennen, sowie die Häufigkeiten der Zuordnungen der Codierungen zu den verschiedenen Kategorien in der Pilotierung, wie z. B. überproportional viele Zuordnungen zu einer einzigen Unterkategorie (Schreier, 2012). Aufbauend auf den Ergebnissen dieser umfassenden Evaluation ist das Kategoriensystem zu optimieren und für die anschließende Hauptcodierung zu finalisieren.

7.2.1.5 Die Hauptcodierung

In der Hauptcodierung wird das gesamte zu untersuchende Material entsprechend dem Kategoriensystem sukzessive codiert. Jede Codiereinheit wird separat betrachtet und bezüglich aller Merkmale des Kategoriensystems eingeordnet. Dabei wird die in der Pilotphase erstellte, finale Version des Kategoriensystems verwendet, und nicht weiter verändert (Schreier, 2012).

Um zu überprüfen, ob diese finale Version des Kategoriensystems besser ist als die vorhergehende aus der Probecodierung, und um die Qualität des finalen Kategoriensystems zu dokumentieren[113], sollte auch im Rahmen der Hauptcodierung ein Teil des Materials doppelt codiert werden. Danach ist zusätzlich zur Berechnung der Intercoderreliabilität im Falle unterschiedlicher Codierungen jeweils eine auszuwählen, da die weitere Auswertung der Codierungen und anschließende Ergebnispräsentation eindeutige Zuordnungen aller Codiereinheiten erfordert. Diese Entscheidung für eine von zwei unterschiedlichen Codierungen ist leicht, wenn bei einer der Codierungen ein Flüchtigkeitsfehler passiert ist. Ansonsten sind verschiedenen Vorgehensweisen denkbar. Es ist möglich, eine der beiden Zuordnungen im Rahmen der Reflexion begründet auszuwählen, sich für die Zuordnung eines bestimmten Codierenden zu entscheiden, die Codiereinheit bezüglich des infrage stehenden Merkmals in eine Restkategorie einzuordnen oder für die Entscheidung eine weitere Person hinzuziehen (Schreier, 2012).

7.2.1.6 Weitere Auswertung und Ergebnisdarstellung

Nach Abschluss der Hauptcodierung werden die vorgenommenen Zuordnungen weiter ausgewertet und mit Blick auf die Forschungsfragen präsentiert.

Mit Blick auf die Forschungsfragen kann es sich anbieten, die **Häufigkeiten der Codierungen** zu analysieren und mittels Tabellen, Diagrammen und ergänzendem Fließtext zu präsentieren. In diesem Rahmen ist es möglich, verschiedene Gruppen, wie etwa Schülerinnen und Schüler

[113] Im Vergleich zu der Pilotphase geht es nicht darum, das Kategoriensystem zu optimieren, sondern die Qualität des Kategoriensystems zu dokumentieren und ggf. Schwierigkeiten aufzudecken, die bei der Interpretation und Diskussion der Ergebnisse zu bedenken sind.

aus unterschiedlichen Klassenstufen, bezüglich der Merkmale zu vergleichen sowie Zusammenhänge zwischen verschiedenen Merkmalen der analysierten Daten zu explorieren. Um eine angemessene Interpretation augenscheinlicher Unterschiede und Zusammenhänge zu unterstützen, können Methoden der Inferenzstatistik angewendet werden (Abschn. 7.2.2).

Für eine vertiefende inhaltliche Analyse kann es sinnvoll sein, **Fallbeschreibungen** durchzuführen, das heißt einzelne Fälle, wie etwa einzelne Beweisversuche, im Detail zu betrachten oder gegenüberzustellen. Ausgehend von den Ergebnissen der qualitativ-strukturierenden Inhaltsanalyse ist es möglich, Fälle mit etwa besonders typischen, extremen oder aus theoretischer Sicht interessanten Ausprägungen auszuwählen sowie mehrere Fälle, die ein Spektrum an Ausprägungen aufzeigen, zusammenzustellen. Diese Fälle können dann ganzheitlich untersucht und beschrieben werden (Schreier, 2012, 2013; Mayring, 2010).

Die Beschreibung einzelner Fälle kann in Zusammenhang mit einer **Typenbildung** erfolgen. Eine Typenbildung ist ein Prozess, in dem aufgrund des Vergleichs zuvor definierter Merkmale einander ähnliche Fälle identifiziert und so zu Gruppen bzw. Typen zusammengefasst werden, dass sich die Fälle innerhalb eines Typus einander möglichst ähnlich sind („interne Homogenität auf der ‚Ebene des Typus'" (Kelle & Kluge, 2010, S. 85, H. i. O.)) und sich die Typen untereinander dagegen möglichst stark unterscheiden („externe Heterogenität auf der ‚Ebene der Typologie'" (Kelle & Kluge, 2010, S. 85, H. i. O.)) (Kelle & Kluge, 2010; Kluge, 2000). „Mittels Typen und Typologien kann eine komplexe Realität auf wenige Gruppen beziehungsweise Begriffe reduziert werden, um sie greifbar, und damit begreifbar zu machen" (Kluge, 1999, S. 85). Eine Typenbildung bietet sich insbesondere an, um große Datensätze übersichtlich darzustellen (Hussy et al., 2013; Kelle & Kluge, 2010). Zudem kann sie dazu beitragen, charakteristische Züge der Fälle und inhaltliche Sinnzusammenhänge aufzudecken (Kelle & Kluge, 2010). Die gebildeten Typen lassen sich etwa unter Berücksichtigung ihrer Merkmalsausprägungen und Häufigkeitsverteilungen, inhaltlicher Sinnzusammenhänge und prototypischer Fälle präsentieren (Kelle & Kluge, 2010).

Eine Typenbildung lässt sich als eine besondere Form einer **Klassifikation** ansehen. Bei einer Klassifikation geht es um ein systematisches Ordnen, oftmals eine hierarchische Untergliederung, mithilfe einer prinzipiell beliebigen Anzahl an disjunkten und exhaustiven Kategorien (Klimke et al., 2020; PONS Langenscheidt GmbH, o. J.; Wirtz, 2021).

> Klassifikation […] ist die Zuordnung von Objekten zu Klassen (Kategorien) anhand der Ähnlichkeit eines best. Merkmals (eindimensionale Klassifikation) oder mehrerer Merkmale (multidimensionale Klassifikation). In der einfachsten Form müssen für eine Klassifikation zwei Grundregeln erfüllt sein. Erstens muss das Merkmal, auf dem die Klassifikation beruht, mind. zwei unterschiedliche Werte besitzen. Zweitens müssen die Klassen exhaustiv und sich gegenseitig ausschließend sein. Klassifikationen bez. aber nicht nur den Prozess der Zuordnung, sondern auch das Ergebnis dieses Prozesses wird Klassifikation genannt. Eine Klassifikation oder Systematik als Resultat ist eine planmäßige Zus.stellung von Klassen oder Kategorien. (Wirtz, 2021, o. S.)

So wie mit einer Typenbildung kann mithilfe einer Klassifikation „eine große Zahl an Informationen bzw. Daten nach bestimmten Kriterien angeordnet, in ihrer Komplexität reduziert und in übersichtlicher Form dargestellt werden" (Stangl, 2023, o. S.). Dabei wird, im Unterschied

zu einer Typenbildung, nicht unbedingt die Bildung weniger, klar distinkter Kategorien angestrebt. Zugleich stehen bei einer Klassifikation, im Vergleich zu einer beliebigen Kategorisierung, neben den Eigenschaften der einzelnen Kategorien immer auch die Beziehungen zwischen verschiedenen Kategorien im Fokus (Straehler-Pohl & Gellert, 2012). Vor allem im Zuge multidimensionaler Klassifikationen lassen sich, so wie bei einer Typenbildung nach Kluge (1999, 2000), Strukturen und inhaltliche Sinnzusammenhänge in Daten herauszuarbeiten. Eine Klassifikation kann möglich und sinnvoll sein, um Ergebnisse zu verdichten, zu strukturieren und zu resümieren, auch wenn eine Typenbildung im Speziellen nicht angemessen ist, etwa weil sie erfordern würde zu trennen, was nicht klar zu trennen ist oder Zusammenfassungen vorzunehmen, durch die wesentliche Informationen verloren gehen.

7.2.2 Statistische Grundlagen

Bei der Durchführung der qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) im Rahmen dieser Arbeit werden Verfahren der Stichprobenziehung und Methoden der Datenanalyse verwendet. Für die Auswertung werden aus der Gesamtmenge der erhobenen Daten Stichproben gezogen. Zudem werden im Rahmen der Entwicklung und Pilotierung des Kategoriensystems Stichprobenziehungen durchgeführt. Um das Kategoriensystem zu evaluieren, werden Verfahren zur Berechnung der Intercoderreliabilität angewendet. Außerdem werden in der weiteren Auswertung der Codierungen Verfahren zur Überprüfung von Unterschieds- und Zusammenhangshypothesen verwendet (Abschn. 7.3). Im Folgenden werden die im Rahmen dieser Arbeit relevanten Verfahren der Stichprobenziehung und Methoden der Datenanalyse erläutert.

7.2.2.1 Arten der Stichprobenziehung

In einer Stichprobenziehung wird aus einer Grundgesamtheit, d. h. der „Menge aller potenziellen Untersuchungsobjekte für eine gegebene Fragestellung“ (Hussy, 2013, S. 118), eine Teilmenge ausgewählt. Diese Auswahl kann auf verschiedene Weisen durchgeführt werden. Im Folgenden wird auf Zufallsstichproben und auf bewusste Stichprobenziehungen eingegangen, weil genau diese bei der Datenanalyse im Rahmen dieser Arbeit relevant sind (Abschn. 7.3.3).

Um die Ergebnisse einer Untersuchung auf die infrage stehende Grundgesamtheit zu übertragen, sind **Zufallsstichproben** am besten geeignet. Dazu gehören u. a. einfache Zufallsstichproben und geschichtete Zufallsstichproben. In einer einfachen Zufallsstichprobe hat jedes Objekt der Grundgesamtheit die gleiche Chance, ausgewählt zu werden. „Eine aus n Objekten bestehende Teilmenge, welche ohne Zurücklegen aus der Grundgesamtheit ausgewählt wird, heißt einfache Zufallsstichprobe, falls alle gleich großen und auf die gleiche Weise selektierten Teilmengen, gleich wahrscheinlich sind“ (Bortz & Schuster, 2010, S. 80). In einer geschichteten Zufallsstichprobe wird die Grundgesamtheit nach bestimmten Merkmalen in Teilmengen, sogenannte Schichten, unterteilt. Anschließend wird aus den gebildeten Teilmengen proportional zur prozentualen Verteilung der Schichtungsmerkmale in der Grundgesamtheit zufällig gezogen (Bortz, 2010). Bezüglich der Größe einer Zufallsstichprobe gilt folgende Tendenz: „Je größer die Stichprobe, desto höher [ist] die Wahrscheinlichkeit, eine Stichprobe erhoben zu

haben, die typisch für die zugehörige Population ist“ (Hussy, 2013, S. 119). Zugleich ist die praktische Umsetzbarkeit der Untersuchung zu bedenken.

Neben einer Zufallsauswahl besteht die Möglichkeit einer **bewussten Stichprobenziehung**.

> Bei der bewussten bzw. absichtsvollen Stichprobenziehung wird die Stichprobe gezielt nach bestimmten Kriterien aus der Grundgesamtheit ausgewählt. Bei Bottom-up-Strategien der bewussten Stichprobenziehung ergeben sich diese Kriterien erst im Untersuchungsverlauf; bei Top-down-Strategien stehen sie zu Untersuchungsbeginn fest. (Schreier, 2013, S. 194)

Ein Bottom-up-Verfahren der bewussten Stichprobenziehung ist die theoretische Stichprobenziehung. Mit dem Ziel, eine möglichst große Variabilität abzudecken, werden so lange neue, möglichst verschiedene Objekte in die Stichprobe aufgenommen, bis die Einbeziehung zusätzlicher Objekte keine neuen Variationen aufzeigt. Ein Top-down-Verfahren, um eine möglichst große Variabilität abzubilden, ist ein qualitativer Stichprobenplan. Vor Beginn der Ziehung werden Faktoren ausgewählt, die die relevanten Eigenschaften der Untersuchungsobjekte wahrscheinlich beeinflussen, und diese Faktoren miteinander kombiniert. Anschließend werden zu jeder der Kombinationsmöglichkeiten, soweit vorhanden, ein oder mehrere Objekte ausgewählt. Bei jeder bewussten Stichprobenziehung ist die Zusammensetzung der Stichprobe, weniger deren Umfang entscheidend (Schreier, 2013).

7.2.2.2 Die Berechnung von Intercoderreliabilität

Bei der Pilotierung und Hauptcodierung im Rahmen einer qualitativ-strukturierenden Inhaltsanalyse wird üblicherweise zu jedem Merkmal ein Wert für die Intercoderreliabilität, d. h. für die Übereinstimmung zwischen den Codierungen mehrerer Personen berechnet (Abschn. 7.2.1.4, 7.2.1.5). Dementsprechend wird auch in Rahmen dieser Arbeit vorgegangen, wobei bei der Pilotierung und der Hauptcodierung jeweils genau zwei Personen involviert sind (Abschn. 7.3.6, 7.3.7). Dazu passend wird bei der folgenden Erläuterung der Berechnung von Intercoderreliabilität davon ausgegangen, dass jede ausgewählte Codiereinheit durch genau zwei verschiedene Personen codiert wurde.

Das einfachste Maß für Intercoderreliabilität ist nach Wirtz und Caspar (2002) die prozentuale Übereinstimmung ($\mathrm{PÜ}$). Sie gibt den prozentualen Anteil der Codiereinheiten an, die von den zwei Personen bezüglich eines fokussierten Merkmals gleich codiert wurden.

$$\mathrm{PÜ} = \frac{\text{Anzahl der von beiden Personen identisch codierten Codiereinheiten}}{\text{Anzahl der insgesamt codierten Codiereinheiten}} \cdot 100\%$$

Die prozentuale Übereinstimmung ist allerdings als Maß für die Übereinstimmungen zwischen den zwei codierenden Personen nur beschränkt aussagekräftig, insbesondere weil „der $\mathrm{PÜ}$ nicht zu entnehmen ist, in welchem Maße der gemessene Wert größer ist als er bei rein zufälligem Beurteilungsverhalten sein müsste“ (Wirtz & Caspar, 2002, S. 50). Durch die Möglichkeit der zufälligen Übereinstimmung ist der gleiche Prozentwert weniger aussagekräftig, wenn es sich um ein Merkmal mit zwei möglichen Ausprägungen handelt, als wenn eine Vielzahl an Ausprägungen in Form von Kategorien unterschieden werden. Vor allem wenn ein Merkmal nur durch zwei mögliche Kategorien beschrieben wird, überschätzt die prozentuale Übereinstimmung die „wahre“ Übereinstimmung (Wirtz & Caspar, 2002). Deshalb wird im Rahmen

dieser Arbeit anstatt der prozentualen Übereinstimmung ein zufallsbereinigtes Maß, und zwar Cohens Kappa, verwendet. Cohens Kappa ist nach Wirtz und Caspar (2002) ein gutes, geeignetes zufallsbereinigtes Übereinstimmungsmaß und hat sich in der Literatur als gängiges zufallsbereinigtes Maß durchgesetzt. Wirtz und Caspar (2002) erläutern:

> Cohens κ (kappa) und Scotts π (pi) basieren auf der prozentualen Übereinstimmung (PÜ) und haben gegenüber dieser den Vorteil, dass sie das Verhältnis der beobachteten zu der bei Zufall erwarteten Übereinstimmung berücksichtigen: κ und π liefern somit eine standardisierte Maßzahl (zwischen -1 und $+1$) für das Ausmaß, in dem die tatsächlich beobachtete Übereinstimmung positiv von der Zufallserwartung abweicht. […] κ und π sind Maße, die direkt zur Quantifizierung der Beobachterübereinstimmung entwickelt wurden. Cohens κ ist das am häufigsten angewandte Maß zur Bestimmung der Übereinstimmung. (Wirtz & Caspar, 2002, S. 55-56)

Die Berechnung des Kappa-Koeffizienten (κ) ist folgendermaßen definiert:

$$\kappa = \frac{\text{Differenz der } \frac{\text{beobachteten PÜ}}{100\%} \text{ und der } \frac{\text{bei Zufall erwarteten PÜ}}{100\%}}{\text{Differenz der } \frac{\text{maximal möglichen PÜ}}{100\%} \text{ und der } \frac{\text{bei Zufall erwarteten PÜ}}{100\%}}$$

$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

Dabei entspricht P_0 dem relativen Anteil der Codiereinheiten, bei denen die zwei Personen identische Codierungen vorgenommen haben und P_e dem erwarteten relativen Anteil der Übereinstimmungen bei zufälligem Codierverhalten. Dieser bei zufälligem Codierverhalten erwartete relative Anteil der Übereinstimmung ist bei Cohens κ durch

$$P_e = \sum_1^n h_{j.} \cdot h_{.j}$$

definiert, wobei n die Anzahl an Kategorien und $h_{j.}$ bzw. $h_{.j}$ die Randsummen der relativen Häufigkeiten der Entscheidungen der zwei Personen für die Kategorie j darstellen.

Die Berechnungen der Intercoderreliabilität werden im Rahmen dieser Arbeit mit IBM SPSS Statistics für Windows (Version 26) durchgeführt. Für die Einschätzung der Güte der Übereinstimmung auf Grundlage der erhaltenen Kappa-Koeffizienten werden im Rahmen dieser Arbeit in Anlehnung an die Ausführungen von Wirtz und Caspar (2002, S. 59) die in Tabelle 4 angegebenen Wertebereiche verwendet.

Kappa-Koeffizient	Übereinstimmung
$0 < \kappa \leq 0{,}4$	gering
$0{,}4 < \kappa \leq 0{,}6$	akzeptabel
$0{,}6 < \kappa \leq 0{,}75$	gut
$0{,}75 < \kappa \leq 1$	sehr gut

Tabelle 4: Wertebereiche des Kappa-Koeffizienten für die Einschätzung der Güte der Übereinstimmung in Anlehnung an Wirtz und Caspar (2002, S. 59)

7.2.2.3 Verfahren zur Prüfung von Unterschieds- und Zusammenhangshypothesen

Dieser Abschnitt beschäftigt sich mit statistischen Verfahren zur Überprüfung von Unterschieds- und Zusammenhangshypothesen. Ziel der Anwendung dieser Verfahren ist im Rahmen dieser Arbeit, Unterschiede oder Zusammenhänge, etwa Unterschiede zwischen Beweisversuchen von Lernenden aus verschiedenen Klassenstufen oder Zusammenhänge zwischen verschiedenen Merkmalen der Beweisversuche, statistisch zu überprüfen und auf diese Weise

eine angemessene Interpretation erhaltener Daten zu unterstützen (Abschn. 7.2.1.6, 7.3.8; Kap. 8). Welches statistische Verfahren zur Überprüfung von Unterschieds- und Zusammenhangshypothesen geeignet ist, hängt u. a. vom Skalenniveau der Daten ab. Im Rahmen der vorliegenden Arbeit sind vor allem Verfahren, die sich auch für nominal skalierte Merkmale eignen, von Bedeutung. Deshalb wird im Folgenden insbesondere der Chi-Quadrat-Test vorgestellt. Außerdem wird für den Vergleich zweier Gruppen in Bezug auf ein ordinal skaliertes Merkmal der Mann-Whitney-U-Test verwendet. Alle Tests werden im Rahmen dieser Arbeit mit IBM SPSS Statistics für Windows (Version 26) durchgeführt.

Verfahren für nominal skalierte Daten

Für die Analyse von Unterschieds- und Zusammenhangshypothesen bei nominal skalierten Merkmalen wird im Rahmen der vorliegenden Arbeit i. d. R. der **Chi-Quadrat-Test** nach Karl Pearson (im Folgenden kurz: Chi²-Test) verwendet. Dieser ist sowohl für Merkmale mit genau zwei möglichen Merkmalsausprägungen (dichotome Merkmale) als auch für Merkmale mit mehr als zwei möglichen Merkmalsausprägungen geeignet (Bortz & Schuster, 2010). Er überprüft, ob ein Zusammenhang bzw. Unterschied statistisch signifikant ist oder als zufällig bezeichnet werden muss.

> Die Nullhypothese postuliert die stochastische Unabhängigkeit [zweier] Merkmale. Die Alternativhypothese fordert einen irgendwie gearteten Zusammenhang zwischen den Merkmalsausprägungen des einen Merkmals und jenen des anderen. (Benesch, 2013, S. 225)

Um die Nullhypothese zu testen, werden die beobachteten Häufigkeiten der Merkmalsausprägungen mit erwarteten Häufigkeiten anhand von Kreuztabellen verglichen. Es werden also zunächst Häufigkeiten berechnet, welche bei der Gültigkeit der Nullhypothese, d. h. im Falle empirischer Unabhängigkeit, zu erwarten wären. Wenn erhebliche Differenzen zwischen diesen erwarteten und den tatsächlich beobachteten Häufigkeiten der Merkmalsausprägungen bestehen, so spricht dies gegen die Nullhypothese (Bortz & Schuster, 2010).

		Ausprägungen von Y						Summe
		y_1	y_2	$\cdots$	y_j	$\cdots$	y_q	
Ausprägungen von X	x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1q}	$n_{1\cdot}$
	x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2q}	$n_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iq}	$n_{i\cdot}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	x_p	n_{p1}	n_{p2}	$\cdots$	n_{ij}	$\cdots$	n_{pq}	$n_{p\cdot}$
Summe		$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot q}$	n

		Ausprägungen von Y						Summe
		y_1	y_2	$\cdots$	y_j	$\cdots$	y_q	
Ausprägungen von X	x_1	$\widetilde{n_{11}}$	$\widetilde{n_{12}}$	$\cdots$	$\widetilde{n_{1j}}$	$\cdots$	$\widetilde{n_{1q}}$	$n_{1\cdot}$
	x_2	$\widetilde{n_{21}}$	$\widetilde{n_{22}}$	$\cdots$	$\widetilde{n_{2j}}$	$\cdots$	$\widetilde{n_{2q}}$	$n_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	x_i	$\widetilde{n_{i1}}$	$\widetilde{n_{i2}}$	$\cdots$	$\widetilde{n_{ij}}$	$\cdots$	$\widetilde{n_{iq}}$	$n_{i\cdot}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	x_p	$\widetilde{n_{p1}}$	$\widetilde{n_{p2}}$	$\cdots$	$\widetilde{n_{pj}}$	$\cdots$	$\widetilde{n_{pq}}$	$n_{p\cdot}$
Summe		$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot q}$	n

Abbildung 77: Kreuztabelle mit beobachteten Häufigkeiten sowie dazugehörige Kreuztabelle mit erwarteten Häufigkeiten bei empirischer Unabhängigkeit zweier Merkmale X und Y mit p bzw. q ($p, q \in \mathbb{N}_{\geq 2}$) möglichen Ausprägungen

Die Abbildung 77 zeigt eine Kreuztabelle mit beobachteten Häufigkeiten und die dazugehörige Kreuztabelle mit den erwarteten Häufigkeiten bei empirischer Unabhängigkeit zweier Merkmale X und Y mit p bzw. q ($p, q \in \mathbb{N}_{\geq 2}$) möglichen Ausprägungen. Dabei bezeichnet n den Stichprobenumfang (im Rahmen der vorliegenden Arbeit konkret eine Anzahl an analysierten

Codiereinheiten), n_{ij} mit $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$ die beobachtete Häufigkeit der Kombination der Merkmalsausprägungen x_i, y_j und $\widetilde{n_{ij}}$ mit $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$ die erwartete Häufigkeit der Kombination der Merkmalsausprägungen x_i, y_j. Die erwarteten Häufigkeiten $\widetilde{n_{ij}}$ der einzelnen Merkmalskombinationen sind in Einklang mit der Definition empirischer Unabhängigkeit zweier Merkmale[114] definiert als

$$\widetilde{n_{ij}} := \frac{n_{i\cdot} n_{\cdot j}}{n},\ i \in \{1, \dots, p\},\ j \in \{1, \dots, q\}.$$

Beim Chi²-Test werden die beobachteten und erwarteten Häufigkeitsverteilungen mittels relativer, quadrierter Differenzen verglichen und zum Chi²-Wert (χ^2) zusammengefasst. Dieser ist definiert, wenn keine der erwarteten Häufigkeiten bzw. keine der Randsummen Null ist.

$$\chi^2 := \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{\left(n_{ij} - \widetilde{n_{ij}}\right)^2}{\widetilde{n_{ij}}} \text{ mit } \widetilde{n_{ij}} := \frac{n_{i\cdot} n_{\cdot j}}{n},\ i \in \{1, \dots, p\},\ j \in \{1, \dots, q\}.$$

Für den einfachen Fall zweier dichotomer Merkmale ($p = q = 2$) gilt folglich:

$$\chi^2 = \frac{n\,(n_{11} n_{22} - n_{12} n_{21})^2}{n_{1\cdot} n_{2\cdot} n_{\cdot 1} n_{\cdot 2}}$$

Der Chi²-Wert (χ^2) ist i. d. R. approximativ Chi-Quadrat-verteilt[115] mit $(p-1)(q-1)$ Freiheitsgraden[116]. Benesch (2013) erläutert:

> Die χ^2-Verteilung assoziiert jeden χ^2-Wert bei gegebener Anzahl der Freiheitsgrade mit einer bestimmten Wahrscheinlichkeit. Diese gibt an, wie wahrscheinlich die empirisch gefundene Häufigkeitsverteilung unter Annahme der Nullhypothese der Unabhängigkeit ist. Je größer der χ^2-Wert, umso unwahrscheinlicher ist es, dass in der Population, aus der die Stichprobe stammt, die Nullhypothese gilt. Ist die Wahrscheinlichkeit (der p-Wert) kleiner als ein vorher festgelegtes Signifikanzniveau, so wird die Nullhypothese verworfen. (Benesch, 2013, S. 225)

Im Rahmen der vorliegenden Arbeit wird als Signifikanzniveau $\alpha = 0{,}05$ bzw. $\alpha = 0{,}01$ verwendet, d. h. eine Wahrscheinlichkeit von 5% bzw. 1% für einen Fehler erster Art in Kauf genommen. Wenn ein Unterschied bzw. Zusammenhang auf dem 5%-Niveau signifikant ist ($p < 0{,}05$), wird dieser im Rahmen der vorliegen Arbeit als signifikant bezeichnet. Wenn ein Unterschied bzw. Zusammenhang auf dem 1%-Niveau signifikant ist ($p < 0{,}01$), wird er im Rahmen der vorliegenden Arbeit als hoch signifikant bezeichnet.

Der Chi²-Test nach Karl Pearson wird im Rahmen dieser Arbeit in der Regel, jedoch nicht immer verwendet. Für den Spezialfall einer 2 x 2-Kreuztabelle zeigt eine Simulationsstudie, dass der **$N - 1$ Chi²-Test** nach Egon Pearson angemessenere p-Werte liefert als der Chi²-Test nach Karl Pearson (Busing et al., 2015). Deshalb wird im Rahmen dieser Arbeit, falls eine 2 x 2 - Konti-

[114] Die Merkmale X und Y heißen empirisch unabhängig, wenn für alle $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$ gilt: $\frac{n_{ij}}{n} = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n}$ (z. B. Bortz & Schuster, 2010).

[115] Informationen zur Chi-Quadrat-Verteilung sind z. B. bei Bourier (2018) zu finden.

[116] „Grundsätzlich gilt, dass die Randsummen der erwarteten Häufigkeiten mit den Randsummen der beobachteten Häufigkeiten übereinstimmen müssen. […] Dies hat zur Konsequenz, dass nicht alle erwarteten Häufigkeiten frei wählbar sind“ (Bortz & Schuster, 2010, S. 140). In jeder Spalte können $p - 1$ erwartete Häufigkeiten frei gewählt werden, in jeder Zeile $q - 1$ erwartete Häufigkeiten, woraus insgesamt $df = (p-1)(q-1)$ Freiheitsgrade resultieren. Im Falle einer 2 x 2 - Kreuztabelle gilt also $df = 1$; wenn eine der vier Häufigkeiten $\widetilde{n_{ij}}$ $(i, j \in \{1, 2\})$ bekannt ist, ergeben sich die anderen drei aus den Randhäufigkeiten (Bortz & Schuster, 2010).

genztafel vorliegt, d. h. zwei dichotome Merkmale betrachtet werden, und die Voraussetzungen für einen Chi²-Test erfüllt sind, der $N-1$ Chi²-Test verwendet. Beim $N-1$ Chi²-Test wird im Vergleich zu der Formel für die Berechnung des Chi²-Werts (χ^2) der Stichprobenumfang n durch $n-1$ ersetzt (Busing et al., 2015). An die obigen Ausführungen zur Berechnung des Chi²-Werts (χ^2) im Fall zweier dichotomer Merkmale ($p = q = 2$) anknüpfend gilt

$$\chi^2_{N-1} = \frac{(n-1)\,(n_{11}n_{22}-n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}} = \frac{n-1}{n}\chi^2.$$

Der $N-1$ Chi²-Test nach Egon Pearson entspricht im Fall zweier dichotomer Merkmale dem Mantel-Haenszel Chi²-Test, dessen Ergebnis in SPSS bei der Durchführung eines Chi²-Tests unter der Bezeichnung „Linear-by-Linear Association" ausgegeben wird (Busing et al., 2015).

Zusätzlich zum Ausnahmefall zweier dichotomer Merkmale sind bei der Nutzung eines Chi²-Tests mehrere Voraussetzungen zu beachten. Der Chi²-Wert (χ^2) und der $N-1$ Chi²-Wert (χ^2_{N-1}) sind ausschließlich definiert, wenn keine der erwarteten Häufigkeiten bzw. keine der Randsummen gleich Null ist. Zudem sollten für die Approximation durch die Chi-Quadrat-Verteilung die erwarteten absoluten Häufigkeiten größer oder gleich fünf sein (Bourier, 2018). Falls eine dieser Voraussetzung nicht erfüllt ist, besteht ein Lösungsansatz darin, mehrere Ausprägungen eines Merkmals inhaltlich passend zusammenzufassen, d. h. die Kreuztabelle etwa von einer 6 x 2 - auf eine 3 x 2 - Tabelle zu reduzieren, und dann den Chi²-Test anzuwenden. Wenn eine der Voraussetzungen für einen Chi²-Test nicht erfüllt ist und eine Reduktion der Kreuztabelle nicht infrage kommt, wird im Rahmen dieser Arbeit stets der **exakte Fisher-Test** (Fisher-Yates-Test) genutzt, der auch bei geringen erwarteten Häufigkeiten anwendbar ist.

Beim exakten Fisher-Test werden so wie beim Chi²-Test die beobachteten absoluten Häufigkeiten in einer Kreuztabelle angeordnet. Auf Grundlage der als festliegend angesehenen Randsummenverteilungen wird dann zuerst die Punktwahrscheinlichkeit des zufälligen Auftretens der beobachteten Häufigkeitsverteilung berechnet. Zusätzlich zu der Wahrscheinlichkeit dieser einen beobachteten Häufigkeitsverteilung werden anschließend noch die Punktwahrscheinlichkeiten für alle möglichen Häufigkeitsverteilungen, die stärker als die beobachtete Häufigkeitsverteilung auf die Alternativhypothese hindeuten, berechnet. Die Summe aller dieser Punktwahrscheinlichkeiten entspricht der Wahrscheinlichkeit (dem p-Wert), dass die beobachtete oder eine noch extreme Verteilung zustande kommt.[117] Ist $p < 0{,}05$ bzw. $p < 0{,}01$, wird analog zur Anwendung des Chi²-Test im Rahmen der vorliegenden Arbeit die Nullhypothese mit einer Fehlerwahrscheinlichkeit von 5% bzw. 1% verworfen.

Hinsichtlich der festgelegten Fehlerwahrscheinlichkeiten und damit der Interpretation erhaltener p-Werte ist beim **multiplen Testen** Vorsicht geboten. Multiples Testen meint, dass an einer Datendatei mehrere Signifikanztests durchgeführt werden. Die Wahrscheinlichkeit dafür, dass bei einem von mehreren Tests die Nullhypothese fälschlicherweise abgelehnt wird, ist deutlich größer als das festgelegte Signifikanzniveau (Bortz & Schuster, 2010). Diese Wahrscheinlichkeit lässt sich reduzieren und kontrollieren, indem man die Grenzen für den p-Wert

[117] Die Durchführung des exakten Fisher-Tests wird für den einfachen Fall eines einseitigen Tests auf Grundlage einer 2x2-Kreuztabelle von Bortz und Lienert (2008) ausführlich erläutert. Darüberhinausgehende Informationen zum exakten Fisher-Test sind bei Bortz et al. (2008) zu finden.

anpasst. Beispielsweise wird bei der Bonferroni-Korrektur für eine Ablehnung auf dem Signifikanzniveau α die Grenze des p-Werts durch die Anzahl an Einzeltests k dividiert, d. h. anstatt bei $p < \alpha$ wird ausschließlich im Fall $p < \frac{\alpha}{k}$ die Nullhypothese verworfen (Bortz & Lienert, 2008). Allerdings müssen die Tests dann sehr geringe p-Werte ergeben, um auf dem festgelegten Signifikanzniveau als statistisch signifikant zu gelten, so dass eventuell vorhandenen Unterschiede bzw. Zusammenhänge nicht erkannt werden (Bortz & Schuster, 2010). Deshalb wird im Rahmen dieser Arbeit durchgehend $p < 0{,}05$ verwendet, unter der Bewusstheit, dass der Alpha-Fehler nicht unter Kontrolle ist und die relativen Häufigkeitsverteilungen und Effektstärken substanziell mit interpretiert werden sollten, vor allem bei p-Werten, die nicht deutlich unterhalb von $p = 0{,}05$ liegen.

Dass die p-Werte stets vorsichtig interpretiert werden, ist außerdem wichtig, weil der p-Wert vom zu wählenden Signifikanztest und dem Stichprobenumfang abhängt, so dass die p-Werte aus verschiedenen Einzeltests nicht unbedingt vergleichbar sind. Beim exakten Fisher-Test ist der p-Wert i. d. R. größer bzw. konservativer als beim Chi²-Test. Zudem gilt bei sämtlichen Tests: Umso größer der Stichprobenumfang n ist, desto kleiner ist der p-Wert (Bortz & Schuster, 2010). „Die Nullhypothese ist bei sehr großen Stichproben gewissermaßen chancenlos, da bei zunehmendem Stichprobenumfang die Teststärke ansteigt und somit auch sehr kleine Effekte mit großer Wahrscheinlichkeit entdeckt werden können“ (Bortz & Schuster, 2010, S. 112). Um die Ergebnisse trotz verschiedener Tests und Stichprobenumfänge vergleichen und die Stärke von Zusammenhängen quantifizieren zu können, wird innerhalb dieser Arbeit der Zusammenhangskoeffizient **Cramérs V** verwendet. Cramérs V nimmt als normiertes Zusammenhangsmaß stets Werte zwischen 0 und 1 an ($0 \leq V \leq 1$) und wird nach Kuckartz et al. (2013) auf Grundlage des Chi²-Werts χ^2, des Stichprobenumfangs n und der Dimension M der p x q-Kreuztabelle folgendermaßen berechnet:

$$V := \sqrt{\frac{\chi^2}{n(M-1)}} \quad \text{mit } M := \min(p; q).$$

Für die Interpretation der Werte von Cramérs V wird im Rahmen dieser Arbeit in Anlehnung an Kuckartz et al. (2013) folgende Einteilung angewendet: $0 \leq V < 0{,}1$: sehr geringe Effektstärke; $0{,}1 \leq V < 0{,}3$: geringe Effektstärke; $0{,}3 \leq V < 0{,}5$: mittlere Effektstärke; $0{,}5 \leq V < 0{,}7$: hohe Effektstärke; $0{,}7 \leq V < 1$: sehr hohe Effektstärke.

Test	Anwendungsbedingung(en)	Berechnung von Cramérs V
Chi²-Test nach Karl Pearson	- p x q – Kreuztabelle mit $p, q \in \mathbb{N}_{\geq 2}$ und $p \geq 3 \vee q \geq 3$ - alle erwarteten absoluten Häufigkeiten größer oder gleich fünf	Berechnung immer Interpretation der Werte von Cramérs V: $0 \leq V < 0{,}1$: sehr geringe Effektstärke $0{,}1 \leq V < 0{,}3$: geringe Effektstärke $0{,}3 \leq V < 0{,}5$: mittlerer Effektstärke $0{,}5 \leq V < 0{,}7$: hohe Effektstärke $0{,}7 \leq V < 1$: sehr hohe Effektstärke
$N - 1$ Chi²-Test nach Egon Pearson	- 2 x 2 – Kreuztabelle - alle erwarteten absoluten Häufigkeiten größer oder gleich fünf	
Exakter Fisher-Test	- eine der erwarteten absoluten Häufigkeiten kleiner als fünf	

Tabelle 5: Übersicht über die Verwendung von Signifikanztests zur Überprüfung von Unterschieds- und Zusammenhangshypothesen sowie über die Berechnung von Cramérs V im Rahmen dieser Arbeit im Falle nominal skalierter Merkmale und falls die ordinale Struktur von Merkmalen nicht in Tests eingehen soll

Zusammenfassend werden für die Analyse von Unterschieds- und Zusammenhangshypothesen im Rahmen der vorliegenden Arbeit, im Falle nominal skalierter Merkmale und falls die ordinale Struktur von Merkmalen nicht in den Test eingehen soll, entsprechend den Angaben in Tabelle 5 Signifikanztests ausgewählt. Die Testergebnisse werden durch die Berechnung von Cramérs V als Maß für die Effektstärke ergänzt.

Verfahren für den Vergleich zweier Gruppen bei ordinal skalierten Merkmalen

Wenn zwei Gruppen in Bezug auf ein ordinal skaliertes Merkmal verglichen werden sollen, wird im Rahmen der vorliegenden Arbeit der **Mann-Whitney-U-Test**[118] (im Folgenden kurz: U-Test) angewendet. Die Verwendung des U-Tests im Falle ordinal skalierter Merkmale hat im Vergleich zum Rückgriff auf den Chi²-Test den Vorteil, dass mehr der in den Daten enthaltenen Informationen berücksichtigt werden. Der U-Test eignet sich für paarweise Vergleiche, d. h. bei 2 x q - Kreuztabellen ($q \in \mathbb{N}_{\geq 2}$). Er kann beispielsweise genutzt werden, um Beweisversuche von Lernenden in den Klassenstufen 5 und 6 bezüglich eines ordinal skalierten Merkmals zu vergleichen.

> Der MANN-WHITNEY-**U-Test** prüft die Nullhypothese, ob zwei zu vergleichende Stichproben aus formgleich (homogen) verteilten Populationen mit identischem Medianwert stammen. Mit formgleich ist jedoch nicht notwendigerweise symmetrisch oder gar normalverteilt gemeint. Wird der U-Test signifikant, ist davon auszugehen, dass sich die Mediane der zugrunde liegenden Populationen unterscheiden (H_1). Dies erfolgt mit Hilfe der Rangplätze, die die Merkmalsträger der beiden Gruppen in Bezug zur gesamten Stichprobe einnehmen. Die Information über die Rangplätze verarbeitet er in eine Prüfgröße, den U-Wert. (Benesch, 2013, S. 211, H. i. O.)

Beim U-Test werden die Merkmalsträger der zwei Stichproben mit den Umfängen n_1 und n_2 in Abhängigkeit der Merkmalsausprägungen in eine gemeinsame Rangreihe gebracht, d. h. jedem Merkmalsträger wird genau ein Rangplatz zwischen 1 und $n = n_1 + n_2$ zugeordnet (Benesch, 2013; Bortz & Lienert, 2008). Wenn mehrere Merkmalsträger die gleiche Merkmalsausprägung aufweisen, d. h. z. B. mehrere Beweisversuche bezüglich eines Merkmals der gleichen Kategorie zugewiesen wurden, wird diesen Merkmalsträgern jeweils der Durchschnitt der für diese Merkmalsträger normalerweise zu vergebenden Rangplätze zugewiesen (Bortz & Schuster, 2010). Anschließend wird für jede der Stichproben die Summe der Rangplätze T_1 bzw. T_2 berechnet (Benesch, 2013; Bortz & Lienert, 2008). Im nächsten Schritt werden zusätzlich die Rangüberschreitungen miteinbezogen. Für jeden Merkmalsträger der ersten Stichprobe wird die Anzahl an Merkmalsträgern aus der zweiten Stichprobe gezählt, die den Rangplatz des fokussierten Merkmalsträgers aus der ersten Stichprobe überschreiten. Die Summe der Rangüberschreitungen ergibt den Prüfwert U_1 (Benesch, 2013). Alternativ lässt sich diese Summe der Rangüberschreitungen U_1 nach Benesch (2013) auch direkt auf Grundlage der Stichprobenumfänge n_1, n_2 und der Summe der Rangplätze T_1 über folgende Formel bestimmen:

$$U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - T_1 \, .$$

[118] Bei den im Rahmen dieser Arbeit vorliegenden ordinal skalierten Daten wird der Mann-Whitney-U-Test gegenüber dem Mediantest, der die Nullhypothese prüft, „ob zwei zu vergleichende Stichproben aus Populationen den identischen Median aufweisen“ (Bortz & Lienert, 2008, S. 137), bevorzugt, weil der Mann-Whitney-U-Test mehr Informationen aus den ordinal skalierten Daten berücksichtigt. Eine Erläuterung beider Tests inklusive ihrer Vor- und Nachteile ist bei Bortz und Lienert (2008) zu finden.

Analog oder über die Beziehung $U_1 + U_2 = n_1 \cdot n_2$ wird zudem die Summe der Rangüberschreitungen U_2 für die zweite Stichprobe bestimmt (Benesch, 2013; Bortz & Lienert, 2008). In der vorliegenden Arbeit sind die Stichprobenumfänge n_1, n_2 stets so groß, dass bei der darauf aufbauenden Signifikanzüberprüfung von der Normalverteilungsapproximation Gebrauch gemacht wird. Als Output erhält man insbesondere einen z-Wert, der den Unterschied zwischen den beiden Stichproben beschreibt, und den daraus resultierenden p-Wert. Je nach erhaltenem p-Wert wird unter Berücksichtigung des festgelegten Signifikanzniveaus $\alpha = 0{,}05$ die Nullhypothese (nicht) verworfen (Benesch, 2013; Bortz & Lienert, 2008; Bortz & Schuster, 2010).[119] Auf Grundlage des z-Werts und des Stichprobenumfangs $n = n_1 + n_2$ wird im Rahmen dieser Arbeit in Ergänzung zum Mann-Whitney-U-Test die Effektstärke r

$$r := \frac{|z|}{\sqrt{n}}$$

berechnet. Diese nimmt genauso wie Cramérs V stets Werte zwischen 0 und 1 an ($0 \leq r \leq 1$) und kann analog interpretiert werden (Rosenthal, 1991).

7.3 Anwendung der Auswertungsmethoden

Auf Grundlage der beschriebenen Auswertungsmethoden wird im Folgenden das genaue Vorgehen bei der Analyse der Beweisversuche der Teilnehmerinnen und Teilnehmer der Mathematik-Olympiade schrittweise erläutert. Dabei stehen die methodischen Überlegungen im Fokus. Das Kategoriensystem an sich, das als Kern der qualitativ-strukturierenden Inhaltsanalyse entwickelt wurde, ist zusammen mit dessen theoretischen Grundlagen in Abschnitt 3.3 zu finden.

7.3.1 Überblick über die gewählte Vorgehensweise

Um die Beweisversuche auszuwerten, wird primär das Verfahren der qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) verwendet. Die Abbildung 78 gibt einen Überblick über die Schritte der Inhaltsanalyse im Rahmen dieser Arbeit. Die bei der Konzeption der Studie festgelegten Forschungsfragen bilden den Ausgangspunkt der Analyse von Beweisversuchen. Für diese Analyse liegen 3400 Beweisversuche vor, so dass es erforderlich ist, auszuwählen und somit die Menge des zu untersuchenden Datenmaterials zu verkleinern. Es wird ein Kategoriensystem erstellt, das einen differenzierten und vergleichenden Blick auf die Beweisversuche ermöglichen soll. Dieses Kategoriensystem wird pilotiert, weiterentwickelt und schließlich genutzt, um alle ausgewählten Beweisversuche zu codieren. Die erhaltenen Codierungen werden weiter ausgewertet und mit Blick auf die Forschungsfragen präsentiert.

[119] Für detaillierte Erläuterungen zum Mann-Whitney-U-Test wird auf die sich ergänzenden Ausführungen von Benesch (2013), Bortz und Lienert (2008) sowie Bortz und Schuster (2010) verwiesen.

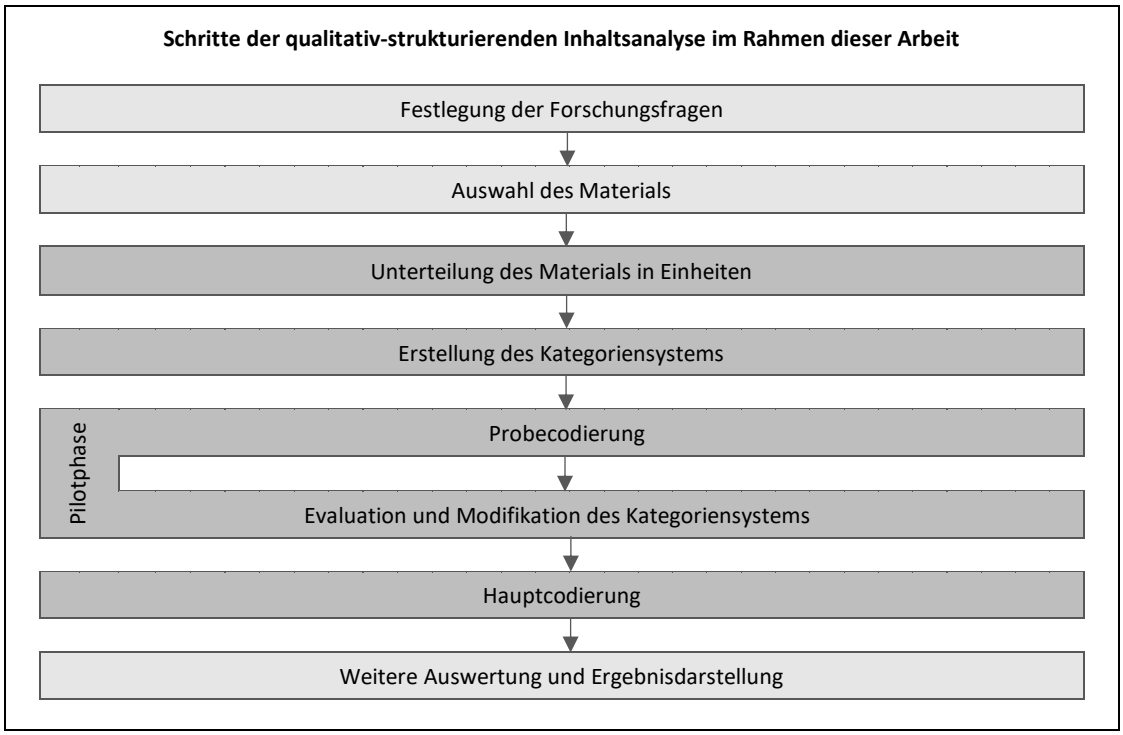


Abbildung 78: Schritte der qualitativ-strukturierenden Inhaltsanalyse im Rahmen dieser Arbeit

7.3.2 Konkretisierung der Forschungsfragen

Die Auswertung der Beweisversuche von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade zielt auf die Beantwortung der Forschungsfragen dieser Arbeit. Es wird untersucht, inwieweit es Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 gelingt, eine mathematische Allaussage mit unendlichem Gültigkeitsbereich aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen (1), welche Charakteristika deren Beweisversuche auszeichnen (2) und inwieweit Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind (3). In Bezug auf alle drei Punkte wird aufgezeigt, welche Gemeinsamkeiten und Unterschiede diesbezüglich mit Blick auf die verschiedenen Klassenstufen bestehen (Abschn. 1.1, 6.1).

7.3.3 Auswahl des Materials

Für die Beantwortung der genannten Forschungsfragen stehen 3400 Beweisversuche zum Teil d) der Beweisaufgabe, die von Lernenden in der Wettbewerbsklausur der Regionalrunde der Mathematik-Olympiade 2016/17 bearbeitet wurde, zur Verfügung (Abschn. 6.2, 7.1). Das sind so viele Beweisversuche, dass es zeitlich nicht möglich ist, alle Daten zu analysieren. Um trotzdem angemessene und differenzierte Aussagen über die Beweisversuche der Teilnehmenden treffen zu können, werden zwei Stichproben gezogen, eine bewusste Stichprobe und eine Zufallsstichprobe (Abschn. 7.2.2.1 zu Arten der Stichprobenziehung).

Die bewusste Stichprobenziehung zielt darauf, einen Einblick in das Spektrum der Beweisversuche zu erhalten, die von Lernenden in den einzelnen Klassenstufen 5 bis 8 formuliert wurden. Um möglichst vielfältige Beweisversuche von Lernenden aus den einzelnen Klassenstufen

zu berücksichtigen, wird die Ziehung pro Klassenstufe separat durchgeführt. Die einzelnen Ziehungen erfolgen anhand eines qualitativen Stichprobenplans, in den der Erfolg der Teilnehmenden bei der Mathematik-Olympiade, deren Teilnahme an mathematischen Förderangeboten und deren Bearbeitungen der Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche sowie das Geschlecht der Lernenden einbezogen wurden, weil es Hinweise darauf gibt, dass diese Faktoren eventuell mit manchen Eigenschaften von Beweisversuchen von Schülerinnen und Schülern korrelieren (u. a. Healy & Hoyles, 2000; Heinze et al., 2007; Tebaartz 2013; Ufer et al., 2009). Nebenbei wurden wegen der leicht differenten Klausurbedingungen und der Unterschiedlichkeit der mathematischen Förderangebote, die in den verschiedenen Regionen etabliert sind, die Regionen Nordrhein-Westfalens, in denen die Studie erfolgte, in den qualitativen Stichprobenplan mit aufgenommen.[120] Anhand des qualitativen Stichprobenplans werden pro Klassenstufe je 50 Schülerinnen und Schüler ausgewählt, die sich in der Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche mit dem Beweisen einer Allaussage bzw. dem Widerlegen einer Allaussage auseinandergesetzt haben.[121] Daraus ergibt sich eine Stichprobengröße von 100 Probanden je Klassenstufe fünf bis acht bzw. von 400 Probanden insgesamt. Alternativ zu der bewussten Stichprobenziehung anhand eines qualitativen Stichprobenplans wäre es möglich, eine theoretische Stichprobenziehung durchzuführen (Abschn. 7.2.2.1). Auch dieses Bottom-Up-Verfahren wäre geeignet, um ein großes Spektrum an Beweisversuchen abzudecken, würde jedoch eine zeitintensive Einsicht in sehr viele der 3400 Beweisversuche erfordern, zumal ergänzende Informationen vorliegen, die das beschriebene Top-down-Verfahren ermöglichen.

Zusätzlich zu der bewussten Stichprobenziehung werden Schülerinnen und Schüler mittels einer Zufallsstichprobe ausgewählt. Ziel der Zufallsstichprobe ist es, Einblicke in die Häufigkeitsverteilungen der Eigenschaften der Beweisversuche von Lernenden in den einzelnen Klassenstufen 5 bis 8 zu erhalten. Dafür werden je 25% der Teilnehmenden aus den Klassenstufen 5 bis 8 ausgewählt (geschichtete Zufallsstichprobe).[122] Die Tabelle 6 zeigt die Anzahlen an zufällig auszuwählenden Schülerinnen und Schülern. In der fünften Klassenstufe sind es 328, in der sechsten 230, in der siebten 192 und in der achten 103 der Lernenden, von denen Datensätze vorliegen. Da die Anzahl an auszuwählenden Personen jeweils auf eine natürliche Zahl aufgerundet wird, handelt es sich in den einzelnen Klassenstufen und insgesamt mit 853 von 3400 um leicht mehr als 25% der Schülerinnen und Schüler.[123]

[120] Die Informationen, die im qualitativen Stichprobenplan berücksichtigt werden, wurden mit dem Fragebogen erhoben, den die Teilnehmenden im Anschluss an die Wettbewerbsklausur erhalten haben (Abschn. 6.2.1, 10.2).

[121] Obwohl die Auswertung im Rahmen der vorliegenden Arbeit auf den Aufgabenteil d) der Beweisaufgabe fokussiert, werden die Stichprobenziehung (sowie das Kategoriensystem) so angelegt, dass darauf aufbauend, nach Abschluss der vorliegenden Arbeit, eine weitere Auswertung der im Rahmen des übergeordneten Forschungsprojekts zum Beweisen bei der Mathematik-Olympiade erhobenen Daten gut möglich ist (Abschn. 7.3.5 zur Entwicklung des Kategoriensystems; Abschn. 10.2 zum übergeordneten Forschungsprojekt).

[122] Die Größe der Zufallsstichprobe wurde unter Abwägung deren Repräsentativität und der praktischen Umsetzbarkeit der Analyse aller ausgewählter Daten festgelegt. Dass pro Klassenstufe ein gleich großer Anteil der Teilnehmenden ausgewählt wird, unterstützt die Verallgemeinerbarkeit der Ergebnisse der Analyse der ausgewählten Beweisversuche auf die insgesamt vorhandenen 3400 Beweisversuche.

[123] Dies ist auch damit zu erklären, dass je Klassenstufe 25% der Datensätze von Lernenden, die im Fragebogen die Aufgabe zum Validieren, Bewerten fiktiver Beweisversuche zu einer wahren Allaussage erhalten haben, und 25% der Datensätze der Lernenden, die im Fragebogen die Aufgabe zum Validieren, Bewerten fiktiver Beweisversuche

Klassenstufe	Anzahl an vorhandenen Datensätzen	Anzahl an auszuwählenden Datensätzen (Zufallsstichprobe, 25%, aufgerundet)
Kl. 5	1309	328
Kl. 6	919	230
Kl. 7	765	192
Kl. 8	407	103
Kl. 5-8	3400	853

Tabelle 6: Zufallsstichprobe: Anzahlen an vorhandenen und auszuwählenden Datensätzen

Die Datensätze aus der bewusst gezogenen Stichprobe und die aus der Zufallsstichprobe zusammen genommen handelt es sich nicht um 1253 $(= 400 + 853)$, sondern um 1155 ausgewählte Schülerinnen und Schüler. Die Gesamtanzahl ist geringer als 1253, weil in der Zufallsstichprobe Datensätze vorkommen, die auch Teil der bewusst gezogenen Stichprobe sind.

7.3.4 Unterteilung des Materials in Einheiten

In der qualitativ-strukturierenden Inhaltsanalyse sind somit 1155 Beweisversuche zum Aufgabenteil d) der für die Studie konzipierten Beweisaufgabe auszuwerten. Dabei wird jeder der Beweisversuche als eine Analyseeinheit und als eine Codiereinheit angesehen. Eine weitere Unterteilung der Beweisversuche in mehrere Codiereinheiten wird nicht vorgenommen, weil sich beispielsweise die Frage, inwieweit es sich bei einem Beweisversuch um einen Beweis der infrage stehenden Aussage handelt, i. A. nur auf Grundlage der gesamten Bearbeitung beurteilen lässt. Ausschließlich die Nummerierung des Aufgabenteils (z. B. „d)“, „Aufgabe 2d“, „Defne:“) und eventuelle Bemerkungen der Art „Auf der nächsten Seite geht’s weiter“ bleiben unberücksichtigt. Wenn es bei der Analyse einzelner Eigenschaften der Beweisversuche erforderlich ist, bestimmte Teile der Beweisversuche zu fokussieren oder separat zu codieren, wird dies an den entsprechenden Stellen des Codierleitfadens festgehalten. Da die Analyse der Beweisversuche zum Aufgabenteil d) unabhängig von weiteren vorliegenden Informationen zu den einzelnen Probanden, d. h. insbesondere unabhängig von den Beweisversuchen zu den anderen Aufgabenteilen a) bis c) erfolgen soll, stimmen die Kontexteinheiten weitgehend mit den Codiereinheiten überein. Neben den Beweisversuchen zum Aufgabenteil d) an sich wird ausschließlich die Klassenstufe bei der Codierung mitberücksichtigt, um die Analyse auf die richtige Version des Aufgabenteils d) zu beziehen (Abschn. 6.2.2 zu den Aufgabenteilen a) bis c) und zu den Versionen des Aufgabenteils d)).

7.3.5 Erstellung des Kategoriensystems

Für die Analyse der Beweisversuche zum Aufgabenteil d) wird deduktiv-induktiv ein Kategoriensystem erstellt. Eine deduktiv-induktive Vorgehensweise bietet sich an, weil sich aus den Forschungsfragen und der mathematikdidaktischen Literatur zahlreiche Ansatzpunkte für die

zu einer falschen Allaussage erhalten haben, gezogen werden. Mit Blick auf eine eventuelle Fortsetzung der Auswertung nach Abschluss der vorliegenden Arbeit, wird bei der Auswahl von Probanden nach der Version der erhaltenen Aufgabe zum Validieren und Bewerten fiktiver differenziert. Es handelt sich also um eine geschichtete Zufallsstichprobe, die je Klassenstufe zwei einfache Zufallsstichproben bzw. insgesamt acht einfache Zufallsstichproben umfasst.

Erstellung des Kategoriensystems ergeben und zugleich der Einbezug eines Teils der erhobenen Beweisversuche für die Passung des Kategoriensystems zu den zu analysierenden Beweisversuchen wichtig ist. Bei der Erstellung des Kategoriensystems werden dementsprechend die Forschungsfragen, vorhandene Theorien und Forschungsergebnisse, die in dieser Arbeit fokussierte Aufgabenstellung und je Klassenstufe 50 (insgesamt $4 \cdot 50 = 200$) Beweisversuche aus der bewussten Stichprobenziehung berücksichtigt.

Die Forschungsfragen zielen auf die Deskription und den Vergleich der Beweisversuche von Teilnehmenden der Mathematik-Olympiade u. a. mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage (Abschn. 1.1, 6.1). Um zu beschreiben, inwieweit es den Teilnehmenden gelungen ist, die vorgegebene Aussage zu beweisen, sollten im Kategoriensystem auf jeden Fall die Korrektheit und Vollständigkeit der Argumentgefüge sowie die verwendete Schlussweise im Kategoriensystem berücksichtigt werden. Vorab stellt sich gerade mit Blick auf Beweisversuche junger Schülerinnen und Schüler die Frage, inwieweit diese überhaupt ein Argumentgefüge mit einem Bezug zu der zu zeigenden Aussage formuliert haben und eine Beweisidee vorkommt, deren Verallgemeinerbarkeit explizit angedeutet ist. Auch dies ist in das Kategoriensystem mit aufzunehmen. Weiterhin erscheint es sinnvoll, die Beweisideen, Eigenschaften angegebener Beispiele, die vorkommenden Darstellungsmittel und die Darstellungsform der Beweisversuche zur Beantwortung der Forschungsfrage nach Charakteristika der Beweisversuche und Anknüpfungspunkten für deren Weiterentwicklung in das Kategoriensystem mit einzubeziehen, wie in Abschnitt 3.3 ausführlich erläutert ist.

Um die Beweisversuche der Teilnehmenden in den einzelnen Klassenstufen 5 bis 8 vergleichen zu können, wird trotz der leicht unterschiedlichen Versionen des Aufgabenteils d), die Teilnehmende in den Klassenstufen 5, 6 bzw. 7, 8 erhalten haben, ein einheitliches Kategoriensystem für die Analyse aller Beweisversuche zum Aufgabenteil d) angestrebt (Abschn. 6.2.2 zu den Versionen des Aufgabenteils d))

Wenn ein erstes Kategoriensystem mit dazugehörigem Codierleitfaden zur Analyse von Beweisversuchen zum fokussierten Aufgabenteil d) erstellt ist, wird davon ausgehend ein Kategoriensystem für die Analyse von Beweisversuchen zum Aufgabenteil c), in dem eine Allaussage mit endlichem Gültigkeitsbereich zu zeigen ist, formuliert.[124] Zudem werden zwei möglichst analoge Kategoriensysteme für die Analyse der Beweisversuche zu den Aufgabenteilen a) und b), in denen jeweils eine Existenzaussage zu beweisen ist, konstruiert. Dabei wird versucht, eine gute Passung zwischen den Kategoriensystemen für die Analyse der Beweisversuche zu Allaussagen und denen für die Analyse der Beweisversuche zu Existenzaussagen zu erreichen. Es wird insbesondere auch das Kategoriensystem zum Aufgabenteil d) so lange weiterentwickelt bis die Optimierung, Angleichung der Kategoriensysteme abgeschlossen erscheint.

[124] Um die Analyse der Beweisversuche zum hier fokussierten Aufgabenteil d) mit Blick auf die Gesamtauswertung des dieser Arbeit übergeordneten Forschungsprojektes möglichst generisch anzulegen und langfristig inter- und intraindividuelle Vergleiche von Beweisversuchen zu den verschiedenen Aufgabenteilen a) bis d) zu ermöglichen, werden auch Kategoriensysteme zu den anderen Aufgabenteilen entwickelt.

7.3.6 Pilotierung des Kategoriensystems

Im Anschluss an die Erstellung des Kategoriensystems erfolgt zunächst eine Pilotphase, um dessen Qualität zu überprüfen und zu optimieren (Abschn. 7.2.1.4). In der Pilotphase wird das Kategoriensystem zum Aufgabenteil d) zunächst auf einen Teil des Materials angewendet. Für diese **Probecodierung** werden je Klassenstufe die 50 bzw. insgesamt die 200 Beweisversuche aus der bewussten Stichprobe ausgewählt, die nicht in die Erstellung des Kategoriensystems einbezogen worden sind, so dass tendenziell vielfältige Beweisversuche Berücksichtigung finden. Die 200 Beweisversuche entsprechen 17% aller zu analysierenden Beweisversuche, was im von Schreier (2012) vorgeschlagenen Rahmen von zehn bis zwanzig Prozent liegt. Die Probecodierung wird von einer studentischen Mitarbeiterin und der Autorin dieser Arbeit durchgeführt. Nach einer kurzen Einarbeitung der studentischen Mitarbeiterin werden alle 200 Beweisversuche von beiden Personen unabhängig voneinander und ohne vorhergehende Einsicht in diese Daten codiert.

Die auf Grundlage der Codierungen berechneten Werte für die Intercoderreliabilität (Cohens Kappa) liegen alle über $\kappa = 0{,}45$ und mehrheitlich über $\kappa = 0{,}75$. Werden für die Interpretation dieser Werte die von Wirtz und Caspar (2002) angelegten Wertegrenzen herangezogen, liegt die Intercoderreliabilität bei allen analysierten Merkmalen im akzeptablen, mehrheitlich im sehr guten Bereich (Abschn. 7.2.2.2 zur Intercoderreliabilität).[125] In Ergänzung zur Berechnung der Intercoderreliabilität werden die unterschiedlichen Codierungen und die während der Probecodierung notierten Schwierigkeiten besprochen und die Validität des Kategoriensystems diskutiert. Im Zuge dieser **Evaluation des Kategoriensystems** entstehen mehrere Ideen zur **Optimierung des Kategoriensystems**. Während der bisherige Entwurf des Kategoriensystems Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz ausschließlich als solche erfasst, wird nach der Probecodierung eine detaillierte Analyse dieser Beweisversuche eingeführt (Abschn. 3.3.10 zur Analyse von Beweisversuchen mit nicht eindeutig kategorisierbarem Ansatz). Zudem werden nach der Probecodierung einzelne Details von Kategoriendefinitionen modifiziert sowie Codierregeln ergänzt und optimiert. Auch werden im Codierleitfaden manche Ankeritems durch noch passendere ersetzt und wenige Ankeritems zur Veranschaulichung des Umgangs mit Grenzfällen ergänzt. In diesem Evaluations- und Modifikationsprozess des Kategoriensystems zum in dieser Arbeit im Fokus stehenden Aufgabenteil d) werden die Kategoriensysteme zu den anderen Aufgabenteile a) bis c) mit berücksichtigt und mitgedacht, wenngleich sich aus Zeitgründen ausschließlich das Kategoriensystem zum Aufgabenteil d) finalisieren lässt (Abschn. 3.3 zum finalen Kategoriensystem).

[125] Da im Rahmen der Hauptcodierung eine erneute Überprüfung erfolgt und im Anschluss an diese Probecodierungen noch Modifikationen am Kategoriensystem vorgenommen werden, wird auf einen detaillierten Bericht der einzelnen Werte verzichtet.

7.3.7 Hauptcodierung

Nach der Pilotierung werden alle 1155 zu untersuchenden Beweisversuche zum Aufgabenteil d) sukzessive anhand des finalen Kategoriensystems codiert.[126] Um die Qualität dieses finalen Kategoriensystems zu überprüfen und zu dokumentieren, wird die Codierung eines Teils dieser Beweisversuche zusätzlich durch einen studentischen Mitarbeiter vorgenommen. Die Überprüfung erfolgt an 15% ($\mathrm{N} = 132$) der Beweisversuche aus der Zufallsstichprobe, so dass Beweisversuche von Lernenden aller Klassenstufen 5 bis 8 vorkommen und sie weder bei der Erstellung des Kategoriensystems noch bei dessen Pilotierung gesichtet worden waren.[127]

Alle Beweisversuche			
Merkmal: Ansatz	$\kappa = 0{,}957$		
Beweisversuche mit begründendem Ansatz: Beweisideen			
Vorkommende Beweisideen (BI)			
Beweisidee	κ	Beweisidee	κ
BI1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*	1,000	BI2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*	1,000
BI3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*	1,000	BI4 *Zahldarstellung im Dezimalsystem*	1,000
BI5 *Reste bei Division einzelner Summanden durch vier oder acht*	0,739	BI6 *Teilbarkeit der Summen zweier Zahlenpaare*	0,969
BI7 *Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen*	1,000	BI8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*	1,000
BI9 *Parität von Summen ungerader oder gerader Zahlen*	0,977	BI10 *Abstände ungerader Zahlen*	0,861
BI11 *vier Zahlen*	0,862	BI12 *Vierer- oder Achterreihe*	0,788
BI13 *Regeln zur Teilbarkeit von Produkten oder Summen durch vier oder acht*	0,923	BI55 *nicht eindeutig kategorisierbar*	0,798
Anzahl und Verknüpfung der Beweisideen		**Qualität der verwendeten Beweisideen**	
Merkmal	κ	Merkmal	κ
Eine vs. mehrere Beweisideen	0,830	Zielführendste Beweisidee	0,926
Verknüpfung der Beweisideen	0,868	Primäre Beweisidee	0,930

Tabelle 7: Intercoderreliabilität (Cohens Kappa κ). Fokus: Ansatz, Beweisideen in Beweisversuchen mit begründendem Ansatz

[126] Um trotz der großen Komplexität des Kategoriensystems und der Vielzahl an zu analysierenden Beweisversuchen eine möglichst einheitliche Codierung zu gewährleisten, wird schrittweise vorgegangen. Es werden zunächst alle Beweisversuche hinsichtlich des gewählten Ansatzes codiert. Danach werden nacheinander Beweisversuche mit begründendem, mit prüfendem, mit weder begründendem noch prüfendem und mit nicht eindeutig kategorisierbarem Ansatz fokussiert. Abschließend werden bei allen diese Beweisversuche die vorkommenden Darstellungsmittel und die Darstellungsform analysiert.

[127] Aus jeder Klassenstufe 5 bis 8 werden die ersten 15% der Beweisversuche aus den beiden einfachen Zufallsstichproben verwendet, die nicht Teil der bewusst gezogenen Stichprobe sind und somit nicht bei der Erstellung oder Pilotierung des Kategoriensystems eingesehen wurden. Insgesamt handelt es sich um 132 Beweisversuche, die doppelt codiert werden. Die Codierung erfolgt in mehreren Schritten (Fußnote 126). Nach der Analyse des Ansatzes wird sofort die Intercoderreliabilität berechnet, anschließend werden die drei unterschiedlichen Codierungen besprochen, sodass die darauf aufbauende detaillierte Analyse der Beweisversuche mit begründendem, prüfendem, weder begründendem noch prüfendem oder nicht eindeutig kategorisierbarem Ansatz auf der Grundlage einer einheitlichen Datenbasis erfolgt. Da ausschließlich wenige der 132 Beweisversuche solche mit prüfendem Ansatz, mit weder begründendem noch prüfendem oder mit nicht eindeutig kategorisierbarem Ansatz sind, wird die Stichprobe an diesen Stellen für eine angemessene Überprüfung des finalen Kategoriensystems erweitert. Es werden je Klassenstufe 30% der Beweisversuche mit prüfendem Ansatz (insg. $\mathrm{N} = 51$) und alle Beweisversuche mit weder begründendem noch prüfendem oder nicht eindeutig kategorisierbarem Ansatz aus der Zufallsstichprobe, die nicht zugleich Teil der bewusst gezogenen Stichprobe sind, (insg. $\mathrm{N} = 24$ bzw. $\mathrm{N} = 13$) doppelt codiert.

Beweisversuche mit begründendem Ansatz: Korrektheit und Vollständigkeit, Beispiele, Schlussweise			
Korrektheit und Vollständigkeit			
Merkmal: Passung zu der zu zeigenden Aussage	$\kappa = 0{,}873$		
Merkmal: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges	$\kappa = 0{,}861$		
Existenz und Passung eines Beispiels zu der zu zeigenden Aussage			
Merkmal: Existenz und Passung eines Beispiels zu der zu zeigenden Aussage	$\kappa = 0{,}906$		
Anzahl, Art und Ort angegebener Beispiele			
Merkmal	κ	Merkmal	κ
Anzahl an angegebenen Beispielen	0,948	Ort eines angegebenen Beispiels: Oberhalb	0,965
Anzahl an angegebenen Beispielen zu der zz. Aussage	0,920	Ort eines angegebenen Beispiels: Innerhalb	0,947
		Ort eines angegebenen Beispiels: Unterhalb	0,961
Anzahl an angegebenen sonstigen Beispielen	0,883	Ort eines angegebenen Beispiels: Nicht eindeutig kategorisierbar	1,000
Systematik angegebener Beispiele	0,936	Ort der angegebenen Beispiele	0,923
Angabe eines besonderen Beispiels	0,982	Angabe vollständig und weitgehend korrekt ausgeführter Beispiele zu der zz. Aussage	0,924
Angabe eines großen Beispiels	1,000		
Beispiele oberhalb des (übrigen) Argumentgefüges		**Beispiele unterhalb des (übrigen) Argumentgefüges**	
Merkmal	κ	Merkmal	κ
Anzahl an angegebenen Beispielen	0,945	Anzahl an angegebenen Beispielen	0,891
Anzahl an angegebenen Beispielen zu der zz. Aussage	0,948	Anzahl an angegebenen Beispielen zu der zz. Aussage	0,836
Anzahl an angegebenen sonstigen Beispielen	0,921	Anzahl an angegebenen sonstigen Beispielen	0,836
Hervorhebung einer Beweisidee	0,946	Hervorhebung einer Beweisidee	0,836
Systematik angegebener Beispiele	0,949	Systematik angegebener Beispiele	0,891
Angabe eines besonderen Beispiels	0,947	Angabe eines besonderen Beispiels	0,836
Angabe eines großen Beispiels	0,945	Angabe eines großen Beispiels	0,888
Angabe vollständig und weitgehend korrekt ausgeführter Beispiele zu der zz. Aussage	0,918	Angabe vollständig und weitgehend korrekt ausgeführter Beispiele zu der zz. Aussage	0,891
Beispiele innerhalb des (übrigen) Argumentgefüges			
Merkmal	κ	Merkmal	κ
Anzahl an angegebenen Beispielen	0,861	Integration der Beispiele in das Argumentgefüge: Notwendiges Beispiel	0,906
Anzahl an angegebenen Beispielen zu der zz. Aussage	0,865		
		Integration der Beispiele in das Argumentgefüge: Fest integriertes Beispiel	0,883
Anzahl an angegebenen sonstigen Beispielen	0,863		
Systematik angegebener Beispiele	0,864	Integration der Beispiele in das Argumentgefüge: Ergänzendes Beispiel	0,906
Angabe eines besonderen Beispiels	0,884		
Angabe eines großen Beispiels	0,922	Integration der Beispiele in das Argumentgefüge: Weglassbares Beispiel	0,906
Angabe vollständig und weitgehend korrekt ausgeführter Beispiele zu der zz. Aussage	0,863		
Schlussweise			
Merkmal: Schlussweise	$\kappa = 0{,}800$		

Tabelle 8: Intercoderreliabilität (Cohens Kappa κ). Fokus: Beweisversuche mit begründendem Ansatz: Korrektheit und Vollständigkeit, Beispiele sowie Schlussweise

Beweisversuche mit prüfendem Ansatz			
Passung eines Beispiels zu der zu zeigenden Aussage			
Merkmal: Passung eines Beispiels zu der zu zeigenden Aussage	$\kappa = 0{,}941$		
Eigenschaften der angegebenen Beispiele			
Merkmal	κ	Merkmal	κ
Anzahl an angegebenen Beispielen	1,000	Erkennbarkeit und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges	0,846
Anzahl an angegebenen Beispielen zu der zu zeigenden Aussage	1,000	Primäre Beweisidee	0,848
Anzahl an angegebenen sonstigen Beispielen	0,964	Systematik angegebener Beispiele	0,893
Passung aller Beispiele zu der zz. Aussage	0,870	Angabe eines besonderen Beispiels	0,928
Existenz, Art einer Angabe eines Wahrheitswerts	1,000	Angabe eines großen Beispiels	1,000
Kein Wahrheitswert angegeben: Verhältnis von Positiv- und Negativbeispielen	1,000	Angabe vollständig und weitgehend korrekt ausgeführter Beispiele zu der zz. Aussage	0,974
Hinweis auf die Allgemeingültigkeit der Aussage	0,897	Hinweis auf die Unzulänglichkeit von Beispielen	1,000
Schlussweise			
Merkmal: Schlussweise	$\kappa = 0{,}822$		

Tabelle 9: Intercoderreliabilität (Cohens Kappa κ). Fokus: Beweisversuche mit prüfendem Ansatz

Beweisversuche mit weder begründendem noch prüfendem Ansatz			
Merkmal	κ	Merkmal	κ
Existenz, Art einer Angabe eines Wahrheitswerts	1,000	Hinweis auf die Allgemeingültigkeit der Aussage	1,000
Art der Bearbeitung	0,902	Hinweis auf die Unzulänglichkeit von Beispielen	1,000

Tabelle 10: Intercoderreliabilität (Cohens Kappa κ). Fokus: Beweisversuche mit weder begründendem noch prüfendem Ansatz

Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz			
Merkmal	κ	Merkmal	κ
Passung der Ausführungen zu der zz. Aussage	0,857	Existenz, Passung eines Beispiels zu der zz. Aussage	0,810
Erkennbarkeit einer Beweisidee	0,762	Anzahl an Beispielen	1,000
Primäre Beweisidee	0,920	Existenz systematischer Beispiele	1,000
Existenz und Art einer Angabe eines Wahrheitswerts	0,857	Existenz eines besonderen Beispiels	1,000
		Existenz eines großen Beispiels	1,000

Tabelle 11: Intercoderreliabilität (Cohens Kappa κ). Fokus: Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz

Beweisversuche mit begründendem, prüfendem, weder begründendem noch prüfendem oder nicht eindeutig kategorisierbarem Ansatz			
Darstellungsmittel (Dstm)			
Merkmal	κ	Merkmal	κ
Symbolische Dstm: Verwendung	1,000	Ikonische Dstm: Verwendung	0,953
Symbolische Dstm: Variable	1,000	Ikonische Dstm: Figurierte Zahlen	0,909
Symbolische Dstm: Anzahl an Variablen	1,000	Ikonische Dstm: Klammern, Pfeile, Linien	0,912
Symbolische Dstm: Definitionen von Variablen	0,748	Ikonische Dstm: Einkreisungen, Unterstr., o. Ä.	0,913
Symbolische Dstm: Symbol mit fester Bedeutung	1,000	Ikonische Dstm: Häkchen, Kreuzchen	0,940
Strukturierende Dstm: Verwendung	0,976	Ikonische Dstm: Tabellen	0,939
Strukturierende Dstm: Mathematische Strukturierungselemente	0,938	Ikonische Dstm: Sonstiges	0,909
Strukturierende Dstm: Stukturierende Begriffe	0,944	Numerische Dstm: Verwendung	0,982
Strukturierende Dstm: Nummerierungen	0,970	Verbale Dstm: Verwendung	0,952
Darstellungsform			
Merkmal: Darstellungsform	$\kappa = 0{,}916$		

Tabelle 12: Intercoderreliabilität (Cohens Kappa κ). Fokus: Darstellungsmittel und Darstellungsform

Die Tabellen 7 bis 12 zeigen die erhaltenen Werte für die Intercoderreliabilität (Abschn. 7.2.2.2 zur Berechnung dieser Werte). Werden für die Interpretation der erhaltenen Werte die von Wirtz und Caspar (2002) angelegten Wertegrenzen für Cohens Kappa (κ) herangezogen, liegt die Intercoderreliabilität bei fast allen analysierten Merkmalen im sehr guten Bereich, vereinzelt im guten Bereich. Dabei deuten die Unterschiede zwischen den erhaltenen Kappa-Koeffizienten an, bei der Codierung welcher Merkmale mehr bzw. weniger Interpretationsspielraum besteht. Beispielsweise lässt sich gut erfassen, ob in einem Beweisversuch mit begründendem Ansatz die einzelnen Beweisideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* und 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* vorkommen. Die Analyse der Korrektheit und Vollständigkeit sowie der Schlussweise bei Beweisversuchen mit begründendem Ansatz ist im Vergleich dazu deutlich anspruchsvoller, wenngleich sich auch hier die Kappa-Koeffizienten im sehr guten Bereich befinden und nahe legen, dass es trotz des Kontinuums an möglichen Ausprägungen gelungen ist, eine gut anwendbare Unterteilung in Kategorien vorzunehmen (Abschn. 3.3.4, 3.3.6 zum Kategoriensystem; Tab. 7.6, 7.7 zu den Kappa-Koeffizienten).

Auffällig ist der mit $\kappa = 0{,}739$ vergleichsweise niedrige Wert bei der Erfassung, ob in einem Beweisversuch mit begründendem Ansatz die Beweisidee 5 *Reste bei Division einzelner Summanden durch vier oder acht* vorkommt (Tab. 7). Der Vergleich der vorgenommenen Codierungen zeigt, dass bei zwei der 87 Beweisversuche mit begründendem Ansatz, die in der zur Überprüfung des Kategoriensystems genutzten Stichprobe vorkommen, eine der beiden codierenden Personen die Beweisidee 5 erkannt hat, die andere Person nicht.[128] Bei den zwei unterschiedlich codierten Beweisversuchen handelt es sich um Grenzfälle, in denen die Beweisidee 5 *Reste bei Division einzelner Summanden durch vier oder acht* keineswegs offensichtlich, sondern beispielsweise an oberhalb des (übrigen) Argumentgefüges angeführten Beispielen zu erkennen ist. Dies macht darauf aufmerksam, dass u. a. die Erkennbarkeit der Beweisidee 5 nicht unbedingt ganz klar und indiskutabel ist.

Des Weiteren ist die Intercoderreliabilität bei den Variablendefinitionen mit $\kappa = 0{,}748$ vergleichsweise niedrig (Tab. 12). Der Vergleich und die Besprechung der unterschiedlich codierten Beweisversuche verdeutlichen, dass es schwierig sein kann zu entscheiden, ob alle Variablen angemessen definiert sind oder dies nicht ganz erfüllt ist. Dies wird bei der Darstellung und Interpretation der Ergebnisse berücksichtigt (insb. Abschn. 8.2.4).

Bei sämtlichen analysierten Merkmalen wird mit den Beweisversuchen, die unterschiedlich codiert wurden, folgendermaßen umgegangen: In einfachen Fällen entscheidet die Autorin dieser Arbeit über die endgültige Codierung, ansonsten wird im gemeinsamen Gespräch der Autorin mit dem studentischen Mitarbeiter die endgültige Codierung, die dann in der weiteren Auswertung und Ergebnispräsentation genutzt wird, festgelegt.

[128] Die prozentuale Übereinstimmung beträgt $\mathrm{PÜ} = 97{,}70\%$. Der vergleichsweise niedrige Kappa-Koeffizient kommt insbesondere dadurch zustande, dass durch die sehr schiefe Randverteilung (in über 90% der Beweisversuche mit begründendem Ansatz kommt die Beweisidee 5 nicht vor) der über die Randsummen der relativen Häufigkeiten der Entscheidungen der zwei Personen definierte relative Anteil der Übereinstimmungen bei zufälligem Codierverhalten ($\mathrm{P_e}$) bei $\mathrm{P_e} = 0{,}9120$ liegt und folglich der relative Anteil der Beweisversuche, bei denen die zwei Personen identische Codierungen vorgenommen haben, ($\mathrm{P_0}$) nicht besonders weit über den relativen Anteil der Übereinstimmung bei zufälligem Codierverhalten ($\mathrm{P_e}$) hinausgeht.

7.3.8 Weitere Auswertung und Ergebnispräsentation

Um ausgehend von den in der Hauptcodierung vorgenommenen Zuordnungen die Forschungsfragen systematisch und differenziert zu beantworten, werden die Forschungsfragen durch mehrere Teilfragen konkretisiert (Abb. 79). Um die Forschungsfragen zu beantworten, wird in der weiteren Auswertung und Ergebnispräsentation zuerst zwischen Beweisversuchen mit begründendem, mit prüfendem, mit weder begründendem noch prüfendem und nicht eindeutig kategorisierbaren Ansatz sowie fehlenden Bearbeitungen differenziert. Danach werden nacheinander die Beweisversuche mit begründendem Ansatz, die mit prüfendem Ansatz und die mit weder begründendem noch prüfendem Ansatz fokussiert und die Ergebnisse der detaillierten Analysen dieser Beweisversuche herauskristallisiert. In der jeweils ersten Teilfrage zu Beweisversuchen mit begründendem, prüfendem bzw. weder begründendem noch prüfendem Ansatz liegt ein besonderes Augenmerk auf dem Gelingen der Beweisversuche, während die nachfolgenden Teilfragen zwei bis vier mehr auf Charakteristika der Beweisversuche und Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis fokussieren. Die fünfte Teilfrage zielt jeweils auf Klassifikationen der Beweisversuche, die es ermöglichen sollen, die einzelnen Ergebnisse zu verdichten, zu strukturieren und zu resümieren. Abschließend werden die Ergebnisse der Teilfragen zu Beweisversuchen mit begründendem, prüfendem bzw. weder begründendem noch prüfendem Ansatz zusammengeführt. Es werden alle analysierten Beweisversuche gemeinsam mit Blick auf (1) den Nachweis der Gültigkeit der zu zeigenden Aussage und (2) Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis klassifiziert, um darzulegen, welche wesentlichen Ergebnisse mit Blick auf die beiden Forschungsfragen zu erkennen sind.

Für die Beantwortung der einzelnen Fragen werden jeweils das Spektrum an vorhandenen Codierungen und deren relative Häufigkeiten analysiert. Um eine angemessene Interpretation erhaltener Daten zu unterstützen, werden scheinbare Unterschiede und Zusammenhänge statistisch überprüft. Dabei wird der Mann-Whitney-U-Test, ein bei ordinal skalierten Merkmalen geeignetes Verfahren, ausschließlich vereinzelt genutzt, da sich bei vielen der infrage kommenden Merkmale, wie der zielführendsten Beweisidee und der Schlussweise bei Beweisversuchen mit begründendem Ansatz, kritisch diskutieren lässt, inwieweit diese ordinal skaliert sind. Deshalb wird die Verwendung von Verfahren, die sich auch für nominal skalierte Daten eigenen, präferiert. Die Ergebnisse aller Analysen werden sukzessive u. a. mittels Diagrammen dargestellt und vorsichtig interpretiert (Abschn. 8.1 - 8.5). An einzelnen Stellen werden zudem mehrere Beweisversuche als exemplarische Fälle ganzheitlich betrachtet, etwa um konkrete Einblicke in ein Spektrum an Ausprägungen zu geben (insb. Abschn. 8.2.3.3). Außerdem werden mehrmals Beweisversuche gruppiert oder klassifiziert, um verschiedene Ergebnisse zu kombinieren, zu strukturieren und zu resümieren (z. B. Abschn. 8.2.1, 8.2.6). Unter einer Klassifikation wird dabei anknüpfend an Wirtz (2021), Strangl (2023) und Straehler-Pohl und Gellert (2012) Folgendes verstanden: Eine Klassifikation ist eine Zuordnung von Objekten zu einer systematisch geordneten, übersichtlichen Zusammenstellung exhaustiver und disjunkten Kategorien, bei der neben den Eigenschaften der einzelnen Kategorien auch die Beziehungen zwischen verschiedenen Kategorien im Fokus stehen. Es werden sowohl der Prozess der Zuordnung als auch dessen Ergebnis als Klassifikation angesehen.

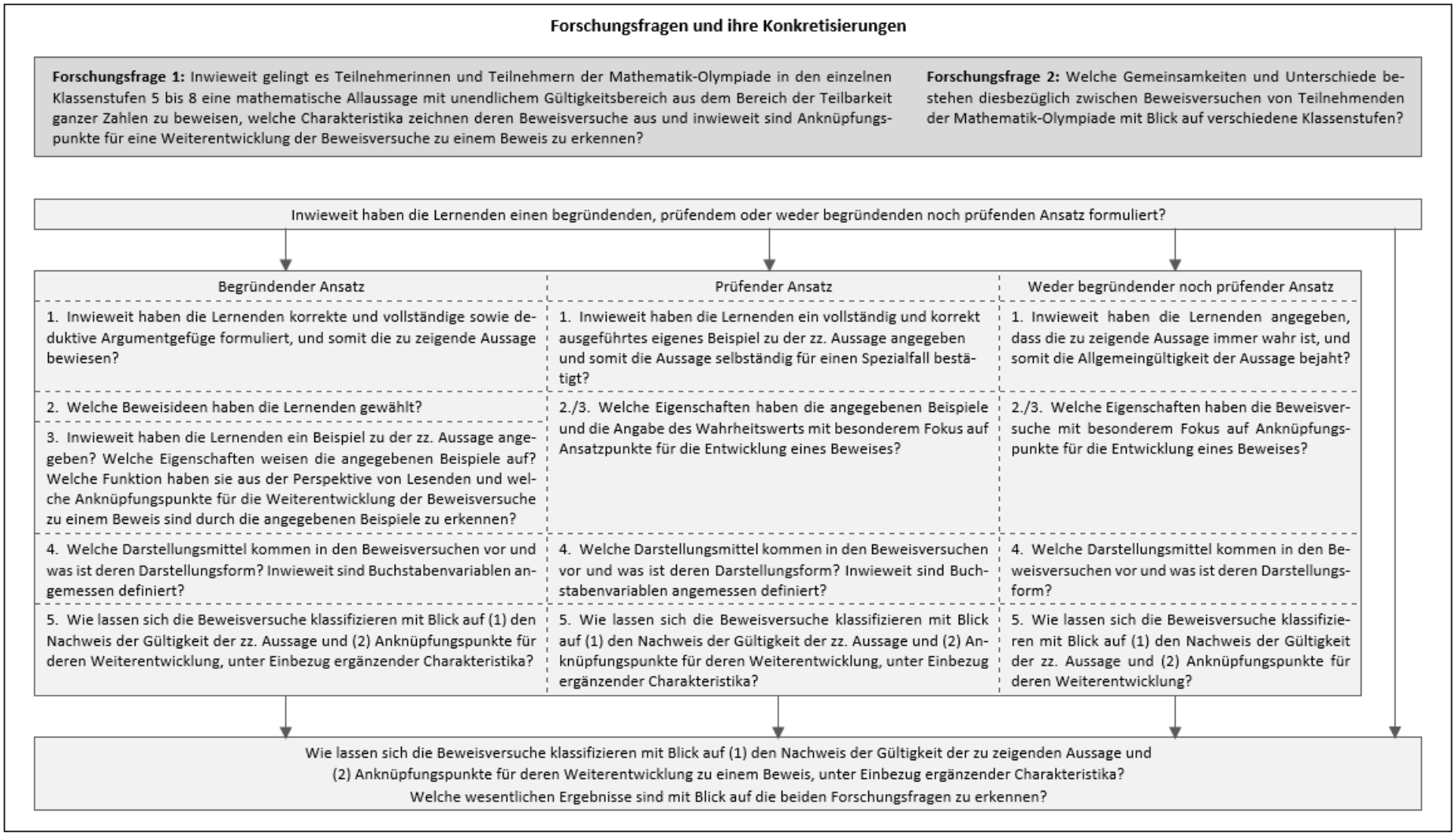


Abbildung 79: Forschungsfragen und ihre Konkretisierungen

7.4 Zusammenfassung des siebten Kapitels

Verlauf der Datenerhebung und Auswertung

Die Erhebung von Beweisversuchen im Rahmen der Regionalrunde der Mathematik-Olympiade 2016/2017 in Nordrhein-Westfalen verlief den Planungen entsprechend. Es liegen **3400 Beweisversuche** zu der in dieser Arbeit fokussierten Allaussage vor.

Für die Analyse dieser Beweisversuche von Teilnehmerinnen und Teilnehmer der Mathematik-Olympiade wird das Verfahren der **qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014)** verwendet. Mit dem Verfahren der qualitativ-strukturierenden Inhaltsanalyse ist es möglich, große Datenmengen in systematischer und zugleich flexibler Form inhaltlich aussagekräftig zu beschreiben und zu vergleichen. Dadurch ist dieses Verfahren für die Untersuchung von Beweisversuchen mit Blick auf die festgelegten Forschungsfragen (Abschn. 1.1, 6.1) optimal geeignet.

In Einklang mit der Grundstruktur des Ablaufs einer qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) bilden die Forschungsfragen den Ausgangspunkt der Analyse von Beweisversuchen. Für diese Analyse werden aus den 3400 Beweisversuchen mittels einer bewussten Stichprobenziehung und einer Ziehung einer Zufallsstichprobe insgesamt 1155 ausgewählt. Jeder Beweisversuch wird als eine Analyse- und als eine Codiereinheit angesehen. Auf Grundlage eines Teils dieser Beweisversuche, der Forschungsfragen, der Aufgabenstellung sowie vorhandener Theorien und Forschungsergebnisse wird deduktiv-induktiv ein Kategoriensystem erstellt. Dieses Kategoriensystem wird pilotiert, finalisiert und schließlich genutzt, um alle ausgewählten Beweisversuche zu codieren.

Um ausgehend von den vorgenommenen Codierungen die Forschungsfragen systematisch und differenziert zu beantworten, werden die Forschungsfragen durch Teilfragen konkretisiert. Es werden die Spektren an vorhandenen Codierungen und deren relative **Häufigkeitsverteilungen analysiert**, die Ergebnisse sukzessive dargestellt und vorsichtig interpretiert. Dabei werden scheinbare **Unterschiede und Zusammenhänge statistisch überprüft** und auf diese Weise eine angemessene Interpretation erhaltener Daten unterstützt. An einzelnen Stellen erfolgen zudem **Fallbeschreibungen**, um detaillierte Einblicke in bestimmte Beweisversuche zu erhalten. Außerdem werden **Klassifikationen** durchgeführt, um erhaltene Ergebnisse zu systematisieren und zu resümieren.

8 Ergebnisse

In diesem Kapitel werden die Ergebnisse der Analyse der Beweisversuche von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade präsentiert. Bei der Präsentation der Ergebnisse beziehen sich Allaussagen, wie z. B. „Keiner der analysierten Beweisversuche mit prüfendem Ansatz …“, und Angaben zu Spektren an Merkmalsausprägungen stets auf alle analysierten Beweisversuche, d. h. diejenigen aus der bewusst gezogenen Stichprobe und der Zufallsstichprobe. Relative Häufigkeitsangaben basieren ausschließlich auf den Beweisversuchen aus der Zufallsstichprobe. Bei den relativen Häufigkeitsangaben ist zu beachten, dass es sich teils um Anteile von Anteilen handelt, beispielsweise um 20% der Beweisversuche mit begründendem Ansatz anstatt um 20% aller Beweisversuche.

Die relativen Häufigkeitsverteilungen werden im Folgenden mithilfe zahlreicher Abbildungen illustriert. Diesbezüglich ist wegen der begrenzten farblichen Darstellungsmöglichkeiten anzumerken, dass in den Abbildungen zu verschiedenen Fragestellungen gleiche Farben und Farbkombinationen vorkommen können, die jedoch unterschiedliche Bedeutungen haben. In vielen Abbildungen sind passend zu den Diagrammen Ergebnisse statistischer Tests angegeben. Dabei können die angegebenen p-Werte von Signifikanztests auf unterschiedlichen Verfahren basieren. Um welches Verfahren es sich handelt, ist stets kompakt mit einer hochgestellten Abkürzung am p-Wert vermerkt („U“: Mann-Whitney-U-Test, „χ^2“: Chi²-Test nach Karl Pearson, „N-1“: N-1 Chi²-Test nach Egon Pearson, „ex“: exakter Fisher-Test; Abschn. 7.2.2.3 zu den Testverfahren). Zugunsten einer guten Lesbarkeit werden nicht alle durchgeführten, sondern i. d. R. nur die statistisch signifikanten Testergebnisse ausgewiesen.[129] Zudem werden nicht alle Details der Tests und alle Testergebnisse, die den Abbildungen entnommen werden können, zusätzlich im Fließtext erwähnt.

Die nachfolgende Ergebnispräsentation wird anhand der in Abschnitt 7.3.8 formulierten Teilfragen strukturiert. Zuerst geht es darum, inwieweit die Teilnehmenden Beweisversuche mit begründendem, prüfendem oder weder begründendem noch prüfendem Ansatz formuliert haben (Abschn. 8.1). Danach werden die Ergebnisse der Detailanalysen der Beweisversuche mit begründendem Ansatz, der Beweisversuche mit prüfendem Ansatz und der Beweisversuche mit weder begründendem noch prüfendem Ansatz präsentiert (Abschn. 8.2-8.4). Auf dieser Basis werden in Abschnitt 8.5 die wesentlichen Ergebnisse der Analyse der Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage, Charakteristika der Beweisversuche und mögliche Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis herausgearbeitet und gebündelt präsentiert. Leserinnen und Lesern, die sich einen schnellen Überblick verschaffen möchten, wird vorgeschlagen, zunächst den Abschnitt 8.5 zu lesen. Danach bietet es sich je nach Interessenlage an, in den Abschnitten 8.1 bis 8.4 ergänzende, detailliert ausgeführte Ergebnisse einzusehen.

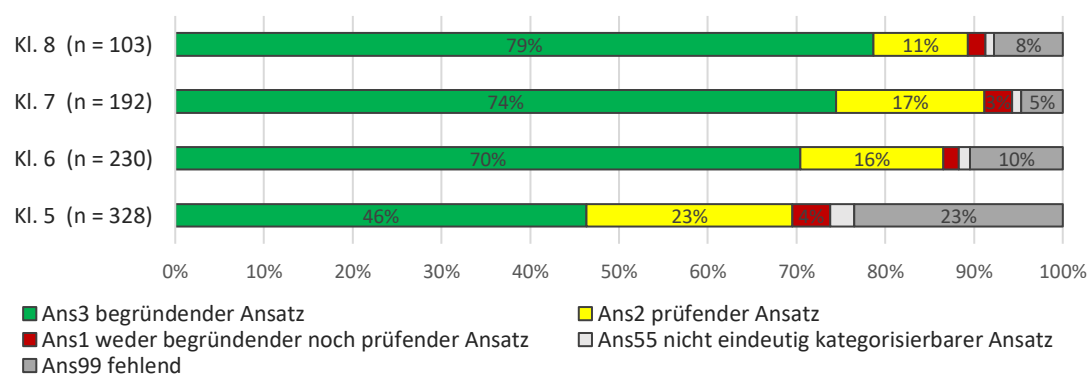
[129] Diese Vorgehensweise ist eher unüblich und zugleich im Rahmen der vorliegenden Arbeit für eine übersichtliche und zügig lesbare Darstellung erforderlich. Die erhaltenen Häufigkeitsverteilungen wurden in systematischer, aufwendiger Form mithilfe statistischer Tests untersucht. Wenn Testergebnisse nicht ausgewiesen sind, kann i. d. R. davon ausgegangen werden, dass diese nicht signifikant sind.

8.1 Begründender, prüfender oder sonstiger Ansatz

Als erster Einblick in die Beweisversuche von Teilnehmenden der Mathematik-Olympiade beschäftigt sich dieser Abschnitt mit der Frage, inwieweit die Teilnehmenden einen begründenden, prüfenden oder weder begründenden noch prüfenden Ansatz formuliert haben. Zur Beantwortung dieser Frage werden die Zuordnungen aller analysierten Beweisversuche zu je genau einer der Oberkategorien *begründender Ansatz*, *prüfender Ansatz*, *weder begründender noch prüfender Ansatz*, *nicht eindeutig kategorisierbarer Ansatz* sowie *fehlend* herangezogen (Abschn. 3.3.2 zu den Kategoriendefinitionen).

Die Abbildung 80 zeigt die relativen Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen. Demnach haben mehr als drei Viertel der Fünftklässlerinnen und Fünftklässler eine Bearbeitung zum Aufgabenteil d) abgegeben, der Anteil an fehlenden Bearbeitungen liegt in der Klassenstufe 5 bei 23%. In den Klassenstufen 6, 7, 8 haben 90% bis 95% der Teilnehmenden eine Bearbeitung zum Aufgabenteil d) abgegeben. Der Anteil an fehlenden Bearbeitungen ist mit 5% bis 10% jeweils statistisch hoch signifikant kleiner als in der Klassenstufe 5 bei geringer Effektstärke. Folglich legen die Daten offen, dass sich die große Mehrheit der Teilnehmenden mit dem Aufgabenteil d) beschäftigt hat. Sie deuten an, dass in den Klassenstufen 6 bis 8 im Vergleich zu der Klassenstufe 5 ein größerer Anteil der Lernenden den Aufgabenteil gelesen und einen Zugang zu der Aufgabenstellung gefunden hat.

Weiterhin ist in Abbildung 80 zu sehen, dass Beweisversuche mit begründendem Ansatz in jeder der Klassenstufen über 45% aller Beweisversuche ausmachen. Dabei ist der Anteil der Beweisversuche mit begründendem Ansatz in der Klassenstufe 5 mit 46% im Vergleich zu den Klassenstufen 6 bis 8 mit je 70% bis 80% statistisch hoch signifikant kleiner bei geringer bis



	p-Werte von Signifikanztests und dazugehörige Effektstärken			
	alle einzelnen Kategorien	Ans3 vs. ¬Ans3	Ans2 vs. ¬Ans2	Ans99 vs. ¬Ans99
Kl. 5 vs. Kl. 6	$p < .001^{ex}, V = .247$	$p < .001^{N-1}, V = .239$	$p = .041^{N-1}, V = .087$	$p < .001^{N-1}, V = .167$
Kl. 5 vs. Kl. 7	$p < .001^{ex}, V = .302$	$p < .001^{N-1}, V = .274$	$p = .078^{N-1}, V = .077$	$p < .001^{N-1}, V = .244$
Kl. 5 vs. Kl. 8	$p < .001^{ex}, V = .278$	$p < .001^{N-1}, V = .276$	$p = .006^{N-1}, V = .133$	$p < .001^{N-1}, V = .168$
Kl. 6 vs. Kl. 7	$p = .218^{ex}, V = .115$	$p = .356^{N-1}, V = .045$	$p = .873^{N-1}, V = .008$	$p = .029^{N-1}, V = .107$

Abbildung 80: Beweisversuche mit begründendem, prüfendem, weder begründendem noch prüfendem und nicht eindeutig kategorisierbarem Ansatz sowie fehlende Bearbeitungen: Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

mittlerer Effektstärke. Während in der Klassenstufe 5 unter der Hälfte der Beweisversuche einen begründenden Ansatz enthalten, haben in den Klassenstufen 6 bis 8 mehr als zwei Drittel der Teilnehmenden einen begründenden Ansatz formuliert. Beweisversuche mit prüfendem Ansatz machen in jeder der Klassenstufen 5 bis 8 mit zwischen 10% und 25% der Zuordnungen einen relevanten Anteil aus, wobei der Anteil in der Klassenstufe 5 statistisch hoch signifikant größer ist als in der Klassenstufe 8 (Abb. 80). Beweisversuche mit weder begründendem noch prüfendem Ansatz sowie Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz kommen ebenfalls in jeder der Klassenstufen 5 bis 8 vor, deren relativer Anteil liegt jeweils unter 5%. Somit weisen die Ergebnisse auf eine große Vielfalt an Beweisversuchen innerhalb der einzelnen Klassenstufen bei gleichzeitiger deutlicher Dominanz von Beweisversuchen mit begründendem Ansatz hin. Von den Teilnehmenden, die sich mit dem Aufgabenteil erkennbar auseinandergesetzt haben, hat über die Hälfte nach allgemeingültigen Argumenten gesucht, was einen guten Anknüpfungspunkt für eine Weiterentwicklung von Beweisversuchen zu einem Beweis darstellt.

Die Einordnung eines Beweisversuchs in die Kategorie *begründender Ansatz* bedeutet allerdings weder, dass das Argumentgefüge durchgehend in allgemeingültiger Form formuliert wurde, noch, dass es korrekt und vollständig ist. Die Oberkategorien *begründender Ansatz*, *prüfender Ansatz* und *weder begründender noch prüfender Ansatz* umfassen jeweils ein großes Spektrum an möglichen Bearbeitungen, das im Folgenden systematisch ausdifferenziert wird (Abschn. 3.3.2, 7.3.8).

8.2 Beweisversuche mit begründendem Ansatz

Dieser Abschnitt fokussiert auf Beweisversuche mit begründendem Ansatz. In Beweisversuchen mit begründendem Ansatz weist das Argumentgefüge einen Bezug zu der infrage stehenden Aussage auf und enthält eine Beweisidee, deren Allgemeingültigkeit oder Verallgemeinerbarkeit in der Bearbeitung explizit angedeutet ist. Solche Beweisversuche können jedoch Fehler oder Lücken enthalten. Auch bezüglich der Schlussweise und Verwendung von Beispielen, der vorkommenden Beweisideen, der Darstellungsmittel und der Darstellungsform lässt die Definition Spielraum (Abschn. 3.3.2). Um diesbezüglich ein differenziertes Bild zu erhalten, wurden die Beweisversuche mit begründendem Ansatz detailliert analysiert.

Im Folgenden werden die Ergebnisse der detaillierten Analyse dieser Beweisversuche präsentiert. Zuerst wird dargelegt, inwieweit die Schülerinnen und Schüler korrekte und vollständige sowie deduktive Argumentgefüge formuliert, und somit die zu zeigende Aussage bewiesen haben (Abschn. 8.2.1). In einem zweiten Schritt wird beschrieben, welche Beweisideen die Lernenden gewählt haben (Abschn. 8.2.2). Danach liegt der Fokus auf der Verwendung von Beispielen. Die Ergebnisse stellen dar, inwieweit Beispiele in den Beweisversuchen mit begründendem Ansatz vorkommen und welche Funktion ihnen zukommt (Abschn. 8.2.3). Im vierten Abschnitt geht es darum, welche Darstellungsmittel in den Beweisversuchen mit begründendem Ansatz vorkommen und was deren Darstellungsform ist (Abschn. 8.2.4). Aufbauend auf den Erkenntnissen zu den einzelnen Charakteristika der Beweisversuche mit begründendem

Ansatz, werden diese abschließend mit Blick auf (1) den Nachweis der Gültigkeit der zu zeigenden Aussage und (2) Anknüpfungspunkte für deren Weiterentwicklung klassifiziert. In diesem Zuge werden auch Zusammenhänge zwischen verschiedenen Eigenschaften der Beweisversuche mit begründendem Ansatz präsentiert (Abschn. 8.2.5, 8.2.6). Am Ende jedes Abschnitts 8.2.1 bis 8.2.6 sind wesentliche Erkenntnisse zusammenfassend formuliert.

Für eine gute Lesbarkeit der Abschnitte 8.2.1 bis 8.2.6 wird an vielen Stellen verkürzend der Ausdruck „Beweisversuche“ für „Beweisversuche mit begründendem Ansatz“ verwendet. Bei sämtlichen Prozentangaben ist zu bedenken, dass sich diese nicht auf alle Beweisversuche, sondern auf die Beweisversuche mit begründendem Ansatz beziehen.

8.2.1 Korrektheit, Vollständigkeit und Schlussweise

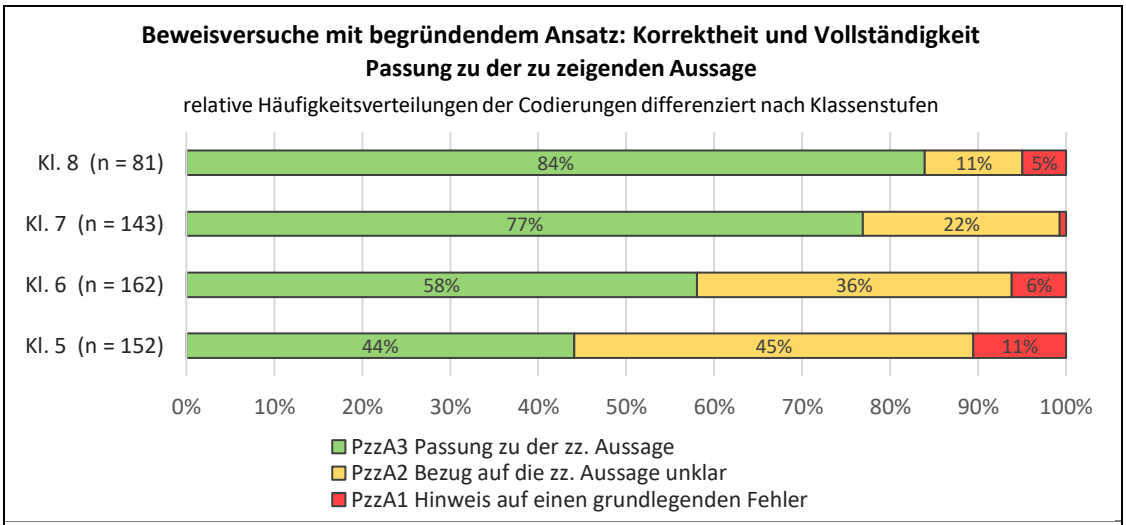
Dieser Abschnitt gibt Einblicke, inwieweit es den Lernenden, die einen Beweisversuch mit begründendem Ansatz abgegeben haben, gelungen ist, die zu zeigende Aussage zu beweisen. Als Beweis werden in dieser Arbeit korrekte und vollständige, deduktive Argumentgefüge bezeichnet (Abschn. 3.4). Dementsprechend stellt sich die Frage, inwieweit die Lernenden korrekte und vollständige sowie deduktive Argumentgefüge formuliert haben. Im Folgenden werden nacheinander die Korrektheit und Vollständigkeit der Beweisversuche (Abschn. 8.2.1.1), die Häufigkeiten der verschiedenen Schlussweisen (Abschn. 8.2.1.2) und das Zusammenspiel der Korrektheit, Vollständigkeit und Schlussweisen beschrieben (Abschn. 8.2.1.3) und wesentliche Ergebnisse zusammengefasst (Abschn. 8.2.1.4).

8.2.1.1 Korrektheit und Vollständigkeit

Zur Beschreibung der Korrektheit und Vollständigkeit wurde jeder Beweisversuch in zwei Schritten codiert (Abschn. 3.3.4). Zuerst wurde die **Passung der Ausführungen zu der zu zeigenden Aussage** analysiert. Die Abbildung 81 zeigt die dazugehörigen relativen Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen.

Mit Blick auf die Klassenstufen 5 bis 7 deuten die Ergebnisse folgende Tendenz an: Je höher die Klassenstufe, desto häufiger geht aus den Beweisversuchen mit begründendem Ansatz eine Passung der Ausführungen zu der zu zeigenden Aussage hervor. In der Klassenstufe 5 passen 44% der Beweisversuche mit begründendem Ansatz aus Expertensicht zu der zu zeigenden Aussage ($\mathrm{PzzA3}$), bei 45% ist der Bezug auf die zu zeigende Aussage unklar ($\mathrm{PzzA2}$) und in 11% ist ein Hinweis auf einen grundlegenden Fehler zu erkennen ($\mathrm{PzzA1}$). Bei mehr als der Hälfte der Beweisversuche mit begründendem Ansatz von Fünftklässlerinnen und Fünftklässlern erscheint es also sinnvoll, für eine Weiterentwicklung der Beweisversuche zuerst die Interpretation der zu zeigenden Aussage sicherzustellen. Im Vergleich dazu passen in der Klassenstufe 7 mehr als 75% der Beweisversuche mit begründendem Ansatz zu der zu zeigenden Aussagen ($\mathrm{PzzA3}$), bei 22% der Beweisversuche ist dieser Bezug unklar ($\mathrm{PzzA2}$) und rund 1% enthalten einen Hinweis auf einen grundlegenden Fehler ($\mathrm{PzzA1}$). Der paarweise Vergleich der relativen Häufigkeitsverteilungen in den Klassenstufen 5 und 6, 5 und 7 sowie 6 und 7 zeigt, dass jeweils in der höheren Klassenstufe die Wahrscheinlichkeit statistisch signifikant größer ist, eine Zuordnung zu einer höheren der drei Kategorien zu erhalten (Abb. 81).

In der Klassenstufe 8 sind die Ergebnisse der Analyse der Passung der Ausführungen zu der zu zeigenden Aussage bei Beweisversuchen mit begründendem Ansatz denen in der Klassenstufe 7 ähnlich, die erkennbaren Unterschiede sind nicht statistisch signifikant. In beiden Klassenstufen 7, 8 liegt der Anteil an Beweisversuchen in der Kategorie PzzA3 *Passung zu der zu zeigenden Aussage* über 75% und ist jeweils statistisch hoch signifikant größer mit $0{,}20 \leq V < 0{,}40$ als in den Klassenstufen 5, 6 (Abb. 81).



p-Werte von Signifikanztests und dazugehörige Effektstärken

	PzzA1, PzzA2, PzzA3
Kl. 5 vs. Kl. 6	$p = .011^{U}$, $r = .143$
Kl. 5 vs. Kl. 7	$p < .001^{U}$, $r = .354$
Kl. 5 vs. Kl. 8	$p < .001^{U}$, $r = .333$
Kl. 6 vs. Kl. 7	$p < .001^{U}$, $r = .220$
Kl. 6 vs. Kl. 8	$p < .001^{U}$, $r = .215$
Kl. 7 vs. Kl. 8	$p = .288^{U}$, $r = .029$

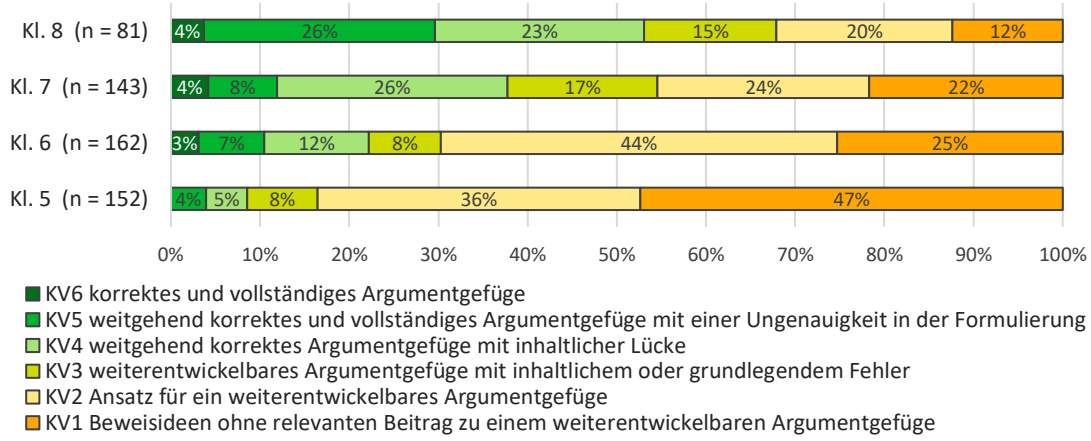
	PzzA3 vs. ¬PzzA3
Kl. 5 vs. Kl. 7	$p < .001^{N-1}$, $V = .335$
Kl. 6 vs. Kl. 7	$p < .001^{N-1}$, $V = .200$
Kl. 5 vs. Kl. 8	$p < .001^{N-1}$, $V = .385$
Kl. 6 vs. Kl. 8	$p < .001^{N-1}$, $V = .259$

	PzzA2 vs. ¬PzzA2
Kl. 5 vs. Kl. 7	$p < .001^{N-1}$, $V = .242$
Kl. 6 vs. Kl. 7	$p = .010^{N-1}$, $V = .147$
Kl. 5 vs. Kl. 8	$p < .001^{N-1}$, $V = .346$
Kl. 6 vs. Kl. 8	$p < .001^{N-1}$, $V = .260$

Abbildung 81: Beweisversuche mit begründendem Ansatz: Passung zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Zur Beschreibung der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz wurden nach der Passung der Ausführungen zu der zu zeigenden Aussage die **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** analysiert. Dabei wurde jeder Beweisversuch mit begründendem Ansatz genau einer von sechs Kategorien zugeordnet (Abschn. 3.3.4). Die Abbildung 82 zeigt die relativen Häufigkeitsverteilungen der Zuordnungen differenziert nach Klassenstufen sowie ergänzende statistische Daten.

In jeder der Klassenstufen 5 bis 8 liegt der Anteil an korrekten und vollständigen Argumentgefügen (KV6) an allen Beweisversuchen mit begründendem Ansatz unter 5%. In der Klassenstufe 5 ist in keinem der analysierten Beweisversuche ein korrektes und vollständiges Argumentgefüge zu erkennen.



	p-Werte von Signifikanztests und dazugehörige Effektstärken				
	KV1 bis KV6 einzeln	KV1 vs. KV2-6	KV1-2 vs. KV3-6	KV1-3 vs. KV4-6	KV5 vs. ¬KV5
Kl. 6 vs. Kl. 5	$p < .001^{U}$, $r = .247$	$p < .001^{N\text{-}1}$, $V = .230$	$p = .004^{N\text{-}1}$, $V = .162$	$p = .001^{N\text{-}1}$, $V = .188$	$p = .188^{N\text{-}1}$, $V = .074$
Kl. 7 vs. Kl. 5	$p < .001^{U}$, $r = .395$	$p < .001^{N\text{-}1}$, $V = .269$	$p < .001^{N\text{-}1}$, $V = .399$	$p < .001^{N\text{-}1}$, $V = .348$	$p = .168^{N\text{-}1}$, $V = .080$
Kl. 7 vs. Kl. 6	$p = .004^{U}$, $r = .167$	$p = .457^{N\text{-}1}$, $V = .043$	$p < .001^{N\text{-}1}$, $V = .246$	$p = .003^{N\text{-}1}$, $V = .170$	$p = .925^{N\text{-}1}$, $V = .005$
Kl. 8 vs. Kl. 5	$p < .001^{U}$, $r = .233$	$p < .001^{N\text{-}1}$, $V = .349$	$p < .001^{N\text{-}1}$, $V = .516$	$p < .001^{N\text{-}1}$, $V = .496$	$p < .001^{N\text{-}1}$, $V = .327$
Kl. 8 vs. Kl. 6	$p < .001^{U}$, $r = .243$	$p = .020^{N\text{-}1}$, $V = .150$	$p < .001^{N\text{-}1}$, $V = .359$	$p < .001^{N\text{-}1}$, $V = .311$	$p < .001^{N\text{-}1}$, $V = .255$
Kl. 8 vs. Kl. 7	$p = .005^{U}$, $r = .224$	$p = .083^{N\text{-}1}$, $V = .116$	$p = .051^{N\text{-}1}$, $V = .131$	$p = .026^{N\text{-}1}$, $V = .149$	$p < .001^{N\text{-}1}$, $V = .250$

Abbildung 82: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Des Weiteren zeigen die Ergebnisse, dass innerhalb jeder der Klassenstufen deutliche Unterschiede hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges bestehen. In allen Klassenstufen kommt jede der Kategorien KV1 bis KV5 vor. Einzelnen Teilnehmenden aus der Klassenstufe 5 ist ein weitgehend korrektes und vollständiges Argumentgefüge (KV5-6) gelungen, in manchen Beweisversuchen mit begründendem Ansatz aus der Klassenstufe 8 ist kein relevanter Beitrag zu einem weiterentwickelbaren Argumentgefüge (KV1) formuliert.

Zugleich deuten die Ergebnisse, vereinfacht gesagt, folgende Tendenz an: Je höher die Klassenstufe, desto besser sind die Beweisversuche mit begründendem Ansatz in Bezug auf die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Der paarweise Vergleich der relativen Häufigkeitsverteilungen mittels des Mann-Whitney-U-Tests zeigt, dass in der jeweils höheren Klassenstufe die Wahrscheinlichkeit statistisch signifikant größer ist, eine Zuordnung zu einer höheren der sechs Kategorien zu erhalten, bei geringer bis mittlerer Effektstärke. In der Klassenstufe 5 gehören fast 50% der Beweisversuche mit begründendem Ansatz in die Kategorie KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge* und 36% in die Kategorie KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge*. Insgesamt enthalten also über 80% der Beweisversuche kein weiterentwickelbares Argumentgefüge (KV1-2). Im Vergleich zu der Klassenstufe 5 ist in der Klassenstufe 6 insbesondere der Anteil der Kategorie KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge* statisch hoch signifikant kleiner. Außerdem ist der Gesamt-

anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge (KV4-6) in der Klassenstufe 6 statistisch hoch signifikant größer als in der Klassenstufe 5. In der Klassenstufe 7 ist der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge (KV4-6) noch größer und statistisch hoch signifikant größer als in beiden Klassenstufen 5, 6. Dabei enthält in jeder der Klassenstufen 5 bis 7 mehr als die Hälfte der Beweisversuche mit einem weitgehend korrekten Argumentgefüge (KV4-6) eine inhaltliche Lücke (KV4). In der Klassenstufe 8 liegt der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge an allen Beweisversuchen mit begründendem Ansatz über 50%, er ist statistisch signifikant größer als in jeder der Klassenstufen 5 bis 7. In ca. der Hälfte dieser Beweisversuche ist ein weitgehend korrektes und vollständiges Argumentgefüge (KV5) zu erkennen. Der Anteil der Beweisversuche in der Kategorie KV 5 an allen Beweisversuchen mit begründendem Ansatz ist in der Klassenstufe 8 im Vergleich zu jeder der Klassenstufen 5, 6, 7 statistisch hoch signifikant größer mit $0{,}25 \leq V < 0{,}33$.

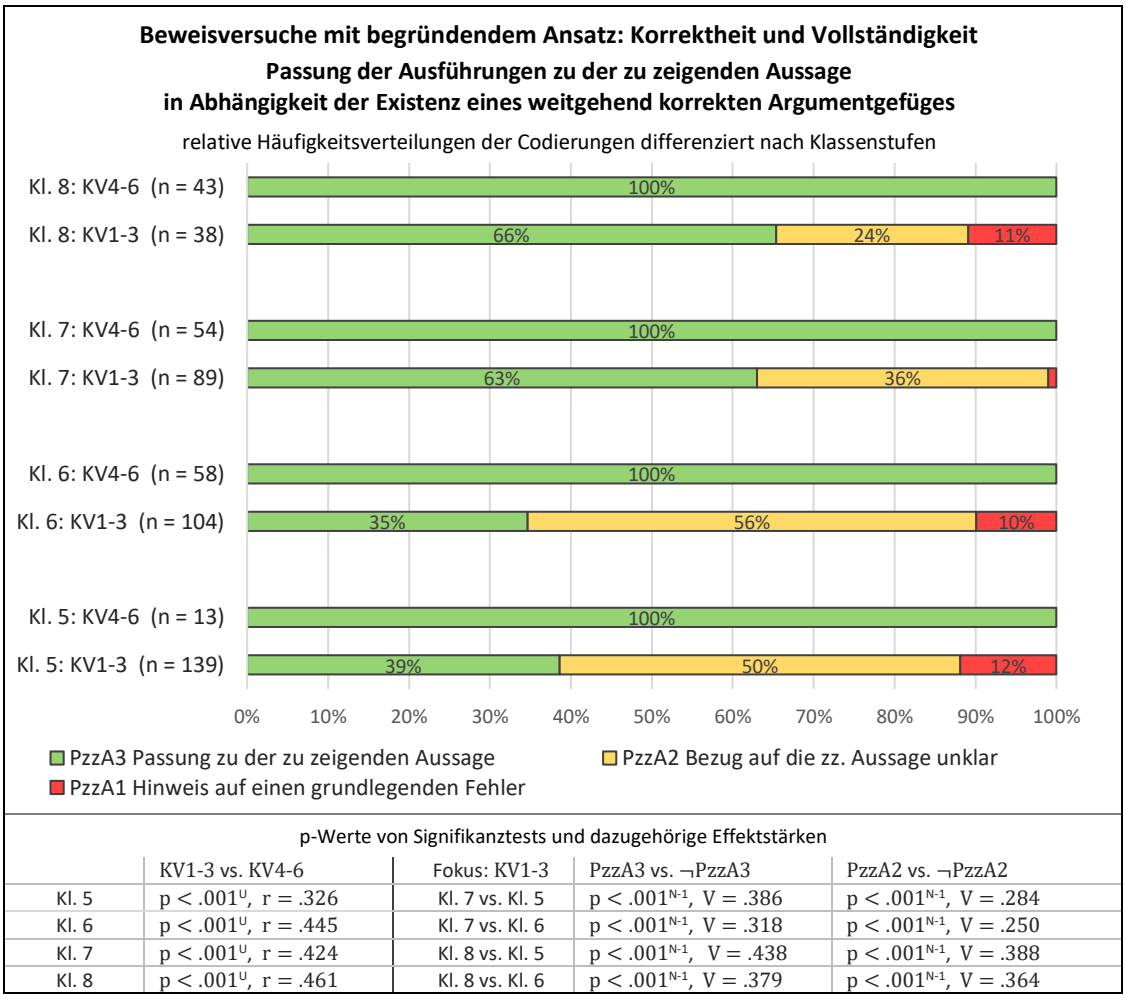
Der Anteil an korrekten und vollständigen Argumentgefügen (KV 6) hingegen liegt in der Klassenstufe 8 so wie in den Klassenstufen 6, 7 bei ca. 4%. Insbesondere auch mit Blick auf Lernende in der Klassenstufe 8 scheint zwischen der Formulierung eines weitgehend korrekten und vollständigen Argumentgefüges mit einer Ungenauigkeit in der Formulierung (KV 5) und der Abgabe eines korrekten und vollständigen Argumentgefüges ein deutlicher Anforderungssprung zu bestehen, obwohl aus fachinhaltlicher Perspektive Beweisversuche in der Kategorie KV 5 ebenfalls „gut" sind, insofern als dass ausschließlich Formulierungslücken oder Formulierungsfehler vorkommen, die die prinzipielle Korrektheit und Vollständigkeit kaum beeinflussen und vom Lernenden nach einem Hinweis wahrscheinlich sofort behoben werden könnten.

Zwischen den beiden Merkmalen zur Beschreibung der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz, d. h. der **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** und der **Passung der Ausführungen zu der zu zeigenden Aussage**, besteht mit Blick auf jede der Klassenstufen 5 bis 8 ein statistisch hoch signifikanter Zusammenhang (Abb. 83).[130] Bei sämtlichen Beweisversuchen mit einem weitgehend korrekten Argumentgefüge (KV4-6) passen die Ausführungen aus Expertensicht zu der zu zeigenden Aussage (PzzA3). Dementsprechend fallen alle Beweisversuche mit einem Hinweis auf einen grundlegenden Fehler oder ohne klar erkennbaren Bezug auf die zu zeigende Aussage (PzzA1-2) in eine der Kategorien KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge* bis KV3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler*.

Zugleich impliziert das Fehlen eines weitgehend korrekten Argumentgefüges (KV1-3) nicht unbedingt, dass ein Beweisversuch mit begründendem Ansatz einen Hinweis auf einen grundlegenden Fehler oder keinen klar erkennbaren Bezug auf die zu zeigende Aussage aufweist

[130] Dieser Zusammenhang ist im Wesentlichen mit der inhaltlichen Nähe der beiden Merkmale zu erklären. Trotzdem erscheint die separate Analyse der Passung zu der zu zeigenden Aussage und der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges wegen der Möglichkeit der Zuordnung von Beweisversuchen mit einem Hinweis auf einen grundlegenden Fehler (PzzA1) oder ohne klar erkennbaren Bezug auf die zu zeigende Aussage (PzzA2) zu jeder der Kategorien KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge* bis KV3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler* als sinnvoll und lohnend.

(PzzA1-2). Auch bei Beweisversuchen mit begründendem Ansatz, aber ohne weitgehend korrektes Argumentgefüge (KV1-3) ist es gut möglich, dass eine Passung zu der zu zeigenden Aussage besteht (PzzA3). Der Anteil der Beweisversuche in der Kategorie *Passung zu der zu zeigenden Aussage* (PzzA3) an den Beweisversuchen mit begründendem Ansatz, aber ohne weitgehend korrektes Argumentgefüge liegt in den Klassenstufen 5, 6 bei 35% bis 40% und in den Kalssenstufen 7, 8 bei ca. 65%. Er ist in den Klassenstufen 7, 8 jeweils statistisch hoch signifikant größer als in den Klassenstufen 5, 6 bei mittlerer Effektstärke (Abb. 83).



p-Werte von Signifikanztests und dazugehörige Effektstärken

	KV1-3 vs. KV4-6	Fokus: KV1-3	PzzA3 vs. ¬PzzA3	PzzA2 vs. ¬PzzA2
Kl. 5	$p < .001^{U}$, $r = .326$	Kl. 7 vs. Kl. 5	$p < .001^{N-1}$, $V = .386$	$p < .001^{N-1}$, $V = .284$
Kl. 6	$p < .001^{U}$, $r = .445$	Kl. 7 vs. Kl. 6	$p < .001^{N-1}$, $V = .318$	$p < .001^{N-1}$, $V = .250$
Kl. 7	$p < .001^{U}$, $r = .424$	Kl. 8 vs. Kl. 5	$p < .001^{N-1}$, $V = .438$	$p < .001^{N-1}$, $V = .388$
Kl. 8	$p < .001^{U}$, $r = .461$	Kl. 8 vs. Kl. 6	$p < .001^{N-1}$, $V = .379$	$p < .001^{N-1}$, $V = .364$

Abbildung 83: Beweisversuche mit begründendem Ansatz: Passung zu der zz. Aussage in Abhängigkeit der Existenz eines weitgehend korrekten Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

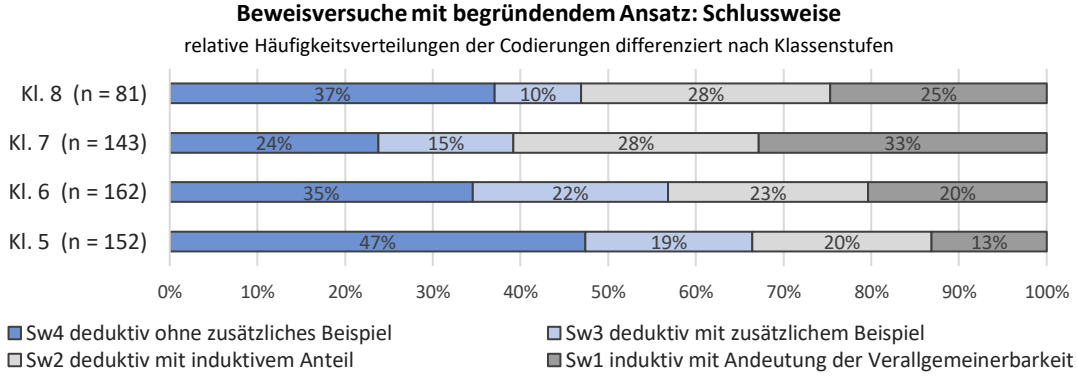
Bezüglich der **Korrektheit und Vollständigkeit** der Beweisversuche mit begründendem Ansatz lassen sich **zusammenfassend** folgende Tendenzen festhalten: Innerhalb jeder der Klassenstufen 5 bis 8 unterscheiden sich die Beweisversuche mit begründendem Ansatz hinsichtlich der Korrektheit und Vollständigkeit deutlich, sowohl mit Blick auf die Passung der Ausführungen zu der zu zeigenden Aussage als auch hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Dabei ist der Anteil der Beweisversuche mit einer klaren Passung zu der zu zeigenden Aussage bezogen auf die Beweisversuche mit begründendem Ansatz, aber keinem weitgehend korrekten Argumentgefüge in den Klassenstufen 7, 8 größer als in den Klassenstufen 5, 6. Die Beweisversuche mit begründendem Ansatz sind in Bezug auf die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüge im Mittel je

höher die Klassenstufe, desto besser. Ein korrektes und vollständiges Argumentgefüge enthalten jedoch je Klassenstufe unter 5% der Beweisversuche mit begründendem Ansatz.

8.2.1.2 Schlussweise

Um herauszuarbeiten, inwieweit den Schülerinnen und Schülern, die einen Beweisversuch mit begründendem Ansatz abgegeben haben, ein Beweis gelungen ist, wurde zusätzlich zur Korrektheit und Vollständigkeit die Schlussweise codiert (Abschn. 3.3.6 zum Kategoriensystem). Die Abbildung 84 zeigt die relativen Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen und ergänzende statistische Daten. In allen Klassenstufen 5 bis 8 kennzeichnet jede der unterschiedenen Schlussweisen Sw1-4 über 10% der Beweisversuche mit begründendem Ansatz. Die Beweisversuche mit begründendem Ansatz sind also bezüglich der Schlussweise in jeder der Klassenstufen vielfältig ausgeprägt.

Bezogen auf die Beweisversuche mit begründendem Ansatz ist in den Klassenstufen 5, 6 der Anteil der Kategorie Sw4 *deduktiv ohne zusätzliches Beispiel* am größten, der Gesamtanteil der Kategorien Sw3-4 *deduktiv* liegt jeweils über 50%.[131] Beweisversuche mit der Schlussweise Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit* kommen am seltensten vor. Im Vergleich zu den Klassenstufen 5, 6 ist in der Klassenstufe 7 der Anteil der Kategorie Sw4 und der Gesamtanteil der Kategorien Sw3-4 statistisch signifikant kleiner und der Anteil der Kategorie Sw1 statistisch signifikant größer. In der Klassenstufe 8 ist in etwa der Hälfte der Beweisversuche die Schlussweise deduktiv (Sw3-4) bzw. zwischen induktiv und deduktiv (Sw1-2).



p-Werte von Signifikanztests und dazugehörige Effektstärken

	alle Kategorien	Sw 4 vs. Sw1-3	Sw 1 vs. Sw2-4	Sw1-2 vs. Sw3-4
Kl. 6 vs. Kl. 5	$p = .218^{X^2}$, V = .140	$p = .021^{N-1}$, V = .130	$p = .089^{N-1}$, V = .096	$p = .079^{N-1}$, V = .099
Kl. 7 vs. Kl. 5	$p < .001^{X^2}$, V = .299	$p < .001^{N-1}$, V = .246	$p < .001^{N-1}$, V = .235	$p < .001^{N-1}$, V = .273
Kl. 7 vs. Kl. 6	$p = .017^{X^2}$, V = .183	$p = .040^{N-1}$, V = .118	$p = .013^{N-1}$, V = .142	$p = .002^{N-1}$, V = .176
Kl. 8 vs. Kl. 5	$p = .009^{X^2}$, V = .204	$p = .131^{N-1}$, V = .099	$p = .027^{N-1}$, V = .146	$p = .004^{N-1}$, V = .190
Kl. 8 vs. Kl. 7	$p = .138^{X^2}$, V = .157	$p = .035^{N-1}$, V = .141	$p = .200^{N-1}$, V = .086	$p = .260^{N-1}$, V = .075

Abbildung 84: Beweisversuche mit begründendem Ansatz: Schlussweise. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

[131] Wie im folgenden Abschnitt 8.2.1.3 deutlich werden wird, enthalten von den Beweisversuchen mit deduktiver Schlussweise aus den Klassenstufen 5, 6 über 80% kein weiterentwickelbares Argumentgefüge. In diesen Beweisversuchen sind die angegebenen Beweisideen teils falsch, sehr nah an der zu zeigenden Aussage oder nicht als mehrschrittiges Argumentgefüge ausgeführt. Beispielsweise lautet der Beweisversuch einer Fünftklässlerin, dass alle Summen durch vier teilbar sind, weil im Ergebnis in der Einerstelle eine 4 oder 6 oder 8 oder 2 steht. Andere Schülerinnen und Schüler haben ausschließlich geschrieben, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen stets in der Viererreihe ist. Manche Teilnehmende haben ausschließlich angegeben, dass die Summe von zwei aufeinanderfolgenden ungeraden Zahlen immer durch vier teilbar ist.

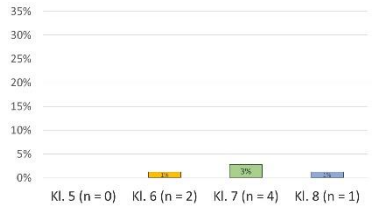
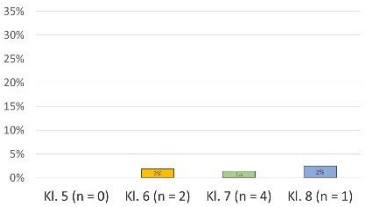
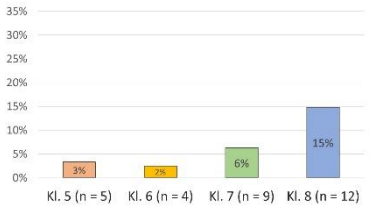
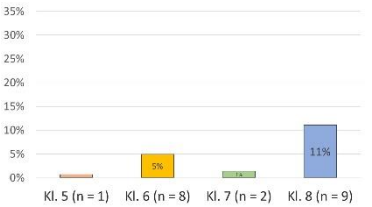
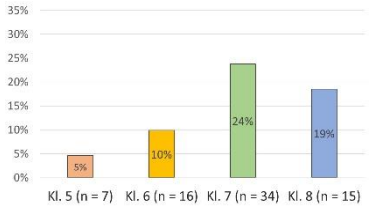
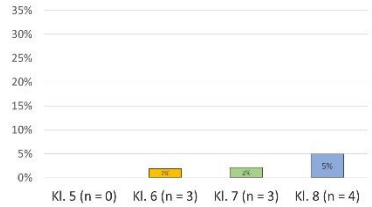
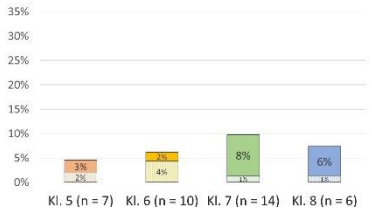
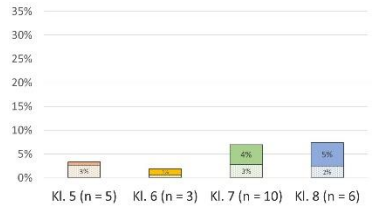
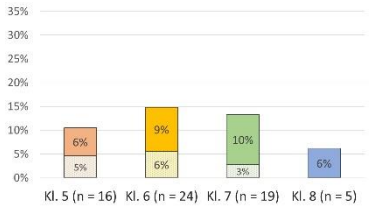
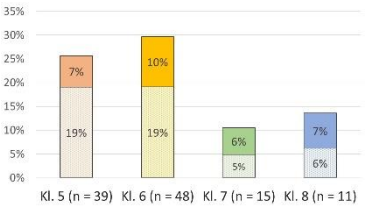
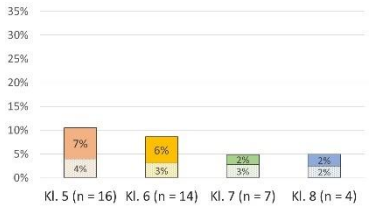
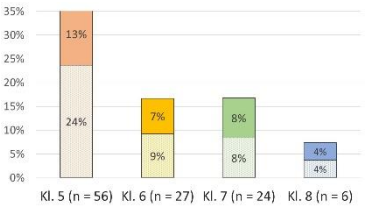
8.2.1.3 Zusammenhänge zwischen der Korrektheit und Vollständigkeit der Beweisversuche mit begründendem Ansatz und der Schlussweise

Nachdem die Ergebnisse der Analyse der Korrektheit und Vollständigkeit der Beweisversuche mit begründendem Ansatz und der Untersuchung der Schlussweise separat dargestellt wurden, werden diese im Folgenden in ihrem Zusammenspiel betrachtet. Primäres Ziel ist die Beantwortung der Frage, inwieweit es den Lernenden, die einen Beweisversuch mit begründendem Ansatz abgegeben haben, gelungen ist, die zu zeigende Aussage zu beweisen. Nebenbei werden Zusammenhänge zwischen der Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz und den Schlussweisen deutlich, die sowohl mit Blick auf den Nachweis der Gültigkeit als auch hinsichtlich möglicher Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis interessant sind.

Als Beweise gelten in dieser Arbeit **korrekte und vollständige, deduktive Argumentgefüge** (Abschn. 3.4). Nicht in der Klassenstufe 5, aber in allen Klassenstufen 6, 7, 8 enthalten ca. 4% der Beweisversuche mit begründendem Ansatz ein korrektes und vollständiges Argumentgefüge. Der Anteil an korrekten, vollständigen und zusätzlich deduktiven Argumentgefügen ($\mathrm{KV}6$,$\mathrm{Sw}3\text{-}4$) an allen Beweisversuchen mit begründendem Ansatz liegt in den Klassenstufen 6 bis 8 jeweils zwischen $0{,}5\%$ und 3%. Folglich zeigen die Daten, dass in den Klassenstufen 6 bis 8 Beweise vorkommen, diese jedoch Ausnahme darstellen.

Während der Anteil an Beweisen gering ist, haben die Analysen der Korrektheit und Vollständigkeit der Beweisversuche sowie der Schlussweise z. B. offengelegt, dass in der Klassenstufe 8 ca. 50% der Beweisversuche mit begründendem Ansatz ein weitgehend korrektes Argumentgefüge enthalten und ca. 50% ein deduktives Argumentgefüge. Um differenzierte und aussagekräftige Einblicke zu geben, inwieweit den Schülerinnen und Schülern, die einen Beweisversuch mit begründendem Ansatz abgegeben haben, ein Beweis gelungen ist, werden im Folgenden die Ausprägungen der Korrektheit und Vollständigkeit der Beweisversuche und der Schlussweise miteinander kombiniert und die Beweisversuche mit begründendem Ansatz auf diese Weise gruppiert.[132] Dabei werden die vier Schlussweisen zu den zwei Kategorien $\mathrm{Sw}1\text{-}2$ *zwischen induktiv und deduktiv* und $\mathrm{Sw}3\text{-}4$ *deduktiv* zusammengefasst. Bei Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise ($\mathrm{Sw}1\text{-}2$) ist mindestens ein Teil der Schlüsse an konkreten Beispielen ausgeführt oder ein die inhaltlichen Ausführungen ergänzendes Beispiel angeführt. Bei Beweisversuchen mit deduktiver Schlussweise $\mathrm{Sw}3\text{-}4$ hingegen könnten alle angegebenen Beispiele, die kein notwendiger Bestandteil des Argumentgefüges sind, weggelassen werden, ohne das Argumentgefüge zu beeinträchtigen. Die Beweisversuche mit deduktiver Schlussweise ($\mathrm{Sw}3\text{-}4$) können prinzipiell, d. h. im Falle eines korrekten und vollständigen Argumentgefüges ($\mathrm{KV}6$), als Beweis gelten (Abschn. 3.3.6 zu den Kategorien; 3.4 zum Beweisbegriff). In Abbildung 85 sind die aus der **Verknüpfung der Ausprägungen der Korrektheit, Vollständigkeit der Beweisversuche und der Schlussweise** resultierenden Gruppen systematisch dargestellt. Bezogen auf jede der Gruppen ist angegeben, ein wie großer Anteil der Beweisversuche aus den einzelnen Klassenstufen 5 bis 8 in diese Gruppe gehört.

[132] Diese Gruppierung wird in den Abschnitt 8.2.5, 8.2.6 als Ausgangspunkt für die Klassifikation der Beweisversuche mit begründendem Ansatz verwendet.

	Sw1-2 ***zwischen induktiv und deduktiv***	**Sw3-4** ***deduktiv***
KV6 ***korrektes und vollständiges Argumentgefüge*** PzzA3 *Passung zu der zu zeigenden Aussage*	 Kl. 5 - 8: n = 7, h = 1%	 Kl. 5 - 8: n = 7, h = 1%
KV5 weitgehend korrektes, vollständiges Arg.gefüge mit einer Ungenauigkeit in der Formulierung PzzA3 *Passung zu der zu zeigenden Aussage*	 Kl. 5 - 8: n = 30, h = 6%	 Kl. 5 - 8: n = 20, h = 4%
KV4 ***weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke*** PzzA3 *Passung zu der zu zeigenden Aussage*	 Kl. 5 - 8: n = 72, h = 13%	 Kl. 5 - 8: n = 10, h = 2%
KV3 ***weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler*** PzzA1-3	 Kl. 5 - 8: n = 37, h = 7%	 Kl. 5 - 8: n = 24, h = 4%
KV2 ***Ansatz für ein weiterentwickelbares Argumentgefüge*** PzzA1-3	 Kl. 5 - 8: n = 64, h = 12%	 Kl. 5 - 8: n = 113, h = 21%
KV1 ***Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge*** PzzA1-3	 Kl. 5 - 8: n = 41, h = 8%	 Kl. 5 - 8: n = 113, h = 21%

PzzA3 Passung zu der zz. Aussage PzzA1-2 keine klare Passung zu der zz. Aussage

Gesamtanzahlen an Beweisversuchen mit begründendem Ansatz
Kl. 5: n = 152 Kl. 6: n = 162 Kl. 7: n = 143 Kl. 8: n = 81 Kl. 5-8: n = 538

n: Gesamtanzahl der Beweisversuche (aus der Zufallsstichprobe) in der fokussierten Gruppe
h: Anteil der Beweisversuche in der fokussierten Gruppe an den n = 538 Beweisversuchen mit begründendem Ansatz

Abbildung 85: Beweisversuche mit begründendem Ansatz: Gruppierung der Beweisversuche bezüglich derer Korrektheit und Vollständigkeit sowie der Schlussweise und dazugehörige Häufigkeitsverteilungen

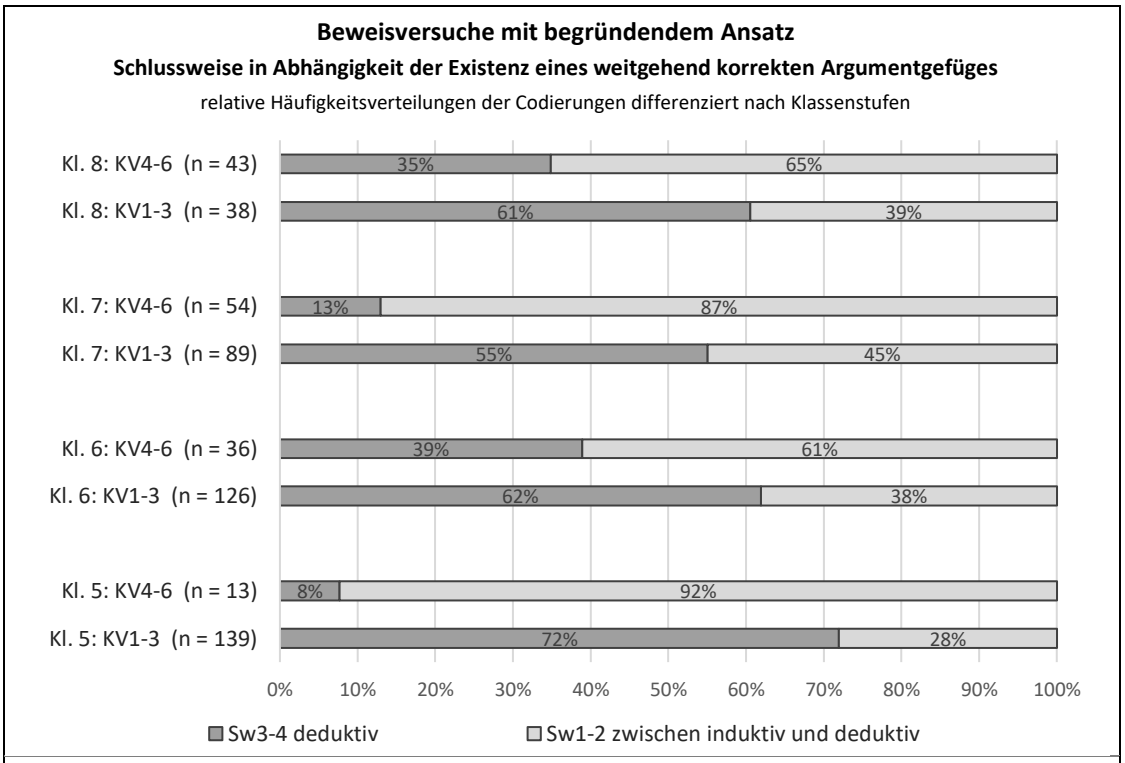
In Abbildung 85 ist, entsprechend den obigen Ausführungen zu sehen, dass je Klassenstufe unter 3% der Beweisversuche mit begründendem Ansatz in die Gruppe KV6, Sw3-4 gehören und somit als Beweis bezeichnet werden können. Wenn man weiter gefasst betrachtet, in welche Gruppen je Klassenstufe die **besten 10% der Beweisversuche mit begründendem Ansatz** gehören, zeigt sich folgendes Bild: In der Klassenstufe 5 enthalten die besten 10% fast alle ein weitgehend korrektes Argumentgefüge (KV4-6), fast immer bei zwischen induktiver und deduktiver Schlussweise (Sw1-2). In den Klassenstufen 6, 7 enthalten die besten 10%, in der Klassenstufe 8 die besten 30% ein weitgehend korrektes und vollständiges Argumentgefüge (KV5-6). Die Schlussweise liegt dabei mehrheitlich zwischen induktiv und deduktiv (Sw1-2).

Weiterhin ist in Abbildung 86 angegeben, **in welchen Gruppen sich besonders viele der Beweisversuche mit begründendem Ansatz befinden**. Je Klassenstufe befinden sich in genau vier der zwölf Gruppen mindestens 10% der Beweisversuche mit begründendem Ansatz. Die Beweisversuche, die in eine dieser Gruppen einzuordnen sind, machen in der fünften Klassenstufe 85%, in der sechsten 78% und in der siebten, achten ca. 58% aller Beweisversuche mit begründendem Ansatz aus. In der Klassenstufe 5 kennzeichnen sich alle vier Gruppen dadurch, dass die Beweisversuche allenfalls einen Ansatz für ein weiterentwickelbares Argumentgefüge (KV1-2) enthalten, während in der Klassenstufe 8 sowohl eine Gruppe mit der Ausprägung KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge* als auch eine mit der Ausprägung KV5 *weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung* zu den Gruppen gehören, in die mindestens 10% der Beweisversuche mit begründendem Ansatz fallen. Es deutet sich an, dass die große Mehrheit der Beweisversuche mit begründendem Ansatz aus den Klassenstufen 5, 6 hinsichtlich der fokussierten Merkmalskombinationen recht homogen sind, die Beweisversuche mit begründendem Ansatz aus den Klassenstufen 7, 8 etwas heterogener.

	Kl. 5 (n = 152)		Kl. 6 (n = 162)		Kl. 7 (n = 143)		Kl. 8 (n = 81)	
1 größter Anteil	KV1, Sw3-4	37%	KV2, Sw3-4	32%	KV4, Sw1-2	24%	KV4, Sw1-2	19%
2 zweitgrößter Anteil	KV2, Sw3-4	26%	KV1, Sw3-4	19%	KV 2, Sw1-2	13%	KV5, Sw1-2	15%
3 drittgrößter Anteil	KV1, Sw1-2	11%	KV2, Sw1-2	16%	KV 2, Sw3-4	10%	KV2, Sw3-4	14%
4 viertgrößter Anteil	KV2, Sw1-2	11%	KV4, Sw1-2	11%	KV 3, Sw1-2	10%	KV5, Sw3-4	11%

Abbildung 86: Beweisversuche mit begründendem Ansatz: Gruppen an Beweisversuchen in Abhängigkeit der Korrektheit, Vollständigkeit und Schlussweise. Besonders häufige Gruppen und deren Anteile differenziert nach Klassenstufen

Zusätzlich zu dieser Homogenitäts- bzw. Heterogenitätstendenz fällt auf, dass bei fast allen Gruppen mit der Ausprägung KV4-6 *weitgehend korrektes Argumentgefüge*, in die mindestens 10% der Beweisversuche mit begründendem Ansatz aus einer der Klassenstufen 5 bis 8 gehören, die Schlussweise zwischen induktiv und deduktiv liegt (Sw1-2, Abb. 86). Die Abbildung 87 zeigt die relativen Häufigkeitsverteilungen der Schlussweisen in Abhängigkeit von der Existenz eines weitgehend korrekten Argumentgefüges in systematischer Form, differenziert nach Klassenstufen. Innerhalb jeder der Klassenstufen 5 bis 8 unterscheiden sich die relativen Häufigkeitsverteilungen der Schlussweisen bei Beweisversuchen in den Kategorien KV1-3 und KV4-6 statistisch signifikant. Bei Beweisversuchen, in denen ein weitgehend korrektes Argumentgefüge (KV4-6) vorkommt, liegt die Schlussweise jeweils mehrheitlich zwischen induktiv und deduktiv, während die hinsichtlich der Korrektheit und Vollständigkeit schwächeren Beweisversuche in der Kategorie KV1-3 mehrheitlich deduktiv formuliert sind.



p-Werte von Signifikanztests und dazugehörige Effektstärken

	KV1-3 vs. KV4-6
Kl. 8	$p = .022^{N-1}, V = .256$
Kl. 7	$p < .001^{N-1}, V = .418$
Kl. 6	$p = .014^{N-1}, V = .193$
Kl. 5	$p < .001^{ex}, V = .381$

	Fokus: KV1-3
Kl. 7 vs. Kl. 5	$p = .009^{N-1}, V = .173$

	Fokus: KV4-6
Kl. 6 vs. Kl. 5	$p = .043^{ex}, V = .299$
Kl. 6 vs. Kl. 7	$p = .005^{N-1}, V = .300$
Kl. 8 vs. Kl. 5	$p = .082^{ex}, V = .254$
Kl. 8 vs. Kl. 7	$p = .014^{N-1}, V = .260$

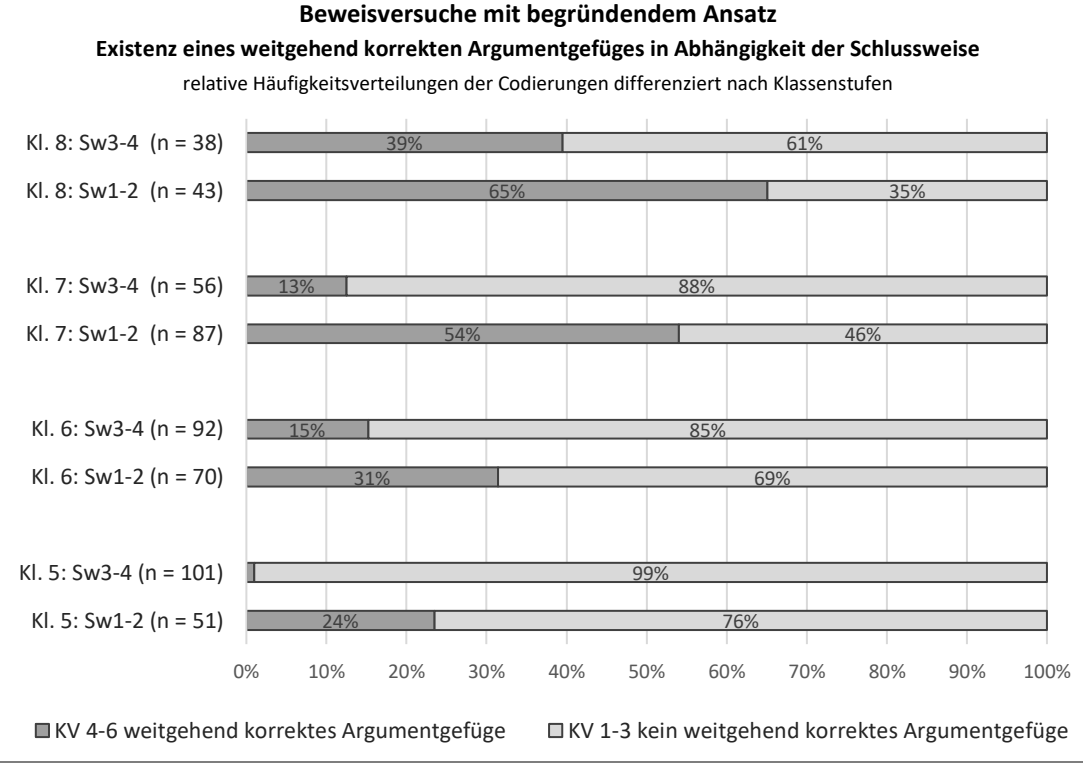
Abbildung 87: Beweisversuche mit begründendem Ansatz: Schlussweise in Abhängigkeit der Existenz eines weitgehend korrekten Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Die nachfolgende Abbildung 88 zeigt umgekehrt die relativen Häufigkeitsverteilungen der **Existenz eines weitgehend korrekten Argumentgefüges in Abhängigkeit von der Schlussweise** differenziert nach Klassenstufen. Sie verdeutlicht, dass in jeder der Klassenstufen 5 bis 8 ein größerer Anteil der Beweisversuche, bei denen die Schlussweise zwischen induktiv und deduktiv (Sw1-2) liegt, ein weitgehend korrektes Argumentgefüge (KV4-6) enthält als es bei Beweisversuchen mit deduktiver Schlussweise (Sw3-4) der Fall ist. Der Unterschied ist, entsprechend den vorhergehenden Ausführungen, jeweils statistisch signifikant (Abb. 87, 88).

Bei den Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise (Sw1-2) liegt der Anteil derjenigen mit einem weitgehend korrekten Argumentgefüge (KV4-6) in den Klassenstufen 5, 6 bei 24% bzw. 31%, in den Klassenstufen 7, 8 bei 54% bzw. 65%. Die Unterschiede zwischen den Klassenstufen 7, 8 und den Klassenstufen 5, 6 sind paarweise statistisch hoch signifikant mit $0{,}22 < V < 0{,}42$. Mit Blick auf die Beweisversuche in der Kategorie Sw1-2 *zwischen induktiv und deduktiv* ist Lernenden in den Klassenstufen 7, 8 eher ein weitgehend korrektes Argumentgefüge gelungen als den Lernenden in den Klassenstufen 5, 6.

Bezogen auf die Beweisversuche deduktiver Schlussweise (Sw3-4) liegt der Anteil derjenigen mit einem weitgehend korrekten Argumentgefüge (KV4-6) in der Klassenstufe 5 bei 1%, in den

Klassenstufen 6, 7 bei etwa 15% und in der Klassenstufe 8 bei knapp 40%. Er ist in der Klassenstufe 8 statistisch hoch signifikant größer mit $0{,}26 < V < 0{,}54$ als in jeder der Klassenstufen 5 bis 7. In der Klassenstufe 8 ist Schülerinnen und Schülern, die deduktiv argumentiert haben, zu einem statistisch hoch signifikant größeren Anteil ein weitgehend korrektes Argumentgefüge gelungen als in den niedrigeren Klassenstufen, wenngleich auch in der Klassenstufe 8 in der Mehrheit der Beweisversuche mit deduktiver Schlussweise kein weitgehend korrektes Argumentgefüge vorkommt.



p-Werte von Signifikanztests und dazugehörige Effektstärken

	Sw1-2 vs. Sw3-4
Kl. 8	$p = .022^{N-1}, V = .256$
Kl. 7	$p < .001^{N-1}, V = .418$
Kl. 6	$p = .014^{N-1}, V = .193$
Kl. 5	$p < .001^{ex}, V = .381$

	Fokus: Sw1-2
Kl. 7 vs. Kl. 5	$p < .001^{N-1}, V = .298$
Kl. 7 vs. Kl. 6	$p = .005^{N-1}, V = .226$
Kl. 8 vs. Kl. 5	$p < .001^{N-1}, V = .419$
Kl. 8 vs. Kl. 6	$p < .001^{N-1}, V = .329$

	Fokus: Sw 3-4
Kl. 6 vs. Kl. 5	$p < .001^{ex}, V = .265$
Kl. 8 vs. Kl. 5	$p < .001^{ex}, V = .537$
Kl. 8 vs. Kl. 6	$p = .003^{N-1}, V = .265$
Kl. 8 vs. Kl. 7	$p = .003^{N-1}, V = .313$

Abbildung 88: Beweisversuche mit begründendem Ansatz: Existenz eines weitgehend korrekten Argumentgefüges in Abhängigkeit der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

8.2.1.4 Zusammenfassung

Der Abschnitt 8.2.1 fokussiert auf die Frage, inwieweit es Lernenden, die einen Beweisversuch mit begründendem Ansatz abgegeben haben, gelungen ist, die zu zeigende Aussage zu beweisen. Die Ergebnisse zeigen, dass in jeder der Klassenstufen 5 bis 8 ein großes Spektrum an Beweisversuchen vorkommt, das von solchen mit einem Hinweis auf einen grundlegenden Fehler und ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge bis hin zu Beweisversuchen mit einem weitgehend korrekten und vollständigen Argumentgefüge reicht. Einzelne Teilnehmende aus den Klassenstufen 6 bis 8 haben einen Beweis, d. h. ein korrektes und vollständiges, deduktives Argumentgefüge formuliert.

Die relativen Häufigkeitsverteilungen der Codierungen deuten zudem folgende Tendenzen an: Je höher die Klassenstufe, desto größer ist der Anteil der Beweisversuche, aus denen eine Passung zu der zu zeigenden Aussage hervorgeht und desto besser sind die Beweisversuche hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. In den Klassenstufen 5, 6 ist die Schlussweise bei der Mehrheit der Beweisversuche deduktiv, in der Klassenstufe 7 liegt sie mehrheitlich zwischen induktiv und deduktiv und in der Klassenstufe 8 ist sie in etwa der Hälfte der Beweisversuche deduktiv bzw. zwischen induktiv und deduktiv. Bezogen auf jede der Klassenstufen 5 bis 8 gilt, dass bei Beweisversuchen, in denen ein weitgehend korrektes Argumentgefüge vorkommt, die Schlussweise meistens zwischen induktiv und deduktiv liegt, während die Beweisversuche mit begründendem Ansatz, aber ohne weitgehend korrektes Argumentgefüge mehrheitlich deduktiv formuliert sind. In jeder der Klassenstufen enthält ein größerer Anteil der Beweisversuche, bei denen die Schlussweise zwischen induktiv und deduktiv liegt, ein weitgehend korrektes Argumentgefüge als es bei Beweisversuchen in der Kategorie deduktiv der Fall ist.

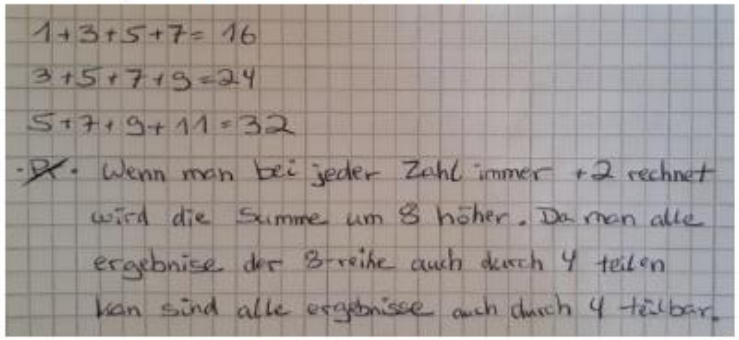
8.2.2 Beweisideen

Dieser Abschnitt beantwortet die Frage, welche Beweisideen die Lernenden gewählt haben. Dazu werden erstens das Spektrum an Beweisideen und zweitens die relativen Häufigkeiten der Beweisideen beschrieben. Die Ergebnisse geben aus einer fachinhaltlich-heuristischen Perspektive Einblicke in die Beweisversuche der Lernenden und lassen Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis erkennen (Abschn. 3.3.3 zum Kategoriensystem und zum Nutzen der Analyse der Beweisideen).

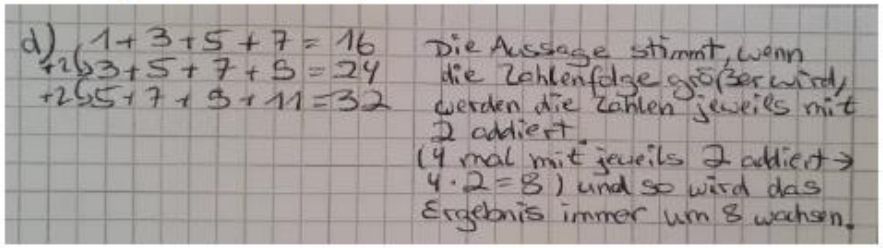
8.2.2.1 Das Spektrum an Beweisideen

Die Analyse der Beweisversuche mit begründendem Ansatz zeigt, dass in jeder der Klassenstufen 5 bis 8 eine große Vielfalt an Beweisideen vorkommt. Fast alle der unterschiedenen Beweisideen tauchen in jeder der Klassenstufen 5 bis 8 auf. Das heißt, in jeder der Klassenstufen gibt es Lernende, die z. B. über die sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante, über die Reste bei Division einzelner Summanden durch vier oder acht oder die Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen argumentiert haben. Die Abbildungen 89, 90 zeigen exemplarisch einzelne Beweisversuche von Teilnehmenden in den Klassenstufen 5 und 8. Die Bearbeitungen 8.2.2a-b basieren auf der Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Zudem wurden der Abstand zweier ungerader Zahlen und die Anzahl an Summanden in beiden Bearbeitungen berücksichtigt. In den Bearbeitungen 8.2.2c-d sind jeweils die Beweisideen 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*, 7 *Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen* und 11 *vier Zahlen* zu erkennen. In den Bearbeitungen 8.2.2e-f wird über die Parität von Summen ungerader Zahlen und eine nicht eindeutig kategorisierbare Beweisidee argumentiert. Die letzten Bearbeitungen 8.2.2g-h basieren auf der Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*. Diese Beweisidee kommt in jeder der Klassenstufen 6 bis 8, aber in keinem der analysierten Beweisversuche aus der Klassenstufe 5 vor. Die Beweisidee 2 bildet hinsichtlich des Spektrums an vorkommenden Beweisideen die einzige Ausnahme.

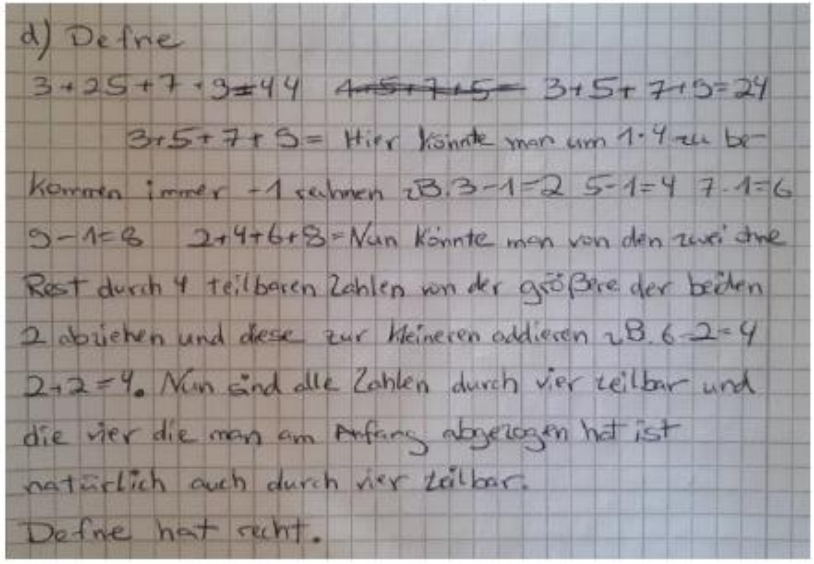
Bearbeitung 8.2.2a: Klassenstufe 5, weiblich (Code: 0512006)



Bearbeitung 8.2.2b: Klassenstufe 8, weiblich (Code: 0811702)



Bearbeitung 8.2.2c: Klassenstufe 5, männlich (Code: 0511206)



Bearbeitung 8.2.2d: Klassenstufe 8, weiblich (Code: 0820303)

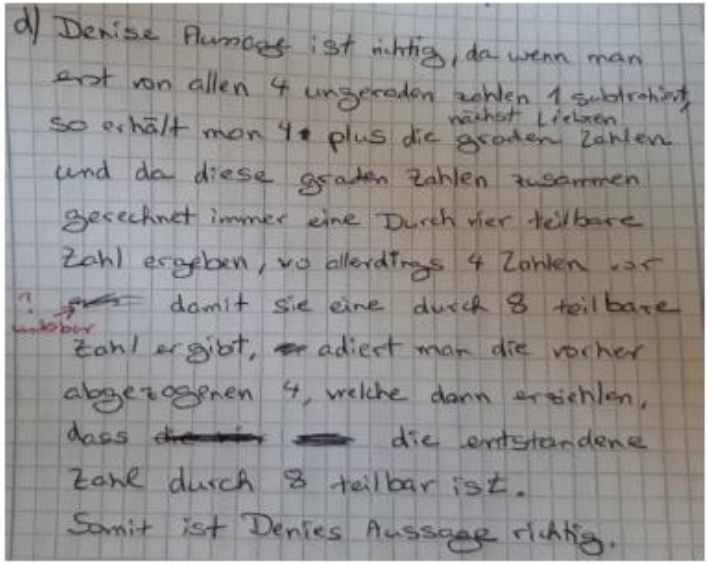
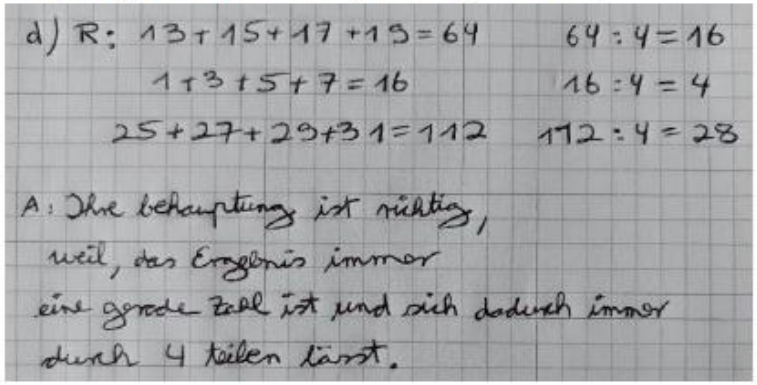
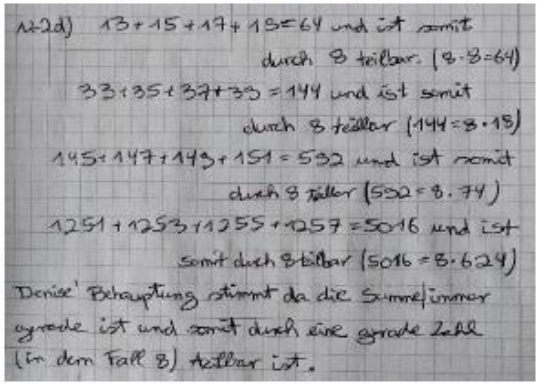


Abbildung 89: Beweisversuche mit begründendem Ansatz aus den Klassenstufen 5, 8 mit BI1,10,11 (oben) bzw. BI7,8,11 (unten)

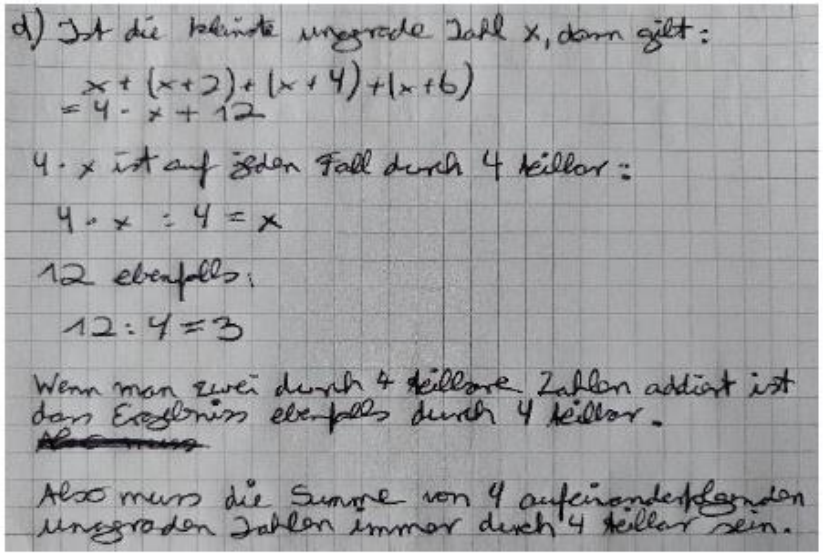
Bearbeitung 8.2.2e: Klassenstufe 5, weiblich (Code: 0521637)



Bearbeitung 8.2.2f: Klassenstufe 8, männlich (Code: 0812108)



Bearbeitung 8.2.2g: Klassenstufe 6, männlich (Code: 0610116)



Bearbeitung 8.2.2i: Klassenstufe 8, männlich (Code: 0810702)

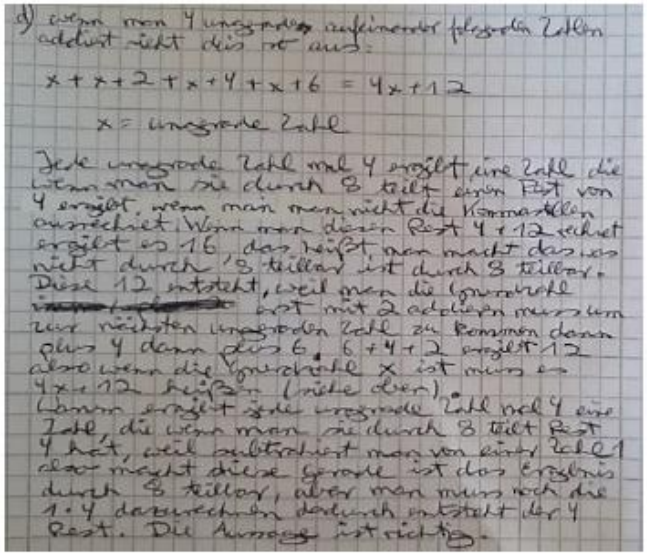


Abbildung 90: Beweisversuche aus den Klassenstufen 5, 8 mit BIp9, BI55 (oben) und Beweisversuche aus den Klassenstufen 6, 8 mit BIp2 (unten)

8.2.2.2 Relative Häufigkeitsverteilungen der Beweisideen

Die Analyse der relativen Häufigkeitsverteilungen dieser vielfältigen Beweisideen zeigt auf, dass einige Ideen besonders häufig vorkommen. Dabei scheinen insbesondere zwischen den Klassenstufen 5, 6 und den Klassenstufen 7, 8 Unterschiede zu bestehen.

Relative Häufigkeiten der einzelnen Beweisideen

Dies geht u. a. aus Abbildung 91 hervor. Diese stellt differenziert nach Klassenstufen dar, in einem wie großen Anteil der Beweisversuche mit begründendem Ansatz die einzelnen Beweisideen vorkommen. In Abbildung 92 sind ergänzende statistische Daten angegeben.

In den **Klassenstufen 5, 6** kommt die Beweisidee 9 *Parität von Summen ungerader oder gerader Zahlen* in mehr als der Hälfte der Beweisversuche mit begründendem Ansatz vor. Wissenselemente zur Parität von Summen ungerader oder gerader Zahlen scheinen den Lernenden gut vertraut zu sein. Die Verwendung solcher Wissenselemente erschien vielen Lernenden eventuell als naheliegend, weil sich die zu zeigende Aussage auf ungerade Zahlen bezieht.

Neben der Beweisidee 9 ist die Beweisidee 11 *vier Zahlen* mit je 40% bis 45% besonders häufig in Beweisversuchen von Lernenden in den Klassenstufen 5, 6 zu finden. Dass es sich um vier Zahlen handelt, ist in der Aufgabenstellung vorgegeben und erschien einigen Lernenden vielleicht durch die Übereinstimmung der Anzahl an Summanden und des Teilers als entscheidend (Abschn. 6.2.2). Zudem ist die Anzahl an Summanden bei einer Argumentation über mehrere andere Beweisideen, wie etwa den Ideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* und 6 *Teilbarkeit der Summen zweier Zahlenpaare*, als ergänzende inhaltliche Überlegung von Bedeutung, was zu der häufigen Verwendung der Beweisidee 11 *vier Zahlen* beigetragen haben kann.

Weiterhin kommt die Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* in knapp einem Fünftel der Beweisversuche mit begründendem Ansatz aus der fünften Klassenstufe und knapp einem Viertel derjenigen aus der sechsten Klassenstufen vor. Die Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* ist somit die ausbaubare Beweisidee, die in den Klassenstufen 5, 6 am häufigsten zu finden ist. Ein möglicher Erklärungsansatz besteht darin, dass die Beweisidee 6 vielleicht oftmals ausgehend von Überlegungen zur Parität von Summen ungerader Zahlen, die in sehr vielen Beweisversuchen vorkommen, gefunden wurde. Die ausbaubare und zugleich zentrale Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, die sich nach den Überlegungen im Rahmen der Aufgabenkonstruktion besonders gut für weitgehend korrektes und vollständiges Argumentgefüge eignet und für Lernende aller Klassenstufen gut zugänglich erscheint (Abschn. 6.2.2), kommt in den Klassenstufen 5, 6 in etwa 8% der Beweisversuche mit begründendem Ansatz vor. Alle anderen zentralen Beweisideen 2 bis 5 sind jeweils in maximal 5% der Beweisversuche mit begründendem Ansatz zu finden.

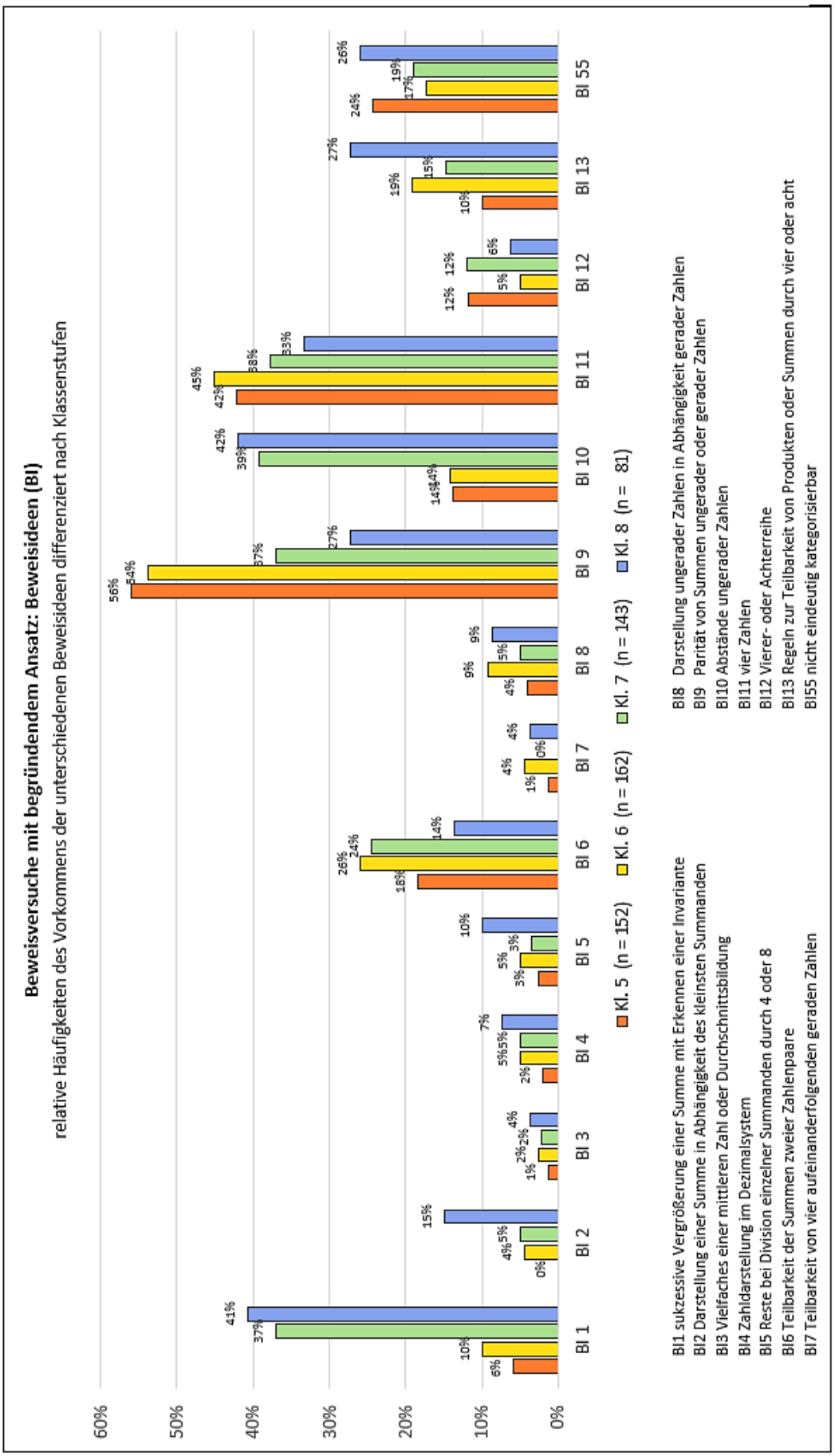


Abbildung 91: Relative Häufigkeiten des Vorkommens der unterschiedenen Beweisideen in den Beweisversuchen mit begründendem Ansatz differenziert nach Klassenstufen

Beweisversuche mit begründendem Ansatz: Beweisideen

relative Häufigkeiten des Vorkommens der einzelnen Beweisideen in Beweisversuchen mit begründendem Ansatz

p-Werte von Signifikanztests und dazugehörige Effektstärken

	Beweisidee 9	Beweisidee 1	Beweisidee 10
Kl. 7 vs. Kl. 5	$p = .001^{N-1}, V = .189$	$p < .001^{N-1}, V = .382$	$p < .001^{N-1}, V = .288$
Kl. 7 vs. Kl. 6	$p = .004^{N-1}, V = .167$	$p < .001^{N-1}, V = .324$	$p < .001^{N-1}, V = .284$
Kl. 8 vs. Kl. 5	$p < .001^{N-1}, V = .275$	$p < .001^{N-1}, V = .431$	$p < .001^{N-1}, V = .316$
Kl. 8 vs. Kl. 6	$p < .001^{N-1}, V = .252$	$p < .001^{N-1}, V = .363$	$p < .001^{N-1}, V = .309$
	Beweisidee 2	**Beweisidee 5**	**Beweisidee 13**
Kl. 8 vs. Kl. 5	$p < .001^{ex}, V = .319$	$p = .027^{ex}, V = .156$	$p = .001^{N-1}, V = .225$
Kl. 8 vs. Kl. 6	$p = .004^{N-1}, V = .184$	$p = .144^{N-1}, V = .093$	$p = .154^{N-1}, V = .092$
Kl. 8 vs. Kl. 7	$p = .011^{N-1}, V = .171$	$p = .072^{ex}, V = .131$	$p = .023^{N-1}, V = .152$

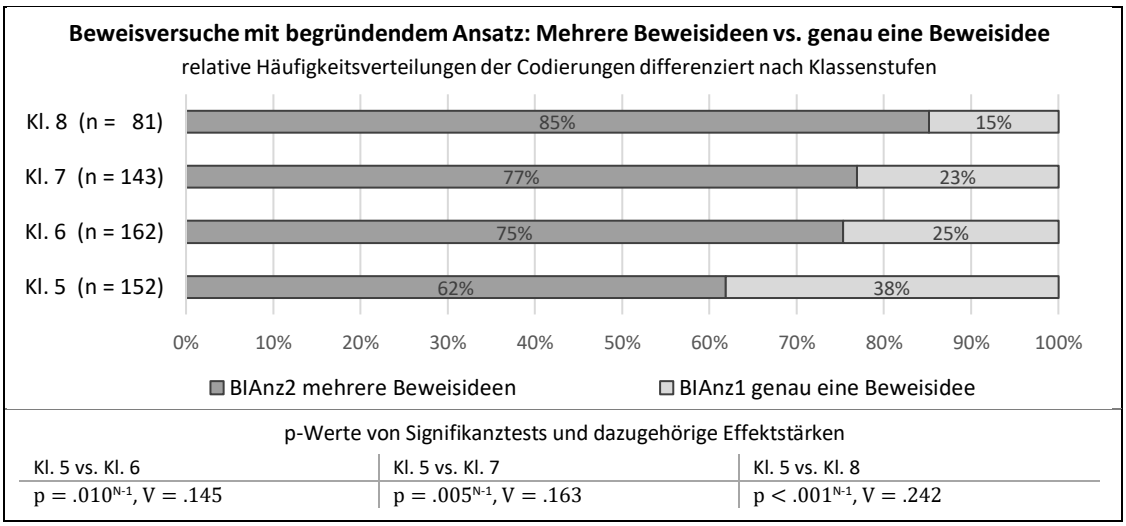
Abbildung 92: Relative Häufigkeiten des Vorkommens der unterschiedenen Beweisideen in den Beweisversuchen mit begründendem Ansatz differenziert nach Klassenstufen: p-Werte von Signifikanztests und dazugehörige Effektstärken

In den **Klassenstufen 7, 8** sind die vier Beweisideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, 9 *Parität von Summen ungerader oder gerader Zahlen*, 10 *Abstände ungerader Zahlen* und 11 *vier Zahlen* sehr häufig. Der Anteil an Beweisversuchen, in denen die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* vorkommt, ist in den Klassenstufe 7, 8 mit ca. 40% im Vergleich zu 5% bis 10% in den Klassenstufen 5, 6 statistisch hoch signifikant größer mit $0{,}32 < V < 0{,}44$. Eventuell haben Teilnehmende aus den Klassenstufen 7, 8 die Beweisidee 1 durch die leicht andere Aufgabenstellung häufiger gefunden (Abschn. 6.2.2). Zumindest ein weitgehend korrektes und vollständiges Argumentgefüge basierend auf der Beweisidee 1 impliziert auch die Verwendung der Beweisideen 10 *Abstände ungerader Zahlen* und 11 *vier Zahlen*. Der Anteil der Beweisversuche mit der Beweisidee 10 an allen Beweisversuchen mit begründendem Ansatz ist ebenfalls in den Klassenstufen 7, 8 jeweils statistisch hoch signifikant größer als in den Klassenstufen 5, 6. Demgegenüber ist der Anteil der Beweisversuche, in denen die Beweisidee 9 *Parität von Summen ungerader oder gerader Zahlen* vorkommt, an allen Beweisversuchen mit begründendem Ansatz in den Klassenstufen 7, 8 jeweils statistisch hoch signifikant kleiner als in jeder der Klassenstufen 5, 6. Vielleicht erscheinen Lernenden Überlegungen zur Parität von Summen ungerader oder gerader Zahlen mit Blick auf den in den Klassenstufen 5, 6 geforderten Beweis der Teilbarkeit durch vier als naheliegender und zielführender als mit Blick auf den in den Klassenstufen 7, 8 geforderte Beweis der Teilbarkeit durch acht.

In der **Klassenstufe 8** im Speziellen sind zudem die Anteile der Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*, 5 *Reste bei Division einzelner Summanden durch vier oder acht* und 13 *Regeln zur Teilbarkeit von Produkten oder Summen durch vier oder acht* auffällig. Die Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* kommt in 13% der Beweisversuche von Achtklässlerinnen und Achtklässlern vor, und damit in einem statistisch signifikant größeren Anteil als in jeder der Klassenstufen 5 bis 7, mit $0{,}17 < V < 0{,}32$. Eventuell wurde die Verwendung der Beweisidee 2 von Achtklässlern und Achtklässlern durch die erfolgte Einführung in den Umgang mit Buchstabenvariablen begünstigt (Abschn. 8.2.4, 8.2.5.2). Die Beweisidee 5 *Reste bei Division einzelner Summanden durch vier oder acht* kommt mit 10% in der Klassenstufe 8 im Vergleich zu den Klassenstufen 5 bis 7 in einem mehr als doppelt so großen Anteil der Beweisversuche mit begründendem Ansatz vor.

Der Unterschied zur Klassenstufe 5 ist statistisch signifikant mit geringer Effektstärke. Ein möglicher Erklärungsansatz lautet, dass die häufigere Verwendung der Beweisidee 5 *Reste bei Division einzelner Summanden durch vier oder acht* mit einer vorhergehenden mathematischen Förderung eines eventuell vergleichsweise großen Anteils der Teilnehmenden in der Klassenstufe 8 zusammenhängt. Ein weiterer Erklärungsansatz besteht darin, dass das Rechnen mit Resten eventuell in den Olympiadeaufgaben für die Klassenstufe 8 häufig hilfreich ist und die Lernenden vielleicht frühere Aufgaben zur Vorbereitung auf die Teilnahme genutzt haben. Mit Blick auf die Beweisidee 13 *Regeln zur Teilbarkeit von Produkten oder Summen durch vier oder acht* ist es z. B. denkbar, dass eine Tendenz zu einer im Vergleich detaillierteren Explizierung von Überlegungen zum etwas häufigeren Vorkommen der Beweisidee 13 in der Klassenstufe 8 im Vergleich zu den Klassenstufen 5 bis 7 beigetragen hat. Die Beweisidee 13 *Regeln zur Teilbarkeit von Produkten oder Summen durch vier oder acht* ist in der Klassenstufe 8 in 27% der Beweisversuche zu identifizieren, in den Klassenstufen 5 bis 7 jeweils in 10% bis 20% der Beweisversuche, wobei die Unterschiede zwischen den Klassenstufen 8, 5 sowie 8, 7 statistisch signifikant sind.

Vor allem bei dem Ergebnis, dass die Beweisideen 9 bis 13, 55 teils in beachtlichen Anteilen der Beweisversuche mit begründendem Ansatz zu finden sind, gilt es die **Möglichkeit der Verwendung mehrerer Beweisideen innerhalb eines Beweisversuchs** zu bedenken. In jeder der Klassenstufen 5 bis 8 kommen in weit über der Hälfte der Beweisversuche mit begründendem Ansatz mehrere Beweisideen vor. Der Anteil an Beweisversuchen mit mehreren Beweisideen liegt in der Klassenstufe 5 bei gut 60%, in den Klassenstufen 6, 7 bei ca. 75% und in der Klassenstufe 8 bei 85%. Dabei sind die Unterschiede zwischen der Klassenstufe 5 und den Klassenstufen 6, 7, 8 statistisch signifikant mit $0{,}14 < V < 0{,}25$ (Abb. 93). Es deutet sich an, dass die Beweisversuche mit begründendem Ansatz von Lernenden in den Klassenstufen 6, 7, 8 teils inhaltlich komplexer sind als einige der Beweisversuche mit begründendem Ansatz aus der Klassenstufe 5. Wenn in einem Beweisversuch mehrere Beweisideen vorkommen, wurden diese je Klassenstufe in 65% bis 80% der Fälle klar als ein einziges zusammenhängendes Argumentgefüge ausgeführt.



p-Werte von Signifikanztests und dazugehörige Effektstärken

Kl. 5 vs. Kl. 6	Kl. 5 vs. Kl. 7	Kl. 5 vs. Kl. 8
$p = .010^{N-1}, V = .145$	$p = .005^{N-1}, V = .163$	$p < .001^{N-1}, V = .242$

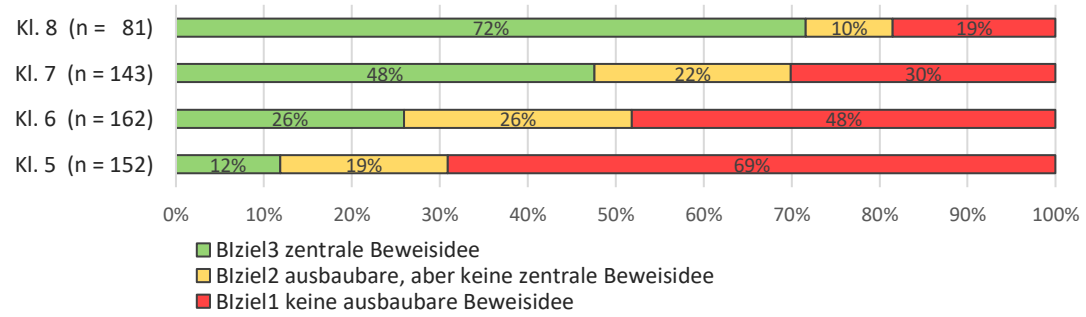
Abbildung 93: Beweisversuche mit begründendem Ansatz: Mehrere Beweisideen vs. genau eine Beweisidee. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests, dazugehörige Effektstärken

Relative Häufigkeitsverteilungen der zielführendsten Beweisidee und der primären Beweisidee

Um die Möglichkeit mehrerer Beweisideen innerhalb eines Beweisversuchs als Einflussfaktor zu relativieren und stärker die Qualität der Beweisideen, die in den Beweisversuchen mit begründendem Ansatz vorkommen, zu fokussieren, wurde die zielführendste Beweisidee kategorisiert und die primäre Beweisidee codiert (Abschn. 3.3.3).

Zur Einordnung der zielführendsten Beweisidee wurden die Kategorien BIziel3 *zentrale Beweisidee*, BIziel2 *ausbaubare, aber keine zentrale Beweisidee* und BIziel1 *keine ausbaubare Beweisidee* gebildet. Ein Beweisversuch enthält eine zentrale Beweisidee, wenn eine der Beweisideen eins bis fünf vorkommt. Die Beweisideen eins bis fünf werden als zentrale Beweisideen bezeichnet, weil ein Beweis der vorgegebenen Aussage i. d. R. eine dieser Ideen erfordert. Weiterhin werden alle Beweisideen eins bis acht als ausbaubare Beweisideen bezeichnet, weil jede von ihnen den Rahmen oder den wesentlichen Ausgangspunkt eines korrekten und vollständigen Argumentgefüges bilden kann. Wenn in einem Beweisversuch keine der Beweisideen eins bis acht vorkommt, enthält er keine ausbaubare Beweisidee (Abschn. 3.3.3).

In Abbildung 94 sind die Ergebnisse der Analyse der **zielführendsten Beweisidee** dargestellt. Sie verdeutlichen erstens, dass in jeder der Klassenstufen 5 bis 8 Beweisversuche mit einer zentralen Beweisidee (BIziel3), Beweisversuche mit einer ausbaubaren, aber keiner zentralen Beweisidee (BIziel2) und auch Beweisversuche ohne eine ausbaubare Beweisidee (BIziel1) mit jeweils über 10% der Zuordnungen relevante Anteile ausmachen. Zugleich ist nach dem Mann-Whitney-U-Test je höher die Klassenstufe, die Wahrscheinlichkeit, eine Zuordnung zu einer höheren der drei Kategorien zu erhalten, statistisch signifikant größer. Insbesondere ist je höher die Klassenstufe, der Anteil der Beweisversuche in der Kategorie BIziel3 *zentrale Beweisidee* desto größer.



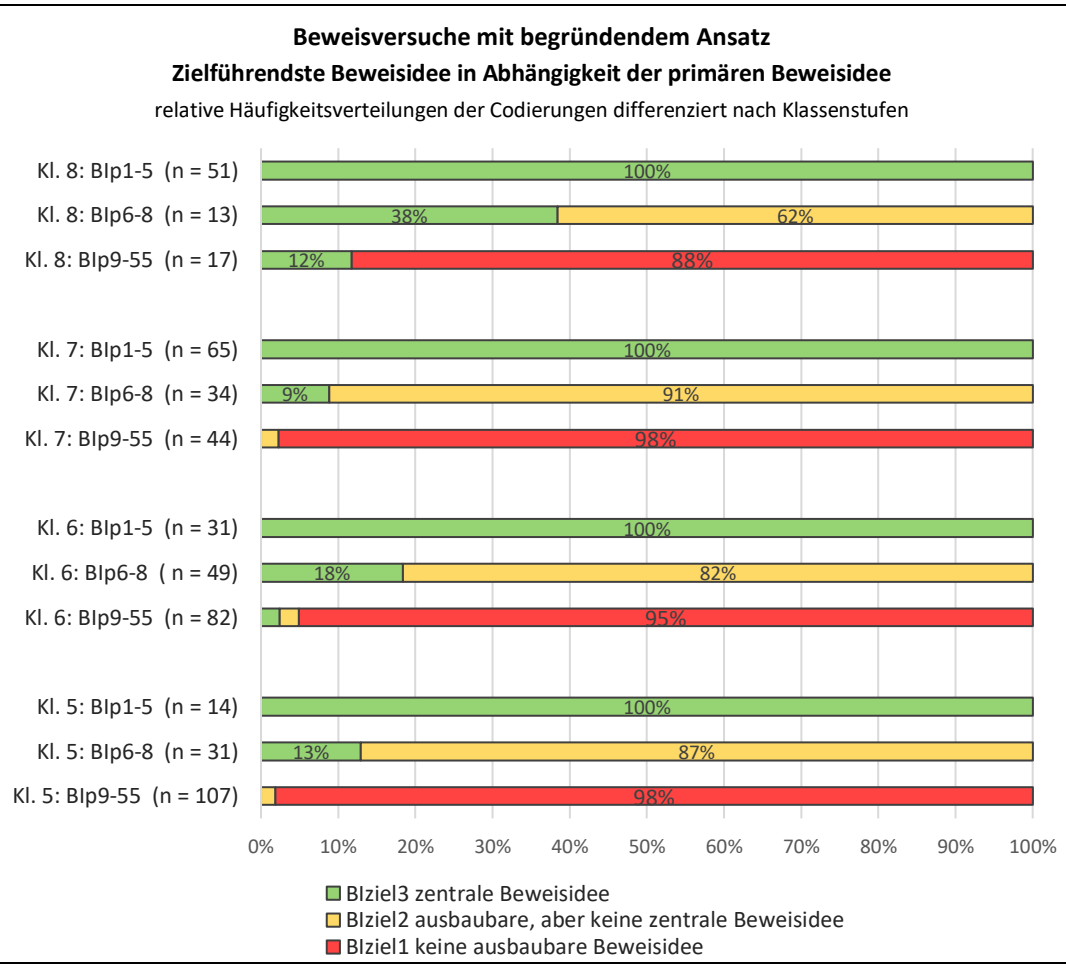
p-Werte von Signifikanztests und dazugehörige Effektstärken

	alle einzelnen Kategorien	BIziel 3 vs. ¬BIziel3	BIziel 2 vs. ¬BIziel2	BIziel 1 vs. ¬BIziel1
Kl. 6 vs. Kl. 5	$p < .001^{U}$, $r = .223$	$p = .002^{N-1}$, $V = .179$	$p = .148^{N-1}$, $V = .082$	$p < .001^{N-1}$, $V = .212$
Kl. 7 vs. Kl. 5	$p < .001^{U}$, $r = .428$	$p < .001^{N-1}$, $V = .393$	$p = .485^{N-1}$, $V = .041$	$p < .001^{N-1}$, $V = .390$
Kl. 7 vs. Kl. 6	$p < .001^{U}$, $r = .227$	$p < .001^{N-1}$, $V = .225$	$p = .471^{N-1}$, $V = .041$	$p = .001^{N-1}$, $V = .184$
Kl. 8 vs. Kl. 5	$p < .001^{U}$, $r = .571$	$p < .001^{N-1}$, $V = .607$	$p = .068^{N-1}$, $V = .120$	$p < .001^{N-1}$, $V = .482$
Kl. 8 vs. Kl. 6	$p < .001^{U}$, $r = .400$	$p < .001^{N-1}$, $V = .438$	$p = .004^{N-1}$, $V = .187$	$p < .001^{N-1}$, $V = .287$
Kl. 8 vs. Kl. 7	$p = .001^{U}$, $r = .213$	$p = .001^{N-1}$, $V = .233$	$p = .019^{N-1}$, $V = .157$	$p = .058^{N-1}$, $V = .127$

Abbildung 94: Beweisversuche mit begründendem Ansatz: Zielführendsten Beweisidee. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Unter dem Begriff „**primäre Beweisidee**“ wird weiterhin betrachtet, welche Beweisidee den Rahmen oder den wesentlichen Ausgangspunkt des vorliegenden Argumentgefüges bildet. Die primäre Beweisidee kann mit der zielführendsten Beweisidee übereinstimmen, muss es aber nicht immer. Wenn in einem Beweisversuch eine der zentralen Beweisideen vorkommt (BIziel3), kann prinzipiell jede der unterschiedenen Beweisideen die primäre Beweisidee bilden (Abschn. 3.3.3). Die empirischen Ergebnisse deuten trotzdem an, dass die primäre Beweisidee oftmals der zielführendsten Beweisidee entspricht. Beispielsweise ist in jeder der Klassenstufen 5 bis 8 in über 70% der Beweisversuche mit einer zentralen Beweisidee (BIziel3) auch eine zentrale Beweisidee die primäre Beweisidee.

Die Abbildung 95 zeigt andersherum die Ergebnisse der Analyse der zielführendsten Beweisidee in Abhängigkeit primärer Beweisideen. Wenn in einem Beweisversuch mit begründendem Ansatz eine der Beweisideen eins bis fünf die primäre Beweisidee ist (BIp1-5), dann kommt eine der zentralen Beweisideen vor (BIziel3). Im Falle einer der primären Beweisideen sechs



p-Werte von Signifikanztests und dazugehörige Effektstärken

Kl. 5	$p < .001^{ex}, V = .894$
Kl. 6	$p < .001^{X^2}, V = .868$
Kl. 7	$p < .001^{X^2}, V = 954$
Kl. 8	$p < .001^{ex}, V = .841$

	Fokus: BIp6-8
Kl. 8 vs. Kl. 5	$p = .020^{ex}, V = .207$
Kl. 8 vs. Kl. 6	$p = .339^{ex}, V = .074$
Kl. 8 vs. Kl. 7	$p = .575^{ex}, V = .050$

Abbildung 95: Beweisversuche mit begründendem Ansatz: Zielführendste Beweisidee in Abhängigkeit der primären Beweisidee. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

bis acht (BIp6-8), enthält der Beweisversuch eine ausbaubare Beweisidee (BIziel2-3). Die Ergebnisse der Analyse der Beweisversuche mit begründendem Ansatz und BIp6-8 verdeutlichen, dass in der Klassenstufe 8 in knapp 40% der Fälle eine der zentralen Beweisideen als unter- oder nachgeordnete Beweisidee ausgeführt ist, in den Klassenstufen 5 bis 7 jeweils in weniger als 20% der Fälle. Wenn in einem Beweisversuch mit begründendem Ansatz eine der Beweisideen 9 bis 13 oder eine nicht eindeutig kategorisierbare Beweisidee die primäre Beweisidee ist (BIp9-55), kann trotzdem eine zentrale oder zumindest ausbaubare Beweisidee ausgeführt sein (BIziel1-3). Aus der Analyse der Beweisversuche geht allerdings hervor, dass je Klassenstufe in über 85% dieser Beweisversuche mit BIp9-55 keine ausbaubare Beweisidee vorkommt. Insgesamt ist festzuhalten, dass in den Beweisversuchen mit begründendem Ansatz aus den Klassenstufen 5 bis 8 nur selten eine noch zielführendere Beweisidee als die primäre Beweisidee vorkommt. Diese Ergebnisse implizieren zum Beispiel, dass die Argumentgefüge in Beweisversuchen mit einer ausbaubaren, aber keiner zentralen primären Beweisidee (BIp6-8) oftmals eine inhaltliche Lücke enthalten (Abschn. 8.2.6.3).

Dadurch, dass die zielführendste Beweisidee und die primäre Beweisidee in Beweisversuchen mit begründendem Ansatz mit Blick auf jede der Klassenstufen in einem statistisch hoch signifikanten Zusammenhang stehen (Abb. 95), stützen und komplettieren die Ergebnisse der Analyse der primären Beweisidee im Wesentlichen bereits dargelegte Erkenntnisse zu Beweisideen in Beweisversuchen mit begründendem Ansatz.

Die Analyse der primären Beweisidee verdeutlicht, dass in allen Klassenstufen fast jede der unterschiedenen Beweisideen als primäre Beweisidee verwendet wurde. Zugleich ist zu sehen, dass je Klassenstufe 60% bis 80% der Beweisversuche mit begründendem Ansatz auf einer von drei der vierzehn unterschiedenen Beweisideen basieren. In Tabelle 13 sind differenziert nach Klassenstufen alle primären Beweisideen aufgeführt, die mindestens ein Zehntel der Beweisversuche mit begründendem Ansatz ausmachen. In den Klassenstufen 5 und 6 handelt es sich um die primären Beweisideen 9 *Parität von Summen ungerader oder gerader Zahlen*, 11 *vier Zahlen* und 6 *Teilbarkeit der Summen zweier Zahlenpaare*. Das heißt, die Beweisideen, die in den Klassenstufen 5 und 6 besonders oft vorkommen, sind zugleich die am häufigsten auftauchenden primären Beweisideen. In den Klassenstufen 5 und 6 beruht die große Mehrheit der Beweisversuche mit begründendem Ansatz auf einer nicht ausbaubaren Beweisidee. In den Klassenstufen 7 und 8 bilden die nicht ausbaubaren Beweisideen 10, 11, 13, die jeweils in großen Anteilen der Beweisversuche mit begründendem Ansatz vorkommen, in jeweils klar unter 10% der Beweisversuche mit begründendem Ansatz die primäre Beweisidee. In der Klassenstufe 7 ist in rund 45%, in der Klassenstufe 8 in 63% der Beweisversuche mit begründendem Ansatz eine der zentralen Beweisideen die primäre Beweisidee, wobei die primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* mit Abstand am häufigsten vorkommt.

	Kl. 5 (n = 152)		Kl. 6 (n = 162)		Kl. 7 (n = 143)		Kl. 8 (n = 81)	
größter Anteil	BIp9	38%	BIp9	25%	BIp1	34%	BIp1	40%
zweitgrößter Anteil	BIp11	18%	BIp6	19%	BIp6	22%	BIp2	10%
drittgrößter Anteil	BIp6	18%	BIp11	17%	BIp9	10%	BIp6	10%

Tabelle 13: Beweisversuche mit begründendem Ansatz: Häufigste primäre Beweisideen differenziert nach Klassenstufen

8.2.2.3 Zusammenfassung

Zusammenfassend zeigt die Analyse der Beweisideen bei Beweisversuchen mit begründendem Ansatz, dass in jeder der Klassenstufen 5 bis 8 in den meisten Beweisversuchen mehrere Beweisideen ausgeführt sind und dass in jeder der Klassenstufen eine große Vielfalt an Beweisideen vorkommt.

Dabei sind in den Klassenstufen 5, 6 vor allem die Beweisideen 9 *Parität von Summen ungerader oder gerader Zahlen* und 11 *vier Zahlen*, in den Klassenstufen 7, 8 vor allem die Beweisideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* und 10 *Abstände ungerader Zahlen* besonders häufig zu finden. Die Beweisversuche mit begründendem Ansatz von Achtklässlerinnen und Achtklässlern zeichnet zudem eine vergleichsweise häufige Verwendung einer der zentralen Beweisideen 2 bis 5 aus. Je höher die Klassenstufe, desto größer ist der Gesamtanteil der Beweisversuche mit einer der zentralen Beweisideen 1 bis 5.

Die primäre Beweisidee ist in der großen Mehrheit der Beweisversuche mit begründendem Ansatz aus den Klassenstufen 5 und 6 eine nicht ausbaubare Beweisidee. In den Klassenstufen 7 und 8 ist in etwa der Hälfte der Beweisversuche mit begründendem Ansatz eine der zentralen Beweisideen die primäre Beweisidee. In allen Klassenstufen 5 bis 8 enthalten die Beweisversuche mit begründendem Ansatz nur selten eine noch zielführendere Beweisidee als die primäre Beweisidee.

8.2.3 Beispiele

Dieser Abschnitt gibt Einblicke in die Existenz, Eigenschaften und Funktion von Beispielen in Beweisversuchen mit begründendem Ansatz.

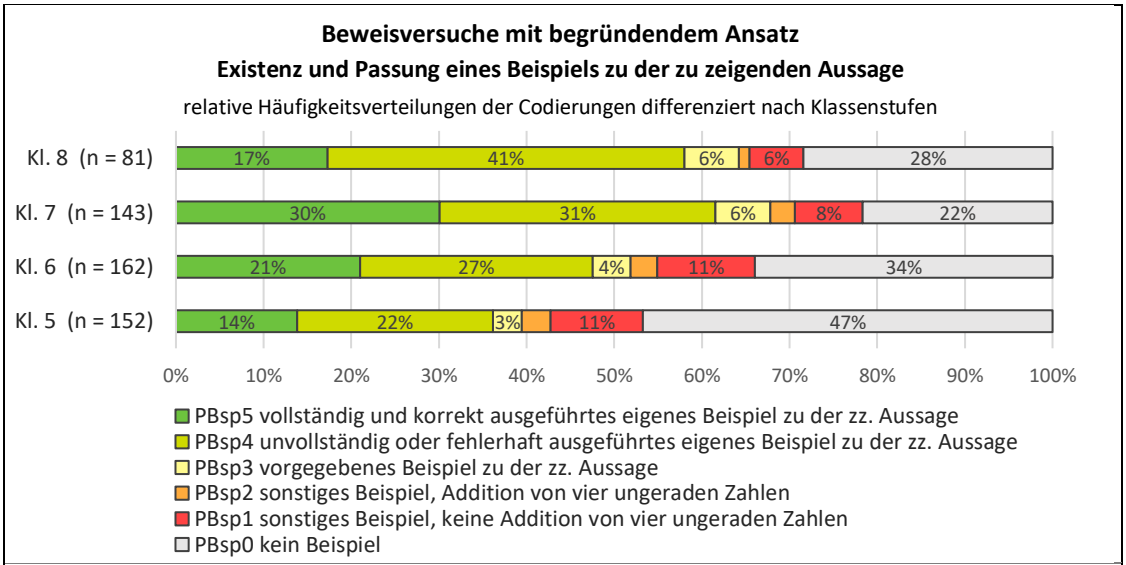
Zuerst wird dargelegt, ob in den Beweisversuchen mit begründendem Ansatz ein Beispiel vorkommt, und insbesondere eines, das die zu zeigende Aussage bestätigt. Dies ist mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage von Bedeutung, insbesondere wenn in Beweisversuchen mit begründendem Ansatz kein weiterentwickelbares Argumentgefüge vorkommt. Den Lernenden ist es ggf. immerhin gelungen, die Aussage für einen Spezialfall zu bestätigen (Abschn. 8.2.5, 8.5.1 bzgl. des Nachweises der Gültigkeit).

Zweitens werden die Beweisversuche mit begründendem Ansatz, die ein Beispiel enthalten, fokussiert und Eigenschaften der angegebenen Beispiele präsentiert. Die Ergebnisse beschreiben, in welcher Form Beispiele in Beweisversuchen mit begründendem Ansatz besonders häufig oder eher selten vorkommen. Dieser Abschnitt trägt zur Beantwortung der Frage nach Charakteristika der Beweisversuche in Bezug auf Beispiele bei.

Abschließend wird anhand einzelner Bearbeitungen aufgezeigt, welche Funktion die Beispiele in den Beweisversuchen aus der Perspektive von Lesenden haben und welche Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis der zu zeigenden Aussage durch die angegebenen Beispiele zu erkennen sind. Dieser Abschnitt ist mit Blick auf die Beantwortung der Frage nach solchen Anknüpfungspunkten von Bedeutung.

8.2.3.1 Existenz und Passung eines Beispiels zu der zu zeigenden Aussage

Der vorliegende Abschnitt beschäftigt sich mit der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage. Es wird dargelegt, inwieweit die Schülerinnen und Schüler ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage angegeben und somit diese Aussage selbstständig für einen Spezialfall bestätigt haben.



p-Werte von Signifikanztests und dazugehörige Effektstärken

	PBsp5 vs. ¬PBsp5
Kl. 7 vs. Kl. 5	$p = .001^{N-1}, V = .197$
Kl. 7 vs. Kl. 8	$p = .035^{N-1}, V = .141$

	PBsp4 vs. ¬PBsp4
Kl. 8 vs. Kl. 5	$p = .003^{N-1}, V = .193$
Kl. 8 vs. Kl. 6	$p = .025^{N-1}, V = .144$

	PBsp4-5 vs. ¬PBsp4-5
Kl. 5 vs. Kl. 6	$p = .042^{N-1}, V = .115$
Kl. 5 vs. Kl. 7	$p < .001^{N-1}, V = .254$
Kl. 5 vs. Kl. 8	$p = .001^{N-1}, V = .210$
Kl. 6 vs. Kl. 7	$p = .014^{N-1}, V = .140$
Kl. 6 vs. Kl. 8	$p = .124^{N-1}, V = .099$

	PBsp1-5 vs. ¬PBsp1-5
Kl. 6 vs. Kl. 5	$p = .021^{N-1}, V = .130$
Kl. 7 vs. Kl. 5	$p < .001^{N-1}, V = .263$
Kl. 7 vs. Kl. 6	$p = .018^{N-1}, V = .136$
Kl. 8 vs. Kl. 5	$p = .007^{N-1}, V = .178$
Kl. 8 vs. Kl. 6	$p = .383^{N-1}, V = .056$
Kl. 8 vs. Kl. 7	$p = .260^{N-1}, V = .075$

Abbildung 96: Beweisversuchen mit begründendem Ansatz: Existenz und Passung eines Beispiels zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Entsprechend Abbildung 96 kommt je Klassenstufe in knapp 15% bis 30% der Beweisversuche mit begründendem Ansatz ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage vor (**PBsp5**). Der Anteil an Beweisversuchen mit einem unvollständig oder fehlerhaft ausgeführten eigenen Beispiel zu der zu zeigenden Aussage (**PBsp4**) liegt je Klassenstufe zwischen 20% und 45%. Solche Beweisversuche sind insbesondere in der Klassenstufe 8 mit über 40% der Codierungen besonders häufig. Sie machen einen statistisch signifikant größeren Anteil an allen Beweisversuchen mit begründendem Ansatz aus als in den Klassenstufen 5, 6. Eventuell hat insbesondere in der Klassenstufe 8 ein relativ großer Anteil der Lernenden ein Beispiel betrachtet, aber unvollständig aufgeschrieben, weil ihnen z. B. die Teilbarkeit einer Summe wie 48 durch acht als offensichtlich erschien.

Wenn man die Kategorien PBsp5 und PBsp4 zusammenfasst, ist zu sehen, dass in der Klassenstufe 5 mit 36% gut ein Drittel der Beweisversuche mit begründendem Ansatz ein eigenes Beispiel zu der zu zeigenden Aussage enthalten. Dieser Anteil der Kategorien **PBsp4-5** ist in der Klassenstufe 5 statistisch signifikant kleiner mit $0{,}11 < V < 0{,}25$ als in jeder der Klassenstufen

6 bis 8. Eigene Beispiele zu der zu zeigenden Aussage sind also in der Klassenstufe 5 bezogen auf die Beweisversuche mit begründendem Ansatz vergleichsweise selten.

Neben den Beweisversuchen mit begründendem Ansatz, in denen ein eigenes Beispiel zu der zu zeigenden Aussage vorkommt, gibt es solche, die das in der Aufgabenstellung vorgegebene Beispiel als einziges Beispiel zu der zu zeigenden Aussage oder ein sonstiges Beispiel enthalten (**PBsp1-3**). Der relative Anteil dieser Beweisversuche liegt je Klassenstufe bei 10% bis 20%.

Insgesamt kommt je Klassenstufe in 50% bis 80%, d. h. in mindestens der Hälfte der Beweisversuche mit begründendem Ansatz, ein Beispiel vor (**PBsp1-5**).[133] Bezogen auf die Klassenstufen 5 bis 7 ist der Anteil der Beweisversuche mit begründendem Ansatz, in denen ein Beispiel vorkommt, je höher die Klassenstufe, desto größer (Abb. 96).

Zusammenfassend betrachtet scheint das Vorkommen eines Beispiels in Beweisversuchen mit begründendem Ansatz durchaus normal zu sein. Je Klassenstufe enthält mindestens die Hälfte der Beweisversuche mit begründendem Ansatz ein Beispiel, in der Klassenstufe 5 ist der Anteil am kleinsten. Je Klassenstufe wurde in ca. 15% bis 30% der Beweisversuche mit begründendem Ansatz ein korrekt und vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage angegeben und somit diese Aussage selbstständig für einen Spezialfall bestätigt. In vielen anderen Beweisversuchen mit begründendem Ansatz kommt ein eigenes, jedoch unvollständig oder fehlerhaft ausgeführtes Beispiel zu der zu zeigenden Aussage vor.

8.2.3.2 Eigenschaften der angegebenen Beispiele

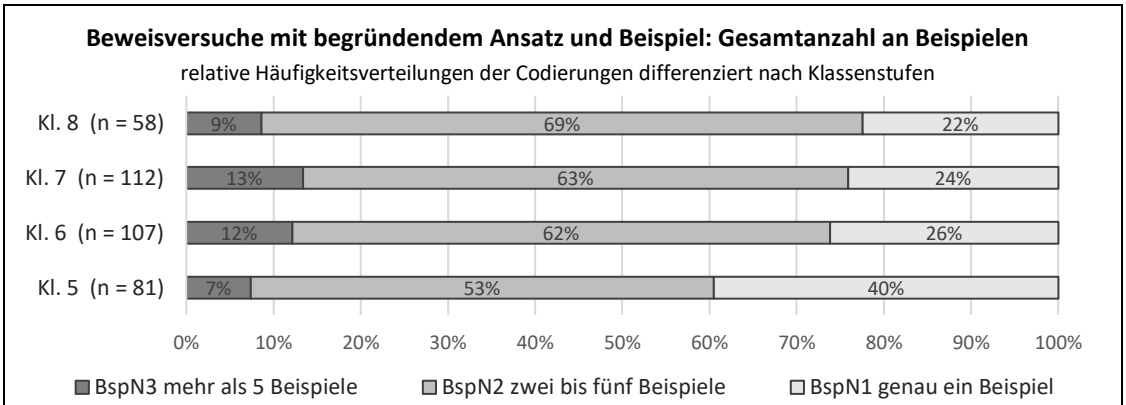
Im vorhergehenden Abschnitt 8.2.3.1 ist deutlich geworden, dass je Klassenstufe in 50% bis 80% der Beweisversuche mit begründendem Ansatz mindestens ein Beispiel vorkommt. Diese Beweisversuche mit begründendem Ansatz und Beispiel werden nun fokussiert und Eigenschaften der angegebenen Beispiele präsentiert. Die Ergebnisse geben einen Einblick, in welcher Form Beispiele in Beweisversuchen mit begründendem Ansatz und Beispiel besonders häufig oder eher selten vorkommen.[134]

Die Abbildung 97 zeigt die Ergebnisse der Analyse der **Gesamtanzahl an Beispielen**, die Beweisversuche mit begründendem Ansatz und Beispiel enthalten. Wenn in einem Beweisversuch mit begründendem Ansatz mindestens ein Beispiel angegeben ist, dann kommen meistens zwei bis fünf Beispiele vor. Der Anteil der Kategorie $\mathrm{BspN2}$ *zwei bis fünf Beispiele* liegt in allen Klassenstufen zwischen 50% und 70%. Beweisversuche mit mehr als fünf Beispielen ($\mathrm{BspN3}$) sind im Vergleich dazu mit je 5% bis 15% seltener.[135] In der Klassenstufe 5 machen Beweisversuche mit genau einem Beispiel ($\mathrm{BspN1}$) einen statistisch signifikant größeren Anteil der Beweisversuche mit begründendem Ansatz und Beispiel aus als in den Klassenstufen 7, 8.

[133] Der Anteil der Beweisversuche mit begründendem Ansatz, bei deren Entwicklung ein Beispiel betrachtet wurde, ist vielleicht größer als aus den schriftlichen Produkten hervorgeht, u. a. weil in der Aufgabenstellung auf dem Arbeitsblatt ein Beispiel vorkommt und die Lernenden eventuell Beispiele nur im Kopf berechnet haben.

[134] In diesem Abschnitt 8.2.3.2 beziehen sich sämtliche relativen Häufigkeitsangaben nicht auf alle Beweisversuche mit begründendem Ansatz, sondern auf die Beweisversuche mit begründendem Ansatz und Beispiel.

[135] Dass die Lernenden eher selten mehr als fünf Beispiele angegeben haben, kann z. B. damit zusammenhängen, dass den Lernenden die Notwendigkeit einer allgemeingültigen Begründung bewusst war, sie anhand weniger Beispiele eine aus ihrer Sicht tragfähige Beweisidee gefunden haben oder dass ihnen im Wettbewerb nur ein begrenztes Zeitfenster zur Verfügung stand.



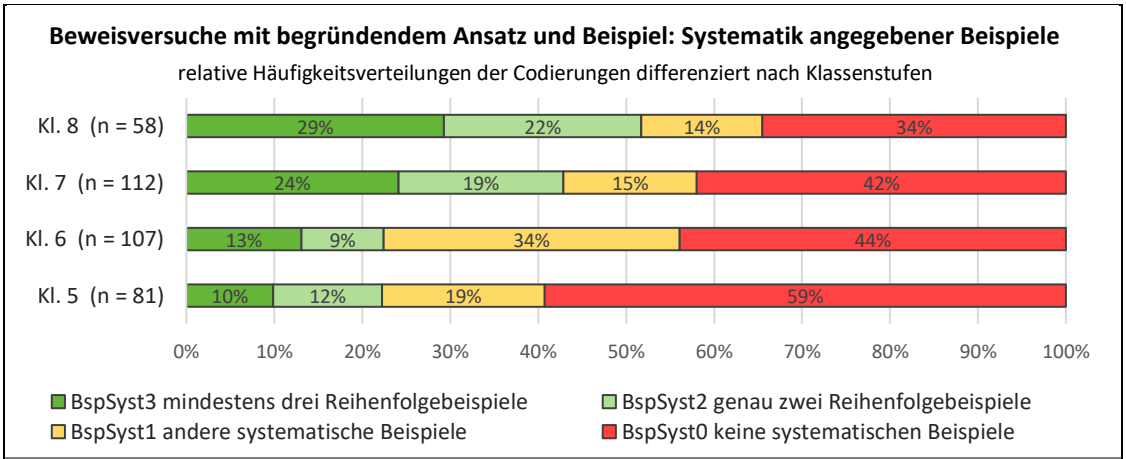
p-Werte von Signifikanztest und dazugehörige Effektstärken

	BspN1 vs. BspN2-3		BspN1 vs. BspN2-3		BspN1 vs. BspN2-3
Kl. 5 vs. Kl. 6	$p = .053^{N-1}, V = .142$	Kl. 5 vs. Kl. 7	$p = .022^{N-1}, V = .165$	Kl. 5 vs. Kl. 8	$p = .034^{N-1}, V = .180$

Abbildung 97: Beweisversuche mit begründendem Ansatz und Beispiel: Gesamtanzahl an Beispielen. Relative Häufigkeitsverteilungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Zusätzlich zu der Anzahl an Beispielen wurde die **Art der angegebenen Beispiele** untersucht. Bei jedem Beweisversuch mit begründendem Ansatz und Beispiel wurde erfasst, inwieweit die Beispiele eine Systematik aufweisen und ein besonderes oder ein großes Beispiel angegeben ist (Abschn. 3.3.5).

In Abbildung 98 sind die Ergebnisse der Analyse der Systematik angegebener Beispiele dargestellt. Sie weisen darauf hin, dass in der Klassenstufe 5 die Beweisversuche mit begründendem Ansatz und Beispiel seltener systematische Beispiele enthalten als in den Klassenstufen 6 bis 8. Der Anteil der Beweisversuche mit systematischen Beispielen (BspSyst1-3) ist in der Klassenstufe 5 mit ca. 40% im Vergleich zu den Klassenstufen 6 bis 8 mit etwa 60% statistisch signifikant geringer mit $0{,}15 < V < 0{,}25$. Die Analyse der Systematik angegebener Beispiele zeigt



p-Werte von Signifikanztest und dazugehörige Effektstärken

	BspSyst1-3 vs BspSyst0	BspSyst0-1 vs. BspSyst2-3	BspSyst0-2 vs. BspSyst3
Kl. 5 vs. Kl. 6	$p = .038^{N-1}, V = .152$		
Kl. 5 vs. Kl. 7	$p = .018^{N-1}, V = .171$	$p = .003^{N-1}, V = .215$	$p = .012^{N-1}, V = .182$
Kl. 5 vs. Kl. 8	$p = .004^{N-1}, V = .244$	$p < .001^{N-1}, V = .306$	$p = .003^{N-1}, V = .250$
Kl. 6 vs. Kl. 7		$p = .001^{N-1}, V = .217$	$p = .037^{N-1}, V = .141$
Kl. 6 vs. Kl. 8		$p < .001^{N-1}, V = .298$	$p = .011^{N-1}, V = .198$

Abbildung 98: Beweisversuchen mit begründendem Ansatz und Beispiel: Systematik angegebener Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests, dazugehörige Effektstärken

zudem, dass in den Klassenstufen 7, 8 in den Beweisversuchen mit begründendem Ansatz und Beispiel wesentlich häufiger Reihenfolgebeispiele, und insbesondere häufiger mindestens drei Reihenfolgebeispiele angegeben sind als in den Klassenstufen 5, 6. Die Unterschiede sind jeweils statistisch signifikant (Abb. 98). Auch wenn man die Beweisversuche mit begründendem Ansatz und mindestens zwei Beispielen fokussiert, ist der Anteil derjenigen mit Reihenfolgebeispielen in den Klassenstufen 7, 8 jeweils statistisch signifikant größer als in den Klassenstufen 5, 6 (jeweils $p < 0{,}05^{N-1}$, $0{,}16 < V < 0{,}36$).

Ein besonderes Beispiel ist je Klassenstufe in rund 50% bis 60% der Beweisversuche mit begründendem Ansatz und Beispiel zu finden. Dabei kommt das kleinste Beispiel am häufigsten vor. Es ist je Klassenstufe in gut einem Drittel bis zu gut der Hälfte der Beweisversuche mit begründendem Ansatz und Beispiel angegeben.

Der Anteil der Beweisversuche, in denen das kleinste Beispiel und Reihenfolgebeispiele vorkommen, an allen Beweisversuchen mit begründendem Ansatz und Beispiel liegt in den Klassenstufen 5, 6 bei 10% bis 15% und ist nach dem N-1 Chi²-Test jeweils statistisch signifikant niedriger mit $0{,}15 < V < 0{,}22$ als in den Klassenstufen 7, 8 mit 30% bis 35%. In den Klassenstufen 7, 8 sind in etwa jedem dritten Beweisversuch mit begründendem Ansatz und Beispiel das kleinste Beispiel und Reihenfolgebeispiele angegeben, in den Klassenstufen 5, 6 in weniger als jedem Sechstel.

Ein großes Beispiel ist je Klassenstufe in 5% bis 15% der Beweisversuche mit begründendem Ansatz und Beispiel zu sehen. Dies bedeutet andersherum, dass die meisten Beweisversuche mit begründendem Ansatz und Beispiel ausschließlich kleine und dadurch tendenziell leicht zu berechnende Beispiele enthalten.

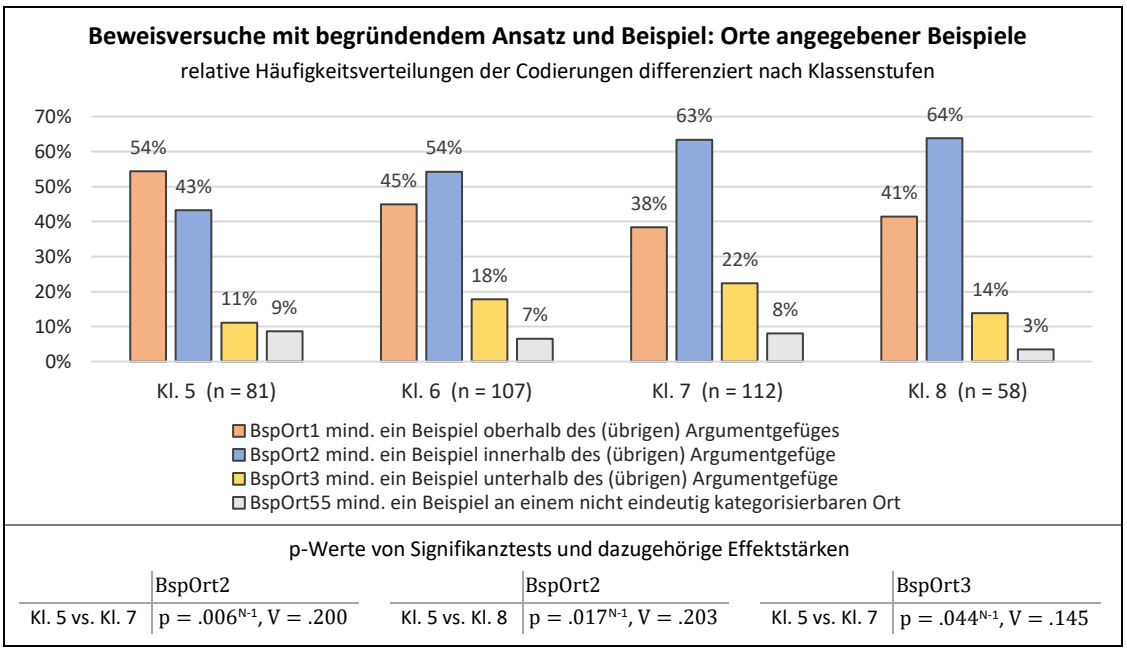
Um die Eigenschaften der Beispiele in Beweisversuchen mit begründendem Ansatz zu beschreiben, wurde des Weiteren der **Ort der angegebenen Beispiele** im Verhältnis zum (übrigen) Argumentgefüge untersucht (Abschn. 3.3.5). Die Abbildung 99 gibt an, in welchem Anteil der Beweisversuche mit begründendem Ansatz und Beispiel sich ein Beispiel oberhalb, innerhalb oder unterhalb des (übrigen) Argumentgefüges befindet.

Erstens lassen die Ergebnisse vermuten, dass häufig ein Beispiel vor der Formulierung des (übrigen) Argumentgefüges ausgeführt wurde. In einem Anteil von 38% bis 54% der Beweisversuche mit begründendem Ansatz und Beispiel ist oberhalb des (übrigen) Argumentgefüges ein Beispiel angegeben. Zudem befinden sich die Beispiele im Falle einer nicht eindeutigen Kategorisierbarkeit meistens links neben dem Argumentgefüge, was andeutet, dass oftmals zumindest ein Teil dieser Beispiele vor dem Argumentgefüge angeführt wurde. Schließlich ist zu bedenken, dass in der Aufgabenstellung ein Beispiel vorgegeben ist und sich Lernende eventuell weitere Beispiele im Kopf überlegt haben, so dass der Anteil der Lernenden, die vor der Formulierung des Argumentgefüges ein Beispiel betrachtet haben, wahrscheinlich noch größer ist, als es die Ergebnisse der Codierung zeigen.

Die Ergebnisse der Analyse des Orts der Beispiele weisen zweitens darauf hin, dass sich Beispiele besonders häufig innerhalb des (übrigen) Argumentgefüges befinden. Der relative Anteil der Beweisversuche mit begründendem Ansatz und Beispiel, in denen ein Beispiel innerhalb

des (übrigen) Argumentgefüges vorkommt, liegt je Klassenstufe zwischen 40% und 65%. Er ist in der Klassenstufe 5 mit 43% am kleinsten und statistisch signifikant kleiner mit geringer Effektstärke als in den Klassenstufen 7, 8 mit über 60%.

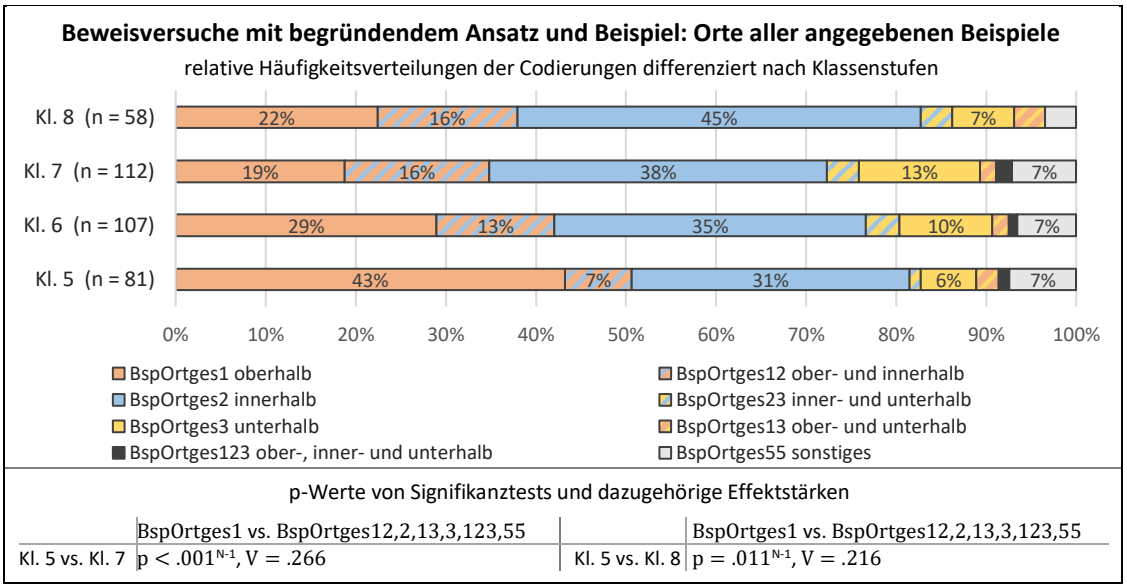
Drittens deuten die Ergebnisse an, dass es in jeder der Klassenstufen 5 bis 8 durchaus vorkommt, dass Lernende nach der Formulierung des (übrigen) Argumentgefüges noch ein Beispiel angeben. Je Klassenstufe enthält ein Zehntel bis zu einem Viertel der Beweisversuche mit begründendem Ansatz und Beispiel ein Beispiel unterhalb des (übrigen) Argumentgefüges. In der Klassenstufe 5 ist der Anteil mit 11% am kleinsten und statistisch signifikant kleiner mit $V = .145$ als in der Klassenstufe 7 mit 22%.



p-Werte von Signifikanztests und dazugehörige Effektstärken

	BspOrt2		BspOrt2		BspOrt3
Kl. 5 vs. Kl. 7	$p = .006^{N-1}, V = .200$	Kl. 5 vs. Kl. 8	$p = .017^{N-1}, V = .203$	Kl. 5 vs. Kl. 7	$p = .044^{N-1}, V = .145$

Abbildung 99: Beweisversuche mit begründendem Ansatz und Beispiel: Orte angegebener Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken



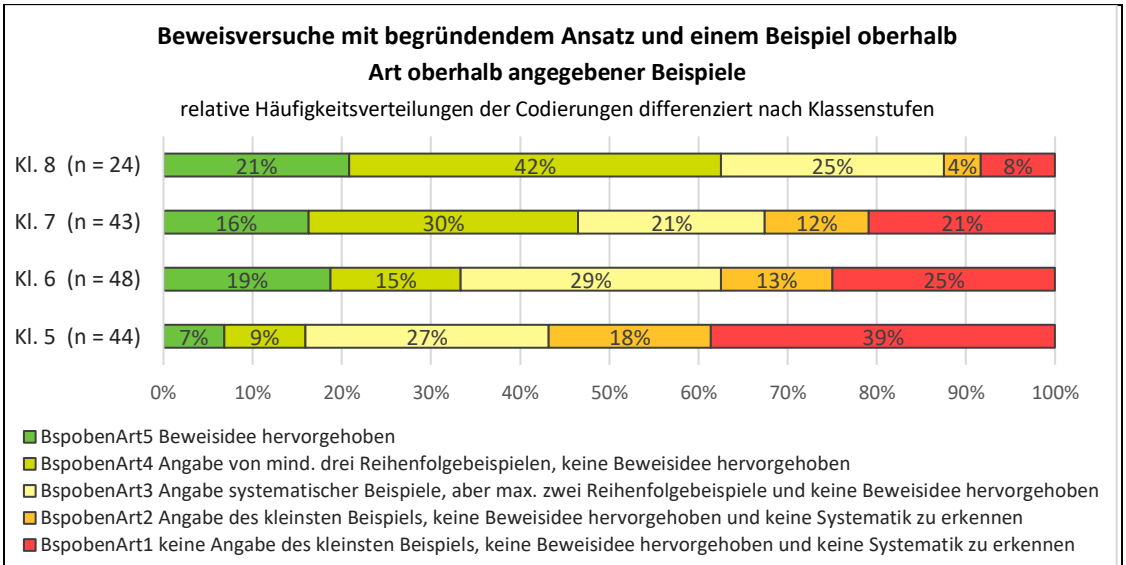
p-Werte von Signifikanztests und dazugehörige Effektstärken

	BspOrtges1 vs. BspOrtges12,2,13,3,123,55		BspOrtges1 vs. BspOrtges12,2,13,3,123,55
Kl. 5 vs. Kl. 7	$p < .001^{N-1}, V = .266$	Kl. 5 vs. Kl. 8	$p = .011^{N-1}, V = .216$

Abbildung 100: Beweisversuche mit begründendem Ansatz und Beispiel: Orte aller angegebenen Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Weiterhin ist in Abbildung 100 bezogen auf alle angegebenen Beispiele beschrieben, inwieweit sich diese ober-, inner- oder unterhalb des (übrigen) Argumentgefüges befinden. Je Klassenstufe 5 bis 8 befinden sich in 70% bis 85% der Beweisversuche mit begründendem Ansatz und Beispiel alle Beispiele oberhalb oder innerhalb des (übrigen) Argumentgefüges (BspOrtges1,12,2). In der Klassenstufe 5 machen die Beweisversuche, in denen alle Beispiele oberhalb des (übrigen) Argumentgefüges angeführt sind (BspOrtges1), mit 43% den größten Anteil aus, einen statistisch signifikant größeren als in den Klassenstufen 7, 8. Außerdem ist in Abbildung 100 zu sehen, dass je Klassenstufe in mindestens 70% der Beweisversuche mit begründendem Ansatz und Beispiel Beispiele ausschließlich oberhalb, innerhalb oder unterhalb des (übrigen) Argumentgefüges, d. h. an einem einzigen Ort, ausgeführt sind.

Bezüglich der Eigenschaften der **Beispiele, die oberhalb des (übrigen) Argumentgefüges angeführt sind**, lassen sich folgende Auffälligkeiten festhalten: Je Klassenstufe ist in über 90% der Fälle ein Beispiel zu der zu zeigenden Aussage und in fast allen Übrigen eine Summe von vier nicht aufeinanderfolgenden ungeraden Zahlen angegeben. Es scheint demnach nahezuliegen, sich eine zu beweisende Aussage zuerst an einem Beispiel zu veranschaulichen. In der Klassenstufe 5 ist in fast der Hälfte der Beweisversuche mit begründendem Ansatz und einem Beispiel oberhalb des (übrigen) Argumentgefüges ausschließlich genau ein Beispiel oberhalb angegeben. Demgegenüber sind in der Klassenstufe 8 in über 90% der Fälle mindestens zwei Beispiele oberhalb angeführt. Dieser Unterschied ist statistisch hoch signifikant ($p = .002^{N-1}$, $V = .379$). Außerdem gibt die Abbildung 101 ergänzende Einblicke in Eigenschaften der oberhalb angeführten Beispiele bei Beweisversuchen mit begründendem Ansatz und einem Beispiel oberhalb des (übrigen) Argumentgefüges. Dass an einem Beispiel oberhalb des (übrigen) Argu-



p-Werte von Signifikanztests und dazugehörige Effektstärken

	BspobenArt4 vs. ¬BspobenArt4	BspobenArt3-5 vs. ¬BspobenArt3-5
Kl. 7 vs. Kl. 5	$p = .013^{N-1}$, $V = .267$	$p = .024^{N-1}$, $V = .244$
Kl. 8 vs. Kl. 5	$p = .003^{ex}$, $V = .385$	$p < .001^{N-1}$, $V = .430$
Kl. 8 vs. Kl. 6	$p = .011^{N-1}$, $V = .301$	$p = .029^{N-1}$, $V = .259$

Abbildung 101: Beweisversuche mit begründendem Ansatz und einem Beispiel oberhalb: Zusammenfassende Betrachtung der Art der oberhalb angegebenen Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

mentgefüges eine Beweisidee hervorgehoben ist, trifft in der Klassenstufe 5 bei 7%, in den Klassenstufen 6 bis 8 bei 16% bis 21% der Beweisversuche mit begründendem Ansatz und einem Beispiel oberhalb des (übrigen) Argumentgefüges zu. In der Klassenstufe 8 ist der Gesamtanteil der Beweisversuche, in denen an den Beispielen oberhalb des (übrigen) Argumentgefüges eine Beweisidee hervorgehoben oder eine Systematik zu erkennen ist ($\mathrm{BspobenArt3\text{-}5}$), an allen Beweisversuchen mit begründendem Ansatz und einem Beispiel oberhalb mit über 85% statistisch signifikant größer als in den Klassenstufen 5, 6.

Im Vergleich zu den Beispielen oberhalb des (übrigen) Argumentgefüges fällt bei den Beweisversuchen mit begründendem Ansatz und einem **Beispiel unterhalb des (übrigen) Argumentgefüges** auf, dass in jeder der Klassenstufen 5 bis 8 in 45% bis 75% der Fälle an den Beispielen unterhalb des (übrigen) Argumentgefüges eine Beweisidee hervorgehoben ist. Abgesehen davon haben die Beispiele, die sich unterhalb des (übrigen) Argumentgefüges befinden, innerhalb jeder der Klassenstufen sehr unterschiedliche Eigenschaften.

Mit Blick auf Beweisversuche mit begründendem Ansatz und einem **Beispiel innerhalb des (übrigen) Argumentgefüges** lässt sich festhalten, dass in mehr als der Hälfte dieser Beweisversuche ein sonstiges Beispiel innerhalb des (übrigen) Argumentgefüges vorkommt. In den Klassenstufen 5, 6 sind oftmals ausschließlich sonstige Beispiele angeführt. Reihenfolgebeispiele sind innerhalb des (übrigen) Argumentgefüges in den Klassenstufen 7, 8 in ca. einem Drittel, in den Klassenstufen 5, 6 in ca. einem Zehntel der Beweisversuche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges zu finden, ein großes Beispiel kommt in allen Klassenstufen mit einer relativen Häufigkeit von deutlich unter 10% noch seltener vor.

Hinsichtlich der Integration der innerhalb angeführten Beispiele in das Argumentgefüge zeigen die Daten folgendes Bild: Ein notwendiges Beispiel kommt in den Klassenstufen 5, 6 in 6% bis 16%, in der Klassenstufe 7 in 39% und in der Klassenstufe 8 in 65% der Beweisversuche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges vor. Dabei sind die Unterschiede zwischen den Klassenstufen 5, 6 und 7, 8 nach dem N-1 Chi²-Test alle statistisch hoch signifikant mit $0{,}26 < \mathrm{V} < 0{,}62$. In der Klassenstufe 8 sind in über 50% der Beweisversuche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges alle diese Beispiele auch notwendig, während dies in den Klassenstufen 5 bis 7 jeweils in unter 25% der Fälle zutrifft. Das bedeutet, dass vor allem in der Klassenstufe 8 die Schlussweise häufig trotz innerhalb angeführter Beispiele deduktiv sein kann. Weiterhin lässt sich hinsichtlich der Integration der innerhalb angeführten Beispiele in das Argumentgefüge festhalten, dass in jeder der Klassenstufen in knapp 80% bis 95% der Fälle mindestens eines der innerhalb angeführten Beispiele ein fester Bestandteil des Argumentgefüges ist oder die Ausführungen inhaltlich ergänzt. Zugleich kann in den Klassenstufen 5 bis 7 in etwa 45% der Beweisversuche mit einem Beispiel innerhalb des Argumentgefüges mindestens eines der innerhalb angeführten Beispiele weggelassen werden, ohne die Struktur des Argumentgefüges oder die im Argumentgefüge dargelegten inhaltlichen Überlegungen zu beeinflussen. Dies bedeutet, dass meistens eines der innerhalb angeführten Beispiele für die Korrektheit und Vollständigkeit des Argumentgefüges relevant ist, und dass in den Klassenstufen 5 bis 7 häufig mehr Beispiele innerhalb des Argumentgefüges angeführt sind als erforderlich.

Zusammenfassend lassen sich mit Blick auf die Frage, in welcher Form Beispiele in Beweisversuchen mit begründendem Ansatz und Beispiel besonders häufig oder eher selten vorkommen, folgende Aspekte herausstellen: In den Beweisversuchen mit begründendem Ansatz und Beispiel kommen meistens zwei bis fünf Beispiele vor. In den Klassenstufen 7, 8 sind im Vergleich zu den Klassenstufen 5, 6 wesentlich öfters Reihenfolgebeispiele angegeben. Ein großes Beispiel kommt in jeder der Klassenstufen 5 bis 8 eher selten vor.

In der Klassenstufe 5 befindet sich in mehr als der Hälfte der Beweisversuche mit begründendem Ansatz und Beispiel mindestens ein Beispiel oberhalb des (übrigen) Argumentgefüges, in den Klassenstufen 6 bis 8 befindet sich in mehr als der Hälfte der Beweisversuche mit begründendem Ansatz und Beispiel mindestens ein Beispiel innerhalb des (übrigen) Argumentgefüges. Je Klassenstufe sind in über 70% der Beweisversuche mit begründendem Ansatz und Beispiel alle Beispiele ober- oder innerhalb des (übrigen) Argumentgefüges angeführt. Es kommt in jeder der Klassenstufen auch vor, dass Lernende ein Beispiel unterhalb des (übrigen) Argumentgefüges angegeben haben.

Wenn oberhalb des (übrigen) Argumentgefüges ein Beispiel angegeben ist, ist in der Klassenstufe 8 besonders häufig, in der Klassenstufe 5 vergleichsweise selten an den oberhalb angeführten Beispielen eine Beweisidee hervorgehoben oder eine Systematik zu erkennen. Wenn innerhalb des (übrigen) Argumentgefüges ein Beispiel angegeben ist, handelt es sich häufig um sonstige Beispiele. In der Klassenstufe 8 sind in ca. der Hälfte der Fälle alle innerhalb angeführten Beispiele notwendige Beispiele. Wenn in einem Beweisversuch mit begründendem Ansatz ein Beispiel unterhalb des (übrigen) Argumentgefüges angegeben ist, ist meistens an mindestens einem der unterhalb angeführten Beispiele eine Beweisidee hervorgehoben.

8.2.3.3 Einblicke in die Funktion angegebener Beispiele aus Sicht von Lesenden

Dieser Abschnitt gibt Einblicke in die Funktion angegebener Beispiele aus der Sicht von Leserinnen und Lesern. Ein besonderer Fokus liegt dabei auf Anknüpfungspunkten für die Weiterentwicklung der Beweisversuche zu einem Beweis. In den vorhergehenden Abschnitten 8.2.3.1, 8.2.3.2 wurden Beispiele in Beweisversuchen mit begründendem Ansatz mit Blick auf den Nachweis der Gültigkeit und Charakteristika der Beweisversuche betrachtet. Allerdings lässt sich aus den Ausführungen kaum folgern, welche Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche sich aus der Verwendung der Beispiele ergeben. Um diese herauszuarbeiten, ist es erforderlich, die einzelnen Beweisversuche im Gesamten zu betrachten. Insbesondere sind neben den Beispielen an sich auch die Schlussweise, die Beweisideen sowie die Korrektheit und Vollständigkeit der Argumentgefüge zu berücksichtigen. Im Folgenden werden einige Funktionen vorgestellt, die Beispiele aus der Sicht von Lesenden in Beweisversuchen mit begründendem Ansatz haben können, und diese Funktionen anhand einzelner Bearbeitungen veranschaulicht, um Einblicke in die Funktionen der Beispiele und verschiedene Anknüpfungspunkte, die sich aus angegebenen Beispielen ergeben können, zu ermöglichen.[136]

[136] Dabei ist zu bedenken, dass diese Überlegungen aus der Sicht einer Leserin geschrieben sind. Warum Lernende Beispiele angegeben haben, kann sehr unterschiedliche Gründe haben, die sich allein anhand der Beweisversuche nicht festmachen lassen und von deren Funktion aus Sicht von Lesenden abweichen können (Abschn. 3.3.5).

Manche Funktionen können Beispielen in sämtlichen Beweisversuchen mit begründendem Ansatz, d. h. **unabhängig von der Schlussweise**, zukommen. Dazu gehört, dass Beispiele zu der zu zeigenden Aussage, die korrekt und vollständig ausgeführt sind, diese Aussage aus der Sicht von Lesenden[137] für einen Spezialfall bestätigen. Sie sind mit Blick auf den Nachweis der Gültigkeit dieser Aussage keineswegs hinreichend, aber ein erster kleiner Schritt, gerade bei Beweisversuchen mit begründendem Ansatz, aber ohne weiterentwickelbares Argumentgefüge (Bearbeitung 8.2.3a; Abschn. 8.2.3.1, 8.2.5.1, 8.5.1).

Bearbeitung 8.2.3a: Klassenstufe 6, weiblich (Code: 0611922)

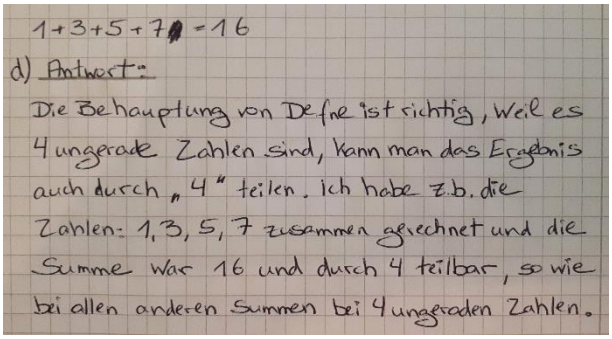


Abbildung 102: Bearbeitungen 8.2.3a (Beweisversuch mit begründendem Ansatz und einem korrekt und vollständig ausgeführten Beispiel zu der zz. Aussage, aber keinem weiterentwickelbaren Argumentgefüge)

Zusätzlich zum Nachweis der Gültigkeit für einen Spezialfall signalisieren korrekt und vollständig ausgeführte eigene Beispiele zu der zu zeigenden Aussage, dass die Lernenden ein solches Beispiel angeben können. Ist in einer Bearbeitung etwa das Beispiel der vier aufeinanderfolgenden ungeraden Zahlen $7, 9, 11, 13$ als Induktionsanfang korrekt und vollständig ausgeführt, lässt sich mit Blick auf eine Weiterentwicklung des Beweisversuchs davon ausgehen, dass der Lernende prinzipiell auch das kleinste Beispiel als korrekten Induktionsanfang aufschreiben könnte. In der Bearbeitung 8.2.3a in Abbildung 102 ist das kleinste Beispiel, aber kein weiterentwickelbares Argumentgefüge ausgeführt. An das Beispiel anknüpfend ließe sich etwa durch eine systematische Betrachtung mehrerer Beispiele die Entwicklung einer ausbaubaren Beweisidee anstoßen.

Bearbeitung 8.2.3b: Klassenstufe 5, weiblich (Code: 0522307)

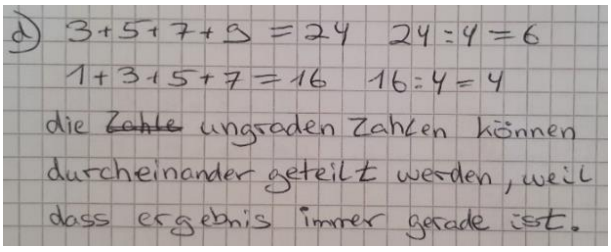


Abbildung 103: Bearbeitung 8.2.3b (Beweisversuch mit begründendem Ansatz und Beispielen, die u. a. vermuten lassen, dass die zu zeigende Aussage korrekt interpretiert wurde und sich die Ausführungen auf die zu zeigende Aussage beziehen)

[137] Der gesamte Abschnitt zur Funktion von Beispielen fokussiert, wie angegeben, auf die Perspektive von Leserinnen und Lesern. Für eine gute Lesbarkeit wird im Folgenden nicht ständig darauf hingewiesen.

Beispiele können in sämtlichen Beweisversuchen mit begründendem Ansatz auch darauf hinweisen, wie die zu zeigende Aussage interpretiert wurde und dass sich eventuell vage gehaltene oder fachlich nicht ganz korrekte Ausführungen auf die zu zeigende Aussage beziehen. Beispielsweise wäre bei der Bearbeitung 8.2.3b in Abbildung 103 ohne die Beispiele fraglich, inwieweit sich die Ausführungen auf eine korrekte Interpretation der zu zeigenden Aussage beziehen. Obwohl die verbalsprachlichen Ausführungen nicht explizit auf die angegebenen Beispiele verweisen, legen sie nahe, dass mit „ungeraden Zahlen" konkret vier aufeinanderfolgende ungerade Zahlen und mit dem Ausdruck „durcheinander geteilt werden" die Teilbarkeit der Summe der Zahlen durch vier gemeint ist. Mit Blick auf Anknüpfungspunkte für eine Weiterentwicklung des Beweisversuchs geht aus der Bearbeitung hervor, dass die Schülerin eigenständig ein Beispiel zu der zu zeigenden Aussage angeben kann und die Begrifflichkeiten der zu zeigenden Aussage tendenziell korrekt interpretiert wurden. Zugleich deutet sich an, dass für eine präzise Versprachlichung inhaltlicher Überlegungen Unterstützung erforderlich wäre.

Zusätzlich zu den beschriebenen Funktionen, die Beispiele in sämtlichen Beweisversuchen mit begründendem Ansatz haben können, werden im Folgenden differenziert nach der Schlussweise Einblicke in weitere mögliche Funktionen von Beispielen in Beweisversuchen mit begründendem Ansatz gegeben. Eine Differenzierung nach Schlussweisen bietet sich an, weil die Funktionen von Beispielen aus der Sicht von Lesenden mit der Schlussweise, d. h. dem Verhältnis angegebener Beispiele zu allgemein gehaltenen inhaltlichen Überlegungen, teilweise zusammenspielen (Abschn. 3.3.6).

In Beweisversuchen mit begründendem Ansatz und der **Schlussweise *deduktiv ohne zusätzliches Beispiel*** (Sw4) sind alle eventuell angegebenen Beispiele notwendige Bestandteile des Argumentgefüges und haben somit eine zentrale Funktion im Argumentgefüge. Beispielsweise sind die Beispiele in der Bearbeitung 8.2.3c Fälle einer vollständigen Fallunterscheidung und dadurch notwendige, nicht weglassbare Bestandteile des Argumentgefüges. Solche notwendigen Beispiele können ebenfalls in Beweisversuchen mit begründendem Ansatz und einer der anderen Schlussweisen Sw1-3 vorkommen. Andersherum gilt jedoch, dass die Funktionen, die Beispiele in Beweisversuchen mit einer der Schlussweisen Sw1-3 einnehmen können, bei denen mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* nicht relevant sind.

Bearbeitung 8.2.3c: Klassenstufe 8, weiblich (Code: 0822002)

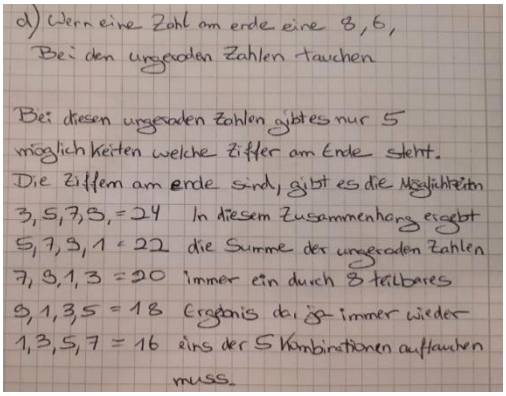


Abbildung 104: Bearbeitung 8.2.3c (Beweisversuch mit begründendem Ansatz, Beispielen und der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel*)

In Beweisversuchen mit der **Schlussweise *deduktiv mit zusätzlichem Beispiel*** (Sw3) ist im Unterschied zu Beweisversuchen mit der Schlussweise Sw4 mindestens ein Beispiel angegeben, das kein notwendiger Bestandteil des Argumentgefüges ist. Dabei könnten alle nicht notwendigen Beispiele weggelassen werden, ohne die deduktive Struktur des Argumentgefüges oder die im Argumentgefüge dargelegten inhaltlichen Überlegungen zu beeinträchtigen (Abschn. 3.3.6). Aus der Sicht von Lesenden können Beispiele in Beweisversuchen mit der Schlussweise Sw3 *deduktiv mit zusätzlichem Beispiel* einen Ausschnitt des ansonsten allgemein gehaltenen Argumentgefüges veranschaulichen und konkretisieren. Zudem können sie Überlegungen durch die Bestätigung für einen Spezialfall stützen. Zum Beispiel ist in der bereits auf Seite 98 angeführten Bearbeitung 3.3.6c die Überlegung, dass das Vierfache einer ungeraden Zahl plus zwölf durch acht teilbar ist, an zwei Beispielen veranschaulicht und wird durch die Beispiele gestützt, wenn auch nicht bewiesen. In der nachfolgenden Bearbeitung 8.2.3d in Abbildung 105 sind Beispiele dazu angegeben, dass die Summe zweier ungerader Zahlen und die Summe zweier gerader Zahlen gerade ist. In beiden Bearbeitungen 3.3.6c und 8.2.3d sind die Beispiele aus fachlicher Perspektive obsolet. Sie sind kein fester Bestandteil des Argumentgefüges und ergänzen die inhaltlichen Überlegungen nicht. Mit Blick auf die Weiterentwicklung dieser Beweisversuche zu einem Beweis wäre es eventuell sinnvoll zu thematisieren, dass die Beispiele die veranschaulichten Überlegungen nicht beweisen und dass die Angabe von Beispielen i. A. nicht notwendig ist, aber etwa bei der Kontrolle der eigenen Überlegungen oder der Entwicklung von Beweisideen helfen kann. Ansonsten stellen die angeführten Beispiele für die Weiterentwicklung der Beweisversuche keinen besonderen Anknüpfungspunkt dar.

Bearbeitung 8.2.3d: Klassenstufe 5, männlich (Code: 0511319)

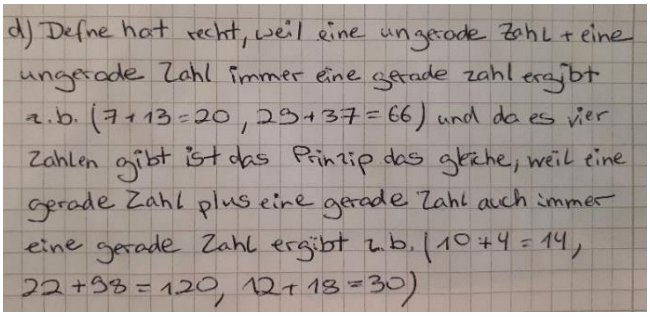


Abbildung 105: Bearbeitung 8.2.3d (Beweisversuch mit begründendem Ansatz und der Schlussweise Sw3 *deduktiv mit zusätzlichem Beispiel*)

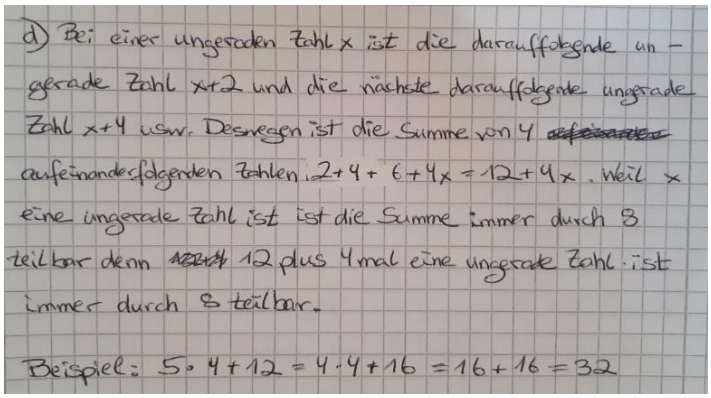
Beweisversuche mit der **Schlussweise *deduktiv mit induktivem Anteil*** (Sw2) enthalten ebenfalls mindestens ein nicht notwendiges Beispiel. Die Schlussweise ist überwiegend deduktiv, wobei das deduktive Argumentgefüge jedoch durch Weglassen aller angegebener Beispiele beeinflusst würde. Es ist ein Teil der Schlüsse an konkreten Beispielen ausgeführt oder ein Beispiel angegeben, welches die Ausführungen inhaltlich ergänzt (Abschn. 3.3.6).

In manchen Beweisversuchen mit der Schlussweise *deduktiv mit induktivem Anteil* wurden aus der Sicht von Lesenden Beispiele anstatt von Fachvokabularien verwendet, um Überlegungen zu kommunizieren. Beispielsweise steht in einem Beweisversuch eines Fünftklässlers: „Ungerade + ungerade ergibt immer gerade. Aber wenn $1 + 1$ oder $3 + 3$ oder $5 + 5$ usw. kommt

eine Zahl heraus die nicht durch 4 teilbar ist. […]“ (Kl. 5, m, Code 0510223). Ein Sechstklässler erklärt: „Ihre Behauptung ist richtig, weil 2, 4, 6 … ungerade Zahlen die man addiert immer gerade sind. […]“ (Kl. 6, m, Code 0620906). Eine Siebtklässlerin schreibt: „Denise hat Recht, denn jede Zahl wird immer um 2, 4, 6, 8 usw. größer […]“ (Kl. 7, w, Code 0710112). Es ist möglich, dass Lernenden allgemeine Fachvokabularien fehlen oder noch nicht abrufbar sind, die Verknüpfung mehrerer Fachtermini schwerfällt, die Versprachlichung mithilfe von Beispielen leichter erscheint oder auch eine Gewohnheit besteht, sich auf diese Weise auszudrücken.[138] Die Beweisversuche belegen zugleich, dass es einigen Lernenden gelingt, mithilfe konkreter Objekte, wie bestimmten Zahlen, Überlegungen zu kommunizieren, wenn die notwendigen sprachlichen Mittel eventuell noch nicht zur Verfügung stehen. Zur Weiterentwicklung ließen sich die Beweisversuche sprachlich unterstützen.

In den nachfolgenden zwei Bearbeitungen 8.2.3e-f mit der Schlussweise *deduktiv mit induktivem Anteil* (Sw2) ist unterhalb des (übrigen) Argumentgefüges ein Beispiel angegeben, welches das Argumentgefüge ergänzt oder präzisiert. Die Bearbeitungen geben exemplarische Einblicke, inwieweit Beispiele aus der Sicht von Lesenden Anknüpfungspunkte bieten können, um inhaltliche Lücke im Argumentgefüge zu schließen.

Bearbeitung 8.2.3e: Klassenstufe 7, männlich (Code: 0711103)



Bearbeitung 8.2.3f: Klassenstufe 5, männlich (Code: 0521836)

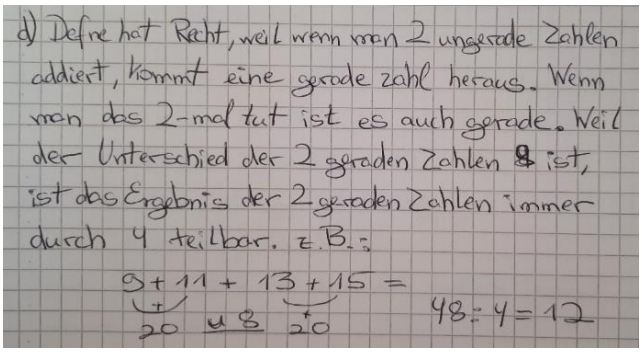


Abbildung 106: Bearbeitungen 8.2.3e-f (Beweisversuche mit begründendem Ansatz, der Schlussweise Sw2 *deduktiv mit induktivem Anteil* und einem Beispiel unterhalb des übrigen Argumentgefüges)

[138] Eine Analyse der Beweisversuche aus sprachlicher Sicht kann im Rahmen dieser Arbeit nicht geleistet werden und steht nicht im Fokus, könnte aber im Anschluss erfolgen.

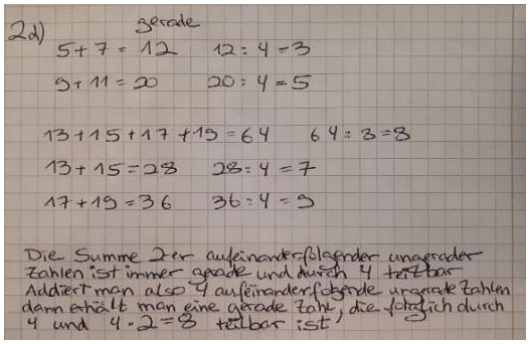
In der Bearbeitung 8.2.3e wird unter Verwendung von Variablen in allgemeingültiger Form über die primäre Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* argumentiert. Die Bearbeitung enthält eine inhaltliche Lücke, weil sie im Vergleich zum relevanten Erwartungshorizont weitgehend offenlässt, warum das Vierfache einer ungeraden Zahl plus zwölf immer durch acht teilbar ist (Abschn. 3.3.4 zur Analyse der Korrektheit und Vollständigkeit sowie Erwartungshorizonten). Zu der Aussage, dass das Vierfache einer ungeraden Zahl plus zwölf immer durch acht teilbar ist, ist unterhalb des (übrigen) Argumentgefüges ein Beispiel angegeben. Eine mögliche Interpretation lautet, dass der Siebtklässler sich an dieser Stelle selbst unsicher war, ob diese Aussage gilt, und deshalb ein Beispiel betrachtet hat. Eine weitere Möglichkeit besteht darin, dass das Beispiel als Stützung (oder als Beweis) herangezogen wurde, wo allgemeingültige Argumente fehlten. An dem Beispiel fällt auf, dass der Term „$5 \cdot 4 + 12$“ nicht direkt berechnet, sondern zunächst so umgeformt wurde, dass beide Summanden durch acht teilbar sind. Dieser Ansatz lässt sich ausbauen und verallgemeinern, weshalb das Beispiel die oberhalb angeführten Überlegungen ergänzt und einen guten Anknüpfungspunkt für die Weiterentwicklung des Beweisversuchs darstellt.

Die Bearbeitung 8.2.3f basiert auf der primären Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare*. Ähnlich wie bei der Bearbeitung 8.2.3e ist unterhalb allgemein gehaltener Ausführungen ein Beispiel angegeben. Es hebt hervor, dass mit den ungeraden Zahlen, die addiert werden, die beiden kleinsten und die beiden größten von vier aufeinanderfolgenden ungeraden Zahlen gemeint sind, und präzisiert somit die allgemein gehaltenen Ausführungen. Im Vergleich zu der vorhergehenden Bearbeitung liefert das Beispiel jedoch keinen direkten Beitrag dazu, die inhaltliche Lücke zu schließen, warum die Summe zweier gerader Zahlen mit Differenz acht durch vier teilbar ist. Zusätzlich stellt sich die Frage, warum die Summe der größten beiden von vier aufeinanderfolgenden ungeraden Zahlen immer um acht größer ist als die Summe der kleinsten beiden von vier aufeinanderfolgenden ungeraden Zahlen. An dem Beispiel ist diesbezüglich nichts hervorgehoben, wenngleich es eine mögliche Herangehensweise wäre, zunächst an dem konkreten Beispiel die Differenz der beiden Summen zu begründen.

In Beweisversuchen mit der Schlussweise *deduktiv mit induktivem Anteil* (Sw2) ist es neben der Angabe eines Beispiels unterhalb des (übrigen) Argumentgefüges u. a. möglich, dass Beispiele oberhalb oder innerhalb des (übrigen) Argumentgefüges ausgeführt sind. Die beiden Bearbeitungen 8.2.3g-h in Abbildung 107 stehen exemplarisch für Beweisversuche mit der Schlussweise Sw2, die aus der Sicht von Lesenden zu einem bedeutsamen Anteil mithilfe von Beispielen entwickelt oder dargestellt wurden. Die beiden Bearbeitungen weisen zudem deutlich darauf hin, dass die Arbeit mit Beispielen trotz allen Potentials keineswegs einen Selbstläufer hin zu einem korrekten und vollständigen Argumentgefüge darstellt, auch nicht bei interessierten und vergleichsweise leistungsstarken Lernenden in den Klassenstufen 7, 8.

So wie die vorhergehende Bearbeitung 8.2.3f basieren auch beide Bearbeitungen 8.2.3g-h auf der Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare*, enthalten jedoch einen inhaltlichen Fehler. In der linken Bearbeitung 8.2.3g einer Achtklässlerin wird ausgehend davon, dass die Summe zweier aufeinanderfolgender ungerader Zahlen durch zwei und vier teilbar ist, über die Rechnung $4 \cdot 2 = 8$ auf die Teilbarkeit durch acht der Gesamtsumme geschlossen. Ob die

Bearbeitung 8.2.3g: Kl. 8, weiblich (Code: 0810101)



Bearbeitung 8.2.3h: Kl. 7, weiblich (Code: 0712101)

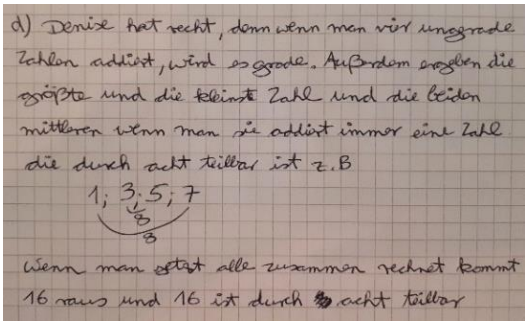


Abbildung 107: Bearbeitungen 8.2.3g-h (Beweisversuche mit begründendem Ansatz und der Schlussweise Sw2 *deduktiv mit induktivem Anteil*, die zu einem bedeutsamen Anteil mithilfe von Beispielen entwickelt oder dargestellt wurden)

Schülerin die Beziehung $4 \cdot 2 = 8$ wegen der Addition zweier durch vier teilbarer Zahlen oder wegen der Teilbarkeit jedes der Zahlenpaare durch zwei und vier als legitimen Grund für den Schluss auf die Teilbarkeit durch acht angesehen hat, lassen die Ausführungen offen. Unabhängig davon ist der angegebene Schluss nicht korrekt. An den oberhalb des (übrigen) Argumentgefüges angeführten Beispielen ist aus Expertensicht die Idee der Addition der kleinsten und der größten beiden von vier aufeinanderfolgenden ungeraden Zahlen sowie die Teilbarkeit durch zwei und vier jedes der Zahlenpaare hervorgehoben. Es ist jedoch kein Hinweis darauf zu erkennen, dass die Gültigkeit der Begründung über die Rechnung $4 \cdot 2 = 8$ hinterfragt oder nach einem anderen Grund für den Schluss auf die Teilbarkeit durch acht gesucht wurde. Die Bearbeitung 8.2.3g macht somit darauf aufmerksam, dass es vorkommen kann, dass Schülerinnen und Schüler an Beispielen eine Beweisidee erarbeiten oder darstellen, die jedoch aus Expertensicht nicht hinreichend für den Beweis einer zu zeigenden Aussage ist. Anstelle einer weitergehenden Betrachtung von Beispielen oder der Recherche einer alternativen Beweisidee werden eventuell plausible, aus Sicht des Lernenden vielleicht tragfähige, jedoch faktisch fehlerhafte Gründe angeführt, um z. B. auf die vorgegebene Behauptung zu schließen. Mit Blick auf eine Weiterentwicklung des vorliegenden Beweisversuchs 8.2.3.g erscheint es sinnvoll, zunächst die Schülerin aufzufordern, zu erklären, was sie sich überlegt hat und herauszufinden, wie die Rechnung $4 \cdot 2 = 8$ gedacht war, bevor man sie etwa durch ein passendes Gegenbeispiel (wie z. B. „4 und 8 sind durch vier teilbar, aber $4 + 8 = 12$ ist nicht durch vier teilbar“ oder „12 ist durch 2 und 4 teilbar, aber nicht durch $2 \cdot 4 = 8$“) auf die Ungültigkeit des Schlusses aufmerksam macht. Der Fehler wäre zu korrigieren und im Vergleich zu den Erwartungshorizonten wäre auch die inhaltliche Lücke zu schließen, warum die Summe von zwei aufeinanderfolgenden ungeraden Zahlen durch vier teilbar ist.

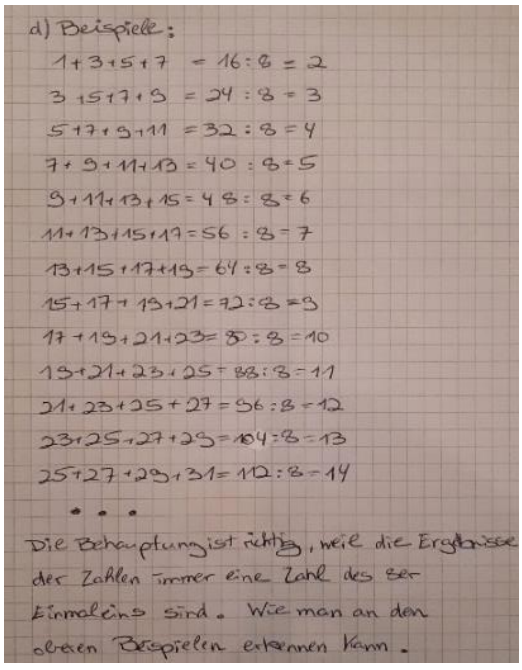
In der rechten Bearbeitung 8.2.3h in Abbildung 107 werden nicht die Summen der beiden kleinsten und der beiden größten, sondern die Summen der beiden „äußeren“ und der beiden „mittleren“ von vier aufeinanderfolgenden ungeraden Zahlen betrachtet. Die Bearbeitung 8.2.3h beginnt mit allgemein gehaltenen Ausführungen, die am Beispiel der vier aufeinanderfolgenden ungeraden Zahlen $1, 3, 5, 7$ veranschaulicht und fortgesetzt werden. Allerdings gilt die Aussage, dass die Summe der beiden „äußeren“ und der beiden „mittleren“ von vier aufeinanderfolgenden ungeraden Zahlen immer durch acht teilbar ist, nur in jedem zweiten Fall (bzw. wenn der kleinste Summand bei der Division durch acht Rest eins oder Rest fünf lässt).

Die Bearbeitung 8.2.3h deutet somit an, dass manche Lernende bei der Arbeit mit Beispielen zu verallgemeinernd unterwegs sind. Beide Bearbeitungen 8.2.3g-h machen darauf aufmerksam, dass bei der Entwicklung und Formulierung eines Beweises mithilfe von Beispielen eine tiefgehende, kritisch-reflexive Auseinandersetzung mit Beispielen zumindest hilfreich, wenn nicht sogar notwendig ist.

So wie Beweisversuche mit der Schlussweise Sw2 *deduktiv mit induktivem Anteil* enthalten Beweisversuche mit der **Schlussweise *induktiv mit Andeutung der Verallgemeinerbarkeit*** (Sw1) mindestens ein nicht notwendiges Beispiel, wobei das deduktive Argumentgefüge durch Weglassen aller nicht notwendiger Beispiele beeinflusst würde. Bei Beweisversuchen in der Kategorie Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit* basiert das Argumentgefüge primär auf Überlegungen, die an konkreten Beispielen durchgeführt werden. Dabei ist die Verallgemeinerbarkeit dieser Überlegungen angedeutet (Abschn. 3.3.6). Im Vergleich zu Beweisversuchen mit einer der Schlussweisen Sw2-4 spielen Beispiele in Beweisversuchen mit der Schlussweise Sw1 also eine besonders dominante Rolle. Überlegungen an konkreten Beispielen sind in vielen dieser Beweisversuche konstitutiv für deren Einordnung in die Kategorie *begründendem Ansatz*.

Die beiden Bearbeitungen 8.2.3i-j in Abbildung 108 veranschaulichen die dominante Rolle von Beispielen in Beweisversuchen mit der Schlussweise Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit*. Zudem deuten sie an, dass manche Lernende durch die Betrachtung von Reihenfolgebeispielen eine Beweisidee gefunden und im Beweisversuch gewinnbringend genutzt haben. Dabei kann die Bearbeitungstiefe deutlich divergieren.

Bearbeitung 8.2.3i: Kl. 8, weiblich (Code: 0811204)



Bearbeitung 8.2.3j: Kl. 8, männlich (Code: 0821501)

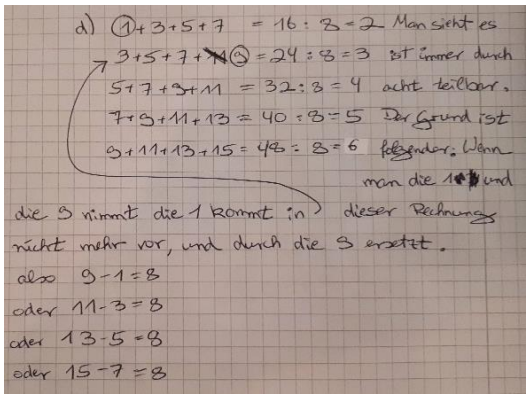


Abbildung 108: Bearbeitungen 8.2.3i-j (Beweisversuche mit begründendem Ansatz und der Schlussweise Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit*, die vermuten lassen, dass durch die Betrachtung von Reihenfolgebeispielen eine Beweisidee gefunden wurde)

In der Bearbeitung 8.2.3i sind dreizehn Reihenfolgebeispiele beginnend mit dem kleinsten Beispiel angegeben. Die Fortsetzbarkeit dieser Auflistung ist durch drei Punkte angedeutet. Darunter steht mit explizitem Verweis auf die Beispiele, dass „die Ergebnisse der Zahlen immer eine Zahl des 8er Einmaleins sind“. Diese Invariante scheint an den angegebenen Beispielen erkannt worden zu sein und mit den Beispielen gestützt zu werden („Wie man an den obigen Beispielen erkennen kann“). In der Bearbeitung gibt es keinen Hinweis darauf, dass die Schülerin eine Notwendigkeit gesehen hat, unter Verwendung allgemeingültiger Wissenselemente zu begründen, warum die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer in der Achterreihe ist. Die Allgemeingültigkeit dieser Idee zu hinterfragen, die Notwendigkeit einer tiefergehenden, allgemeingültigen Begründung zu motivieren und diese (z. B. ausgehend von den angegebenen Beispielen) zu entwickeln, wären die nächsten Schritte.

In der Bearbeitung 8.2.3j sind, so wie in der linken Bearbeitung, Reihenfolgebeispiele beginnend mit dem kleinsten Beispiel angegeben. Es handelt sich jedoch nicht um dreizehn, sondern um fünf Beispiele. Zudem wird anhand der Beispiele unter Gebrauch numerischer, ikonischer und verbaler Darstellungsmittel erläutert, dass bei den nächstgrößeren vier aufeinanderfolgenden ungeraden Zahlen anstatt der kleinsten Zahl der vorhergehenden vier aufeinanderfolgenden ungeraden Zahlen stets eine um acht größere Zahl hinzukommt. Inwieweit der Schüler allgemeingültig begründen könnte, warum immer eine um acht größere Zahl hinzukommt, oder er dies ausschließlich an den angegebenen Beispielen festmacht, lässt die Bearbeitung offen. Mit Blick auf die Weiterentwicklung des Beweisversuchs zu einem Beweis erscheint es sinnvoll, dies und auch die Relevanz der Teilbarkeit der kleinsten Summe von vier aufeinanderfolgenden ungeraden natürlichen Zahlen durch acht zu besprechen. Die Ausführungen wären zu ergänzen und zu einem deduktiven Argumentgefüge weiterzuentwickeln.

Die abschließende Bearbeitung 8.2.3k in Abbildung 109 macht besonders deutlich, dass an Beispielen ausgeführte Überlegungen konstitutiv für die Einordnung eines Beweisversuchs in die Oberkategorie *begründendem Ansatz* sein können. In Abbildung 109 ist ein Beweisversuch eines Fünftklässlers zu sehen, dem es gelungen ist, an Beispielen ein weitgehend korrektes und vollständiges Argumentgefüge zu formulieren, und zwar nicht basierend auf der primären Beweisidee 1, sondern auf der primären Beweisidee 3 *Vielfaches einer mittleren Zahl oder Durch-*

Bearbeitung 8.2.3k: Klassenstufe 5, männlich (Code: 0521720)

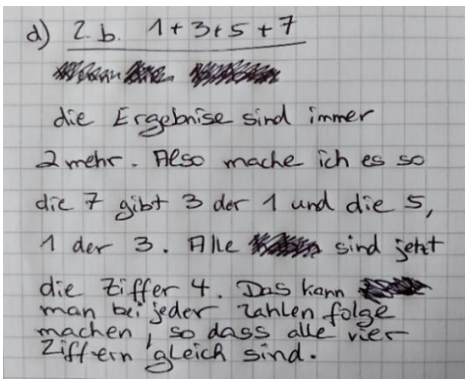


Abbildung 109: Bearbeitung 8.2.3k (Beweisversuch mit begründendem Ansatz, der Schlussweise Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit* und einem weitgehend korrekten und vollständigen Argumentgefüge basierend auf der Beweisidee 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*, der von einem Fünftklässler verfasst wurde)

schnittsbildung. Die Schlussweise ist induktiv mit Andeutung der Verallgemeinerbarkeit (Sw1). Der Schüler hat seine Beweisideen an einem Beispiel erläutert und vor allem im abschließenden Satz die Verallgemeinerbarkeit derselben angedeutet. Die Bearbeitung lässt offen, dass und warum aus der Möglichkeit der Umformung von vier aufeinanderfolgenden ungeraden Zahlen zu vier gleichen (ganzen) Zahlen die gewünschte Teilbarkeit durch vier folgt. Ansonsten erscheint es vor dem Hintergrund der guten sachinhaltlichen Qualität der Ausführungen als nächster Schritt sinnvoll, die Ablösung der Überlegungen von den spezifischen Besonderheiten des Beispiels und eine Verallgemeinerung anzuregen. Dabei ist zu bedenken, dass mit der Verallgemeinerung eine vom Lernenden eventuell noch nicht vollzogene Ausschärfung der am Beispiel dargestellten Überlegungen einhergehen muss. Zudem können größere Schwierigkeiten bei der Versprachlichung auftreten, wie es die im Beweisversuch vorkommenden, nicht optimal passenden Begriffe „Ergebnise“ und „Ziffer“ vermuten lassen.

Insgesamt betrachtet können die in Beweisversuchen mit begründendem Ansatz angegebenen Beispiele aus der Sicht von Lesenden verschiedenartige Funktionen haben. Sie können z. B. darauf hinweisen, wie die zu zeigende Aussage interpretiert wurde, ein notwendiger Bestandteil des Argumentgefüges sein, allgemein gehaltene Ausführungen veranschaulichen, ergänzen oder präzisieren und Anknüpfungspunkte bieten, um inhaltliche Lücken im Argumentgefüge zu schließen oder Fehler zu korrigieren. Beispiele in Beweisversuchen mit begründendem Ansatz können nahelegen, dass der Beweisversuch zu einem bedeutsamen Anteil mithilfe von Beispielen entwickelt wurde. An Beispielen ausgeführte Überlegungen können konstitutiv für die Einordnung einer Bearbeitung in die Kategorie *begründender Ansatz* sein. Welche Funktion Beispielen in Beweisversuchen mit begründendem Ansatz aus der Sicht von Lesenden zukommt, hängt teilweise mit der Schlussweise zusammen.

8.2.3.4 Zusammenfassung

Die Analyse der Beispiele bei Beweisversuchen mit begründendem Ansatz zeigt, dass in allen Klassenstufen in jeder sechsten bis hin zu jeder dritten Bearbeitung mit begründendem Ansatz die infrage stehende Aussage für einen Spezialfall bestätigt wird. Je Klassenstufe kommt in über der Hälfte der Beweisversuche mit begründendem Ansatz irgendein Beispiel vor.

In den Beweisversuchen mit begründendem Ansatz und Beispiel sind meistens zwei bis fünf Beispiele angeführt. In den Klassenstufen 7, 8 sind im Vergleich zu den Klassenstufen 5, 6 wesentlich häufiger Reihenfolgebeispiele angegeben. Ein großes Beispiel kommt in jeder der Klassenstufen 5 bis 8 eher selten in einem Beweisversuch mit begründendem Ansatz vor.

In allen Klassenstufen befinden sich die Beispiele meistens ober- oder innerhalb des (übrigen) Argumentgefüges. Dass an einem oberhalb angeführten Beispiel eine Beweisidee hervorgehoben ist, trifft je Klassenstufe bei weniger als einem Viertel der Beweisversuche mit begründendem Ansatz und einem Beispiel oberhalb des (übrigen) Argumentgefüges zu.

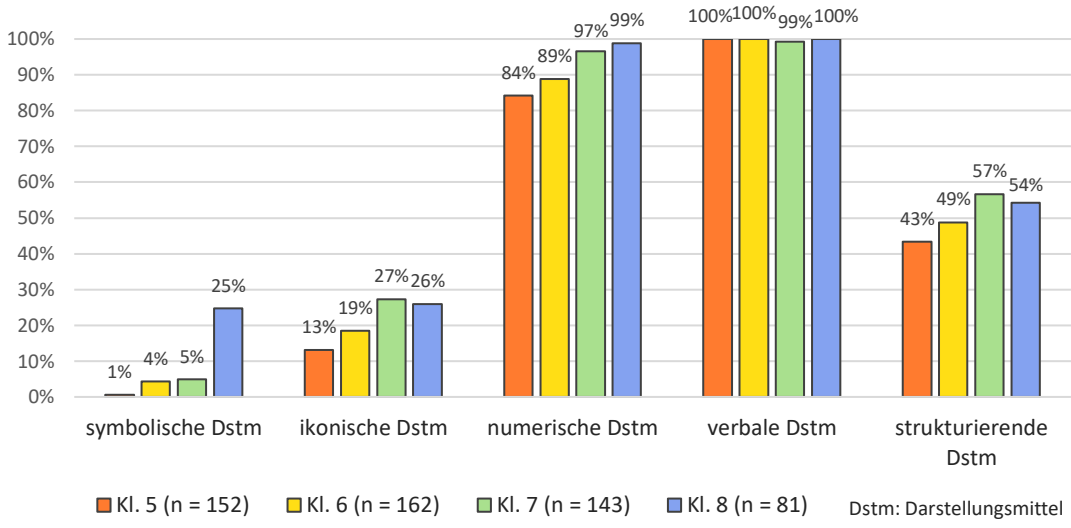
Die in Beweisversuchen mit begründendem Ansatz angegebenen Beispiele können aus der Sicht von Lesenden z. B. allgemein gehaltene Ausführungen veranschaulichen, ergänzen oder präzisieren und Anknüpfungspunkte bieten, um ein Argumentgefüge weiterzuentwickeln.

8.2.4 Darstellungsmittel und Darstellungsformen

Dieser Abschnitt beantwortet die Frage, welche Darstellungsmittel in den analysierten Beweisversuchen mit begründendem Ansatz vorkommen und was deren Darstellungsform ist.[139] Mit Blick auf die symbolischen Darstellungsmittel wird ergänzend dargelegt, inwieweit Buchstabenvariablen angemessen definiert sind. Das zugrundeliegende Kategoriensystem ist in Abschnitt 3.3.11 zu finden.

8.2.4.1 Vorkommende Darstellungsmittel

Dieser Abschnitt fokussiert auf die Frage, welche Darstellungsmittel in Beweisversuchen mit begründendem Ansatz vorkommen. Die Abbildung 110 zeigt die Anteile an Beweisversuchen mit symbolischen, ikonischen, numerischen, verbalen bzw. strukturierenden Darstellungsmitteln an allen Beweisversuchen mit begründendem Ansatz differenziert nach Klassenstufen.



p-Werte der Signifikanztests und dazugehörige Effektstärken

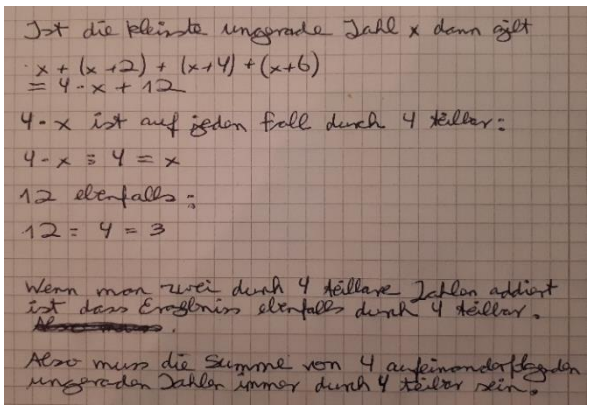
	symbolisch Dstm		ikonische Dstm		numerische Dstm
Kl. 8 vs. Kl. 5	$p < .001^{N-1}, V = .400$	Kl. 7 vs. Kl. 5	$p = .002^{N-1}, V = 176$	Kl. 7 vs. Kl. 5	$p < .001^{N-1}, V = .206$
Kl. 8 vs. Kl. 6	$p < .001^{N-1}, V = .306$	Kl. 8 vs. Kl. 5	$p = .015^{N-1}, V = 160$	Kl. 8 vs. Kl. 5	$p < .001^{N-1}, V = .224$
Kl. 8 vs. Kl. 7	$p < .001^{N-1}, V = .292$			Kl. 7 vs. Kl. 6	$p = .012^{N-1}, V = 144$
			strukturierende Dstm	Kl. 8 vs. Kl. 6	$p = .007^{N-1}, V = 173$
		Kl. 7 vs. Kl. 5	$p = .023^{N-1}, V = 132$		

Abbildung 110: Beweisversuche mit begründendem Ansatz: Vorkommende Darstellungsmittel. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

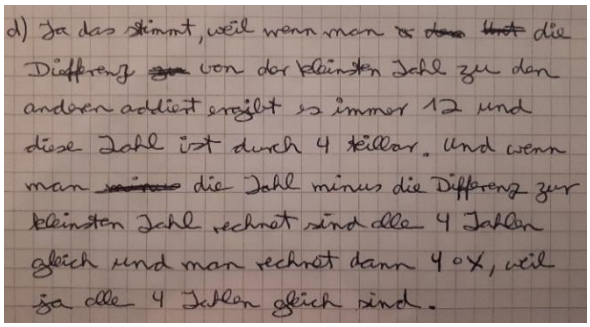
[139] Bei den Ausführungen zu Darstellungsmitteln und Darstellungsformen ist zu bedenken, dass es sich primär um Einblicke handelt, obwohl bei der Berechnung der Intercoderreliabilität fast durchgehend sehr gute Werte herausgekommen sind (Abschn. 3.3.11, 7.3.7).

Symbolische Darstellungsmittel sind in den Klassenstufen 5 bis 7 in bis zu fünf Prozent der Beweisversuche mit begründendem Ansatz zu finden. In der Klassenstufe 5 gibt es einen Beweisversuch, in dem Kästchen als Platzhalter für eine Zahl verwendet werden. Mit einer eigensprachlichen Ressource wurde eine symbolische Darstellung angegeben, die noch nicht in konventionalisierter Form bekannt ist. Buchstabenvariablen, wie sie üblicherweise verwendet werden, und Symbole mit fester Bedeutung, wie etwa Quantoren und das Zeichen für Teilbarkeit, die den Lernenden nicht unbedingt aus der Schule vertraut sind, kommen in keinem der analysierten Beweisversuche von Fünftklässlerinnen und Fünftklässlern vor. In den Klassenstufen 6, 7 werden in fast allen Beweisversuchen mit begründendem Ansatz und symbolischen Dar-

Bearbeitung 8.2.4a: Klassenstufe 6, männlich (Code: 0610116)



Bearbeitung 8.2.4b: Klassenstufe 6, männlich (Code: 0610455)



Bearbeitung 8.2.4c: Klassenstufe 6, männlich (Code: 0622009)

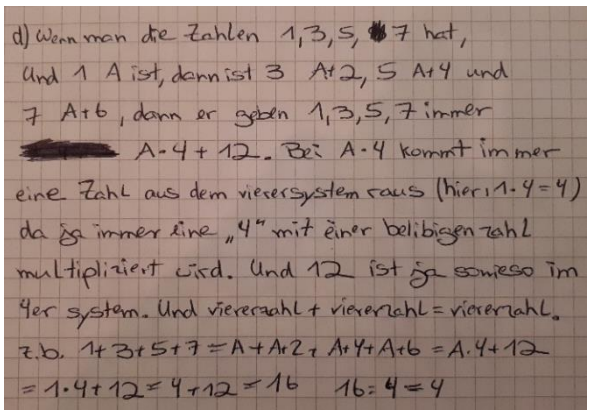


Abbildung 111: Bearbeitungen 8.2.4a-c (Beweisversuche mit begründendem Ansatz und Buchstabenvariablen, die am Beispiel der Klassenstufe 6 die Unterschiedlichkeit der Variablennutzung veranschaulichen)

stellungsmitteln Variablen genutzt, als Symbole mit fester Bedeutung sind ausschließlich Folgerungspfeile zu finden. In beiden Klassenstufen unterscheiden sich die Beweisversuche mit begründendem Ansatz, in denen eine Buchstabenvariable vorkommt, hinsichtlich des Umfangs der Nutzung von Variablen und des dabei angelegten Abstraktionsgrades deutlich (Bearbeitungen 8.2.4a-c in Abb. 111). Auch hinsichtlich der Existenz und Korrektheit von Definitionen der verwendeten Variablen zeigt sich in beiden Klassenstufen 6, 7 ein gemischtes Bild, das von Beweisversuchen ohne Variablendefinition bis hin zu solchen mit korrekten Definitionen aller Variablen reicht.

In der Klassenstufe 8 ist der Anteil an Beweisversuchen mit symbolischen Darstellungsmitteln an allen Beweisversuchen mit begründendem Ansatz statistisch hoch signifikant größer als in jeder der Klassenstufen 5 bis 7 (Abb. 110). In der Klassenstufe 8 kommen in einem Viertel der Beweisversuche mit begründendem Ansatz symbolische Darstellungsmittel vor. In 85% dieser Beweisversuche ist mindestens eine Variable enthalten. Die angegebenen Variablen sind teils angemessen definiert, teils ansatzweise definiert, teils nicht definiert. Symbole mit fester Bedeutung enthalten in der Klassenstufe 8 gut 5% aller Beweisversuche mit begründendem Ansatz. Meistens handelt es sich um Folgerungspfeile. In einem der analysierten Beweisversuche kommt das zahlentheoretische Symbol für Kongruenz vor, in einem anderen kommen die Symbole $\in$ und $\mathbb{N}$ vor.

Der Anteil der Beweisversuche mit **ikonischen Darstellungsmitteln** an allen Beweisversuchen mit begründendem Ansatz liegt in der Klassenstufe 5 bei 13%, in der Klassenstufe 6 bei 18% und in den Klassenstufen 7, 8 bei leicht mehr als 25%. Er ist in den Klassenstufen 7, 8 ca. doppelt so groß wie in der Klassenstufe 5 und statistisch signifikant größer (Abb. 110). Dabei zeigt sich hinsichtlich der Art der ikonischen Darstellungsmittel in jeder der Klassenstufen ein ähnliches Bild. Je Klassenstufe wurden in rund 35% bis 55% der Beweisversuche mit begründendem Ansatz und ikonischen Darstellungsmitteln ikonische Darstellungsmittel der Art *Klammern, Pfeile, Linien* verwendet. Ebenfalls in je rund 35% bis 55% der Fälle sind ikonische Darstellungsmittel der Art *Einkreisungen, Unterstreichungen o. Ä.* zu finden. Figurierte Zahlen kommen je Klassenstufe in maximal 5% der Beweisversuche mit begründendem Ansatz und ikonischem Darstellungsmittel vor.

Der Anteil an Beweisversuchen mit **numerischen Darstellungsmitteln** an allen Beweisversuchen mit begründendem Ansatz beträgt in den Klassenstufen 5, 6 etwa 85%. Dieser Anteil ist in den Klassenstufen 7, 8 mit jeweils über 95% statistisch signifikant größer als in den Klassenstufen 5, 6, wenn auch mit geringer Effektstärke (Abb. 110).

Bearbeitung 8.2.4d: Klassenstufe 7, männlich (Code: 0721811)

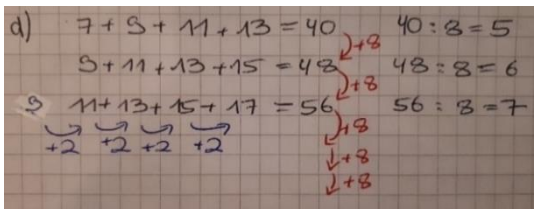


Abbildung 112: Bearbeitung 8.2.4d (Beweisversuch mit begründendem Ansatz, in dem ausschließlich numerische und ikonische Darstellungsmittel vorkommen)

Verbale Darstellungsmittel sind in fast allen analysierten Beweisversuchen mit begründendem Ansatz enthalten. Die einzige Ausnahme bildet der in Abbildung 112 zu sehende Beweisversuch mit begründendem Ansatz aus der Klassenstufe 7, in dem ausschließlich numerische und ikonische Darstellungsmittel vorkommen.

Der Anteil der Beweisversuche mit **strukturierenden Darstellungsmitteln** an allen Beweisversuchen mit begründendem Ansatz liegt je Klassenstufe zwischen 40% und 60%. Er ist in der fünften Klassenstufe mit 43% am kleinsten und in der siebten Klassenstufe mit 57% am größten. Dieser Unterschied zwischen den Klassenstufen 5, 7 ist statistisch signifikant (Abb. 110).

Mathematische Strukturierungselemente im Speziellen sind in der Klassenstufe 5 in keinem, in den Klassenstufen 6, 7 in rund 1% und in der Klassenstufe 8 in rund 7% der analysierten Beweisversuche mit begründendem Ansatz zu finden. Dabei ist in einzelnen Beweisversuchen aus der Klassenstufe 8 ein Beweisabschluss gekennzeichnet, während dies in den Klassenstufen 5 bis 7 nicht der Fall ist. Strukturierende Begriffe kommen je Klassenstufe in 25% bis 40% der Beweisversuche mit begründendem Ansatz vor, wobei die Begriffe „Beispiel“ und „Antwort“ sowie entsprechende Abkürzungen besonders häufig sind. Eine Nummerierung enthält keiner der analysierten Beweisversuche mit begründendem Ansatz aus der Klassenstufe 5. In der Klassenstufe 6 sind in rund 1% der Fälle Beispiele nummeriert. In den Klassenstufen 7, 8 liegt der Anteil der Beweisversuchen mit einer Nummerierung an allen Beweisversuchen mit begründendem Ansatz bei ca. 5%, wobei in manchen Beweisversuchen nicht Beispiele, sondern Lösungsschritte nummeriert sind.

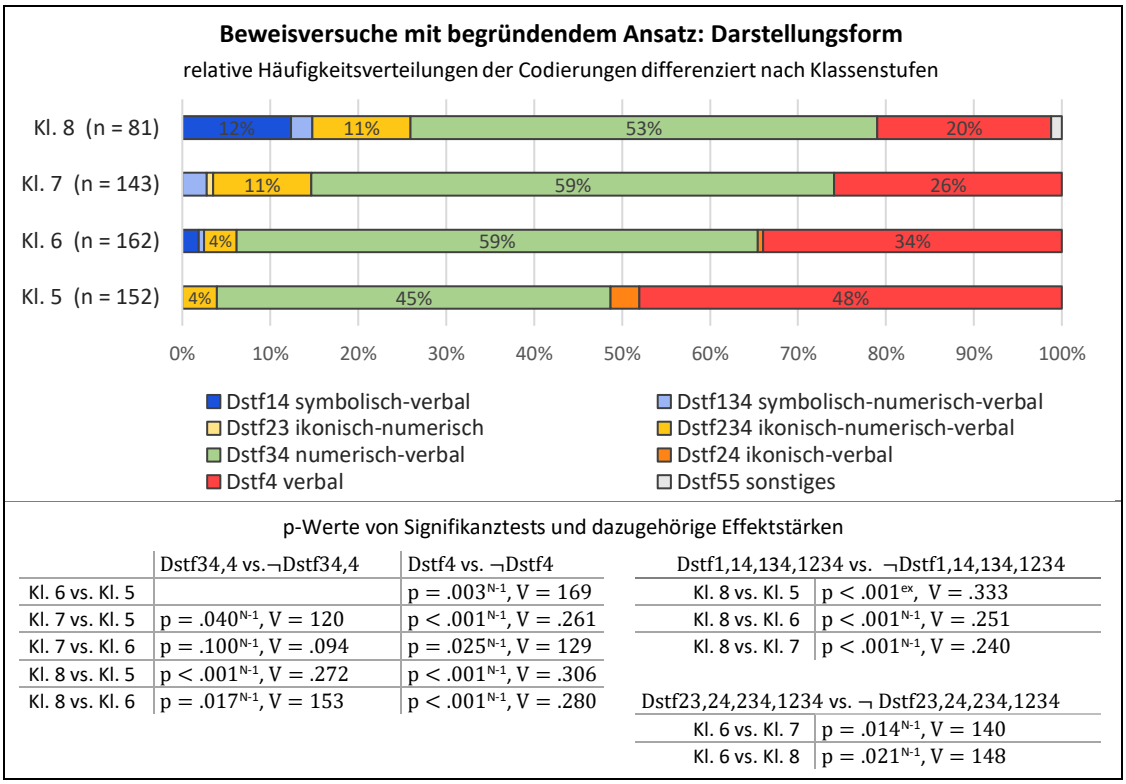
8.2.4.2 Darstellungsform

Darauf aufbauend, welche Darstellungsmittel in den Beweisversuchen mit begründendem Ansatz vorkommen, fokussiert dieser Abschnitt auf die Darstellungsform der Beweisversuche mit begründendem Ansatz. Die Abbildung 113 zeigt die diesbezüglichen relativen Häufigkeitsverteilungen der Codierungen.[140]

Beweisversuche mit der Darstellungsform ***verbal*** ($\mathrm{Dstf}4$) machen in der Klassenstufe 5 etwa die Hälfte, in der Klassenstufe 6 rund ein Drittel, in der Klassenstufe 7 ca. ein Viertel und in der Klassenstufe 8 rund ein Fünftel aller Beweisversuche mit begründendem Ansatz aus. Der Anteil der Beweisversuche mit der Darstellungsform *verbal* ist in der Klassenstufe 5 statistisch signifikant größer als in jeder der Klassenstufen 6 bis 8 und in der Klassenstufe 6 als in den Klassenstufen 7, 8.

In allen Klassenstufen 5 bis 8 ist die Darstellungsform in etwa der Hälfte der Beweisversuche mit begründendem Ansatz **numerisch-verbal** ($\mathrm{Dstf}34$) und in über 70% der Beweisversuche mit begründendem Ansatz **numerisch-verbal oder verbal** ($\mathrm{Dstf}34{,}4$). In den Klassenstufen 5, 6 ist dieser Anteil mit je knapp 95% am größten und statistisch signifikant größer mit $0{,}15 < \mathrm{V} < 0{,}28$ als in der Klassenstufe 8 mit knapp 75%.

[140] Der Übersichtlichkeit halber sind in Abbildung 113 ausschließlich Darstellungsformen separat aufgeführt, die in mehreren der analysierten Beweisversuche vorkommen. Alle anderen Darstellungsformen sind in der Kategorie *sonstiges* zusammengefasst.



p-Werte von Signifikanztests und dazugehörige Effektstärken

	Dstf34,4 vs.¬Dstf34,4	Dstf4 vs. ¬Dstf4
Kl. 6 vs. Kl. 5		$p = .003^{N-1}, V = 169$
Kl. 7 vs. Kl. 5	$p = .040^{N-1}, V = 120$	$p < .001^{N-1}, V = .261$
Kl. 7 vs. Kl. 6	$p = .100^{N-1}, V = .094$	$p = .025^{N-1}, V = 129$
Kl. 8 vs. Kl. 5	$p < .001^{N-1}, V = .272$	$p < .001^{N-1}, V = .306$
Kl. 8 vs. Kl. 6	$p = .017^{N-1}, V = 153$	$p < .001^{N-1}, V = .280$

Dstf1,14,134,1234 vs. ¬Dstf1,14,134,1234	
Kl. 8 vs. Kl. 5	$p < .001^{ex}, V = .333$
Kl. 8 vs. Kl. 6	$p < .001^{N-1}, V = .251$
Kl. 8 vs. Kl. 7	$p < .001^{N-1}, V = .240$

Dstf23,24,234,1234 vs. ¬ Dstf23,24,234,1234	
Kl. 6 vs. Kl. 7	$p = .014^{N-1}, V = 140$
Kl. 6 vs. Kl. 8	$p = .021^{N-1}, V = 148$

Abbildung 113: Beweisversuche mit begründendem Ansatz: Darstellungsform. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Dass die Darstellungsform **ikonisch** ist (Dstf2), trifft auf keinen der analysierten Beweisversuche mit begründendem Ansatz zu, die Darstellungsform hat in allen analysierten Beweisversuchen mit begründendem Ansatz maximal einen ikonischen Anteil. Die Beweisversuche mit einer Darstellungsform mit ikonischem Anteil (Dstf23,234,24,1234) machen in den Klassenstufen 5, 6 etwa 5% und in den Klassenstufen 7, 8 ca. 12% aller Beweisversuche mit begründendem Ansatz aus. Die Unterschiede zwischen der Klassenstufe 6 und den Klassenstufen 7, 8 sind statistisch signifikant.

Wenn man die in Abschnitt 8.2.4.1 beschriebenen Anteile der Beweisversuche mit ikonischen Darstellungsmitteln mitberücksichtigt, ist Folgendes zu sehen: In allen Klassenstufen ist in weniger als der Hälfte der Beweisversuche mit einem ikonischen Darstellungsmittel eine sinnbezogene Einheit überwiegend ikonisch dargestellt. Diese Diskrepanz liegt u. a. daran, dass in einigen Beweisversuchen ausschließlich Begriffe wie „immer“, „richtig“ oder „aufeinanderfolgend“ unterstrichen sind, was zwar die Relevanz einzelner Aspekte hervorhebt, jedoch keine eigene sinnbezogene Einheit darstellt.

Dass eine sinnbezogene Einheit überwiegend **symbolisch** dargestellt ist (Dstf1,14,134, 1234), trifft in der Klassenstufe 5 bei keinem der analysierten Beweisversuche zu. In den Klassenstufen 6, 7 ist dies in rund 3% der Beweisversuche mit begründendem Ansatz der Fall. Im Vergleich zu jeder der Klassenstufen 5 bis 7 ist in der Klassenstufe 8 mit rund 15% der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil an allen Beweisversuchen mit begründendem Ansatz statistisch hoch signifikant größer mit $0{,}24 \leq V < 0{,}34$.

8.2.4.3 Zusammenfassung

Hinsichtlich der **Darstellungsmittel** in Beweisversuchen mit begründendem Ansatz lassen sich im Wesentlichen folgende Tendenzen festhalten: In so gut wie allen Beweisversuchen mit begründendem Ansatz kommen verbale Darstellungsmittel vor. In den meisten Fällen sind zusätzlich numerische Darstellungsmittel zu finden, während ikonische und insbesondere symbolische Darstellungsmittel wesentlich seltener sind. Strukturierende Darstellungsmittel wurden in etwa der Hälfte der Beweisversuche mit begründendem Ansatz verwendet, oftmals sind strukturierende Begriffe angegeben. Die **Darstellungsform** ist je Klassenstufe bei etwa der Hälfte der Beweisversuche mit begründendem Ansatz numerisch-verbal. In den Klassenstufen 5, 6 sind zudem Beweisversuche mit der Darstellungsform *verbal* besonders häufig, während diese in den Klassenstufen 7, 8 statistisch signifikant seltener vorkommen. In der Klassenstufe 7 machen ikonisch-numerisch-verbal dargestellte Beweisversuche, in der Klassenstufe 8 ikonisch-numerisch-verbal sowie symbolisch-numerisch-verbal dargestellte Beweisversuche jeweils mehr als ein Zehntel der Beweisversuche mit begründendem Ansatz aus.

8.2.5 Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage

Nachdem die Beweisversuche mit begründendem Ansatz nacheinander hinsichtlich ihrer Korrektheit und Vollständigkeit, der Schlussweise, der Beweisideen, der Verwendung von Beispielen, der Darstellungsmittel und der Darstellungsform beschrieben wurden, werden im vorliegenden und im nachfolgenden Abschnitt mehrere Eigenschaften dieser Beweisversuche zusammen betrachtet, um die Beweisversuche mit begründendem Ansatz abschließend zu klassifizieren. Die Beweisversuche werden klassifiziert und nicht typisiert, weil eine Typenbildung erfordern würde, Beweisversuche möglichst differenten Typen zuzuordnen, obwohl die Beweisversuche mit begründendem Ansatz ein Kontinuum an möglichen Ausprägungen aufweisen. Außerdem impliziere die Bildung weniger Typen, Beweisversuche so sehr zusammenfassen, dass etwa mit Blick auf den Nachweis der Gültigkeit wesentliche Informationen verloren gingen. Mit einer Klassifikation ist es möglich, mit dem Kontinuum an vorhandenen Ausprägungen angemessen umzugehen, alle relevanten Ausprägungen einzubeziehen und trotzdem Ergebnisse zu verdichten, zu strukturieren sowie abschließend zu resümieren (Abschn. 7.2.1.6, 7.3.8 zu Typenbildungen und Klassifikationen).

Im vorliegenden Abschnitt liegt der Fokus auf der Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Diese Klassifikation wird zuerst eingeführt und begründet (Abschn. 8.2.5.1). Danach lässt sich anhand der Häufigkeitsverteilungen der gebildeten Gruppen aufzeigen, wie die Beweisversuche von Lernenden in den verschiedenen Klassenstufen in Hinblick auf den Nachweis der Gültigkeit ausgeprägt sind, was für die einzelnen Klassenstufen diesbezüglich typische Beweisversuche auszeichnet und welche Arten an Beweisversuchen eher Ausnahmen darstellen (Abschn. 8.2.5.2). Am Ende ist eine kurze Zusammenfassung zu finden (Abschn. 8.2.5.3).

8.2.5.1 Einführung und Begründung der Klassifikation

Als Beweise werden im Rahmen dieser Arbeit alle korrekten und vollständigen, deduktiven Argumentgefüge angesehen (Abschn. 3.4). Um Einblicke zu geben, inwieweit es den Lernenden, die einen Beweisversuch mit begründendem Ansatz abgegeben haben, gelungen ist, die zu zeigende Aussage zu beweisen, wurden dementsprechend die Beweisversuche mit begründendem Ansatz in Abschnitt 8.2.1 hinsichtlich ihrer Korrektheit, Vollständigkeit und Schlussweise gruppiert. Dabei wurde zwischen den sechs Kategorien KV1 *kein relevanter Beitrag zu einem weiterentwickelbaren Argumentgefüge* bis KV6 *korrektes und vollständiges Argumentgefüge* sowie den Schlussweisen Sw1-2 *zwischen induktiv und deduktiv* und Sw3-4 *deduktiv* differenziert, so dass aus der Verknüpfung der Ausprägungen zwölf Gruppen resultieren, die Gruppen KV1, Sw1-2 bis KV6, Sw3-4.

Für die Klassifikation der Beweisversuche mit Blick auf den Nachweis der Gültigkeit (Abb. 114) werden von diesen Gruppen diejenigen mit der Merkmalsausprägung KV1-2 *allenfalls Ansatz für ein weiterentwickelbares Argumentgefüge* zusammengefasst. Mit Blick auf den Nachweis der Gültigkeit macht es keinen großen Unterschied, ob ein Beweisversuch die Ausprägung KV1

<table>
<tr><th></th><th>Sw1-2 zwischen induktiv und deduktiv</th><th>Sw3-4 deduktiv</th></tr>
<tr><td>KV5-6 weitgehend korrektes und vollständiges Argumentgefüge
PzzA3 Passung zu der zz. Aussage</td><td>NwG3.5: KV5-6,¬(KV6,Sw3-4)
weitgehend korrektes und vollständiges Argumentgefüge, aber kein Beweis</td><td>NwG3.6: KV6, Sw3-4
korrektes und vollständiges, deduktives Argumentgefüge</td></tr>
<tr><td>KV4 weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke
PzzA3 Passung zu der zz. Aussage</td><td colspan="2">NwG3.4: KV4
weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke</td></tr>
<tr><td>KV3 weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler
PzzA1-3</td><td colspan="2">NwG3.3: KV3
weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler</td></tr>
<tr><td rowspan="3">KV1-2 kein weiterentwickelbares Argumentgefüge
PzzA1-3</td><td colspan="2">NwG3.2: KV1-2, PBsp5
kein weiterentwickelbares Argumentgefüge,
aber korrekt und vollständig ausgeführtes eigenes Beispiel zu der zz. Aussage</td></tr>
<tr><td colspan="2">NwG3.1: KV1-2, PBsp3-4
kein weiterentwickelbares Argumentgefüge,
Beispiel zu der zz. Aussage, aber kein korrekt, vollständig ausgeführtes eigenes Beispiel</td></tr>
<tr><td colspan="2">NwG3.0: KV1-2, PBsp1-2
kein weiterentwickelbares Argumentgefüge
und kein Beispiel zu der zz. Aussage</td></tr>
</table>

Abbildung 114: Beweisversuche mit begründendem Ansatz: Nachweis der Gültigkeit. Klassifikation der Beweisversuche

oder KV2 aufweist, weil in beiden Fällen kein weiterentwickelbares Argumentgefüge vorkommt. Weiterhin wird bei den Beweisversuchen mit KV1-2 nicht zwischen den Schlussweisen unterschieden, weil diejenigen mit deduktiver Schlussweise mit Blick auf den Nachweis der Gültigkeit nicht unbedingt „besser" sind. Anstelle einer Differenzierung zwischen KV1 und KV2 sowie den Schlussweisen Sw1-2 und Sw3-4 wird mangels eines weiterentwickelbaren Argumentgefüges berücksichtigt, inwieweit die Ausführungen die zu zeigende Aussage für einen Spezialfall bestätigen. In Abhängigkeit der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage werden die Gruppen NwG3.0, NwG3.1 und NwG3.2 gebildet, wobei in Beweisversuchen der Art NwG3.2 im Vergleich zu denen aus den Gruppen NwG3.0-3.1 ein korrekt und vollständig ausgeführtes, eigenes Beispiel zu der zu zeigenden Aussage vorkommt, d. h. die Aussage für einen Spezialfall bestätigt wird.[141]

Beweisversuche mit einer der Ausprägungen KV3-6 enthalten ein weiterentwickelbares Argumentgefüge. Von diesen Beweisversuchen werden zunächst diejenigen mit der Merkmalsausprägung KV3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler* separiert, weil sie sachinhaltlich fehlerhaft und dadurch, unabhängig von der Schlussweise, mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage klar unzureichend sind. Die in Abschnitt 8.2.1 gebildeten Gruppen KV3,Sw1-2 und KV3, Sw3-4 werden also zu der Gruppe NwG3.3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler* zusammengefasst.

Bei den Beweisversuchen mit einem weitgehend korrekten Argumentgefüge (KV4-6) hängt die Einordnung in eine der Kategorien KV4, KV5, KV6 von den festgelegten Erwartungen an ein korrektes und vollständiges Argumentgefüge ab (Abschn. 3.3.4). Beweisversuche mit der Ausprägung KV4 *weitgehend korrektes mehrschrittiges Argumentgefüge mit inhaltlicher Lücke* sind im Vergleich zum relevanten Erwartungshorizont sachinhaltlich unvollständig, und reichen unabhängig von der Schlussweise nicht für den Nachweis der Gültigkeit aus. Sie werden daher alle zu der Gruppe NwG3.4 zusammengefasst.[142]

Beweisversuche mit einer der Merkmalsausprägungen KV5 und KV6 enthalten im Vergleich zum relevanten Erwartungshorizont allenfalls Unzulänglichkeiten in der Formulierung, weshalb diese Beweisversuche hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges als sehr gut angesehen werden können. Allerdings gelten von diesen Beweisversuchen nur diejenigen mit einem korrekten, vollständigen und zudem deduktiven Argumentgefüge als Beweis der zu zeigenden Aussage (Abschn. 3.4). Mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage werden dementsprechend die Beweisversuche mit einem korrekten und vollständigen, deduktiven Argumentgefüge bzw. der Merkmalskombination KV6, Sw3-4 als separate Gruppe NwG3.6 und alle übrigen Beweisversuche mit der Merkmalsausprägung KV5-6 als Gruppe NwG3.5 ausgewiesen.

[141] In den Gruppenbezeichnungen, wie z. B. „NwG3.0", gibt die linke Ziffer „3" an, dass es sich um Beweisversuche aus der Oberkategorie Ans3 *begründender Ansatz* handelt, während die rechte Ziffer, wie z. B. „ 0" bei „NwG3.0" die Ausprägung dieser Beweisversuche mit Blick auf den Nachweis der Gültigkeit beschreibt.

[142] Nebenbei machen Beweisversuche in der Gruppe KV4, Sw3-4 ausschließlich marginale Anteile an allen Beweisversuchen mit begründendem Ansatz aus (Abschn. 8.2.1), so dass es auch aus dieser Perspektive als wenig sinnvoll erscheint, sie als separate Gruppen auszuweisen.

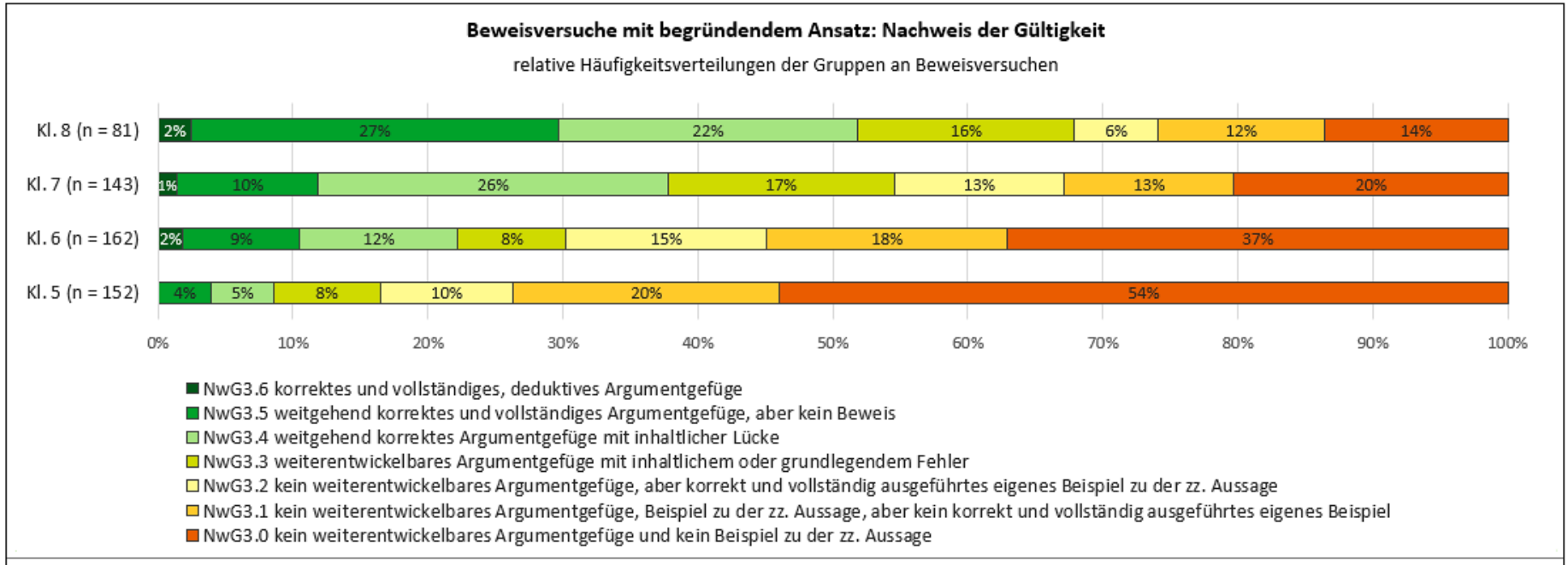
8.2.5.2 Relative Häufigkeitsverteilungen der Beweisversuche

Die relativen Häufigkeitsverteilungen der mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage gebildeten Gruppen an Beweisversuchen mit begründendem Ansatz unterscheiden sich zwischen den **Klassenstufen 5 bis 8** statistisch signifikant (Abb. 115). Der paarweise Vergleich der Häufigkeitsverteilungen mittels des Mann-Whitney-U-Tests zeigt, dass in der jeweils höheren Klassenstufe die Wahrscheinlichkeit statistisch signifikant größer ist, eine Zuordnung zu einer besseren Gruppe zu erhalten, bei geringer bis hoher Effektstärke.[143]

In der **Klassenstufe 5** enthalten über 50% der Beweisversuche mit begründendem Ansatz allenfalls einen Ansatz für ein weiterentwickelbares Argumentgefüge und kein Beispiel zu der zu zeigenden Aussage ($\mathrm{NwG3.1.1}$). Der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge ($\mathrm{NwG3.4\text{-}3.6}$) an allen Beweisversuchen mit begründendem Ansatz liegt unter 10%. Im Vergleich zu der Klassenstufe 5 ist in der **Klassenstufe 6** der Anteil der Beweisversuche in der Gruppe $\mathrm{NwG3.1.1}$ mit etwa einem Drittel anstatt über der Hälfte der Zuordnungen statistisch hoch signifikant kleiner. Der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge ($\mathrm{NwG3.4\text{-}3.6}$) ist mit knapp einem Viertel im Vergleich zu unter 10% statistisch hoch signifikant größer. In der **Klassenstufe 7** enthält ein Fünftel und damit ein im Vergleich zu beiden Klassenstufen 5, 6 statistisch hoch signifikant kleinerer Anteil der Beweisversuche mit begründendem Ansatz weder ein Beispiel zu der zu zeigenden Aussage noch ein weiterentwickelbares Argumentgefüge ($\mathrm{NwG3.1.1}$). Der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge ($\mathrm{NwG3.4\text{-}3.6}$) an allen Beweisversuchen mit begründendem Ansatz ist mit über einem Drittel der Zuordnungen statistisch hoch signifikant größer als in beiden Klassenstufen 5, 6. In der **Klassenstufe 8** ist insbesondere der Anteil der Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge ($\mathrm{NwG3.5\text{-}3.6}$) mit knapp 30% im Vergleich zu unter 15% statistisch hoch signifikant größer als in jeder der Klassenstufen 5 bis 7. Vereinfacht zusammengefasst deuten die Ergebnisse also folgende Tendenz an: Je höher die Klassenstufe, desto besser sind die Beweisversuche mit begründendem Ansatz in Bezug auf den Nachweis der Gültigkeit der zu zeigenden Aussage.

Dieser Tendenz zum Trotz ist ebenfalls festzuhalten, dass **in jeder der Klassenstufen 5 bis 8** mindestens sechs der sieben unterschiedenen Gruppen vorkommen. Dies bedeutet, dass das Spektrum an Beweisversuchen mit begründendem Ansatz hinsichtlich des Nachweises der Gültigkeit in jeder der Klassenstufen 5 bis 8 groß ist. Es gibt Teilnehmende aus der Klassenstufe 5, die ein weitgehend korrektes und vollständiges Argumentgefüge formuliert haben, während in manchen Beweisversuchen aus der Klassenstufe 8 allenfalls ein Ansatz für ein weiterentwickelbares Argumentgefüge formuliert ist und kein Beispiel zu der zu zeigenden Aussage vorkommt. Der Anteil der Beweisversuche mit einem korrekten und vollständigen, deduktiven Argumentgefüge ($\mathrm{NwG3.6}$) an allen Beweisversuchen mit begründendem Ansatz, d. h. der Anteil der Beweisversuche, die die zu zeigende Aussage beweisen, liegt je Klassenstufe unter 3%.

[143] Die nun folgende nähere Beschreibung der relativen Häufigkeitsverteilungen wird kurzgehalten, weil die erkennbaren Tendenzen teils aus den bereits in den Abschnitten 8.2.1, 8.2.3 beschriebenen Ergebnissen der Analyse der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges, der Schlussweisen und der Verwendung von Beispielen hervorgehen.



p-Werte von Signifikanztests und dazugehörige Effektstärken

	Alle einzelnen Gruppen	NwG3.1.1 vs. ¬NwG3.1.1	NwG3.1 vs. ¬NwG3.1	NwG3.1-3.2 vs. NwG3.3-3.6	NwG3.1-3.3 vs. NwG3.4-3.6	NwG3.1-3.4 vs. NwG3.5-3.6
Kl. 5 vs. Kl. 6	$p < .001^{U}$, $r = .208$	$p = .003^{N\text{-}1}$, $V = .170$	$p = .001^{N\text{-}1}$, $V = .195$	$p = .004^{N\text{-}1}$, $V = .162$	$p = .001^{N\text{-}1}$, $V = .188$	$p = .026^{N\text{-}1}$, $V = .126$
Kl. 5 vs. Kl. 7	$p < .001^{U}$, $r = .429$	$p < .001^{N\text{-}1}$, $V = .347$	$p < .001^{N\text{-}1}$, $V = .409$	$p < .001^{N\text{-}1}$, $V = .399$	$p < .001^{N\text{-}1}$, $V = .348$	$p = .011^{N\text{-}1}$, $V = .148$
Kl. 5 vs. Kl. 8	$p < .001^{U}$, $r = .519$	$p < .001^{N\text{-}1}$, $V = .393$	$p < .001^{N\text{-}1}$, $V = .460$	$p < .001^{N\text{-}1}$, $V = .516$	$p < .001^{N\text{-}1}$, $V = .486$	$p < .001^{N\text{-}1}$, $V = .365$
Kl. 6 vs. Kl. 7	$p < .001^{U}$, $r = .223$	$p = .001^{N\text{-}1}$, $V = .184$	$p < .001^{N\text{-}1}$, $V = .222$	$p < .001^{N\text{-}1}$, $V = .246$	$p = .003^{N\text{-}1}$, $V = .170$	$p = .700^{N\text{-}1}$, $V = .022$
Kl. 6 vs. Kl. 8	$p < .001^{U}$, $r = .339$	$p < .001^{N\text{-}1}$, $V = .243$	$p < .001^{N\text{-}1}$, $V = .275$	$p < .001^{N\text{-}1}$, $V = .359$	$p < .001^{N\text{-}1}$, $V = .486$	$p < .001^{N\text{-}1}$, $V = .365$
Kl. 7 vs. Kl. 8	$p = .010^{U}$, $r = .173$	$p = .209^{N\text{-}1}$, $V = .084$	$p = .279^{N\text{-}1}$, $V = .073$	$p = .051^{N\text{-}1}$, $V = .131$	$p = .041^{N\text{-}1}$, $V = .137$	$p = .001^{N\text{-}1}$, $V = .220$

Abbildung 115: Beweisversuche mit begründendem Ansatz: Nachweis der Gültigkeit. Relative Häufigkeitsverteilungen der Gruppen an Beweisversuchen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

8.2.5.3 Zusammenfassung

Eine Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage bietet die Chance, die diesbezüglichen Ergebnisse der detaillierten Analyse der Beweisversuche mit begründendem Ansatz zusammenzuführen, zu verdichten, zu strukturieren und zu resümieren. In der mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage gebildeten Klassifikation werden je nach Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges, Schlussweise und eventuell der Existenz, Passung eines Beispiels zu der zu zeigenden Aussage sieben ordinal skalierte Gruppen an Beweisversuchen mit begründendem Ansatz differenziert.

Die relativen Häufigkeitsverteilungen dieser Gruppen an Beweisversuchen deuten vereinfacht gesagt an, dass die Beweisversuche mit begründendem Ansatz in Bezug auf den Nachweis der Gültigkeit der zu zeigenden Aussage im Mittel mit höherer Klassenstufe besser sind. Dabei liegt jedoch der Anteil der Beweisversuche mit einem korrekten und vollständigen, deduktiven Argumentgefüge an allen Beweisversuchen mit begründendem Ansatz, d. h. der Anteil der Beweisversuche, die die zu zeigende Aussage beweisen, je Klassenstufe unter 3%. Das Spektrum an vorkommenden Gruppen ist in jeder der Klassenstufen 5 bis 8 groß.

8.2.6 Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf deren Weiterentwicklung zu einem Beweis

Dieser Abschnitt zielt auf eine Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Dazu werden die Beweisversuche mit begründendem Ansatz zuerst anhand zweier Merkmale, der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise, gruppiert, weil die Ausprägungen dieser beiden Merkmale mit Blick auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis von grundlegender Bedeutung sind (Abschn. 8.2.6.1). Auf dieser Gruppierung aufbauend werden Zusammenhänge zwischen den genannten zwei Merkmalen und weiteren Eigenschaften der Beweisversuche, wie z. B. der primären Beweisidee und der Darstellungsform, präsentiert (Abschn. 8.2.6.2 – 8.2.6.5).[144] Unter Berücksichtigung dieser Zusammenhänge werden die Beweisversuche anschließend bezüglich ihrer Anknüpfungspunkte für die Weiterentwicklung zu einem Beweis klassifiziert (Abschn. 8.2.6.6). Die Klassifikation beschreibt eine Möglichkeit, die Beweisversuche mit begründendem Ansatz zusammenfassend mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung einzuordnen. Zudem ermöglicht sie aufzuzeigen, was diesbezüglich in den einzelnen Klassenstufen 5 bis 8 häufige Ausprägungen sind, und welche wesentlichen Gemeinsamkeiten und Unterschiede zwischen den Klassenstufen bestehen (Abschn. 8.2.6.7). Der umfangreiche Abschnitt 8.2.6 wird abschließend zusammengefasst (Abschn. 8.2.6.8).

[144] Während mögliche Zusammenhänge zwischen verschiedenen Eigenschaften der Beweisversuche mit begründendem Ansatz systematisch und umfassend untersucht wurden, werden im Fließtext top-down bestimmte Eigenschaften der Beweisversuche mit begründendem Ansatz fokussiert und deren Zusammenspiel präsentiert, um den Umfang der Ausführungen in einem vertretbaren Rahmen zu halten. Weiterhin liegt der Fokus in den Abschnitten 8.2.6.2 bis 8.2.6.5 auf dem Zusammenspiel der verschiedenen Eigenschaften, nicht auf deren Häufigkeiten in den einzelnen Klassenstufen. Wenn bedeutende Unterschiede zwischen den Klassenstufen 5, 6, 7, 8 bestehen, wird dies im Fließtext erwähnt.

8.2.6.1 Grundlegende Gruppierung: Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und die unterschiedenen Schlussweisen

Die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und die Schlussweise sind für die Beschreibung von Anknüpfungspunkten für die Weiterentwicklung von Beweisversuchen mit begründendem Ansatz zu einem Beweis von grundlegender Bedeutung, weil sich Beweise insbesondere durch ein korrektes und vollständiges Argumentgefüge und deduktives Schließen charakterisieren (Abschn. 3.4). In Abschnitt 8.2.1 wurden die Beweisversuche mit begründendem Ansatz bereits in Abhängigkeit der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise gruppiert.

Im Vergleich zu dieser Gruppierung werden für die Klassifikation der Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis die Ausprägungen KV5 *weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung* und KV6 *korrektes und vollständiges Argumentgefüge* zusammengefasst, da je Klassenstufe unter 5% der Beweisversuche mit begründendem Ansatz ein korrektes und vollständiges Argumentgefüge (KV6) enthalten und zudem Beweisversuche mit beiden Ausprägungen allenfalls kleine Änderungen, Ergänzungen oder Verallgemeinerungen dargestellter Überlegungen erfordern, um als Beweis zu gelten.

Weiterhin wird im Vergleich zu der Gruppierung in Abschnitt 8.2.1 zwischen den Schlussweisen Sw3 *deduktiv mit zusätzlichem Beispiel* und Sw4 *deduktiv ohne zusätzliches Beispiel* differenziert, weil mit diesen beiden Schlussweisen unterschiedliche Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis verknüpft sein können.

Insgesamt werden somit je nach Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und Schlussweise die Gruppen KV1, Sw1-2 bis KV5-6, Sw4 gebildet. Um herauszuarbeiten, wie sich diese Gruppen für die Klassifikation der Beweisversuche weiter zusammenfassen lassen, und andererseits, welche Binnendifferenzierungen sinnvoll sind, werden im Folgenden weitere Eigenschaften der Beweisversuche mit begründendem Ansatz in Zusammenhang mit der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweisen betrachtet. Diese Eigenschaften sind die Passung der Ausführungen zu der zu zeigenden Aussage, die primäre Beweisidee, die Existenz, die Art und der Ort von Beispielen sowie die Darstellungsform.

8.2.6.2 Detailanalyse: Passung zu der zu zeigenden Aussage

In diesem Abschnitt wird in Abhängigkeit der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise betrachtet, inwieweit die Ausführungen in Beweisversuchen mit begründendem Ansatz zu der zu zeigenden Aussage passen oder auf einen grundlegenden Fehler hinweisen. Dazu sind in Abbildung 116 relative Häufigkeitsverteilungen von Codierungen der Beweisversuche mit begründendem Ansatz dargestellt.[145]

[145] Bei der Betrachtung einzelner Eigenschaften der Beweisversuche mit begründendem Ansatz in Abhängigkeit der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise wird jeweils mithilfe einer Abbildung dargestellt, wie diese Eigenschaften bei den Beweisversuchen in den einzelnen Gruppen KV1, Sw1-2 bis KV5-6, Sw4 ausgeprägt sind, um Muster oder Ausnahmen zu erkennen und näher analysieren zu können. Bei der Interpretation der Abbildungen ist zu beachten, dass die absoluten Anzahlen an Beweisversuchen

Abbildung 116: Beweisversuche mit begründendem Ansatz: Passung zu der zu zeigenden Aussage in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

in den einzelnen Gruppen KV1, Sw1-2 bis KV5-6, Sw4 teils klein sind, sodass die zu sehenden Ergebnisse eventuell nicht das Gesamtgeschehen bei einer größeren Stichprobe widerspiegeln würden.

Von den **Beweisversuchen mit begründendem Ansatz, aber ohne weiterentwickelbares Argumentgefüge (KV1-2)** enthalten knapp 10% einen Hinweis auf einen grundlegenden Fehler bzw. eine nicht ganz korrekte Interpretation dieser Aussage. In ca. 45% der Beweisversuche beziehen sich die Ausführungen klar auf die zu zeigenden Aussage, in rund 45% der Beweisversuche ist die Passung zu der zu zeigenden Aussage unklar. Dabei fällt auf, dass von den Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* über 75% in die Kategorie PzzA2 *Bezug auf die zu zeigende Aussage unklar* gehören. Ein großer Teil dieser Beweisversuche ist so vage oder unvollständig, dass unklar ist, inwieweit sich die Ausführungen auf die zu zeigende Aussage beziehen. Von den Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und einer der anderen Schlussweisen Sw1-3 enthalten 14% einen Hinweis auf einen grundlegenden Fehler, bei rund 67% passen die Ausführungen zu der zu zeigenden Aussage. Der Anteil der Kategorie PzzA3 Passung zu der z*u zeigenden Aussage* ist statistisch hoch signifikant größer mit $V = 0{,}457$ wie im Falle der Schlussweise Sw4. Dies kann damit zusammenhängen, dass alle Beweisversuche mit Sw1-3 im Vergleich zu denen mit Sw4 mindestens ein Beispiel enthalten und Beispiele bei der Analyse der Passung der Ausführungen zu der zu zeigenden Aussage miteinbezogen wurden (Abschn. 3.3.4, 3.3.5). Die Beispiele können aus der Sicht von Lesenden z. B. darauf hindeuten, dass sich eventuell nicht ganz präzise Ausführungen auf die zu zeigende Aussage beziehen (Abschn. 8.2.3.3).

So wie bei Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge (KV1-2) sind auch bei **Beweisversuchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler (KV3)** hinsichtlich der Passung der Ausführungen zu der zu zeigenden Aussage alle drei Ausprägungen PzzA1-3 möglich. Dabei liegt der Anteil der Beweisversuche mit einem Hinweis auf einen grundlegenden Fehler (PzzA1) unter 5%. Wenn ein Hinweis auf einen grundlegenden Fehler vorliegt oder der Bezug auf die zu zeigende Aussage unklar ist, ist dies i. d. R. in Zusammenhang mit dem Fehler, wie z. B. einem nicht hinreichend legitimierten Schluss auf die Behauptung, zu sehen.

In **Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge (KV4-6)** beziehen sich die Ausführungen entsprechend Abbildung 116 und den Erläuterungen in Abschnitt 8.2.1 auf die zu zeigende Aussage.

Zusammenfassend sind insbesondere folgende Ergebnisse hervorzuheben: Während aus allen Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge der Bezug auf die zu zeigende Aussage hervorgeht, ist vor allem bei Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* der Bezug auf die zu zeigende Aussage oftmals unklar. Von den Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und einer der Schlussweise Sw1-3 enthalten mehr als 10% einen Hinweis auf einen grundlegenden Fehler.

8.2.6.3 Detailanalyse: Primäre Beweisideen

In diesem Abschnitt wird die primäre Beweisidee in Beweisversuchen mit begründendem Ansatz in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise untersucht. In Abbildung 117 ist bezogen auf jede der Gruppen KV1, Sw1-2 bis KV5-6, Sw4 angegeben, welche primären Beweisideen vorkommen und in einem wie großen Anteil der Beweisversuche sie zu finden sind.[146] Im Folgenden werden die in Abbildung 117 erkennbaren Regelmäßigkeiten sukzessive erläutert. Dabei wird zuerst von den primären Beweisideen ausgegangen, anschließend von der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und abschließend von den Schlussweisen. Am Ende dieses Abschnitts befindet sich ein Fazit, in dem wesentliche Ergebnisse der detaillierten Analysen zusammengefasst sind.

Primäre Beweisideen

Die relativen Häufigkeiten der primären Beweisideen sind in Abbildung 117 farbig dargestellt. Beispielsweise ist der Anteil der Beweisversuche mit der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* in hellgrün dargestellt. Es ist zu sehen, dass die Beweisversuche mit der primären Beweisidee 1 fast alle ein weitentwickelbares Argumentgefüge und besonders häufig ein weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke enthalten sowie dass die Schlussweise meistens zwischen induktiv und deduktiv liegt. Während die primäre Beweisidee in einem großen Anteil der Beweisversuche mit begründendem Ansatz vorkommt, kennzeichnet die primäre Beweisidee 4 *Zahldarstellung im Dezimalsystem* bezogen auf jede der Gruppen KV1, Sw1-2 bis KV5-6, Sw4 unter 10% aller Beweisversuche. Der Anteil der Beweisversuche mit der primären Beweisidee 4 ist daher entsprechend Fußnote 146 in den grau markierten Anteilen berücksichtigt. Zudem ist unterhalb jedes Balkendiagramms angegeben, welche primären Beweisideen in den analysierten Beweisversuchen der fokussierten Gruppe vorkommen. Die Beweisidee 4 kommt demnach in Beweisversuchen ohne weiterentwickelbares Argumentgefüge (KV1-2) und in solchen mit einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler (KV3) vor, jedoch nicht in solchen mit einem weitgehend korrekten Argumentgefüge (KV4-6), obwohl die zu zeigende Aussage ausgehend von der Beweisidee 4 bewiesen werden kann. Diese und weitere Auffälligkeiten werden im Folgenden näher ausgeführt.

Primäre Beweisidee 1

Von den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* enthalten rund 71% ein weitgehend korrektes Argumentgefüge bei zwischen induktiver und deduktiver Schlussweise.

In ca. der Hälfte der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 ist ein weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke (KV4) zu finden.

[146] Zugunsten der Übersichtlichkeit sind in den Streifendiagrammen ausschließlich die primären Beweisideen, die bezogen auf mindestens eine der Gruppen in über 10% der analysierten Beweisversuche vorkommen, separat aufgeführt und die anderen primären Beweisideen in einer Restkategorie zusammengefasst.

Dies ist insofern nachvollziehbar, als dass sämtliche Bearbeitungen mit der primären Beweisidee 1 dort einzuordnen sind, bei denen nicht begründet ist, warum sich die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer um acht vergrößert, oder als Induktionsanfang nicht das kleinste Beispiel, sondern etwa das Beispiel der Zahlen 5, 7, 9, 11 ausgeführt ist.

Rund ein Drittel der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 enthalten ein weitgehend korrektes und vollständiges Argumentgefüge (KV5-6). Dieser Anteil ist statistisch hoch signifikant größer als bei den übrigen Beweisversuchen mit begründendem Ansatz und einer ausbaubaren primären Beweisidee ($p = 0{,}002^{N-1}$, $V = 0{,}185$). Ein möglicher Grund besteht darin, dass für ein weitgehend korrektes und vollständiges Argumentgefüge die Verwendung der Beweisidee 1 zusammen mit elementaren Beweisideen, wie den Beweisideen 10 *Abstände ungerader Zahlen* und 11 *vier Zahlen*, ausreicht (Abschn. 6.2.2). Zudem stammen die Beweisversuche mit der primären Beweisidee 1 zu gut 75% von Lernenden aus den Klassenstufen 7, 8 (Abschn. 8.2.2). Die Beweisversuche mit begründendem Ansatz aus den Klassenstufen 7, 8 sind hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges im Durchschnitt besser als die aus den Klassenstufen 5 und 6 (Abschn. 8.2.1).

Insgesamt enthalten 95% der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 ein weiterentwickelbares Argumentgefüge (KV3-6). Die primäre Beweisidee 1 gehört zusammen mit den primären Beweisideen 2 und 5 zu denjenigen mit den größten Anteilen an Beweisversuchen mit einem weiterentwickelbaren Argumentgefüge.

Von den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe bei Erkennen einer Invariante* gehören unter 20% in die Kategorie Sw4 *deduktiv ohne zusätzliches Beispiel*, über 80% in eine der anderen Kategorien Sw1-3, meistens in eine der Kategorien Sw1-2. In jeder der Klassenstufen 6 bis 8 ist der Anteil der Beweisversuche mit zwischen induktiver und deduktiver Schlussweise (Sw1-2) an allen Beweisversuchen mit begründendem Ansatz im Fall der primären Beweisidee 1 statistisch hoch signifikant größer als im Fall einer anderen primären Beweisidee (jeweils $p < 0.01^{N-1}$, $0{,}27 < V < 0{,}46$). Die primäre Beweisidee 1 kann besonders gut ausgehend von Reihenfolgebeispielen entwickelt werden, wobei es fraglich erscheint, inwieweit Lernende die Notwendigkeit sehen, ihre tendenziell an Beispielen durchgeführten Überlegungen zusätzlich allgemeingültig zu formulieren (Abschn. 6.2.2). Dass in der Klassenstufe 5 kein statistisch signifikanter Unterschied erkennbar ist, kann damit zusammenhängen, dass in dieser Klassenstufe die primäre Beweisidee 1 einen vergleichsweise kleinen Anteil der Beweisversuche mit begründendem Ansatz ausmacht und die Schlussweise in fast allen Beweisversuchen mit einem weitgehend korrekten Argumentgefüge, d. h. unabhängig von der primären Beweisidee, zwischen induktiv und deduktiv liegt (Abschn. 8.2.1, 8.2.2).

Primäre Beweisidee 2

Beweisversuche mit begründendem Ansatz und der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* kommen in den Klassenstufen 6, 7, 8 vor (Abschn. 8.2.2). Die aus der Klassenstufe 6 enthalten alle ein weitgehend korrektes und voll-

ständiges Argumentgefüge (KV5-6). Von denjenigen aus den Klassenstufen 7, 8 enthalten weniger als die Hälfte ein weitgehend korrektes und vollständiges Argumentgefüge (KV5-6), aber etwa zwei Drittel mindestens ein weitgehend korrektes Argumentgefüge (KV4-6). Dies kann entsprechend den Ausführungen in Abschnitt 6.2.2 damit zusammenhängen, dass für einen Beweis der in den Klassenstufen 5, 6 zu zeigenden Teilbarkeit durch vier im Wesentlichen die Beweisidee 2 und ein zwei- bis dreischrittiges Argumentgefüge ausreichen, während für einen Beweis der in den Klassenstufen 7, 8 zu zeigenden Teilbarkeit durch acht eine weitere ausbaubare Beweisidee und ein komplexeres Argumentgefüge erforderlich sind. Darauf, dass die primäre Beweisidee 2 vor allem von eher leistungsstarken oder mathematisch geförderten Teilnehmenden genutzt wurde, deuten die erhobenen Daten nicht hin (Abschn. 10.4).

Die Schlussweise ist in mehr als der Hälfte der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 2 deduktiv. In knapp einem Fünftel der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 2 wird die zu zeigende Aussage bewiesen.

Primäre Beweisideen 3, 5, 7, 8

Die Beweisversuche mit einer der ausbaubaren primären Beweisideen 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*, 5 *Reste bei Division einzelner Summanden durch vier oder acht*, 7 *Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* sind sowohl bezüglich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges als auch hinsichtlich der Schlussweise vielfältig. Dabei fällt aber auf, dass in ca. 60% der Beweisversuche mit der primären Beweisidee 5 *Reste bei Division einzelner Summanden durch vier oder acht* ein weitgehend korrektes und vollständiges Argumentgefüge formuliert ist, was in Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion steht (Abschn. 6.2.2). Zudem fällt auf, dass alle analysierten Beweisversuche mit der primären Beweisidee 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* und der Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* ein weitgehend korrektes und vollständiges, deduktives Argumentgefüge enthalten. Wie an späterer Stelle deutlich wird, hatten über 90% der Teilnehmenden, die einen Beweisversuche mit der primären Beweisidee 5 oder einen mit der primären Beweisidee 8 und der Beweisidee 2 verfasst haben, an einem mathematischen Förderangebot teilgenommen (Abschn. 10.4.2.3).

Primäre Beweisidee 4

Ausgehend von der primären Beweisidee 4 *Zahldarstellung im Dezimalsystem*, einer der zentralen Beweisideen, ist ein Beweis der zu zeigenden Aussage möglich. In Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion, kommt aber in keinem der analysierten Beweisversuche mit begründendem Ansatz und der primären Beweisidee 4 ein Beweis der zu zeigenden Aussage vor, de facto auch kein weitgehend korrektes Argumentgefüge (KV4-6). Die Beweisversuche mit begründendem Ansatz und der primären Beweisidee 4 enthalten fast alle maximal einen Ansatz für ein weiterentwickelbares Argumentgefüge (KV1-2).

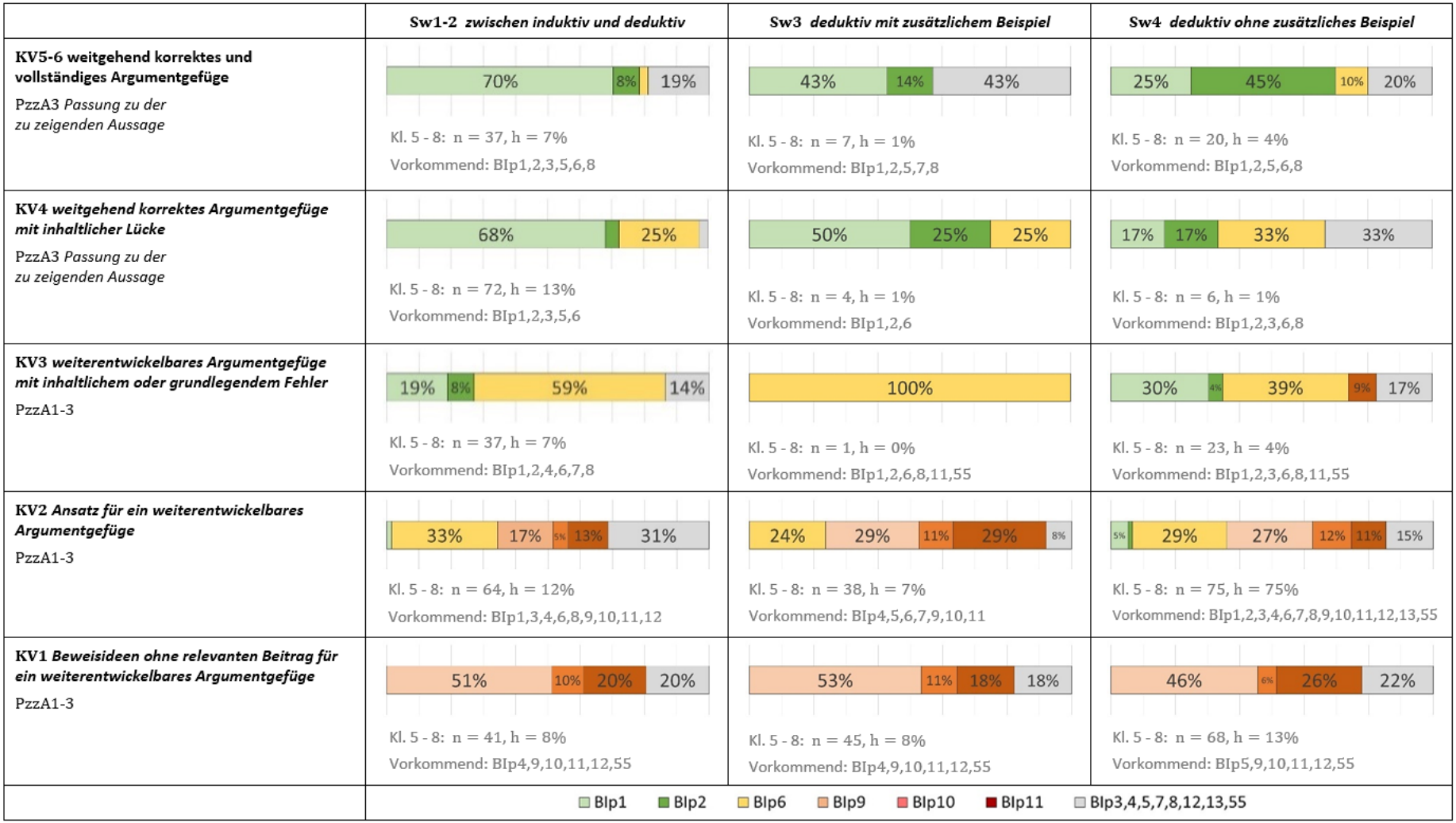


Abbildung 117: Beweisversuche mit begründendem Ansatz: Primäre Beweisideen in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise. Spektrum an vorkommenden Ausprägungen und relative Häufigkeitsverteilungen der Codierungen

Primäre Beweisidee 6

Von den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* enthalten in jeder der Klassenstufen 5 bis 8 mit etwa 20% der Zuordnungen einige ein weitgehend korrektes Argumentgefüge (KV4-6), jedoch allenfalls einzelne ein weitgehend korrektes und vollständiges Argumentgefüge (KV5-6), was ebenfalls in Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion steht (Abschn. 6.2.2). Pro Klassenstufe gehören etwa 50% der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 6 in Kategorie KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge*. Es wurde also vergleichsweise häufig z. B. ausschließlich angeführt, dass die Summe von zwei aufeinanderfolgenden ungeraden Zahlen durch vier teilbar ist, oder in den Klassenstufen 7, 8 so über die Parität von Summen argumentiert, dass aus Expertensicht die Beweisidee 6 als übergeordnete Beweisidee erkennbar ist, ohne dass ein weiterentwickelbares Argumentgefüge formuliert ist.

Bezüglich der Schlussweise in Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 6 fällt auf, dass von denjenigen mit zwischen induktiver und deduktiver Schlussweise knapp ein Drittel ein weitgehend korrektes Argumentgefüge enthalten, während dies ausschließlich auf rund ein Zehntel derjenigen mit deduktiver Schlussweise zutrifft. Zudem ist bei den Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise der Anteil derjenigen ohne weiterentwickelbares Argumentgefüge (KV1-2) mit rund einem im Vergleich zu zwei Dritteln der Zuordnungen kleiner. Beide Unterschiede sind statistisch signifikant mit geringer bis mittlerer Effektstärke (N-1 Chi²-Tests, $0.01 \leq p \leq 0.15$, $0.23 < V < 0.33$). Die Beweisversuche mit begründendem Ansatz, primärer Beweisidee 6 und zwischen induktiver und deduktiver Schlussweise sind hinsichtlich der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges tendenziell besser als diejenigen mit deduktiver Schlussweise.

Primäre Beweisideen 9 bis 55

Ausgehend von fast jeder der primären Beweisideen 9 bis 13, 55 lässt sich maximal ein Ansatz für ein weiterentwickelbares Argumentgefüge formulieren (Abschn. 3.3.3, 6.2.2), der relative Anteil der Beweisversuche mit einer der primären Beweisideen 9 bis13, 55 in Kategorie KV1-2 beträgt über 98%. In Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion in Abschnitt 6.2.2 kommt kein weitgehend korrektes Argumentgefüge vor, obwohl dies ausgehend von der Beweisidee 11 *vier Zahlen* möglich ist.

Weiterhin fällt auf, dass in über 70% der Beweisversuche mit einer der primären Beweisideen 9 *Parität von Summen ungerader oder gerader Zahlen*, 10 *Abstände ungerader Zahlen* und 11 *vier Zahlen* die Schlussweise deduktiv ist, meistens deduktiv ohne zusätzliches Beispiel. Das lässt vermuten, dass diese Beweisideen einigen Lernenden so naheliegend erschienen oder ihnen Wissenselemente zur Parität von Summen so gut bekannt sind, dass sie die Beweisideen ohne eine intensive Beschäftigung mit einem Beispiel ausgeführt haben, und dass Beispiele gegebenenfalls weniger für die Entwicklung oder Darstellung einer Beweisidee als z. B. zur Stützung der zu zeigenden Aussage oder aus Gewohnheit angegeben wurden.

Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges

Im Folgenden werden die vorkommenden primären Beweisideen und deren relative Häufigkeitsverteilungen in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges betrachtet. Dazu ist in Abbildung 117 zeilenweise zu schauen. Beispielsweise sind die Beweisversuche mit begründendem Ansatz in Kategorie KV2 *Ansatz für ein weiterentwickelbares Argumentgefüges* hinsichtlich der vorkommenden primären Beweisideen besonders vielfältig. Dabei beruhen von den Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge (KV1-2) mehr als ein Drittel auf der primären Beweisidee 9 *Parität von Summen ungerader oder gerader Zahlen*. Bei der Mehrheit der Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler (KV3) ist die primäre Beweisidee die Idee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* oder die Idee 6 *Teilbarkeit der Summen zweier Zahlenpaare*. Von den Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) basieren über 70% auf einer der Beweisideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* und 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*. Diese und weitere Auffälligkeiten werden im Folgenden erläutert.

KV1-2 *kein weiterentwickelbares Argumentgefüge*

In Beweisversuchen mit begründendem Ansatz, aber ohne weiterentwickelbares Argumentgefüge (KV1-2) sind **alle unterschiedenen Beweisideen 1 bis 13, 55** als primäre Beweisidee möglich. Bei den Beweisversuchen mit der Merkmalsausprägung KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge* taucht tatsächlich jede der unterschiedenen Beweisideen als primäre Beweisidee auf. In den analysierten Beweisversuchen mit der Merkmalsausprägung KV1 *Beweisideen ohne relevanten Beitrag für ein weiterentwickelbares Argumentgefüge* hingegen kommen die primären Beweisideen 1 bis 8 fast nicht vor. Da im Falle einer ausbaubaren Beweisidee als primärer Beweisidee (BIp1-8) meistens zumindest ein Ansatz für ein weiterentwickelbares Argumentgefüge erkennbar ist, fallen die Beweisversuche mit BIp1-8 i. d. R. nicht in Kategorie KV1 bzw. mindestens in Kategorie KV2 (Abschn. 3.3.4, 6.2.2).

In jeder der Gruppen mit der Merkmalsausprägung KV1-2 *kein weiterentwickelbares Argumentgefüge* gehören die primären Beweisideen 9 *Parität von Summen ungerader oder gerader Zahlen* und 11 *vier Zahlen* zu den drei häufigsten primären Beweisideen. Auf der **primären Beweisidee 9** basieren über 20% aller Beweisversuche mit begründendem Ansatz (Abschn. 8.2.2). Allerdings lässt sich auf Grundlage dieser Idee allenfalls ein Ansatz für ein weiterentwickelbares Argumentgefüge (KV1-2) formulieren (Abschn. 3.3.3, 6.2.2), was die Häufung an Beweisversuchen mit der primären Beweisidee 9 in den Gruppen mit KV1-2 erklärt. Ähnlich verhält es sich bei den Beweisversuchen mit der **primären Beweisidee 11** *vier Zahlen*, die rund 12% aller Beweisversuche mit begründendem Ansatz ausmachen, aber fast immer in eine der Gruppen mit der Merkmalsausprägung KV1-2 fallen (Abschn. 6.2.2, 8.2.2).

Zusätzlich zu den primären Beweisideen 9, 11 macht die **primäre Beweisidee 6** *Teilbarkeit der Summen zweier Zahlenpaare* in jeder der Gruppen mit der Merkmalsausprägung KV2 *Ansatz*

für ein weiterentwickelbares Argumentgefüge mit etwa 30% einen großen Anteil aus. Die Beweisidee 6 ist in etwa einem Fünftel aller Beweisversuche mit begründendem Ansatz die primäre Beweisidee. Sie gehört zu den ausbaubaren Beweisideen, so dass die Beweisversuche i. d. R. mindestens in Kategorie $\mathrm{KV2}$ einzuordnen sind, wobei bei der primären Beweisidee 6 der Anteil der Kategorie $\mathrm{KV2}$ im Vergleich zu den meisten anderen ausbaubaren Beweisideen groß ist, wie es in obigen Ausführungen zur primären Beweisidee 6 erklärt ist.

Wenn man bei den Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge ($\mathrm{KV1\text{-}2}$) ergänzend nach der **Passung der Ausführungen zu der zu zeigenden Aussage** differenziert, sind zwei weitere Auffälligkeiten zu erkennen. Erstens ist bei den Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und einem Hinweis auf einen grundlegenden Fehler die primäre Beweisidee 9 *Summen oder Produkte ungerader Zahlen* statistisch signifikant häufiger als bei denen ohne einen Hinweis auf einen grundlegenden Fehler ($\mathrm{p} = .022^{\mathrm{N\text{-}1}}$, $\mathrm{V} = .126$). Zweitens ist bei den Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und einem Hinweis auf einen grundlegenden Fehler der Anteil der ausbaubaren primären Beweisideen mit rund einem Zehntel im Vergleich zu gut einem Viertel kleiner als bei denen ohne einen Hinweis auf einen grundlegenden Fehler. Diese gegenläufigen Tendenzen sind plausibel. Wenn ein Beweisversuch einen Hinweis auf einen grundlegenden Fehler enthält, weil beispielsweise die Summe von vier nicht aufeinanderfolgenden ungeraden Zahlen betrachtet wird, dann können Wissenselemente zur Parität von Summen ungerader Zahlen angewendet werden, während mehrere der ausbaubaren Beweisideen nicht möglich sind.

KV3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler*

Bei den Beweisversuchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler ($\mathrm{KV3}$) umfasst das Spektrum an möglichen primären Beweisideen die Ideen 1 bis 8, 11, 55 (Abschn. 3.3.3, 6.2.2). Fast jede diese primären Beweisideen kommt in analysierten Beweisversuchen mit einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler vor (Abb. 117). Besonders häufig sind dabei die primären Beweisideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* und 6 *Teilbarkeit der Summen zweier Zahlenpaare*, die jeweils in einem großen Anteil aller Beweisversuche mit begründendem Ansatz die primäre Beweisidee darstellen und sich als ausbaubare Beweisideen für ein korrektes und vollständiges Argumentgefüge eignen, wobei deren Nutzung auch aus theoretischer Sicht fehleranfällig erscheint (Abschn. 6.2.2, 8.2.2).

KV4-6 *weitgehend korrektes Argumentgefüge*

Bei Beweisversuchen mit einem weitgehend korrekten Argumentgefüge ($\mathrm{KV4\text{-}6,Sw1\text{-}4}$) sind prinzipiell die **primären Beweisideen 1 bis 8 und 11** möglich. In Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion und obigen Ausführungen zu den verschiedenen primären Beweisideen kommen viele dieser primären Beweisideen vor, die Ideen 4 *Zahldarstellung im Dezimalsystem* und 11 *vier Zahlen* jedoch nicht (Abb. 117, Abschn. 6.2.2).

Ebenfalls in Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion und obigen Ausführungen gehören in jeder der Gruppen mit der Merkmalsausprägung KV4 *weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke* die **primären Beweisideen 1, 6** zu den drei häufigsten primären Beweisideen, und die primäre Beweisidee 1, aber nicht die primäre Beweisidee 6 auch in jeder der Gruppen mit der Merkmalsausprägung KV5-6 *weitgehend korrektes und vollständiges Argumentgefüge* (Abb. 117).

Zusätzlich zu den primären Beweisideen 1 und 6 kommt die **primäre Beweisidee 2** *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* in einem großen Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge vor(Abb. 117). Diese Beweisidee ist ausschließlich in etwa 5% der Beweisversuche mit begründendem Ansatz aus den Klassenstufen 6, 7 und in rund 12% derjenigen aus der Klassenstufen 8 die primäre Beweisidee. Darauf aufbauend wurde aber in mehr als drei Viertel der Fälle ein weitgehend korrektes Argumentgefüge formuliert. Die anderen möglichen primären Beweisideen 3, 4, 5, 7, 8, 11 sind alle in einem noch kleineren Anteil der Beweisversuche mit begründendem Ansatz zu finden oder wurden seltener als weitgehend korrektes Argumentgefüge ausgeführt.

Schlussweisen

Abschließend werden die vorkommenden primären Beweisideen und deren relative Häufigkeitsverteilungen in Abhängigkeit der Schlussweise betrachtet und besondere Auffälligkeiten dargestellt. Dabei werden auch die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges berücksichtigt. Bezogen auf Abbildung 117 ist also primär spaltenweise und ergänzend zeilenweise zu schauen.

Die Beweisversuche der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* und einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) basieren zu über 50% auf der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*. Das kann damit zusammenhängen, dass Beweisversuche mit einer anderen primären Beweisidee vergleichsweise selten die Merkmalskombination **Sw4, KV5-6** aufweisen. Außerdem kommt jede der primären Beweisideen 3, 5, 7, 8, die sich gut für ein weitgehend korrektes und vollständiges, deduktives Argumentgefüge eignen, in allen Klassenstufen in weniger als 5% der Beweisversuche mit begründendem Ansatz vor (Abschn. 6.2.2, 8.2.2).

Bezogen auf alle Beweisversuche mit **deduktiver Schlussweise (Sw3-4)** lässt sich festhalten, dass in etwa 60% dieser Beweisversuche eine der nicht ausbaubaren Beweisideen die primäre Beweisidee ist, meistens eine der Ideen 9 *Parität von Summen ungerader und gerader Zahlen* und 11 *vier Zahlen*. Ein Erklärungsansatz besteht darin, dass, wie beschrieben, die Beweisideen 9, 11 in einem großen Anteil der Beweisversuche mit begründendem Ansatz die primäre Beweisidee darstellen, im Falle dieser primären Beweisideen die Schlussweise meistens deduktiv ist und vor allem die primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* zwar ebenfalls in einem großen Anteil der Beweisversuche mit begründendem Ansatz ausgeführt ist, aber bei zwischen induktiver und deduktiver Schlussweise. Von den Beweisversuchen mit **zwischen induktiver und deduktiver Schlussweise (Sw1-2)** enthalten ca. 70% eine ausbaubare primäre Beweisidee, meistens eine der Ideen 1, 6.

Fazit: Primäre Beweisidee und Korrektheit, Vollständigkeit, Schlussweise

Die Betrachtung der primären Beweisidee in Zusammenhäng mit der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges hebt hervor, dass sich nicht alle primären Beweisideen für ein korrektes und vollständiges Argumentgefüge oder zumindest ein weiterentwickelbares Argumentgefüge eignen. Zudem fällt auf, dass insbesondere in Beweisversuchen mit begründendem Ansatz und einer der primären Beweisideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*, 5 *Reste einzelner Summanden bei Division durch vier oder acht* häufig ein weitgehend korrektes und vollständiges Argumentgefüge oder zumindest ein weiterentwickelbares Argumentgefüge ausgeführt ist. Demgegenüber kommt in fast keinem der Beweisversuche mit begründendem Ansatz und einer der primären Beweisideen 4, 9, 10, 11, 12, 13, 55 ein weiterentwickelbares Argumentgefüge vor.

Die Beweisversuche mit begründendem Ansatz in der Kategorie KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge* sind bezüglich der primären Beweisidee besonders vielfältig, aber z. B. auch bei denen mit einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) kommen sieben verschiedene Beweisideen vor. Zugleich basieren mehr als die Hälfte der Beweisversuche mit begründendem Ansatz und einem Ansatz für ein weiterentwickelbares Argumentgefüge sowie mehr als die Hälfte derjenigen mit einem weitgehend korrekten und vollständigen Argumentgefüge auf ausschließlich drei der unterschiedenen Beweisideen.

Die Beweisversuche mit zwischen induktiver und deduktiver Schlussweise basieren mehrheitlich auf einer ausbaubaren Beweisidee, die Beweisversuche mit der Schlussweise *deduktiv ohne zusätzliches Beispiel* mehrheitlich auf einer nicht ausbaubaren Beweisidee. Insbesondere bei den Beweisversuchen mit der Schlussweise *deduktiv ohne zusätzliches Beispiel* und einem weitgehend korrekten und vollständigen Argumentgefüge macht die primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* einen großen Anteil aus.

Insgesamt betrachtet deuten die Ergebnisse auf Zusammenhänge zwischen primären Beweisideen, der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise hin, von denen die meisten in Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion stehen.

8.2.6.4 Detailanalyse: Existenz, Art und Ort von Beispielen

In diesem Abschnitt wird die Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz in Zusammenhang mit der Schlussweise und der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüge betrachtet. Ergänzend werden die Passung der Ausführungen zu der zu zeigenden Aussage und die primäre Beweisidee berücksichtigt. Bei der nun folgenden Betrachtung der Verwendung von Beispielen in Zusammenhang mit weiteren Eigenschaften von Beweisversuchen mit begründendem Ansatz liegt der Fokus nacheinander auf der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage, der Art angegebener Beispiele und dem Ort von Beispielen. Abschließend werden wesentlichen Ergebnisse dieser Detailanalysen zusammengefasst.

Existenz und Passung eines Beispiels zu der zu zeigenden Aussage

In Abbildung 118 ist für jede der unterschiedenen Gruppen KV1, Sw1-2 bis KV5-6, Sw4 dargestellt, inwieweit in den Beweisversuchen mit begründendem Ansatz ein Beispiel und insbesondere ein korrekt und vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage vorkommt. Dabei sind die Anteile der Beweisversuche ohne Beispiel in grau dargestellt, die Anteile derjenigen mit Beispiel in Grün-, Gelb- und Rottönen. Es ist beispielsweise zu sehen, dass in über 90% der Beweisversuche mit der Schlussweise Sw4 kein Beispiel angegeben ist, während alle Beweisversuche mit einer der Schlussweisen Sw1-3 mindestens ein Beispiel enthalten. Dieser primär definitionsbedingte Unterschied ist nach dem N-1 Chi²-Test statistisch hoch signifikant mit $V = 0{,}952$ (Abschn. 3.3.6 zu den Definitionen der Schlussweisen).

Ergänzend ist zu sehen, dass 30% der Beweisversuche mit der **Schlussweise Sw4** und einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6, Sw4) ein Beispiel zu der zu zeigenden Aussage enthalten, während mit Blick auf alle anderen Beweisversuche mit der Schlussweise Sw4 (KV1-4, Sw4) Beispiele zu der zu zeigenden Aussage sehr selten sind. Dieser Unterschied zwischen Beweisversuchen mit der Schlussweise Sw4 und mit bzw. ohne weitgehend korrektes und vollständiges Argumentgefüge ist nach dem exakten Fisher-Test statistisch hoch signifikant mit $V = 0{,}335$. Die Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge enthalten genau dann ein Beispiel, wenn dies eine der verwendeten Beweisideen erfordert. Diejenigen mit der Schlussweise Sw4 basieren fast alle auf der primären Beweisidee 1. Im Falle der Verwendung der Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* ist die Angabe eines Beispiels für ein weitgehend korrektes und vollständiges Argumentgefüge notwendig. In allen analysierten Beweisversuchen mit der Schlussweise Sw4, der Beweisidee 1 und einem Beispiel zu der zu zeigenden Aussage ist tatsächlich ein weitgehend korrektes und vollständiges Argumentgefüge formuliert.

Von den Beweisversuchen mit einer der **Schlussweisen Sw1-3** enthalten rund 75% ein eigenes Beispiel zu der zu zeigenden Aussage (PBsp4-5). Auffällig sind dabei die Beweisversuche mit einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler (KV3, Sw1-3). Bei ihnen ist der Anteil derjenigen mit einem eigenen Beispiel zu der zu zeigenden Aussage (Bsp4-5) im Vergleich zu den übrigen Beweisversuchen mit einer der Schlussweisen Sw1-3 statistisch hoch signifikant kleiner mit $p < 0.001^{N\text{-}1}$, $V = 0{,}191$. Es ist denkbar, dass in vielen dieser Beweisversuche der inhaltliche oder grundlegende Fehler mit dem Fehlen eines eigenen Beispiels zu der zu zeigenden Aussage zusammenhängt.

Weiterhin unterscheiden sich die Beweisversuche mit einer der Schlussweisen Sw1-3 und keinem weiterentwickelbaren Argumentgefüge (KV1-2) je nach Existenz eines Hinweises auf einen grundlegenden Fehler hinsichtlich der Passung eines Beispiels zu der zu zeigenden Aussage statistisch hoch signifikant mit $p < 0{,}001^{\chi^2}$, $V = 0.470$. Von den Beweisversuchen mit Sw1-3, KV1-2 und einem Hinweis auf einen grundlegenden Fehler enthalten über 70% ausschließlich sonstige Beispiele (PBsp2) und keiner ein korrekt und vollständig ausgeführtes eigenes Beispiel (PBsp5). Von den Beweisversuchen mit Sw1-3, KV1-2 und keinem Hinweis auf einen grundlegenden Fehler enthalten über 80% ein Beispiel zu der zu zeigenden Aussage (PBsp3-5), über 35% ein korrekt und vollständig ausgeführtes eigenes Beispiel (PBsp5). Andersherum betrach-

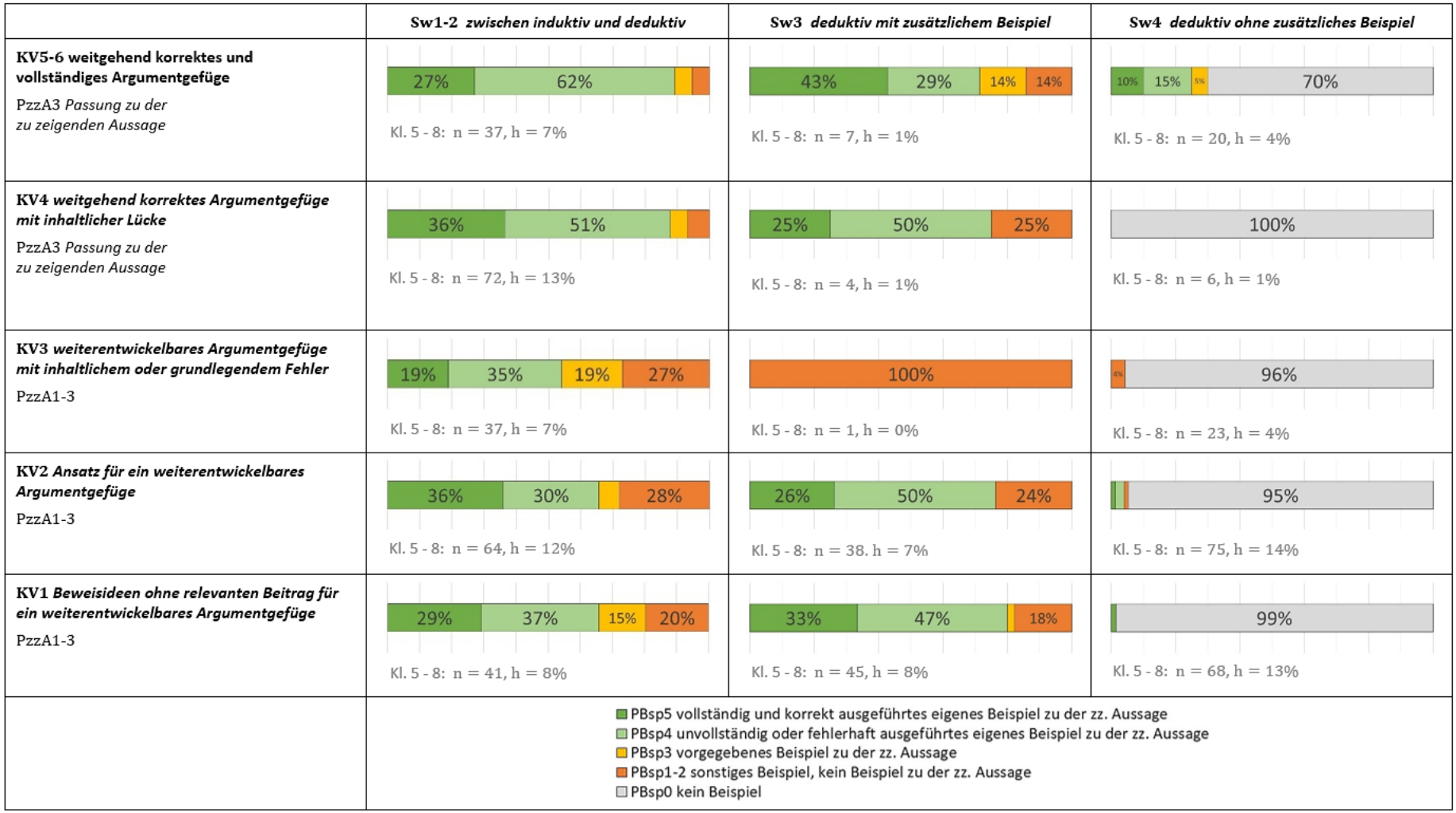


Abbildung 118: Beweisversuche mit begründendem Ansatz: Existenz und Passung eines Beispiels zu der zu zeigenden Aussage in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen

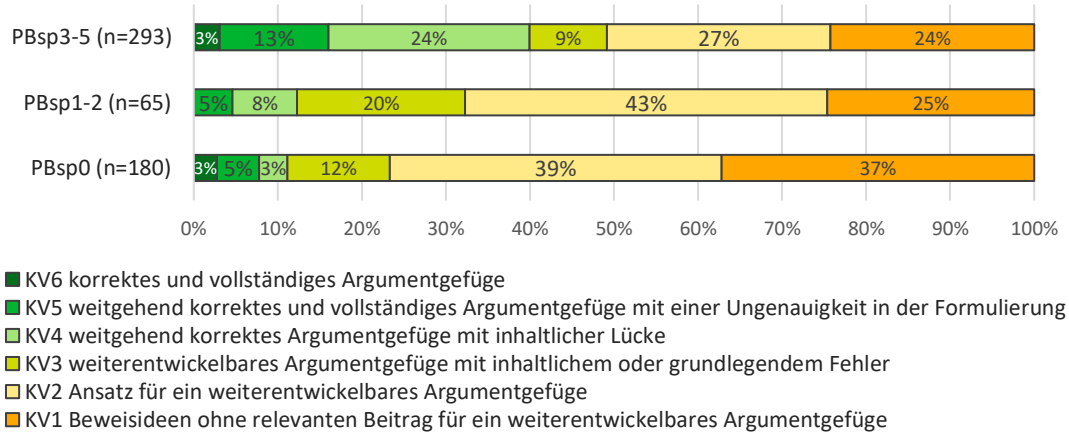
tet enthalten von den Beweisversuchen mit Sw1-3, KV1-2 und keinem Beispiel zu der zu zeigenden Aussage (PBsp1-2) rund 42% einen Hinweis auf einen grundlegenden Fehler, während dieser Anteil bei den Beweisversuchen mit Sw1-3, KV1-2 und einem Beispiel zu der zu zeigenden Aussage (PBsp3-5) rund 5% beträgt und statistisch hoch signifikant kleiner ist mit $p < 0{,}001^{N-1}$, $V = 0{,}456$.

In Ergänzung dazu zeigt die Abbildung 119 mit Blick auf alle Beweisversuche mit begründendem Ansatz, dass bei denen mit einem Beispiel zu der zu zeigenden Aussage (PzzA3-5) nach dem Mann-Whitney-U-Test die Wahrscheinlichkeit größer ist, bezüglich der **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** zu einer höheren der Kategorien zu gehören als bei denen ohne Beispiel (PzzA0) sowie als bei denen mit einem Beispiel, aber ausschließlich sonstigen Beispielen (PzzA1-2). Bei den Beweisversuchen mit begründendem Ansatz und einem Beispiel zu der zu zeigenden Aussage ist mit rund 40% im Vergleich zu je ca. 10% insbesondere der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge (KV4-6) statistisch hoch signifikant größer. Sowohl von den Beweisversuchen mit begründendem Ansatz und ohne Beispiel als auch von denen mit sonstigen Beispielen enthalten knapp 90% kein weitgehend korrektes Argumentgefüge. Vereinfacht gesagt sind die Beweisversuche mit begründendem Ansatz und einem Beispiel zu der zu zeigenden Aussage hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges im Mittel besser als diejenigen ohne Beispiel zu der zu zeigenden Aussage.

Beweisversuche mit begründendem Ansatz

Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in Abhängigkeit der Existenz und Passung eines Beispiels zu der zz. Aussage

relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen



p-Werte von Signifikanztests und dazugehörige Effektstärken

	alle einzelnen Kategorien	KV1-3 vs. KV4-6
PBsp1-2 vs. PBsp3-5	$p < ,001^{U}$, $r = .138$	$p < ,001^{N-1}$, $V = .211$
PBsp0 vs. PBsp3-5	$p < ,001^{U}$, $r = .247$	$p < ,001^{N-1}$, $V = .308$

Abbildung 119: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in Abhängigkeit der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Zusammengefasst heben die Ergebnisse hinsichtlich der Existenz und Passung eines korrekt und vollständig ausgeführten Beispiels zu der zu zeigenden Aussage in Beweisversuchen mit begründendem Ansatz hervor, dass die Beweisversuche mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* sehr selten ein Beispiel enthalten, mit Ausnahme derjenigen mit einem weitgehend korrekten und vollständigen Argumentgefüge und einer Beweisidee wie der Beweisidee 1, die ein Beispiel zu der zu zeigenden Aussage erfordert. Insbesondere fällt auf, dass in allen analysierten Beweisversuchen mit der Schlussweise $Sw4$, der Beweisidee 1 und einem Beispiel zu der zu zeigenden Aussage tatsächlich ein weitgehend korrektes und vollständiges Argumentgefüge formuliert ist. Bezüglich der Beweisversuche mit einer der Schlussweisen Sw1-3 lässt sich hervorheben, dass diese alle ein Beispiel enthalten und meistens mindestens ein eigenes Beispiel zu der zu zeigenden Aussage angegeben ist. Die Beweisversuche mit Sw1-3, KV1-2 und einem Hinweis auf einen grundlegenden Fehler im Speziellen enthalten in der Regel ausschließlich sonstige Beispiele. Außerdem sind, vereinfacht gesagt, die Beweisversuche mit begründendem Ansatz und einem Beispiel zu der zu zeigenden Aussage hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges im Mittel besser als diejenigen ohne Beispiel und als diejenigen mit einem Beispiel, aber ausschließlich sonstigen Beispielen.

Art der Beispiele

Nachdem es im vorhergehenden Abschnitt um die Existenz und Passung eines Beispiels zu der zu zeigenden Aussage ging, werden in diesem Abschnitt Eigenschaften aller angegebener Beispiele betrachtet. Dabei geht es zuerst um die Existenz von Reihenfolgebeispielen und des kleinsten Beispiels, danach um die Existenz eines großen Beispiels.

Art der Beispiele: Reihenfolgebeispiele und das kleinste Beispiel

Die Abbildung 120 zeigt, inwieweit die Beweisversuche in den einzelnen Gruppen KV1, Sw1-2 bis KV5-6,Sw4 Reihenfolgebeispiele und das kleinste Beispiel enthalten. Im jeweils oberen Balkendiagramm sind der Anteil der Beweisversuche mit Reihenfolgebeispielen und dem kleinsten Beispiel in dunkelblau, der Anteil der übrigen Beweisversuche mit Beispiel in hellblau und der Anteil der Beweisversuche ohne Beispiel in grau dargestellt. Es ist zu sehen, dass fast keiner der Beweisversuche mit der **Schlussweise Sw4** *deduktiv ohne zusätzliches Beispiel* Reihenfolgebeispiele und das kleinste Beispiel enthält. Dies erscheint plausibel, da im Falle der Schlussweise Sw4 allenfalls notwendige Beispiele vorkommen. Bei der fokussierten Aussage können Reihenfolgebeispiele und das kleinste Beispiel zusammen genommen fast nur dann notwendige Beispiele sein, wenn sie als Fälle einer vollständigen Fallunterscheidung im Kontext der Beweisidee 4 *Zahldarstellung im Dezimalsystem* ausgeführt werden.

Mit Blick auf die Beweisversuche mit einer der **Schlussweisen Sw1-3** fällt auf, dass in jeder der Gruppen an Beweisversuchen mit einem weitgehend korrekten Argumentgefüge (KV4-6, Sw1-3) je mindestens 25% der Beweisversuche Reihenfolgebeispiele und das kleinste Beispiel enthalten, in jeder der Gruppen an Beweisversuchen ohne weitgehend korrektes Argumentgefüge (KV1-3, Sw1-3) unter 25%.

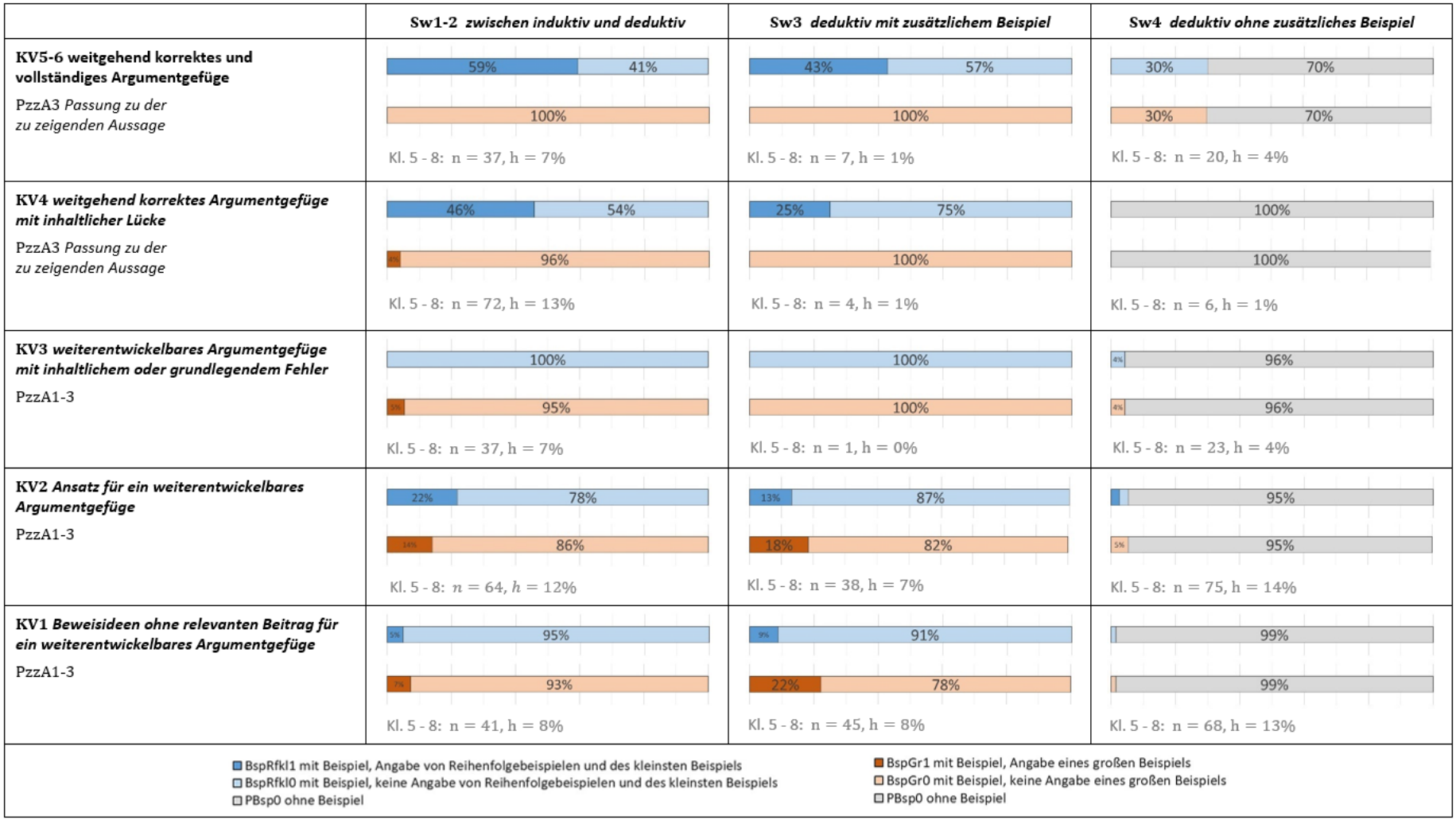
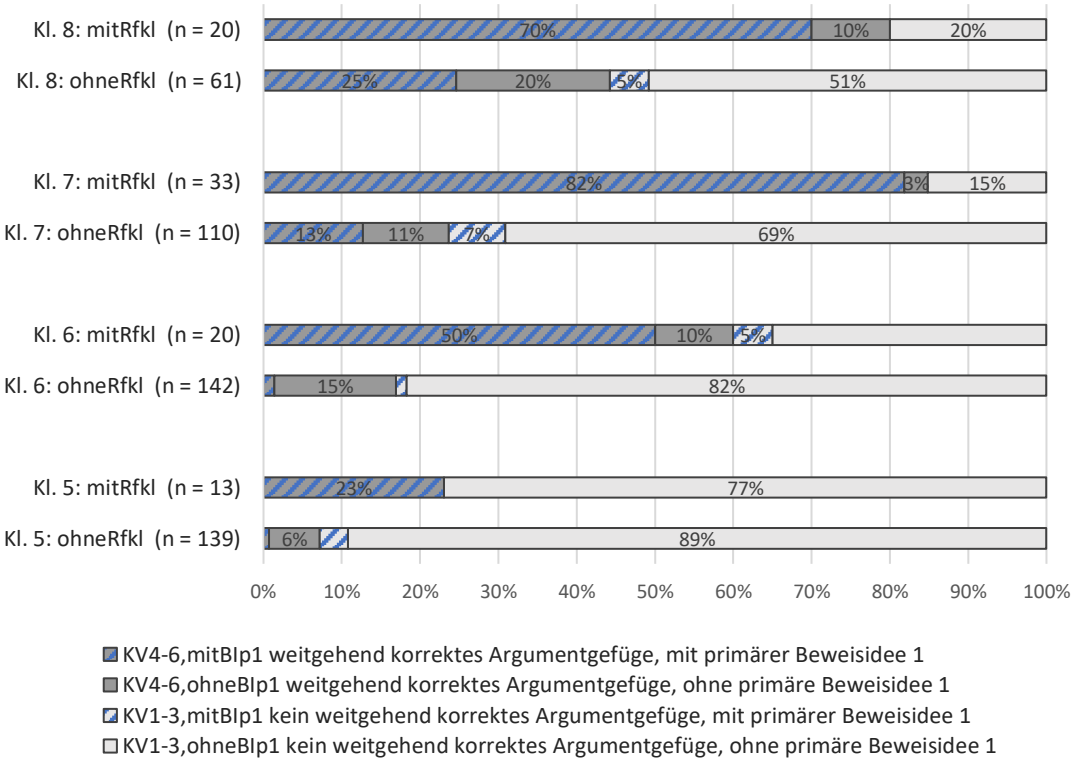


Abbildung 120: Beweisversuche mit begründendem Ansatz: Existenz und Art angegebener Beispiele in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen



p-Werte von Signifikanztests und dazugehörige Effektstärken

	KV1-3 vs. KV4-6
Kl. 5: mitRfkl vs. ohneRfkl	$p = .085^{ex},\ V = .159$
Kl. 6: mitRfkl vs. ohneRfkl	$p < .001^{ex},\ V = .341$
Kl. 7: mitRfkl vs. ohneRfkl	$p < .001^{N-1}, V = .532$
Kl. 8: mitRfkl vs. ohneRfkl	$p = .006^{N-1}, V = .309$

Fokus mitRfkl: KV1-3 vs. KV4-6	
Kl. 5 vs. Kl. 7	$p < .001^{ex},\ V = .593$
Kl. 5 vs. Kl. 8	$p = .001^{N-1}, V = .563$

Fokus mitRfkl: mitBIp1 vs. ohneBIp1	
Kl. 5 vs. Kl. 7	$p < .001^{ex},\ V = .555$
Kl. 5 vs. Kl. 8	$p = .009^{N-1}, V = .459$

Abbildung 121: Beweisversuche mit begründendem Ansatz: Existenz eines weitgehend korrekten Argumentgefüges und der primären Beweisidee 1 in Abhängigkeit der Existenz von Reihenfolgebeispielen und des kleinsten Beispiels. Relative Häufigkeitsverteilungen der Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Eine **Differenzierung nach Klassenstufen** (Abb. 121) zeigt, dass in jeder der Klassenstufen 6 bis 8 die Beweisversuche mit Reihenfolgebeispielen und dem kleinsten Beispiel (mitRfkl) zu einem statistisch hoch signifikant größeren Anteil mit $0{,}30 < V < 0{,}53$ ein weitgehend korrektes Argumentgefüge enthalten als die übrigen Beweisversuche mit begründendem Ansatz (ohneRfkl). Dabei scheint diese Art an Beispielen in der Klassenstufe 5 seltener gewinnbringend genutzt worden zu sein als in den Klassenstufen 7, 8. Während in den Klassenstufen 7, 8 etwa drei Viertel der Beweisversuche mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel ein weitgehend korrektes Argumentgefüge enthalten, sind es in der Klassenstufe 5 ca. ein Viertel und damit statistisch hoch signifikant weniger. Wenn man zusätzlich die Verwendung der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Er-*

kennen einer Invariante berücksichtigt, fällt Folgendes auf: Von den Beweisversuchen mit Reihenfolgebeispielen und dem kleinsten Beispiel aus der Klassenstufe 5 basiert ein weniger als halb so großer, statistisch hoch signifikant kleinerer Anteil auf der primären Beweisidee 1 als in den Klassenstufen 7, 8. Von den Beweisversuchen mit Reihenfolgebeispielen, dem kleinsten Beispiel und der primären Beweisidee 1 enthalten in allen Klassenstufen über 90% ein weitgehend korrektes Argumentgefüge (KV4-6), von den Beweisversuchen mit Reihenfolgebeispielen und dem kleinsten Beispiel, aber ohne primäre Beweisidee 1 sind es jeweils maximal 33%.

In Ergänzung dazu ist explizit hervorzuheben, dass in jeder der Klassenstufen 5 bis 8 Beweisversuche mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel häufig die primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* enthalten, statistisch hoch signifikant häufiger als die anderen Beweisversuche mit begründendem Ansatz ($p < .001^{N-1}, V = .489$). Trotzdem basierten die Beweisversuche mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel auf vielfältigen primäre Beweisideen. Zusätzlich zur primären Beweisidee 1 kommen im analysierten Sample auch die primären Beweisideen BIp4,5,6,7,9,10,11,12 in Beweisversuchen mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel vor.

Insgesamt sind bezüglich der Existenz von Reihenfolgebeispielen und des kleinsten Beispiels folgende Tendenzen zu erkennen: Beweisversuche mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel kommen vor allem in den Gruppen an Beweisversuchen mit einem weitgehend korrekten Argumentgefüge und einer der Schlussweisen Sw1-3 vor. Die Beweisversuche mit Reihenfolgebeispielen und dem kleinsten Beispiel enthalten häufiger ein weitgehend korrektes Argumentgefüge und häufiger die primäre Beweisidee 1 als die anderen Beweisversuche mit begründendem Ansatz. Wenn in Beweisversuchen mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel zusätzlich die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* die primäre Beweisidee bildet, dann enthalten diese fast immer ein weitgehend korrektes Argumentgefüge.[147] In der Klassenstufe 5 sind im Vergleich zu den Klassenstufen 7, 8 der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge und der Anteil der Beweisversuche mit der primären Beweisidee 1 an allen Beweisversuchen mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel kleiner.

Art der Beispiele: Großes Beispiel

Neben Reihenfolgebeispielen und dem kleinsten Beispiel lassen sich u. a. große Beispiele als eine besondere Art an Beispielen ansehen. In obiger Abbildung 120 ist angegeben, inwieweit die Beweisversuche in den einzelnen Gruppen KV1, Sw1-2 bis KV5-6, Sw4 ein großes Beispiel enthalten. Im jeweils unteren Balkendiagramm sind der Anteil der Beweisversuche mit einem großen Beispiel in dunkelrot, der Anteil der übrigen Beweisversuche mit Beispiel in hellrot und der Anteil der Beweisversuche ohne Beispiel in grau dargestellt. Es ist zu sehen, dass in keinem der analysierten Beweisversuche mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel*

[147] Dabei ist zu bedenken, dass es sich nicht unbedingt um Kausalzusammenhänge handelt. Die Ergebnisse lassen u. a. offen, inwieweit zuerst Reihenfolgebeispiele betrachtet und dadurch die Beweisidee 1 gefunden wurde.

ein großes Beispiel vorkommt. Dies erscheint plausibel, weil in den Beweisversuchen mit der Schlussweise Sw4 allenfalls notwendige Beispiele angeführt sind und die Verwendung eines großen Beispiels als notwendiges Beispiel, d. h. beispielsweise als Induktionsanfang, zumindest aus Expertensicht weder sinnvoll noch naheliegend erscheint.

Außerdem kommt in keinem der analysierten Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6, Sw1-4) ein großes Beispiel vor. Rund 85% der Beweisversuche mit begründendem Ansatz und einem großen Beispiel enthalten allenfalls einen Ansatz für ein weiterentwickelbares Argumentgefüge (KV1-2). Der Anteil der Beweisversuche ohne weiterentwickelbares Argumentgefüge ist bei den Beweisversuchen mit begründendem Ansatz und einem großen Beispiel statistisch hoch signifikant größer als bei denen ohne großes Beispiel ($p = 0{,}003^{N-1}$, $V = 0{,}127$).

In mehr als der Hälfte der Beweisversuche mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und einem großen Beispiel ist die Schlussweise *deduktiv mit zusätzlichem Beispiel* (Sw3). Dies lässt vermuten, dass zumindest in diesen Beweisversuchen ein großes Beispiel eher angeführt und genutzt wurde, um die Gültigkeit der Aussage zu veranschaulichen, zu überprüfen oder zu stützen, als dazu eine Beweisidee zu generieren.

Ort der Beispiele

Die Abbildung 122 gibt Einblicke in den Ort der Beispiele in Beweisversuchen mit begründendem Ansatz differenziert nach den Gruppen KV1, Sw1-2 bis KV5-6, Sw4. Bezogen auf jede dieser Gruppen ist in blau dargestellt, in einem wie großen Anteil der Beweisversuche mindestens ein Beispiel vorkommt und sich alle vorkommenden Beispiele oberhalb des (übrigen) Argumentgefüges befinden. Zudem ist bezogen auf jede der Gruppen in grün dargestellt, ein wie großer Anteil der Beweisversuche mindestens ein Beispiel innerhalb des (übrigen) Argumentgefüges enthält. Im Folgenden wird zuerst auf die Anteile der Beweisversuche mit allen Beispielen oberhalb des (übrigen) Argumentgefüges und ergänzend auf Beweisversuche mit mindestens einem Beispiel oberhalb des (übrigen) Argumentgefüges eingegangen. Danach stehen die in Abbildung 122 erkennbaren Anteile der Beweisversuche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges im Fokus.

Ort der Beispiele: Alle Beispiele oberhalb des (übrigen) Argumentgefüges

Der Anteil der Beweisversuche in der Kategorie BspOrtges1 *alle Beispiele oberhalb des (übrigen) Argumentgefüges* liegt entsprechend Abbildung 122 in den Gruppen mit KV1-2,Sw3 *kein weiterentwickelbares Argumentgefüge bei deduktiver Schlussweise ohne zusätzliches Beispiel* über 50%. In jeder der Gruppen mit einem weiterentwickelbaren Argumentgefüge (KV3-6) liegt dieser Anteil unter 20%. Beweisversuche mit begründendem Ansatz und Beispiel, in denen alle Beispiele oberhalb des (übrigen) Argumentgefüges angegeben sind, enthalten statistisch hoch signifikant seltener ein weiterentwickelbares Argumentgefüge als die übrigen Beweisversuche mit begründendem Ansatz ($p < 0{,}001^{N-1}$, $V = 0{,}152$). Sie sind hinsichtlich der Existenz eines weiterentwickelbaren Argumentgefüges im Durchschnitt wenig tragfähig.

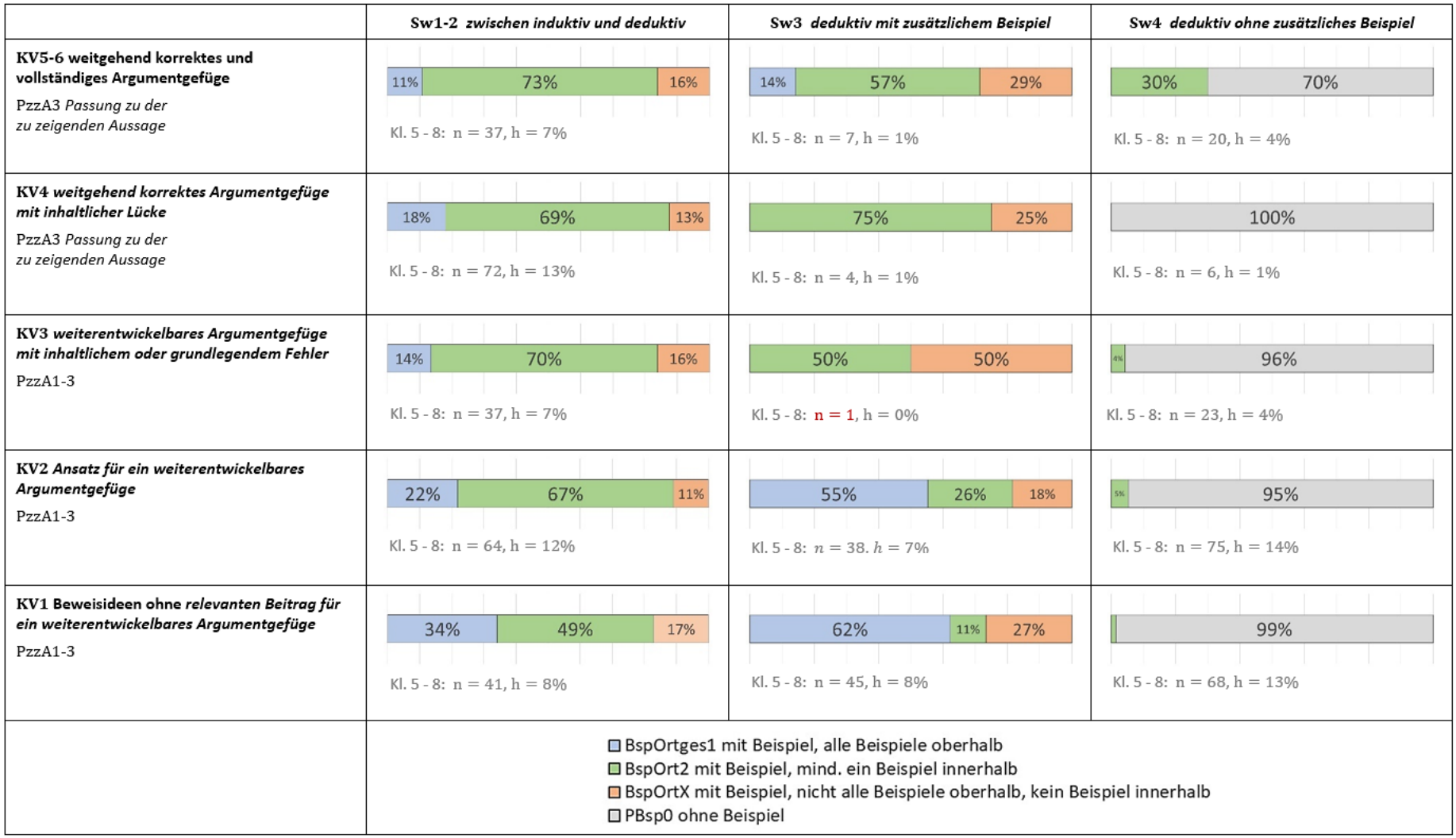


Abbildung 122: Beweisversuche mit begründendem Ansatz: Existenz und Ort angegebener Beispiele in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen

Dies bedeutet nicht, dass sämtliche Beweisversuche mit BspOrtges1 hinsichtlich der Existenz eines weiterentwickelbaren Argumentgefüges schwach sind. In 23% der Beweisversuche mit BspOrtges1 ist ein weiterentwickelbares Argumentgefüge (KV3-6) formuliert, in 18% ein weitgehend korrektes Argumentgefüge (KV4-6). In allen diesen Beweisversuchen mit BspOrtges1 und einem weitgehend korrekten Argumentgefüge sind oberhalb des (übrigen) Argumentgefüges Reihenfolgebeispiele angeführt. Die primäre Beweisidee ist in 75% der Fälle die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, ansonsten meistens die Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare*. Die Schlussweise liegt in 90% der Beweisversuche mit BspOrtges1 und einem weitgehend korrekten Argumentgefüge zwischen induktiv und deduktiv. Im Vergleich dazu enthalten von den Beweisversuchen mit BspOrtges1 und keinem weiterentwickelbaren Argumentgefüge (KV1-2) rund 30% Reihenfolgebeispiele. In keinem dieser Beweisversuche, d. h. auch in keinem mit Reihenfolgebeispielen, ist die primäre Beweisidee die Beweisidee 1. Die Schlussweise ist in ca. der Hälfte der Fälle deduktiv mit zusätzlichem Beispiel (Sw3).

Die Ergebnisse weisen also darauf hin, dass die Beweisversuche mit begründendem Ansatz und Beispiel, in denen sich alle Beispiele oberhalb des (übrigen) Argumentgefüges befinden, meistens kein weiterentwickelbares Argumentgefüge enthalten. Dabei scheint es jedoch auf die Art und die Nutzung der oberhalb des (übrigen) Argumentgefüges angeführten Beispielen anzukommen. Insbesondere ist in sämtlichen analysierten Beweisversuchen, in denen sich zwar alle Beispiele oberhalb des (übrigen) Argumentgefüges befinden, aber Reihenfolgebeispiele und als primäre Beweisidee die Beweisidee 1 ausgeführt sind, ein weitgehend korrektes Argumentgefüge (KV4-6) formuliert.

Ort der Beispiele: Mindestens ein Beispiel oberhalb des (übrigen) Argumentgefüges

Dass es auf die Art und Nutzung von Beispielen ankommt, deutet auch der Vergleich der Beweisversuche mit mindestens einem Beispiel oberhalb des (übrigen) Argumentgefüges mit den übrigen Beweisversuchen mit begründendem Ansatz an.

Von den Beweisversuchen, bei denen an einem Beispiel oberhalb des (übrigen) Argumentgefüges eine **Beweisidee hervorgehoben** ist, enthalten über 60% ein weitgehend korrektes Argumentgefüge und über 85% zumindest ein weiterentwickelbares Argumentgefüge. Diese Beweisversuche, bei denen an einem oberhalb des (übrigen) Argumentgefüges angeführten Beispiel eine Beweisidee hervorgehoben ist, enthalten im Vergleich zu den übrigen Beweisversuchen mit begründendem Ansatz zu einem statistisch hoch signifikant größeren Anteil ein weiterentwickelbares Argumentgefüge bzw. ein weitgehend korrektes Argumentgefüge (jeweils $p < 0{,}001^{N-1}$, $0{,}17 < V < 0{,}20$). Wenn man zusätzlich die primären Beweisideen berücksichtigt, fällt auf, dass in allen Beweisversuchen, bei denen an einem oberhalb des (übrigen) Argumentgefüges angeführten Beispiel eine Beweisidee hervorgehoben ist, eine ausbaubare Beweisidee die primäre Beweisidee ist. In fast der Hälfte der Fälle handelt es sich um die primäre Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare*, in weiteren rund 25% um die primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Aber auch fast jede der anderen ausbaubaren primären Beweisideen (BIp2,3,4,5,7,8) kommt vor.

Von den Beweisversuchen mit **Reihenfolgebeispielen** oberhalb des (übrigen) Argumentgefüges enthalten 45% ein weitgehend korrektes Argumentgefüge und 55% zumindest ein weiterentwickelbares Argumentgefüge. Diese Beweisversuche mit Reihenfolgebeispielen oberhalb des (übrigen) Argumentgefüges enthalten im Vergleich zu den übrigen Beweisversuchen mit begründendem Ansatz zu einem statistisch hoch signifikant größeren Anteil ein weitgehend korrektes Argumentgefüge bzw. ein weiterentwickelbares Argumentgefüge (jeweils $\mathrm{p} < 0{,}001^{\text{N-1}}$, $0{,}13 < \mathrm{V} < 0{,}24$). Bezüglich der primären Beweisideen fällt auf, dass mehr als die Hälfte der Beweisversuche mit Reihenfolgebeispielen oberhalb des (übrigen) Argumentgefüges auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* beruhen.

Von den Beweisversuchen, in denen oberhalb des (übrigen) Argumentgefüges das **kleinste Beispiel** steht, enthalten 42% ein weitgehend korrektes Argumentgefüge. Dieser Anteil ist bei den Beweisversuchen mit dem kleinsten Beispiel oberhalb des (übrigen) Argumentgefüges nach dem nach dem N-1 Chi²-Test statistisch hoch signifikant größer mit $\mathrm{V} = 0{,}141$ als bei den übrigen Beweisversuchen mit begründendem Ansatz. In rund der Hälfte der Beweisversuche mit dem kleinsten Beispiel oberhalb des (übrigen) Argumentgefüges ist die primäre Beweisidee die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*.

Von den Beweisversuchen mit einem **großen Beispiel** oberhalb des (übrigen) Argumentgefüges hingegen enthalten 95% kein weiterentwickelbares Argumentgefüge. Diese Beweisversuche mit einem großen Beispiel oberhalb des (übrigen) Argumentgefüge enthalten in den Klassenstufen 7, 8 zu einem statistisch hoch signifikant größeren Anteil kein weiterentwickelbares Argumentgefüge als die übrigen Beweisversuche mit begründendem Ansatz (jeweils $p < 0{,}01$, $0{,}22 < V < 0{,}34$).

Ort der Beispiele: Mindestens ein Beispiel innerhalb des (übrigen) Argumentgefüges

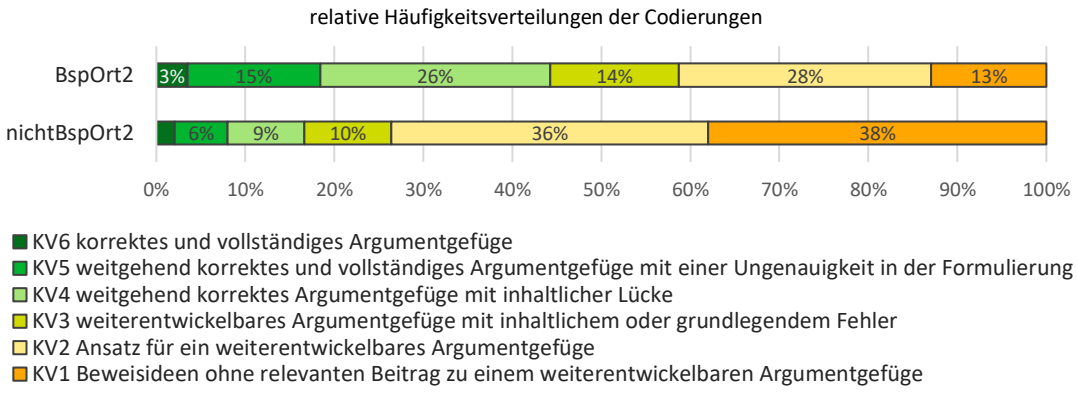
Nachdem die Beweisversuche mit begründendem Ansatz und einem Beispiel oberhalb des (übrigen) Argumentgefüges betrachtet wurden, stehen nun die Beweisversuche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges im Fokus. Deren Anteil ist in Abbildung 122 bezogen auf jede der Gruppen $\mathrm{KV1}, \mathrm{Sw1\text{-}2}$ bis $\mathrm{KV5\text{-}6}, \mathrm{Sw4}$ in grün dargestellt.

Es ist zu sehen, dass, wenn in einem Beweisversuch mit der **Schlussweise $\mathbf{Sw4}$** *deduktiv ohne zusätzliches Beispiel* ein Beispiel vorkommt, sich immer eines innerhalb des (übrigen) Argumentgefüges befindet. Dieses Ergebnis ist vor dem Hintergrund, dass alle angegeben Beispiele notwendige Beispiele sind, plausibel.

Von den Beweisversuchen mit begründendem Ansatz und einer der **Schlussweisen $\mathbf{Sw1\text{-}3}$** enthalten rund 55% ein Beispiel innerhalb des (übrigen) Argumentgefüges, wobei der Anteil in den Gruppen $\mathrm{KV1\text{-}2}, \mathrm{Sw3}$ vergleichsweise niedrig und in den Gruppen $\mathrm{KV2\text{-}5}, \mathrm{Sw1\text{-}2}$ und $\mathrm{KV4}, \mathrm{Sw3}$ vergleichsweise hoch ist.

Andersherum betrachtet gilt: In über 80% der Beweisversuche mit begründendem Ansatz und einem Beispiel innerhalb des (übrigen) Argumentgefüges liegt die **Schlussweise** zwischen induktiv und deduktiv ($\mathrm{Sw1\text{-}2}$), in 11% ist sie deduktiv mit zusätzlichem Beispiel ($\mathrm{Sw3}$), in 6%

deduktiv ohne zusätzliches Beispiel (Sw4). Der Anteil der Beweisversuche mit zwischen induktiver und deduktiver Schlussweise ist bei den Beweisversuchen mit begründendem Ansatz und einem Beispiel innerhalb des (übrigen) Argumentgefüges nach dem N-1 Chi²-Test statistisch hoch signifikant größer mit $V = 0{,}556$ als bei den Beweisversuchen mit begründendem Ansatz und ohne Beispiel innerhalb des (übrigen) Argumentgefüges.



p-Werte von Signifikanztests und dazugehörige Effektstärken

alle Kategorien KV1-6	KV1 vs. KV2-6	KV1-2 vs. KV3-6	KV1-3 vs. KV4-6
$p < .001^{U}$, $r = .394$	$p < .001^{N\text{-}1}$, $V = .268$	$p < .001^{N\text{-}1}$, $V = .321$	$p < .001^{N\text{-}1}$, $V = .306$

Abbildung 123: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in Abhängigkeit der Existenz eines Beispiels innerhalb des (übrigen) Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

In der ergänzenden Abbildung 123 ist bezogen auf alle Beweisversuche mit begründendem Ansatz die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in Abhängigkeit der Existenz mindestens eines Beispiels innerhalb des (übrigen) Argumentgefüges dargestellt. Bei den Beweisversuchen mit einem Beispiel innerhalb des (übrigen) Argumentgefüges ist im Vergleich zu den übrigen Beweisversuchen mit begründendem Ansatz nach dem Mann-Whitney-U-Test die Wahrscheinlichkeit statistisch hoch signifikant größer, hinsichtlich der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in eine höhere der Kategorien **KV1 bis KV6** zu gehören. Von den Beweisversuchen mit begründendem Ansatz und einem Beispiel innerhalb des (übrigen) Argumentgefüges enthalten ca. 10% keinen relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge (KV1), ca. 60% ein weiterentwickelbares Argumentgefüge (KV3-6). Wenn mindestens ein Beispiel innerhalb des (übrigen) Argumentgefüges angegeben ist, dann gehören die Beweisversuche mit begründendem Ansatz insbesondere statistisch hoch signifikant seltener in die Kategorie KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge* und enthalten statistisch hoch signifikant häufiger ein weiterentwickelbares Argumentgefüge.[148]

[148] En détail gilt nach dem Mann-Whitney-U-Test bezogen auf jede der Klassenstufen 5 bis 8, dass bei den Beweisversuchen mit einem Beispiel innerhalb des (übrigen) Argumentgefüges im Vergleich zu den (übrigen) Beweisversuchen mit begründendem Ansatz die Wahrscheinlichkeit statistisch signifikant größer ist, hinsichtlich der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in eine höhere der Kategorien KV1 bis KV6 zu gehören (jeweils $p \leq 0{,}026^{U}$; $0{,}17 < r < 0{,}36$). In jeder der Klassenstufen 5 bis 7 ist der Anteil der Beweisversuche

Es ist denkbar, dass bei den Beweisversuchen mit einem Beispiel innerhalb des (übrigen) Argumentgefüges vergleichsweise häufig ein Beispiel bewusst genutzt wurde, um eine Beweisidee zu generieren oder zu präsentieren und dadurch die Schlussweise vergleichsweise häufig zwischen induktiv und deduktiv liegt. Von den Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise enthält ein größerer Anteil ein weitgehend korrektes Argumentgefüge als von denjenigen mit deduktiver Schlussweise (Abschn. 8.2.1). Die Nutzung eines Beispiels zur Entwicklung oder Darstellung einer Beweisidee könnte zudem eine intensive Beschäftigung mit der Aufgabe implizieren, was meint, dass sich die Bearbeitung nicht etwa auf die Angabe eines Beispiels und einer kurzen Plausibilitätsüberlegung beschränkt hat. Daraus könnte resultieren, dass die Beweisversuche vergleichsweise selten keinen relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge enthalten.

Primäre Beweisideen und die Existenz, die Art und der Ort von Beispielen

In den vorhergehenden Ausführungen ist z. B. deutlich geworden, dass bei den Beweisversuchen mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel der Anteil der Beweisversuche mit der primären Beweisidee 1 größer ist als bei den übrigen Beweisversuchen mit begründendem Ansatz. Andersherum, d. h. ausgehend von den primären Beweisideen betrachtet, fallen Beweisversuche mit begründendem Ansatz einer der primären Beweisideen 1, 4, 12 hinsichtlich der Existenz, der Art und dem Ort von Beispielen auf.

In Abbildung 124 sind mehrere Unterschiede zwischen Beweisversuchen mit begründendem Ansatz mit der primären Beweisidee 1 und denen mit einer der primären Beweisideen 2 bis 8 sowie denen mit einer der primären Beweisideen 9 bis 13, 55 dargestellt. Von den Beweisversuchen mit begründendem Ansatz und der **primären Beweisidee 1** *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* enthalten knapp 85% ein Beispiel zu der zu zeigenden Aussage, von denen mit einer der primären Beweisideen 2 bis 8 sowie von denen mit einer der primären Beweisideen 9 bis 13, 55 jeweils etwa 50%. In rund 70% der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 kommen Reihenfolgebeispiele vor, dabei sind in mehr als der Hälfte der Fälle drei oder mehr Reihenfolgebeispiele angegeben, in ebenfalls mehr als der Hälfte der Fälle Reihenfolgebeispiele und das kleinste Beispiel. Im Vergleich dazu enthalten von den Beweisversuchen mit begründendem Ansatz und einer der primären Beweisideen 2 bis 8 bzw. 9 bis 13, 55 ca. 10% Reihenfolgebeispiele, der Anteil derjenigen mit drei oder mehr Reihenfolgebeispiele sowie derjenigen mit Reihenfolgebeispielen und dem kleinsten Beispiel ist jeweils noch geringer. Bezüglich des Orts der Beispiele fällt auf, dass sich in knapp einem Drittel der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 oberhalb des (übrigen) Argumentgefüges Reihenfolgebeispiele befinden, während dies auf weniger als ein Zehntel der Beweisversuchen mit begründendem Ansatz und ei-

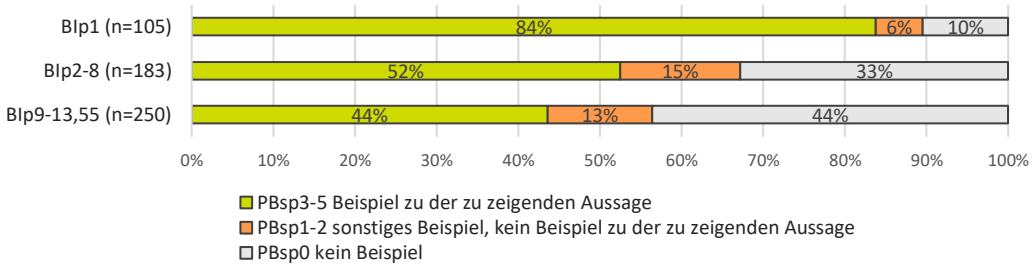
in der Kategorie $\mathrm{KV}1$ bei den Beweisversuchen mit einem Beispiel innerhalb des (übrigen) Argumentgefüges im Vergleich zu den (übrigen) Beweisversuchen statistisch signifikant kleiner (jeweils $p \leq 0{,}026^{N\text{-}1}$; $0{,}16 < V \leq 0{,}30$). In jeder der Klassenstufen 5, 7, 8 sind der Anteil der Beweisversuche mit einem weiterentwickelbaren Argumentgefüge ($\mathrm{KV}3\text{-}6$) und der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge ($\mathrm{KV}4\text{-}6$) bei den Beweisversuchen mit einem Beispiel innerhalb des (übrigen) Argumentgefüges im Vergleich zu den übrigen Beweisversuchen mit begründendem Ansatz statistisch signifikant kleiner (jeweils $p \leq 0{,}005^{N\text{-}1}$; $0{,}28 < V \leq 0{,}38$).

Eigenschaften von Beweisversuchen mit begründendem Ansatz differenziert nach primären Beweisideen

Fokus: Primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*

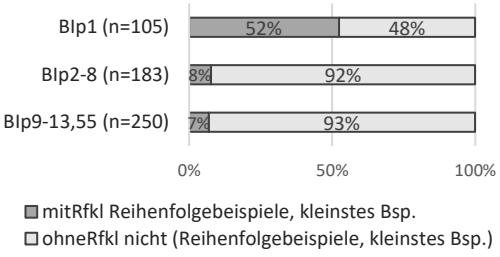
relative Häufigkeitsverteilungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Existenz eines Beispiels zu der zu zeigenden Aussage oder eines sonstigen Beispiels



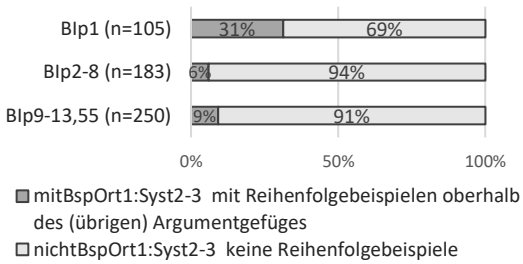
	PBsp0 vs. PBsp1-5	PBsp0 vs. PBsp3-5
BIp1 vs. BIp2-8	$p < .001^{N-1}$, $V = .249$	$p < .001^{N-1}$, $V = .314$
BIp1 vs. BIp9-13,55	$p < .001^{N-1}$, $V = .320$	$p < .001^{N-1}$, $V = .369$

Existenz von Reihenfolgebeispielen zusammen mit dem kleinsten Beispiel



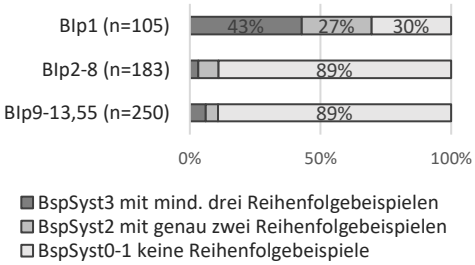
	mitRfkl vs. ohneRfkl
BIp1 vs. BIp2-8	$p < .001^{N-1}$, $V = .499$
BIp1 vs. BIp9-13,55	$p < .001^{N-1}$, $V = .517$

Existenz von Reihenfolgebeispielen oberhalb des (übrigen) Argumentgefüges



	mitBspOrt1: Syst2-3 vs. nichtBspOrt1: Syst2-3
BIp1 vs. BIp2-8	$p < .001^{N-1}$, $V = .340$
BIp1 vs. BIp9-13,55	$p < .001^{N-1}$, $V = .278$

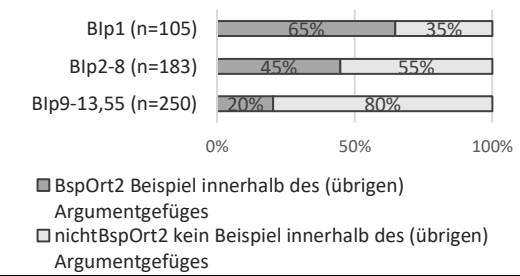
Existenz zweier oder mehr Reihenfolgebeispiele



	BspSyst2-3 vs. BspSyst0-1
BIp1 vs. BIp2-8	$p < .001^{N-1}$, $V = .603$
BIp1 vs. BIp9-13,55	

	BspSyst3 vs. BspSyst0-2
BIp1 vs. BIp2-8	$p < .001^{N-1}$, $V = .499$
BIp1 vs. BIp9-13,55	$p < .001^{N-1}$, $V = .449$

Existenz eines Beispiels innerhalb des (übrigen) Argumentgefüges



	BspOrt2 vs. nichtBspOrt2
BIp1 vs. BIp2-8	$p = .001^{N-1}$, $V = .192$
BIp1 vs. BIp9-13,55	$p < .001^{N-1}$, $V = .429$

Abbildung 124: Beweisversuche mit begründendem Ansatz: Existenz, Art und Ort von Beispielen differenziert nach primären Beweisideen. Relative Häufigkeitsverteilungen der Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

ner der primären Beweisideen 2 bis 8 bzw. 9 bis 13, 55 zutrifft. Zudem ist zu sehen, dass sich in knapp zwei Drittel der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 mindestens ein Beispiel innerhalb des (übrigen) Argumentgefüges befindet, bei den Be-

weisversuchen mit begründendem Ansatz und einer der primären Beweisideen 2 bis 8 in etwa der Hälfte, bei denen mit einer der primären Beweisideen 9 bis 13, 55 in einem Fünftel der Fälle. Dabei sind sämtliche aufgeführten Unterschiede zwischen den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 und denen mit einer der primären Beweisideen 2 bis 8 bzw. einer der primären Beweisideen 9 bis 13, 55 statistisch hoch signifikant mit geringer bis hoher Effektstärke (Abb. 124).

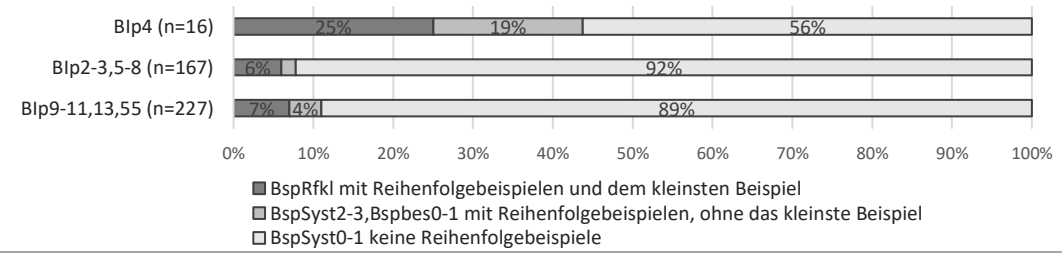
In Ergänzung zu den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 sind Beweisversuche mit begündendem Ansatz und der primären Beweisidee 4 *Zahldarstellung im Dezimalsystem* sowie Beweisversuche mit begründendem Ansatz und der primären Beweisidee 12 *Vierer- oder Achterreihe* hervorzuheben. In Abbildung 125 ist dargestellt, inwieweit Beweisversuche mit begründendem Ansatz und der primären Beweisidee 4 bzw. 12 Reihenfolgebeispiele enthalten.

Beweisversuche mit begründendem Ansatz und der **primären Beweisidee 4** *Zahldarstellung im Dezimalsystem* machen 3% aller Beweisversuche mit begründendem Ansatz aus. Die Beweisversuche mit der primären Beweisidee 4 fallen dadurch wenig ins Gewicht. Trotzdem ist festzuhalten, dass auch bei Beweisversuchen mit der primären Beweisidee 4 der Anteil der Beweisversuche mit Reihenfolgebeispielen mit ca. 45% und insbesondere der Anteil derjenigen mit Reihenfolgebeispielen und dem kleinsten Beispiel mit rund 25% vergleichsweise groß ist. Die Unterschiede zu den anderen Beweisversuchen mit begründendem Ansatz und einer der primären Beweisideen 2 bis 8 sowie zu denen mit einer der primären Beweisideen 9 bis 13, 55 sind statistisch hoch signifikant (Abb. 125).

Beweisversuche mit begründendem Ansatz: Existenz von Reihenfolgebeispielen und des kleinsten Beispiels

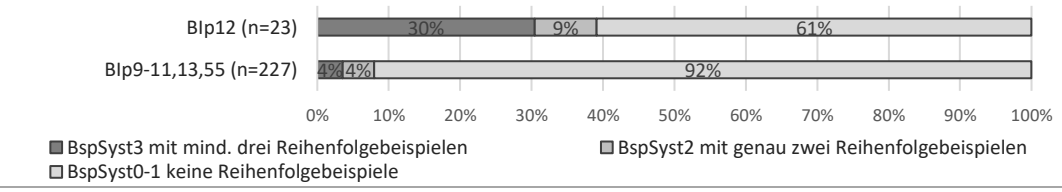
relative Häufigkeitsverteilungen von Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Fokus: Primäre Beweisidee 4 *Zahldarstellung im Dezimalsystem*



	BspRfkl vs. ¬BspRfkl	BspSyst2-3 vs. BspSyst0-1
BIp4 vs. BIp2-8	$p = .006^{ex}$, V = .202	$p < .001^{ex}$, V = .326
BIp4 vs. BIp9-13,55	$p = .001^{ex}$, V = .208	$p < .001^{ex}$, V = .292

Fokus: Primäre Beweisidee 12 *Vierer- oder Achterreihe*



	BspSyst2-3 vs. BspSyst0-1	BspSyst3 vs. BspSyst0-2
BIp12 vs. BIp9-11,13,55	$p < .001^{ex}$, V = .291	$p < .001^{ex}$, V = .328

Abbildung 125: Beweisversuche mit begründendem Ansatz: Existenz von Reihenfolgebeispielen und des kleinsten Beispiels (Fokus: Primäre Beweisideen 4, 12). Relative Häufigkeitsverteilungen von Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Beweisversuche mit begründendem Ansatz und der **primären Beweisidee 12** *Vierer- oder Achterreihe* machen rund 4% der Beweisversuche mit begründendem Ansatz aus und rund 9% derjenigen mit einer der primären Beweisideen 9 bis 13, 55. Die Beweisversuche mit der primären Beweisidee 12 fallen dadurch ebenfalls wenig ins Gewicht. Es fällt aber auf, dass mit rund 40% der Anteil derjenigen mit Reihenfolgebeispielen und mit rund 30% der Anteil derjenigen mit drei oder mehr Reihenfolgebeispielen vergleichsweise groß ist. Beide Anteile sind statistisch hoch signifikant größer als bei anderen Beweisversuchen mit begründendem Ansatz und einer der primären Beweisideen 9 bis 13, 55 (Abb. 125). Dabei befinden sich die Reihenfolgebeispiele in allen analysierten Beweisversuchen mit begründendem Ansatz, der primären Beweisidee 12 und Reihenfolgebeispielen oberhalb des (übrigen) Argumentgefüges. Der Anteil derjenigen mit einem Beispiel innerhalb des (übrigen) Argumentgefüges ist mit ca. 15% im Vergleich zu 65% bei den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 eher klein. Es lässt sich vermuten, dass im Falle der primären Beweisidee 12 *Vierer- oder Achterreihe* häufig die Invariante, dass sich die Summe von vier aufeinanderfolgenden ungeraden Zahlen immer in der Achter- bzw. Viererreihe befindet, an Reihenfolgebeispielen erkannt, jedoch nicht hinterfragt und insbesondere nicht näher begründet wurde.

Diese Auffälligkeiten hinsichtlich der Beispielnutzung bei Beweisversuchen mit begründendem Ansatz und einer der **primären Beweisideen 1, 4, 12** stehen in Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion (Abschn. 6.2.2). Beispielsweise ist die vergleichsweise häufige Existenz von Reihenfolgebeispielen plausibel, weil alle drei Beweisideen im Wesentlichen auf Invarianten beruhen können, die sich besonders gut anhand von Reihenfolgebeispielen erkennen lassen.

Fazit: Beispiele und Korrektheit, Vollständigkeit, Schlussweise, primäre Beweisidee

Ausgehend von der Analyse der Verwendung von Beispielen in Beweisversuchen mit begründendem Ansatz zusammen mit der Schlussweise, der Korrektheit und Vollständigkeit der Beweisversuche und der primären Beweisidee, lassen sich u. a. folgende Auffälligkeiten und Tendenzen festhalten:

Beweisversuche mit begründendem Ansatz und einer der Schlussweisen Sw1-3 enthalten per Definition ein Beispiel. Beweisversuche mit begründendem Ansatz und der **Schlussweise** Sw4 *deduktiv ohne zusätzliches Beispiel* hingegen können allenfalls notwendige Beispiele enthalten. In über 90% dieser Beweisversuche kommt kein Beispiel vor. Falls ein notwendiges Beispiel angegeben ist, ist die primäre Beweisidee meistens die Beweisidee 1 und ein weitgehend korrektes und vollständiges Argumentgefüge formuliert.

Beweisversuche mit begründendem Ansatz und einem Beispiel zu der zu zeigenden Aussage, solche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges sowie solche mit Reihenfolgebeispielen und dem kleinsten Beispiel sind hinsichtlich der **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** eher stark. Beweisversuche mit begründendem Ansatz ohne Beispiel oder ausschließlich mit sonstigen Beispielen, solche mit einem großen Beispiel und solche mit allen Beispielen oberhalb des (übrigen) Argumentgefüges sind

hinsichtlich der Korrektheit und Vollständigkeit der Beweisversuche im Durchschnitt eher schwach. Dabei scheint es jedoch auf den Umgang mit den Beispielen anzukommen.

Zudem scheinen die Existenz, die Art und der Ort von Beispielen in einem gewissen Maße mit **primären Beweisideen** zusammenzuhängen, wobei die Ergebnisse zu den Überlegungen im Rahmen der Aufgabenkonstruktion passen. Beispielsweise enthalten die Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* wesentlich häufiger Reihenfolgebeispiele und das kleinste Beispiel als diejenigen mit einer anderen ausbaubaren primären Beweisidee.

8.2.6.5 Detailanalyse: Darstellungsformen

In diesem Abschnitt wird die Darstellungsform der Beweisversuche mit begründendem Ansatz in Zusammenhang mit der Schlussweise und der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüge betrachtet. Zudem werden die Passung der Ausführungen zu der zu zeigenden Aussage, die primäre Beweisidee und die Verwendung von Beispielen berücksichtigt. Die Ergebnisse deuten an, dass sich insbesondere die Beweisversuche mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* von denjenigen mit begründendem Ansatz und einer anderen Schlussweise mit Blick auf die Darstellungsformen unterscheiden. Zudem sind Differenzen zwischen den Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6), einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler oder einer inhaltlichen Lücke (KV3-4) und ohne weiterentwickelbares Argumentgefüge (KV1-2) zu erkennen. Auch ist die Darstellungsform bei einzelnen primären Beweisideen auffällig. Die wesentlichen Ergebnisse der Detailanalyse der Darstellungsform zusammen mit weiteren Eigenschaften von Beweisversuchen mit begründendem Ansatz sind am Ende dieses Abschnitts zusammengefasst.

Für die Detailanalyse ist in Abbildung 126 bezogen auf jede der Gruppen KV1, Sw1-2 bis KV5-6, Sw4 angegeben, welche Darstellungsformen vorkommen und einen wie großen Anteil der Beweisversuche sie ausmachen. Dabei sind die Darstellungsformen zugunsten der Übersichtlichkeit und Aussagekraft der Ergebnisse folgendermaßen gruppiert:

- Dstf1,14,134,1234 *Darstellungsformen mit symbolischem Anteil*[149]
- Dstf23,234 *Darstellungsformen ohne symbolischen, aber mit numerisch-ikonischem Anteil*
- Dstf34 *Darstellungsform numerisch-verbal*
- Dstf4,24 *Darstellungsformen verbal, ikonisch-verbal*[150]

[149] Die Bezeichnung „Darstellungsformen mit symbolischem Anteil“ wird im Folgenden bei allen Beweisversuche verwendet, in denen mindestens eine sinnbezogene Einheit überwiegend symbolisch dargestellt ist (z. B. Dstf1 *symbolisch*, Dstf14 *symbolisch-verbal*). Die Bezeichnung „Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil“ wird analog verstanden.

[150] Die Darstellungsform ist in den analysierten Beweisversuchen mit Dstf4,24 fast immer verbal, fast nie ikonisch-verbal (Abschn. 8.2.4).

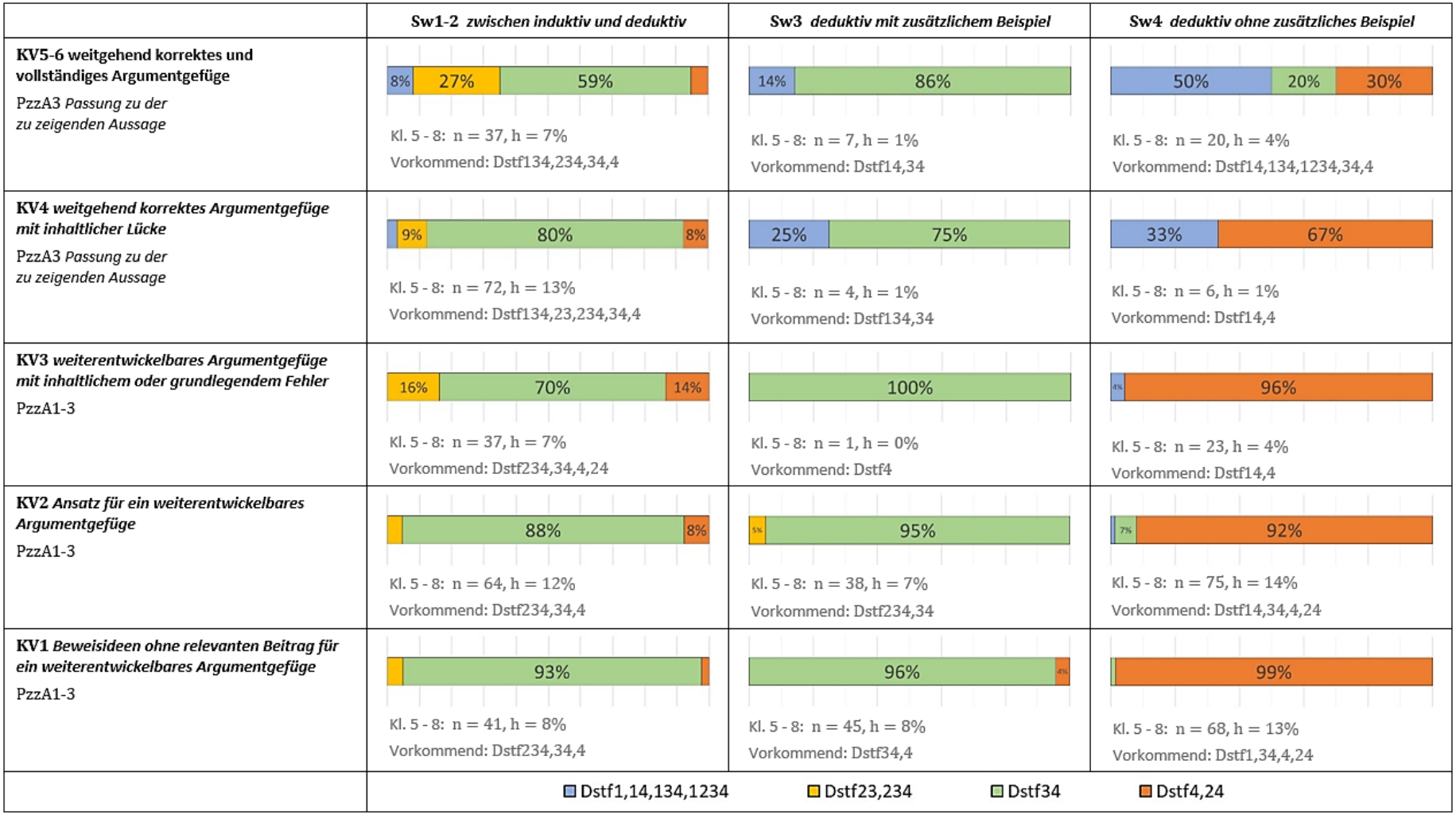
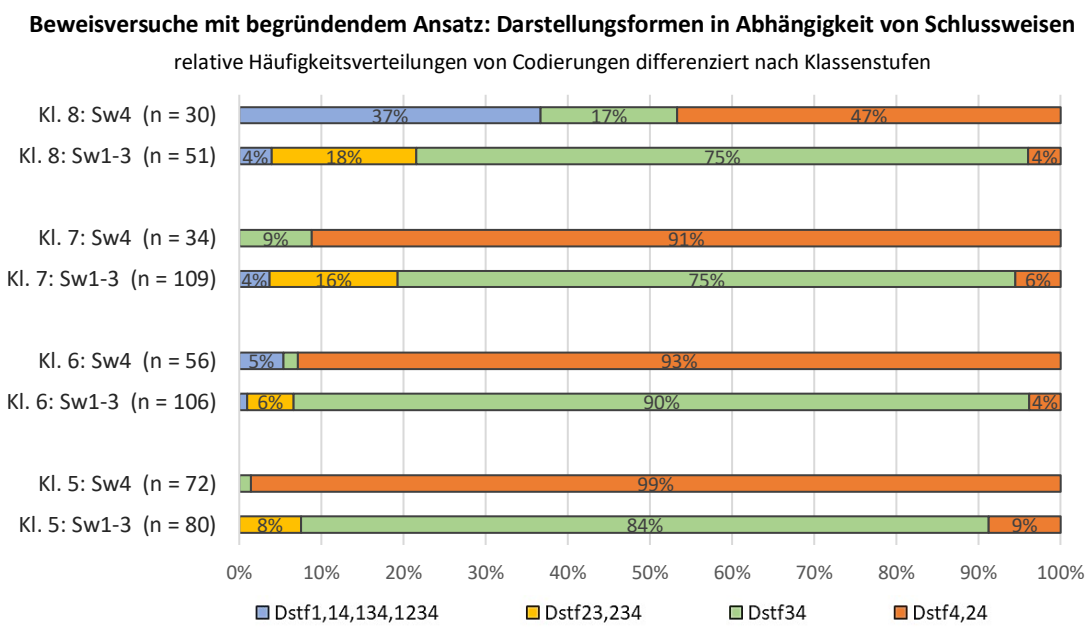


Abbildung 126: Beweisversuche mit begründendem Ansatz: Darstellungsformen in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise. Relative Häufigkeitsverteilungen von Codierungen

Schlussweisen und Darstellungsformen

Im Folgenden stehen zuerst Zusammenhänge zwischen Schlussweisen und Darstellungsformen im Fokus. In Abbildung 126 ist folglich primär spaltenweise zu schauen. Wenn man die relativen Häufigkeitsverteilungen der unterschiedenen Darstellungsformen je nach Schlussweise vergleicht, fällt auf, dass die große Mehrheit der Beweisversuche mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* in die Kategorie Dstf4,24 *Darstellungsformen verbal, ikonisch-verbal* gehört, während im Fall der Schlussweise *Sw*1-2 *zwischen induktiv und deduktiv* sowie der Schlussweise *Sw*3 *deduktiv mit zusätzlichem Beispiel* die Darstellungsform *numerisch-verbal* (Dstf34) dominiert. Der Anteil der **Darstellungsform numerisch-verbal** ist in jeder der Klassenstufen 5 bis 8 unter den Beweisversuchen mit einer der Schlussweisen Sw1-3 statistisch hoch signifikant größer mit hoher bis sehr hoher Effektstärke als unter den Beweisversuchen mit der Schlussweise Sw4. Zugleich ist je Klassenstufe der Anteil der **Darstellungsformen verbal, ikonisch-verbal** bei den Beweisversuchen mit einer der Schlussweisen Sw1-3 statistisch hoch signifikant kleiner mit hoher bis sehr hoher Effektstärke als bei den Beweisversuchen mit der Schlussweise Sw4, wie die ergänzende Abbildung 127 zeigt.[151]



p-Werte von Signifikanztests und dazugehörige Effektstärken

	Dstf34 vs. ¬Dstf34	Dstf4,24 vs. ¬Dstf4,24	Dstf1,14,134,1234 vs. ¬Dstf1,14,134,1234
Kl. 8	$p < .001^{N-1}, V = .560$	$p < .001^{N-1}, V = .518$	$p < .001^{ex}, V = .431$
Kl. 7	$p < .001^{N-1}, V = .576$	$p < .001^{N-1}, V = .833$	
Kl. 6	$p < .001^{N-1}, V = .850$	$p < .001^{N-1}, V = .891$	
Kl. 5	$p < .001^{N-1}, V = .827$	$p < .001^{N-1}, V = .898$	

	Kl. 8 vs. Kl. 5	Kl. 8 vs. Kl. 6	Kl. 8 vs. Kl. 7
Sw4: Dstf4,24 vs. ¬Dstf4,24	$p < .001^{N-1}, V = .635$	$p < .001^{N-1}, V = .521$	$p < .001^{N-1}, V = .487$

Abbildung 127: Beweisversuche mit begründendem Ansatz: Darstellungsformen in Abhängigkeit von Schlussweisen. Relative Häufigkeitsverteilungen von Codierungen differenziert nach Kl. sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

[151] Die Beweisversuche mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* sind in der Klassenstufe 8 zu 47%, in den Klassenstufen 5 bis 7 zu über 90% verbal oder ikonisch-verbal dargestellt. Dieser Unterschied zwischen den einzelnen Klassenstufen 5 bis 7 und der Klassenstufe 8 ist jeweils statistisch hoch signifikant mit mittlerer bis hoher Effektstärke (Abb. 126).

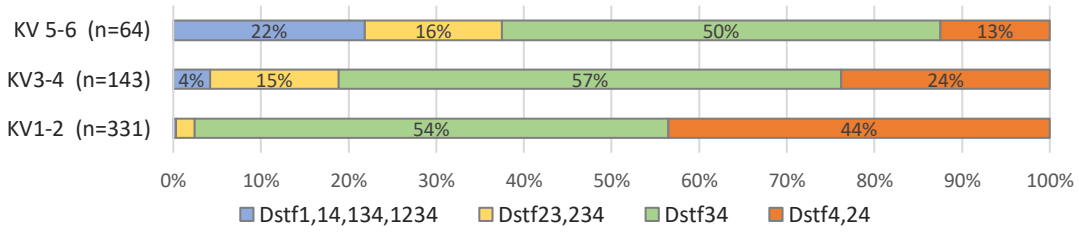
Die beschriebenen gegenläufigen Tendenzen bezüglich der Darstellungsformen Dstf34 *numerisch-verbal* und Dstf4,24 *verbal, ikonisch-verbal* sind in Zusammenhang mit der Angabe von Beispielen zu sehen. In jedem der Beweisversuche mit einer der Schlussweisen Sw1-3 muss per Definition mindestens ein Beispiel vorkommen, während Beweisversuche mit der Schlussweise Sw4 allenfalls Beispiele enthalten, die notwendige Bestandteile des Argumentgefüges sind (Abschn. 3.3.6). Wenn in einem Beweisversuch mit begründendem Ansatz ein Beispiel vorkommt, dann ist je Klassenstufe in rund 75% bis 90% der Fälle die Darstellungsform numerisch-verbal, während die Darstellungsformen *verbal, ikonisch-verbal* zusammen maximal 10% ausmachen. Dies erklärt den hohen Anteil der Kategorie Dstf34 bzw. den geringen Anteil der Kategorie Dstf4,24 unter den Beweisversuchen mit einer der Schlussweisen Sw1-3. Von den Beweisversuchen mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* enthalten in den Klassenstufen 5 bis 7 jeweils unter 10% ein Beispiel, in der Klassenstufe 8 rund 23%, was ergänzend die dortigen Anteile der Kategorie Dstf34 *numerisch-verbal* plausibilisiert.

Korrektheit, Vollständigkeit und Darstellungsformen

In diesem Abschnitt wird die Darstellungsform von Beweisversuchen mit begründendem Ansatz in Abhängigkeit derer Korrektheit und Vollständigkeit, d. h. insbesondere der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges betrachtet. In Abbildung 126 ist also primär zeilenweise zu schauen. So wie die in Abbildung 126 zu sehenden relativen Häufigkeitsverteilungen der Darstellungsformen vermuten lassen, unterscheiden sich die analysierten Beweisversuche mit begründendem Ansatz und keinem weiterentwickelbaren Argumentgefüge (KV1-2), einem weiterentwickelbaren Argumentgefüge mit einem inhaltlichen oder grundlegenden Fehler oder einer inhaltlichen Lücke (KV3-4) bzw. einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) paarweise statistisch hoch signifikant mit $0{,}29 \leq V \leq 0{,}51$ (Abb. 128).

Beweisversuche mit begründendem Ansatz: Darstellungsformen in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges

relative Häufigkeitsverteilungen von Codierungen



p-Werte von Signifikanztests und dazugehörige Effektstärken

	alle einzelnen Kategorien	Dstf4,24 vs. ¬Dstf4,24
KV1-2 vs. KV3-4	$p < .001^{ex}$, $V = .317$	$p < .001^{N-1}$, $V = .187$
KV1-2 vs. KV5-6	$p < .001^{ex}$, $V = .508$	$p < .001^{N-1}$, $V = .235$
KV3-4 vs. KV5-6	$p < .001^{N-1}$, $V = .291$	$p = .063^{N-1}$, $V = .130$
	Dstf1,14,134,1234 vs. ¬Dstf1,14,134,1234	Dstf23,234 vs. ¬Dstf23,234
KV1-2 vs. KV3-4	$p = .004^{ex}$, $V = .148$	$p < .001^{N-1}$, $V = .245$
KV1-2 vs. KV5-6	$p < .001^{ex}$, $V = .416$	$p < .001^{ex}$, $V = .245$
KV3-4 vs. KV5-6	$p < .001^{N-1}$, $V = .277$	$p = .861^{N-1}$, $V = .012$

Abbildung 128: Beweisversuche mit begründendem Ansatz: Darstellungsformen in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Relative Häufigkeitsverteilungen von Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Von den Beweisversuchen mit begründendem Ansatz, aber **keinem weiterentwickelbaren Argumentgefüge (KV1-2)** gehören etwa die Hälfte in die Kategorie Dstf4,24 *Darstellungsformen verbal, ikonisch-verbal*, während dies bei den Beweisversuchen mit einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler oder inhaltlicher Lücke (KV3-4) auf ca. ein Viertel, bei den Beweisversuchen mit einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) auf ca. ein Achtel der Fälle zutrifft. Der Anteil der Kategorie Dstf4,24 ist bei den Beweisversuchen mit KV1-2 statistisch hoch signifikant größer als bei den Beweisversuchen mit KV3-4 bzw. KV5-6 (Abb. 128).

Wenn man bei den Beweisversuchen mit begründendem Ansatz, aber allenfalls einem Ansatz für ein weiterentwickelbares Argumentgefüge (KV1-2) weiter differenziert, zum einen zwischen den Kategorien KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge* und KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge*, zum zweiten je nach Existenz eines Hinweises auf einen grundlegenden Fehler (PzzA1 vs. PzzA2-3) sind hingegen kaum Unterschiede und insbesondere keine statistisch signifikanten Unterschiede zu erkennen.

Bei den Beweisversuchen mit begründendem Ansatz und einem **weiterentwickelbaren Argumentgefüge** (KV3-6) beträgt der Anteil der Darstellungsformen ohne symbolischen, aber mit ikonisch-numerischem Anteil rund 15%, während diese Darstellungsformen bei den Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge (KV1-2) kaum vorkommen. Der Anteil der Beweisversuche mit einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil ist bei den Beweisversuchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler oder inhaltlicher Lücke (KV3-4) sowie bei denen mit einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) statistisch hoch signifikant größer mit $V = 0{,}245$ als bei denen ohne weiterentwickelbares Argumentgefüge (KV1-2; Abb. 128).

Bei den Beweisversuchen mit begründendem Ansatz und einem **weitgehend korrekten und vollständigen Argumentgefüge** (KV5-6) ist mit 22% der Anteil derjenigen mit einer Darstellungsform mit symbolischem Anteil am größten. Im Vergleich dazu beträgt er bei den Beweisversuchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler oder einer inhaltlichen Lücke rund 4%, bei denen ohne weiterentwickelbares Argumentgefüge (KV1-2) unter 1%. Die Unterschiede sind paarweise statistisch hoch signifikant (Abb. 128). Bei den Beweisversuchen mit begründendem Ansatz, einem weitgehend korrekten und vollständigen Argumentgefüge und der Schlussweise *deduktiv ohne zusätzliches Beispiel* im Speziellen beträgt der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil rund 50% (Abb. 126). Er ist statistisch signifikant größer als bei den übrigen Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge ($p = .001^{ex}$, $V = .459$) sowie als bei den anderen Beweisversuchen mit der Schlussweise *deduktiv ohne zusätzliches Beispiel* ($p = .001^{ex}$, $V = .560$).

Darstellungsform und Korrektheit, Vollständigkeit, Schlussweise, Beweisideen

Im Folgenden werden die Schlussweise und die Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz in Abhängigkeit derer Darstellungsform betrachtet. Die Anteile der unterschiedenen Darstellungsformen sind in obiger Abbildung 126 farbig dargestellt. Die Anteile der Darstellungsformen mit symbolischem Anteil sind blau gefärbt, die Anteile der Darstellungsformen ohne symbolischen, aber mit ikonisch-numerischem Anteil gelb, die Anteile der Darstellungsform numerisch-verbal grün und die Anteile der Darstellungsformen verbal, ikonisch-verbal rot. Es ist zu sehen, dass die Beweisversuche mit begründendem Ansatz und der **Darstellungsform *numerisch-verbal*** hinsichtlich der Korrektheit und Vollständigkeit und der Schlussweise vielfältig ausgeprägt sind. Demgegenüber lassen sich bei den Beweisversuchen mit begründendem Ansatz und einer der Darstellungsformen *verbal, ikonisch-verbal*, bei denen mit einer Darstellungsform mit symbolischem Anteil und bei denen mit einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil deutliche Tendenzen zu erkennen. Diese Tendenzen werden im Folgenden näher ausgeführt.

Beweisversuche mit begründendem Ansatz und einer der **Darstellungsformen *verbal, ikonisch-verbal*** erscheinen, vereinfacht gesagt, in Bezug auf die Schlussweise weitgehend homogen und in Bezug auf die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie die primäre Beweisidee eher schwach.

Bei rund 90% der Beweisversuche mit begründendem Ansatz und einer der Darstellungsformen *verbal, ikonisch-verbal* ist die Schlussweise $\mathrm{Sw4}$ *deduktiv ohne zusätzliches Beispiel*. Der Anteil dieser Schlussweise ist statistisch hoch signifikant größer als bei den Beweisversuchen mit begründendem Ansatz und einer anderen Darstellungsform ($\mathrm{p} < .001^{\mathrm{N\text{-}1}}, \mathrm{V} = .829$).

Von den Beweisversuchen mit begründendem Ansatz und einer der Darstellungsformen *verbal, ikonisch-verbal* enthalten mehr als drei Viertel kein weiterentwickelbares Argumentgefüge ($\mathrm{KV1\text{-}2}$). Dieser Anteil ist statistisch hoch signifikant größer als bei den übrigen Beweisversuchen mit begründendem Ansatz ($\mathrm{p} < .001^{\mathrm{N\text{-}1}}, \mathrm{V} = .237$).

Insgesamt enthalten rund 73% der Beweisversuche mit begründendem Ansatz und einer der Darstellungsformen *verbal, ikonisch-verbal* kein weiterentwickelbares Argumentgefüge bei deduktiver Schlussweise ohne zusätzliches Beispiel. In ca. 80% der Fälle ist eine der elementaren Beweisideen 9 bis $13, 55$ die primäre Beweisidee. Dieser Anteil an Beweisversuchen ohne ausbaubare Beweisidee als primäre Beweisidee ist statistisch hoch signifikant größer als bei den Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und der Schlussweise *deduktiv ohne zusätzliches Beispiel*, aber einer anderen Darstellungsform ($\mathrm{p} = .001^{\mathrm{ex}}, \mathrm{V} = .316$).

Beweisversuche mit begründendem Ansatz und einer **Darstellungsform mit symbolischem Anteil** kennzeichnet, vereinfacht zusammengefasst, die Zugehörigkeit zur Klassenstufe 8, die Schlussweise $Sw4$ *deduktiv ohne zusätzliches Beispiel* und die Existenz eines weitgehend korrekten Argumentgefüges.

Beweisversuche mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil sind vor allem in der Klassenstufe 8 zu finden (Abschn. 8.2.4; Abb. 127 auf S. 314). Sie machen in der Klassenstufe 8 unter den Beweisversuchen mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* einen statistisch hoch signifikant größeren Anteil aus als unter den restlichen Beweisversuchen mit begründendem Ansatz aus der Klassenstufe 8. Zudem machen sie unter den Beweisversuchen mit der Schlussweise $Sw4$ aus der Klassenstufe 8 einen statistisch hoch signifikant größeren Anteil aus als unter den Beweisversuchen mit der Schlussweise Sw4 aus jeder der Klassenstufen 5 bis 7.

Zudem fällt auf, dass über 90% der Beweisversuche mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil ein weitgehend korrektes Argumentgefüge (KV4-6) enthalten. In jeder der Klassenstufen 6 bis 8 ist der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge bei den Beweisversuchen mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil statistisch signifikant größer als bei den anderen Beweisversuchen mit begründendem Ansatz (Kl. 6: $p = .002^{ex}$, $V = .298$; Kl. 7: $p = .019^{N-1}$, $V = .218$; Kl. 8: $p = .013^{N-1}$, $V = .276$). In den Klassenstufen 6, 8 ist zudem der Anteil der Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge bei den Beweisversuchen mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil statistisch signifikant größer als bei den anderen Beweisversuchen mit begründendem Ansatz (Kl. 6: $p < .001^{ex}$, $V = .465$; Kl. 8: $p = .002^{ex}$, $V = .379$). Dabei ist zu bedenken, dass bei der in dieser Arbeit fokussierten Aussage elementare Kompetenzen im Umgang mit Algebra ausreichen (Abschn. 6.2.2). Zudem ist ein Zusammenhang zur Nutzung der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* zu erkennen. Von den Beweisversuchen mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil basieren 76% auf der primären Beweisidee 2. Ausgehend von dieser Beweisidee erscheint es mit Blick auf Lernende aller Klassenstufen 5 bis 8 gut möglich, ein weitgehend korrektes Argumentgefüge (KV4-6) zu formulieren (Abschn. 6.2.2). Schließlich ist es denkbar, dass viele der Teilnehmenden, die einen Beweisversuch mit einer Darstellungsform mit symbolischem Anteil formuliert haben, mathematisch gefördert wurden (Abschn. 10.4.2).

Viele der Beweisversuche mit begründendem Ansatz und einer **Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil** charakterisiert die Existenz eines weiterentwickelbaren Argumentgefüges bei zwischen induktiver und deduktiver Schlussweise und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*.

Von den Beweisversuchen mit begründendem Ansatz und einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil enthalten über 80% ein weiterentwickelbares Argumentgefüge (KV3-6). Dieser Anteil ist nach dem N - 1 Chi²-Test statistisch hoch signifikant größer mit $V = 0{,}244$ als bei den übrigen Beweisversuchen mit begründendem Ansatz. Es lässt sich vermuten, dass die Darstellung einer sinnbezogenen Einheit mit ikonischen Mitteln eine gewisse sachinhaltliche Auseinandersetzung mit der zu zeigenden Aussage impliziert, so dass in den meisten Fällen ein weiterentwickelbares Argumentgefüge gelungen ist.

Mit der Existenz eines weiterentwickelbaren Argumentgefüges einhergehend ist in über 85% der Beweisversuche mit begründendem Ansatz und einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil eine ausbaubare Beweisidee die primäre Beweisidee ($\mathrm{BIp1\text{-}8}$). Bei den Beweisversuchen mit begründendem Ansatz und einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil kommen die primären Beweisideen $1, 2, 4, 5, 6, 7, 9, 11$ vor, wobei die primären Beweisideen $1, 6$ klar dominieren. Mit einem Anteil von knapp 60% der Beweisversuche mit begründendem Ansatz und einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil ist die primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* besonders häufig. Der Anteil der Beweisversuche mit der primären Beweisidee 1 ist statistisch hoch signifikant größer als unter allen übrigen Beweisversuchen mit begründendem Ansatz ($\mathrm{p} < .001^{\mathrm{N\text{-}1}}, \mathrm{V} = .267$) sowie als unter den übrigen Beweisversuchen mit begründendem Ansatz und einer ausbaubaren Beweisidee als primärer Beweisidee ($\mathrm{p} < .001^{\mathrm{N\text{-}1}}, \mathrm{V} = .226$).

In rund 95% der Beweisversuche mit begründendem Ansatz und einer der Darstellungsformen ohne symbolischen, aber mit ikonisch-numerischem Anteil liegt die Schlussweise zwischen induktiv und deduktiv ($\mathrm{Sw1\text{-}2}$). Es lässt sich vermuten, dass oftmals mit ikonischen Darstellungsmitteln Überlegungen an Beispielen ausgeführte und nicht zusätzlich allgemeingültig aufgeschrieben wurden.

Von allen Beweisversuchen mit begründendem Ansatz und einer der Darstellungsformen ohne symbolischen, aber mit ikonisch-numerischem Anteil enthalten mehr als die Hälfte ein weiterentwickelbares Argumentgefüge bei zwischen induktiver und deduktiver Schlussweise und basieren auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*.

Primäre Beweisideen und Darstellungsformen

In den vorhergehenden Ausführungen ist u. a. deutlich geworden, dass ein großer Anteil der Beweisversuche mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil auf der primären Beweisidee 2 beruhen und ein großer Anteil derjenigen mit einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil auf der primären Beweisidee 1. Andersherum, d. h. von den primären Beweisideen ausgehend, betrachtet, zeigt sich folgendes Bild:

Bei der Mehrheit der Beweisversuche mit begründendem Ansatz und der **primären Beweisidee $\mathbf{1}$** *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* ist die Darstellungsform numerisch-verbal. Rund 21% der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 haben eine **Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil**. Dieser Anteil der Darstellungsformen ohne symbolischen, aber mit ikonisch-numerischem Anteil ist im Falle der primären Beweisidee 1 statistisch hoch signifikant größer als bei den übrigen Beweisversuchen mit begründendem Ansatz ($\mathrm{p} < .001^{\mathrm{N\text{-}1}}, \mathrm{V} = .267$) sowie, im Speziellen, als bei den übrigen Beweisversuchen mit begründendem Ansatz und einer ausbaubaren primären Beweisidee ($\mathrm{p} < .001^{\mathrm{N\text{-}1}}, \mathrm{V} = .226$). Ergänzend fällt auf, dass die Schluss-

weise in allen analysierten Beweisversuchen mit begründendem Ansatz, der primären Beweisidee 1 und einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil zwischen induktiv und deduktiv liegt. Der Anteil der Schlussweise *zwischen induktiv und deduktiv* ist bei den Beweisversuchen mit begründendem Ansatz, der primären Beweisidee 1 und einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil statistisch hoch signifikant größer als bei denen mit der primären Beweisidee 1 und einer anderen Darstellungsform ($p = .006^{ex}$, $V = .265$). Hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges ist kein statistisch signifikanter Unterschied zwischen Beweisversuchen mit begründendem Ansatz, der primären Beweisidee 1 und einer Darstellungsformen ohne symbolischen, aber mit ikonisch-numerischem Anteil sowie denen mit der primären Beweisidee 1 und einer anderen Darstellungsform zu erkennen.

Bei den Beweisversuchen mit begründendem Ansatz und der **primären Beweisidee 2** *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* beträgt der Anteil der Darstellungsformen mit symbolischem Anteil rund 70%. Mit Blick auf jede der Klassenstufen 6, 7, 8 hat bei mehr als der Hälfte der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 2 die Darstellungsform symbolischen Anteil. Die Schlussweise ist bei drei Viertel der Beweisversuche mit begründendem Ansatz, der primären Beweisidee 2 und einer Darstellungsform mit symbolischem Anteil deduktiv, bei einem Viertel liegt sie zwischen induktiv und deduktiv. Dabei ist jedoch die geringe Anzahl an Beweisversuchen mit der primären Beweisidee 2 zu bedenken, der genannte Unterschied zwischen denen mit bzw. ohne Darstellungsform mit symbolischem Anteil ist nicht statistisch signifikant ($p = .066^{ex}$, $V = .438$).

Bei den Beweisversuchen mit begründendem Ansatz und einer der übrigen ausbaubaren **primären Beweisideen 3 bis 8** beträgt der Anteil der Darstellungsform *numerisch-verbal* rund 60%. Die Darstellungsformen mit symbolischem Anteil machen 3% aus, die Darstellungsformen ohne symbolischen, aber mit ikonisch-numerischem Anteil rund 6% und die Darstellungsformen *verbal, ikonisch-verbal* rund 31%. Dabei sind zwischen den einzelnen primären Beweisideen 3, 4, 5, 6, 7, 8 keine statistisch signifikanten Unterschiede zu erkennen. Es ist lediglich auffällig, dass die einzelnen Beweisversuche mit der primären Beweisidee 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* und der Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* alle eine Darstellungsform mit symbolischem Anteil haben bei deduktiver Schlussweise. Die Beweisversuche mit der primären Beweisidee 8 und der Beweisidee 2 erscheinen denen mit der primären Beweisidee 2 ähnlich.

Mit Blick auf die Beweisversuche mit begründendem Ansatz und einer der nicht ausbaubaren Beweisideen 9 bis 13, 55 als primärer Beweisidee ist die Darstellungsform in ca. der Hälfte der Fälle verbal oder ikonisch-verbal, in ca. der Hälfte der Fälle numerisch-verbal. Die Darstellungsformen mit symbolischem oder ikonisch-numerischem Anteil kommen kaum vor. Bei den Beweisversuchen mit begründendem Ansatz und einer der **primären Beweisideen 9 bis 13, 55** ist der Anteil der Darstellungsformen *verbal, ikonisch-verbal* statistisch hoch signifikant größer ($p < .001^{N-1}$, $V = .224$), der Anteil der Darstellungsformen mit symbolischem oder ikonisch-numerischem Anteil statistisch hoch signifikant geringer ($p < .001^{N-1}$, $V = .267$) als bei den Beweisversuchen mit begründendem Ansatz und einer der primären Beweisideen 1 bis 8.

Fazit: Darstellungsform und Schlussweise, Beispiele, Korrektheit, Vollständigkeit, primäre Beweisidee

Die Ergebnisse der Analyse der Darstellungsformen zusammen mit weiteren Eigenschaften der Beweisversuche mit begründendem Ansatz weisen darauf hin, dass insbesondere zwischen Schlussweisen und Darstellungsformen Zusammenhänge bestehen sowie zwischen einzelnen primären Beweisideen und Darstellungsformen.

Die Darstellungsform ist bei den meisten Beweisversuchen mit der **Schlussweise** Sw4 *deduktiv ohne zusätzliches Beispiel* verbal, bei den meisten anderen Beweisversuchen mit begründendem Ansatz numerisch-verbal, was im Zusammenhang mit den Definitionen der Schlussweisen und der Angabe von Beispielen zu sehen ist.

Eine Ausnahme bilden dabei die Beweisversuche mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* und einem weitgehend korrekten und vollständigen Argumentgefüge. Bei rund der Hälfte dieser Beweisversuche hat die Darstellungsform symbolischen Anteil. Die Beweisversuche mit begründendem Ansatz und einer **Darstellungsform mit symbolischem Anteil** enthalten mehrheitlich ein weitgehend korrektes Argumentgefüge bei deduktiver Schlussweise. Etwa drei Viertel von ihnen basieren auf der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*. Bei den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 2 beträgt der Anteil der Darstellungsformen mit symbolischem Anteil rund 70%. Außerdem haben alle analysierten Beweisversuche mit der primären Beweisidee 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* und der Beweisidee 2 eine Darstellungsform mit symbolischem Anteil.

Von den Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise und einem weiterentwickelbaren Argumentgefüge hat ca. ein Sechstel eine **Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil**. Von den Beweisversuchen mit begründendem Ansatz und einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil enthalten mehr als drei Viertel ein weiterentwickelbares Argumentgefüge, fast immer bei zwischen induktiver und deduktiver Schlussweise. Die primäre Beweisidee ist in mehr als der Hälfte der Beweisversuche mit begründendem Ansatz und einer Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Von den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 hat rund ein Fünftel eine Darstellungsform ohne symbolischen, aber mit ikonisch-numerischem Anteil.

Von den Beweisversuchen mit begründendem Ansatz und einer der **Darstellungsformen *verbal, ikonisch-verbal*** enthalten ca. drei Viertel kein weiterentwickelbares Argumentgefüge bei deduktiver Schlussweise ohne zusätzliches Beispiel. Bei den meisten dieser Beweisversuche ist eine der nicht ausbaubaren Beweisideen die primäre Beweisidee. Von den Beweisversuchen mit einer der nicht ausbaubaren primären Beweisideen haben ca. die Hälfte eine der Darstellungsformen *verbal, ikonisch-verbal*.

8.2.6.6 Klassifikation: Einführung und Begründung

Auf Grundlage der vorhergehenden Ausführungen zu Eigenschaften von Beweisversuchen mit begründendem Ansatz wird im Folgenden eine Klassifikation dieser Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis eingeführt und begründet. Dabei wird sukzessive erläutert, welche Eigenschaften von Beweisversuchen mit begründendem Ansatz als Klassifikationsmerkmale berücksichtigt werden und wie die in Abbildung 129 zu sehende Klassifikation aufgebaut ist.

Klassifikation der Beweisversuche mit begründendem Ansatz
mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

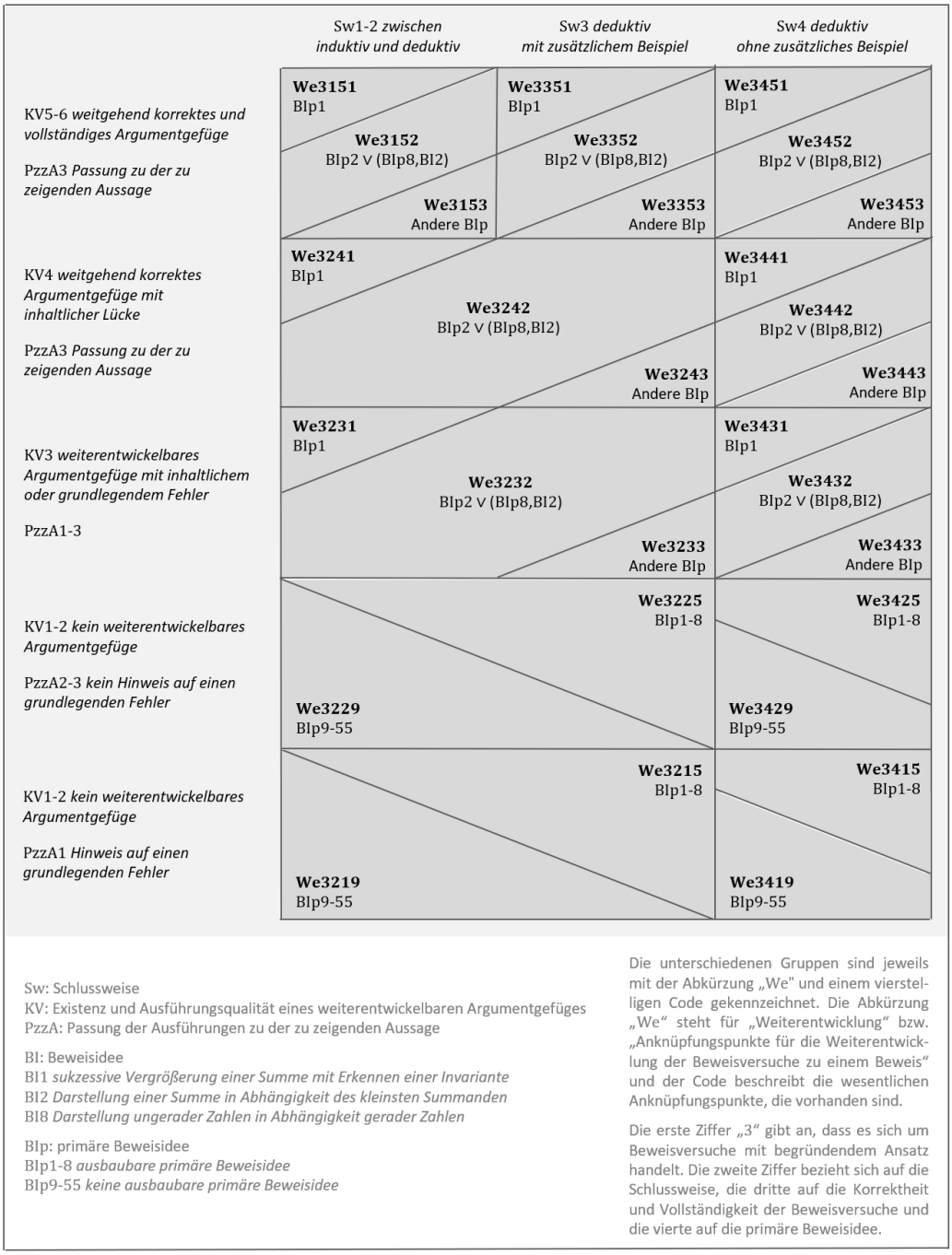


Abbildung 129: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

Korrektheit, Vollständigkeit und Schlussweise

Mit Blick auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche mit begründendem Ansatz zu einem Beweis der zu zeigenden Aussage sind die **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** und die **Schlussweise** von grundlegender Bedeutung, weshalb diese beiden Eigenschaften bei der Klassifikation der Beweisversuche auf jeden Fall zu berücksichtigen sind.

Was die **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** betrifft, werden die Kategorien KV1 *Beweisideen ohne relevanten Beitrag für ein weiterentwickelbares Argumentgefüge* und KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge* zusammengefasst. Beiden Kategorien KV1, KV2 charakterisiert die Nichtexistenz eines weiterentwickelbaren Argumentgefüges. Zudem unterscheiden sich die Beweisversuche in diesen beiden Kategorien vergleichsweise wenig in Bezug auf weitere Eigenschaften, wie z. B. die Darstellungsform. Auffällig ist vielmehr, dass von den Beweisversuchen mit begründendem Ansatz, aber ohne weiterentwickelbares Argumentgefüge (KV1-2) diejenigen ohne einen Hinweis auf einen grundlegenden Fehler (PzzA2-3) im Vergleich zu denen mit einen Hinweis auf einen grundlegenden Fehler (PzzA1) häufiger ein Beispiel zu der zu zeigenden Aussage und häufiger eine ausbaubare Beweisidee als primäre Beweisidee enthalten. Außerdem deuten die Ausführungen in Beweisversuchen im Falle eines Hinweises auf einen grundlegenden Fehler (PzzA1) darauf hin, dass für deren Weiterentwicklung zu einem Beweis zunächst grundlegend Begrifflichkeiten der zu zeigenden Aussage geklärt werden sollten, während dies bei den anderen Beweisversuchen ohne weiterentwickelbares Argumentgefüge (PzzA2-3) nicht unbedingt der Fall ist. Aus diesen Gründen erscheint es sinnvoll, bei der Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis anstatt zwischen den Kategorien KV1 und KV2 bei den Beweisversuchen ohne weiterentwickelbares Argumentgefüge (KV1-2) zwischen denen mit bzw. ohne einen Hinweis auf einen grundlegenden Fehler zu differenzieren.

Bezüglich der **Schlussweise** in Beweisversuchen mit begründendem Ansatz haben die vorhergehenden Ausführungen verdeutlicht, dass sich die Beweisversuche mit der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* hinsichtlich mehrerlei Eigenschaften, wie der Existenz eines Beispiels und der Darstellungsform, deutlich von den Beweisversuchen mit begründendem Ansatz und einer der Schlussweisen Sw1-3 unterscheiden. Beweisversuche mit den Schlussweisen Sw1-2 *zwischen induktiv und deduktiv* und Sw3 *deduktiv mit zusätzlichem Beispiel* haben hingegen u. a. gemeinsam, dass sie alle mindestens ein Beispiel enthalten und mehrheitlich die Darstellungsform numerisch-verbal aufweisen. Daher wird mit Blick auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche mit begründendem Ansatz zwischen Beweisversuchen mit deduktiver Schlussweise ohne zusätzliches Beispiel (Sw4) und denen mit einer der anderen Schlussweisen Sw1-3 differenziert.

Bei den Beweisversuchen mit einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) im Speziellen erscheint eine zusätzliche Unterscheidung zwischen den Schlussweisen Sw1-2 und Sw3 als sinnvoll, weil bei den Beweisversuchen mit der Schlussweise Sw3 *deduktiv mit zusätzlichem Beispiel*, jedoch nicht bei denen mit der zwischen induktiver und deduktiver

Schlussweise (Sw1-2), zusätzlich zur Angabe eines weitgehend korrekten und vollständigen Argumentgefüges auch alle Überlegungen allgemeingültig formuliert wurden. Dies bedeutet, dass bei Beweisversuchen mit der Schlussweise Sw3 auch bezüglich der Schlussweise nicht unbedingt Weiterentwicklungsbedarf besteht. Zugleich werden die Beweisversuche mit der Ausprägung KV5-6, Sw3 nicht mit denjenigen mit deduktiver Schlussweise ohne zusätzliches Beispiel (KV5-6, Sw4) zusammengefasst, weil die Angabe eines Beispiels in den Beweisversuchen mit KV5-6, Sw3 aus Sicht der Lernenden vielfältige Gründe haben kann, wie z. B. die Stützung einer allgemeingültig formulierten Überlegung oder die Vorstellung, dass Beispiele notwendige Bestandteile eines Beweises sind, was mit Blick auf Ansatzpunkte für eine Förderung relevant ist.

Wenn man die Beweisversuche mit begründendem Ansatz in der beschriebenen Form je nach Korrektheit, Vollständigkeit und Schlussweise gruppiert, ergeben sich somit **elf Gruppen** an Beweisversuchen, die sich hinsichtlich dieser Klassifikationsmerkmale unterscheiden. Darauf aufbauend stellt sich die Frage, inwieweit diese elf Gruppen weiter unterteilt werden sollten, um bezüglich ihrer Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis noch homogenere Gruppen zu bilden, so dass die Zugehörigkeit eines Beweisversuchs zu einer bestimmten Gruppe noch aussagekräftiger ist.

Primäre Beweisidee

In der vorhergehenden Ausführungen zu Zusammenhängen zwischen verschiedenen Eigenschaften von Beweisversuchen mit begründendem Ansatz (insb. Abschn. 8.2.6.3 – 8.2.6.5) ist deutlich geworden, dass sich die Beweisversuche mit begründendem Ansatz und der **primären Beweisidee 1** *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* in mehrerlei Hinsicht von anderen Beweisversuchen mit begründendem Ansatz unterscheiden. Beispielsweise sind in Beweisversuchen mit der primären Beweisidee 1 besonders häufig Reihenfolgebeispiele angegeben und die Darstellungsform ist verhältnismäßig oft ikonisch-verbal. Zudem liegt Beweisversuchen mit einem korrekten und vollständigen Argumentgefüge und der primären Beweisidee 1 aus Expertensicht das Verfahren der vollständigen Induktion zugrunde, während es sich im Falle fast jeder der anderen primären Beweisideen i. d. R. um direkte Beweise handelt. Nebenbei ist die primäre Beweisidee 1 in allen bereits gebildeten Gruppen mit Beweisversuchen mit einem weiterentwickelbaren Argumentgefüge (KV3-6) eine der drei häufigsten primären Beweisideen. Die Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 erscheinen also mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung als besonders und zugleich von ihrer relativen Häufigkeit her als relevante Teilgruppe. Daher wird fokussiert auf die Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge die Verwendung der primären Beweisidee 1 als Klassifikationsmerkmal aufgenommen.

Hinsichtlich mehrerlei Eigenschaften besonders sind ebenfalls die Beweisversuche mit begründendem Ansatz und der **primären Beweisidee 2** *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*. Insbesondere machen bei den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 2 die Darstellungsformen mit symbolischem Anteil rund 70% aus. Zudem gehört die primäre Beweisidee 2 in jeder der Gruppen mit einem weitgehend

korrekten Argumentgefüge und deduktiver Schlussweise zu den drei häufigsten primären Beweisideen. In den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 2 kommt teils auch die Beweisidee 8 vor. Zugleich enthält ein Teil der Beweisversuche mit der **primären Beweisidee 8** *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* auch die Beweisidee 2. Alle analysierten Beweisversuche mit der primären Beweisidee 8 und der Beweisidee 2 haben ebenfalls eine Darstellungsform mit symbolischem Anteil und enthalten ein weitgehend korrektes Argumentgefüge bei deduktiver Schlussweise. In Anbetracht dieser Besonderheiten wird die Verwendung der primären Beweisidee 2 oder der primären Beweisidee 8 zusammen mit der Beweisidee 2 als ergänzendes Klassifikationsmerkmal aufgenommen. Analog zur primären Beweisidee 1 wird die Verwendung der primären Beweisidee 2 oder der primären Beweisidee 8 zusammen mit der Beweisidee 2 bei allen Beweisversuchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge (KV3-6) als Klassifikationsmerkmal berücksichtigt.

Bei den Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge (KV1-2) erscheint die Berücksichtigung der Verwendung der primären Beweisideen 1, 2, 8 wenig sinnvoll, weil sie in fast keinem dieser Beweisversuche vorkommen. Mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung erscheint es aber durchaus als lohnend, bei den Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge (KV1-2) zwischen denen mit und denen ohne ausbaubare primäre Beweisidee zu differenzieren. Wenn eine **ausbaubare primäre Beweisidee** erkennbar ist, kann diese für die Entwicklung eines Beweises genutzt, d. h. im Rahmen der Weiterentwicklung des Beweisversuchs beibehalten werden. Ist hingegen eine der Beweisideen 9 bis 13, 55 die primäre Beweisidee, dann kommt in den Beweisversuchen i. d. R. keine ausbaubare Beweisidee vor, sondern muss für einen Beweis der zu zeigenden Aussage entwickelt werden. Daher wird bei den Beweisversuchen mit begründendem Ansatz, aber allenfalls einem Ansatz für ein weiterentwickelbares Argumentgefüge die Existenz einer ausbaubaren primären Beweisidee als ergänzendes Klassifikationsmerkmal aufgenommen.

Existenz, Art und Ort von Beispielen sowie Darstellungsformen

Anstelle primärer Beweisideen als Klassifikationsmerkmale etwa die Existenz von Reihenfolgebeispielen, den Ort von Beispielen oder bestimmte Darstellungsformen zu berücksichtigen, erscheint weniger sinnvoll. Auch eine Berücksichtigung solcher Eigenschaften als zusätzliche Klassifikationsmerkmale, zusätzlich zu primären Beweisideen, erscheint nicht als lohnend.

Beispielsweise deutet die **Existenz von Reihenfolgebeispielen** zwar darauf hin, dass die Lernenden systematisch vorgegangen sind, geht aber nicht automatisch mit weiteren Eigenschaften der Beweisversuche einher. Die Beweisversuche mit begründendem Ansatz und Reihenfolgebeispielen sind im Vergleich zu denen mit der primären Beweisidee 1 heterogen, weil Reihenfolgebeispiele in Beweisversuchen mit fast jeder der unterschiedenen primären Beweisideen vorkommen. Mit Blick auf das Ziel der Bildung möglichst homogener Gruppen ist die Existenz von Reihenfolgebeispielen daher kein optimales Klassifikationsmerkmal.

Die Beweisversuche mit begründendem Ansatz und einem **großen Beispiel** machen mit unter 10% ausschließlich einen kleinen Anteil der Beweisversuche mit begründendem Ansatz aus. Viele von ihnen sind zudem hinsichtlich der Korrektheit und Vollständigkeit des Argumentgefüges und der Schlussweise ähnlich ausgeprägt, sodass die Frage nach der Existenz eines großen Beispiels bereits z. T. durch die vorgenommene Gruppierung der Beweisversuche mit begründendem Ansatz in Abhängigkeit ihrer Korrektheit, Vollständigkeit und Schlussweise in der Klassifikation berücksichtigt ist. Außerdem enthält die große Mehrheit der Beweisversuche mit begründendem Ansatz und einem großen Beispiel kein weiterentwickelbares Argumentgefüge und auch aus theoretischer Sicht sind große Beispiele bei der in dieser Arbeit fokussierten Aussage nicht per se ein besonders guter Anknüpfungspunkt für die Entwicklung eines Beweises. Aus diesen Gründen wird die Existenz eines großen Beispiels nicht als Klassifikationsmerkmal berücksichtigt.

Der **Ort von Beispielen** in Beweisversuchen mit begründendem Ansatz kann als ein Anhaltspunkt dienen, inwieweit vor, während oder nach der Formulierung des (übrigen) Argumentgefüges Beispiele angegeben wurden. Wo genau sich ein Beispiel im Verhältnis zum (übrigen) Argumentgefüge befindet, erscheint jedoch z. B. für die Entwicklung einer ausbaubaren Beweisidee nicht als entscheidend. Diesbezüglich ist es relevanter, inwieweit überhaupt ein Beispiel zu der zu zeigenden Aussage vorkommt. Die Frage nach der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage ist in der Klassifikation durch die Unterscheidung der Schlussweisen und der Beweisversuche mit bzw. ohne Hinweis auf einen grundlegenden Fehler bereits indirekt berücksichtigt.

Welche **Darstellungsform** ein Beweisversuch mit begründendem Ansatz aufweist, hängt entsprechend den vorhergehenden Ausführungen mit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise zusammen. Zudem besteht bei Beweisversuchen mit begründendem Ansatz ein Zusammenhang zwischen den Darstellungsformen mit symbolischem Anteil und der primären Beweisidee 2 sowie den Darstellungsformen ohne symbolischen, aber mit ikonisch-numerischem Anteil und der primären Beweisidee 1. Aus der Zugehörigkeit von Beweisversuchen mit begründendem Ansatz zu den bereits in Abhängigkeit der Korrektheit und Vollständigkeit, der Schlussweise und primärer Beweisideen gebildeten Gruppen geht dadurch z. B. hervor, ob die Beweisversuche i. d. R. die Darstellungsform *verbal* aufweisen. Eine Berücksichtigung der Darstellungsform als weiteres Klassifikationsmerkmal würde kaum zusätzliche Informationen liefern.

Aufbau der Klassifikation

Insgesamt werden folglich in der Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis der zu zeigenden Aussage 29 Gruppen unterschieden. Diese Gruppen sind in Abbildung 129 auf Seite 322 schematisch dargestellt und jeweils mit der Abkürzung „We" und einem vierstelligen Code gekennzeichnet. Die Abkürzung „We“ steht für „Weiterentwicklung“ bzw. „Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis“ und der Code beschreibt die wesentlichen Anknüpfungspunkte, die vorhanden sind.

Die erste, linke Ziffer des vierstelligen Codes ist bei jeder der Gruppen die „3“. Sie hebt hervor, dass die Beweisversuche zur Oberkategorie Ans3 *begründender Ansatz* gehören. Dies impliziert insbesondere, dass die Lernenden versucht haben, eine allgemeingültige oder verallgemeinerbare Begründung der zu zeigenden Aussage zu formulieren. In den Beweisversuchen ist mindestens eine Beweisidee ausgeführt.

Die zweite Ziffer von links gibt ergänzend die Schlussweise an. Ist die zweite Ziffer die „1“, dann liegt die Schlussweise bei allen Beweisversuchen in der Gruppe zwischen deduktiv und induktiv. Ist die zweite Ziffer die „2“, dann liegt die Schlussweise zwischen induktiv und deduktiv oder ist deduktiv mit zusätzlichem Beispiel. Ist die zweite Ziffer die „3“, dann ist die Schlussweise deduktiv mit zusätzlichem Beispiel. Ist die zweite Ziffer die „4“, dann ist die Schlussweise deduktiv ohne zusätzliches Beispiel.

Von unten nach oben unterscheiden sich die Gruppen hinsichtlich der Korrektheit und Vollständigkeit der Beweisversuche. Wie die Beweisversuche diesbezüglich ausgeprägt sind, ist im Code mit der dritten Ziffer von links berücksichtigt. Ist diese Ziffer etwa die „4“, dann enthalten alle Beweisversuche in der Gruppe ein weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke (KV4), entsprechend den Angaben im Klassifikationsschema (Abb. 129).

Die vierte bzw. rechte Ziffer der Gruppenbezeichnungen bezieht sich auf die primäre Beweisidee. Ist die vierte Ziffer die „1“, basieren alle Beweisversuche in der Gruppe auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Ist die vierte Ziffer die „2“, enthalten alle Beweisversuche in der Gruppe die primäre Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* oder die primäre Beweisidee „8“ *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* zusammen mit der Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*. Ist die vierte Ziffer die „3“, basieren alle Beweisversuche auf anderen als den genannten Beweisideen, in der Regel auf einer der Beweisideen 3 bis 7 (BI3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*, BI4 *Zahldarstellung im Dezimalsystem*, BI5 *Reste bei Division einzelner Summanden durch vier oder acht*, BI6 *Teilbarkeit der Summen zweier Zahlenpaare*, BI7 *Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen*). Gruppen mit der „1“, „2“ oder „3“ als rechte Ziffer sind entsprechend den vorhergehenden Überlegungen ausschließlich im Fall der Existenz eines weiterentwickelbaren Argumentgefüges (KV3-6) ausgewiesen und im Klassifikationsschema durch diagonale Linien von unten nach oben voneinander abgegrenzt.

Bei den Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge (KV1-2) wird mit der vierten bzw. rechten Ziffer zwischen solchen mit und ohne ausbaubare primäre Beweisidee differenziert. Die Beweisversuche in den Gruppen mit der rechten Ziffer „5“ beruhen auf einer ausbaubaren primäre Beweisidee (BIp1-8), die in den Gruppen mit der rechten Ziffer „9“ nicht (BIp9-55). Die Gruppen mit der rechten Ziffer „5“ bzw. „9“ sind mit diagonalen Linien so abgetrennt, dass sich die mit der Ziffer „5“ jeweils rechts oben befinden. Die Beweisversuche in diesen Gruppen sind hinsichtlich Anknüpfungspunkten für deren Weiterentwicklung zu einem Beweis etwas tragfähiger ausgeprägt, weil die primäre Beweisidee bei der Weiterentwicklung prinzipiell beibehalten werden kann.

Insgesamt betrachtet umfasst beispielsweise die Gruppe We3429 alle Beweisversuche mit begründendem Ansatz, deduktiver Schlussweise ohne zusätzliches Beispiel, ohne Hinweis auf einen grundlegenden Fehler, aber auch ohne weiterentwickelbares Argumentgefüge und ohne ausbaubare primäre Beweisidee. Im Vergleich dazu umfasst die Gruppe We3241 alle Beweisversuche mit begründendem Ansatz, einer der Schlussweisen Sw1-3 und einem weitgehend korrekten Argumentgefüge mit inhaltlicher Lücke basierend auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Beide Gruppen We3429 und We3241 machen in einzelnen Klassenstufen große Anteile der Beweisversuche mit begründendem Ansatz aus, wie im Folgenden beschrieben wird.

8.2.6.7 Klassifikation: Häufigkeitsverteilungen und exemplarische Bearbeitungen

Die Abbildungen 135 bis 138 auf den Seiten 336, 337 geben Einblicke in die relativen Häufigkeiten der mit Blick auf Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche mit begründendem Ansatz gebildeten Gruppen an Beweisversuchen differenziert nach den Klassenstufen 5 bis 8. Dabei sind die relativen Häufigkeiten der Gruppen anhand der Grautöne, mit denen die einzelnen Gruppen hinterlegt sind, zu erkennen. Je dunkler der Grauton ist, desto größer ist der Anteil der Beweisversuche in der fokussierten Gruppe an allen Beweisversuchen mit begründendem Ansatz. Die auf diese Weise dargestellten relativen Häufigkeitsangaben basieren primär auf den Ergebnissen der Analyse der Beweisversuche aus der Zufallsstichprobe. Um aber auch das Spektrum an vorkommenden Gruppen angemessen abzubilden, wurden bei den Null-Prozent-Angaben ergänzend die Beweisversuche aus der bewusst gezogenen Stichprobe berücksichtigt. Wenn eine Gruppe, wie z. B. die Gruppe We3353 in der Klassenstufe 7, nicht in der Zufallsstichprobe, aber in der bewusst gezogenen Gruppe vorkommt, wird der relative Anteil h dieser Gruppe in die Kategorie $0\% < h < 5\%$ eingeordnet.

Überblicksartig sieht man anhand der Abbildungen 135 bis 138 zum Beispiel, dass in der Klassenstufe 5 rund die Hälfte, in der Klassenstufe 8 ca. drei Viertel der unterschiedenen Gruppen vorkommen. Außerdem ist zu erkennen, welche Gruppen an Beweisversuchen in den einzelnen Klassenstufen 5 bis 8 besonders große Anteile ausmachen. Darüberhinausgehend deutet sich an, dass sich die Beweisversuche von der Tendenz her je höher die Klassenstufe, desto weiter oben rechts im Klassifikationsschema befinden, wie beim Vergleich der relativen Häufigkeitsverteilungen der Gruppen in den einzelnen Klassenstufen näher ausgeführt wird.

Im Folgenden werden zuerst die Gruppen, die in den einzelnen Klassenstufen 5 bis 8 besonders große Anteile ausmachen und somit auch hinsichtlich Anknüpfungspunkten für eine Weiterentwicklung von Beweisversuchen besonders charakteristisch sind, in den Fokus gestellt. Diese Gruppen, die die größten Anteile ausmachen, werden allgemein und anhand exemplarischer Bearbeitungen beschrieben. Danach werden auf Grundlage des Vergleichs der relativen Häufigkeitsverteilungen der Gruppen in den einzelnen Klassenstufen 5 bis 8 mehrere Auffälligkeiten herausgestellt, die mit Blick auf Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche mit begründendem Ansatz relevant sind.

Klassenstufe 5

In der Klassenstufe 5 machen die Beweisversuche in den Gruppen We3229 und We3429 besonders große Anteile der Beweisversuche mit begründendem Ansatz aus (Abb. 135). Rund 60% aller Beweisversuche mit begründendem Ansatz aus der Klassenstufe 5 gehören zu genau diesen zwei Gruppen. Dabei ist die Gruppe **We3429** diejenige, in die mit rund 35% bzw. gut einem Drittel der Zuordnungen die meisten der Beweisversuche mit begründendem Ansatz aus der Klassenstufe 5 gehören. Diese Beweisversuche mit begründendem Ansatz, die zu dieser Gruppe gehören, enthalten kein weiterentwickelbares Argumentgefüge. Die primäre Beweisidee ist eine der nicht ausbaubaren Beweisideen 9 bis 13, 55. Oftmals handelt es sich um die Beweisidee 9 *Parität von Summen ungerader oder gerader Zahlen*, so wie es auch in der Bearbeitung 8.2.6a in Abbildung 130 der Fall ist. Die Aussage „wen man $4 \cdot$ eine ungerade Zahl nimmt ist das ergebniss immer gerade“ aus Bearbeitung 8.2.6a lässt sich nicht als relevanter Bestandteil eines Beweises der zu zeigenden Aussage nutzen. Eventuell liegt dieser Aussage aber das Wissenselement zugrunde, dass die Summe zweier ungerader Zahlen gerade ist, was als Bestandteil eines Beweises der zu zeigenden Aussage verwendet werden kann. Auch ist es möglich, dass sich der Schüler überlegt hat, dass die Summe der kleinsten und die der größten beiden Summanden durch zwei teilbar ist, d. h. die Ausführungen implizit auf der ausbaubaren Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* beruhen. Auf Grundlage der Beweisidee 6 ließe sich ein Beweis der zu zeigenden Aussage entwickeln. Dies veranschaulicht, dass in den Beweisversuchen in der Gruppe We3429 zwar allenfalls ein Ansatz für ein weiterentwickelbares Argumentgefüges basierend auf einer nicht ausbaubaren Beweisideen vorkommt, den Ausführungen aber vielleicht Überlegungen zugrunde liegen, die als Anknüpfungspunkte für die Weiterentwicklung des Beweisversuchs genutzt werden könnten.

In den Beweisversuchen in der Gruppe We3429 ist zwar kein Hinweis auf einen grundlegenden Fehler zu erkennen, der genaue Bezug der Ausführungen auf die zu zeigende Aussage ist aber oftmals nicht ganz klar. Beispielsweise lässt sich bei der Bearbeitung 8.2.6a nicht ganz sicher sagen, ob die Summe oder das Produkt von ungeraden Zahlen gemeint ist sowie ob von der Teilbarkeit durch zwei auf die Teilbarkeit durch vier geschlossen oder die Teilbarkeit durch zwei als zu zeigende Behauptung angesehen wurde. Obwohl die Beweisversuche in der Gruppe We3429 keinen Hinweis auf einen grundlegenden Fehler enthalten, erscheint für deren Weiterentwicklung die Klärung der zu zeigenden Aussage von Bedeutung.

Bei allen Beweisversuchen mit begründendem Ansatz in der Gruppe We3429 ist die Schlussweise deduktiv ohne zusätzliches Beispiel. So wie in der exemplarischen Bearbeitung 8.2.6a kommt in fast keinem dieser Beweisversuche ein Beispiel vor. Während die Ausführungen ein

Bearbeitung 8.2.6a: Klassenstufe 5, männlich (Code: 0510652)

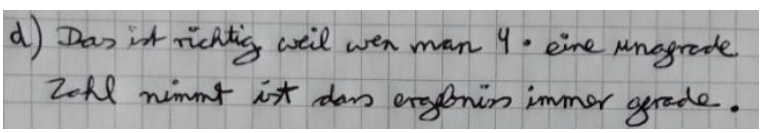


Abbildung 130: Bearbeitung 8.2.6a (We3429 *begründender Ansatz, Schlussweise deduktiv ohne zusätzliches Beispiel, kein weiterentwickelbares Argumentgefüge, kein Hinweis auf einen grundlegenden Fehler, keine ausbaubare primäre Beweisidee*)

Bewusstsein für die Notwendigkeit einer allgemeingültigen Begründung andeuten, wäre es für die Lernenden eventuell neu, verallgemeinerbare Überlegungen anhand von Beispielen zu erläutern und darzustellen.

Weiterhin ist, so wie in der Bearbeitung 8.2.6a, in rund 95% der Beweisversuche in der Gruppe We3429 die Darstellungsform verbal. Mit Blick auf eine Weiterentwicklung der Beweisversuche ist nicht davon auszugehen, dass die Lernenden mit der Verwendung symbolischer oder ikonischer Darstellungsmittel gut vertraut sind.

Neben den Beweisversuchen in der Gruppe We3429 machen die Beweisversuche in der Gruppe **We3229** in der Klassenstufe 5 ca. 25% aller Beweisversuche mit begründendem Ansatz aus. Auch diese Beweisversuche mit begründendem Ansatz enthalten kein weiterentwickelbares Argumentgefüge. Dabei beziehen sich die Ausführungen in den meisten dieser Beweisversuche sehr wahrscheinlich genau auf die zu zeigende Aussage. Das heißt, dass es in der Regel keinen Anhaltspunkt dafür gibt, dass bei der Weiterentwicklung des Beweisversuchs bei der Interpretation der zu zeigenden Aussage angesetzt werden muss.

Die primäre Beweisidee ist bei den Beweisversuchen in der Gruppe We3229 immer eine der nicht ausbaubaren Beweisideen, in rund der Hälfte der Fälle die Beweisidee 9 *Parität von Summen ungerader oder gerader Zahlen* und ansonsten meistens die Beweisidee 11 *vier Zahlen*. Die Schlussweise ist zwischen induktiv und deduktiv oder deduktiv mit zusätzlichem Beispiel. In jedem Beweisversuch in der Gruppe We3229 kommt dementsprechend ein Beispiel vor. Meistens sind mindestens zwei Beispiele angegeben und dabei mindestens ein Beispiel zu der zu zeigenden Aussage. Oftmals sind also, so wie in der Bearbeitung 8.2.6b in Abbildung 131, Beispiele ausgeführt, die als Anknüpfungspunkte für die Entwicklung einer ausbaubaren Beweisidee verwendet werden könnten. Zugleich gibt es in vielen der Beweisversuche in der Gruppe We3229 Hinweise darauf, dass die Beispiele noch zielführender ausgewählt und genutzt hätten werden können. Systematische Beispiele und im Speziellen Reihenfolgebeispiele kommen eher selten vor. In keinem der analysierten Beweisversuche in der Gruppe We3229 ist an einem Beispiel oberhalb des (übrigen) Argumentgefüges eine Beweisidee hervorgehoben.

Bearbeitung 8.2.6b: Klassenstufe 5, weiblich (Code: 0521637)

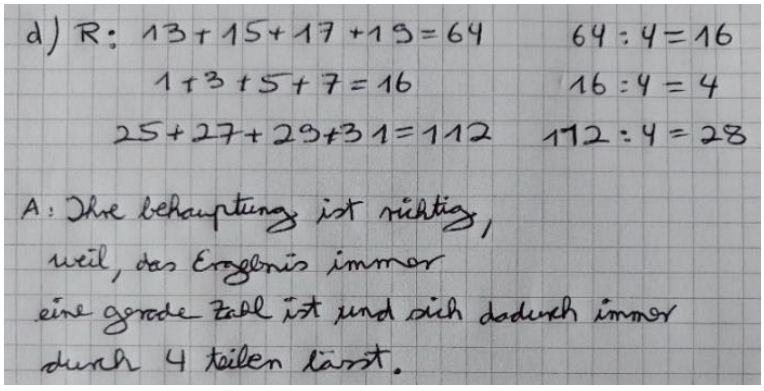


Abbildung 131: Bearbeitung 8.2.6b (We3229 *begründender Ansatz, Schlussweise induktiv mit Andeutung der Verallgemeinerbarkeit, deduktiv mit induktivem Anteil oder deduktiv mit zusätzlichem Beispiel, kein weiterentwickelbares Argumentgefüge, kein Hinweis auf einen grundlegenden Fehler, keine ausbaubare primäre Beweisidee*)

Mit der Existenz von Beispielen einhergehend ist die Darstellungsform in über 90% der Beweisversuche in der Gruppe We3229 numerisch-verbal. Mit Blick auf eine Weiterentwicklung der Beweisversuche lässt sich nicht davon ausgehen, dass die Lernenden mit der Verwendung symbolischer oder ikonischer Darstellungsmittel gut vertraut sind.

Klassenstufe 6

Die beschriebenen Gruppen **We3429** und **We3229** sind auch in der Klassenstufe 6 diejenigen, die die größten Anteile der Beweisversuche mit begründendem Ansatz ausmachen (Abb. 136). Der Anteil der Gruppe We3429 beträgt in der Klassenstufe 6 rund 18%, der Anteil der Gruppe We3229 rund 28%.

Neben diesen beiden Gruppen gehören in der Klassenstufe 6 rund 10% der Beweisversuche in die Gruppe **We3225**. Die Beweisversuche in der Gruppe We3225 kennzeichnet im Vergleich zu denen in der Gruppe We3229 insbesondere, dass die primäre Beweisidee eine ausbaubare Beweisidee ist, sodass die primäre Beweisidee als wesentlicher Anknüpfungspunkt bei der Weiterentwicklung des Beweisversuchs genutzt werden kann. Bei ca. 60% der Beweisversuche in der Gruppe We3225 handelt es sich um die Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare*, bei rund 23% um die Beweisidee 4 *Zahldarstellung im Dezimalsystem*. Beispielsweise ist in der Bearbeitung 8.2.6c in Abbildung 132 die Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* die primäre Beweisidee. Sie ist aber nicht ganz klar zu erkennen, weil ausschließlich die Teilbarkeit der Summe zweier aufeinanderfolgender Zahlen durch vier ausgeführt ist. Die Beweisidee 6 ließe sich ausgehend von den Ausführungen des Lernenden weiter herausarbeiten. Zudem wäre mithilfe einer der zentralen Beweisideen zu begründen, warum die Summe zweier aufeinanderfolgender ungerader Zahlen immer durch vier teilbar ist.

Bearbeitung 8.2.6c: Klassenstufe 6, männlich (Code: 0622117)

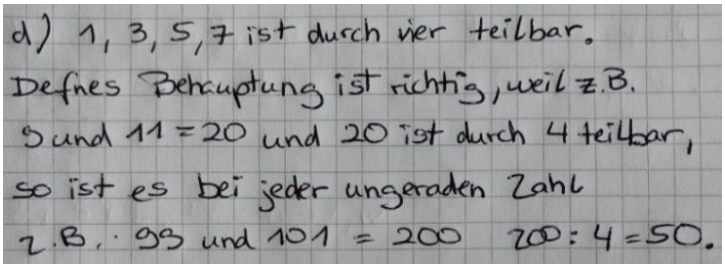


Abbildung 132: Bearbeitung 8.2.6c (We3225 *begründender Ansatz, Schlussweise induktiv mit Andeutung der Verallgemeinerbarkeit, deduktiv mit induktivem Anteil oder deduktiv mit zusätzlichem Beispiel, kein weiterentwickelbares Argumentgefüge, kein Hinweis auf einen grundlegenden Fehler, ausbaubare primäre Beweisidee*)

Beim Vergleich der Beweisversuche in der Gruppe **We3225** mit denen in der Gruppe **We3229** ist ergänzend zu sehen, dass in einem statistisch hoch signifikant größeren Anteil als in der Gruppe We3229 an Beispielen oberhalb des (übrigen) Argumentgefüges eine Beweisidee hervorgehoben ist ($p = .006^{ex}$, $V = .249$). Dabei ist der Gesamtanteil der Beweisversuche mit allen Beispielen oberhalb des (übrigen) Argumentgefüges mit ca. 25% im Vergleich zu 50% statistisch hoch signifikant geringer ($p = .004^{N-1}$, $V = .224$) und der Anteil der Beweisversuche mit mindestens einem Beispiel innerhalb des (übrigen) Argumentgefüges mit ca. 60%

im Vergleich zu 34% statistisch hoch signifikant größer ($p = .005^{N-1}$, $V = .220$). Die Schlussweise liegt mit ca. 75% im Vergleich zu rund 50% in der Gruppe $We3229$ statistisch hoch signifikant häufiger zwischen induktiv und deduktiv ($p = .009^{N-1}$, $V = .204$). Insgesamt deuten die Ergebnisse an, dass im Falle von Beweisversuchen in der Gruppe $We3225$, so wie etwa im Falle der Bearbeitung 8.2.6c, oftmals Beispiele genutzt wurden, um eine Beweisidee zu entwickeln oder darzustellen. Mit Blick auf die Weiterentwicklung eines Beweisversuchs aus der Gruppe $We3225$ lässt sich diese Arbeit mit Beispielen ggf. als Potential ansehen. Eine Möglichkeit bestünde darin, bei der Entwicklung eines korrekten und vollständigen Argumentgefüges zunächst bewusst im Konkreten zu bleiben, erst in einem späteren Schritt die Weiterentwicklung zu einem deduktiven Argumentgefüge anzustreben.

Klassenstufe 7

In der Klassenstufe 7 gehören, so wie in der Klassenstufe 6, in jede der Gruppen $We3249$, $We3229$, $We3225$ mindestens 10% der Beweisversuche mit begründendem Ansatz (Abb. 136, 137). Die Beweisversuche in diesen Gruppen machen in der Klassenstufe 7 insgesamt ca. 40% aller Beweisversuche mit begründendem Ansatz aus.

Zudem sind in der Klassenstufe 7 Beweisversuche, die in die Gruppe $\mathbf{We3241}$ gehören, besonders häufig (Abb. 137). Sie machen knapp 20% aller Beweisversuche mit begründendem Ansatz aus. Beweisversuche in der Gruppe $We3241$ passen erkennbar zu der zu zeigenden Aussage und enthalten ein weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke basieren auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Alle diese Beweisversuche beruhen aus Expertensicht auf dem Prinzip der vollständigen Induktion. Es wurde jedoch meistens mindestens ein Einzelfall nicht berücksichtigt oder die erkannte Invariante im Vergleich zum relevanten Erwartungshorizont unzureichend begründet, weshalb die Ausführungen eine inhaltliche Lücke aufweisen. In der Bearbeitung 8.2.6d in Abbildung 133 zum Beispiel kommt der Fall der aufeinanderfolgenden ungeraden Zahlen 1, 3, 5, 7 nicht vor, wenngleich der Satz „Wenn ich also angefangen von 24 (eine Zahl die sich durch 8 teilen lässt) immer 8 dazu addiere, wird dann auch das gewonnene Ergebnis durch 8 teilbar sein“ ein Bewusstsein für die Relevanz des Induktionsanfangs andeutet. Das fehlende Beispiel ließe sich auf Grundlage der Überlegung, dass die zu zeigende Aussage für alle Fälle gelten soll, ergänzen. Auch könnte solch ein Beweisversuch als Anlass genommen werden, um im Sinne einer über die Weiterentwicklung der Beweisversuche hinausgehenden mathematischen Förderung das Prinzip der vollständigen Induktion zu thematisieren. Zusätzlich zu dem unberücksichtigten Einzelfall lassen die Ausführungen in der Bearbeitung 8.2.6d offen, warum „der Abstand zwischen den Summen“ immer acht beträgt.

So wie in der Bearbeitung 8.2.6d liegt in über 95% aller Beweisversuche in der Gruppe $We3241$ die Schlussweise zwischen induktiv und deduktiv, in ca. 80% der Fälle ist sie induktiv mit Andeutung der Verallgemeinerbarkeit. Alle Beweisversuche in der Gruppe $We3241$ enthalten notwendigerweise ein Beispiel zu der zu zeigenden Aussage. In über 95% der Beweisversuche sind Reihenfolgebeispiele zu finden, wobei meistens, so wie in der Bearbeitung 8.2.6d, mindestens drei Reihenfolgebeispiele ausgeführt sind. Die Reihenfolgebeispiele können bei Bedarf hilfreich sein, um die Invariante hinreichend zu begründen, d. h. diese Lücke zu

Bearbeitung 8.2.6d: Klassenstufe 7, weiblich (Code: 0721406)

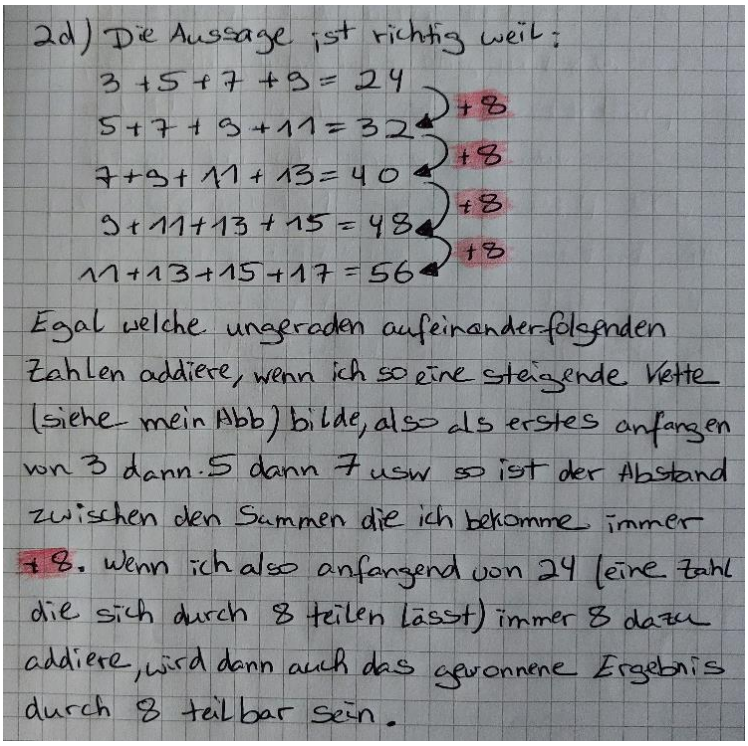


Abbildung 133: Bearbeitung 8.2.6d (We3241 *begründender Ansatz, Schlussweise induktiv mit Andeutung der Verallgemeinerbarkeit, deduktiv mit induktivem Anteil oder deduktiv mit zusätzlichem Beispiel, weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke, primäre Beweisidee* 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*)

schließen. Dazu ließe sich beispielsweise im Falle der Bearbeitung 8.2.6d die Verwendung ikonischer Darstellungsmittel ausbauen. Die Darstellungsform ist in ca. 25% der Beweisversuche in der Gruppe We3241 ikonisch-numerisch-verbal, in fast allen anderen numerisch-verbal.

Die Entwicklung eines allgemeinen Bildes mit dem Ziel, die zu zeigende Aussage zu beweisen, d. h. insbesondere deduktiv zu argumentieren, erscheint hingegen nicht primär naheliegend. Schließlich kommen als ikonische Darstellungsmittel in den analysierten Beweisversuchen, d. h. insbesondere auch in denen in der Gruppe We3241, fast nie figurierte Zahlen vor. Für einen Beweis basierend auf der primären Beweisidee 1 mithilfe von Pfeilen, Linien, o. Ä. ist die Verwendung von Variablen erforderlich, und Variablen sind Lernenden zu Beginn der Klassenstufe 7 in der Regel nicht bekannt. Es ist davon auszugehen, dass eine fast ausschließliche Verwendung verbaler Darstellungsmittel bei der Formulierung deduktiver Argumentgefüge am ehesten an die Erfahrungen und Kompetenzen der Lernenden anknüpfen würde. Dabei kann sprachliche Unterstützung erforderlich sein. Es ist möglich, dass ein versierter Umgang mit Wortvariablen und eine präzise Verwendung von Fachvokabularien den Schülerinnen und Schülern nicht ganz selbstständig gelingt. Mit Blick auf die Formulierung eines deduktiven Argumentgefüges ist zudem zu bedenken, dass den Lernenden vielleicht die Wichtigkeit eines deduktiven Argumentgefüges nicht bekannt ist, die Schlussweise ist, wie angegeben, bei der Mehrheit der Beweisversuche in der Gruppe We3241 induktiv mit Andeutung der Verallgemeinerbarkeit.

Klassenstufe 8

In der Klassenstufe 8 machen die Beweisversuche in der Gruppe **We3229**, so wie in jeder der Klassenstufen 5 bis 7, über 10% aller Beweisversuche mit begründendem Ansatz aus (Abb. 135-138). Der Anteil der Gruppe **We3241**, der auch in der Klassenstufe 7 besonders groß ist, beträgt in der Klassenstufe 8 rund 16%.

Zusätzlich gehören in der Klassenstufe 8 mit rund 14% der Zuordnungen ebenfalls mehr als ein Zehntel der Beweisversuche mit begründendem Ansatz in die Gruppe **We3151** (Abb. 138). Die Beweisversuche in der Gruppe We3151 basieren auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* und enthalten ein weitgehend korrektes und vollständiges Argumentgefüge. Im Vergleich zu den Beweisversuchen in der Gruppe We3241 ist keine inhaltliche Lücke vorhanden, was unter anderem impliziert, dass ein weitgehend korrektes und vollständiges Beispiel zu der zu zeigenden Aussage angegeben sein muss. In über 80% der Beweisversuche in der Gruppe We3151 sind das kleinste Beispiel und Reihenfolgebeispiele ausgeführt. Die Schlussweise liegt per Definition bei den Beweisversuchen in der Gruppe We3151 zwischen induktiv und deduktiv, meistens ist sie deduktiv mit induktivem Anteil. Damit die Beweisversuche in der Gruppe We3151 als Beweis bezeichnet werden können, fehlt in der Regel nur wenig. Es ist erforderlich, die relevanten sachinhaltlichen Ausführungen vollständig zu verallgemeinern, so dass die Schlussweise deduktiv ist, und eventuell Formulierungsfehler oder Formulierungslücken zu beseitigen. Beispielsweise wird in der Bearbeitung 8.2.6e in Abbildung 134 die Vergrößerung der Summe um acht am Beispiel der aufeinanderfolgenden ungeraden Zahlen 1, 3, 5, 7 und 3, 5, 7, 9 festgemacht, wobei die Verallgemeinerung bereits weitgehend angelegt ist. Die Bearbeitung enthält zusätzlich einen Formulierungsfehler, der zu korrigieren wäre („$1 + 3 + 5 + 7 = 16 : 8 = 2$“). In diesem Zuge ließe sich auch die Formulierung „Da es bei $1 + 3 + 5 + 7$ funktioniert […]“ präzisieren.

Bearbeitung 8.2.6e: Klassenstufe 8, männlich (Code: 0821408)

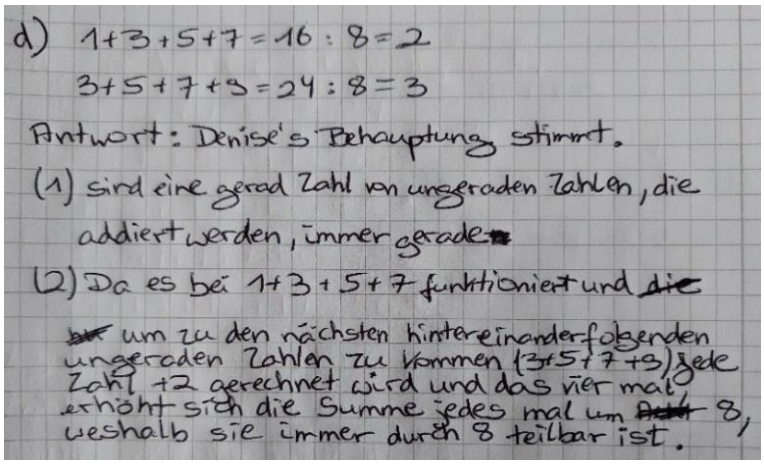


Abbildung 134: Bearbeitung 8.2.6e (We3151 *begründender Ansatz, Schlussweise zwischen induktiv und deduktiv, weitgehend korrektes und vollständiges Argumentgefüge, primäre Beweisidee* 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*)

Vergleich der Klassenstufen 5 bis 8

In Ergänzung zu den vorhergehenden Ausführungen lassen sich auf Grundlage des Vergleichs der relativen Häufigkeitsverteilungen der Gruppen in den einzelnen Klassenstufen 5 bis 8 mehrere Auffälligkeiten herausstellen, die mit Blick auf Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche mit begründendem Ansatz relevant sind.

Beispielsweise ist zu betonen, dass es in der Klassenstufe 8 Beweisversuche mit begründendem Ansatz gibt, deren Ausprägung für die Klassenstufe 5 charakteristisch ist, und umgekehrt. De facto kommen in jeder der Klassenstufen 5 bis 8 Beweisversuche mit begründendem Ansatz vor, die in die Gruppe We3249 *Schlussweise deduktiv ohne zusätzliches Beispiel, kein Hinweis auf einen grundlegenden Fehler, kein weiterentwickelbares Argumentgefüge, keine ausbaubare primäre Beweisidee* gehören, sowie solche, die der Gruppe We3151 *zwischen induktive und deduktive Schlussweise, weitgehend korrektes und vollständiges Argumentgefüge, primäre Beweisidee* 1 zuzuordnen sind (Abb. 135-138). Das **Spektrum an vorkommenden Beweisversuchen mit begründendem Ansatz** ist mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis in jeder der Klassenstufen 5 bis 8 groß.

Dabei deutet der Vergleich der relativen Häufigkeitsverteilungen der mit Blick auf Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche mit begründendem Ansatz zu einem Beweis gebildeten Gruppen an Beweisversuchen an, dass die Mehrheit der Beweisversuche mit begründendem Ansatz aus der Klassenstufe 5 mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis weitgehend homogen ist und **je höher die Klassenstufe, desto heterogener**. Von den 29 mit Blick auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche mit begründendem Ansatz zu einem Beweis gebildeten Gruppen kommen in der fünften Klassenstufe 14, in der sechsten und siebten Klassenstufe 18 und in der achten Klassenstufen 23 im analysierten Sample vor (Abb. 135-138). Dabei gehören in der Klassenstufe 5 rund 60% der Beweisversuche mit begründendem Ansatz zu einer der beiden Gruppen We3229, We3429 und weitere rund 15% in eine der beiden Gruppen We3225, We3425. Insgesamt kennzeichnet rund drei Viertel der Beweisversuche mit begründendem Ansatz aus der Klassenstufe 5, dass allenfalls ein Ansatz für ein weiterentwickelbares Argumentgefüge, wenn auch kein Hinweis auf einen grundlegenden Fehler zu erkennen ist. Demgegenüber macht in den Klassenstufen 7, 8 keine der Gruppen einen Anteil von 20% oder mehr aus. Zudem sind die Gruppen, in die mindestens 5% der Beweisversuche mit begründendem Ansatz fallen, unterschiedlicher als in der Klassenstufe 5, weil dazu nicht nur Gruppen an Beweisversuchen ohne weiterentwickelbares Argumentgefüge gehören, sondern z. B. auch die Gruppe We3151 mit Beweisversuchen mit einem weitgehend korrekten und vollständigen Argumentgefüge (Abb. 135, 137-138).

Die Beweisversuche in den **Gruppen We3225,3229,3425,3429** machen zusammen genommen in jeder der Klassenstufen 5 bis 8 mehr als ein Viertel aller Beweisversuche mit begründendem Ansatz aus. Zwar enthalten die Beweisversuche in diesen Gruppen keinen Hinweis auf einen grundlegenden Fehler, aber auch kein weiterentwickelbares Argumentgefüge, so dass die vorhandenen Ausführungen mit Blick auf die Entwicklung eines Beweises nur als erster Ansatz genutzt werden können.

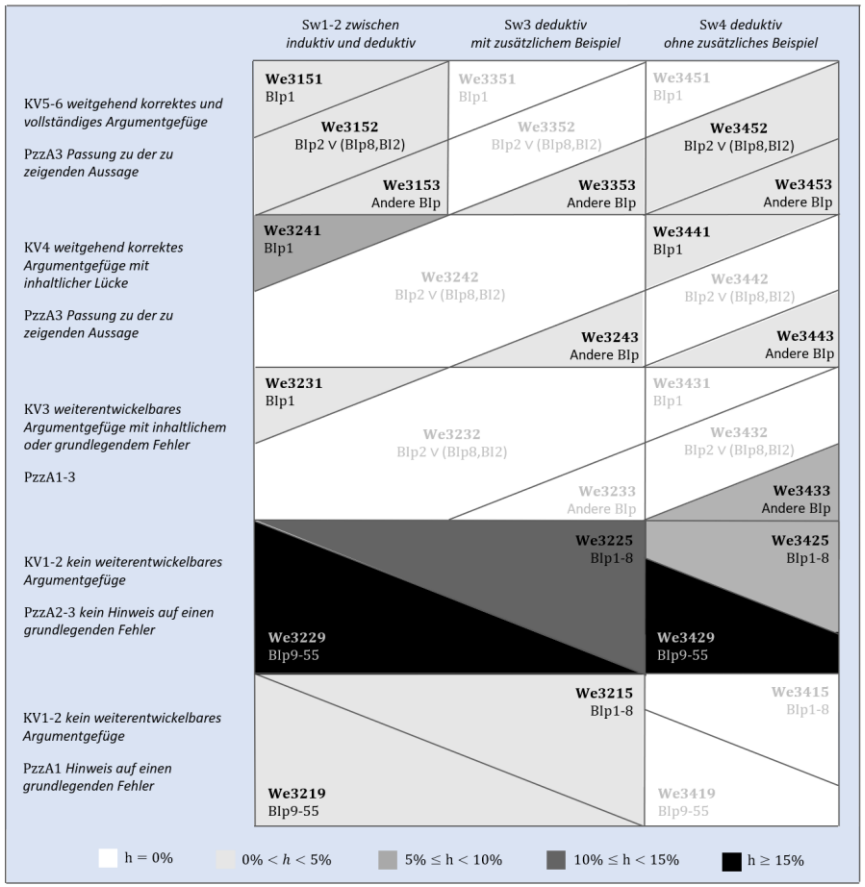


Abbildung 136: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Häufigkeitsverteilung der Beweisversuche (Fokus: Klassenstufe 6)

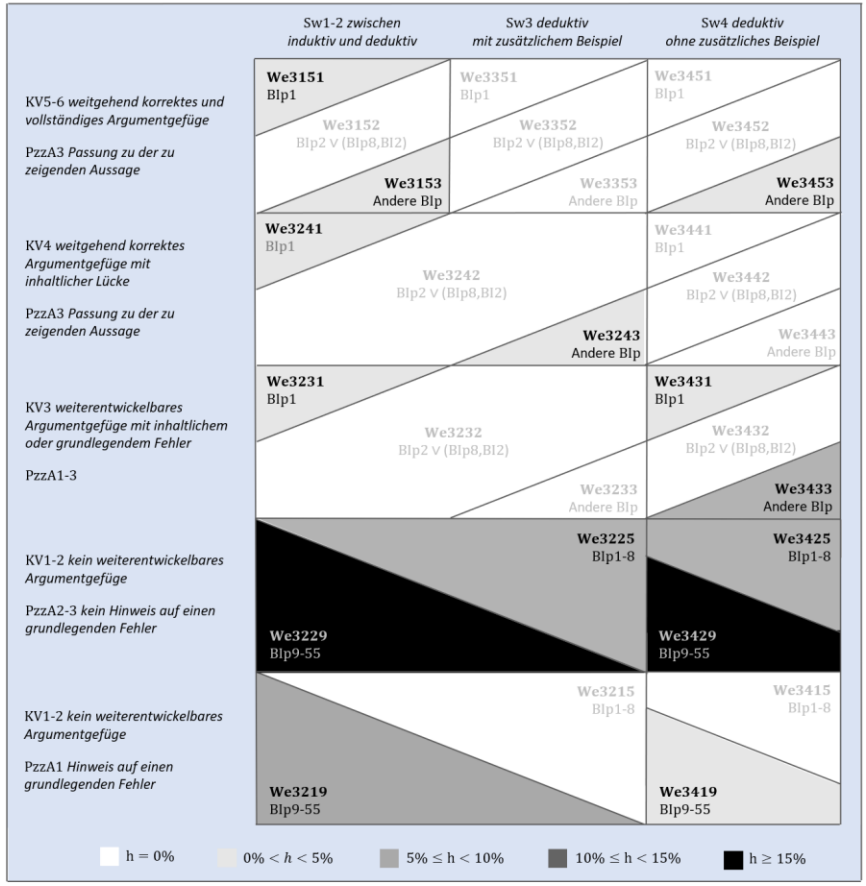


Abbildung 135: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Häufigkeitsverteilung der Beweisversuche (Fokus: Klassenstufe 5)

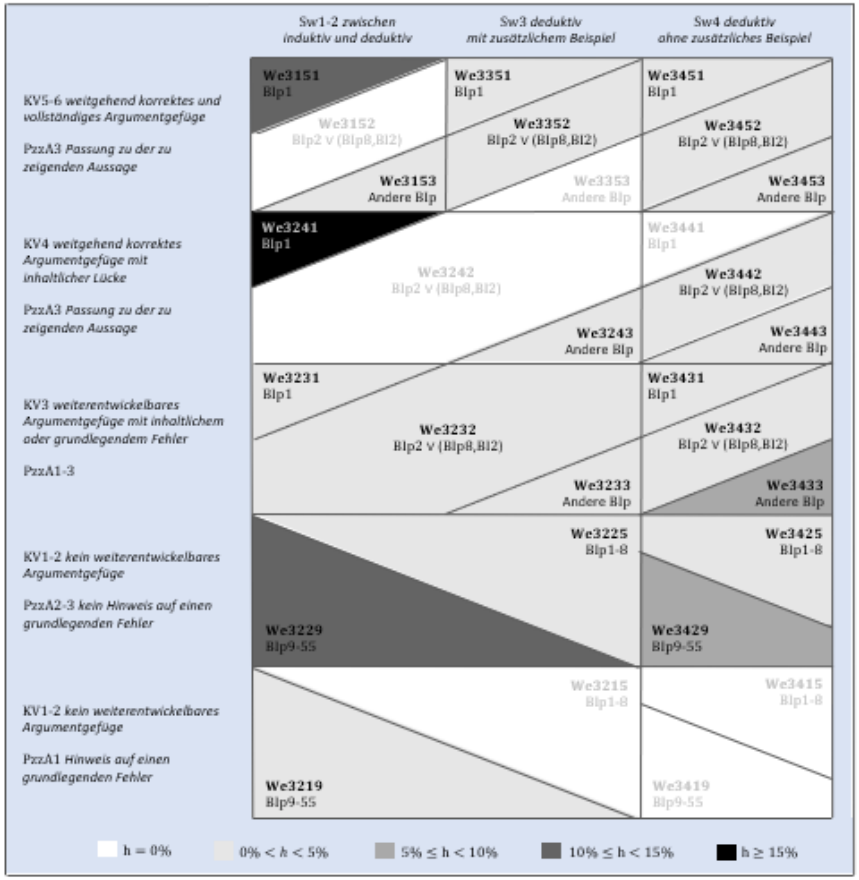


Abbildung 138: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Häufigkeitsverteilung der Beweisversuche (Fokus: Klassenstufe 8)

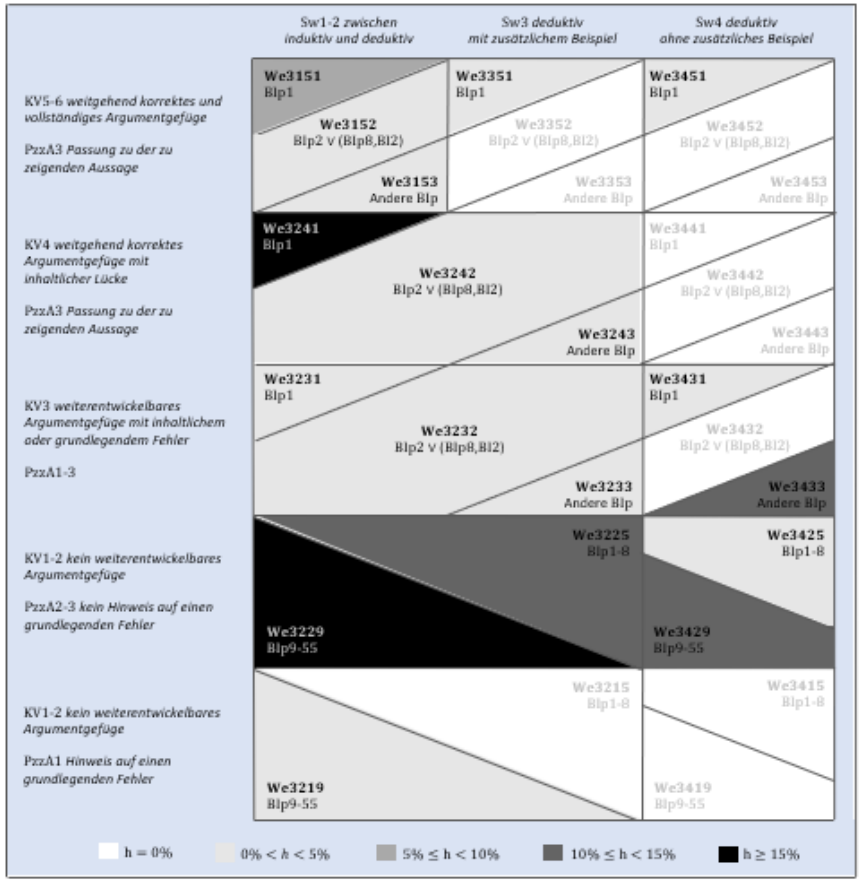


Abbildung 137: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Häufigkeitsverteilung der Beweisversuche (Fokus: Klassenstufe 7)

Der Gesamtanteil der Gruppen We3225,3229,3425,3429 ist in der fünften Klassenstufe mit 74% am höchsten, gefolgt von der sechsten Klassenstufe mit 64%, der siebten Klassenstufe mit 45% und der achten Klassenstufe mit 27%. Mit Ausnahme der Differenz zwischen den Klassenstufen 5 und 6 sind alle erkennbaren Unterschiede statistisch hoch signifikant (jeweils $p < 0{,}01^{N-1}$, $0{,}17 < V < 0{,}45$). Dies deutet an, dass es sich hierbei um eine Ausgangslage handelt, die in den Klassenstufen 5, 6 und auch in der Klassenstufe 7 ganz besonders häufig vorkommt. Dabei ist in jeder der Klassenstufen 5 bis 8 in 20% bis 37% der Beweisversuche in den Gruppen We3225,3229,3425,3429 die primäre Beweisidee eine ausbaubare Beweisidee, sodass die primäre Beweisidee bei der Weiterentwicklung der Beweisversuche gut aufgegriffen und prinzipiell beibehalten werden kann.

Neben den Gruppen We3225,3229,3425,3429 machen in den Klassenstufen 7, 8 die Beweisversuche in den **Gruppen We3241,3151** einen großen Anteil der Beweisversuche mit begründendem Ansatz aus. In den Klassenstufen 7, 8 gehören mehr als ein Viertel der Beweisversuche mit begründendem Ansatz in eine der beiden Gruppen We3241,3151, in den Klassenstufen 5, 6 jeweils etwa ein Zwanzigstel. Der Gesamtanteil dieser zwei Gruppen ist in beiden Klassenstufen 7, 8 statistisch hoch signifikant größer als in den Klassenstufen 5, 6 (jeweils $p < .001^{N-1}$, $0{,}26 < V < 0{,}40$). Diese Auffälligkeit geht mit der häufigen Verwendung der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* in den Klassenstufen 7, 8 einher (Abschn. 8.2.2, 8.2.6.3). Im Falle eines Beweisversuchs in einer der Gruppen We3241,3151 liegt die Vermutung nahe, dass der Schülerin bzw. dem Schüler auch bei Vorlage anderer Aussagen aus dem Themenbereich der Teilbarkeit ganzer Zahlen, zu denen sich mithilfe einer systematischen Beispielnutzung und des Invarianzprinzips ein Beweis entwickeln lässt, zumindest ein weitgehend korrektes Argumentgefüges gelingt.

Nach den Ausführungen in Abschnitt 8.2.5.2 sind die Beweisversuche mit begründendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage, vereinfacht gesagt, je höher die Klassenstufe, desto besser. Damit einhergehend sind in den Beweisversuchen mit begründendem Ansatz im Mittel **je höher die Klassenstufe, desto tragfähigere Anknüpfungspunkte** für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen. Beispielsweise enthält je höher die Klassenstufe, ein desto größerer Anteil der Beweisversuche mit begründendem Ansatz ein weitgehend korrektes Argumentgefüge (Abschn. 8.2.5.2). Die Existenz eines weitgehend korrekten Argumentgefüges impliziert, dass die vorhandenen Überlegungen bei der Entwicklung eines Beweises im Wesentlichen beibehalten werden können, nur eventuell zu ergänzen und zu verallgemeinern sind.

8.2.6.8 Zusammenfassung

Um in möglichst aussagekräftiger und kompakter Form zu beschreiben, wie die Beweisversuche mit begründendem Ansatz von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis ausgeprägt sind, wurden die Beweisversuche mit begründendem Ansatz klassifiziert.

Für die Klassifikation wurden die Beweisversuche mit begründendem Ansatz zuerst in Abhängigkeit der **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges**

sowie der **Schlussweise** gruppiert, weil die Ausprägungen dieser beiden Merkmale mit Blick auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis von grundlegender Bedeutung sind. Insgesamt wurden fünfzehn Gruppen KV1,Sw1-2 bis KV5-6,Sw4 gebildet.

Um darauf aufbauend herauszuarbeiten, wie sich diese Gruppen für die Klassifikation der Beweisversuche weiter zusammenfassen lassen, und andererseits, welche Binnendifferenzierungen sinnvoll sind, wurden weitere Eigenschaften der Beweisversuche mit begründendem Ansatz in Zusammenhang mit der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweisen betrachtet. Diese Eigenschaften sind die Passung der Ausführungen zu der zu zeigenden Aussage, die primäre Beweisidee, die Existenz, die Art und der Ort von Beispielen sowie die Darstellungsform.

Bezogen auf die **Passung der Ausführungen zu der zu zeigenden Aussage** ist grundlegend festzuhalten: Aus allen Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge geht der Bezug auf die zu zeigende Aussage hervor. Vor allem bei den Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel* ist der Bezug auf die zu zeigende Aussage oftmals unklar. Von den Beweisversuchen mit begründendem Ansatz, keinem weiterentwickelbaren Argumentgefüge und einer der Schlussweise Sw1-3 enthalten mehr als 10% einen Hinweis auf einen grundlegenden Fehler.

Weiterhin deuten die Ergebnisse auf Zusammenhänge zwischen **primären Beweisideen**, der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise hin, von denen die meisten in Einklang mit den Überlegungen im Rahmen der Aufgabenkonstruktion stehen. Beispielsweise enthalten mehr als zwei Drittel der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* ein weitgehend korrektes Argumentgefüge bei zwischen induktiver und deduktiver Schlussweise. Von denen mit der primären Beweisidee 2 *Darstellung einer Summe mit Erkennen einer Invariante* sowie von denen mit der primären Beweisidee 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* und der Beweisidee 2 enthalten mehr als zwei Drittel ein weitgehend korrektes Argumentgefüge bei deduktiver Schlussweise. Auch die Beweisversuche mit der primären Beweisidee 5 *Reste bei Division einzelner Summanden durch vier oder acht* enthalten mehrheitlich ein weitgehend korrektes Argumentgefüge, während in keinem der Beweisversuche mit begründendem Ansatz und einer der primären Beweisideen 4, 9-13, 55 ein weiterentwickelbares Argumentgefüge vorkommt.

In Bezug auf die **Beispielnutzung** sind folgende Tendenzen hervorzuheben: Beweisversuche mit begründendem Ansatz und einem Beispiel zu der zu zeigenden Aussage, solche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges sowie solche mit Reihenfolgebeispielen und dem kleinsten Beispiel sind hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges im Durchschnitt eher stark. Beweisversuche mit begründendem Ansatz ohne Beispiel oder ausschließlich mit sonstigen Beispielen, solche mit einem großen Beispiel und solche mit allen Beispielen oberhalb des (übrigen) Argumentgefüges sind hinsichtlich der der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges

im Durchschnitt eher schwach. Dabei scheint es jedoch auf den Umgang mit den Beispielen anzukommen.

Zudem scheinen die Existenz, die Art und der Ort von Beispielen in einem gewissen Maße mit primären Beweisideen zusammenzuhängen. Beispielsweise enthalten die Beweisversuche mit begründendem Ansatz und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* häufiger Reihenfolgebeispiele und häufiger ein Beispiel oberhalb des (übrigen) Argumentgefüges als die Beweisversuche mit begründendem Ansatz und einer anderen ausbaubaren primären Beweisidee.

Was die **Darstellungsform** betrifft, ist unter anderem folgende Auffälligkeit zu sehen: Bei mehr als zwei Drittel der Beweisversuche mit der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* sowie bei allen analysierten Beweisversuchen mit der primären Beweisidee 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* und der Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* hat die Darstellungsform symbolischen Anteil. Bei den übrigen Beweisversuchen mit begründendem Ansatz hingegen kommen die Darstellungsformen mit symbolischem Anteil kaum vor.

Unter Berücksichtigung der skizzierten und weiterer Zusammenhänge zwischen verschiedenen Eigenschaften von Beweisversuchen mit begründendem Ansatz werden in der **Klassifikation** der Beweisversuche mit begründendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis 29 Gruppen unterschieden. Diese 29 Gruppen an Beweisversuchen mit begründendem Ansatz differieren insbesondere hinsichtlich der Schlussweise, der Korrektheit und Vollständigkeit der Argumentgefüge und der primären Beweisidee.

In der Klassenstufe 5 gehören rund 60% aller Beweisversuche mit begründendem Ansatz zu einer der Gruppen We3229,3429. Die Beweisversuche in diesen Gruppen enthalten zwar keinen Hinweis auf einen grundlegenden Fehler, aber auch kein weiterentwickelbares Argumentgefüge und keine ausbaubare primäre Beweisidee. In der Klassenstufe 6 beträgt der Anteil der Beweisversuche mit begründendem Ansatz in diesen beiden Gruppen etwa 50%. Weitere rund 15% gehören zu einer der Gruppen We3225,3429 mit Beweisversuchen ohne weiterentwickelbares Argumentgefüge, aber mit einer ausbaubaren primären Beweisidee, sodass die primäre Beweisidee als wesentlicher Anknüpfungspunkt bei der Weiterentwicklung des Beweisversuchs genutzt werden kann. Neben den genannten Gruppen We3225,3229,3425,3429 machen in den Klassenstufen 7 und 8 auch die Gruppen We3241,3151 große Anteile aus. Die Beweisversuche in den Gruppen We3241,3151 enthalten ein weitgehend korrektes Argumentgefüge basierend auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, fast immer bei zwischen induktiver und deduktiver Schlussweise.

Insgesamt betrachtet erscheint das Spektrum an vorkommenden Beweisversuchen mit begründendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis in jeder der Klassenstufen 5 bis 8 groß und je höher die Klassenstufe, desto größer. Zugleich sind in den Beweisversuchen mit begründendem Ansatz im Mittel je höher die Klassenstufe, desto tragfähigere Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen.

8.3 Beweisversuche mit prüfendem Ansatz

Nachdem Beweisversuche mit begründendem Ansatz ausführlich analysiert wurden, fokussiert dieser Abschnitt auf Beweisversuche mit prüfendem Ansatz. Beweisversuche mit prüfendem Ansatz können nicht der Kategorie *begründender Ansatz* zugeordnet werden, enthalten aber mindestens ein Beispiel, das einen Bezug zu der infrage stehenden Aussage aufweist (Abschn. 3.3.2). Es wurde zumindest versucht, einen Spezialfall der Aussage zu betrachten. Dabei können z. B. hinsichtlich der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage Unterschiede zwischen verschiedenen Beweisversuchen mit prüfendem Ansatz bestehen. In den Abschnitten 8.3.1 bis 8.3.4 werden differenzierte Einblicke in die Ausprägungen der Beweisversuche mit prüfendem Ansatz gegeben.

Zuerst wird dargelegt, inwieweit die Schülerinnen und Schüler ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage angegeben und somit die zu zeigende Aussage für einen Spezialfall bestätigt haben. Dies ist mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage von Bedeutung (Abschn. 8.3.1). Danach werden Eigenschaften der angegebenen Beispiele und der Angabe des Wahrheitswerts beschrieben. Es wird dargelegt, welche Eigenschaften in den Beweisversuchen von Lernenden in den Klassenstufen 5 bis 8 besonders häufig oder eher selten vorkommen und herausgearbeitet, welche Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind (Abschn. 8.3.2). Abschließend werden die Beweisversuche mit Blick auf den Nachweis der Gültigkeit sowie mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung klassifiziert (Abschn. 8.3.3, 8.3.4). Der Ausdruck „Beweisversuche“ wird dabei stets im Sinne von „Beweisversuche mit prüfendem Ansatz“ verwendet.

8.3.1 Passung eines Beispiels zu der zu zeigenden Aussage

In diesem Abschnitt wird dargelegt, inwieweit Teilnehmende der Mathematik-Olympiade in Beweisversuchen mit prüfendem Ansatz ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage angegeben und somit die zu zeigende Aussage selbstständig für einen Spezialfall bestätigt haben. Dazu wurde jeder Beweisversuch mit prüfendem Ansatz genau einer von fünf Kategorien zugeordnet (Abschn. 3.3.7). Die Abbildung 139 zeigt die erhaltenen relativen Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen.

Entsprechend Abbildung 139 kommt je Klassenstufe in 35% bis 50% der Beweisversuche mit prüfendem Ansatz ein vollständig und korrekt ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage vor ($\mathrm{PBsp5}$). Folglich haben pro Klassenstufe zwischen einem Drittel und der Hälfte der Teilnehmenden, die einen Beweisversuch mit prüfendem Ansatz abgegeben haben, die zu zeigende Aussage eigenständig für einen Spezialfall bestätigt.

Der Gesamtanteil der Beweisversuche mit einem Beispiel zu der zu zeigenden Aussage ($\mathrm{PBsp3\text{-}5}$) an allen Beweisversuchen mit prüfendem Ansatz liegt je Klassenstufe bei ca. 65% bis 90%.[152] In den meisten anderen Beweisversuchen mit prüfendem Ansatz wurde eine Summe

[152] Die in Abbildung 139 zu sehenden Unterschiede zwischen der Klassenstufe 7 und den Klassenstufen 5, 6, 8 sind nicht statistisch signifikant.

von vier nicht aufeinanderfolgenden ungeraden Zahlen betrachtet (PBsp2), so dass die Vermutung nahe liegt, dass die Lernenden ein Beispiel zu der zu zeigenden Aussage angeben wollten, diese jedoch nicht ganz korrekt interpretiert und dadurch de facto kein Beispiel zu der zu zeigenden Aussage angegeben haben.

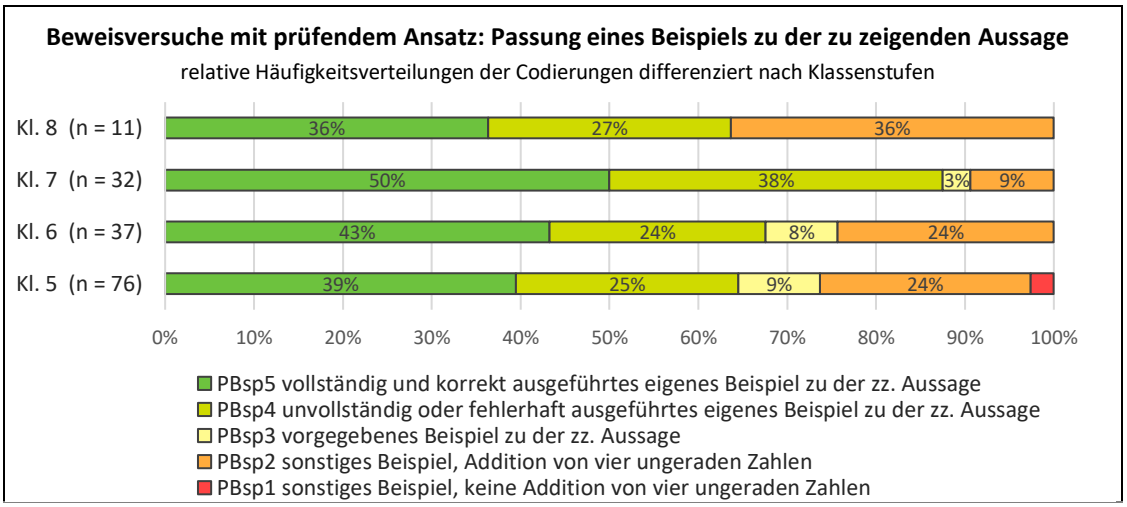


Abbildung 139: Beweisversuche mit prüfendem Ansatz: Passung eines Beispiels zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen

8.3.2 Eigenschaften der Beweisversuche mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

Dieser Abschnitt beschäftigt sich mit den Eigenschaften der Beispiele und der Angabe eines Wahrheitswerts in Beweisversuchen mit prüfendem Ansatz. Er zielt auf die Beantwortung der Teilfrage, welche Eigenschaften die angegebenen Beispiele und die Angabe eines Wahrheitswerts mit besonderem Fokus auf Ansatzpunkte für die Entwicklung eines Beweises haben.

Die Eigenschaften der angegebenen Beispiele und der Angabe eines Wahrheitswerts sind vielfältig, daher wird für eine aussagekräftige und konsistente Darstellung der Eigenschaften eine Perspektive festgelegt, die mit Blick auf die Fragestellung der Arbeit insbesondere auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche mit prüfendem Ansatz zu einem Beweis fokussiert. In Abschnitt 8.3.2.1 werden Eigenschaften fokussiert, die Anhaltspunkte hinsichtlich der Interpretation der zu zeigenden Aussage liefern. Der Abschnitt 8.3.2.2 beschäftigt sich mit sachinhaltlichen und heuristischen Anknüpfungspunkten für die Entwicklung eines korrekten und vollständigen Argumentgefüges. In Abschnitt 8.3.2.3 geht es um Hinweise auf die verwendete Schlussweise und in Abschnitt 8.3.2.4 um die Darstellungsmittel und Darstellungsformen. Die Ergebnisse aller Abschnitte werden genutzt, um die Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis zu klassifizieren (Abschn. 8.3.4).

Innerhalb der Abschnitte 8.3.2.1 bis 8.3.2.4 erfolgt eine primär deskriptive Ergebnispräsentation. Es werden zunächst immer Eigenschaften der Beispiele und der Angabe des Wahrheitswerts beschrieben und in einem nächsten Schritt interpretiert. Die Ergebnisse zeigen deskriptiv, welche Eigenschaften in den Beweisversuchen mit prüfendem Ansatz von Lernenden in den

Klassenstufen 5 bis 8 besonders häufig oder eher selten vorkommen, und können zudem über Anknüpfungspunkte für die Entwicklung eines Beweises informieren.

8.3.2.1 Interpretation der zu zeigenden Aussage

Für die Entwicklung eines Beweises ist eine korrekte Interpretation der zu zeigenden Aussage von grundlegender Bedeutung. Deshalb werden im vorliegenden Abschnitt Eigenschaften der Beweisversuche mit prüfendem Ansatz beschrieben, die diesbezüglich Hinweise bieten. Die relevanten Eigenschaften sind die Passung der Ausführungen zu der zu zeigenden Aussage, die Existenz und Art eines angegebenen Wahrheitswerts und Hinweise auf die Berücksichtigung der Frage nach der Allgemeingültigkeit der Aussage (Abschn. 3.3.7 zum Kategoriensystem).

Passung der Ausführungen zu der zu zeigenden Aussage

Die Ergebnisse der Analyse der Passung der Ausführungen in Beweisversuchen mit prüfendem Ansatz zu der zu zeigenden Aussage deuten an, inwieweit die Begrifflichkeiten der zu zeigenden Aussage korrekt interpretiert wurden. Je Klassenstufe ist in über 80% der Beweisversuche mit prüfendem Ansatz entweder der Bezug auf die zu zeigende Aussage oder aber ein Hinweis auf einen grundlegenden Fehler zu erkennen. Ein Hinweis auf einen grundlegenden Fehler ist jeweils in 35% bis 55% der Beweisversuche mit prüfendem Ansatz zu sehen (Abb. 140).

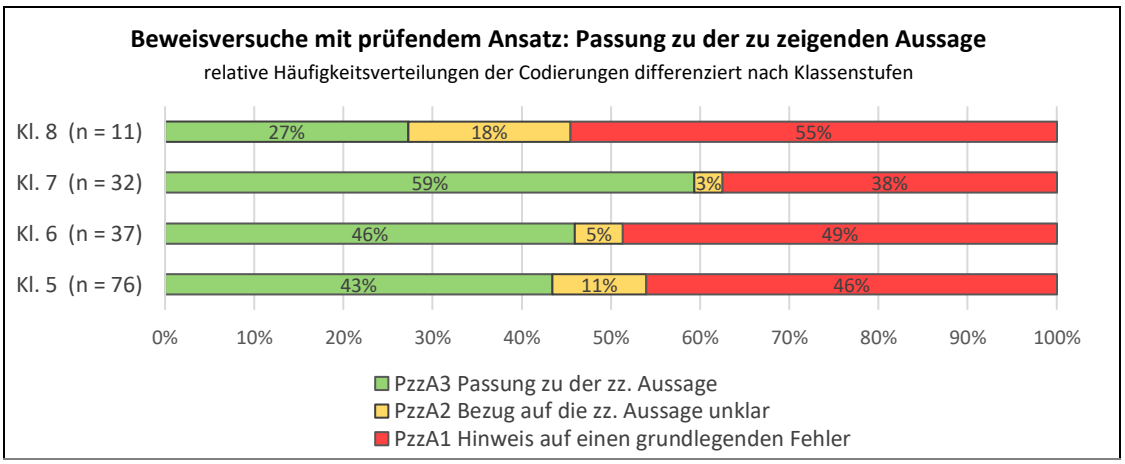


Abbildung 140: Beweisversuche mit prüfendem Ansatz: Passung der Ausführungen zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen

Existenz und Art eines angegebenen Wahrheitswerts

Ein angegebener Wahrheitswerts kann darauf hinweisen, inwieweit die infrage stehende Aussage entsprechend der Aufgabenstellung als gültig angesehen wird. Die Ergebnisse der Erfassung der Existenz, Art eines angegebenen Wahrheitswerts sind in allen Klassenstufen 5 bis 8 ähnlich (Abb. 141). In jeweils über 70% der Beweisversuche mit prüfendem Ansatz ist ein Wahrheitswert angegeben. Von den Beweisversuchen mit prüfendem Ansatz und einem Wahrheitswert enthalten jeweils über 50% eine Affirmation der Aussage, die meisten anderen jeweils eine Negation. In den Klassenstufen 5 bis 7 ist jeweils in einem Anteil von etwa 5% aller Beweisversuche mit prüfendem Ansatz ein Wahrheitswert angegeben, der nicht eindeutig als Affirmation oder als Negation der Aussage bezeichnet werden kann.

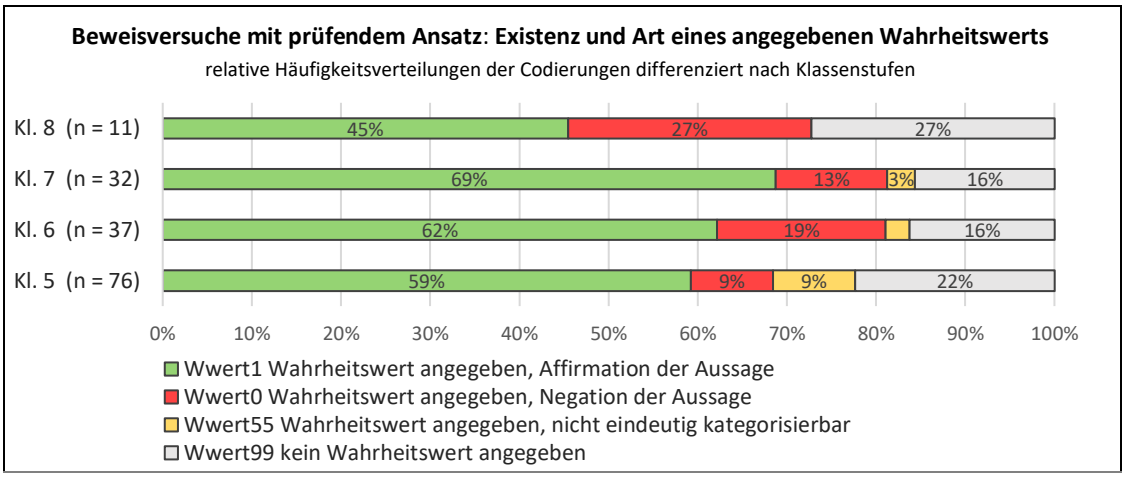


Abbildung 141: Beweisversuche mit prüfendem Ansatz: Existenz und Art eines angegebenen Wahrheitswerts. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen

Um bei Beweisversuchen, in denen kein Wahrheitswert angegeben ist, eine Tendenz auszumachen, ob die Aussage eher als gültig oder als ungültig angesehen wird, lässt sich die Anzahl und Reihenfolge von Positiv- und Negativbeispielen betrachten (Abschn. 3.3.7). Von den Beweisversuchen mit prüfendem Ansatz, in denen kein Wahrheitswert angegeben ist, enthalten 85% ausschließlich Positivbeispiele. In allen anderen analysierten Fällen kommt genau ein Negativbeispiel vor, das zudem das letzte Beispiel ist. Wenn man diese Beweisversuche, in denen kein Wahrheitswert angegeben ist, denen in den Kategorien *Affirmation der Aussage* bzw. *Negation der Aussage* zuordnet, liegt der Anteil an Beweisversuchen mit einer Affirmation der Aussage in jeder Klassenstufe zwischen gut 70% und 85%, der Anteil der Beweisversuche mit einer Negation der Aussage zwischen gut 10% und 30% und der Anteil derjenigen mit einem nicht eindeutig kategorisierbaren Wahrheitswert bei maximal 10%.[153]

Hinweise auf die Berücksichtigung der Frage nach der Allgemeingültigkeit

Eine Bewusstheit für die Frage nach der Allgemeingültigkeit der zu zeigenden Aussage gehört zu einer korrekten Interpretation dieser Aussage und ist erforderlich, um die Notwendigkeit einer allgemeingültigen Begründung zu verstehen. Daher beschäftigt sich dieser Abschnitt damit, inwieweit die Frage nach der Allgemeingültigkeit der zu zeigenden Aussage den Ausführungen der Lernenden zufolge bei der Bearbeitung bedacht wurde. In der Klassenstufe 5 enthalten gut 25%, in der Klassenstufe 6 knapp 40%, in der Klassenstufen 7 rund 45% und in der Klassenstufe 8 knapp 55% der Beweisversuche mit prüfendem Ansatz einen Hinweis auf die Berücksichtigung der Frage nach der Allgemeingültigkeit der Aussage. Meistens ist ein Hinweis auf die Allgemeingültigkeit der Aussage an sich zu erkennen. In einem der analysierten Beweisversuche mit prüfendem Ansatz aus der Klassenstufe 7 kommt ein Hinweis auf die Unzulänglichkeit von Beispielen vor. Dabei ist zu bedenken, dass die Analyse insbesondere bei den Beweisversuchen ohne irgendeinen Hinweis offenlässt, ob die Frage nach der Allgemeingültigkeit

[153] Bei dieser Zuordnung ist jedoch Vorsicht geboten. Dass in einem Beweisversuch mit prüfendem Ansatz kein Wahrheitswert angegeben ist, kann vielfältige Gründe haben, wie z. B. ein Abbruch aus Zeitgründen oder weil keine Beweisidee gefunden wurde, oder dass die Angabe obsolet erschien, weil die Gültigkeit in der Aufgabenstellung vorgegeben war oder die Ungültigkeit mit Blick auf ein Negativbeispiel als offensichtlich angesehen wurde, es könnte aber auch eine Unsicherheit vorliegen.

bedacht wurde. Es ist möglich, dass der tatsächliche Anteil an Lernenden, die diese Frage bedacht haben, größer ist als angegeben (Abschn. 3.3.7).

Zusammenfassende Betrachtung: Interpretation der zu zeigenden Aussage

Um zusammenfassende Aussagen hinsichtlich der Interpretation der zu zeigenden Aussage zu treffen, werden die Ergebnisse der drei Analysen entsprechend Abbildung 142 miteinander kombiniert und die Beweisversuche mit prüfendem Ansatz gruppiert.

		Existenz und Art einer Angabe eines Wahrheitswerts			
		Wahrheitswert angegeben			Wwert99 *kein Wahrheitswert angegeben*
		Wwert1 *Affirmation der Aussage*	Wwert0 *Negation der Aussage*	Wwert55 *nicht eind. kategorisierbar*	
Passung der Ausführungen zu der zz. Aussage	PzzA3 *Passung zu der zz. Aussage*	IntzzA3 (+Allg.[1]) 13% ($n = 20$) IntzzA2 22% ($n = 35$)	IntzzA1 (+Allg.[1]) 0% ($n = 0$) IntzzA1 0% ($n = 0$)	IntzzA1 (+Allg.[1]) 0% ($n = 0$) IntzzA1 0% ($n = 0$)	IntzzA3 (+Allg.[1]) 1% ($n = 2$) IntzzA2 10% ($n = 15$)
	PzzA2 *Bezug auf die zz. Aussage unklar*	IntzzA2 (+Allg.[1]) 1% ($n = 2$) IntzzA2 4% ($n = 7$)	IntzzA1 (+Allg.[1]) 0% ($n = 0$) IntzzA1 0% ($n = 0$)	IntzzA1 (+Allg.[1]) 0% ($n = 0$) IntzzA1 0% ($n = 0$)	IntzzA2 (+Allg.[1]) 1% ($n = 1$) IntzzA2 2% ($n = 3$)
	PzzA1 *Hinweis auf einen grundlegenden Fehler*	IntzzA1 (+Allg.[1]) 8% ($n = 13$) IntzzA1 11% ($n = 18$)	IntzzA1 (+Allg.[1]) 10% ($n = 15$) IntzzA1 4% ($n = 6$)	IntzzA1 (+Allg.[1]) 1% ($n = 1$) IntzzA1 5% ($n = 8$)	IntzzA1 (+Allg.[1]) 1% ($n = 1$) IntzzA1 6% ($n = 9$)

[1] +Allg: mit einem Hinweis auf die Berücksichtigung der Frage nach der Allgemeingültigkeit

Abbildung 142: Gruppierung der Beweisversuche mit prüfendem Ansatz mit Blick auf die Interpretation der zu zeigenden Aussage. Gebildete Gruppen und deren Häufigkeiten

Die Abkürzungen IntzzA3, IntzzA2, IntzzA1 in Abbildung 142 bezeichnen die gebildeten Gruppen. Bei Beweisversuchen mit prüfendem Ansatz in der Gruppe **IntzzA3** legen die Ausführungen nahe, dass die zu zeigende Aussage korrekt interpretiert wurde. Bei diesen Beweisversuchen mit prüfendem Ansatz passen die vorhandenen Ausführungen zu der zu zeigenden Aussage, es ist die Gültigkeit der zu zeigenden Aussage angegeben und ein Hinweis auf die Frage nach der Allgemeingültigkeit zu sehen (Bearbeitung 8.3.2a in Abb. 143). Der relative Anteil der Beweisversuche in dieser Gruppe an allen Beweisversuchen mit prüfendem Ansatz beträgt je Klassenstufe bis zu 25%. Dabei ist der Unterschied zwischen der Klassenstufen 5 mit 9% und der Klassenstufe 7 mit 25% statistisch signifikant ($p = .012^{ex}$, $V = .267$). Dies deutet an, dass aus Beweisversuchen mit prüfendem Ansatz aus der Klassenstufe 7 eher die korrekte Interpretation der zu zeigenden Aussage hervorgeht als aus denen aus der Klassenstufe 5.

Bearbeitung 8.3.2a: Klassenstufe 6, männlich (Code: 0620109)

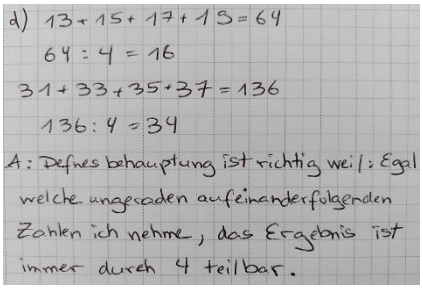


Abbildung 143: Bearbeitung 8.3.2a (IntzzA3 *Andeutung einer korrekten Interpretation der zu zeigenden Aussage*)

In die Gruppe **IntzzA2** gehören je Klassenstufe ca. 35% bis 45% der Beweisversuche, zum Beispiel die Bearbeitung 8.3.2b in Abbildung 144. Aus diesen Beweisversuchen geht die korrekte Interpretation der zu zeigenden Aussage nicht klar hervor, insbesondere ist häufig kein Hinweis auf die Frage nach der Allgemeingültigkeit zu erkennen. Im Vergleich zu den Beweisversuchen in der Gruppe IntzzA1 gibt es aber keinen Anhaltspunkt für Zweifel an einer korrekten Interpretation.

Bearbeitung 8.3.2b: Klassenstufe 5, weiblich (Code: 0521725)

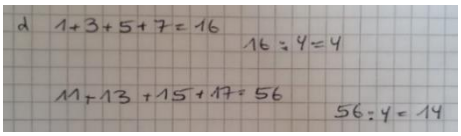


Abbildung 144: Bearbeitung 8.3.2b (IntzzA2 *Passung der Ausführungen zu der zz. Aussage bei teilweise unklarer Interpretation*)

Die in die Gruppe **IntzzA1** einzuordnenden Beweisversuche, wie etwa die Bearbeitung 8.3.2c in Abbildung 145, enthalten einen Hinweis auf eine nicht ganz korrekte Interpretation der zu zeigenden Aussage oder einen inhaltlichen Fehler. Sie lassen vermuten, dass für deren Weiterentwicklung zu einem Beweis zunächst ein fehlerhaftes Beispiel oder die Interpretation der Begrifflichkeiten der zu zeigenden Aussage zu korrigieren wäre. Dies trifft in jeder der Klassenstufen 5 bis 8 bei 35% bis 55%, d. h. einem Drittel bis gut der Hälfte der Beweisversuche mit prüfendem Ansatz zu.

Bearbeitung 8.3.2c: Klassenstufe 5, weiblich (Code: 0522133)

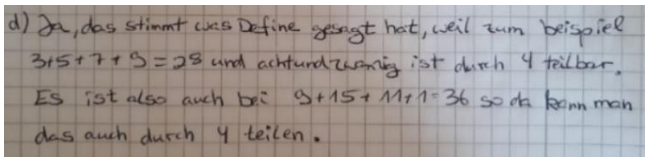


Abbildung 145: Bearbeitung 8.3.2c (IntzzA1 *Hinweis auf eine nicht ganz korrekte Interpretation der zz. Aussage oder einen inhaltlichen Fehler*)

Insgesamt besteht in allen Klassenstufen bei über 70% der Beweisversuche mit prüfendem Ansatz hinsichtlich der Interpretation der zu zeigenden Aussage zumindest Interpretationsspielraum (IntzzA1-2). Dies deutet an, dass im Falle von Beweisversuchen mit prüfendem Ansatz oftmals eine Klärung der Begrifflichkeiten oder der Allgemeingültigkeit der zu zeigenden Aussage für eine Weiterentwicklung der Beweisversuche wichtig ist.

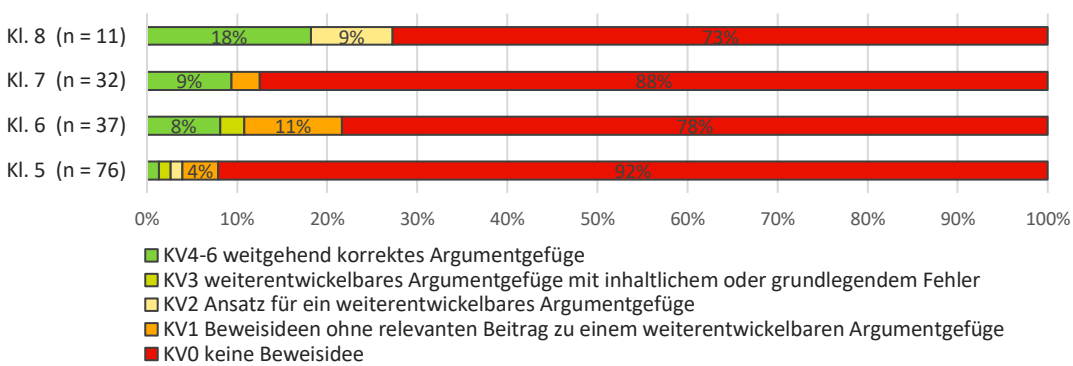
8.3.2.2 Entwicklung eines korrekten und vollständigen Argumentgefüges

In Beweisversuchen mit prüfendem Ansatz kann an Beispielen ein weiterentwickelbares Argumentgefüge ausgeführt sein, und, auch wenn kein weiterentwickelbares Argumentgefüge vorkommt, lassen sich teils sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges finden (Abschn. 3.3.7). Daher beschäftigt sich dieser Abschnitt mit Eigenschaften der Beweisversuche mit prüfendem Ansatz, die Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges sein können. Zu diesen Eigenschaften zählen die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges, die Systematik angegebener Beispiele, die Existenz und Art eines besonderen Beispiels, die Existenz eines vollständig und weitgehend korrekt ausgeführten Beispiels, die Existenz eines großen Beispiels und die Anzahl an Beispielen. Die genannten Eigenschaften werden im Folgenden zuerst einzeln betrachtet und danach miteinander kombiniert.

Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges

In diesem Abschnitt wird beschrieben, inwieweit in den Beweisversuchen mit prüfendem Ansatz aus Expertensicht ein weiterentwickelbares Argumentgefüges erkennbar ist, das an Beispielen ausgeführt wurde. Die Abbildung 146 zeigt die dazugehörigen relativen Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen.

In allen Klassenstufen 5 bis 8 ist in ca. 7% bis 27% der Beweisversuche mit prüfendem Ansatz aus Expertensicht eine Beweisidee zu erkennen (KV1-4). Insbesondere ist in jeder der Klassenstufen in mindestens einem der Beweisversuche ein weitgehend korrektes Argumentgefüge an Beispielen ausgeführt (KV4-6). Dabei ist der Anteil der Beweisversuche mit einem weitgehend korrekten Argumentgefüge an allen Beweisversuchen mit prüfendem Ansatz in der Klassenstufe 8 statistisch signifikant größer als in der Klassenstufe 5 (Abb. 146).



p-Werte von Signifikanztests und dazugehörige Effektstärken

	KV1-4 vs. KV0		KV4-6 vs. KV0-3
Kl. 8 vs. Kl. 5	$p = .083^{ex}$, $V = .211$	Kl. 8 vs. Kl. 5	$p = .041^{ex}$, $V = .307$

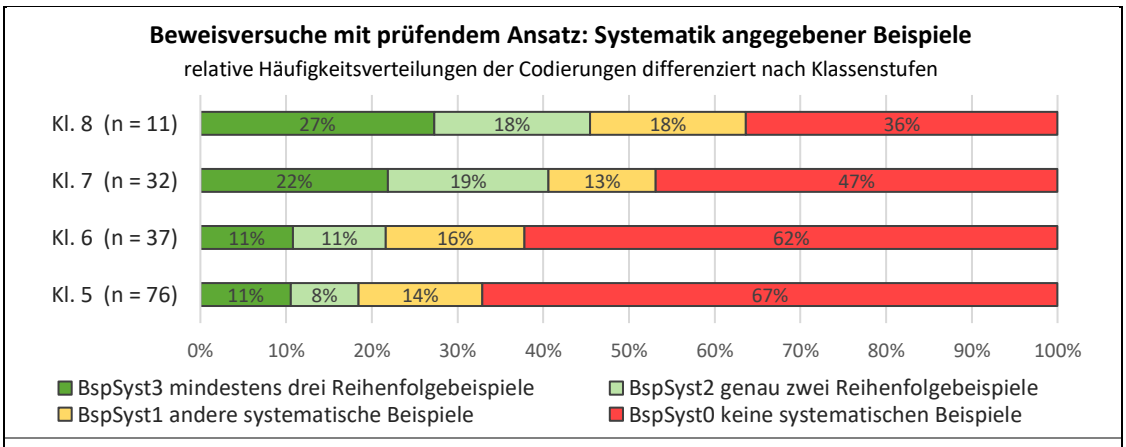
Abbildung 146: Beweisversuche mit prüfendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Alle Beweisversuche mit prüfendem Ansatz, in denen ein weitgehend korrektes Argumentgefüge zu erkennen ist (KV4-6), basieren auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Bei den meisten Beweisversuchen, in denen kein weitgehend korrektes Argumentgefüge, aber zumindest ein Ansatz für ein weiterentwickelbares Argumentgefüge zu erkennen ist (KV2-3), handelt es sich um die Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare*. Die Beweisversuche mit prüfendem Ansatz mit mindestens einer Beweisidee, aber keinem relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge (KV1) beruhen auf vielfältigen Beweisideen, u. a. den Ideen 9 *Parität von Summen ungerader oder gerader Zahlen* und 10 *Abstände ungerader Zahlen*.

Bei diesen Ergebnissen ist zu bedenken, dass die Beweisideen aus der Perspektive der Schülerinnen und Schüler latent sein können. Zugleich ist es vor allem bei Beweisversuchen, in denen Reihenfolgebeispiele ausgeführt wurden, gut möglich, dass einige der Teilnehmenden zum Zeitpunkt der Bearbeitung hätten begründen können, warum die Aussage gilt, dies aber nicht notiert haben. Letzteres wird durch die Analyse der Systematik der angegebenen Beispiele berücksichtigt.

Systematik angegebener Beispiele

Die Angabe systematischer Beispiele ist mit Blick auf die Entwicklung einer Beweisidee sowohl aus sachinhaltlicher als auch aus heuristischer Perspektive positiv hervorzuheben. Anhand systematischer Beispiele und vor allem anhand von Reihenfolgebeispielen lassen sich sehr gut Invarianten erkennen, die für den Beweis der zu zeigenden Aussage genutzt werden können. Zudem deutet die Existenz systematischer Beispiele systematisches Probieren an, was als Strategie bei der Entwicklung einer Beweisidee aufgegriffen werden kann. In diesem Abschnitt wird deshalb dargelegt, inwieweit die Beweisversuche mit prüfendem Ansatz systematische Beispiele enthalten. Die Abbildung 147 zeigt die relativen Häufigkeitsverteilungen der vorgenommenen Codierungen differenziert nach Klassenstufen.



p-Werte von Signifikanztests und dazugehörige Effektstärken

BspSyst2-3 vs. BspSyst0-1		BspSyst2-3 vs. BspSyst0-1		BspSyst1-3 vs. BspSyst0	
Kl. 7 vs. Kl. 5	$p = .027^{ex}, V = .234$	Kl. 8 vs. Kl. 5	$p = .057^{ex}, V = .217$	Kl. 7 vs. Kl. 5	$p = .050^{ex}, V = .189$
Kl. 7 vs. Kl. 6	$p = .089^{ex}, V = .206$	Kl. 8 vs. Kl. 6	$p = .140^{ex}, V = .225$	Kl. 8 vs. Kl. 5	$p = .090^{ex}, V = .212$

Abbildung 147: Beweisversuche mit prüfendem Ansatz: Systematik angegebener Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Reihenfolgebeispiele (BspSyst2-3) kommen in den Klassenstufen 5, 6 in ca. 20%, in den Klassenstufen 7, 8 in gut 40% der Beweisversuche mit prüfendem Ansatz vor. Der Unterschied zwischen den Klassenstufen 5, 7 ist statistisch signifikant (Abb. 147). In den Beweisversuchen mit prüfendem Ansatz und Reihenfolgebeispielen ist statistisch hoch signifikant häufiger ein weitgehend korrektes Argumentgefüge zu erkennen als in den anderen Beweisversuchen mit prüfendem Ansatz ($p = 0{,}009^{ex}$, $V = 0{,}232$), das Argumentgefüge basiert in allen analysierten Fällen auf der Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Zugleich sind in den Beweisversuchen mit prüfendem Ansatz, in denen die primäre Beweisidee 1 erkennbar ist, statistisch signifikant häufiger Reihenfolgebeispiele angegeben als in den anderen Beweisversuchen mit prüfendem Ansatz ($p = 0{,}019^{ex}$, $V = 0{,}206$). Zusätzlich zu den Beweisversuchen mit prüfendem Ansatz und Reihenfolgebeispielen gibt es in allen Klassenstufen solche, die andere systematische Beispiele enthalten (BspSyst1; Abb. 147). Der Gesamtanteil der Beweisversuche mit systematischen Beispielen (BspSyst1-3) an allen Beweisversuchen mit prüfendem Ansatz beträgt in der fünften und sechsten Klassenstufe etwa 35%, in der siebten 53% und in der achten 64% (Abb. 147).

Dabei ist zu berücksichtigen, dass systematische Beispiele nur in Beweisversuchen mit mindestens zwei Beispielen vorkommen können. Wenn man die Beweisversuche mit prüfendem Ansatz und mindestens zwei Beispielen fokussiert, liegt deren Anteil in allen Klassenstufen über 50%. Dies deutet an, dass in allen Klassenstufen die Mehrheit der Teilnehmenden, die in einem Beweisversuch mit prüfendem Ansatz mehr als ein Beispiel angegeben haben, systematisch vorgegangen sind.

Existenz und Art eines besonderen Beispiels

Zusätzlich zu systematischen Beispielen können auch einzelne besondere Beispiele die Erarbeitung eines korrekten und vollständigen Argumentgefüges unterstützen. Ein solches Beispiel ist bei der fokussierten Aussage das kleinste Beispiel. Die Betrachtung des kleinsten Beispiels kann für einen Beweis der zu zeigenden Aussage z. B. als Induktionsanfang erforderlich sein. Zudem sind noch andere Beispiele potenziell besonders hilfreich, weil etwa die Betrachtung der Summe von vier gleichen ungeraden Zahlen zur Entwicklung einer ausbaubaren Beweisidee wesentlich beitragen kann. Daher fokussiert dieser Abschnitt auf die Frage, ob in den Beweisversuchen das kleinste Beispiel oder ein anderes besonderes Beispiel vorkommen.

Die Analyseergebnisse zeigen, dass je Klassenstufe 5 bis 8 ein bis zwei Drittel der Beweisversuche mit prüfendem Ansatz das kleinste Beispiel enthalten (Abb. 148). Insbesondere kommt in fast allen Beweisversuchen mit prüfendem Ansatz und einem weitgehend korrekten Argumentgefüge das kleinste Beispiel vor. Von den Beweisversuchen mit prüfendem Ansatz und Reihenfolgebeispielen enthalten über 80% das kleinste Beispiel. Der Anteil der Beweisversuche mit dem kleinsten Beispiel ist bei den Beweisversuchen mit prüfendem Ansatz und Reihenfolgebeispielen statistisch hoch signifikant größer als bei den übrigen Beweisversuchen mit prüfendem Ansatz ($p < .001^{N-1}$, $V = .444$). Andersherum betrachtet enthalten von den Beweisversuchen mit prüfendem Ansatz und dem kleinsten Beispiel ca. die Hälfte Reihenfolgebeispiele und rund 10% ein weitgehend korrektes Argumentgefüge.

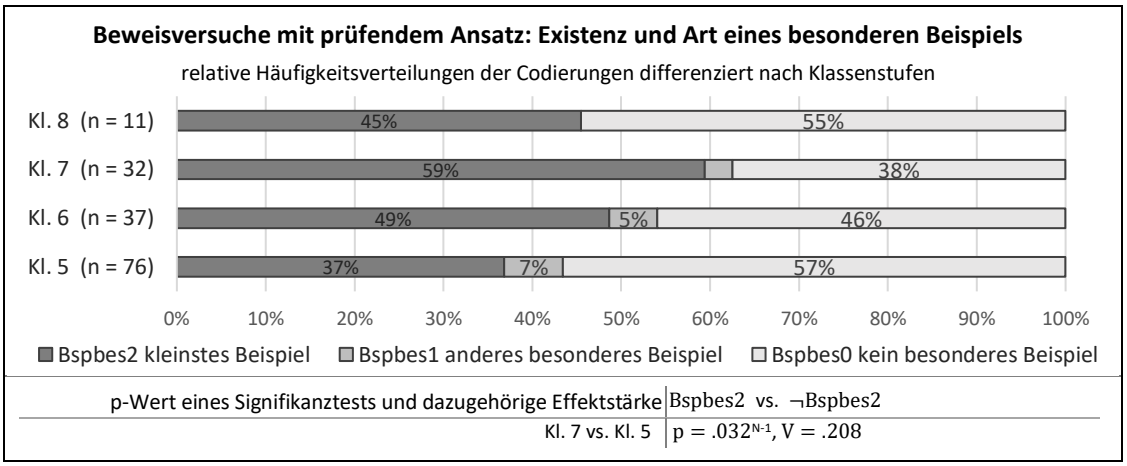


Abbildung 148: Beweisversuche mit prüfendem Ansatz: Existenz und Art eines besonderen Beispiels. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Wert eines Signifikanztests und dazugehörige Effektstärke

Existenz vollständig und weitgehend korrekt ausgeführter Beispiele

Neben Reihenfolgebeispielen oder etwa dem kleinsten Beispiel kann prinzipiell jedes vollständig und weitgehend korrekt ausgeführte Beispiel zu der zu zeigenden Aussage das Finden einer Beweisidee unterstützen, da daran Zusammenhänge zwischen vier aufeinanderfolgenden ungeraden Zahlen und dem Ergebnis der Division der Summe dieser Zahlen durch vier (Kl. 5, 6) bzw. acht (Kl. 7, 8) erkannt werden können. Von den Beweisversuchen mit prüfendem Ansatz enthalten je Klassenstufe ein bis zwei Drittel ein vollständig und weitgehend korrekt ausgeführtes Beispiel zu der zu zeigenden Aussage. Insbesondere ist je Klassenstufe in mehr als der Hälfte der Beweisversuche mit prüfendem Ansatz, in denen ein Beispiel zu der zu zeigenden Aussage vorkommt, mindestens eines vollständig und weitgehend korrekt ausgeführt.

Anzahl an Beispielen und die Existenz eines großen Beispiels

Krumsdorf (2017) kommt in seiner Studie zu dem Ergebnis, dass nicht die Betrachtung möglichst vieler Beispiele, sondern die Betrachtung weniger und insbesondere auch nicht gut vorstellbarer Beispiele, wie etwa großer Beispiele, für die Entwicklung eines beispielgebundenen Beweises besonders günstig ist (Abschn. 5.2).

Die Ergebnisse der Analyse der **Anzahl an Beispielen** im Rahmen dieser Arbeit zeigen, dass je Klassenstufe zwischen 50% und 80% der Beweisversuche mit prüfendem Ansatz zwei bis fünf Beispiele enthalten. Mehr als fünf Beispiele wurden zumindest in den Klassenstufen 5 bis 7 selten angegeben, der Anteil der Beweisversuche mit mehr als fünf Beispielen an allen Beweisversuchen mit prüfendem Anteil liegt in diesen Klassenstufen jeweils bei 1% bis 5%.

Ein **großes Beispiel** ist in den Klassenstufen 5, 6 in gut 10% und in den Klassenstufen 7, 8 in ca. 20% der analysierten Beweisversuche mit prüfendem Ansatz zu finden. In den Klassenstufen 6 bis 8 handelt es sich dabei fast immer um Beweisversuche, in denen mindestens zwei Beispiele zu der zu zeigenden Aussage angegeben sind, während in der Klassenstufe 5 das große Beispiel meistens das einzige angegebene Beispiel darstellt. Zudem fällt auf, dass in keinem der analysierten Beweisversuche mit prüfendem Ansatz und einem großen Beispiel eine Beweisidee zu erkennen ist.

Zusammenfassende Betrachtung: Entwicklung eines korrekten und vollständigen Argumentgefüges

Um auf Grundlage der vorgestellten Ergebnisse zu Beweisversuchen mit prüfendem Ansatz zusammenfassende Aussagen hinsichtlich Anknüpfungspunkten für die Entwicklung eines korrekten und vollständigen Argumentgefüges zu treffen, werden mehrere der analysierten Eigenschaften miteinander kombiniert (Abb. 149). Dafür werden die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges zusammen mit primären Beweisideen, die Systematik angegebener Beispiele und die Angabe eines besonderen Beispiels ausgewählt. Die Existenz vollständig und weitgehend korrekt ausgeführter Beispiele wird nicht berücksichtigt, weil sie ebenfalls mit Blick auf die Interpretation der zu zeigenden Aussage relevant ist und somit diesbezüglich miteinbezogen wurde (Abschn. 3.3.7, 8.3.2.1). Die Gesamtanzahl an Beispielen bleibt ebenfalls unberücksichtigt, weil weniger die Anzahl an Beispielen als die Art der Beispiele für die Entwicklung eines korrekten und vollständigen Argumentgefüges relevant erscheint. Die Existenz eines großen Beispiels wird nicht berücksichtigt, weil große Beispiele nicht unbedingt ein besonders guter Anknüpfungspunkt für die Entwicklung eines korrekten und vollständigen Argumentgefüges sind, zumal in keinem der analysierten Beweisversuche mit prüfendem Ansatz und einem großen Beispiel eine Beweisidee zu erkennen ist.

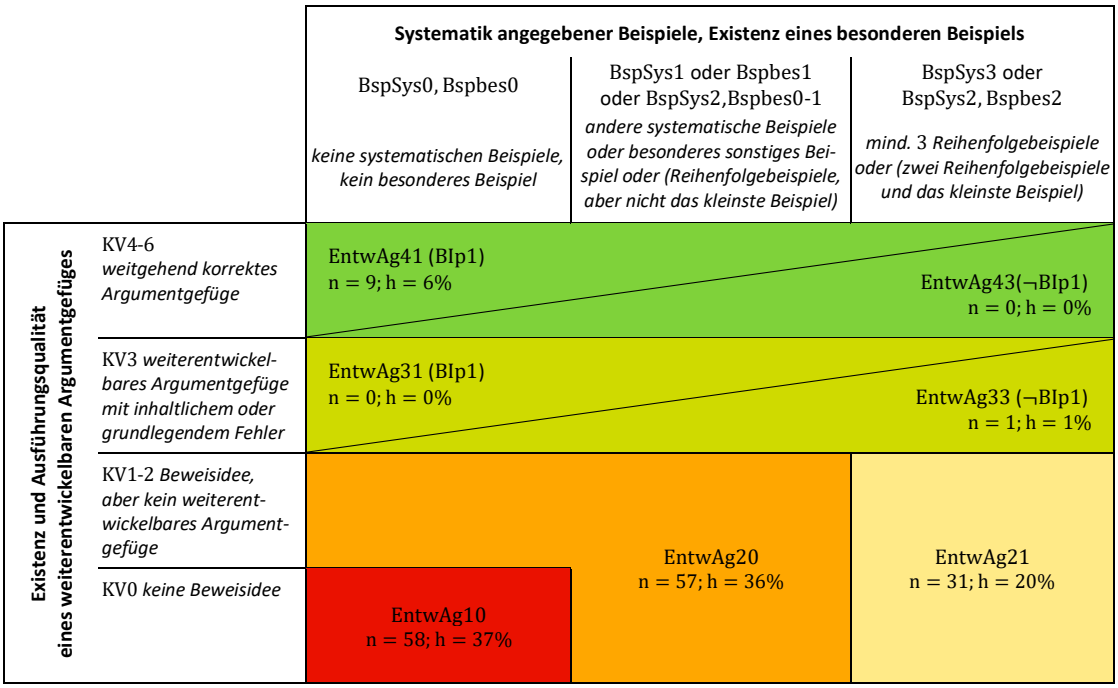


Abbildung 149: Gruppierung der Beweisversuche mit prüfendem Ansatz mit Blick auf die Entwicklung eines korrekten und vollständigen Argumentgefüges. Gebildete Gruppen und deren Häufigkeiten

Wenn man die ausgewählten Eigenschaften von Beweisversuchen mit prüfendem Ansatz entsprechend Abbildung 149 miteinander kombiniert und gruppiert, ergeben sich sieben Gruppen. Diese Gruppen werden im Folgenden beschrieben, begründet und veranschaulicht.

Beweisversuche mit prüfendem Ansatz in der Gruppe **EntwAg41** enthalten ein weitgehend korrektes Argumentgefüge basierend auf der primären Beweisidee 1 *sukzessive Vergrößerung*

einer Summe mit Erkennen einer Invariante. In der Mehrheit dieser Beweisversuche sind Reihenfolgebeispiele und das kleinste Beispiel ausgeführt. In ca. der Hälfte der Fälle ist die Darstellungsform ikonisch-numerisch-verbal, in den anderen numerisch-verbal. Beweisversuche in der Gruppe EntwAg41 erscheinen folglich in mehrerlei Hinsicht als homogen, sodass mit der Erkennbarkeit eines weitgehend korrekten Argumentgefüges basierend auf der primären Beweisidee 1 gut eingrenzbare Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges einhergehen. Beispielsweise wird in der Bearbeitung 8.3.2d in Abbildung 150 über die Vergrößerung der Summe um acht argumentiert. Ähnlich wie bei vielen Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 wäre für ein korrektes und vollständiges Argumentgefüge vor allem diese Invariante zu begründen. Zudem wäre für einen Beweis der zu zeigenden Aussage eine Verallgemeinerung des Argumentgefüges erforderlich, wobei sich ausgehend von den schriftlichen Ausführungen nicht ganz sicher sagen lässt, inwieweit der Schüler die Verallgemeinerbarkeit bereits kognitiv realisiert hat.

Bearbeitung 8.3.2d: Klassenstufe 6, männlich (Code: 0611225)

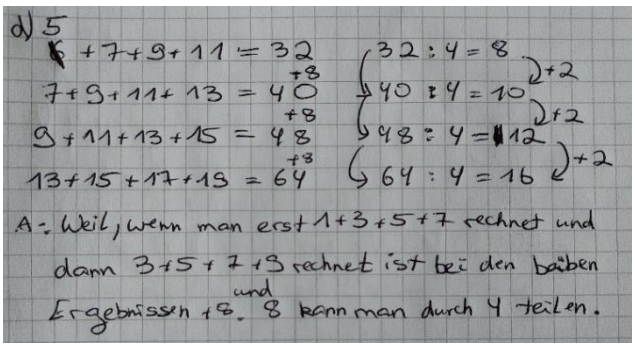


Abbildung 150: Bearbeitung 8.3.2d (EntwAg41 *weitgehend korrektes Argumentgefüge basierend auf der primären Beweisidee* 1)

Weiterhin ist es möglich, dass in Beweisversuchen mit prüfendem Ansatz ein weiterentwickelbares Argumentgefüge basierend auf der primären Beweisidee 1 erkennbar ist, jedoch ein inhaltlicher oder grundlegender Fehler vorkommt, was durch die Bildung der Gruppe **EntwAg31** berücksichtigt wird. Beispielsweise kann es sein, dass Lernende Summen von vier gleichen Zahlen betrachten und behaupten, die Aussage sei korrekt, weil sich die Summe immer um vier vergrößert.

In die Gruppen **EntwAg43** und **EntwAg33** werden Beweisversuche mit prüfendem Ansatz eingeordnet, in denen ein weitgehend korrektes Argumentgefüge bzw. ein weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler basierend auf einer anderen primären Beweisidee als der Beweisidee 1 zu erkennen ist. Auch auf Grundlage jeder der Beweisidee 2 bis 8 ist ein Beweisversuche mit prüfendem Ansatz und einem weiterentwickelbaren Argumentgefüge möglich (Abschn. 6.2.2). Die Abbildung 151 zeigt exemplarisch einen Beweisversuch mit prüfendem Ansatz, der in die Gruppe EntwAg33 gehört. Er basierend auf der primären Beweisidee 6 *Teilbarkeit der Summe zweier Zahlenpaare* und enthält ein weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler. In der Bearbeitung wird von der Teilbarkeit zweier Zahlen durch zwei auf die Teilbarkeit durch vier geschlossen, was ein inhaltlicher Fehler ist. Zudem lässt der Bezug auf nicht aufeinanderfolgende ungerade

Zahlen vermuten, dass der Ausdruck „aufeinanderfolgende ungerade Zahlen“ nicht ganz korrekt interpretiert oder nicht vollständig berücksichtigt wurde.

Bearbeitung 8.3.2e: Klassenstufe 5, männlich (Code: 0512220)

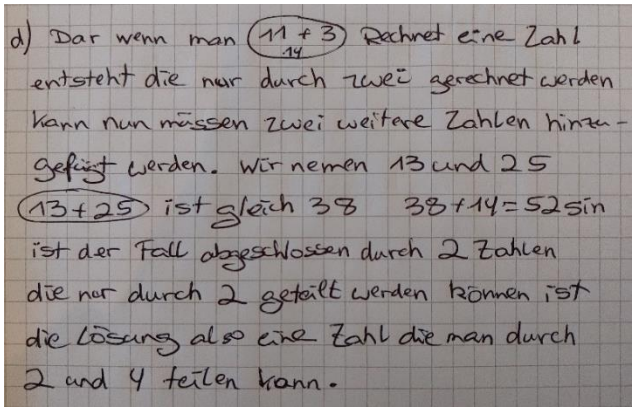


Abbildung 151: Bearbeitung 8.3.2e (EntwAg33 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler basierend auf einer anderen als der primären Beweisidee* 1)

Neben Beweisversuchen mit prüfendem Ansatz und einem erkennbaren weiterentwickelbaren Argumentgefüge wird mit der Gruppe **EntwAg21** berücksichtigt, dass es auch Beweisversuche mit prüfendem Ansatz gibt, in denen zwar kein weiterentwickelbares Argumentgefüge vorkommt, aber Reihenfolgebeispiele und das kleinste Beispiel oder mindestens drei Reihenfolgebeispiele angegeben sind. In diesen Beweisversuchen, wie etwa der Bearbeitung 8.3.2f in Abbildung 152, ist ein weiterentwickelbares Argumentgefüge basierend auf der primären Beweisidee 1 angelegt und eventuell von dem Lernenden bereits subjektiv realisiert worden. Schließlich lässt sich insbesondere die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* sehr gut anhand von Reihenfolgebeispielen erkennen.

Bearbeitung 8.3.2f: Klassenstufe 6, männlich (Code: 0621813)

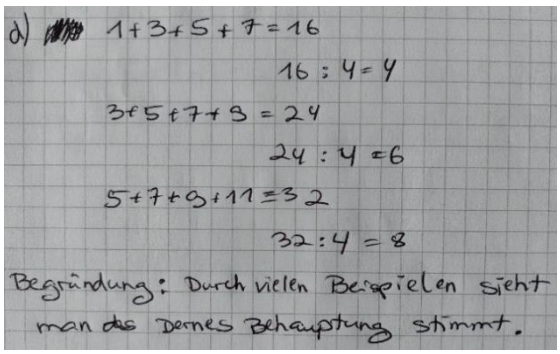


Abbildung 152: Bearbeitung 8.3.2f (EntwAg21 *Anlage eines weiterentwickelbaren Argumentgefüges*)

Alle Beweisversuche mit prüfendem Ansatz in einer der Gruppen EntwAg21,31,32,41,42 haben gemeinsam, dass aus Expertensicht ein weiterentwickelbares Argumentgefüges erkennbar oder angelegt ist. Aber auch wenn dies nicht so klar der Fall ist, können in Beweisversuchen mit prüfendem Ansatz sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges vorkommen, was durch die Bildung der Gruppe **EntwAg20** berücksichtigt wird (Abb. 149). Beispielsweise ist in der Bearbeitung

8.3.2g in Abbildung 153 die Überlegung, dass der Abstand zwischen zwei aufeinanderfolgenden Summanden zwei beträgt, hervorgehoben. In der Bearbeitung 8.3.2h in Abbildung 154 kommt das kleinste Beispiel vor, zudem lassen sich die zwei Beispiele als systematische Beispiele bezeichnen. So wie bei diesen Bearbeitungen ist in allen Beweisversuchen in der Gruppe EntwAg20 zwar kein weiterentwickelbares Argumentgefüge, aber dennoch eine besondere sachinhaltliche Überlegung, ein besonderes Beispiel oder eine besondere Vorgehensweise erkennbar. Diese Überlegungen, Beispiele und Vorgehensweisen stellen mögliche Anknüpfungspunkte für die Entwicklung einer ausbaubaren Beweisidee und schließlich eines korrekten, vollständigen Argumentgefüges dar.

Bearbeitung 8.3.2g: Klassenstufe 6, weiblich (Code: 0611706)

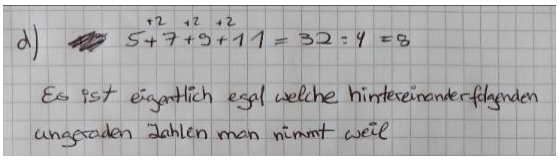


Abbildung 153: Bearbeitungen 8.3.2g (EntwAg20 *sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges*)

Bearbeitung 8.3.2h: Klassenstufe 5, weiblich (Code: 0521725)

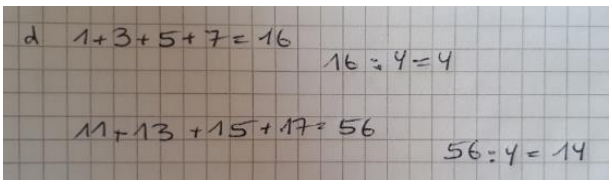


Abbildung 154: Bearbeitungen 8.3.2h (EntwAg20 *sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges*)

Des Weiteren gibt es Beweisversuche mit prüfendem Ansatz ohne ein weiterentwickelbares Argumentgefüge und ohne besondere Anknüpfungspunkte für die Entwicklung einer ausbaubaren Beweisidee. Beispielsweise kommen in der Bearbeitung 8.3.i in Abbildung 155 keine Beweisidee, kein besonderes Beispiel und keine systematischen Beispiele vor. Solche Beweisversuche werden entsprechend Abbildung 149 in der Gruppe **EntwAg10** zusammengefasst.[154]

Bearbeitung 8.3.2i: Klassenstufe 5, männlich (Code: 0510212)

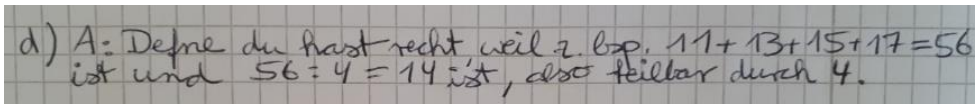


Abbildung 155: Bearbeitung 8.3.2i (EntwAg10 *keine besonderen Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges*)

[154] Insbesondere mit Blick auf Beweisversuche in der Gruppe EntwAg10 ist anzumerken, dass ausschließlich codiert wurde, was aus Expertensicht zu erkennen ist. Es ist möglich, dass Lernende sich eine Beweisidee gedacht oder eine Invariante erkannt haben oder aus Ihrer Sicht systematisch vorgegangen sind, dies jedoch aus den schriftlichen Beweisversuchen nicht hervorgeht. Zudem kann, so wie in der Bearbeitung 8.3.2i, durchaus ein vollständig und korrekt ausgeführtes Beispiel zu der zu zeigenden Aussage vorkommen, was – wie bereits erläutert – für die Entwicklung einer ausbaubaren Beweisidee genutzt werden könnte. Deshalb wird als Gruppennummer nicht EntwAg00, sondern EntwAg10 gewählt und im Gruppennamen der vorsichtige Ausdruck „keine besonderen Anknüpfungspunkte“ verwendet.

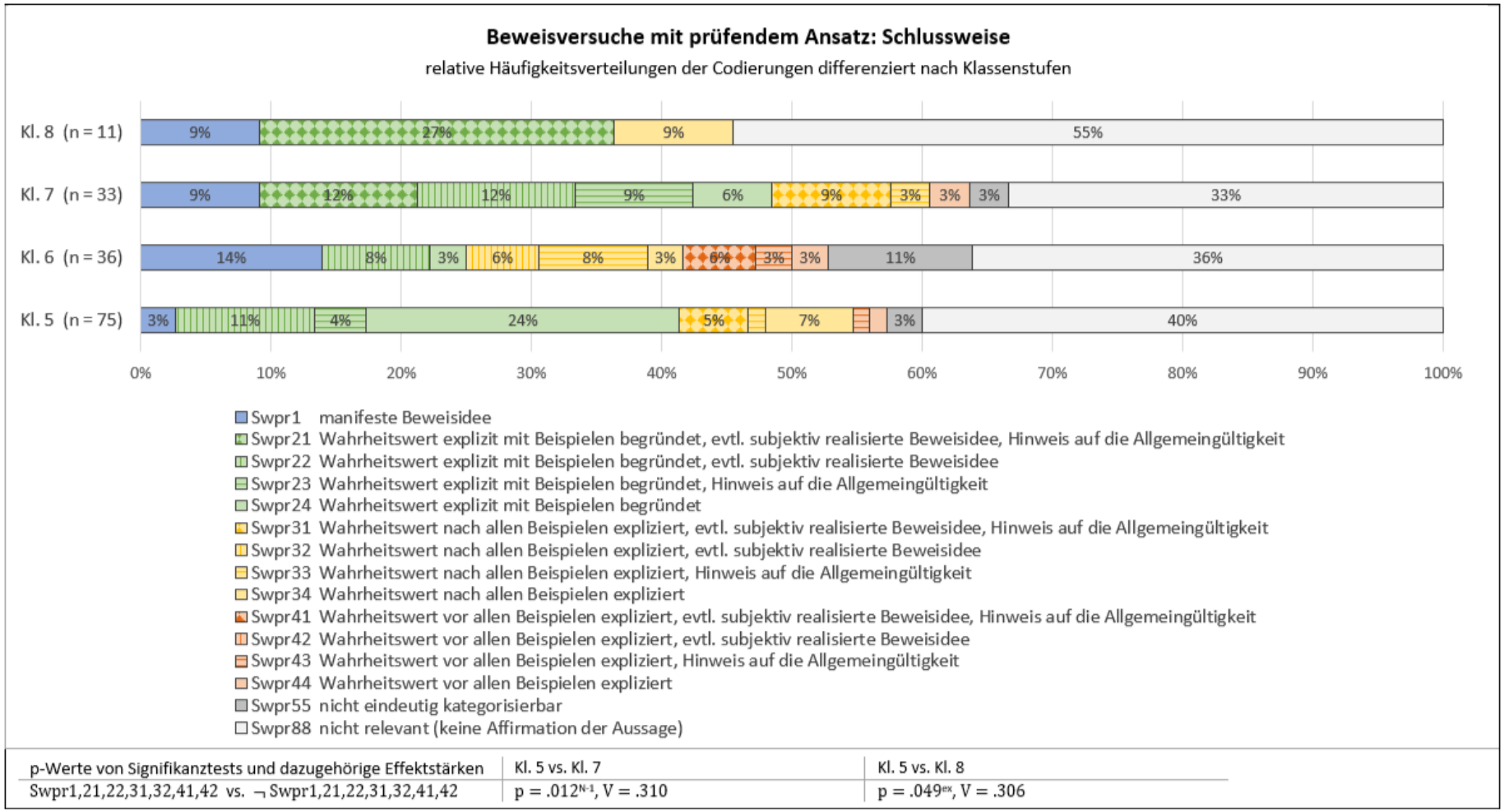
Was die **relativen Häufigkeitsverteilungen** der Zuordnungen der Beweisversuche mit prüfendem Ansatz zu den gebildeten Gruppen betrifft, lassen sich folgende Auffälligkeiten festhalten: In jeder der Klassenstufen gibt es Beweisversuche mit prüfendem Ansatz und einem weitgehend korrekten Argumentgefüge basierend auf der primären Beweisidee 1 (EntwAg41). In die anderen Gruppen mit einem weiterentwickelbaren Argumentgefüge (EntwAg31,33,43) gehört (fast) keiner der analysierten Beweisversuchen mit prüfendem Ansatz. Die obige Bearbeitung 8.3.2e in Abbildung 151 bildet diesbezüglich die einzige Ausnahme. Insgesamt ist in jeder der Klassenstufen 5 bis 8 in einzelnen und bis zu 20% aller Beweisversuche mit prüfendem Ansatz bereits ein weiterentwickelbares Argumentgefüge vorhanden. Außerdem zeigen die relativen Häufigkeitsverteilungen, dass in allen Klassenstufen 5 bis 8 in ca. 55% bis 80% und somit in der Mehrheit der Beweisversuche mit prüfendem Ansatz zumindest sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges zu erkennen sind (EntwAg20-43). Diese Anknüpfungspunkte, wie z. B. ein besonderes Beispiel, eine besondere Vorgehensweise, eine besondere sachinhaltliche Überlegung oder auch ein erkennbares Argumentgefüge, können für die Weiterentwicklung der Beweisversuche aufgegriffen werden.

8.3.2.3 Schlussweise

In diesem Abschnitt steht die Schlussweise im Fokus. Auch die verwendete Schlussweise ist mit Blick auf die Weiterentwicklung eines Beweisversuchs mit prüfendem Ansatz zu einem Beweis von Bedeutung. Während beispielsweise im Falle eines rein induktiven Schlusses vielleicht erst einmal die Unzulänglichkeit dieser Schlussweise und die Notwendigkeit einer allgemeingültigen Begründung herauszuarbeiten wären, ist im Falle einer erkannten Invariante, die zu einem Schluss auf die Gültigkeit geführt hat, eventuell nur eine Versprachlichung eines gedachten Argumentgefüges erforderlich. Anhand der in rein schriftlicher Form vorliegenden Beweisversuche lassen sich jedoch ausschließlich Hinweise auf die Schlussweise beschreiben und anschließend deuten. Unter dem Begriff „Schlussweise“ wird deshalb in dieser Arbeit bei Beweisversuchen mit prüfendem Ansatz das Verhältnis angegebener Beispiele zur Angabe der Gültigkeit der Aussage verstanden und mithilfe von Kategorien beschrieben (Abschn. 3.3.8).

Die Abbildung 156 zeigt die relativen Häufigkeitsverteilungen der Zuordnungen der Beweisversuche zu den Kategorien, die zur Beschreibung der Schlussweise entwickelt wurden. Die Ergebnisse verdeutlichen die **Vielfalt** der Beweisversuche mit prüfendem Ansatz in Bezug auf die Schlussweise. Insbesondere in den Klassenstufen 5 bis 7 kommt eine große Anzahl unterschiedlicher Kategorien vor.

Innerhalb dieser Vielfalt ist eine Dominanz der Kategorie **Swpr2** ***Wahrheitswert explizit mit Beispielen begründet*** zu erkennen. In den Klassenstufen 5, 7, 8 sind in mehr als einem Viertel aller Beweisversuche mit prüfendem Ansatz Beispiele explizit als Grund für die Gültigkeit dargestellt. Dass Beispiele als Grund für die Allgemeingültigkeit der Aussage dargestellt sind, ohne dass ein Anhaltspunkt für eine eventuell subjektiv realisierte Beweisidee zu erkennen ist (Swpr23) und die Ausführungen folglich einen induktiven Schluss besonders nahelegen, trifft aber je Klassenstufe bei weniger als einem Zehntel aller Beweisversuche mit prüfendem Ansatz zu.



p-Werte von Signifikanztests und dazugehörige Effektstärken	Kl. 5 vs. Kl. 7	Kl. 5 vs. Kl. 8
Swpr1,21,22,31,32,41,42 vs. ¬ Swpr1,21,22,31,32,41,42	$p = .012^{N-1}$, V = .310	$p = .049^{ex}$, V = .306

Abbildung 156: Beweisversuche mit prüfendem Ansatz: Schlussweise. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

Weiterhin sieht man in Abbildung 156, dass die Beweisversuche mit **manifester Beweisidee (Swpr1)** in der Klassenstufe 5 unter 5% der Beweisversuche mit prüfendem Ansatz ausmachen, in den Klassenstufen 6 bis 8 jeweils knapp 10% bis 15%. Der Gesamtanteil an Beweisversuchen mit einer Affirmation der Aussage und einer manifesten oder eventuell subjektiv realisierten Beweisidee (Swpr1,21,22,31,32,41,42) beträgt in der Klassenstufe 5 ca. 20%, in der Klassenstufe 6 ca. 30% und in den Klassenstufen 7, 8 je 35% bis 40%. Er ist in den Klassenstufen 7, 8 jeweils statistisch signifikant größer als in der Klassenstufe 5 (Abb. 156).

Zusammenfassend deuten die Ergebnisse mit Blick auf die Weiterentwicklung der Beweisversuche mit prüfendem Ansatz an, dass bezüglich der Schlussweise vielfältige Ausgangslagen vorkommen. Häufig sind Beispiele explizit als Grund für die Gültigkeit dargestellt, wobei die Ausführungen aber meistens nicht klar auf einen induktiven Schluss hindeuten. In den Klassenstufen 7, 8 sind in mehr als einem Drittel der Beweisversuche mit prüfendem Ansatz eine Affirmation der Aussage und eine manifeste oder eventuell subjektiv realisierte Beweisidee zu erkennen.

8.3.2.4 Darstellungsmittel und Darstellungsformen

In diesem Abschnitt wird beschrieben, welche Darstellungsmittel in den Beweisversuchen mit prüfendem Ansatz vorkommen und was deren Darstellungsform ist. Dies gibt Hinweise darauf, inwieweit die Lernenden, die einen Beweisversuch mit prüfendem Ansatz abgegeben haben, mit der Nutzung der verschiedenen Darstellungsmittel vertraut sind.

Die Ergebnisse der Analyse der Darstellungsmittel in Beweisversuchen mit prüfendem Ansatz zeigen ein weitgehend homogenes Bild. In jedem der analysierten Beweisversuche mit prüfendem Ansatz kommen **numerische** Darstellungsmittel vor und der Anteil der Beweisversuche mit verbalen Darstellungsmitteln an allen Beweisversuchen mit prüfendem Ansatz liegt in jeder Klassenstufe zwischen 85% und 95%. **Ikonische** Darstellungsmittel sind in den Klassenstufen 5 bis 7 in ca. 10% bis 20%, in der Klassenstufe 8 in 36% der Beweisversuche mit prüfendem Ansatz zu finden. Dabei kommen keine figurierten Zahlen, aber alle anderen unterschiedenen Arten an ikonischen Darstellungsmitteln vor. **Symbolische** Darstellungsmittel wurden in keinem der analysierten Beweisversuche mit prüfendem Ansatz verwendet. **Strukturierende Begriffe** sind in jeder der Klassenstufen in mehr als einem Viertel der Beweisversuche mit prüfendem Ansatz zu finden. Oftmals sind die Bearbeitungen in die Angabe von Beispielen oder „Rechnungen“ und die Angabe einer Antwort gegliedert. In einzelnen Beweisversuchen mit prüfendem Ansatz sind Beispiele als „Beweis“ oder als „Begründung“ ausgewiesen.

Die **Darstellungsform** ist je Klassenstufe bei ca. 70% bis 85% der Beweisversuche mit prüfendem Ansatz als numerisch-verbal zu bezeichnen und bei ca. 10% bis 15% als numerisch. In den Klassenstufen 5, 6 kommen auch Beweisversuche mit der Darstellungsform *verbal* vor, deren relativer Anteil an allen Beweisversuchen mit prüfendem Ansatz liegt bei 1% bis 6%. In allen Klassenstufen gibt es Beweisversuche mit prüfendem Ansatz und der Darstellungsform ikonisch-numerisch-verbal. Deren relativer Anteil an allen Beweisversuchen mit prüfendem Ansatz ist in der Klassenstufe 8 mit 18% am größten und statistisch signifikant größer als in der

Klassenstufe 5 mit 1% ($\mathrm{p} = .041^{\mathrm{ex}}$, $\mathrm{V} = .307$). In der Hälfte der Beweisversuche mit der Darstellungsform ikonisch-numerisch-verbal ist die primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* zu erkennen. Bei den Beweisversuchen mit prüfendem Ansatz und der primären Beweisidee 1 ist der Anteil der Darstellungsform ikonisch-numerisch-verbal mit 40% statistisch hoch signifikant größer als bei den anderen Beweisversuchen mit prüfendem Ansatz ($\mathrm{p} = 0{,}001^{\mathrm{ex}}$, $\mathrm{V} = 0{,}414$).

Mit Blick auf Anknüpfungspunkte für eine **Weiterentwicklung der Beweisversuche** mit prüfendem Ansatz zu einem Beweis lassen sich folgende Tendenzen festhalten: Beweisversuche mit prüfendem Ansatz sind in der Regel numerisch-verbal dargestellt. Symbolische Darstellungsmittel kommen nicht vor. Ikonische Darstellungsmittel sind selten, am ehesten in Beweisversuchen mit prüfendem Ansatz aus der Klassenstufe 8 und fast nie in solchen aus der Klassenstufe 5 zu finden. Die Beweisversuche mit prüfendem Ansatz sind teils explizit in die Angabe von Beispielen und einer Antwort gegliedert, was für einen Beweis der zu zeigenden Aussage anzupassen wäre.

8.3.3 Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit

Aufbauend auf der Betrachtung der Passung eines Beispiels zu der zu zeigenden Aussage und vieler weiterer Eigenschaften der Beweisversuche mit prüfendem Ansatz (Abschn. 8.3.1, 8.3.2) werden die Beweisversuche mit prüfendem Ansatz im vorliegenden Abschnitt 8.3.3 mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage und im nachfolgenden Abschnitt 8.3.4 in Hinblick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis klassifiziert. Zudem werden die relativen Häufigkeiten der dabei gebildeten Gruppen präsentiert.

8.3.3.1 Einführung und Begründung der Klassifikation

Mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage enthalten Beweisversuche mit prüfendem Ansatz kein korrektes und vollständiges, deduktives Argumentgefüge. Es kann aber ein an Beispielen ausgeführtes, korrektes und vollständiges Argumentgefüge vorhanden sein. Zudem ist es möglich, dass ein Beispiel angegeben ist, welches die zu zeigende Aussage für einen Spezialfall bestätigt, oder dass wenigstens angegeben ist, dass diese Aussage gilt. Die angesprochenen drei Merkmale, die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges, die Existenz und Passung eines Beispiels zu der zu zeigenden Aussage und die Existenz und Art eines angegebenen Wahrheitswerts, werden für die Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit kombiniert.

Die Abbildung 157 zeigt die Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Entsprechend den Ausführungen in Abschnitt 8.3.2.2 enthalten einige der analysierten Beweisversuche mit prüfendem Ansatz ein weitgehend korrektes Argumentgefüge oder ein weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler. Diese Beweisversuche bewegen sich an der Grenze zu solchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge. Zudem ist in fast allen analysierten Fällen mindestens ein eigenes Beispiel zu der zu zeigenden Aussage

Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage
NwG2.4: KV4-6 *weitgehend korrektes Argumentgefüge*
NwG2.3: KV3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler*
NwG2.2: KV0-2, PBsp5 *korrekt und vollständig ausgeführtes eigenes Beispiel zu der zz. Aussage*
NwG2.1: KV0-2, PBsp3-4 *Beispiel zu der zz. Aussage, aber kein korrekt und vollständig ausgeführtes eigenes Beispiel*
NwG2.0.2: KV0-2, PBsp1-2, Wwert1 *kein Beispiel zu der zz. Aussage, aber Angabe der Gültigkeit*
NwG2.0.1: KV0-2, PBsp1-2, Wwert0,55,99 *kein Beispiel zu der zz. Aussage und keine Angabe der Gültigkeit*

Abbildung 157: Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage

angegeben. Daher erscheint es angemessen, die Beweisversuche mit prüfendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage als besonders hochwertig zu deklarieren. Da es mit Blick auf den Nachweis der Gültigkeit einen qualitativen Unterschied macht, inwieweit das Argumentgefüge korrekt ist, wird in der Klassifikation zwischen den Gruppen NwG2.4 *weitgehend korrektes Argumentgefüge* und NwG2.3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler* differenziert.[155]

Die anderen Gruppen NwG2.0-2.2 umfassen Beweisversuche mit prüfendem Ansatz ohne weiterentwickelbares Argumentgefüge, eventuell ist ein bestätigendes Beispiel angeführt. Dabei ist es mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage von Bedeutung, ob ein korrekt und vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage angegeben ist, ein Beispiel zu der zeigenden Aussage vorkommt, aber kein korrekt und vollständig ausgeführtes eigenes Beispiel oder ausschließlich sonstige Beispiele angegeben sind. Dementsprechend werden in der Klassifikation die Gruppen NwG2.2 *korrekt und vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage*, NwG2.1 *Beispiel zu der zu zeigenden Aussage, aber kein korrekt und vollständig ausgeführtes eigenes Beispiel* und NwG2.0 *kein Beispiel zu der zu zeigenden Aussage* unterschieden.

Vor allem wenn kein Beispiel zu der zu zeigenden Aussage vorkommt, d. h. die Beispiele vordergründig nichts zum Nachweis der Gültigkeit der zu zeigenden Aussage beitragen, stellt sich die Frage, ob wenigstens die Gültigkeit dieser Aussage entsprechend der Aufgabenstellung expliziert ist. Daher wird bei den Beweisversuchen mit prüfendem Ansatz, aber keinem Beispiel zu der zu zeigenden Aussage zwischen denen mit und denen ohne Affirmation der Aussage differenziert (NwG2.0.2 vs NwG2.0.1). Somit ergeben sich insgesamt sechs ordinal skalierte Gruppen NwG2.0.1 bis NwG2.4 (Abb. 157).

[155] In den Gruppenbezeichnungen, wie z. B. „NwG2.4“, gibt die linke Ziffer „2“ an, dass es sich um Beweisversuche aus der Oberkategorie Ans2 *prüfender Ansatz* handelt, während die rechte Ziffer, wie z. B. „4" bei „NwG2.4“ die Ausprägung dieser Beweisversuche mit Blick auf den Nachweis der Gültigkeit beschreibt.

8.3.3.2 Relative Häufigkeitsverteilungen der Beweisversuche

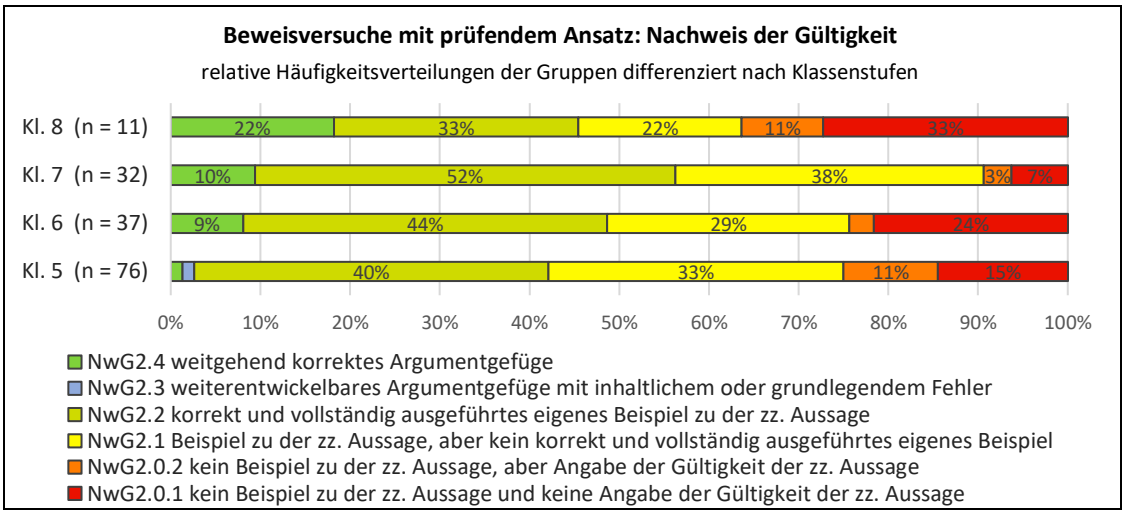


Abbildung 158: Beweisversuche mit prüfendem Ansatz: Nachweis der Gültigkeit. Relative Häufigkeitsverteilungen der Gruppen differenziert nach Klassenstufen

Die relativen Häufigkeitsverteilungen dieser Gruppen sind in den **Klassenstufen 5, 6** ähnlich (Abb. 158). Beweisversuche mit prüfendem Ansatz und einem weiterentwickelbaren Argumentgefüge (NwG2.3-2.4) kommen in beiden Klassenstufen vor und machen jeweils unter 10% aller Beweisversuche mit prüfendem Ansatz aus. Beweisversuche mit prüfendem Ansatz der Art NwG2.2 *korrekt und vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage* machen jeweils mit rund 40% den größten Anteil aus. In rund 30% der Beweisversuche mit prüfendem Ansatz aus den Klassenstufen 5, 6 ist ein Beispiel zu der zu zeigenden, aber kein korrekt und vollständig ausgeführtes eigenes Beispiel (NwG2.1.2) angegeben, in rund 25% ist kein Beispiel zu der zu zeigenden Aussage und meistens auch keine Angabe der Gültigkeit der Aussage zu finden (NwG2.0.1-2.0.2). Letzteres bedeutet, dass ein Viertel der Beweisversuche mit prüfendem Ansatz faktisch keinen oder fast keinen Beitrag zum Nachweis der Gültigkeit der zu zeigenden Aussage liefern.

Im Vergleich zu den Klassenstufen 5, 6 ist in der **Klassenstufe 7** der Anteil der Beweisversuche ohne weiterentwickelbares Argumentgefüge und ohne Beispiel zu der zu zeigenden Aussage (NwG2.0) an allen Beweisversuchen mit prüfendem Ansatz mit knapp 10% kleiner. In der **Klassenstufe 8** ist im Vergleich zu den Klassenstufen 5, 6 der Anteil der Beweisversuchen mit einem weitgehend korrekten Argumentgefüge (NwG2.4) größer. Allerdings ist keiner der in Abbildung 158 zu sehenden Unterschiede statistisch signifikant.

Insgesamt lässt sich also vor allem festhalten, dass in **jeder der Klassenstufen 5 bis 8** in etwa der Hälfte der Beweisversuche mit prüfendem Ansatz ein weiterentwickelbares Argumentgefüge oder ein korrekt, vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage vorkommt. In jeweils ca. 10% bis 35% der Beweisversuche mit prüfendem Ansatz ist kein einziges Beispiel zu der zu zeigenden Aussage, kein weiterentwickelbares Argumentgefüge und meistens auch keine Angabe der Gültigkeit der Aussage zu erkennen. Dies bedeutet, dass je Klassenstufe etwa ein Zehntel bis ein Drittel der Beweisversuche mit prüfendem Ansatz (fast) keinen Beitrag zum Nachweis der Gültigkeit der zu zeigenden Aussage enthalten.

8.3.3.3 Zusammenfassung

In der Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage werden sechs Gruppen NwG2.0.1 bis NwG2.4 differenziert. Beweisversuche in den Gruppen NwG2.3-2.4 charakterisiert die Existenz eines weiterentwickelbaren Argumentgefüges. In Beweisversuchen in der Gruppe NwG2.2 kommt kein weiterentwickelbares Argumentgefüge, aber ein korrekt und vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage vor. Beweisversuche in den Gruppen NwG2.0.1-2.0.2 enthalten kein weiterentwickelbares Argumentgefüge und kein Beispiel zu der zu zeigenden Aussage.

In jeder der Klassenstufen 5 bis 8 enthält etwa die Hälfte der Beweisversuche mit prüfendem Ansatz ein weiterentwickelbares Argumentgefüge oder ein korrekt und vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage. In jeder der Klassenstufen 5 bis 8 ist in etwa einem Zehntel bis zu einem Drittel der Beweisversuche mit prüfendem Ansatz (fast) kein Beitrag zum Nachweis der Gültigkeit der zu zeigenden Aussage zu erkennen.

8.3.4 Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf deren Weiterentwicklung zu einem Beweis

In diesem Abschnitt werden die Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis klassifiziert. Die Ergebnisse deuten unter anderem an, dass die Beweisversuche mit prüfendem Ansatz aus den Klassenstufen 5 und 6 sowie die aus den Klassenstufen 7 und 8 mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis ähnlich sind.

8.3.4.1 Einführung und Begründung der Klassifikation

In Abschnitt 8.3.2 wurden bereits die Ergebnisse der Analyse zahlreicher Eigenschaften der Beweisversuche mit prüfendem Ansatz präsentiert und hinsichtlich möglicher Anknüpfungspunkte für deren Weiterentwicklung interpretiert, so dass bei der Klassifikation direkt an diese Überlegungen angeknüpft werden kann. Die zum einen mit Blick auf die Interpretation der zu zeigenden Aussage und zum zweiten mit Blick auf die Entwicklung einer Beweisidee gebildeten Gruppen werden als relevante **Vergleichsdimensionen** miteinander kombiniert. Die Schlussweise, die vorkommenden Darstellungsmittel und die Darstellungsform werden in der Klassifikation nicht als Vergleichsdimensionen berücksichtigt. Die Schlussweise in Beweisversuchen mit prüfendem Ansatz und einer Affirmation der Aussage hängt mit den erkennbaren Anknüpfungspunkten für die Entwicklung eines korrekten und vollständigen Argumentgefüges zusammen, sodass es nicht sinnvoll ist, die Schlussweise als weitere Vergleichsdimension zu berücksichtigen. Bezüglich der Verwendung von Darstellungsmitteln sind die Beweisversuche mit prüfendem Ansatz weitgehend homogen, weshalb auch nicht die Existenz bestimmter Darstellungsmittel oder Darstellungsformen als Vergleichsdimensionen berücksichtigt werden. Die Ausprägungen der Beweisversuche mit prüfendem Ansatz bezüglich der Interpretation der zu zeigenden Aussage und bezüglich Anknüpfungspunkten für die Entwicklung eines korrekten und vollständigen Argumentgefüges bilden folglich die einzigen Vergleichsdimensionen.

Wie die beiden Vergleichsdimensionen verknüpft werden, ist in Abbildung 159, der **Klassifikation** der Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis, zu sehen. Die in Abbildung 159 dargestellten Gruppen unterscheiden sich von links nach rechts hinsichtlich der Existenz sachinhaltlicher oder heuristischer Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argument-

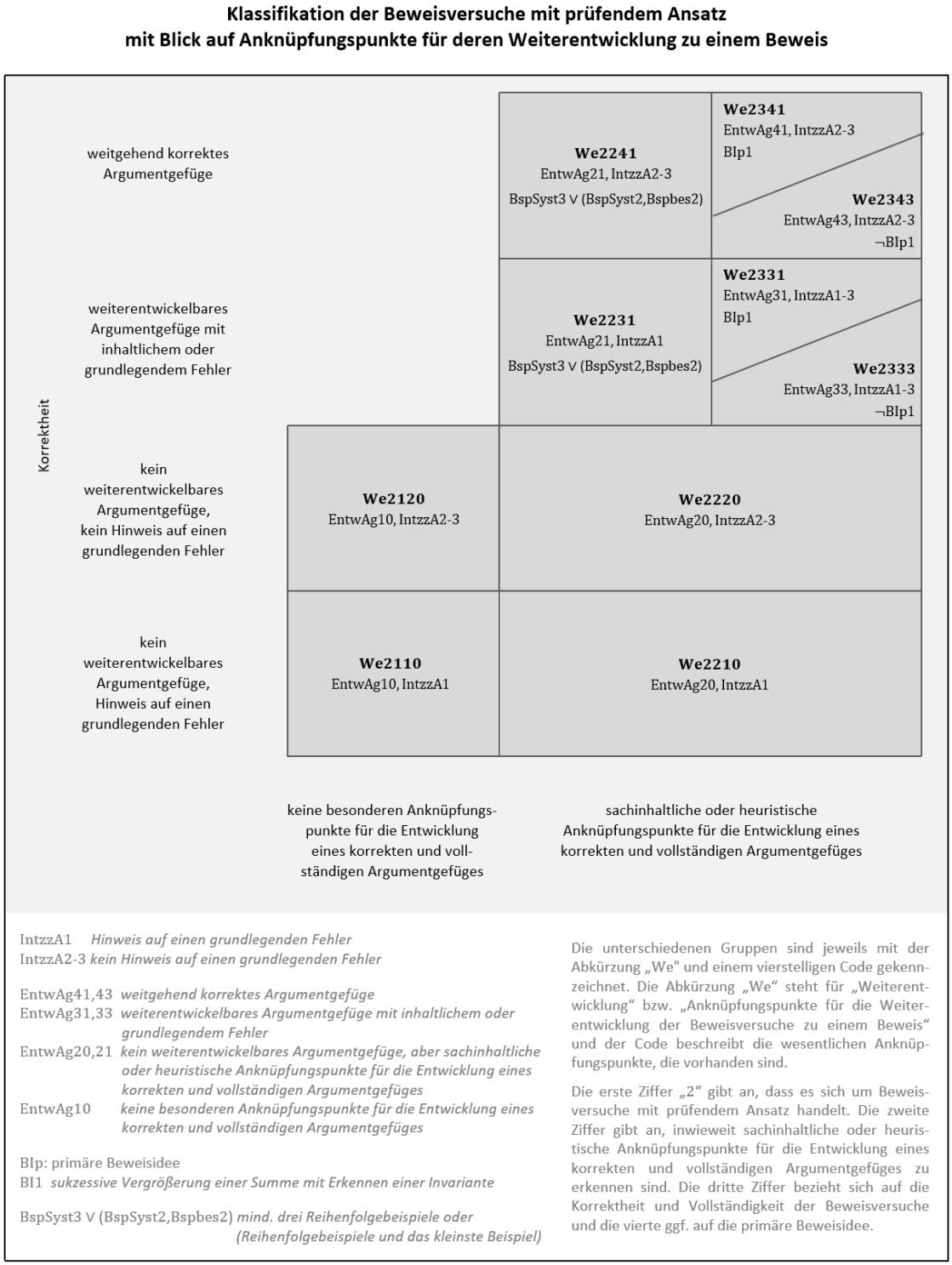


Abbildung 159: Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

gefüges. Von unten nach oben differieren die Gruppen hinsichtlich der Korrektheit der Beweisversuche, im Sinne der Passung der Ausführungen zu der zu zeigenden Aussage und der Existenz, Korrektheit eines weiterentwickelbaren Argumentgefüges. Mittels diagonaler Linien wird zwischen Beweisversuchen mit der primären Beweisidee 1 oder einer anderen primären Beweisidee unterschieden. Beispielsweise enthalten die Beweisversuche mit prüfendem Ansatz in der unten links angeordneten Gruppe We2110 kein weiterentwickelbares Argumentgefüge und es sind auch keine besonderen sachinhaltlichen oder heuristischn Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges zu erkennen. Die Beweisversuche in der Gruppe We2110 enthalten alle einen Hinweis auf einen grundlegenden Fehler, was darauf hindeutet, dass für die Entwicklung eines Beweises bei der Interpretation der zu zeigenden Aussage anzusetzen wäre. Bei den Beweisversuchen mit prüfendem Ansatz in den oben rechts angeordneten Gruppen We2341,2343 passen die Ausführungen zu der zu zeigenden Aussage. Die Beweisversuche in diesen Gruppen enthalten alle ein weitgehend korrektes Argumentgefüge, im Fall der Gruppe We2341 basierend auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, im Fall der Gruppe We2343 basierend auf einer anderen primären Beweisidee als der Idee 1.

Die Angabe „BspSyst3 v (BspSyst2,Bspbes2)“ bei den Gruppen We2231,2241 bedeutet, dass die Beweisversuche mindestens drei Reihenfolgebeispiele oder zwei Reihenfolgebeispiele und das kleinste Beispiel enthalten. In Beweisversuchen in diesen Gruppen ist kein weiterentwickelbares Argumentgefüge zu erkennen, aber aus Expertensicht angelegt und eventuell vom Lernenden bereits subjektiv realisiert worden.

So wie in der Klassifikation der Beweisversuche mit begründendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis (Abschn. 8.2.6.6) sind alle Gruppen mit der Abkürzung „We“ und einem vierstelligen Code gekennzeichnet. Die erste, linke Ziffer des Codes ist bei allen Gruppen die „2“ und hebt hervor, dass es sich um Beweisversuche in der Oberkategorie Ans2 *prüfender Ansatz* handelt. Die zweite Ziffer von links gibt an, inwieweit in den Beweisversuchen in den einzelnen Gruppen sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges zu erkennen sind. Die dritte Ziffer beschreibt die Korrektheit der Beweisversuche und die vierte Ziffer ggf. die primäre Beweisidee.

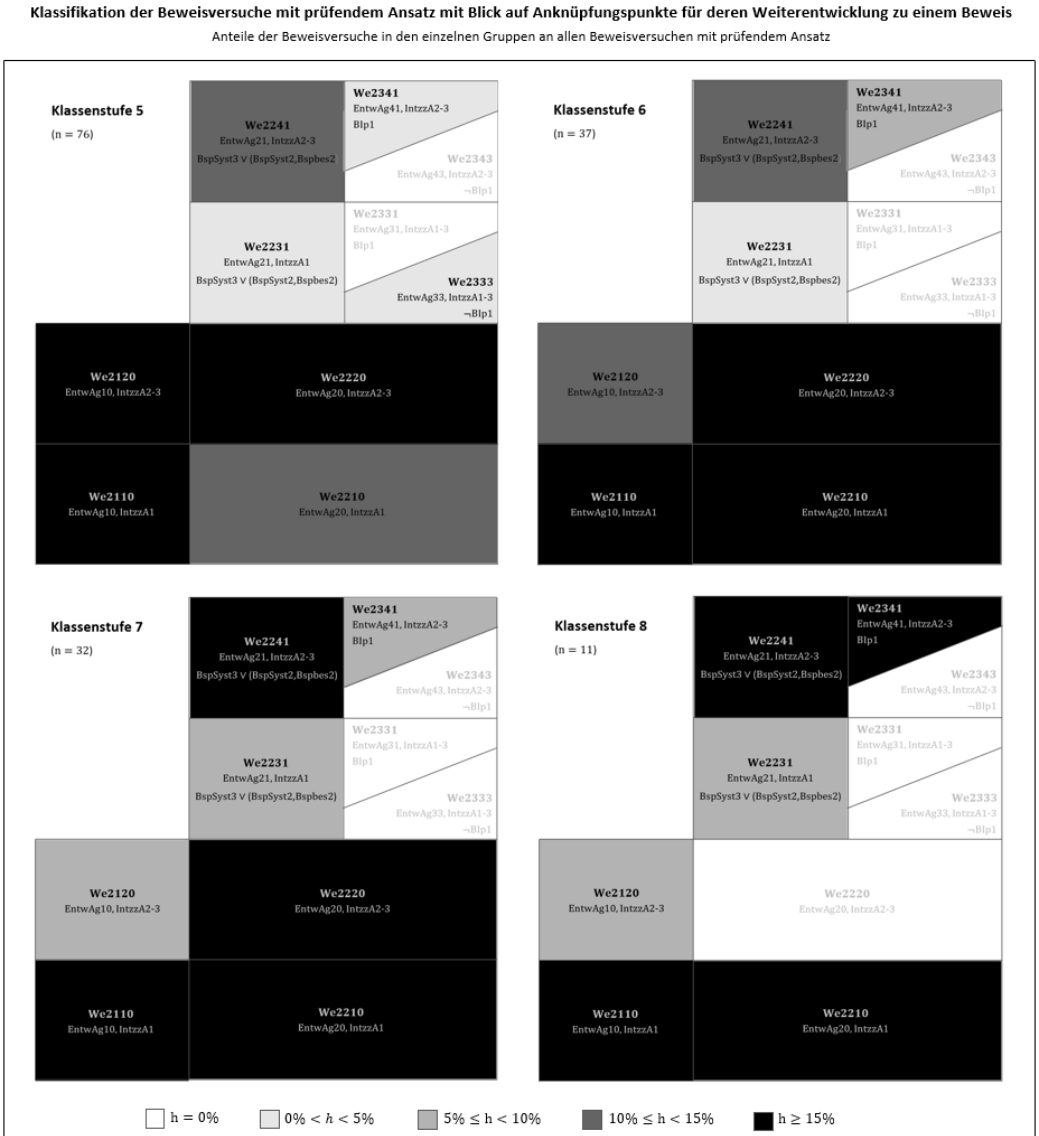
8.3.4.2 Relative Häufigkeitsverteilungen der Beweisversuche

Die relativen Häufigkeiten der unterschiedenen Gruppen gehen im Wesentlichen aus den Ausführungen in den Abschnitten 8.3.2.1, 8.3.2.2 und aus der Abbildung 160 hervor. Sie deuten an, dass sich die Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis in allen Klassenstufen recht ähnlich sind, und ganz besonders in den Klassenstufen 5, 6 sowie den Klassenstufen 7, 8 (Abb. 160).

Beispielsweise gibt es in jeder der Klassenstufen 5 bis 8 Beweisversuche mit prüfendem Ansatz und einem weitgehend korrekten Argumentgefüge (We2341,2343). In allen Klassenstufen sind in ca. 55% bis 80% der Beweisversuche mit prüfendem Ansatz sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argument-

gefüges zu erkennen (We2210,2220,2231,2241,2331, 2333,2341,2343). Je Klassenstufe machen Beweisversuche mit einem Hinweis auf einen grundlegenden Fehler und keinem besonderen Anknüpfungspunkt für die Entwicklung eines solchen Argumentgefüges (We2110) knapp 20% bis 30% der Beweisversuche mit prüfendem Ansatz aus.

Mit Blick auf die beiden Klassenstufen 5 und 6 sowie die Klassenstufen 7 und 8 ist kein statistisch signifikanter Unterschied zwischen den relativen Häufigkeitsverteilungen zu erkennen. In der Klassenstufe 7 ist vor allem der Anteil der Beweisversuche mit prüfendem Ansatz, in denen zwar kein weiterentwickelbares Argumentgefüge erkennbar, aber angelegt ist (We2231,2241), statistisch signifikant größer als in jeder der Klassenstufen 5, 6 (Abb. 160).



p-Werte von Signifikanztests und dazugehörige Effektstärken

	alle Gruppen	We2341 vs. ¬We2341	We2231,2241 vs. ¬We2231,2241	We2231-2343 vs. ¬We2231-2343
Kl. 7 vs. Kl. 5	$p = .066^{\chi^2}$, V = .350	$p = .077^{ex}$, V = .195	$p = .008^{N-1}$, V = .257	$p = .001^{N-1}$, V = .310
Kl. 7 vs. Kl. 6	$p = .350^{ex}$, V = .315	$p = 1.000^{ex}$, V = .022	$p = .022^{N-1}$, V = .278	$p = .028^{N-1}$, V = .267
Kl. 8 vs. Kl. 5	$p = .063^{ex}$, V = .378	$p = .041^{ex}$, V = .307	$p = .074^{ex}$, V = .116	$p = .045^{ex}$, V = .233
Kl. 8 vs. Kl. 6	$p = .321^{ex}$, V = .139	$p = .454^{ex}$, V = .312	$p = .361^{ex}$, V = .155	$p = .140^{ex}$, V = .225

Abbildung 160: Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den einzelnen Gruppen an allen Beweisversuchen mit prüfendem Ansatz sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

8.3.4.3 Zusammenfassung

Für die Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis werden zehn Gruppen gebildet, die hinsichtlich der Existenz sachinhaltlicher und heuristischer Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges, der Korrektheit der Beweisversuche und der primären Beweisidee differieren. Die Häufigkeitsverteilungen der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen mit prüfendem Ansatz deuten an, dass die Beweisversuche mit prüfendem Ansatz aus den Klassenstufen 5 und 6 sowie die aus den Klassenstufen 7 und 8 mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis ähnlich ausgeprägt sind. Dabei sind in jeder der Klassenstufen 5 bis 8 in mehr als der Hälfte der Beweisversuche mit prüfendem Ansatz sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges zu erkennen, die für die Weiterentwicklung der Beweisversuche aufgegriffen werden könnten. In jeder der Klassenstufen enthalten ein Drittel bis zu gut der Hälfte der Beweisversuche mit prüfendem Ansatz einen Hinweis auf einen grundlegenden Fehler und somit darauf, dass eventuell die Interpretation der zu zeigenden Aussage für eine Weiterentwicklung zu korrigieren wäre.

8.4 Beweisversuche mit weder begründendem noch prüfendem Ansatz

Dieser Abschnitt fokussiert auf Beweisversuche mit weder begründendem noch prüfendem Ansatz. Diese Beweisversuche deuten an, dass sich die Schülerin bzw. der Schüler mit der Aufgabenstellung beschäftigt hat, wobei weder eine Beweisidee noch ein Beispiel zu erkennen ist. Trotzdem können die Beweisversuche mit weder begründendem noch prüfendem Ansatz prinzipiell vielfältige Eigenschaften aufweisen (Abschn. 3.3.9, 3.3.11). Die folgenden Ausführungen geben Einblicke in die relativen Häufigkeiten verschiedener Eigenschaften der Beweisversuche mit weder begründendem noch prüfendem Ansatz und in mögliche Klassifikationen dieser Beweisversuche. Dabei wird wegen der geringen Anzahl an Beweisversuchen mit weder begründendem noch prüfendem Ansatz nicht durchgehend zwischen den Klassenstufen 5 bis 8 differenziert. Nur wenn deutliche Unterschiede zu erkennen sind, wird dies vermerkt. Die Beweisversuche aus der bewusst gezogenen Stichprobe und der Zufallsstichprobe zusammen genommen handelt es sich in insgesamt 36 Fällen um Bearbeitungen mit weder begründendem noch prüfendem Ansatz. Der relative Anteil der Beweisversuche mit weder begründendem noch prüfendem Ansatz an allen Beweisversuchen aus der Zufallsstichprobe liegt in jeder der Klassenstufen 5 bis 8 unter 5% (Abschn. 8.1).

8.4.1 Angabe der Allgemeingültigkeit

In diesem Abschnitt wird dargelegt, inwieweit in den Beweisversuchen mit weder begründendem noch prüfendem Ansatz die Allgemeingültigkeit der zu zeigenden Aussage (oder einer nicht ganz korrekten Interpretation derselben) expliziert ist. Dazu wurde bei allen Beweisversuchen mit weder begründendem noch prüfendem Ansatz die Angabe eines Wahrheitswerts

untersucht (Abschn. 3.3.9), mit folgendem Ergebnis: In rund einem Drittel der Beweisversuche mit weder begründendem noch prüfendem Ansatz ist die Allgemeingültigkeit der Aussage postuliert und in rund einem Viertel ist die Gültigkeit der Aussage angegeben, ohne dass ein Hinweis auf deren Allgemeingültigkeit zu erkennen ist. Insgesamt wird die Aussage in ca. 60% der Beweisversuche mit weder begründendem noch prüfendem Ansatz als gültig bezeichnet, während in den übrigen rund 40% keine Angabe der Gültigkeit vorkommt.

8.4.2 Arten der Bearbeitung und Darstellungsmittel, Darstellungsformen

Um nähere Einblicke zu erhalten, welche Eigenschaften die Beweisversuche mit weder begründendem noch prüfendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für die Entwicklung eines Beweises aufweisen, wurden verschiedene Arten der Bearbeitungen differenziert sowie die vorkommenden Darstellungsmittel und die Darstellungsform analysiert (Abschn. 3.3.9, 3.3.11).

Die Ergebnisse zeigen, dass neun der zehn unterschiedenen **Arten der Bearbeitung** in den 36 analysierten Beweisversuchen mit weder begründendem noch prüfendem Ansatz vorkommen. Trotz der geringen Anzahl an Beweisversuchen tauchen auch innerhalb der einzelnen Klassenstufen viele der unterschiedenen Kategorien auf. Beispielsweise gehört in der Klassenstufe 6 jede der sechs Bearbeitungen in eine andere Kategorie. Die Beweisversuche mit weder begründendem noch prüfendem Ansatz scheinen also recht individuell ausgeprägt zu sein, was die Kombination verschiedener Eigenschaften betrifft.

Ein einziger der analysierten Beweisversuche mit weder begründendem noch prüfendem Ansatz befindet sich in der Kategorie WNart3.3 *Angabe der Aussage oder eines Wahrheitswerts, keine klare Affirmation*, während mehr als die Hälfte der Beweisversuche einer der Arten an Bearbeitungen zugehören, die auf die Gültigkeit der Aussage hindeuten. Mit Blick auf elementare Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche lässt sich somit festhalten, dass die Ausführungen in Beweisversuchen mit weder begründendem noch prüfendem Ansatz häufig andeuten, dass die Aussage entsprechend der Aufgabenstellung als wahr angesehen wird, wenngleich dies nicht unbedingt eine korrekte Interpretation der Begrifflichkeiten dieser Aussage impliziert.

Gut ein Drittel der untersuchten Beweisversuche mit weder begründendem noch prüfendem Ansatz legt nahe, dass die Lernenden die Idee hatten, einen Grund für die Gültigkeit der Aussage anzugeben oder ein Beispiel zu betrachten. Auch daran ließe sich im Falle einer Weiterentwicklung der Beweisversuche anknüpfen.

Außerdem fällt auf, dass ca. die Hälfte der Beweisversuche mit weder begründendem noch prüfendem Ansatz in eine der Kategorien gehören, die unter der Oberkategorie WNart2 *abgebrochene Bearbeitung* zusammengefasst sind. Bei einem großen Anteil der Beweisversuche mit weder begründendem noch prüfendem Ansatz sollte also mit Blick auf deren Weiterentwicklung zu einem Beweis geklärt werden, warum die Bearbeitung abgebrochen wurde.

Die Ergebnisse der Analyse der **Darstellungsmittel und Darstellungsform** zeigen, dass bei fast allen Beweisversuchen mit weder begründendem noch prüfendem Ansatz die Darstellungsform verbal ist.[156] In weniger als der Hälfte der Bearbeitungen kommen neben verbalen auch numerische Darstellungsmittel vor. Symbolische und ikonische Darstellungsmittel sind in keinem der analysierten Beweisversuche mit weder begründendem noch prüfendem Ansatz zu finden. In jeder der Klassenstufen gibt es einzelne Beweisversuche mit weder begründendem noch prüfendem Ansatz, in denen einer der strukturierenden Begriffe „Antwort:“, „Lösung:“, „Lösungsweg:“ und „Rechnung:“ vorkommt. Andere strukturierende Darstellungsmittel sind in den analysierten Beweisversuchen mit weder begründendem noch prüfendem Ansatz nicht zu finden.

Zusammenfassend lässt sich also festhalten, dass in Beweisversuchen mit weder begründendem noch prüfendem Ansatz i. d. R. nur verbale Darstellungsmittel und eventuell strukturierende Begriffe vorkommen. Über 50% der Beweisversuche mit weder begründendem noch prüfendem Ansatz deuten an, dass die zu zeigende Aussage als gültig angesehen wird oder dass die Lernenden ein Beispiel oder einen Grund angeben wollten. Näher betrachtet sind die Beweisversuche mit weder begründendem noch prüfendem Ansatz jedoch vielfältig und lassen oftmals vieles offen, wie z. B. warum die Bearbeitung scheinbar abgebrochen wurde oder ob die Begrifflichkeiten der zu zeigenden Aussage tatsächlich korrekt interpretiert wurden.

8.4.3 Klassifikation der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit

Aufbauend auf den beschriebenen Ergebnissen zu Beweisversuchen mit weder begründendem noch prüfendem Ansatz werden diese im vorliegenden Abschnitt 8.4.3 mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage und im nachfolgenden Abschnitt 8.4.4 mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis klassifiziert.

Mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage ist in Beweisversuchen mit weder begründendem noch prüfendem Ansatz allenfalls die Allgemeingültigkeit dieser Aussage expliziert. Da die Allgemeingültigkeit in der Aufgabenstellung vorgegeben ist und eventuell bei der Angabe der Gültigkeit mitgedacht, wenn auch nicht angegeben wurde, werden, in Analogie zur Klassifikation der Beweisversuche mit prüfendem Ansatz, alle Beweisversuche mit einer Angabe der Gültigkeit zusammen gruppiert (NwG1.0.2). Zudem gibt es Beweisversuche mit weder begründendem noch prüfendem Ansatz ohne Angabe der Gültigkeit (NwG1.0.1) und durchgestrichenen Bearbeitungen (NwG99). Durchgestrichene Bearbeitungen werden separiert, weil in diesem Fall keinerlei Beitrag zum Nachweis der Gültigkeit abgegeben wurde. Insgesamt gehören rund 60% der Beweisversuche mit weder begründendem noch prüfendem Ansatz zu der Gruppe NwG1.0.2 *Angabe der Gültigkeit* und ca. 20% zu jeder der beiden anderen Gruppen.

[156] Die Ergebnisse zu Darstellungsmitteln und Darstellungsformen beziehen sich auf die Beweisversuche mit weder begründendem noch prüfendem Ansatz, die nicht in die Kategorie WNart2.4 *abgebrochene Bearbeitung, durchgestrichene Ausführungen* fallen (Abschn. 3.3.9, 3.3.11 zum Kategoriensystem).

8.4.4 Klassifikation der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf deren Weiterentwicklung zu einem Beweis

In diesem Abschnitt werden die Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis klassifiziert und die Gruppen anhand einzelner Beweisversuche exemplifiziert.

Als ein elementarer Anknüpfungspunkt für eine Weiterentwicklung kann in Beweisversuchen mit weder begründendem noch prüfendem Ansatz die Gültigkeit der zu zeigenden Aussage angegeben sein. Außerdem können die Ausführungen in Beweisversuchen mit weder begründendem noch prüfendem Ansatz nahelegen, dass die Lernenden zumindest die Idee hatten, einen Grund für die Gültigkeit der Aussage anzugeben oder ein Beispiel zu betrachten. In durchgestrichenen Bearbeitungen können ausgeführte Beispiele oder Beweisideen verborgen sein (Abschn. 3.3.9). Bezüglich der Verwendung von Darstellungsmitteln sind die Beweisversuche mit weder begründendem noch prüfendem Ansatz weitgehend homogen, es sind keine besonderen Anknüpfungspunkte für deren Weiterentwicklung zu sehen (Abschn. 8.4.2). Auf diesen Überlegungen aufbauend werden entsprechend Abbildung 161 **fünf Gruppen** an Beweisversuchen mit weder begründendem noch prüfendem Ansatz unterschieden.

Klassifikation der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

We1120 *Affirmation der Aussage, kein Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*	We1220 *Affirmation der Aussage, Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*
We1110 *keine Affirmation der Aussage, kein Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*	We1210 *keine Affirmation der Aussage, Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*

We5500 *nicht eindeutig kategorisierbar*

Die unterschiedenen Gruppen sind jeweils mit der Abkürzung „We" und einem vierstelligen Code gekennzeichnet. Die Abkürzung „We“ steht für „Weiterentwicklung“ bzw. „Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis“ und der Code beschreibt die wesentlichen Anknüpfungspunkte, die vorhanden sind.

Mit dem Code „5500“ werden die nicht eindeutig kategorisierbaren Beweisversuche gekennzeichnet. Bei allen anderen Gruppen gibt die erste Ziffer „1“ an, dass es sich um Beweisversuche mit weder begründendem noch prüfendem Ansatz handelt. Die zweite Ziffer indiziert, ob ein Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels zu erkennen ist. Die dritte Ziffer bezieht sich auf Existenz einer Affirmation der Aussage. Anknüpfend an die Klassifikationen der Beweisversuche mit begründendem Ansatz (Abschn. 8.2.6.6) und der Beweisversuche mit prüfendem Ansatz (Abschn. 8.3.4.1) ist zudem eine vierte Ziffer angegeben. Diese ist in der vorliegenden Klassifikation „0“, weil in Beweisversuchen mit weder begründendem noch prüfendem Ansatz keine Beweisidee vorkommt.

Abbildung 161: Klassifikation der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

In die Gruppe **We1220** *Affirmation der Aussage, Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels* gehört knapp die Hälfte der Beweisversuche mit weder begründendem noch prüfendem Ansatz. Beispielsweise ist in der Bearbeitung 8.4.4a in Abbildung 162 die Allgemeingültigkeit der Aussage als Grund für deren Gültigkeit angeführt.

Bearbeitung 8.4.4a: Klassenstufe 6, männlich (Code: 0610427)

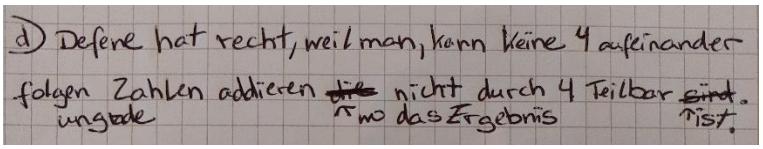


Abbildung 162: Bearbeitung 8.4.4a (We1220 *Affirmation der Aussage, Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*)

In etwa einem Viertel der Beweisversuche mit weder begründendem noch prüfendem Ansatz ist die Gültigkeit der Aussage angegeben, aber kein Hinweis zu erkennen, dass die Lernenden einen Grund oder ein Beispiel angeben wollten (**We1120**). Beispielsweise deutet der Satz „Ja mann kann sie durch 4 Teilen" in der Bearbeitung 8.4.4b in Abbildung 163 an, dass die Aussage von dem Schüler als gültig angesehen wird, ohne dass es irgendeinen Anhaltspunkt dafür gibt, dass er die Idee hatte, einen Grund oder ein Beispiel anzuführen.

Bearbeitung 8.4.4b: Klassenstufe 5, männlich (Code: 0510805)

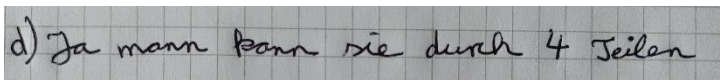
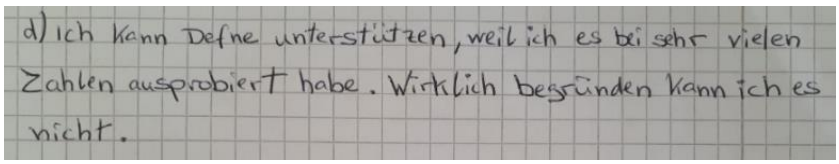


Abbildung 163: Bearbeitung 8.4.4b (We1120 *Affirmation der Aussage, kein Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*)

Bearbeitung 8.4.4c: Klassenstufe 6, männlich (Code: 0510805)



Bearbeitung 8.4.4d: Klassenstufe 5, männlich (Code: 0521619)

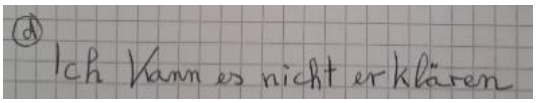


Abbildung 164: Bearbeitung 8.4.4c (We1120 *keine Affirmation der Aussage, Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*), Bearbeitung 8.4.4d (We1110 *keine Affirmation der Aussage, kein Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*)

Im Vergleich zu den Beweisversuchen mit weder begründendem noch prüfendem Ansatz in den Gruppen We1120,1220 ist in denen in den Gruppen **We1110,1210** die Gültigkeit der Aussage nicht angegeben. Trotzdem ist es prinzipiell möglich, dass ein Hinweis darauf zu erkennen ist, dass die Schülerin bzw. der Schüler die Idee hatte, einen Grund für die Gültigkeit der Aussage anzuführen oder ein Beispiel zu betrachten.

Die Bearbeitung 8.4.4c in Abbildung 164 ist die Einzige aus dem analysierten Sample, die in die Gruppe **We1120** *keine Affirmation der Aussage, Idee der Angabe eines Grundes oder eines Beispiels* gehört. Mit ca. einem Viertel der Zuordnungen kommt es häufiger vor, dass in Beweisversuchen mit weder begründendem noch prüfendem Ansatz, in denen die Gültigkeit der Aussage nicht angegeben ist, auch kein Hinweis auf die Idee der Angabe eines Grundes oder

eines Beispiels zu erkennen ist (**We1110**). Mit Blick auf die Weiterentwicklung der Beweisversuche zu einem Beweis deuten diese Beweisversuche, wie z. B. die Bearbeitung 8.4.4d in Abbildung 164, als Anknüpfungspunkt ausschließlich an, dass irgendeine zumindest ansatzweise Auseinandersetzung mit dem Aufgabenteil erfolgt ist.

Mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis einen Sonderfall bilden Beweisversuche mit weder begründendem noch prüfendem Ansatz der Art *durchgestrichene Bearbeitung*, weil die durchgestrichenen Ausführungen sehr vielfältig sein können (Abschn. 3.3.9). Es ist zum Beispiel möglich, dass es sich um einen Beweisversuch mit begründendem Ansatz handeln würde, wenn die Ausführungen nicht durchgestrichen wären. Die eigentlich vorhandenen Anknüpfungspunkte für eine Weiterentwicklung des Beweisversuchs können folglich über die in anderen Beweisversuchen mit weder begründendem noch prüfendem Ansatz zu sehenden Anknüpfungspunkte hinausgehen. Deshalb werden diese ca. 25% der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis als nicht eindeutig kategorisierbar bezeichnet (**We5500**). Beispielsweise gehört die Bearbeitung 8.4.4e in Abbildung 165 zu der Gruppe We5500. Prinzipiell werden sämtliche Beweisversuche mit weder begründendem noch prüfendem Ansatz, die sich keiner der anderen Gruppen eindeutig zuordnen lassen, der Gruppe We5500 zugewiesen.

Bearbeitung 8.4.4e: Klassenstufe 8, männlich (Code: 0811218)

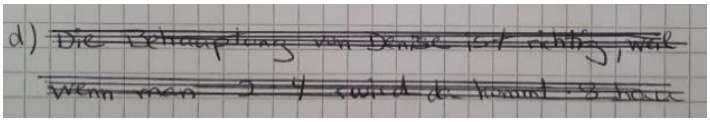


Abbildung 165: Bearbeitung 8.4.4e (We5500 *nicht eindeutig kategorisierbar*)

Insgesamt werden also mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis fünf Gruppen an Beweisversuchen mit weder begründendem noch prüfendem Ansatz unterschieden. Dabei gehört etwa die Hälfte der Beweisversuche mit weder begründendem noch prüfendem Ansatz in die Gruppe We1220 *Affirmation der Aussage, Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*. Dies deutet an, dass auch in vielen Beweisversuchen mit weder begründendem noch prüfendem Ansatz wenn auch sehr elementare Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu erkennen sind.

8.5 Eigenschaften aller Beweisversuche

Im vorliegenden Abschnitt werden die Forschungsfragen dieser Arbeit mit Blick auf alle analysierten Beweisversuche beantwortet. Dazu werden die Erkenntnisse zu Häufigkeiten von Beweisversuchen mit begründendem, prüfendem oder weder begründendem noch prüfendem Ansatz (Abschn. 8.1) und die Ergebnisse der separaten Analysen der Beweisversuche mit begründendem Ansatz, der Beweisversuche mit prüfendem Ansatz und der Beweisversuche mit weder begründendem noch prüfendem Ansatz zusammengeführt. Die Ergebnisse der separaten Analysen bestimmter Beweisversuche werden im Kontext aller analysierter Beweisversuche relativiert und gebündelt mit Blick auf die Forschungsfragen präsentiert.

1. Inwieweit gelingt es Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 eine mathematische Allaussage mit unendlichem Gültigkeitsbereich aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen, welche Charakteristika zeichnen deren Beweisversuche aus und welche Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis sind zu erkennen?
2. Welche Gemeinsamkeiten und Unterschiede bestehen diesbezüglich zwischen Beweisversuchen von Teilnehmenden der Mathematik-Olympiade mit Blick auf verschiedene Klassenstufen?

Im Folgenden liegt der Fokus zuerst auf dem Nachweis der Gültigkeit der zu zeigenden Aussage, d. h. der Teilfrage, inwieweit es den Teilnehmenden gelungen ist, eine mathematische Allaussage mit unendlichem Gültigkeitsbereich aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen (Abschn. 8.5.1). Anschließend wird dargelegt, welche Charakteristika die Beweisversuche von Teilnehmenden der Mathematik-Olympiade auszeichnen und welche Gemeinsamkeiten, Unterschiede diesbezüglich zwischen den Klassenstufen 5 bis 8 bestehen (Abschn. 8.5.2). Nach diesen zwei primär deskriptiven Abschnitten werden die Beweisversuche drittens mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis betrachtet und vorsichtig interpretiert (Abschn. 8.5.3).

Bei der Lektüre dieses Abschnitts 8.5 kann der Eindruck entstehen, dass sich Informationen wiederholen, sowohl im Vergleich zu den Abschnitten 8.1 bis 8.4 als auch innerhalb dieses Abschnitts 8.5. Dies liegt u. a. an der Zielsetzung, die in den Abschnitten 8.1 bis 8.4 herausgearbeiteten Ergebnisse zu resümieren und gebündelt zu präsentieren. Zudem lassen sich Wiederholungen durch die Aufgliederung in die drei Abschnitte 8.5.1 bis 8.5.3 nicht vermeiden. Beispielsweise hängen Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche (Abschn. 8.5.3) damit zusammen, inwieweit den Lernenden ein Beweis der zu zeigenden Aussage gelungen ist (Abschn. 8.5.1) und welche Charakteristika die Beweisversuche etwa hinsichtlich der Beispielnutzung aufweisen (Abschn. 8.5.2). Dieser Nachteil einer gewissen Redundanz wird zugunsten einer systematischen, differenzierten Beschreibung und Interpretation der Beweisversuche in Kauf genommen.

8.5.1 Fokus: Nachweis der Gültigkeit

In diesem Abschnitt liegt der Fokus auf dem Nachweis der Gültigkeit der zu zeigenden Aussage. Es wird zusammenfassend dargelegt, inwieweit es den Teilnehmenden der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 gelungen ist, die vorgegebene mathematische Allaussage mit unendlichem Gültigkeitsbereich zu beweisen, und welche Gemeinsamkeiten und Unterschiede diesbezüglich zwischen Beweisversuchen von Teilnehmenden aus den verschiedenen Klassenstufen bestehen.

Dazu werden sämtliche Beweisversuche zum Aufgabenteil d) mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage klassifiziert.[157] Die in den Abschnitten 8.2.5, 8.3.3, 8.4.3 vorgenommenen Klassifikationen der Beweisversuche mit begründendem, prüfendem bzw.

[157] Als Klassifikation wird dabei eine Zuordnung von Objekten zu einer systematisch geordneten, übersichtlichen Zusammenstellung exhaustiver und disjunkten Kategorien verstanden, bei der neben den Eigenschaften der einzelnen Kategorien auch die Beziehungen zwischen verschiedenen Kategorien im Fokus stehen. Die Beweisversuche werden klassifiziert und nicht typisiert, um nicht zu trennen, was nicht klar zu trennen ist oder Zusammenfassungen vorzunehmen, durch die wesentliche Informationen verloren gehen (Abschn. 7.2.1.6, 7.3.8, 8.2.5).

weder begründendem noch prüfendem Ansatz werden zusammengeführt und die Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz sowie die fehlenden Bearbeitungen in die Klassifikation integriert. Im Folgenden wird zuerst mit einer Beschreibung der Klassifikation ein Überblick gegeben, welche Ausprägungen Beweisversuche hinsichtlich des Nachweises der Gültigkeit der zu zeigenden Aussage aufweisen können, und eine Möglichkeit aufgezeigt, wie sich Beweisversuche diesbezüglich einordnen lassen (Abschn. 8.5.1.1). Darauf aufbauend wird die zentrale Frage, inwieweit es den Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 gelungen ist, die vorgegebene mathematische Allaussage mit unendlichem Gültigkeitsbereich zu beweisen, beantwortet (Abschn. 8.5.1.2). Die Ergebnisse werden abschließend resümiert (Abschn. 8.5.1.3).

8.5.1.1 Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage

In Abbildung 164 ist die entwickelte Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigende Aussage schematisch dargestellt. Im unteren Teil des Schemas sind alle fehlenden und alle durchgestrichenen Bearbeitungen als Gruppe NwG99 von den anderen Beweisversuchen mit begründendem, prüfendem, weder begründendem noch prüfendem oder nicht eindeutig kategorisierbarem Ansatz separiert.[158] Im Falle einer fehlenden oder durchgestrichenen Bearbeitung ist in einem Beweisversuch keinerlei Beitrag zum Nachweis der Gültigkeit der zu zeigenden Aussage zu erkennen. Bei allen anderen Gruppen an Beweisversuchen, die sich im oberen Teil des Schemas befinden, sind zumindest irgendwelche Ausführungen zu der zu zeigenden Aussage oder einer nicht ganz korrekten Interpretation derselben vorhanden.

Die Gruppen an Beweisversuchen im oberen Teil des Schemas unterscheiden sich waagerecht, von links nach rechts, hinsichtlich des Ansatzes und der Schlussweise. In Beweisversuchen mit weder begründendem noch prüfendem Ansatz ist mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage allenfalls angegeben, dass die Aussage allgemeingültig ist. Die Beweisversuche mit prüfendem Ansatz enthalten ein Beispiel, bei dem ein Bezug zu der zu zeigenden Aussage zu erkennen ist. Beweisversuche mit begründendem Ansatz enthalten eine Beweisidee, deren Allgemeingültigkeit oder Verallgemeinerbarkeit in der Bearbeitung explizit angedeutet ist. Dabei kann die Schlussweise deduktiv sein oder zwischen induktiv und deduktiv liegen. Von unten nach oben unterscheiden sich die Gruppen an Beweisversuchen hinsichtlich der Existenz einer Angabe der Gültigkeit der zu zeigenden Aussage, der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage sowie der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges.

Ausführliche Erläuterungen zu den mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage gebildeten Gruppen an Beweisversuchen mit begründendem Ansatz sind in Abschnitt

[158] Der Gruppe NwG99 *fehlend oder durchgestrichen* werden alle Beweisversuche aus der Oberkategorie Ans99 *fehlend* sowie die Beweisversuche mit weder begründendem noch prüfendem Ansatz aus der Kategorie WNArt2.4 *abgebrochene Bearbeitung, durchgestrichene Ausführungen* zugewiesen.

8.2.5 zu finden. Für entsprechende Erläuterungen zu den Beweisversuchen mit prüfendem Ansatz siehe Abschnitt 8.3.3 und zu denen mit weder begründendem noch prüfendem Ansatz Abschnitt 8.4.3.

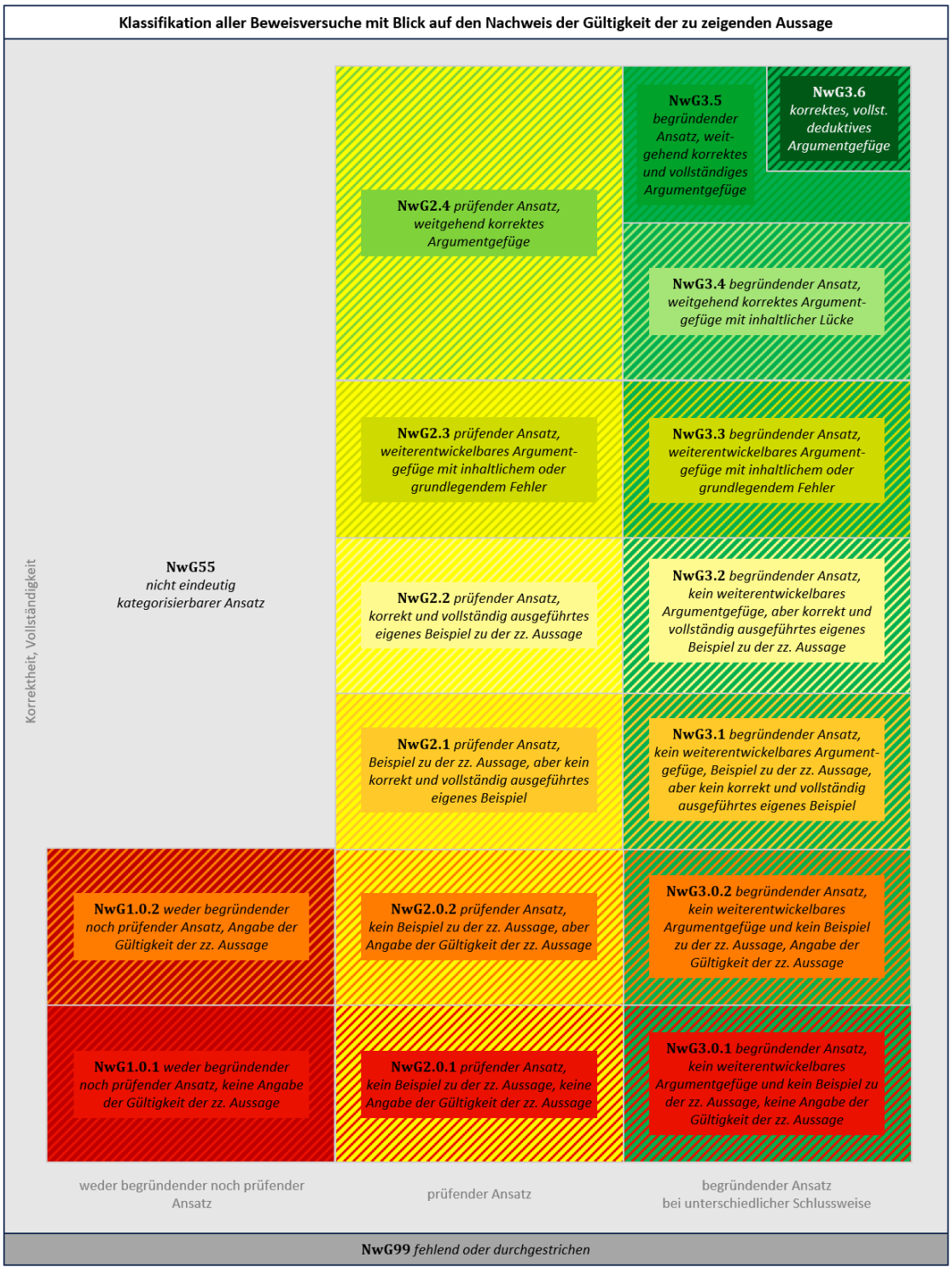


Abbildung 166: Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage

Aus der Berücksichtigung beider Dimensionen, des gewählten Ansatzes zusammen mit der Schlussweise einerseits und, in einem weiten Sinne, der Korrektheit und Vollständigkeit andererseits, ergibt sich, dass die oben rechts verorteten Beweisversuche in der Gruppe NwG3.6 *begründender Ansatz, korrektes und vollständiges deduktives Argumentgefüge* hinsichtlich des Nachweises der Gültigkeit der zu zeigenden Aussage am tragfähigsten sind. Beweisversuche in der Gruppe NwG3.6 werden im Rahmen dieser Arbeit als Beweis bezeichnet (Abschn. 3.4, 8.2.5). Demgegenüber sind die unten links verorteten Beweisversuche in der Gruppe NwG1.0.1 *weder begründender noch prüfender Ansatz, keine klare Angabe der Gültigkeit der zu zeigenden Aussage* abgesehen von den fehlenden und durchgestrichenen Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage am wenigsten tragfähig.

Im oberen Teil des Schemas ist außen etwas hellgrau gefärbter Platz, welcher verdeutlicht, dass sich nicht alle Beweisversuche eindeutig kategorisieren lassen (Abb. 167). Dabei können die Beweisversuche in der Gruppe *NwG*55 *nicht eindeutig kategorisierbarer Ansatz* vielfältig sein, weshalb sie im Klassifikationsschema nicht eindeutig verortet sind. Die Größe der Kästchen und die Größe des hellgrauen Bereichs im Schema sagen nichts über die Häufigkeiten der unterschiedenen Gruppen aussagt. Wie groß der Anteil der Beweisversuche ist, den die einzelnen Gruppen ausmachen, ist in den nachfolgenden Abbildungen zu sehen.

8.5.1.2 Häufigkeitsverteilungen der Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage

Die Abbildungen 167 und 168 zeigen die relativen Häufigkeitsverteilungen der mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage gebildeten Gruppen an Beweisversuchen. Dabei zeigt die Abbildung 167 zwei Teildiagramme, die in der detaillierten Abbildung 168 kombiniert sind. Anhand der Häufigkeitsverteilungen, die in den Abbildungen 167, 168 zu sehen sind, wird im Folgenden dargelegt, welche Ausprägungen die Beweisversuche von Teilnehmenden der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage aufweisen. Zudem wird herausgearbeitet, welche Gemeinsamkeiten und Unterschiede diesbezüglich zwischen den Klassenstufen bestehen.

Einzelne Klassenstufen 5 bis 8

In der **Klassenstufe 5** sind Beweisversuche, die fehlend oder durchgestrichen sind (NwG99), sowie solche mit begründendem Ansatz, allenfalls einem Ansatz für ein weiterentwickelbares Argumentgefüge und keinem Beispiel zu der zu zeigenden Aussage (NwG3.1.1) besonders häufig. Diese beiden Gruppen machen jeweils rund ein Viertel, zusammen ca. die Hälfte aller Beweisversuche aus der Klassenstufe 5 aus.

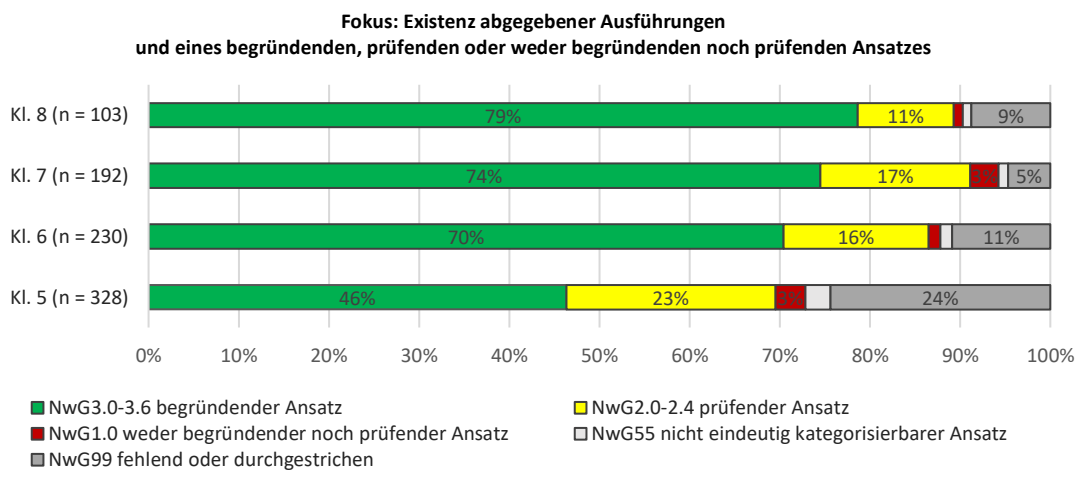
In insgesamt ca. 60% der Beweisversuche aus der **Klassenstufe 5** kommt weder ein weiterentwickelbares Argumentgefüge noch ein Beispiel zu der zu zeigenden Aussage vor (NwG1.0-2.1.1,3.1.1,99). Außerdem ist in keinem der analysierten Beweisversuche aus der Klassenstufe 5 ein Beweis formuliert (NwG3.6). In manchen, aber unter 5% aller Beweisversuche ist ein begründender Ansatz und ein weitgehend korrektes Argumentgefüge (NwG3.4-3.6) zu finden. Mit Blick auf die Klassenstufe 5 sind somit alle Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge als besonders gelungen zu bezeichnen.

In der **Klassenstufe 6** machen Beweisversuche mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge rund die Hälfte aller Beweisversuche aus der Klassenstufe 6 aus und können somit als besonders charakteristisch für diese Klassenstufe angesehen werden (NwG3.1-3.2). In ca. einem Viertel aller Beweisversuche ist ein begründender Ansatz, aber weder ein weiterentwickelbares Argumentgefüge noch ein Beispiel zu der zu zeigenden Aussage formuliert (NwG3.1.1). In einem weiteren Viertel ist ein begründender Ansatz, kein weiterentwickelbares Argumentgefüge, aber ein Beispiel zu der zu zeigenden Aussage angegeben (NwG3.1.2,3.2).

In insgesamt über 40% der Beweisversuche aus der Klassenstufe 6 kommt weder ein weiterentwickelbares Argumentgefüge noch ein Beispiel zu der zu zeigenden Aussage vor (NwG1.0-2.1.1,3.1.1,99). Ein begründender Ansatz und ein weitgehend korrektes Argumentgefüge (NwG3.4-3.6) ist in ca. einem Sechstel der Beweisversuche aus der Klassenstufe 6 zu finden. Einzelnen Teilnehmenden ist ein Beweis der zu zeigenden Aussage gelungen (NwG3.6).

Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage

relative Häufigkeitsverteilungen der Gruppen differenziert nach Klassenstufen



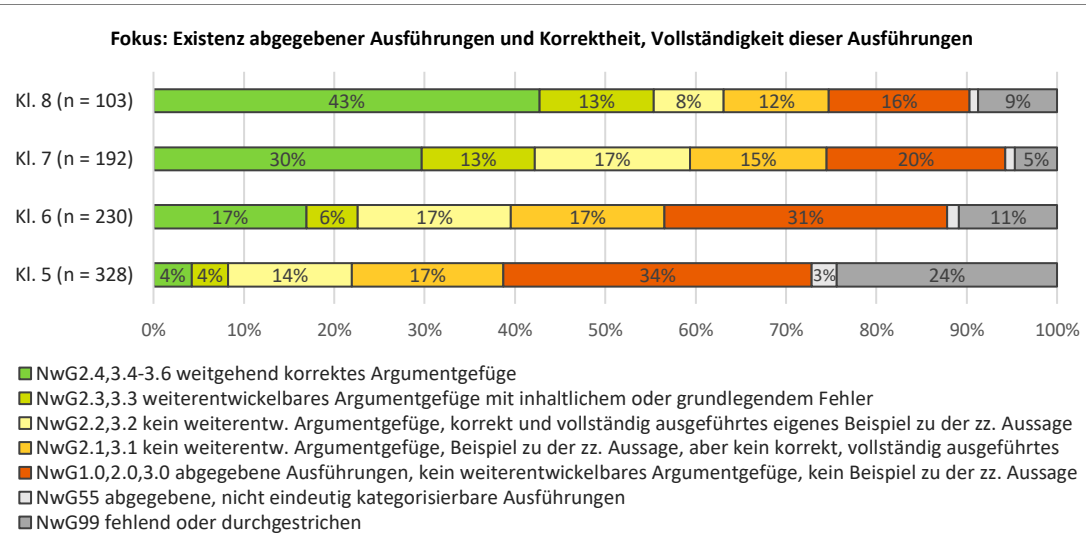


Abbildung 167: Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Gruppen mit Fokus auf (1) der Existenz abgegebener Ausführungen und eines begründenden, prüfenden oder weder begründenden noch prüfenden Ansatzes sowie (2) der Existenz abgegebener Ausführungen und der Korrektheit, Vollständigkeit dieser Ausführungen

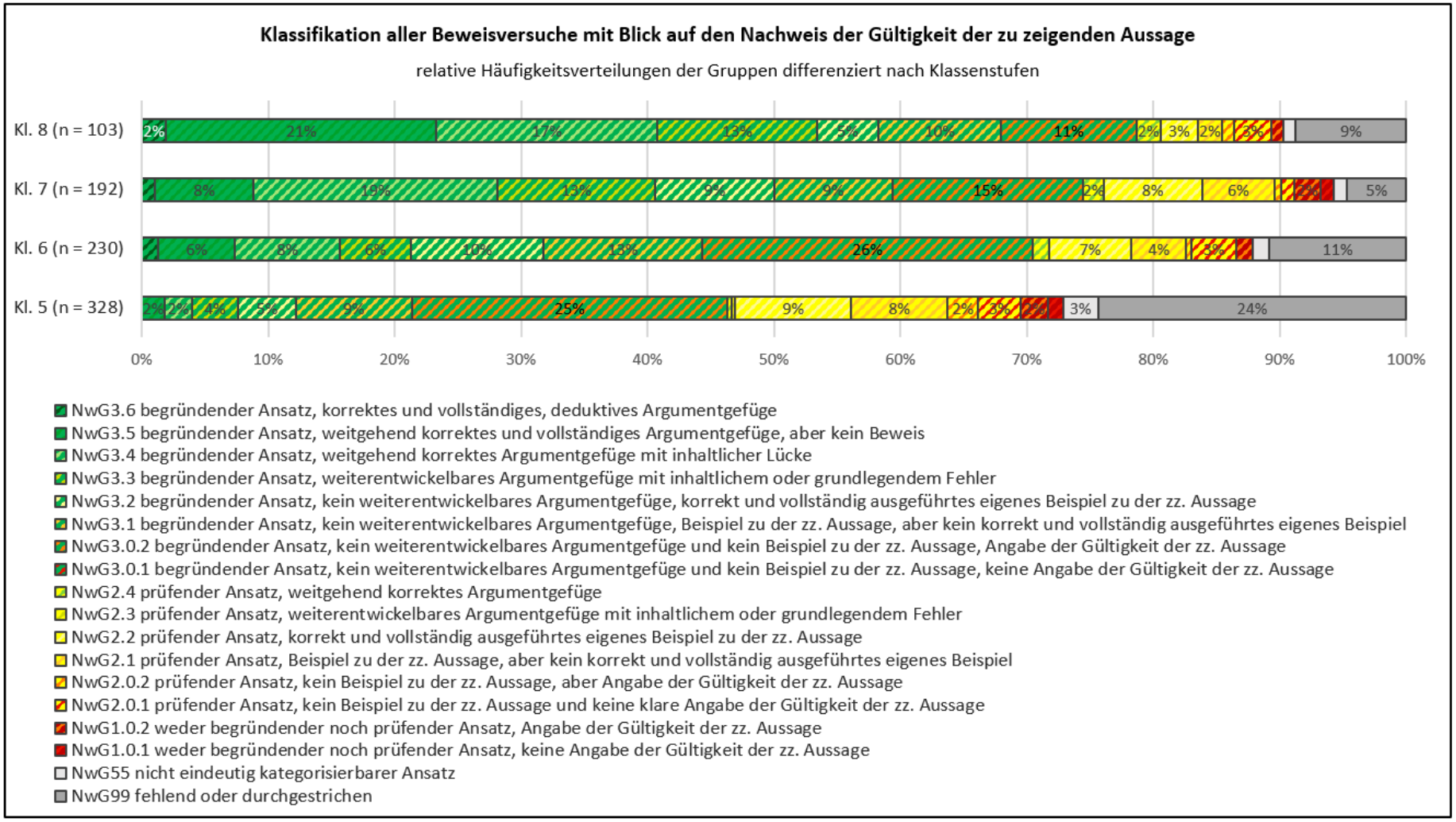


Abbildung 168: Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Gruppen differenziert nach Klassenstufen

Mit Blick auf die **Klassenstufe 7** lassen sich Beweisversuche mit begründendem Ansatz, aber ohne weiterentwickelbares Argumentgefüge (NwG3.1-3.2) sowie Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler oder einer inhaltlichen Lücke (NwG3.3-3.4) als charakteristisch ansehen. Sie machen jeweils rund ein Drittel aller Beweisversuche aus dieser Klassenstufe aus.

Weder ein weiterentwickelbares Argumentgefüge noch ein Beispiel zu der zu zeigenden Aussage (NwG1.0-2.1.1,3.1.1,99) ist in ca. einem Viertel der Beweisversuche aus der Klassenstufe 7 zu finden. In ebenfalls ca. einem Viertel der Beweisversuche aus der Klassenstufe 7 ist ein begründender Ansatz und ein weitgehend korrektes Argumentgefüge (NwG3.4-3.6) formuliert, ein begründender Ansatz und ein weitgehend korrektes und vollständiges Argumentgefüge (NwG3.5-3.6) in knapp einem Zehntel. Einzelnen Teilnehmenden ist ein Beweis der zu zeigenden Aussage gelungen (NwG3.6).

Für die **Klassenstufe 8** sind mit über der Hälfte der Zuordnungen Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge (NwG3.3-3.6) gewissermaßen typisch. Ein begründender Ansatz und ein weitgehend korrektes Argumentgefüge (NwG3.4-3.6) kommen in ca. 40% aller Beweisversuche vor, ein begründender Ansatz und ein weitgehend korrektes und vollständiges Argumentgefüge in rund 23%. Einzelnen Teilnehmenden aus der Klassenstufe 8 ist ein Beweis der zu zeigenden Aussage gelungen (NwG3.6). Zugleich ist in ca. einem Viertel der Beweisversuche aus der Klassenstufe 8 weder ein weiterentwickelbares Argumentgefüge noch ein Beispiel zu der zu zeigenden Aussage formuliert.

Gemeinsamkeiten und Unterschiede zwischen den Klassenstufen

In Ergänzung zur separaten Betrachtung der einzelnen Klassenstufen 5 bis 8 werden im Folgenden wesentliche Gemeinsamkeiten und Unterschiede, die die Beweisversuche aus den verschiedenen Klassenstufen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage aufweisen, dargelegt. Diese Gemeinsamkeiten und Unterschiede werden jeweils zunächst kompakt formuliert und anschließend näher ausgeführt.

In jeder der <u>Klassenstufen 5 bis 8</u> sind die Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage vielfältig. Dabei ist in jeder der Klassenstufen in über 5% der Beweisversuche ein weiterentwickelbares Argumentgefüge zu finden und in mindestens 25% aller Beweisversuche weder ein Beispiel zu der zu zeigenden Aussage noch ein weiterentwickelbares Argumentgefüge zu erkennen.

In jeder der Klassenstufen 5 bis 8 kommen fast alle mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage gebildeten Gruppen an Beweisversuchen vor. In jeder der Klassenstufen enthalten manche Beweisversuche einen begründenden Ansatz und ein weitgehend korrektes und vollständiges Argumentgefüge, in anderen ist die Aussage ausschließlich für einen Spezialfall bestätigt oder allenfalls die Gültigkeit der zu zeigenden Aussage angegeben. In jeder der Klassenstufen kommt es auch vor, dass Lernende an Beispielen ein weiterentwickelbares Argumentgefüge formuliert haben (NwG2.3-2.4). Insgesamt betrachtet ist das Spektrum an vorkommenden Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage in allen Klassenstufen 5 bis 8 sehr groß und facettenreich.

Dabei ist in jeder der Klassenstufen in mindestens 5% aller Beweisversuche ein weiterentwickelbares Argumentgefüge zu erkennen. Jeweils 10% bis 20% aller Beweisversuche enthalten kein weiterentwickelbares Argumentgefüge, aber ein korrekt und vollständig ausgeführtes eigenes Beispiel zu der zu zeigenden Aussage. Das heißt, dass in einem Zehntel bis zu einem Fünftel aller Beweisversuche die zu zeigende Aussage im Wesentlichen für einen Spezialfall bestätigt wird. Zudem gilt mit Blick auf jede der Klassenstufen, dass in mindestens 25% der Beweisversuche weder ein weiterentwickelbares Argumentgefüge noch ein Beispiel zu der zu zeigenden Aussage angegeben ist (NwG1.0-2.1.1,3.1.1,99). Mindestens ein Viertel der Beweisversuche tragen allenfalls ansatzweise zum Nachweis der Gültigkeit der Aussage bei.

Die Beweisversuche aus der Klassenstufe 5 unterscheiden sich u. a. durch den vergleichsweise großen Anteil an fehlenden oder durchgestrichenen Beweisversuchen sowie den vergleichsweise kleinen Anteil an Beweisversuchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge deutlich von denen in den Klassenstufen 6 bis 8.

Entsprechend obiger Abbildung 167 gehört in der Klassenstufe 5 ca. jeder vierte Beweisversuch zu der Gruppe NwG99 *fehlend oder durchgestrichen*, in den Klassenstufen 6 bis 8 maximal jeder Zehnte. Der Anteil der fehlenden oder durchgestrichenen Beweisversuche ist in der Klassenstufe 5 statistisch hoch signifikant größer als in jeder der Klassenstufen 6 bis 8. Andersherum betrachtet haben in der Klassenstufe 5 ca. 75% und in den Klassenstufen 6 bis 8 rund 90% bis 95% der Teilnehmenden Ausführungen zum Aufgabenteil d) abgegeben, was eine Voraussetzung darstellt, damit ein Beweisversuch einen Beitrag zum Nachweis der Gültigkeit der zu zeigenden Aussage liefern kann.

Der Anteil der Beweisversuche mit einem begründenden Ansatz (NwG3.1-3.6) beträgt in der Klassenstufe 5 unter 50%, in den Klassenstufen 6 bis 8 jeweils 70% bis 80%. Er ist in der Klassenstufe 5 statistisch hoch signifikant kleiner als in jeder der höheren Klassenstufen mit geringer bis mittlerer Effektstärke. Insbesondere beträgt in der Klassenstufe 5 der Anteil der Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge (NwG3.3-3.6) unter 10%, in den Klassenstufen 6 bis 8 zwischen 20% und 55%. Auch dieser Unterschied ist jeweils statistisch hoch signifikant mit geringer bis mittlerer Effektstärke (Abb. 167; Tab.14). Nebenbei kommt ausschließlich in der Klassenstufe 5 in keinem einzigen der analysierten Beweisversuche ein Beweis der zu zeigenden Aussage (NwG3.6) vor.

Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage Fokus: Klassenstufe 5; p-Werte von Signifikanztests und dazugehörige Effektstärken			
	NwG99 vs. NwG1.0-3.6,55	NwG1.0-2.3,55 vs. NwG3.1-3.6,99	NwG3.1-3.6 vs. NwG1.0-2.3,55,99
Kl. 5 vs. Kl. 6	$p < .001^{N-1}, V = .167$	$p = .003^{N-1}, V = .125$	$p < .001^{N-1}, V = .239$
Kl. 5 vs. Kl. 7	$p < .001^{N-1}, V = .244$	$p = .020^{N-1}, V = .102$	$p < .001^{N-1}, V = .274$
Kl. 5 vs. Kl. 8	$p < .001^{N-1}, V = .168$	$p = .001^{N-1}, V = .161$	$p < .001^{N-1}, V = .276$
	NwG3.3-3.6 vs.¬NwG3.3-3.6	NwG3.4 vs.¬NwG3.4	NwG3.5,3.6 vs.¬NwG3.5,3.6
Kl. 5 vs. Kl. 6	$p < .001^{N-1}, V = .199$	$p = .001^{N-1}, V = .143$	$p < .001^{N-1}, V = .138$
Kl. 5 vs. Kl. 7	$p < .001^{N-1}, V = .400$	$p < .001^{N-1}, V = .297$	$p < .001^{N-1}, V = .165$
Kl. 5 vs. Kl. 8	$p < .001^{N-1}, V = .502$	$p < .001^{N-1}, V = .280$	$p < .001^{N-1}, V = .360$

Tabelle 14: Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Fokus: Klassenstufe 5; p-Werte von Signifikanztests und dazugehörige Effektstärken

Die Beweisversuche aus den <u>Klassenstufen 7, 8</u> enthalten im Vergleich zu denen aus den Klassenstufen 5, 6 häufiger ein weiterentwickelbares Argumentgefüge und seltener weder ein Beispiel zu der zu zeigenden Aussage noch ein weiterentwickelbares Argumentgefüge.

In den Klassenstufen 7, 8 ist in 40% bis 55% aller Beweisversuche ein weiterentwickelbares Argumentgefüge zu finden, in den Klassenstufen 5, 6 in unter 25%. Der Anteil der Beweisversuche mit einem weiterentwickelbaren Argumentgefüge ist in den Klassenstufen 7, 8 statistisch hoch signifikant größer als in den Klassenstufen 5, 6. Der Anteil der Beweisversuche ohne weiterentwickelbares Argumentgefüge und ohne Beispiel zu der zu zeigenden Aussage ($\mathrm{NwG}1.0\text{-}2.1.1,3.1.1,99$) ist demgegenüber in den Klassenstufen 7, 8 mit ca. 25% statistisch hoch signifikant geringer als in der Klassenstufe 6 mit etwa 40% und als in der Klassenstufe 5 mit ca. 60% (Abb. 167-168; Tab. 15).

Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage Fokus: Klassenstufen 5, 6 vs. Klassenstufen 7, 8; p-Werte von Signifikanztests und dazugehörige Effektstärken		
	NwG2.3-2.4,3.4-3.6 vs.¬NwG2.3-2.4,3.4-3.6	NwG1.0-2.1.1,3.1.1,99 vs.¬NwG1.0-2.1.1,3.1.1,99
Kl. 7 vs. Kl. 5	$p < .001^{N-1}$, $V = .404$	$p = .001^{N-1}$, $V = .176$
Kl. 7 vs. Kl. 6	$p < .001^{N-1}$, $V = .210$	$p = .001^{N-1}$, $V = .168$
Kl. 8 vs. Kl. 5	$p < .001^{N-1}$, $V = .507$	$p = .005^{N-1}$, $V = .153$
Kl. 8 vs. Kl. 6	$p < .001^{N-1}$, $V = .322$	$p = .004^{N-1}$, $V = .159$

Tabelle 15: Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Fokus: Klassenstufen 5, 6 vs. Klassenstufen 7, 8; p-Werte von Signifikanztests und dazugehörige Effektstärken

Die Beweisversuche aus der <u>Klassenstufe 8</u> zeichnen sich gegenüber denen aus den Klassenstufen 5 bis 7 durch einen vergleichsweise großen Anteil an Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge aus.

In der Klassenstufe 8 enthalten rund 21% der Beweisversuche einen begründenden Ansatz und ein weitgehend korrektes und vollständiges Argumentgefüge, wenn auch kein korrektes und vollständiges, deduktives Argumentgefüge ($\mathrm{NwG}3.5$). In den Klassenstufen 5 bis 7 hingegen beträgt der Anteil der Gruppe $\mathrm{NwG}3.5$ jeweils unter 10%. Der Anteil der Beweisversuche in der Gruppe $\mathrm{NwG}3.5$ ist in der Klassenstufe 8 statistisch hoch signifikant größer als in den anderen Klassenstufen 5 bis 7. Die Beweisversuche in den Gruppen $\mathrm{NwG}3.5\text{-}3.6$ zusammen betrachtet, geht in der Klassenstufe 8 aus ca. jedem vierten Beweisversuch bis auf eventuell kleine Details oder eine nicht rein deduktive Schlussweise die Gültigkeit der zu zeigenden Aussage hervor (Abb. 167; Tab. 16).

Zudem ist in der Klassenstufe 8 der Gesamtanteil der Beweisversuche mit einem weiterentwickelbaren Argumentgefüge ($\mathrm{NwG}2.3\text{-}2.4,3.3\text{-}3.6$) an allen Beweisversuchen mit über 50% im Vergleich zu ca. 10% bis 40% statistisch signifikant größer als in jeder der Klassenstufen 5 bis 7, wobei der Unterschied zur Klassenstufe 7 allein auf den größeren Anteil der Gruppe $\mathrm{NwG}3.5$ zurückzuführen ist (Abb. 167-168; Tab. 16).

Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage Fokus: Klassenstufe 8, p-Werte von Signifikanztests und dazugehörige Effektstärken		
	NwG3.5 vs.¬NwG3.5	NwG2.3-2.4,3.3-3.6 vs. ¬NwG2.3-2.4,3.3-3.6
Kl. 8 vs. Kl. 5	$p < .001^{N-1}$, $V = .338$	$p < .001^{N-1}$, $V = .507$
Kl. 8 vs. Kl. 6	$p < .001^{N-1}$, $V = .227$	$p < .001^{N-1}$, $V = .322$
Kl. 8 vs. Kl. 7	$p = .001^{N-1}$, $V = .195$	$p = .031^{N-1}$, $V = .126$

Tabelle 16: Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Fokus: Klassenstufe 8; p-Werte von Signifikanztests und dazugehörige Effektstärken

Je höher die Klassenstufe, desto größer ist der Anteil der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge.

In der Klassenstufe 5 enthalten unter 5% aller Beweisversuche einen begründenden Ansatz und ein weitgehend korrektes Argumentgefüge, in der Klassenstufe 6 sind es ca. 15%, in der Klassenstufe 7 rund 28% und in der Klassenstufe 8 gut 40%. Die Unterschiede zwischen den einzelnen Klassenstufen 5 bis 8 sind alle statistisch signifikant (Abb. 167; Tab. 17).

Des Weiteren sind die Beweisversuche mit begründendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage je höher die Klassenstufe, desto tragfähiger, wie in Abschnitt 8.2.5 ausgeführt ist.

Anteile der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge p-Werte von Signifikanztests und dazugehörige Effektstärken			
	NwG3.4-3.6 vs. ¬NwG3.4-3.6		NwG3.4-3.6 vs. ¬NwG3.4-3.6
Kl. 6 vs. Kl. 5	$p < .001^{N-1}, V = .203$	Kl. 8 vs. Kl. 5	$p < .001^{N-1}, V = .471$
Kl. 7 vs. Kl. 5	$p < .001^{N-1}, V = .348$	Kl. 8 vs. Kl. 6	$p < .001^{N-1}, V = .274$
Kl. 7 vs. Kl. 6	$p = .002^{N-1}, V = .152$	Kl. 8 vs. Kl. 7	$p = .027^{N-1}, V = .129$

Tabelle 17: Anteile der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge: p-Werte von Signifikanztests und dazugehörige Effektstärken

Ein Beweis ist in keinem der analysierten Beweisversuche aus der Klassenstufe 5 formuliert. Einzelne, aber jeweils unter 3% der Teilnehmenden aus den Klassenstufen 6 bis 8 haben die zu zeigende Aussage bewiesen.

Von den analysierten Beweisversuchen aus der Klassenstufe 5 gehört keiner zu der Gruppe NwG3.6 *korrektes und vollständiges, deduktives Argumentgefüge*, d. h. es ist in keinem dieser Beweisversuche ein Beweis zu finden.

In den Klassenstufen 6, 7 beträgt der Anteil der Beweisversuche in der Gruppe NwG3.6 jeweils 1%, in der Klassenstufe 8 beträgt er 2%. Folglich kommen in jeder der Klassenstufen 6 bis 8 Beweise vor, wenngleich diese Ausnahmen darstellen.

8.5.1.3 Zusammenfassung

Um darzulegen, inwieweit es den Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 gelungen ist, die vorgegebene mathematische Allaussage mit unendlichem Gültigkeitsbereich zu beweisen, wurden alle Beweisversuche klassifiziert. In der Klassifikation sind die fehlenden und die durchgestrichenen Bearbeitungen als Gruppe separiert. Die anderen Gruppen mit Beweisversuchen, die nicht fehlend oder durchgestrichen sind, unterscheiden sich hinsichtlich des Ansatzes und der Schlussweise. Zudem sind die Gruppen in Abhängigkeit der Existenz einer Angabe der Gültigkeit der zu zeigenden Aussage, der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage sowie der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges, d. h. in einem weiten Sinne in Abhängigkeit der Korrektheit und Vollständigkeit der Beweisversuche, differenziert.

Die Häufigkeitsverteilungen der mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage gebildeten Gruppen an Beweisversuchen zeigen u. a. Folgendes: Ein Beweis ist in keinem der analysierten Beweisversuche aus der Klassenstufe 5 formuliert. Einzelne, aber jeweils unter 3% der Teilnehmenden aus den Klassenstufen 6 bis 8 haben die zu zeigende Aussage

bewiesen. Dabei ist in jeder der Klassenstufen in mindestens 5% aller Beweisversuche ein weiterentwickelbares Argumentgefüge zu erkennen. In jeweils einem Zehntel bis zu einem Fünftel aller Beweisversuche wird die zu zeigende Aussage im Wesentlichen für einen Spezialfall bestätigt. Je Klassenstufe ist in mindestens einem Viertel aller Beweisversuche weder ein Beispiel zu der zu zeigenden Aussage noch ein weiterentwickelbares Argumentgefüge zu erkennen, so dass die Beweisversuche allenfalls ansatzweise zum Nachweis der Gültigkeit der zu zeigenden Aussage beitragen.

In der Klassenstufe 5 machen Bearbeitungen, die fehlend oder durchgestrichen sind, rund ein Viertel aller Beweisversuche aus. Im Vergleich zur Klassenstufe 5 ist in den Beweisversuchen aus der Klassenstufe 6 häufiger ein begründender Ansatz zu erkennen. Die Beweisversuche aus den Klassenstufen 7, 8 enthalten verglichen mit denen aus beiden Klassenstufen 5 und 6 häufiger ein weiterentwickelbares Argumentgefüge bei begründendem Ansatz. Die Beweisversuche aus der Klassenstufe 8 zeichnen sich gegenüber denen aus den Klassenstufen 5 bis 7 durch einen vergleichsweise großen Anteil an Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge aus. Die Beweisversuche mit begründendem Ansatz im Speziellen sind mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage je höher die Klassenstufe, desto tragfähiger.

8.5.2 Fokus: Charakteristika der Beweisversuche

In diesem Abschnitt liegt der Fokus auf der Teilfrage, welche Charakteristika die Beweisversuche von Teilnehmenden der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 auszeichnen und welche Gemeinsamkeiten, Unterschiede diesbezüglich zwischen diesen Klassenstufen bestehen. Dazu wird nacheinander auf verschiedene Eigenschaften und Kombinationen von Eigenschaften der Beweisversuche eingegangen und aufgezeigt, welche besonders häufig oder aber besonders selten vorkommen. Dennoch sind die Beweisversuche in jeder der Klassenstufen 5 bis 8 sehr vielfältig, wie u. a. in den Abschnitt 8.2, 8.5.3 deutlich wird.

8.5.2.1 Ansatz, Korrektheit, Vollständigkeit und Schlussweise

Zunächst ist herauszustellen, dass aus einem großen Teil der Beweisversuche der **Bezug auf die zu zeigende Aussage** nicht klar hervorgeht. Insgesamt machen die Beweisversuche mit einer nicht ganz klaren Interpretation der zu zeigenden Aussage oder einem Hinweis auf einen grundlegenden Fehler in der Klassenstufe 5 ca. 75%, in der Klassenstufe 6 knapp 60% und in den Klassenstufen 7, 8 etwa 35% aller Beweisversuche aus. Der Anteil ist in der Klassenstufe 5 statistisch hoch signifikant größer als in jeder der Klassenstufen 6 bis 8 und in der Klassenstufe 6 als in jeder der Klassenstufen 7, 8 (jeweils $p < 0{,}001^{N-1}$, $0{,}19 < V < 0{,}41$).

Aus der Analyse des Ansatzes und der Klassifikation der Beweisversuche hinsichtlich des Nachweises der Gültigkeit der zu zeigenden Aussage (Abschn. 8.1; 8.5.1) geht weiterhin hervor, dass in der Klassenstufe 5 das **Fehlen einer Bearbeitung** mit einem Anteil von fast 25% an allen Beweisversuchen aus der Klassenstufe 5 eine häufige und somit charakteristische Ausprägung darstellt. Demgegenüber haben von den Teilnehmenden, die sich mit dem Aufgabenteil er-

kennbar auseinandergesetzt haben, in jeder der Klassenstufen über die Hälfte einen **Beweisversuch mit begründendem Ansatz** formuliert. Insbesondere sind Beweisversuche mit begründendem Ansatz, aber ohne weiterentwickelbares Argumentgefüge mit einem Anteil von 25% bis 50% an allen Beweisversuchen für jede der Klassenstufen 5 bis 8 charakteristisch. Für die Klassenstufe 8 sind mit ca. einem Viertel der Zuordnungen ebenfalls Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge typisch, für die Klassenstufe 7 mit ca. einem Viertel der Zuordnungen solche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge.

Bezüglich der **Schlussweisen** lässt sich festhalten, dass es sich je Klassenstufe bei ca. 30% bis 45% aller Beweisversuche um Beweisversuche mit deduktiver Schlussweise handelt. Für die Klassenstufen 7, 8 sind zudem Beweisversuche mit zwischen induktiver und deduktiver Schlussweise charakteristisch. Sie machen in den Klassenstufen 7, 8 fast die Hälfte aller Beweisversuche aus, in der Klassenstufe 6 gut ein Drittel und in der Klassenstufe 5 ca. ein Siebtel. Der Anteil der Beweisversuche mit zwischen induktiver und deduktiver Schlussweise an allen Beweisversuchen ist in der Klassenstufe 5 statistisch signifikant kleiner als in jeder der Klassenstufen 6 bis 8 (jeweils $\mathrm{p} < 0{,}001^{\mathrm{N-1}}$, $0{,}17 < \mathrm{V} < 0{,}33$) und in der Klassenstufe 6 als in den Klassenstufen 7, 8 (jeweils $\mathrm{p} < 0{,}05^{\mathrm{N-1}}$, $0{,}11 < \mathrm{V} < 0{,}17$).

Die Beweisversuche mit zwischen induktiver und deduktiver Schlussweise enthalten in den Klassenstufen 7, 8 mehrheitlich ein weitgehend korrektes Argumentgefüge (Abschn. 8.2.1.3). Der Anteil der **Beweisversuche mit zwischen induktiver und deduktiver Schlussweise und einem weitgehend korrekten Argumentgefüge** an allen Beweisversuchen liegt in den Klassenstufen 7, 8 bei ca. 25% und ist jeweils statistisch hoch signifikant größer als in den Klassenstufen 5 und 6 (jeweils $\mathrm{p} < 0{,}001^{\mathrm{N-1}}$, $0{,}21 < \mathrm{V} < 0{,}34$).

Die Beweisversuche mit deduktiver Schlussweise enthalten in jeder der Klassenstufen 5 bis 8 mehrheitlich kein weitgehend korrektes Argumentgefüge (Abschn. 8.2.1.3). In jeder der Klassenstufen sind **Beweisversuche mit deduktiver Schlussweise, aber keinem weitgehend korrekten Argumentgefüge** mit einem Anteil von 20% bis 35% an allen Beweisversuchen relativ häufig.

8.5.2.2 Beweisideen

Einhergehend mit der Häufung von Beweisversuchen mit begründendem Ansatz kann die **Erkennbarkeit mindestens einer Beweisidee** als ein Charakteristikum vieler Beweisversuche angesehen werden. In der Klassenstufe 5 ist in rund 50% aller Beweisversuche eine Beweisidee zu erkennen, in den Klassenstufen 6 bis 8 in ca. 75% bis 85%. Die Unterschiede zwischen der Klassenstufe 5 und den Klassenstufen 6 bis 8 sind statistisch hoch signifikant (jeweils $\mathrm{p} < 0{,}001^{\mathrm{N-1}}$, $0{,}25 < \mathrm{V} < 0{,}28$).

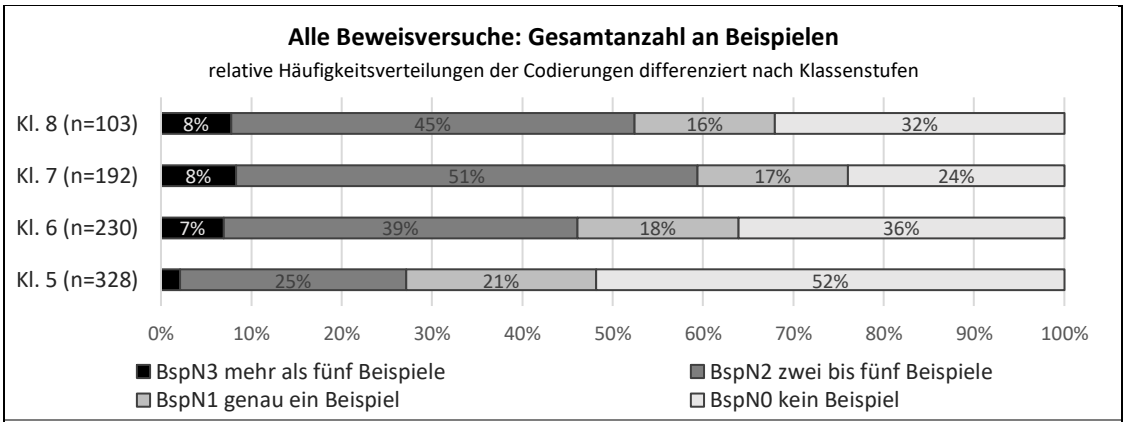
Die häufigste **primäre Beweisidee** ist in den Klassenstufen 7, 8 die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* (Abschn. 8.2.2, 8.3.2.2). In diesen Klassenstufen handelt es sich bei ca. 30% aller Beweisversuche um einen Beweisversuch basierend auf der primären Beweisidee 1, in den Klassenstufen 5, 6 machen diese Beweisversuche mit ca. 5% einen statistisch hoch signifikant kleineren Anteil aus (jeweils $\mathrm{p} < 0{,}001^{\mathrm{N-1}}$,

$0{,}25 < V < 0{,}42$). Für Beweisversuche aus den Klassenstufen 5, 6 sind die primären Beweisideen 6, 9, 11 besonders charakteristisch. Sie machen in der Klassenstufe 5 gut ein Drittel, in der Klassenstufe 6 fast die Hälfte aller Beweisversuche aus und sind damit in beiden Klassenstufen statistisch signifikant häufiger als in den Klassenstufen 7, 8 (jeweils $p < 0{,}035^{N-1}$, $0{,}19 < V < 0{,}32$). Dabei ist ein Beweis der zu zeigenden Aussage auf Grundlage der primären Beweisidee 9 nicht möglich und ausgehend von einer der primären Beweisideen 6, 11 in keinem der analysierten Beweisversuche gelungen (Abschn. 3.3.3, 6.2.2, 8.2.6.3, 8.3.2.2).

Von den einzelnen Beweisversuchen aus den Klassenstufen 6 bis 8, die ein korrektes und vollständiges, deduktives Argumentgefüge enthalten bzw. als Beweis bezeichnet werden können, enthalten über 70% die **Beweisidee 2** ***Darstellung einer Summe in Abhängigkeit des kleinsten Summanden***, sodass die Existenz der Beweisidee 2 als charakteristisch für diese Beweisversuche erscheint (Abschn. 8.2.6.3).

8.5.2.3 Beispiele

In ca. der Hälfte aller Beweisversuche aus der Klassenstufe 5 und in 60% bis 75% derjenigen aus den Klassenstufen 6 bis 8 ist **mindestens ein Beispiel** angegeben. Mit Fokus auf Beweisversuche, die nicht fehlend oder durchgestrichen sind, liegt der Anteil der Beweisversuche mit mindestens einem Beispiel in jeder der Klassenstufen zwischen 60% und 80%. Häufig ist ein Beispiel zu der zu zeigenden Aussage, aber nicht unbedingt ein korrekt und vollständig ausgeführtes eigenes Beispiel angegeben (Abschn. 8.2.3.1, 8.3.1).



p-Werte von Signifikanztests und dazugehörige Effektstärken

alle Beweisversuche	BspN1-3 vs.¬BspN1-3	BspN2-3 vs.¬BspN2-3
Kl. 5 vs. Kl. 6	$p < .001^{N-1}$, $V = .156$	$p < .001^{N-1}$, $V = .196$
Kl. 5 vs. Kl. 7	$p < .001^{N-1}$, $V = .273$	$p < .001^{N-1}$, $V = .219$
Kl. 5 vs. Kl. 8	$p < .001^{N-1}$, $V = .167$	$p < .001^{N-1}$, $V = .229$
Beweisversuche, die nicht fehlend oder durchgestrichen sind	**BspN1-3 vs.¬BspN1-3**	**BspN2-3 vs.¬BspN2-3**
Kl. 5 vs. Kl. 6	$p = .071^{N-1}$, $V = .085$	$p = .002^{N-1}$, $V = .148$
Kl. 5 vs. Kl. 7	$p < .001^{N-1}$, $V = .174$	$p < .001^{N-1}$, $V = .219$
Kl. 5 vs. Kl. 8	$p = .060^{N-1}$, $V = .162$	$p = .001^{N-1}$, $V = .189$

Abbildung 169: Alle Beweisversuche: Gesamtanzahl an Beispielen. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

In Bezug auf die **Gesamtanzahl an Beispielen** deutet sich an, dass die Existenz von zwei bis fünf Beispielen insbesondere für Beweisversuche in den Klassenstufen 6 bis 8 mit einer relativen Häufigkeit von ca. 40% bis 50% charakteristisch ist, eine extensive Angabe von Beispielen in

Form einer Angabe von mehr als fünf Beispielen hingegen eher untypisch. In der Klassenstufe 5 ist die Existenz von zwei oder mehr Beispielen vergleichsweise selten, auch wenn man die Beweisversuche, die nicht fehlend oder durchgestrichen sind, fokussiert (Abb. 169).

Ein **großes Beispiel** ist je Klassenstufe in 5% bis 15% aller Beweisversuche mit Beispiel angegeben. Häufiger kommt das **kleinste Beispiel** vor. Das kleinste Beispiel enthalten in den Klassenstufen 5, 6 etwa 40%, in den Klassenstufen 7, 8 gut 50% aller Beweisversuche mit Beispiel, die Unterschiede zwischen den Klassenstufen 5 und 7, 8 sind statistisch signifikant (jeweils $p < .030^{N-1}$, $V \approx .145$).

Der Anteil der Beweisversuche mit **systematischen Beispielen** an allen Beweisversuchen mit mindestens zwei Beispielen beträgt in jeder der Klassenstufen 5 bis 8 zwischen 65% und 85%. Es deutet sich somit an, dass die meisten Lernenden, die mindestens zwei Beispiele angegeben haben, systematisch vorgegangen sind, wobei diejenigen aus den Klassenstufen 7, 8 häufiger Reihenfolgebeispiele gewählt haben. **Reihenfolgebeispiele** sind in den Klassenstufen 5, 6 in ca. 35%, in den Klassenstufen 7, 8 in ca. 60% der Beweisversuche mit mindestens zwei Beispielen zu finden, die Unterschiede zwischen den einzelnen Klassenstufen 5, 6 und 7, 8 sind statistisch signifikant (jeweils $p \leq 0{,}006^{N-1}$, $0{,}19 < V < 0{,}33$). Dabei ist die Existenz von Reihenfolgebeispielen oftmals in Zusammenhang mit der Verwendung der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* zu sehen (Abschn. 8.2.6.4, 8.3.2.2).

Bezüglich des **Orts der Beispiele in Beweisversuchen mit begründendem Ansatz und Beispiel** lässt sich unter anderem festhalten, dass je Klassenstufe in 70% bis 85% der Fälle alle Beispiele oberhalb oder innerhalb des (übrigen) Argumentgefüges angeführt sind (Abschn. 8.3.2.2). Mit Blick auf **Beweisversuche mit prüfendem Ansatz und einer Affirmation der Aussage** fällt unter anderem auf, dass in den Klassenstufen 5, 7, 8 jeweils in mehr als der Hälfte der Fälle angegebene Beispiele explizit als Grund für die Gültigkeit der zu zeigenden Aussage dargestellt sind (Abschn. 8.3.2.3).

8.5.2.4 Darstellungsmittel und Darstellungsform

In diesem Abschnitt liegt der Fokus auf den vorkommenden Darstellungsmitteln und der Darstellungsform. Für die Beweisversuche der Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 erscheint vor allem die Verwendung **verbaler** und **numerischer** Darstellungsmittel als charakteristisch. Von den Beweisversuchen, die nicht fehlend oder durchgestrichen sind, enthalten 98% verbale Darstellungsmittel und 91% numerische Darstellungsmittel. In jeder der Klassenstufen ist die Darstellungsform bei 55% bis 65% der Beweisversuche, die nicht fehlend oder durchgestrichen sind, numerisch-verbal. Insbesondere haben sehr viele der Beweisversuche mit prüfendem Ansatz sowie derjenigen mit begründendem Ansatz und einer der Schlussweisen $Sw1$-3 die Darstellungsform numerisch-verbal (Abschn. 8.2.6.5, 8.3.2.4). Bei je Klassenstufe ca. 20% bis 35% der Beweisversuche, die nicht fehlend oder durchgestrichen sind, ist die Darstellungsform verbal. Die Darstellungsform *verbal* ist insbesondere für Beweisversuche mit weder begründendem noch prüfendem Ansatz und für Beweisversuche mit der Schlussweise $Sw4$ *deduktiv ohne zusätzliches Beispiel*, aber keinem weitgehend korrekten und vollständigen Argumentgefüge charakteristisch (Abschn. 8.2.6.5, 8.4.2).

Mathematische Strukturierungselemente und Symbole mit fester Bedeutung, die den Lernenden nicht unbedingt aus der Schule bekannt sind, kommen in den Beweisversuchen der Lernenden kaum vor. Darüberhinausgehend kann mit Blick auf die Klassenstufen 5 bis 7 eine weitgehende Nichtexistenz von **symbolischen** Darstellungsmitteln und Darstellungsformen mit symbolischem Anteil als charakteristisch angesehen werden. Beweisversuche mit einer Darstellungsform mit symbolischem Anteil kommen in der Klassenstufe 5 nicht vor. In den Klassenstufen 6, 7 liegt deren Anteil unter 3%. Alle analysierten Beweisversuche mit einer Darstellungsform mit symbolischem Anteil aus den Klassenstufen 6, 7 sind Beweisversuche mit begründendem Ansatz, der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und einem weitgehend korrekten Argumentgefüge (insb. Abschn. 8.2.6.5, 8.3.2.4, 8.4.2).

In der Klassenstufe 8 machen Beweisversuche mit einer Darstellungsform mit symbolischem Anteil rund 13% aller Beweisversuche aus. Der Anteil dieser Beweisversuche ist in der Klassenstufe 8 statistisch hoch signifikant größer als in jeder der Klassenstufen 5 bis 7 (jeweils $p < 0{,}001^{ex}$, $0{,}21 < V < 0{,}32$). Bei mehr als zwei Drittel der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil aus der Klassenstufe 8 handelt es sich um Beweisversuche mit deduktiver Schlussweise, einem weitgehend korrekten Argumentgefüge und einer der primären Beweisideen $1, 2, 8$ (insb. Abschn. 8.2.6.5, 8.3.2.4, 8.4.2).

Mit Blick auf jede der Klassenstufen 5 bis 8 gilt: Bei den Beweisversuchen mit einer Darstellungsform mit **ikonischem** Anteil ist die Darstellungsform meistens ikonisch-numerisch-verbal und in keinem der analysierten Fälle rein ikonisch. Als ikonische Darstellungsmittel kommen fast nie figurierte Zahlen, sondern vor allem Klammern, Pfeile, Linien, Einkreisungen, Unterstreichungen und Ähnliches vor. Beweisversuche mit einer Darstellungsform mit ikonischem Anteil sind in den Klassenstufen 5, 6 mit je ca. 4% selten. In den Klassenstufen 7, 8 beträgt deren Anteil ca. 10% und ist jeweils statistisch signifikant größer als in der Klassenstufe 5 (jeweils $p < 0{,}01^{N-1}$, $0{,}11 < V < 0{,}15$).

Strukturierende Darstellungsmittel kommen in den Klassenstufe 5 bis 8 in ca. 30% bis 50% aller Beweisversuche vor. Der Anteil ist in der Klassenstufe 5 statistisch hoch signifikant geringer als in jeder der Klassenstufen 6 bis 8 ($p \leq 0{,}001^{N-1}$, $0{,}14 < V < 0{,}24$). Strukturierende Begriffe im Speziellen wurden je Klassenstufe in 20% bis 35% aller Beweisversuche verwendet.

8.5.2.5 Zusammenfassung

In jeder der **Klassenstufen 5 bis 8** kennzeichnet einen großen Anteil der Beweisversuche die Existenz eines begründenden Ansatzes und damit einhergehend die Erkennbarkeit mindestens einer Beweisidee. Im Speziellen sind für jede der Klassenstufen 5 bis 8 Beweisversuche mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge mit einem Anteil von 25% bis 50% an allen Beweisversuchen charakteristisch. Zudem sind für jede der Klassenstufen Beweisversuche mit deduktiver Schlussweise mit einem Anteil von über 25% charakteristisch. Des Weiteren ist festzuhalten, dass je Klassenstufe in etwa 50% bis 80% aller Beweisversuche ein Beispiel vorkommt. In den Beweisversuchen mit Beispiel sind meistens zwei bis fünf Beispiele, oftmals systematische Beispiele und häufig das kleinste Beispiel angegeben. Die

Darstellungsform der Beweisversuche ist in jeder der Klassenstufen 5 bis 8 besonders häufig numerisch-verbal. Mit Blick auf die Klassenstufen 5 bis 7 ist eine weitgehende Nichtexistenz von symbolischen Darstellungsmitteln und Darstellungsformen mit symbolischem Anteil als charakteristisch anzusehen.

Bezogen auf die einzelnen Klassenstufen 5 bis 8 lassen sich folgende zusätzlich Besonderheiten herausstellen: In der **Klassenstufe 5** machen fehlende Bearbeitungen rund ein Viertel aller Beweisversuche aus. Aus rund drei Viertel aller Beweisversuche aus der Klassenstufe 5 geht der Bezug auf die zu zeigende Aussage nicht klar hervor, manchmal ist ein Hinweis auf einen grundlegenden Fehler zu erkennen. Zudem fällt auf, dass ausschließlich in der Klassenstufe 5 die Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*, Darstellungsformen mit symbolischem Anteil und mathematische Strukturierungselemente in keinem der analysierten Beweisversuche vorkommen. Außerdem enthalten die Beweisversuche mit begründendem Ansatz aus der Klassenstufe 5 besonders häufig keine ausbaubare Beweisidee.

Für beide **Klassenstufen 5, 6** sind die primären Beweisideen 6 *Teilbarkeit der Summen zweier Zahlenpaare*, 9 *Parität von Summen ungerader oder gerader Zahlen* und 11 *vier Zahlen* besonders charakteristisch. Zudem ist in der Klassenstufe 6, ähnlich wie in der Klassenstufe 5, bei mehr als der Hälfte aller Beweisversuche keine klare Passung der Ausführungen zu der zu zeigenden Aussage zu erkennen.

Im Vergleich zu den Klassenstufen 5, 6 charakterisiert die Beweisversuche aus den **Klassenstufen 7, 8** u. a. ein häufigeres Vorkommen von Reihenfolgebeispielen und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Zudem hat ein größerer Anteil der Beweisversuche eine Darstellungsform mit ikonisch-numerischem Anteil und ein größerer Anteil ist mit zwischen induktiver und deduktiver Schlussweise. Dabei gibt es Hinweise darauf, dass diese besonderen Eigenschaften der Beweisversuche aus den Klassenstufen 7, 8 miteinander zusammenhängen. Weiterhin zeichnet ca. ein Viertel der Beweisversuche aus der Klassenstufe 7 die Existenz eines weitgehend korrekten Argumentgefüges aus, in der Klassenstufe 8 die Existenz eines weitgehend korrekten, vollständigen Argumentgefüges.

8.5.3 Fokus: Anknüpfungspunkte für die Weiterentwicklung

Dieser Abschnitt beschäftigt sich mit Anknüpfungspunkten für eine Weiterentwicklung der Beweisversuche zu einem Beweis. Es wird zusammenfassend dargelegt, inwieweit in den Beweisversuchen der Teilnehmenden der Mathematik-Olympiade Anknüpfungspunkte für deren Weiterentwicklung zu erkennen sind. Dabei wird sowohl auf die einzelnen Klassenstufen 5 bis 8 als auch auf Gemeinsamkeiten und Unterschiede zwischen den Klassenstufen eingegangen. Um die Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung in aussagekräftiger und kompakter Form zu beschreiben, werden die Beweisversuche klassifiziert. Die mit Fokus auf Beweisversuche mit begründendem, prüfendem bzw. weder begründendem noch prüfendem Ansatz bereits entwickelten Klassifikationen (Abschn. 8.2.6.6, 8.3.4, 8.4.4) werden zusammengeführt und die Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz sowie die fehlenden Bearbeitungen in die Klassifikation integriert. Im Folgenden wird die Klassifika-

tion der Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis zuerst beschrieben (Abschn. 8.5.3.1). Danach wird ausgehend von den relativen Häufigkeiten der in dieser Klassifikation vorkommenden Gruppen dargestellt, wie die analysierten Beweisversuche aus den Klassenstufen 5 bis 8 mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung ausgeprägt sind (Abschn. 8.5.3.2). Abschließend werden die Ausführungen zusammengefasst (Abschn. 8.5.3.3).

8.5.3.1 Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

In Abbildung 170 ist die Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis schematisch dargestellt. Dabei sind alle Gruppen an Beweisversuchen mit der Abkürzung „We“ und einem vierstelligen Code deklariert. Die Abkürzung „We“ steht für „Weiterentwicklung“ bzw. „Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis“. Der Code kennzeichnet die wesentlichen Anknüpfungspunkte, die vorhanden sind.

Der Gruppe **We9900** umfasst alle fehlenden Bearbeitungen bzw. alle Beweisversuche in der Oberkategorie Ans99 *fehlend*. Diese sind im unteren Teil des Schemas als Gruppe separiert, weil im Falle von fehlenden Bearbeitungen keine schriftlichen Ausführungen vorhanden sind, an die für die Entwicklung eines Beweises angeknüpft werden kann. Es ist unklar, inwieweit die Lernenden die Aufgabenstellung gelesen haben, warum der Aufgabenteil nicht bearbeitet oder die Bearbeitung abgebrochen wurde und wie die Lernenden bei einer intensiven Beschäftigung mit dem Aufgabenteil zurechtkämen. Anders als in der Klassifikation der Beweisversuche mit Blick auf den Nachweis der Gültigkeit in Abschnitt 8.5.1 sind ausschließlich die fehlenden Bearbeitungen im unteren Teil des Schemas separiert, die durchgestrichenen Beweisversuche im oberen Teil des Schemas integriert. Anders als bei fehlenden Bearbeitungen kann im Falle durchgestrichener Beweisversuche davon ausgegangen werden, dass sich die Lernenden mit dem Aufgabenteil beschäftigt haben und irgendwelche Überlegungen erfolgt sind, die sich für die Entwicklung eines Beweises aufgreifen lassen. Daher werden die durchgestrichenen Bearbeitungen nicht zusammen mit den fehlenden Bearbeitungen gruppiert.

Bei allen Gruppen **We1110 bis We3453** verdeutlicht die erste, linke Ziffer, ob die Beweisversuche in der Oberkategorie Ans1 *weder begründender noch prüfender Ansatz*, Ans2 *prüfender Ansatz* oder Ans3 *begründender Ansatz* sind. In Ergänzung dazu gibt die zweite Ziffer im Fall von Beweisversuchen mit weder begründendem noch prüfendem Ansatz an, ob ein Hinweis auf die Idee der Angabe eines Beispiels oder eines Grundes erkennbar ist. Bei den Gruppen mit Beweisversuchen mit prüfendem Ansatz indiziert die zweite Ziffer, inwieweit sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges zu erkennen sind. Bei den Gruppen an Beweisversuchen mit begründendem Ansatz ist mit der zweiten Ziffer die Schlussweise gekennzeichnet. Die beiden linken Ziffern beschreiben somit den Ansatz und die Nähe der Beweisversuche zu solchen mit einem anderen Ansatz. Im Klassifikationsschema sind die Gruppen We1110 bis We3453 je nach Ansatz und Nähe zu Beweisversuchen mit einem anderen Ansatz von links nach recht angeordnet (Abb. 170).

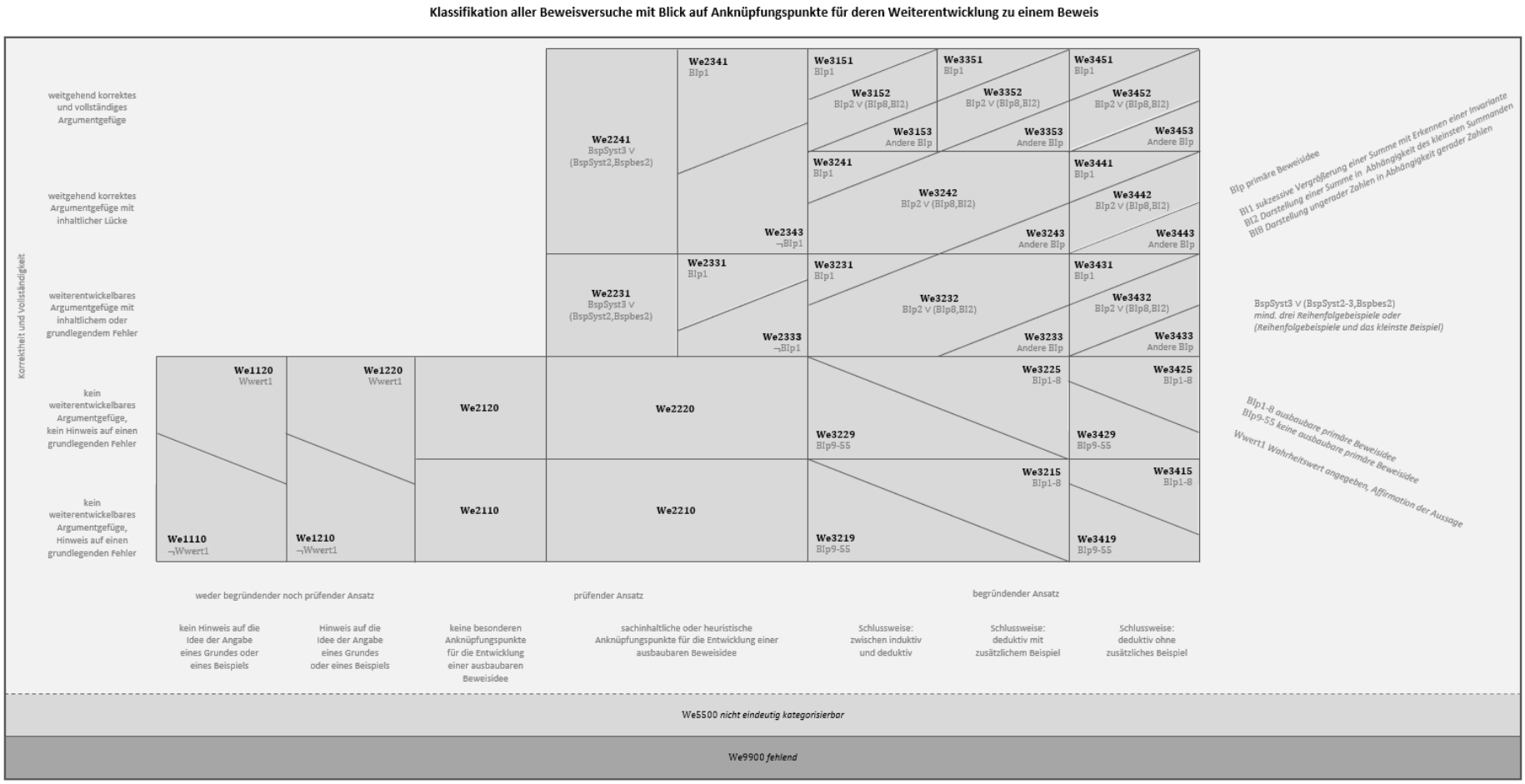


Abbildung 170: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

Von unten nach oben differieren die Gruppen We1110 bis We3453 hinsichtlich der Korrektheit und Vollständigkeit, das heißt konkret der Passung der Ausführungen zu der zu zeigenden Aussage sowie der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Wie die Beweisversuche in den Gruppen diesbezüglich ausgeprägt sind, ist im Code mit der dritten Ziffer von links angegeben. Ist die dritte Ziffer von links beispielsweise eine „1“, dann enthalten die Beweisversuche einen Hinweis auf einen grundlegenden Fehler und kein weiterentwickelbares Argumentgefüge.[159] Ist die dritte Ziffer von links eine „4“, dann enthalten sie zumindest ein weitgehend korrektes Argumentgefüge.

Dabei können die Argumentgefüge auf unterschiedlichen primären Beweisideen basieren. Die vierte, rechte Ziffer der Codes beschreibt die primäre Beweisidee. Beispielsweise basieren die Beweisversuche in Gruppen mit der rechten Ziffer „1“ alle auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Demgegenüber basieren die Beweisversuche in Gruppen mit der rechten Ziffer „9“ auf einer der nicht ausbaubaren Beweisideen 9 bis 13, 55. Ist die letzte Ziffer die „0“, dann ist in den Beweisversuchen keine Beweisidee zu erkennen. Im Klassifikationsschema sind die Gruppen an Beweisversuchen, die sich ausschließlich hinsichtlich der primären Beweisidee unterscheiden, mit diagonalen Linien separiert. Verläuft die diagonale Linie von oben nach unten, sind in den Beweisversuchen in der oben rechts befindlichen Gruppe etwas fortgeschrittenere Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis zu erkennen. Verläuft die diagonale Linie von unten nach oben, dann sind die Beweisversuche in den Gruppen im Wesentlichen anders.

Ausführliche Erläuterungen zu den mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage gebildeten Gruppen mit Beweisversuchen mit begründendem Ansatz sind in Abschnitt 8.2.6 zu finden. Für entsprechende Erläuterungen zu den Gruppen mit Beweisversuchen mit prüfendem Ansatz siehe Abschnitt 8.3.4. Zu den Gruppen mit Beweisversuchen mit weder begründendem noch prüfendem Ansatz siehe Abschnitt 8.4.4.

Zusätzlich zu den Gruppen We9900 und We1110 bis We3453 ist im Klassifikationsschema die Gruppe **We5500** *nicht eindeutig kategorisierbar* ausgewiesen. Diese Gruppe umfasst alle Beweisversuche in der Oberkategorie Ans55 *nicht eindeutig kategorisierbarer Ansatz*, alle durchgestrichenen Bearbeitungen und alle Beweisversuche, die nicht fehlend sind, aber auch nicht klar einer der Gruppen We1110 bis We3453 zugeordnet werden können.

Abschließend ist zu erwähnen, dass die **Größe der Kästchen**, mit denen die einzelnen Gruppen im Klassifikationsschema (Abb. 170) dargestellt sind, nichts über die Anzahlen an Beweisversuchen in den unterschiedenen Gruppen aussagt. Auf die Häufigkeitsverteilungen der Beweisversuche wird im folgenden Abschnitt eingegangen.

[159] Eine Ausnahme bilden diesbezüglich die Gruppen We1110 und We1210 mit Beweisversuchen mit weder begründendem noch prüfendem Ansatz. Da die Beweisversuche in diesen Gruppen keine Affirmation der Aussage, aber nicht unbedingt einen Hinweis auf einen grundlegenden Fehler enthalten, sind sie im Klassifikationsschema von denen mit einer Affirmation der Aussage (We1120,1220) mit diagonalen Linien abgegrenzt.

8.5.3.2 Erkennbare Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis

Die nachfolgenden Abbildungen 171 bis 174 und die Tabelle 18 beschreiben die Häufigkeitsverteilungen der mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis gebildeten Gruppen an Beweisversuchen. In den **Abbildungen 171 bis 174** verdeutlichen die Grautöne, mit denen die einzelnen Gruppen hinterlegt sind, die relativen Anteile der Beweisversuche in diesen Gruppen an allen Beweisversuchen. Je dunkler der Grauton ist, desto größer ist der Anteil der Beweisversuche, die sich in der fokussierten Gruppe befinden. Dabei basieren die relativen Häufigkeitsangaben primär auf den Ergebnissen der Analyse der Beweisversuche aus der Zufallsstichprobe. Um aber auch das Spektrum an vorkommenden Gruppen möglichst angemessen abzubilden, wurden bei den Null-Prozent-Angaben ergänzend die Beweisversuche aus der bewusst gezogenen Stichprobe berücksichtigt. Eine Gruppe ist weiß hinterlegt ($\mathrm{h} = 0\%$), wenn keiner der Beweisversuche aus der Zufallsstichprobe und auch keiner der Beweisversuche aus der bewusst gezogenen Stichprobe in diese Gruppe gehört. Ansonsten ist die Gruppe zumindest mit hellgrauer Farbe markiert ($0\% < \mathrm{h} < 5\%$).

Überblicksartig sieht man anhand der Abbildungen 171 bis 174 zum Beispiel, dass in jeder der Klassenstufen 5 bis 8 mehr als die Hälfte der unterschiedenen Gruppen vorkommen und die Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis vielfältig ausgeprägt sind. Zudem lassen die relativen Häufigkeitsverteilungen der Gruppen erkennen, dass die Beweisversuche tendenziell je höher die Klassenstufe, desto häufiger in eine der im Klassifikationsschema oben rechts befindlichen Gruppen gehören. Dies deutet an, dass in den Beweisversuchen im Mittel je höher die Klassenstufe, desto tragfähigere Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis zu erkennen sind. Außerdem heben die Abbildungen 171 bis 174 hervor, in welche der gebildeten Gruppen sehr viele der analysierten Beweisversuche gehören und welche Ausprägungen der Beweisversuche also mit Blick auf deren Weiterentwicklung zu einem Beweis besonders häufige Ausgangslagen darstellen. Die in den Abbildungen 171 bis 174 erkennbaren Tendenzen ergänzend, lässt sich der **Tabelle 18** entnehmen, einen wie großen Anteil genau die Beweisversuche in den einzelnen Gruppen an allen Beweisversuchen ausmachen.

Ausgehend von den Häufigkeitsverteilungen, die in den Abbildungen 171 bis 174 und in der Tabelle 18 dargestellt sind, und auf Grundlage der Ergebnisse zu Charakteristika der Beweisversuche (z. B. Abschn. 8.5.2) sowie unter Berücksichtigung der Ergebnisse zu Zusammenhängen zwischen verschiedenen Eigenschaften der Beweisversuche (z. B. Abschn. 8.2.6) wird **im Folgenden** differenziert dargelegt, welche Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade zu einem Beweis der zu zeigenden Aussage in den Beweisversuchen erkennbar sind. Dabei werden zuerst die Beweisversuche aus allen Klassenstufen 5 bis 8 gemeinsam betrachtet und danach Besonderheiten der Beweisversuche in den einzelnen Klassenstufen.

Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis

relative Häufigkeiten der Beweisversuche in den gebildeten Gruppen differenziert nach Klassenstufen

	We9900	We1110	We1120	We1210	We1220	We2110	We2120	We2210	We2220	We2231	We2241	We2331	We2333	We2341	We2343
Kl. 5 (n = 328)	23,5%	1,2%	0,0%	1,2%	0,9%	6,4%	4,0%	3,0%	5,8%	0,9%	2,4%	0,0%	0,3%	0,0%	0,0%
Kl. 6 (n = 230)	10,4%	0,0%	0,0%	0,4%	0,9%	3,5%	1,7%	3,9%	3,5%	0,4%	1,7%	0,0%	0,0%	1,3%	0,0%
Kl. 7 (n = 192)	4,7%	1,0%	0,0%	0,0%	2,1%	3,1%	1,0%	1,6%	3,1%	1,6%	4,7%	0,0%	0,0%	1,6%	0,0%
Kl. 8 (n = 103)	7,8%	0,0%	0,0%	0,0%	1,0%	2,9%	1,0%	1,9%	0,0%	1,0%	1,9%	0,0%	0,0%	1,9%	0,0%

	We3219	We3215	We3229	We3225	We3231	We3232	We3233	We3241	We3242	We3243	We3151	We3152	We3153	We3351	We3352
Kl. 5 (n = 328)	4,0%	0,0%	11,3%	3,4%	0,3%	0,0%	0,0%	0,6%	0,0%	1,8%	0,6%	0,0%	0,9%	0,0%	0,0%
Kl. 6 (n = 230)	3,5%	0,4%	19,6%	7,0%	0,4%	0,0%	0,0%	3,9%	0,0%	3,5%	0,9%	0,9%	0,9%	0,0%	0,0%
Kl. 7 (n = 192)	0,5%	0,0%	13,0%	8,3%	2,1%	0,5%	0,0%	13,6%	2,1%	3,6%	5,7%	0,5%	0,5%	1,0%	0,0%
Kl. 8 (n = 103)	1,9%	0,0%	8,7%	3,9%	1,9%	1,9%	0,0%	12,6%	0,0%	1,0%	10,7%	0,0%	0,0%	1,0%	1,0%

	We3353	We3419	We3415	We3429	We3425	We3431	We3432	We3433	We3441	We3442	We3443	We3451	We3452	We3453	We5500
Kl. 5 (n = 328)	0,0%	0,6%	0,0%	16,2%	3,4%	0,6%	0,0%	2,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,3%	3,7%
Kl. 6 (n = 230)	1,3%	0,0%	0,0%	12,6%	6,1%	0,0%	0,0%	5,2%	0,4%	0,0%	0,4%	0,0%	2,2%	1,3%	1,7%
Kl. 7 (n = 192)	0,0%	0,0%	0,0%	9,4%	2,6%	1,6%	0,0%	8,3%	0,0%	0,0%	0,0%	1,0%	0,0%	0,0%	1,0%
Kl. 8 (n = 103)	0,0%	1,0%	1,0%	4,9%	3,9%	1,9%	1,0%	5,8%	0,0%	1,0%	2,9%	2,9%	3,9%	1,9%	1,9%

Hinweis: Die relativen Häufigkeitsangaben basieren primär auf den Ergebnissen der Analyse der Beweisversuche aus der Zufallsstichprobe. Um aber das Spektrum an vorkommenden Gruppen angemessen abzubilden, wurden bei den Null-Prozent-Angaben ergänzend die Beweisversuche aus der bewusst gezogenen Stichprobe berücksichtigt. Die in grau dargestellten Null-Prozent-Angaben bedeuten, dass weder in der Zufallsstichprobe noch in der bewusst gezogenen Stichprobe ein Beweisversuch aus der fokussierten Klassenstufe vorkommt, der in die fokussierte Gruppe gehört. Im Falle einer in schwarz dargestellten Null-Prozent-Angabe kommt in der Zufallsstichprobe kein Beweisversuche aus der fokussierten Klassenstufe vor, der in die angegebene Gruppe gehört, aber in der bewusst gezogenen Stichprobe (Abschn. 7.2.2.1, 7.3.3 zur Stichprobenziehung).

Tabelle 18: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Relative Häufigkeiten der Beweisversuche in den gebildeten Gruppen differenziert nach Klassenstufen

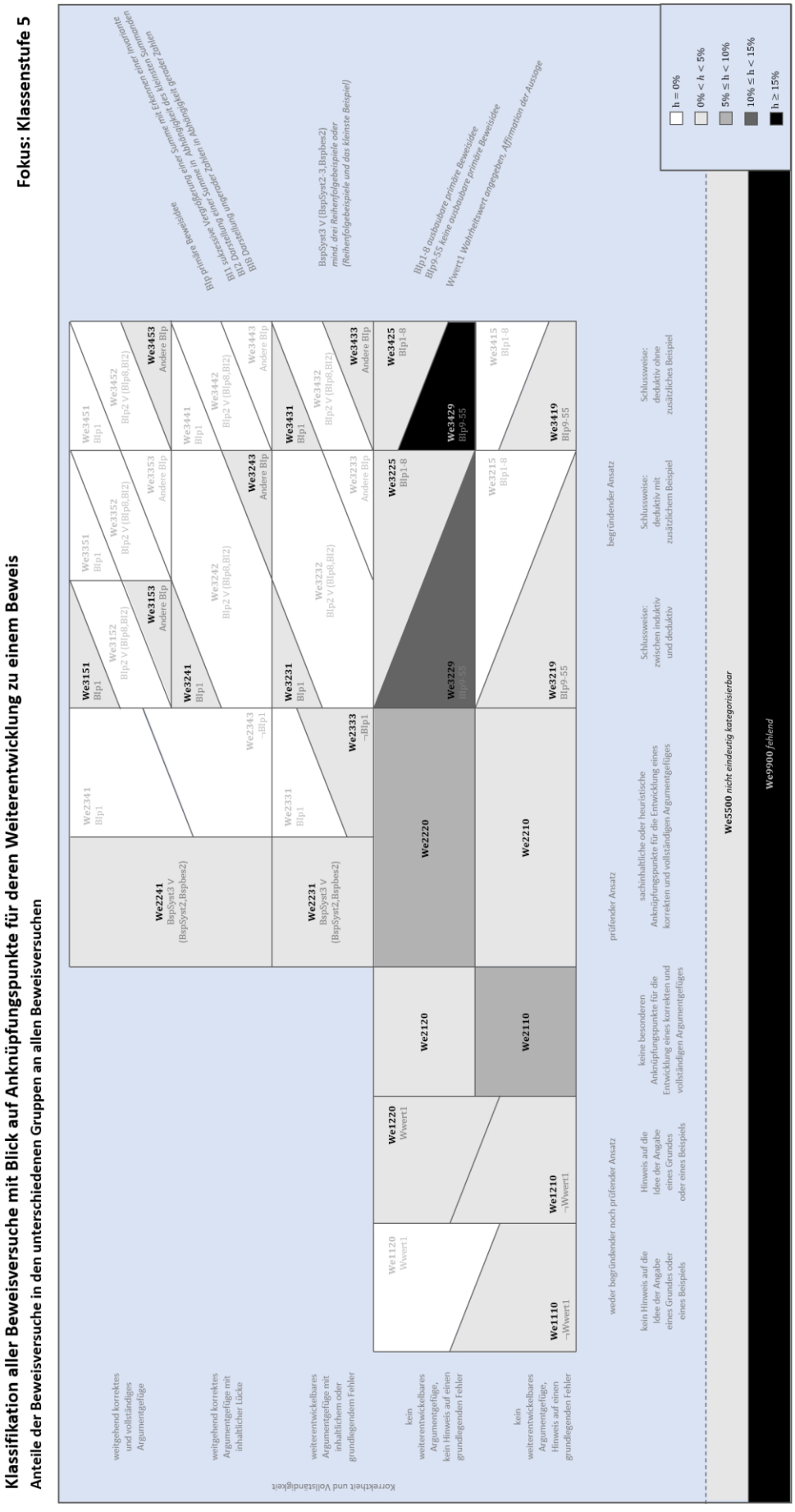


Abbildung 171: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 5)

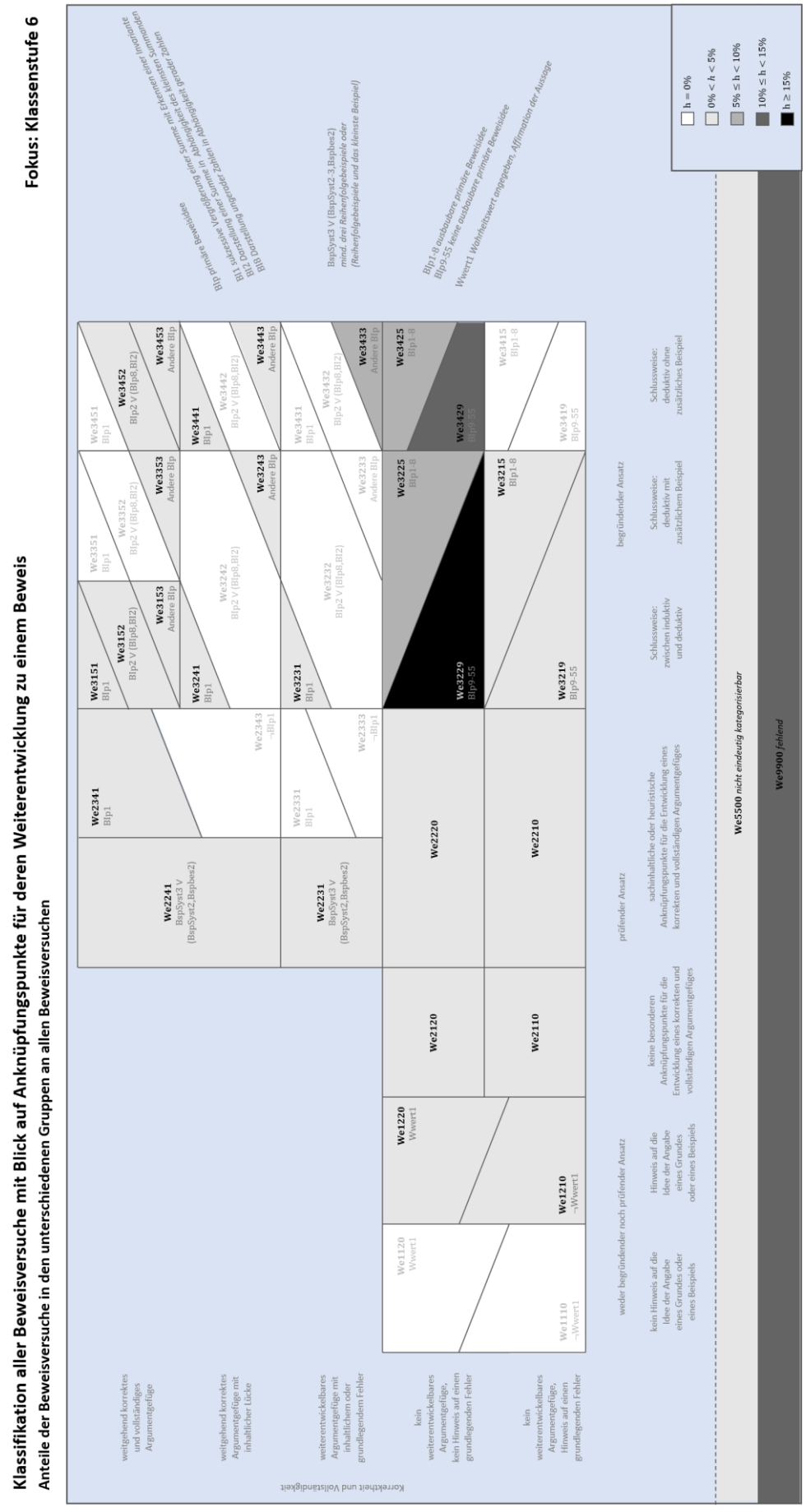


Abbildung 172: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 6)

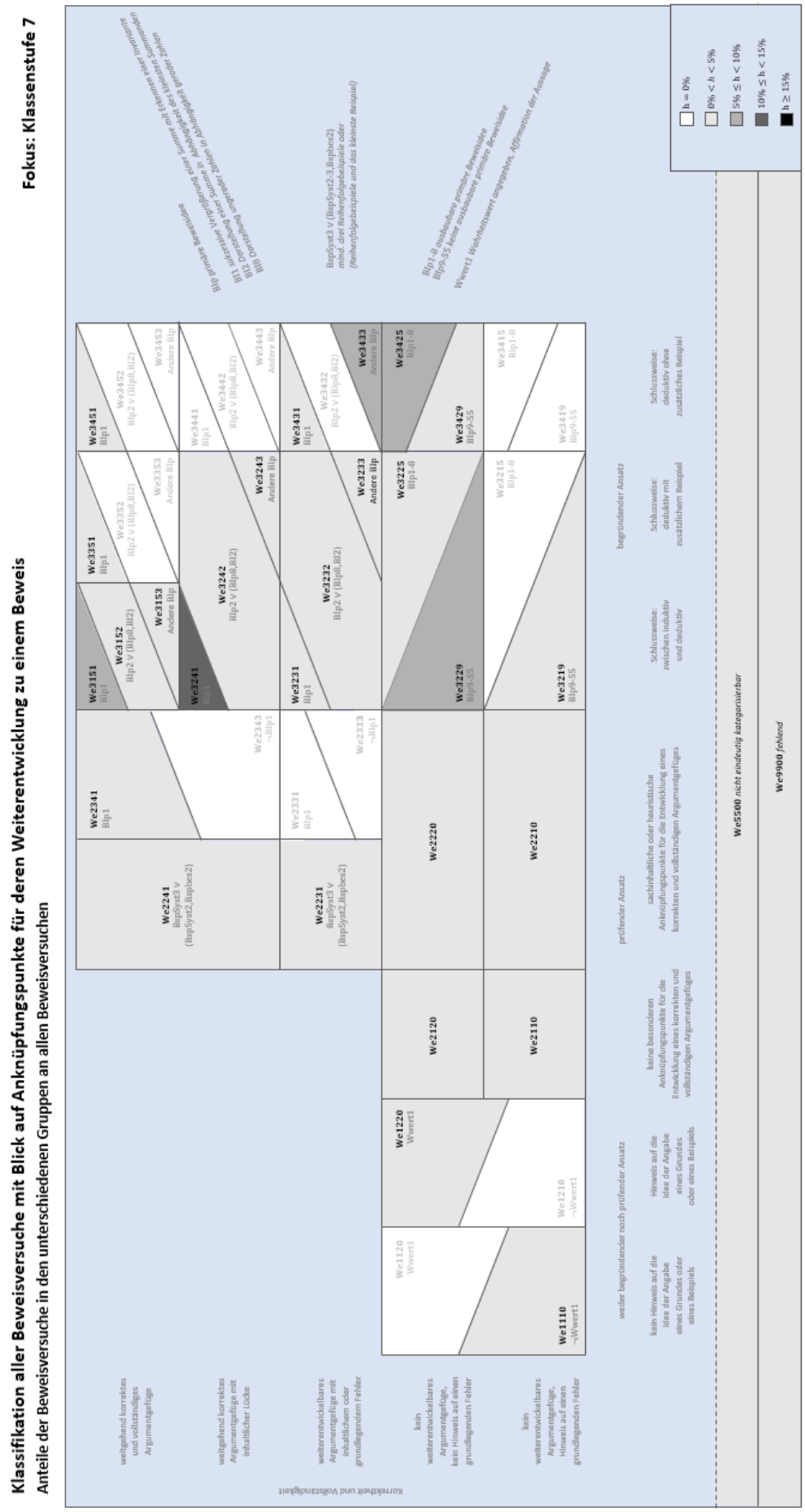


Abbildung 173: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 7)

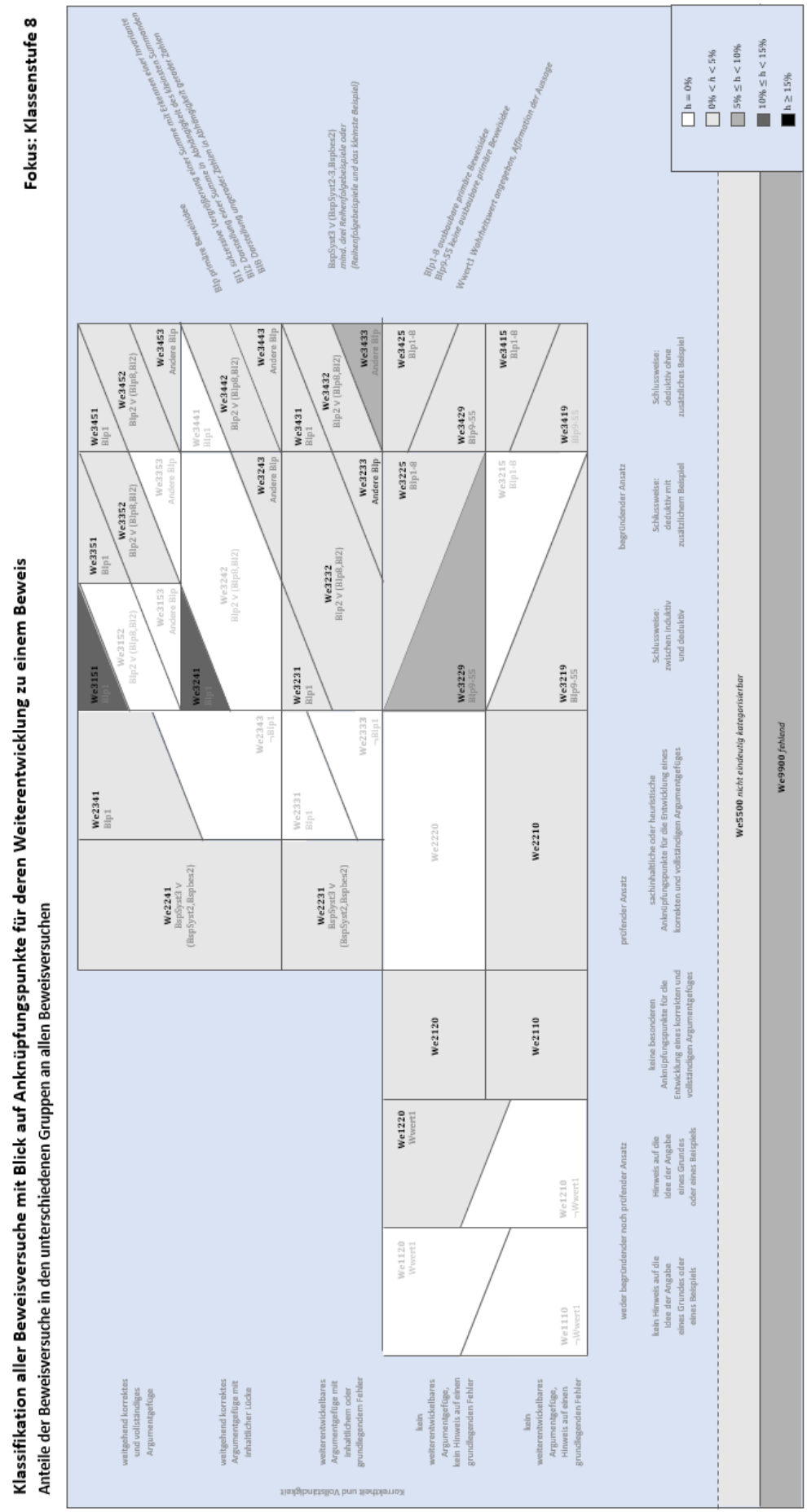


Abbildung 174: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 8)

Klassenstufen 5 bis 8

Mehrere Ergebnisse indizieren, dass in vielen Beweisversuchen von Teilnehmenden der Mathematik-Olympiade aus jeder der Klassenstufen 5 bis 8 durchaus gute Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis vorhanden sind. Beispielsweise zeigen die relativen Häufigkeitsverteilungen der Gruppen, entsprechend den Ausführungen in Abschnitt 8.1, dass der Anteil der Gruppe **We9900** ***fehlend*** in jeder der Klassenstufen 5 bis 8 unter 25%, in den Klassenstufen 6 bis 8 bei maximal 10% liegt. Die Daten legen also offen, dass sich die große Mehrheit der Teilnehmenden mit dem Aufgabenteil d) beschäftigt hat und folglich irgendwelche sachinhaltlichen Überlegungen zum Aufgabenteil d) erfolgt sind, die für eine Weiterentwicklung der Beweisversuche aufgegriffen werden können. Je Klassenstufe hat von den Teilnehmenden, die sich erkennbar mit dem Aufgabenteil auseinandergesetzt haben, über die Hälfte einen **Beweisversuch mit begründendem Ansatz** formuliert (We3125-3453). Dies bedeutet, dass die Teilnehmenden nach allgemeingültigen oder verallgemeinerbaren Argumenten gesucht haben, was einen möglichen Anknüpfungspunkt für eine Weiterentwicklung von Beweisversuchen zu einem Beweis darstellt. Dabei gibt es in jeder der Klassenstufen, d. h. insbesondere auch in der Klassenstufe 5, Lernende, denen ein Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge gelungen ist, sowie weitere, die einen Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge mit inhaltlicher Lücke formuliert haben. Zudem kommen in jeder der Klassenstufen **Beweisversuche mit prüfendem Ansatz** und einem weitgehend korrekten Argumentgefüge vor (We2341-2343). In jeder der Klassenstufen sind in mehr als der Hälfte der Beweisversuche mit prüfendem Ansatz zumindest sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung einer ausbaubaren Beweisidee zu erkennen (Abschn. 8.3.4). Diese Ergebnisse sprechen dafür, durchaus Lernende zu Beginn der Sekundarstufe I an das Beweisen elementarer Allaussagen heranzuführen.

Klassenstufe 5

Mit Blick auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche und die Klassenstufe 5 im Speziellen erscheinen die relativen Anteile der Gruppen We9900,3229,3429 sowie Ergebnisse bezüglich der Schlussweise, der Beispielnutzung und der Darstellungsformen von besonderer Relevanz.

In der Klassenstufe 5 gehören ca. ein Viertel der Beweisversuche zu der Gruppe **We9900** *fehlend* (Abb. 171; Tab. 18). Wie bereits ausgeführt, ist im Falle fehlender Bearbeitungen unklar, warum diese fehlen. Für die Entwicklung eines Beweises kann es z. B. erforderlich sein, beim Lesen und der korrekten Interpretation der zu zeigenden Aussage anzusetzen.

Neben den fehlenden Bearbeitungen geht in der Klassenstufe 5 auch im Falle vorhandener Ausführungen meistens nicht ganz klar aus den Beweisversuchen hervor, dass die zu zeigende Aussage richtig interpretiert wurde (Abschn. 8.5.2), sodass für eine Weiterentwicklung der Be-

weisversuche der Frage, was zu beweisen ist, zumindest kurz Aufmerksamkeit geschenkt werden sollte. Dies gilt zum Beispiel für viele der Beweisversuche in der Gruppe WE3429 (Abschn. 8.2.6.7).

Beweisversuche in den Gruppen **We3229,3429** machen jeweils über 10% aller Beweisversuche aus der Klassenstufe 5 aus (Abb. 171; Tab. 18). Die Beweisversuche in diesen Gruppen enthalten einen begründenden Ansatz, aber kein weiterentwickelbares Argumentgefüge. Die Ausführungen deuten ein Bewusstsein für die Notwendigkeit einer allgemeingültigen Begründung an, wobei für eine Weiterentwicklung der Beweisversuche zu einem Beweis die vorhandenen Ausführungen zu hinterfragen und insbesondere eine ausbaubare Beweisidee zu finden wäre. Es ist möglich, dass in den Beweisversuchen Wissenselemente erkennbar sind, die als Bestandteil eines Beweises zu der zu zeigenden Aussage verwendet werden können, oder dass den Ausführungen implizit inhaltliche Überlegungen zugrunde liegen, die gute Anknüpfungspunkte für die Weiterentwicklung des Beweisversuchs bilden. Fast kein Beweisversuch in der Gruppe We3249, aber alle in der Gruppe We3229 enthalten ein Beispiel, welches eventuell als Anknüpfungspunkt für die Weiterentwicklung der Beweisversuche genutzt werden kann (Abschn. 8.2.6.7).

In der Klassenstufe 5 enthalten unter 10% aller Beweisversuche ein weiterentwickelbares Argumentgefüge, noch seltener ein weitgehend korrektes Argumentgefüge (Abb. 168 in Abschn. 8.5.1.2; Tab. 18). Wenn ein weitgehend korrektes Argumentgefüge erkennbar ist, liegt die Schlussweise in über 90% der analysierten Fälle zwischen induktiv und deduktiv. In der Klassenstufe 5, und auch in jeder der Klassenstufen 6 bis 8, ist unter den Beweisversuchen mit **zwischen induktiver und deduktiver Schlussweise** der Anteil derjenigen mit einem **weitgehend korrekten Argumentgefüge** statistisch signifikant größer als unter den Beweisversuchen mit deduktiver Schlussweise (Abschn. 8.2.1.3). Zudem lassen beispielsweise die Ausführungen in der Bearbeitung 8.2.3k auf Seite 271, in der ein weitgehend korrektes und vollständiges Argumentgefüge bei zwischen induktiver und deduktiver Schlussweise zu erkennen ist, vermuten, dass für eine Weiterentwicklung dieser Ausführungen zu einem deduktiven Argumentgefüge zumindest sprachliche Unterstützung erforderlich wäre und dass dem Fünftklässler kein weitgehend korrektes und vollständiges Argumentgefüge gelungen wäre, wenn er versucht hätte, rein deduktiv zu argumentieren (Abschn. 8.2.3.3). Diese Ergebnisse werfen die Frage auf, inwieweit insbesondere in der Klassenstufe 5 die Formulierung deduktiver Argumentgefüge angestrebt werden sollte oder vielmehr eine produktive Beispielnutzung gefördert und Beweisversuche mit einem weitgehend korrekten Argumentgefüge und zwischen induktiver und deduktiver Schlussweise als sehr gute Beweisversuche akzeptiert werden sollten.

Die Ergebnisse der vorliegenden Arbeit deuten an, dass sich die **Beispielnutzung** von Teilnehmenden der Mathematik-Olympiade in der **Klassenstufe 5**, d. h. dieser besonders jungen Lernenden, in mehrerlei Hinsicht von der Beispielnutzung von Teilnehmenden in den Klassenstufen 7, 8 und teils auch von denen in der Klassenstufe 6 unterscheidet. Im Vergleich zu den Klassenstufen 6 bis 8 kommt in der Klassenstufe 5 statistisch signifikant seltener ein Beispiel vor. Insbesondere ist der Anteil der Beweisversuche mit zwei oder mehr Beispielen geringer (Abschn. 8.5.2).

Bei den Beweisversuchen mit Beispiel ist der Anteil derjenigen mit dem kleinsten Beispiel mit 37% in der Klassenstufe 5 im Vergleich zu 44% in der Klassenstufe 6 und rund 53% in den Klassenstufen 7, 8 geringer. Zudem sind, auch fokussiert auf die Beweisversuche mit mindestens zwei Beispielen, der Anteil der Beweisversuche mit systematischen Beispielen und der Anteil der Beweisversuche mit Reihenfolgebeispielen in der Klassenstufe 5 kleiner als in den Klassenstufe 7, 8 (Abschn. 8.5.2).

Mit Blick auf Beweisversuche mit begründendem Ansatz fällt weiterhin auf, dass von den Beweisversuchen mit Reihenfolgebeispielen und dem kleinsten Beispiel aus der Klassenstufe 5 ein kleinerer Anteil ein weitgehend korrektes Argumentgefüge und ein kleinerer Anteil die primäre Beweisidee 1 enthalten als in jeder der Klassenstufen 6 bis 8. Dabei sind die Unterschiede zu den Klassenstufen 7, 8 statistisch hoch signifikant mit mittlerer bis hoher Effektstärke (Abschn. 8.2.6). Während zudem in jeder der Klassenstufen 6 bis 8 die Beweisversuche mit begründendem Ansatz, Reihenfolgebeispielen und dem kleinsten Beispiel wesentlich häufiger ein weitgehend korrektes Argumentgefüge enthalten als die anderen Beweisversuche mit begründendem Ansatz, ist dieser Unterschied in der Klassenstufe 5 nicht statistisch signifikant (Ab schn. 8.2.6.4).

Des Weiteren ist in der Klassenstufe 5 mit 43% im Vergleich zu ca. 63% in den Klassenstufen 7, 8 der Anteil der Beweisversuche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges an allen Beweisversuchen mit begründendem Ansatz und Beispiel statistisch hoch signifikant kleiner (Abschn. 8.2.3.2). Dabei sind die Beweisversuche mit begründendem Ansatz und einem Beispiel innerhalb des (übrigen) Argumentgefüges hinsichtlich der Existenz, Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in jeder der Klassenstufen, vereinfacht gesagt, besser als die anderen Beweisversuche mit begründendem Ansatz (Abschn. 8.2.6.4). In den Beweisversuchen mit begründendem Ansatz und einem Beispiel aus der Klassenstufe 5 sind mit 43% im Vergleich zu knapp 30% in der Klassenstufe 6 und ca. 20% in den Klassenstufen 7, 8 häufiger alle Beispiele oberhalb des (übrigen) Argumentgefüges angegeben. Die Unterschiede zwischen der Klassenstufe 5 und den Klassenstufen 7, 8 sind statistisch hoch signifikant (Abschn. 8.2.3.2). In der Klassenstufe 5 sind in den Beweisversuchen mit begründendem Ansatz und einem Beispiel oberhalb des (übrigen) Argumentgefüges mit ca. 40% im Vergleich zu ca. 65% in den Klassenstufen 6, 7 und knapp 90% in der Klassenstufe 8 seltener oberhalb des (übrigen) Argumentgefüges systematische Beispiele angegeben oder eine Beweisidee hervorgehoben. Die Unterschiede zu den Klassenstufen 7, 8 sind statistisch signifikant (Abschn. 8.2.3.2). Dabei deuten die Ergebnisse der vorliegenden Studie an, dass sich die Beweisversuche mit begründendem Ansatz je nach Art und Nutzung oberhalb angeführter Beispiele hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges unterscheiden (Abschn. 8.2.6.4). Auch verdeutlicht die Analyse der Beweisversuche mit prüfendem Ansatz, dass in den Beweisversuchen mit prüfendem Ansatz und Reihenfolgebeispielen statistisch hoch signifikant häufiger ein weitgehend korrektes Argumentgefüge zu erkennen ist als in den anderen Beweisversuchen mit prüfendem Ansatz (Abschn. 8.3.2.2).

Insgesamt liefern die Ergebnisse der vorliegenden Arbeit empirische Hinweise darauf, dass Aspekte der Beispielnutzung u. a. mit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in Beweisversuchen mit begründendem oder prüfendem Ansatz zusammenhängen. Die Ergebnisse zeigen zugleich, dass in den Beweisversuchen aus der Klassenstufe 5 seltener zwei oder mehr Beispiele angegeben sind als in den höheren Klassenstufen 6 bis 8 und deuten darauf hin, dass die Auswahl der Beispiele und deren Nutzung weniger zierführend erfolgte vor allem als in den Klassenstufen 7, 8.

Was die **Darstellungsmittel** betrifft, lässt sich in der Klassenstufe 5 insbesondere an die Nutzung verbaler und numerischer Darstellungsmittel anknüpfen. Die Darstellungsformen mit symbolischem Anteil kommen bei den analysierten Beweisversuchen aus der Klassenstufe 5 nicht vor und die Darstellungsformen mit ikonischem Anteil sind selten (Abschn. 8.5.2). Es lässt sich mit Blick auf eine Weiterentwicklung der Beweisversuche aus der Klassenstufe 5 in der Regel nicht davon ausgehen, dass die Lernenden mit der Verwendung symbolischer Darstellungsmittel vertraut sind oder dass sie von sich aus ikonische Darstellungsmittel nutzen und ihnen ein produktiver Einsatz ikonischer Darstellungsmittel gelingt.

Klassenstufe 6

Die Ausführungen zu Darstellungsmitteln gelten ebenfalls für die Klassenstufe 6. Zudem sind in der Klassenstufe 6, so wie in der Klassenstufe 5, die drei Gruppen **We9900**, **We3229** und **We3429** diejenigen, in die die meisten Beweisversuche gehören. Insgesamt sind es rund 43% (Abb. 172; Tab. 18). Neben den Beweisversuchen in diesen drei Gruppen machen in der Klassenstufe 6 die Beweisversuche in den Gruppen **We3225,3425** rund 13% aus, einen doppelt so großen und statistisch signifikant größeren Anteil als in der Klassenstufe 5 (Abb. 171-172; Tab. 18; $\mathrm{p} = .011^{\mathrm{N-1}}$, $\mathrm{V} = .107$). Die Beweisversuche in den Gruppen $\mathrm{We}3225,3425$ kennzeichnet im Vergleich zu denen in den Gruppen $\mathrm{We}3229,3429$, dass als primäre Beweisidee eine ausbaubare Beweisidee zu erkennen ist (Abb. 172; Abschn. 8.2.6.6). Den Lernenden ist es gelungen, eine klar über die zu zeigende Aussage hinausgehende Beweisidee zu entwickeln, wenngleich diese eventuell vage, nicht als mehrschrittiges Argumentgefüge oder nur ansatzweise korrekt dargelegt ist. Die ausbaubare Beweisidee kann prinzipiell für eine Weiterentwicklung des Beweisversuchs genutzt, d. h. beibehalten, bei Bedarf noch klarer herausgearbeitet oder korrigiert und zu einem vollständigen Argumentgefüge ausgebaut werden.

Mit Blick auf eine Weiterentwicklung der Beweisversuche in den Gruppen $\mathrm{We}3225,3245$, und auch derjenigen in den Gruppen $\mathrm{We}3215, 3415$, lässt sich jedoch hinterfragen, inwieweit die erkennbare ausbaubare Beweisidee beibehalten oder die Entwicklung einer anderen ausbaubaren Beweisidee angestrebt werden sollte. Als primäre Beweisidee kommt in den Beweisversuchen in diesen Gruppen meistens eine der Beweisideen 6 T*eilbarkeit der Summen zweier Zahlenpaare* und 4 *Zahldarstellung im Dezimalsystem* vor. Dabei ist basierend auf der primären Beweisidee 4 in keinem der analysierten Beweisversuche ein weitgehend korrektes Argumentgefüge ausgeführt, ausgehend von der primären Beweisidee 6 kommt kein weitgehend korrektes und vollständiges Argumentgefüge vor. Die Ergebnisse der detaillierten Analyse der Beweisversuche mit begründendem Ansatz machen darauf aufmerksam, dass sich deutliche Unter-

schiede hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges auch zwischen verschiedenen primären Beweisideen zeigen, die sich prinzipiell alle für einen Beweis der zu zeigenden Aussage eignen (Abschn. 8.2.6.4). Beweisversuche, die zu einer der Gruppen $\mathrm{We}3225{,}3245$ gehören, sind in der Klassenstufe 6 im Vergleich zur Klassenstufe 5, wie angegeben, häufiger. Die Überlegungen bezüglich der Weiterentwicklung solcher Beweisversuche erscheinen deshalb mit Blick auf die Klassenstufe 6 besonders relevant, gelten aber unabhängig von der Klassenstufe.

Klassenstufen 7, 8

Die Beweisversuche aus den Klassenstufen 7, 8 kennzeichnet gegenüber denjenigen aus den Klassenstufen 5, 6, dass in einem statistisch hoch signifikant größeren Anteil der Beweisversuche die Ausführungen klar zu der zu zeigenden Aussage passen. Zudem enthalten die Beweisversuche statistisch hoch signifikant häufiger ein weiterentwickelbares Argumentgefüge (Abschn. 8.5.1, 8.5.2). Die Analyseergebnisse deuten an, dass in den Beweisversuchen aus den Klassenstufen 7, 8 im Mittel tragfähigere Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind als in den Klassenstufen 5, 6.

Mit jeweils über 10% der Zuordnungen machen in beiden Klassenstufen 7, 8 Beweisversuche, die zu der Gruppe $\mathbf{We3241}$ gehören, einen großen Anteil aus (Abb. 173-174; Tab. 18). Diese Beweisversuche sind, entsprechend den Erläuterungen in Abschnitt 8.2.6.7, mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis recht homogen. Beispielsweise enthalten sie alle ein weitgehend korrektes Argumentgefüge basierend auf der primären Beweisidee 1 *sukzessiver Vergrößerung einer Summe mit Erkennen einer Invariante*.

Auch über die Häufung an Beweisversuchen in der Gruppe $\mathrm{We}3241$ hinausgehend fällt auf, dass in den Klassenstufen 7, 8 Beweisversuche mit der primären Beweisidee 1 oft vorkommen. Sie machen in diesen Klassenstufen mehr als ein Drittel der Beweisversuche mit begründendem Ansatz aus (Abschn. 8.2.2). Zudem ist der Anteil der Beweisversuche mit Reihenfolgebeispielen in den Klassenstufen 7, 8 besonders groß, insbesondere im Falle von Beweisversuchen mit begründendem Ansatz (Abschn. 8.2.3.2, 8.3.2.2, 8.5.2). In den Klassenstufen 7, 8 ist in mehr als zwei Drittel aller Beweisversuche ein Beispiel angegeben. Insgesamt lassen die Ergebnisse der Analyse der Beweisversuche vermuten, dass viele Teilnehmende der Mathematik-Olympiade aus den Klassenstufen 7, 8 bei Aufgaben wie dem hier fokussierten Aufgabenteil d) ein Beispiel und insbesondere auch für die Entwicklung einer Beweisidee besonders geeignete systematische Beispiele betrachten, nach Invarianten suchen und ihnen die Nutzung des Invarianzprinzips sowie die Formulierung eines weitgehend korrekten Argumentgefüges gelingt. Daran anknüpfend kann es erforderlich sein, das Argumentgefüge zu vervollständigen und zu einem deduktiven Argumentgefüge weiterzuentwickeln, damit die Ausführungen als Beweis angesehen werden.

Klassenstufe 8

In der Klassenstufe 8 gehören zusätzlich zum großen Anteil an Beweisversuchen in der Gruppe $\mathrm{We}3241$ rund 11% zu der Gruppe $\mathbf{We3151}$ (Abb. 174; Tab. 18). Die Beweisversuche in der

Gruppe We3151 sind denen in der Gruppe We3241 ähnlich, u. a. weil das Argumentgefüge ebenfalls auf der primären Beweisidee 1 basiert. Dabei enthalten die Beweisversuche in der Gruppe We3241 ein weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke, die in der Gruppe We3151 ein weitgehend korrektes und vollständiges Argumentgefüge (Abb. 174; Abschn. 8.2.6.6). Auch über die Beweisversuche in der Gruppe We3151 hinausgehend sind in der Klassenstufe 8 **Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge** häufig, während der Anteil der Beweisversuche, in denen ein Beweis formuliert ist, unter 3% liegt (Abschn. 8.5.1.2). Folglich scheint auch mit Blick auf Teilnehmende der Mathematik-Olympiade in der Klassenstufe 8 zwischen der Entwicklung eines weitgehend korrekten und vollständigen Argumentgefüges und der Abgabe eines korrekten und vollständigen deduktiven Argumentgefüges ein deutlicher Anforderungssprung zu bestehen.

Rund 88% der Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge aus der Klassenstufe 8 enthalten einen Formulierungsfehler oder eine Formulierungslücke. Mit Blick auf die Weiterentwicklung dieser Beweisversuche stellt sich die Frage, inwieweit diese Ungenauigkeiten dadurch bedingt sind, dass Ausführungen nicht kontrolliert wurden oder dass die Lernenden diese kleinen Mängel nicht als solche erkannt haben. Zudem ist zu bedenken, dass nicht die Formulierungsfehler, aber die Formulierungslücken in einem gewissen Maße Ansichtssache sind, weil sich die Vollständigkeit eines Argumentgefüges nicht rein objektiv beurteilen lässt (Abschn. 3.1.2, 3.3.4). Diese Problematik ist Lernenden, insbesondere auch Teilnehmenden der Mathematik-Olympiade in der Klassenstufe 8, denen ein weitgehend korrektes und vollständiges Argumentgefüge gelungen ist, nicht unbedingt bekannt. Möchte man Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge mit einer Formulierungslücke zu einem vollständigen Argumentgefüge weiterentwickeln, sollten ggf. auch Aspekte des mathematischen Beweisbegriffs thematisiert werden, um angemessene Vorstellungen dazu, was ein Beweis ist, zu fördern und den Eindruck, die Vollständigkeit eines Argumentgefüges präzise beurteilen zu sollen, sowie damit eventuell einhergehende Unsicherheiten und Gefühle der Überforderung zu vermeiden.

Weiterhin liegt die Schlussweise in der Mehrheit der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge aus der Klassenstufe 8 zwischen induktiv und deduktiv (Tab. 18). Damit die Beweisversuche als Beweis bezeichnet werden können, sind die Ausführungen zu einem deduktiven Argumentgefüge weiterzuentwickeln. Dafür stehen Lernenden in der Klassenstufe 8 prinzipiell auch elementare symbolische Darstellungsmittel zur Verfügung. Mit Blick auf eine Weiterentwicklung von Beweisversuchen aus der Klassenstufe 8, in denen keine symbolischen Darstellungsmittel, insbesondere keine Variablen, vorkommen, erscheint trotzdem eine Beschränkung auf verbale, numerische und eventuell ikonische Darstellungsmittel als sinnvoll, um bestmöglich an die vorhandenen Beweisversuche anzuknüpfen, zumal in einigen Beweisversuchen aus der Klassenstufe 8 nur noch kleine Teile des vorhandenen Argumentgefüges zu einem deduktiven Argumentgefüge weiterentwickelt werden müssen (Abschn. 8.2.6.7).

Dadurch, dass in der Klassenstufe 8 im Vergleich zu den Klassenstufen 5 bis 7 vielfältige Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge formuliert wurden, erscheint das **Spektrum an vorkommenden Beweisversuchen** mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung in der Klassenstufe 8 besonders groß (Abb. 171-174). In jeder der Klassenstufen 5 bis 8 sind mehr die Hälfte der unterschiedenen Gruppen an Beweisversuchen, in denen kein weiterentwickelbares Argumentgefüge bei begründendem Ansatz erkennbar ist, von Bedeutung. Beispielsweise gibt es in jeder der Klassenstufen fehlende Bearbeitungen (We9900), Beweisversuche mit weder begründendem noch prüfendem Ansatz, einer Angabe der Gültigkeit und einem Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels (We1220), Beweisversuche mit prüfendem Ansatz, einem Hinweis auf einen grundlegenden Fehler und keinem besonderen Anknüpfungspunkt für die Entwicklung eines korrekten und vollständigen Argumentgefüges (We2110), Beweisversuche mit prüfendem Ansatz und einem weitgehend korrekten Argumentgefüge basierend auf der primären Beweisidee 1 (We2341) sowie Beweisversuche mit deduktiver Schlussweise ohne zusätzliches Beispiel, keinem Hinweis auf einen grundlegenden Fehler, aber auch keinem weiterentwickelbaren Argumentgefüge und keiner ausbaubaren Beweisidee als primärer Beweisidee (We3429).

Beweisversuche aus den Klassenstufen 6, 7 erscheinen mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung etwas vielfältiger als die aus der Klassenstufe 5, u. a. weil auch Beweisversuche mit der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* oder der primären Beweisidee 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* und der Beweisidee 2 vorkommen (Abb. 171-174; Tab. 18; Abschn. 8.2.2, 8.3.2.2, 8.2.6.7). In der Klassenstufe 5 sind rund 38%, in den Klassenstufen 6, 7 ca. 50% und in der Klassenstufe 8 gut 70% der 21 unterschiedenen Gruppen an Beweisversuchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge (We3321-3543) im analysierten Sample zu finden (Abb. 171-174). In der Klassenstufe 8 kommen im Vergleich zu den Klassenstufen 5 bis 7 insbesondere auch vielfältige Beweisversuche mit einem weiterentwickelbaren Argumentgefüge und der Schlussweise *deduktiv ohne zusätzliches Beispiel* vor (Abb. 171-174). Damit einhergehend basieren die Beweisversuche mit einer Darstellungsform mit symbolischem Anteil nicht alle auf der primären Beweisidee 2, sondern auf fünf unterschiedlichen primären Beweisideen (Abschn. 8.2.6.5). Auch sind in einzelnen Beweisversuchen Symbole mit fester Bedeutung, wie das zahlentheoretische Symbol für Kongruenz, zu finden, die den Lernenden nicht unbedingt aus dem Schulunterricht bekannt sind (Abschn. 8.2.4, 8.2.6.5). Inwieweit Aspekte dieser Vielfalt an Beweisversuchen mit einem besonderen Erfolg im Wettbewerb oder einer besonderen mathematischen Förderung zusammenhängen und inwieweit somit bestimmte Ausprägungen der Beweisversuche für unterschiedliche Teilgruppen der Teilnehmenden der Mathematik-Olympiade mehr oder minder relevant erscheinen, ließe sich nach Abschluss der vorliegenden Arbeit aus den im Rahmen dieser Arbeit erhobenen Daten herausarbeiten (Abschn. 10.1, 10.2, 10.4).

8.5.3.3 Zusammenfassung

Um die Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis zusammenfassend zu beschreiben, wurden die Beweisversuche klassifiziert. In der Klassifikation unterscheiden sich die Gruppen an Beweisversuchen insbesondere je nach Ansatz und Nähe der Beweisversuche zu solchen mit einem anderen Ansatz, je nach Korrektheit und Vollständigkeit der Beweisversuche und je nach primärer Beweisidee.

Die Ergebnisse der Klassifikation und der Analyse von Charakteristika der Beweisversuche verdeutlichen, dass in vielen Beweisversuchen aus jeder der **Klassenstufen 5 bis 8** durchaus gute Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis vorhanden sind. Beispielsweise haben je Klassenstufe von den Teilnehmenden, die sich erkennbar mit dem Aufgabenteil auseinandergesetzt haben, mehr als die Hälfte einen Beweisversuch mit begründendem Ansatz formuliert. Dies bedeutet, dass die Teilnehmenden nach allgemeingültigen oder verallgemeinerbaren Argumenten gesucht haben, was einen möglichen Anknüpfungspunkt für eine Weiterentwicklung von Beweisversuchen zu einem Beweis darstellt.

Was die **Klassenstufe 5** betrifft, werfen die Ergebnisse die Frage auf, inwieweit die Formulierung deduktiver Argumentgefüge angestrebt werden sollte oder vielmehr eine produktive Beispielnutzung gefördert und Beweisversuche mit einem weitgehend korrekten Argumentgefüge und zwischen induktiver und deduktiver Schlussweise als sehr gute Beweisversuche angesehen werden sollten. Die Ergebnisse der vorliegenden Arbeit deuten an, dass sich die Beispielnutzung von Teilnehmenden der Mathematik-Olympiade in der Klassenstufe 5 in mehrerlei Hinsicht von der Beispielnutzung von Teilnehmenden in den Klassenstufen 7, 8 und teils auch von denen in der Klassenstufe 6 unterscheidet. Die Ergebnisse zeigen, dass in der Klassenstufe 5 seltener Beispiele und insbesondere seltener zwei oder mehr Beispiele angegeben sind als in den höheren Klassenstufen. Auch deuten sie an, dass die Auswahl der Beispiele und deren Nutzung in der Klassenstufe 5 weniger zielführend erfolgt ist vor allem als in den Klassenstufen 7, 8.

In beiden **Klassenstufen 5 und 6** legen es viele Beweisversuche nahe, dass für deren Weiterentwicklung der Frage, was zu beweisen ist, zumindest kurz Aufmerksamkeit geschenkt werden sollte. Weiterhin machen in beiden Klassenstufen Beweisversuche, die in eine der Gruppen $\mathrm{We}3229,3429$ gehören, etwa 30% aller Beweisversuche aus. Die Beweisversuche in den Gruppen $\mathrm{We}3229,3429$ enthalten einen begründenden Ansatz, keinen Hinweis auf einen grundlegenden Fehler, aber auch kein weiterentwickelbares Argumentgefüge und keine ausbaubare primäre Beweisidee. Die Ausführungen deuten ein Bewusstsein für die Notwendigkeit einer allgemeingültigen Begründung an, wobei für eine Weiterentwicklung der Beweisversuche zu einem Beweis die vorhandenen Ausführungen zu hinterfragen und insbesondere eine ausbaubare Beweisidee zu finden wäre. In der Klassenstufe 5 gehören weitere 7%, in der Klassenstufe 6 weitere 13% aller Beweisversuche zu einer der Gruppen $\mathrm{We}3225$ und $\mathrm{We}3425$. Im Vergleich zu den Beweisversuchen in den Gruppen $\mathrm{We}3229{,}3429$ ist eine ausbaubare primäre Beweisidee zu erkennen. Die ausbaubare Beweisidee kann prinzipiell für eine Weiterentwicklung des Beweisversuchs genutzt, d. h. beibehalten, bei Bedarf noch klarer herausgearbeitet oder korrigiert und zu einem vollständigen Argumentgefüge ausgebaut werden.

Was die Darstellungsmittel betrifft, lässt sich in beiden Klassenstufen 5 und 6 insbesondere an die Nutzung verbaler und numerischer Darstellungsmittel anknüpfen. Es lässt sich in der Regel nicht davon ausgehen, dass die Lernenden mit der Verwendung symbolischer Darstellungsmittel vertraut sind oder dass sie von sich aus ikonische Darstellungsmittel nutzen und ihnen ein produktiver Einsatz ikonischer Darstellungsmittel gelingt. In den Klassenstufen 7 und 8 kommen Darstellungsformen mit ikonisch-numerischem Anteil etwas öfter vor, jedoch vorwiegend in Verbindung mit einer bestimmten Beweisidee. Dass bei der Weiterentwicklung an eine Verwendung symbolischer Darstellungsmittel angeknüpft werden kann, ist in der Klassenstufe 8 bei ca. einem Achtel der Beweisversuche der Fall.

In den Beweisversuchen aus den **Klassenstufen 7 und 8** sind im Mittel tragfähigere Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen als in den Klassenstufen 5 und 6. Die Beweisversuche aus den Klassenstufen 7 und 8 enthalten im Vergleich zu denen aus den Klassenstufen 5 und 6 statistisch hoch signifikant häufiger ein weiterentwickelbares Argumentgefüge. Wenn ein weiterentwickelbares Argumentgefüge vorkommt, können die vorhandenen Ausführungen für einen Beweis der zu zeigenden Aussage im Wesentlichen beibehalten werden. Von den Beweisversuchen aus den Klassenstufen 7 und 8 gehören etwa 20% zu einer der Gruppen $\mathrm{We}3241{,}3151$. Beweisversuche in diesen Gruppen enthalten ein weitgehend korrektes Argumentgefüge basieren auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* fast immer bei zwischen induktiver und deduktiver Schlussweise. Solche Beweisversuche stellen somit eine recht häufige Ausgangslage dar.

Darüberhinausgehend lassen die Analyseergebnisse vermuten, dass viele Teilnehmende der Mathematik-Olympiade aus den Klassenstufen 7 und 8 bei Aufgaben wie dem hier fokussierten Aufgabenteil d) ein Beispiel und auch für die Entwicklung einer Beweisidee besonders geeignete systematische Beispiele betrachten, nach Invarianten suchen und ihnen die Nutzung des Invarianzprinzips sowie die Formulierung eines weitgehend korrekten Argumentgefüges gelingt. Daran anknüpfend kann es erforderlich sein, das Argumentgefüge zu vervollständigen und zu einem deduktiven Argumentgefüge weiterzuentwickeln, damit die Ausführungen als Beweis angesehen werden.

Mit Blick auf die **Klassenstufe 8** ist herauszustellen, dass im Vergleich zu den Klassenstufen 5 bis 7 die Beweisversuche statistisch hoch signifikant häufiger einen begründenden Ansatz und ein weitgehend korrektes und vollständiges Argumentgefüge enthalten, aber ausschließlich einzelne Beweisversuche als Beweis der zu zeigenden Aussage bezeichnet werden können. Bei gut einem Fünftel der Beweisversuche aus der Klassenstufe 8 wären für deren Weiterentwicklung zu einem Beweis Verallgemeinerungen durchzuführen oder Details zu optimieren.

In der Klassenstufe 8 wurden viele verschiedene Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge formuliert. Zugleich kommen so wie in den Klassenstufen 5 bis 7 viele verschiedene Beweisversuche, in denen kein weiterentwickelbares Argumentgefüge bei begründendem Ansatz zu erkennen ist, vor. Dadurch erscheinen die Be-

weisversuche aus der Klassenstufe 8 mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis besonders heterogen, im Vergleich dazu in der Klassenstufe 5 am homogensten.

8.6 Zusammenfassung des achten Kapitels

Ergebnisse

Die Ergebnisse der Analyse der Beweisversuche von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade legen dar, inwieweit es diesen Lernenden gelungen ist, die zu zeigende Aussage zu beweisen. Sie zeigen, dass in jeder der Klassenstufen 5 bis 8 mindestens 25% der Beweisversuche allenfalls ansatzweise zum **Nachweis der Gültigkeit der zu zeigenden Aussage** beitragen. Zugleich kommen in jeder der Klassenstufen, d. h. auch in der Klassenstufe 5, Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge vor. Je höher die Klassenstufe, desto größer ist der Anteil der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge. In der Klassenstufe 8 enthält gut jeder fünfte Beweisversuch ein weitgehend korrektes und vollständiges Argumentgefüge. Einzelne, aber jeweils unter 3% der Teilnehmenden aus den Klassenstufen 6 bis 8 haben die zu zeigende Aussage bewiesen. In der Klassenstufe 5 ist in keinem der analysierten Beweisversuche ein Beweis formuliert.

Außerdem beschreiben die Analyseergebnisse, welche **Charakteristika** die Beweisversuche von Teilnehmenden der Mathematik-Olympiade auszeichnen. Beispielsweise wird in jeder der Klassenstufen 5 bis 8 ein großer Anteil der Beweisversuche durch die Existenz eines begründenden Ansatzes und damit einhergehend die Erkennbarkeit mindestens einer Beweisidee gekennzeichnet. Des Weiteren ist festzuhalten, dass je Klassenstufe in etwa 50% bis 80% aller Beweisversuche ein Beispiel vorkommt. In den Beweisversuchen mit Beispiel sind meistens zwei bis fünf Beispiele, oftmals systematische Beispiele und häufig das kleinste Beispiel angegeben. Die Darstellungsform der Beweisversuche ist in jeder der Klassenstufen 5 bis 8 besonders häufig numerisch-verbal.

In der Klassenstufe 5 sind fehlende Bearbeitungen häufig. Zudem erscheinen Beweisversuche mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge für beide Klassenstufen 5 und 6 besonders charakteristisch. Im Vergleich zu den Klassenstufen 5 und 6 enthalten die Beweisversuche aus den Klassenstufen 7 und 8 häufiger ein weitgehend korrektes Argumentgefüge. Zudem sind häufiger Reihenfolgebeispiele und die primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* ausgeführt. Auch sind der Anteil der Beweisversuche mit zwischen induktiver und deduktiver Schlussweise sowie der Anteil der Beweisversuche mit einer Darstellungsform mit ikonisch-numerischem Anteil größer. Dabei gibt es Hinweise darauf, dass diese Charakteristika der Beweisversuche aus den Klassenstufen 7, 8 miteinander zusammenhängen.

Auf der Analyse von Charakteristika der Beweisversuche aufbauend, beschreiben die Ergebnisse drittens, inwieweit **Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis** in den Beweisversuchen zu erkennen sind. Beispielsweise legen es viele Beweisversuche aus den Klassenstufen 5 und 6 nahe, dass für deren Weiterentwicklung der Frage, was zu beweisen ist, zumindest kurz Aufmerksamkeit geschenkt werden sollte. Ein relevanter Anteil der Beweisversuche aus den Klassenstufen 5 und 6 deutet ein Bewusstsein für die Notwendigkeit einer allgemeingültigen Begründung an, wobei für eine Weiterentwicklung der Beweisversuche zu einem Beweis die vorhandenen Ausführungen teils grundlegend zu überarbeiten wären. Oftmals wäre es erforderlich eine neue, ausbaubare Beweisidee zu generieren. Was die Darstellungsmittel betrifft, lässt sich in den Klassenstufen 5 und 6, aber auch in den Klassenstufen 7 und 8 insbesondere an die Nutzung verbaler und numerischer Darstellungsmittel anknüpfen. Dabei sind in den Beweisversuchen aus den Klassenstufen 7 und 8 im Mittel fortgeschrittenere Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen als in den Klassenstufen 5 und 6. Insbesondere kommt in den Beweisversuchen aus den Klassenstufen 7 und 8 häufiger ein weitgehend korrektes Argumentgefüge vor, sodass vorhandene Ausführungen bei der Entwicklung eines Beweises im Wesentlichen beibehalten werden können. Bei gut einem Fünftel der Beweisversuche aus der Klassenstufe 8 wären für deren Weiterentwicklung zu einem Beweis ausschließlich Verallgemeinerungen durchzuführen oder Details zu optimieren. Dabei erscheinen die Beweisversuche aus der Klassenstufe 8 mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis besonders heterogen, im Vergleich dazu in der Klassenstufe 5 am homogensten.

Im Rahmen der Auswertung der Ergebnisse der Codierung der Beweisversuche sind zwei **Klassifikationen** entstanden, die systematisch darstellen, welche Ausprägungen Beweisversuche mit Blick auf den Nachweis der Gültigkeit sowie mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis aufweisen können. Zudem zeigen sie Möglichkeiten auf, Beweisversuche diesbezüglich vergleichend einzuordnen, ohne alle deren Eigenschaften en détail zu analysieren.

9 Diskussion

In diesem Kapitel werden nacheinander das Design und die Durchführung der im Rahmen dieser Arbeit erfolgten Studie (Abschn. 9.1), die Vorgehensweise bei der Auswertung der erhobenen Daten (Abschn. 9.2) und die damit erzielten Ergebnisse (Abschn. 9.3) diskutiert. Dabei einfließende Überlegungen zu Grenzen dieser Arbeit und möglichen Anschlussfragen bieten Anregungen für zukünftige Forschungsarbeiten.

9.1 Diskussion des Designs und der Durchführung der Studie

Leistungen und Limitationen der vorliegenden Arbeit, die sich im Wesentlichen aus Aspekten des Designs und der Durchführung der Studie ergeben, werden in diesem Abschnitt beschrieben. Dabei wird zuerst auf die Untersuchungspopulation der Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 eingegangen (Abschn. 9.1.1). Anschließend stehen der Ort, Umfang, Ablauf und Zeitpunkt der Datenerhebung im Fokus (Abschn. 9.1.2). Die Datenerhebung wurde im Rahmen einer Regionalrunde der Mathematik-Olympiade durchgeführt. Dafür wurde in jeder der Klassenstufen 5 bis 8 eine der regulären Wettbewerbsaufgaben durch eine spezifisch fokussierte Beweisaufgabe ersetzt. Auf Stärken und Schwächen dieser Beweisaufgabe und die Fokussierung auf einen Aufgabenteil wird in einem nächsten Schritt eingegangen (Abschn. 9.1.3). Danach wird grundlegend die Fokussierung auf Beweisversuche, d. h. in schriftlicher Form vorliegender Bearbeitungen von Beweisaufgaben, diskutiert (Abschn. 9.1.4).

9.1.1 Die Untersuchungspopulation: Teilnehmerinnen und Teilnehmer der Mathematik-Olympiade in den Klassenstufen 5 bis 8

Die vorliegende Arbeit fokussiert auf Teilnehmerinnen und Teilnehmer der Mathematik-Olympiade. Teilnehmende der Mathematik-Olympiade sind eine interessante und wenig beachtete Untersuchungspopulation. Die Mathematik-Olympiade hat lange Tradition und ist heutzutage nicht mehr wegzudenken. Trotzdem steht die mathematikdidaktische Forschung diesbezüglich noch am Anfang (Abschn. 1.1, 6.1). Insbesondere sind der Autorin dieser Dissertation keine forschungsbasierten Arbeiten zu Beweiskompetenzen von Teilnehmenden der Mathematik-Olympiade bekannt.[160] Dabei handelt es sich um Schülerinnen und Schüler, denen man u. a. mit Blick auf den wissenschaftlichen Nachwuchs viel Beachtung schenken sollte. Die vorliegende Arbeit verbindet die etablierte mathematikdidaktische Forschung zum Beweisen mit der bisher kaum betrachteten Gruppe der Teilnehmenden der Mathematik-Olympiade.

Da sich die Ergebnisse dieser Arbeit demnach primär auf Teilnehmende der Mathematik-Olympiade beziehen, ist insbesondere bei der Verallgemeinerung der festgestellten Klassenunterschiede Vorsicht geboten. Es ist erstens zu beachten, dass es sich um eine Querschnitts- und

[160] Eine Ausnahme bildet eine erste eigene Untersuchung, die im Rahmen der wissenschaftlichen Hausarbeit im Lehramtsstudium durchgeführt wurde (Tebaartz, 2013; Tebaartz, 2015; Tebaartz & Lengnink, 2015; Abschn. 1.1).

nicht um eine Längsschnittstudie handelt, weshalb nicht von einer „Entwicklung“, sondern ausschließlich von „Unterschieden“ zu sprechen ist. Zweitens nehmen in Nordrhein-Westfalen (NRW), wo die Datenerhebung stattgefunden hat, üblicherweise viele Schülerinnen und Schüler aus der Klassenstufe 5, etwas weniger aus den Klassenstufen 6 und 7 und deutlich weniger aus der Klassenstufe 8 teil (Abschn. 2.2.2, 6.2). In Einklang mit dieser Tendenz standen für die Auswertung 1309 Beweisversuche aus der fünften, 919 aus der sechsten, 765 aus der siebten und 407 aus der achten Klassenstufe zur Verfügung (Abschn. 7.1). Bezogen auf die Gruppe der Teilnehmenden der Regionalrunde der Mathematik-Olympiade in NRW sind entsprechend den Ergebnissen der vorliegenden Arbeit Unterschiede zwischen den einzelnen Klassenstufen 5 bis 8 festzuhalten. Bei einer Verallgemeinerung der Ergebnisse ist jedoch Vorsicht geboten, weil eventuell eine unterschiedliche „Selektion“ der Teilnehmenden zu den festgestellten Differenzen beigetragen hat. Es kann sein, dass sich insbesondere in der Klassenstufe 8 fast nur besonders leistungsstarke Schülerinnen und Schüler für die Mathematik-Olympiade interessieren und die Qualifikation für die Regionalrunde schaffen, während insbesondere in der Klassenstufe 5 ein größerer Anteil der Lernenden an der Mathematik-Olympiade teilnimmt und vielleicht einem größeren Anteil der Teilnehmenden die Qualifikation für die Regionalrunde gelingt.

Des Weiteren lassen sich die Ergebnisse der vorliegenden Arbeit auch grundsätzlich nicht auf alle Schülerinnen und Schüler in den Klassenstufen 5 bis 8 ausweiten. Bei den Teilnehmenden der Regionalrunde der Mathematik-Olympiade handelt es sich um eine Positivauswahl. Die Teilnahme an der Regionalrunde beruht normalerweise auf Freiwilligkeit und setzt in der Regel eine gute Bearbeitung der Aufgaben der Schulrunde der Mathematik-Olympiade voraus (Abschn. 2.2). Da die teilnehmenden Schülerinnen und Schüler vermutlich einen gewissen Ehrgeiz mitbringen, ist zudem davon auszugehen, dass viele Teilnehmende versucht haben, die als Teil der Wettbewerbsklausur gestellte Beweisaufgabe so gut es geht zu bearbeiten. Die dabei entstandenen Beweisversuche geben einen Einblick, inwieweit leistungsstarken Lernenden aus den verschiedenen Klassenstufen der Beweis der gestellten Beweisaufgabe gelingen kann (Abschn. 8.5.1). So wie es auch Forschungsergebnisse von Moya et al. (2015) andeuten, lässt sich vermuten, dass das erhaltene Spektrum an Beweisversuchen größer ist als es etwa im Rahmen einer Erhebung in einer Schulklasse der Fall wäre. Durch die Fokussierung auf Teilnehmende der Mathematik-Olympiade zeigen die Ergebnisse besonders gut, was in den einzelnen Klassenstufen 5 bis 8 möglich ist, und veranschaulichen zugleich, dass die vorkommenden Beweisversuche sehr vielfältig sein können. Es ist davon auszugehen, dass einige der Ausprägungen der Beweisversuche der Teilnehmenden der Mathematik-Olympiade auch im Falle einer Untersuchung etwa mit Lernenden aus einer Schulklasse vorkämen, nur mit anderen Häufigkeitsverteilungen.

9.1.2 Ort, Umfang, Ablauf und Zeitpunkt der Datenerhebung

Die Beweisversuche von Teilnehmenden der Mathematik-Olympiade wurden im Rahmen der Regionalrunde der Mathematik-Olympiade im November 2016 in Nordrhein-Westfalen erhoben (Abschn. 6.2).

Die **Mathematik-Olympiade** ist für eine Erhebung von Beweisversuchen ein sehr geeignetes, authentisches Setting. Die Aufforderung, eine mathematische Allaussage zu beweisen, und der Anspruch der Formulierung eines korrekten und vollständigen, deduktiven Argumentgefüges stehen in Einklang mit den regulären Wettbewerbsanforderungen (Abschn. 2.3, 10.3).

Die **Regionalrunde der Mathematik-Olympiade** ist, anders als die Schulrunde, als Klausurwettbewerb konzipiert (Abschn. 2.2). Dadurch ist die Regionalrunde besser geeignet, um Durchführungsobjektivität zu gewährleisten. Zugleich ist das Spektrum an Teilnehmerinnen und Teilnehmern bei der Regionalrunde im Vergleich zur darauffolgenden Landesrunde größer. Die Regionalrunde ist somit im Vergleich zu nachfolgenden Wettbewerbsrunden besser geeignet, um einen Überblick über Charakteristika von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade zu erhalten. Diese Überlegung stützend sind die erhaltenen Beweisversuche äußerst vielfältig (z. B. Abschn. 8.2.2, 8.4.2, 9.5.3). Insgesamt erscheint die Regionalrunde der Mathematik-Olympiade also geeignet.

Für die Datenerhebung im Rahmen einer Regionalrunde wurde das Bundesland **Nordrhein-Westfalen** (NRW) gewählt. Die Autorin dieser Arbeit ist mit den dortigen Strukturen gut vertraut und die vorhandenen Strukturen sind für die Durchführung einer solchen Erhebung gut geeignet. Zugleich sind die Teilnehmerzahlen in der Regionalrunde einer jeden Mathematik-Olympiade in NRW so hoch, dass auf den Einbezug weiterer Bundesländer verzichtet werden konnte (Abschn. 2.2.2 zur Mathematik-Olympiade in NRW). Die Ergebnisse dieser Arbeit beschreiben folglich Charakteristika von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade in Nordrhein-Westfalen. Da alle Probandinnen und Probanden aus Nordrhein-Westfalen stammen und sich die curricularen Vorgaben in den verschiedenen Bundesländern unterscheiden, sind die Ergebnisse nicht unmittelbar auf Teilnehmende der Mathematik-Olympiade in der gesamten Bundesrepublik übertragbar.

Insgesamt wurden von gut $\mathbf{3400}$ **Teilnehmerinnen und Teilnehmern** der Mathematik-Olympiade in den Klassenstufen 5 bis 8 Beweisversuche erhoben (Abschn. 7.1, 10.2). Die Erhebung von Beweisversuchen vieler Teilnehmerinnen und Teilnehmer war sinnvoll, um sowohl das Spektrum an vorkommenden Ausprägungen als auch deren relative Häufigkeitsverteilungen differenziert beschreiben und Beweisversuche von Teilnehmenden unterschiedlicher Klassenstufen vergleichen zu können. Eine etwas kleinere Datenerhebung wäre dafür aber auch ausreichend gewesen. Die Regionalrunde der Mathematik-Olympiade wird in NRW separat in 55 Regionen durchgeführt. Die Verantwortlichen aus 44 Regionen haben sich auf die Anfrage der Autorin hin positiv zurückgemeldet, sodass es zu dieser großen Datenerhebung gekommen ist.

Für die Datenerhebung wurde in den 44 Regionen Nordrhein-Westfalens eine spezifisch fokussierte Beweisaufgabe als Teil der regulären Wettbewerbsklausur der Regionalrunde gestellt. Die teilnehmenden Schülerinnen und Schüler sollten diese Aufgabe, so wie die gesamte Wettbewerbsklausur, allein bearbeiten. Bei der Mathematik-Olympiade wird grundsätzlich viel Wert auf eine eigenständige Bearbeitung und einen fairen Ablauf gelegt (Abschn. 2.2). Um eine möglichst objektive **Durchführung der Datenerhebung** zu gewährleisten, wurden die Verantwortlichen in den einzelnen Regionen zusätzlich explizit darum gebeten, in der Klausur keine

Fragen zu beantworten, die sich auf die Bearbeitung der Beweisaufgabe beziehen (Abschn. 6.2.1). Da die Regionalrunde nicht in allen Regionen NRWs am gleichen Tag, sondern binnen dreier aufeinanderfolgender Tage stattfindet, wurden die Verantwortlichen entsprechend der üblichen Praxis explizit aufgefordert, sämtliche Dokumente, d. h. auch die Aufgabenblätter, am Ende der Klausur einzusammeln und Lösungen frühestens am Samstagabend herauszugeben (Abschn. 2.2.2, 6.2.1). Inwieweit tatsächlich keinerlei Hilfestellungen gegeben wurden und z. B. auch keine Kommunikation zwischen Teilnehmenden aus verschiedenen Regionen stattgefunden hat, lässt sich nicht nachprüfen.

Die Regionalrunde der Mathematik-Olympiade findet immer **Mitte November**, also in etwa in der Mitte des ersten Schulhalbjahres statt. Dies bedeutet, dass sich die Teilnehmenden aus den Klassenstufen 5 bis 8 eher am Anfang als am Ende ihrer jeweiligen Klassenstufe befanden. Dies ist unter anderem beim Vergleich der Ergebnisse der vorliegenden Arbeit mit denen anderer empirischer Studien und mit den curricularen Vorgaben in Abschnitt 9.3 zu bedenken. Zudem impliziert die frühe Datenerhebung, dass vielleicht auch viele Teilnehmende aus der Klassenstufe 8 zum Zeitpunkt der Datenerhebung noch nicht allzu vertraut mit elementarer Algebra waren. Dabei wurde durch die Auswahl der Klassenstufen 5 bis 8 unter anderem versucht, Beweisversuche von Lernenden mit unterschiedlichen Darstellungsmöglichkeit miteinzubeziehen (Abschn. 1.1, 6.1). In der Klassenstufe 8 beträgt der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil rund 13%, bei der Mehrheit der Beweisversuche ist die Darstellungsform *numerisch-verbal* (Abschn. 8.5.2.4). In knapp 70% der Beweisversuche aus der Klassenstufe 8 ist mindestens ein Beispiel anführt (Abschn. 8.2.5.3). Die Ergebnisse der vorliegenden Arbeit lassen offen, inwieweit diese und weitere Eigenschaften der Beweisversuche damit zusammenhängen, dass es sich um Teilnehmende am Anfang der Klassenstufe 8 handelte. Es stellt sich die Frage, inwieweit die Beweisversuche bei den Teilnehmenden der Regionalrunde aus der Klassenstufe 9, d. h. nach der Klassenstufe 8, anders ausgefallen wären. Diese Frage ließe sich auf Grundlage der weiteren bereits erhobenen Daten beantworten (Abschn. 10.2, 10.4.1).

9.1.3 Die Fokussierung auf eine Beweisaufgabe aus dem Themenbereich der Teilbarkeit ganzer Zahlen und die Konzeption dieser Aufgabe

Die im Rahmen dieser Arbeit untersuchten Beweisversuche basieren alle auf einer Beweisaufgabe aus dem Themenbereich der Teilbarkeit ganzer Zahlen (Abschn. 1.1, 6.2). Die Ergebnisse der vorliegenden Arbeit beziehen sich also auf den **Themenbereich der Teilbarkeit ganzer Zahlen** und lassen nicht unbedingt Schlussfolgerungen auf Beweisversuche zu Aussagen aus anderen Themengebieten, wie z. B. der Elementargeometrie, zu, wie es auch Forschungsergebnisse von Brunner et al. (2022) nahelegen. Der Bereich der Teilbarkeit ganzer Zahlen wurde ausgewählt, u. a. weil es in diesem Themenbereich möglich ist, auf sehr elementarem Niveau und ohne weitreichende fachinhaltliche Vorkenntnisse Beweise zu führen (Abschn. 1.1, 6.1). In Einklang mit dieser Überlegung haben zu der im Rahmen dieser Arbeit fokussierten Allaussage aus dem Bereich der Teilbarkeit ganzer Zahlen in jeder der Klassenstufen 5 bis 8 zumindest einzelne Teilnehmende ein weitgehend korrektes und vollständiges Argumentgefüge formuliert. In je-

der der Klassenstufen 6 bis 8 ist einzelnen Lernenden ein Beweis gelungen (Abschn. 8.5.1). Zugleich ist in jeder der Klassenstufen ein großes Spektrum an Beweisversuchen entstanden (z. B. Abschn. 8.2.2, 8.4.2, 8.5.3), was ein Hinweis darauf ist, dass die zu zeigende Aussage gut gewählt war. Dafür spricht auch, dass die gewählte Aussage in mehrerlei Hinsicht für Allaussagen aus dem Themenbereich der Teilbarkeit ganzer Zahlen charakteristisch ist. Beispielsweise sind verschiedene Vorformen mathematischer Beweise und Beweise ohne Variablennutzung möglich. Ein Beweis ohne Variablennutzung, basierend auf inhaltlich-anschaulichen Überlegungen erscheint dabei tendenziell einfacher als eine rein symbolische Herangehensweise. Zudem können bei der gewählten Aussage diverse Beweisideen verwendet werden, die sich auch bei zahlreichen anderen Allaussagen aus dem Themenbereich der Teilbarkeit ganzer Zahlen anbieten (Abschn. 6.2.2). Weiterhin ist hervorzuheben, dass bei der Konstruktion der Aufgabe gezielt Schwierigkeitsparameter kontrolliert und u. a. die Anforderungen im Bereich des Bereichswissens und im Umgang mit Algebra niedrig gehalten wurden, um eine Verzerrung über nicht im Fokus stehende Schwierigkeiten zu verhindern (Abschn. 6.2.2).

An der **Aufgabenstellung** diskutieren lässt sich das angegebene Beispiel. Um eine korrekte Interpretation der Aufgabenstellung zu unterstützen, wurde der zentrale Ausdruck „vier aufeinanderfolgende Zahlen“ an einem begründet ausgewählten Beispiel veranschaulicht („z. B. $5 + 7 + 9 + 11 = 32$“ (Abb. 72 in Abschn. 6.2.2)). In jeder der Klassenstufen 5 bis 8 kommen Beweisversuche mit einem Hinweis auf einen grundlegenden Fehler vor, die darauf hindeuten, dass es wichtig und gut war, ein Beispiel anzugeben. Durch dieses Beispiel haben vielleicht weniger Schülerinnen und Schüler in ihrem Beweisversuch ein Beispiel angegeben als es der Fall gewesen wäre, wenn in der Aufgabenstellung kein Beispiel gestanden hätte (Abschn. 8.2.3.1). Zudem kann das angegebene Beispiel u. a. dazu beigetragen haben, dass in manchen Beweisversuchen basierend auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* kein Induktionsanfang vorkommt oder als Induktionsanfang ein anderes als das kleinste Beispiel angeführt wurde. Daher wäre es anstelle eines Beispiels zu der zu zeigenden Aussage wahrscheinlich noch besser gewesen, etwa so wie in Abbildung 175 möglichst fokussiert den Ausdruck „aufeinanderfolgende ungerade Zahlen“ anhand sonstiger Beispiele zu erklären.

Formulierung des Aufgabenteils d) ohne Angabe eines Beispiels zu der zu zeigenden Aussage

Klassenstufen 5, 6	Klassenstufen 7, 8
Defne berechnet Summen von aufeinanderfolgenden ungeraden Zahlen. Sie behauptet: „Die Summe von **vier <u>aufeinanderfolgenden</u> ungeraden Zahlen** ist immer durch 4 teilbar.“ Hilf Denise zu begründen, warum ihre Behauptung richtig ist.	Denise berechnet Summen von aufeinanderfolgenden ungeraden Zahlen. Sie behauptet: „Die Summe von **vier <u>aufeinanderfolgenden</u> ungeraden Zahlen** ist immer durch 8 teilbar.“ Hilf Denise zu beweisen, dass ihre Behauptung richtig ist.
Hinweis: Beispielsweise handelt es sich bei den Zahlen $9, 11, 13$ um drei aufeinanderfolgende ungerade Zahlen, die Zahlen $9, 11, 15$ hingegen sind keine drei aufeinanderfolgenden ungeraden Zahlen.	*Hinweis:* Beispielsweise handelt es sich bei den Zahlen $9, 11, 13$ um drei aufeinanderfolgende ungerade Zahlen, die Zahlen $9, 11, 15$ hingegen sind keine drei aufeinanderfolgenden ungeraden Zahlen.

Abbildung 175: Formulierung des Aufgabenteils d) ohne Angabe eines Beispiels zu der zu zeigenden Aussage

Entsprechend Abbildung 175 haben die Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5, 6 im Vergleich zu denen in den Klassenstufen 7, 8 eine leicht unterschiedliche

Aufgabenstellung erhalten: Teilnehmende in den Klassenstufen 5 und 6 sollten für die Summe von vier aufeinanderfolgenden ungeraden Zahlen die Teilbarkeit durch vier, Teilnehmende in den Klassenstufen 7 und 8 die Teilbarkeit durch acht zeigen. Zudem wurde als Operator in den Klassenstufen 5, 6 „begründen“, in den Klassenstufen 7, 8 „beweisen“ verwendet. Damit die Beweisversuche der Teilnehmenden aus den verschiedenen Klassenstufen vergleichbar sind, wurde für alle Klassenstufen eine möglichst ähnliche Aufgabenstellung gewählt. Da auf die einzelnen Klassenstufen abgestimmte Aufgaben ein wesentliches Merkmal der Mathematik-Olympiade sind und die für die Studie im Rahmen dieser Arbeit entwickelte Aufgabe zu den regulären Wettbewerbsanforderungen passen sollte, wurden die genannten zwei leicht verschiedenen Versionen gestellt (Abschn. 6.2.2). Dies hat zur Folge, dass eventuell manche Unterschiede zwischen den Beweisversuchen aus den Klassenstufen 5, 6 und denen aus den Klassenstufen 7, 8 auf die leicht veränderte Aufgabenstellung zurückzuführen sind. Unter anderem ist, so wie bei der Aufgabenkonstruktion vermutet, der Anteil der Beweisversuche mit der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* an allen Beweisversuchen mit begründendem Ansatz in den Klassenstufen 5, 6 kleiner als in den Klassenstufen 7, 8, auch wenn man die Beweisversuche mit begründendem Ansatz und Reihenfolgebeispielen oberhalb des (übrigen) Argumentgefüges fokussiert (Abschn. 6.2.2.2, 8.2.2.2, 8.2.6.4). Was die Wahl des Operators betrifft, weist eine Untersuchung mit Erstsemesterstudierenden darauf hin, dass die Wahl des Operators in Beweisaufgaben durchaus Auswirkungen auf die Bearbeitungen dieser Aufgaben hat. Die Ergebnisse der Studie deuten an, dass beim Operator „begründen“ bei manchen Beweisaufgaben der Anteil an Beweisversuchen mit begründendem Ansatz vergleichsweise groß ist und im Falle eines Beweisversuchs mit begründendem Ansatz tendenziell viele Wörter und selten Buchstabenvariablen verwendet werden (Kempen et al., 2016, 2019). Inwieweit Aspekte der Aufgabenstellung, wie das angegebene Beispiel, der vorgegebene Teiler und der gewählte Operator, tatsächlich die Beweisversuche der Teilnehmenden beeinflusst haben, lässt sich nicht klären.

Im Rahmen der vorliegenden Arbeit wurden Beweisversuche zu **einer einzigen Allaussage** untersucht. Diese Allaussage wurde, wie beschrieben, so konstruiert, dass sie in mehrerlei Hinsicht für den Themenbereich der Teilbarkeit ganzer Zahlen charakteristisch erscheint (Abschn. 6.2.2.2). Trotzdem können ausgehend von den Ergebnissen der Analyse der Beweisversuche zu dieser Aufgabe nicht uneingeschränkt allgemeine Aussagen über Beweisversuche der Probanden zu Allaussagen mit unendlichem Gültigkeitsbereich aus diesem Themengebiet getroffen werden, wie etwa die vorhergehenden Ausführungen zu Details der formulierten Aufgabenstellung und Forschungsergebnisse von Vogt (2021) vermuten lassen.

Weiterhin ist grundlegend zu bedenken, dass es sich bei der in dieser Arbeit fokussierten Aussage um eine **Allaussage mit unendlichem Gültigkeitsbereich** handelt. Es stellt sich die Frage, inwieweit Beweisversuche der Teilnehmenden zu Allaussagen mit endlichem Gültigkeitsbereich und solche zu Existenzaussagen ähnliche Eigenschaften aufweisen. Die Auswertung weiterer bereits erhobener Daten würde diesbezüglich Einblicke geben (Abschn. 1.1, 6.2.2, 10.2).

Bei den analysierten Beweisversuchen handelt es sich um Bearbeitungen des **Aufgabenteils d)** einer Aufgabe bestehend aus vier voneinander unabhängigen Beweisaufgaben (Abschn. 6.2.2).

Die in dieser Arbeit fokussierte Beweisaufgabe wurde als letzter Aufgabenteil gestellt, weil sie im Vergleich zu den anderen konstruierten Beweisaufgaben von den objektiven Anforderungsstrukturen her am anspruchsvollsten ist. Die Aufgabenteile a) bis d) wurden so gestaltet, dass die Beschäftigung mit einzelnen Aufgabenteilen möglichst keine Vor- oder Nachteile für die Bearbeitung eines anderen Aufgabenteils mit sich bringt und die Bearbeitungen der Aufgabenteile separat analysiert werden können. Es ist zu vermuten, dass viele Teilnehmende die Aufgabenteile a) bis d) in der angegebenen Reihenfolge bearbeitet haben, u. a. da die regulär vom Aufgabenausschuss des Mathematik-Olympiaden e. V. gestellten Aufgaben teils aus aufeinander aufbauenden Aufgabenteilen bestehen. Eventuell haben manche Teilnehmende etwa wegen Schwierigkeiten bei einem der Aufgabenteile a) bis c) aufgegeben, so dass der Anteil an fehlenden Bearbeitungen kleiner ausgefallen wäre, wenn der in dieser Arbeit fokussierte Aufgabenteil als Teil a) herausgegeben worden wäre. Zudem ist es z. B. denkbar, dass sich manche Teilnehmende in diesem Fall länger und intensiver mit dem Aufgabenteil beschäftigt hätten, so dass deren Beweisversuche etwas anders ausgefallen wären.

9.1.4 Die Fokussierung auf Beweisversuche

Wie vielfach erwähnt, liegt der Fokus der vorliegenden Arbeit auf Beweisversuchen, das heißt in schriftlicher Form vorliegenden Bearbeitungen von Beweisaufgaben. Die Fokussierung auf Beweisversuche steht in Einklang mit dem Anliegen, einen Überblick zu erhalten, inwieweit Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 der Beweis einer mathematischen Allaussage mit unendlichem Gültigkeitsbereich gelingt, und welche Charakteristika deren Beweisversuche auszeichnen. Auch lassen sich anhand von Beweisversuchen Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis erkennen. Beispielsweise sind es mit Blick auf eine Weiterentwicklung eines Beweisversuchs grundlegende Informationen, ob es sich um einen Beweisversuch mit begründendem oder mit prüfendem Ansatz handelt, inwieweit ein weiterentwickelbares Argumentgefüge erkennbar ist und was die primäre Beweisidee ist (insb. Abschn. 8.2.6, 8.3.4, 8.5.3). Zugleich lässt die Fokussierung auf Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis teils viel Interpretationsspielraum, weil die erhobenen Daten keine Einsicht in die Beweisprozesse und die den Beweisversuchen zugrundeliegenden Denkweisen geben. Beispielsweise ist im Falle fehlender Bearbeitungen unklar, inwieweit die Lernenden die Aufgabenstellung gelesen und korrekt interpretiert haben und wie die Lernenden ggf. mit dem Aufgabenteil zurechtkämen (Abschn. 8.5.3.1). Zudem lassen die Ausführungen bei einigen Beweisversuchen offen, ob sie sich auf eine ganz korrekte Interpretation der zu zeigenden Aussage beziehen (Abschn. 8.2.6.2). Weiterhin können Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge durchaus inhaltliche Überlegungen zugrunde liegen, die potenziell gute Anknüpfungspunkte für die Weiterentwicklung des Beweisversuchs bilden, aber nicht klar erkennbar sind (Abschn. 8.2.6.7, 8.5.3.2; Abels & Knipping, 2020). Bei Beweisversuchen mit prüfendem Ansatz stellt sich mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis in der Regel die Frage, inwieweit den Lernenden die Unzulänglichkeit von Beispielen für den Beweis der zu zeigenden Aussage bewusst ist (Abschn. 3.3.8, 8.3.2.3, 10.2). Außerdem kann nach Krumsdorf (2017) konkret Gesagtes allgemeiner gedacht sein und eine allge-

mein formulierte Aussage muss nicht unbedingt mit einer subjektiven Realisierung des Allgemeingültigen einhergehen, so dass grundsätzlich bei der Interpretation sowohl im konkreten gehaltener als auch allgemein formulierter Aussagen Vorsicht geboten ist (Abschn. 5.2.2).

Um eine angemessene Interpretation der Beweisversuche zu unterstützen, haben alle Teilnehmenden direkt im Anschluss an die Wettbewerbsklausur einen Fragebogen erhalten, der insbesondere eine Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche umfasst (Abschn. 6.2.1, 10.2). Die Erstellung von Videodateien zu Bearbeitungsprozessen wäre im Rahmen des authentischen Settings der Wettbewerbsklausur nicht in sinnvoller Form möglich gewesen. Ergänzende Interviews nach der Wettbewerbsklausur erschienen aus organisatorischen Gründen schwierig, zumal die wenigen Teilnehmenden vor der Klausur, d. h. unabhängig von ihren Beweisversuchen hätten ausgewählt werden müssen. Schließlich hätte eine zusätzliche Analyse von Video- oder Audiodateien den Rahmen dieser Dissertation überschritten.

9.2 Diskussion der Vorgehensweise bei der Datenauswertung

Für die Auswertung der erhobenen Beweisversuche wurde das Verfahren der **qualitativ-strukturierenden Inhaltsanalyse** nach Schreier (2012, 2014) angewendet (Kap. 7). Dieses passt zum Ziel einer vergleichenden Deskription sowie zu den erhobenen Daten und den vorhandenen theoretischen Grundlagen. Schließlich handelt es sich um ein datenreduzierendes Verfahren, das sich für eine valide, systematische und detaillierte Beschreibung sowie einen Vergleich umfangreichen Datenmaterials eignet. Zudem ist es im Rahmen einer qualitativ-strukturierenden Inhaltsanalyse möglich und üblich, an vorhandene Theorien und Forschungsergebnisse anzuknüpfen, aber auch neue zu generieren.

Bei der Durchführung der qualitativ-strukturierenden Inhaltsanalyse wurden zunächst von den 3400 vorliegenden Beweisversuchen mittels einer bewussten **Stichprobenziehung** und der Ziehung einer Zufallsstichprobe Beweisversuche für die Auswertung ausgewählt (Abschn. 7.1, 7.2.2.1, 7.3.3). Diese doppelte Stichprobenziehung ermöglichte es, sowohl einen Überblick über das Spektrum der erhobenen Beweisversuche zu erhalten als auch die Häufigkeitsverteilungen der Eigenschaften der Beweisversuche von Teilnehmenden in den einzelnen Klassenstufen 5 bis 8 zu beschreiben und zu vergleichen. Zudem war die bewusste Stichprobenziehung günstig, um ein Kategoriensystem zu entwickeln, das sich für alle zu untersuchenden Beweisversuche eignet.

Kern einer jeden qualitativ-strukturierenden Inhaltsanalyse ist die Anwendung eines **Kategoriensystems** auf das gesamte zu untersuchende Material. Die Güte des Kategoriensystems ist folglich von zentraler Bedeutung für die Qualität der damit erhaltenen Ergebnisse. Deshalb wird in Abschnitt 9.2.1 auf das Kategoriensystem näher eingegangen.

Die weitere Auswertung der durch die Anwendung des Kategoriensystems erhaltenen Codierungen führte zu einer **Klassifikation der Beweisversuche** mit Blick auf den Nachweis der Gültigkeit und einer Klassifikation der Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis (insb. Abschn. 7.2.1.6, 7.3.8, 8.5.1, 8.5.3). Der Nutzen dieser Klassifikationen wird in Abschnitt 9.2.2 thematisiert.

Das **Kategoriensystem** inklusive des damit einhergehenden Begriffsnetzes und die beiden **Klassifikationen** können als die wesentlichen Beiträge der vorliegenden Dissertation zur mathematikdidaktischen Theoriebildung und Diskussion angesehen werden.

9.2.1 Diskussion des Kategoriensystems

Grundlegendes **Charakteristikum** des im Rahmen dieser Arbeit entwickelten Kategoriensystems ist die eindeutige Unterscheidung der Oberkategorien *begründender Ansatz*, *prüfender Ansatz*, *weder begründender noch prüfender Ansatz*, *nicht eindeutig kategorisierbarer Ansatz*, *fehlend*. Diese Unterscheidung knüpft an vorhandene mathematikdidaktische Konzepte an und ermöglicht, das gesamte Kontinuum an denkbaren und vorhandenen Beweisversuchen zu Allaussagen global zu strukturieren (Abschn. 3.3.2). Zudem ist die getrennte Analyse verschiedener Eigenschaften von Beweisversuchen als Charakteristikum des Kategoriensystems hervorzuheben. Durch diese getrennte Analyse von Eigenschaften, wie z. B. der primären Beweisidee und der Darstellungsform, lassen sich Beweisversuche je nach Erkenntnisinteresse aus verschiedenen Perspektiven beschreiben. Nebenbei können durch die getrennte Analyse verschiedener Eigenschaften Zusammenhänge zwischen diesen Eigenschaften untersucht werden. Insbesondere mit Blick auf die mathematikdidaktische Theoriebildung wäre es interessant zu beforschen, inwieweit sich die im Rahmen dieser Arbeit festgestellten Zusammenhänge auch in Beweisversuchen anderer Zielgruppen und in Beweisversuchen zu anderen Allaussagen zeigen.

Das Kategoriensystem dieser Arbeit wurde **deduktiv-induktiv entwickelt**. Es knüpft an zahlreiche vorhandene mathematikdidaktische Theorien und Kategoriensysteme an. Zugleich berücksichtigt es die Grenzen des allein anhand von Beweisversuchen Sichtbarem und passt zu den vielfältigen Beweisversuchen von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8. Das Kategoriensystem wurde auf Intercoderreliabilität überprüft und erwies sich in diesem Rahmen als geeignet (Abschn. 7.3.7). Die bei der Entwicklung des Kategoriensystems getroffenen Überlegungen und Entscheidungen wurden offengelegt und begründet (Abschn. 3.3, 7.3). Während des gesamten Entwicklungsprozesses des Kategoriensystems wurden kontinuierlich Teile des Systems mit anderen Wissenschaftlerinnen und Wissenschaftlern diskutiert. Es erfolgten regelmäßige Dissertationsbesprechungen mit der Betreuerin dieser Arbeit, Präsentationen und Diskussionen im Doktorandenkolloquium des Instituts für Didaktik der Mathematik der JLU Gießen, Vorträge auf Jahrestagungen der Gesellschaft für Didaktik der Mathematik, bei der „Conference on Problem Solving in Mathematics Education“ (ProMath) im Jahr 2015 und beim „International Congress of the World Federation of National Mathematics Competitions“ im Jahr 2018. **Am Kategoriensystem hervorzuheben** ist, dass für eine transparente, einheitliche und vergleichbare Beschreibung der Vollständigkeit von Argumentgefügen detaillierte Erwartungshorizonte entwickelt wurden (Abschn. 3.3.4). Außerdem ist festzuhalten, dass ausschließlich ein kleiner Anteil der analysierten Beweisversuche in eine Restkategorie, wie z. B. *nicht eindeutig kategorisierbarer Ansatz*, eingeordnet werden musste (z. B. Abschn. 8.1). Auch kommt es kaum vor, dass sich Beweisversuche in einzelnen Unterkategorien häufen, ohne dass dies eine besondere Aussagekraft hat (Kap. 8). Eine Ausnahme bilden diesbezüglich die zahlreichen Beweisversuche mit der Darstellungsform *numerisch-verbal*. Wie bei der Einführung des Kategoriensystems betont, ist der Exaktheitsanspruch bei der Analyse der

Darstellungsmittel und der Darstellungsform im Vergleich zu allen anderen Analysen reduziert, wenngleich bei der Berechnung der Intercoderreliabilität fast durchgehend sehr gute Werte vorliegen (Abschn. 3.3.11, 7.3.7). Die vorgenommenen Abgrenzungen der verschiedenen Darstellungsformen sind vor dem Hintergrund des Kontinuums von rein numerisch bis hin zu rein verbal dargestellten Beweisversuchen sinnvoll und eine detailliertere Differenzierung zwischen verschiedenen Beweisversuchen mit der Darstellungsform *numerisch-verbal* ist mit Blick auf Anknüpfungspunkte für die Weiterentwicklung dieser Beweisversuche nicht erforderlich. Die Darstellungsmittel und die Darstellungsform könnten aber noch einmal in einem separaten Forschungsprojekt mit einer neuen theoretischen Fundierung und darauf aufbauenden Kategorisierung untersucht werden, um diesbezüglich fundierte und detaillierte Erkenntnisse zu erhalten.

Im Zuge der Entwicklung des Kategoriensystems ist ein **Begriffsnetz** für die Beschreibung von Beweisversuchen zu Allaussagen entstanden. Dieses umfasst eher grundlegende Begriffe wie „Beweis“, „Beweisversuch“, „Beweisversuch mit begründendem Ansatz“ und „Begründung“ bis hin zu Details, wie etwa der Unterscheidung zwischen „grundlegenden Fehlern“, „inhaltlichen Fehlern“ und „Formulierungsfehlern“. Über die genaue Bedeutung der zentralen Begriffe „Beweis“ und „Begründung“ besteht national wie international keine Einigkeit, so dass es erforderlich war, diese für den Gebrauch im Rahmen der vorliegenden Arbeit zu konkretisieren (Abschn. 3.1.2, 3.3.2, 3.4). Zudem war es erforderlich, etwa Termini zur Beschreibung der Schlussweise in Beweisversuchen mit begründendem Ansatz, bei denen diese nicht rein deduktiv ist, neu einzuführen. Das entstandene Begriffsnetz kann als ein Vorschlag verstanden werden, der eine differenzierte Beschreibung von Beweisversuchen zu Allaussagen ermöglicht und zur Diskussion um zentrale Begrifflichkeiten beiträgt. Die auf diesem Begriffsnetz basierenden Kategorien wurden anhand einer Vielzahl authentischer Beweisversuche von Schülerinnen und Schülern aus den Klassenstufen 5 bis 8 illustriert und konkretisiert. Diese Beweisversuche können sowohl in der Arbeit mit Schülerinnen und Schülern als auch in der Aus- und Fortbildung von Lehrkräften eingesetzt werden.[161]

Das Kategoriensystem wurde auf die Analyse der im Rahmen dieser Arbeit erhobenen Beweisversuche aus den Klassenstufen 5 bis 8 abgestimmt. Da Schülerinnen und Schüler zu Beginn der Klassenstufen 5 bis 7 üblicherweise nicht mit Buchstabenvariablen vertraut sind, wird nicht bei der Analyse der Korrektheit von Beweisversuchen, sondern im Rahmen der Beschreibung der verwendeten Darstellungsmittel miterfasst, inwieweit vorkommende Variablen angemessen definiert sind (Abschn. 3.3.4, 3.3.11). Bei der Verwendung des entwickelten Kategoriensystems für die Analyse von Beweisversuchen anderer **Zielgruppen**, wie z. B. von Studierenden, kann es sinnvoll sein, das Kategoriensystem diesbezüglich zu modifizieren. Abgesehen davon ist das Kategoriensystem für die Analyse beliebiger Beweisversuche zu der in dieser Arbeit fokussierten Allaussage geeignet.

[161] Vielen Dank an Prof. Dr. Katja Lengnink für diesen Hinweis.

Für die Nutzung des Kategoriensystems bei **anderen Allaussagen mit unendlichem Gültigkeitsbereich** sind die aufgabenspezifischen Teile anzupassen. Das sind insbesondere die unterschiedenen Beweisideen und die Erwartungshorizonte zu korrekten und vollständigen Argumentgefügen.

Neben dem Kategoriensystem zum in dieser Arbeit fokussierten Aufgabenteil d) der spezifisch fokussierten Aufgabe, die Teilnehmende der Mathematik-Olympiade im Rahmen einer Wettbewerbsklausur erhalten haben, wurden auch Ansätze für die Kategoriensysteme zu den Aufgabenteilen a) bis c) konzipiert (Abschn. 7.3). Im Aufgabenteil c) sollten die Teilnehmenden eine Allaussage mit endlichem Gültigkeitsbereich aus dem Themenbereich der Teilbarkeit ganzer Zahlen beweisen, in den Aufgabenteilen a), b) jeweils eine Existenzaussage (Abschn. 6.2). Die Kategoriensysteme zu den **vier Aufgabenteilen a) bis d)** wurden möglichst analog angelegt, um intraindividuelle Vergleiche von Beweisversuchen zu verschiedenen Aussagen zu ermöglichen (Abschn. 7.3). Ein Kategoriensystem, das sich für die Analyse von Beweisversuchen zu Existenz- und Allaussagen eignet, ist der Autorin dieser Arbeit nicht bekannt. Die Fertigstellung und Anwendung der Kategoriensysteme zu allen Aufgabenteilen würde einen ersten Schritt hin zur Entwicklung eines solchen aussagenübergreifenden Analysewerkzeugs und zu aussagekräftigen intraindividuellen Vergleichen von Beweisversuchen zu verschiedenen Aussagen darstellen. Außerdem wäre es möglich, die Ergebnisse der Verwendung dieser Kategoriensysteme mit denen aus anderen Studien zu vergleichen.

9.2.2 Diskussion der Klassifikationen

Die weitere Auswertung der im Rahmen dieser Arbeit erhaltenen Codierungen führte zu einer Klassifikation der Beweisversuche mit Blick auf den Nachweis der Gültigkeit und einer Klassifikation der Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis (insb. Abschn. 7.2.1.6, 7.3.8, 8.5.1, 8.5.3). Die beiden Klassifikationen geben einen geordneten Überblick darüber, wie Beweisversuche zu der im Rahmen dieser Arbeit fokussierten Allaussage mit Blick auf den Nachweis der Gültigkeit dieser Aussage und mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis ausgeprägt sein können. Die Klassifikationen sind zudem geeignet, um Beweisversuche zu der im Rahmen dieser Arbeit fokussierten Allaussage diesbezüglich zusammenfassend zu beschreiben und zu vergleichen. Außerdem können sie eine gute Grundlage für qualifizierte Rückmeldungen an Lernende und eine gezielte Unterstützung von Schülerinnen und Schülern darstellen.

Im Vergleich zu einer detaillierten Analyse von Beweisversuchen anhand des Kategoriensystems geht die Einordnung von Beweisversuchen anhand der Klassifikationen schneller, erscheint dadurch praxistauglicher. Außerdem erscheint eine Einordnung von Beweisversuchen anhand der Klassifikationen zielgerichteter, wenn der Nachweis der Gültigkeit der zu zeigenden Aussage oder Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche im Fokus stehen. Eine mögliche Herangehensweise an die Analyse von Beweisversuchen könnte darin bestehen, diese anhand der Klassifikationen zunächst übergeordnet einzuordnen und zu sortieren und anschließend gezielt Eigenschaften von Beweisversuchen mithilfe des Kategoriensystems näher zu beschreiben.

Innovativ ist an den beiden Klassifikationen insbesondere die mehrdimensionale Sichtweise, die einen multiperspektivischen Blick auf die Beweisversuche von Lernenden und deren Förderung ermöglicht. Außerdem werden die Klassifikationen dem Kontinuum an denkbaren und vorkommenden Ausprägungen von Beweisversuchen gerecht, u. a. da Grenzbereiche explizit ausgewiesen sind. Die Klassifikationen wurden angelehnt am Konzept der empirisch begründeten Typenbildung nach Kluge (1999, 2000) entwickelt. Im Vergleich zu einer Typenbildung als besonderer Form einer Klassifikation wurde aber nicht versucht, Ausprägungen von Beweisversuchen klar zu trennen, die nicht klar zu trennen sind. Auch wurde es soweit möglich vermieden, Zusammenfassungen vorzunehmen, durch die wesentliche Informationen verloren gehen würden. Dadurch scheinen die Klassifikationen die Beweisversuche in angemessenerer Form zu beschreiben als eine Typologie oder eine eindimensionale ordinale Kategorisierung dies leisten könnte. Zugleich sind die entwickelten Klassifikationen etwa im Vergleich zu eindimensionalen ordinalen Kategorisierungen komplexer und die statistischen Auswertungsmöglichkeiten in Bezug auf die mit den Klassifikationen erhaltenen Ergebnisse eingeschränkter.

Die mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage entwickelte Klassifikation lässt sich ebenfalls bei anderen Allaussagen nutzen. Hingegen ist die Klassifikation von Beweisversuchen mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis insbesondere durch die Berücksichtigung der primären Beweisideen aufgabenspezifisch. Es könnte in ergänzenden Studien für weitere Aufgaben eine solche Klassifikation entwickelt werden, um im Vergleich der Klassifikationen übergreifende Muster zu sehen und davon ausgehend zu einer aufgabenunspezifischen Klassifikation zu kommen. Dabei stellt sich die grundlegende Frage, inwieweit auch in anderen Studien primäre Beweisideen mit weiteren Eigenschaften von Beweisversuchen zusammenhängen und als wesentlich für eine solche Klassifikation erscheinen.

9.3 Diskussion der erhaltenen Ergebnisse

Die Ergebnisse der vorliegenden Arbeit beschreiben entsprechend den Forschungsfragen, inwieweit es Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 gelungen ist, die vorgegebene mathematische Allaussage aus dem Bereich der Teilbarkeit ganzer Zahlen zu beweisen. Zudem legen sie dar, welche Charakteristika die Beweisversuche der Teilnehmenden auszeichnen und welche Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind. Dabei zeigen sie auf, welche Gemeinsamkeiten und Unterschiede mit Blick auf die verschiedenen Klassenstufen bestehen (insb. Abschn. 6.1, 8.5). Beweisversuche von Teilnehmenden der Mathematik-Olympiade wurden bisher nicht beforscht. Außerdem wurden Beweisversuche von Lernenden in den Klassenstufen 5, 6 und auch von solchen am Anfang der Klassenstufe 7 bisher kaum untersucht. Die Ergebnisse der vorliegenden Arbeit ergänzen somit den bisherigen Forschungsstand.

Zugleich bewegt sich diese Arbeit in einem etablierten Forschungsfeld. Insbesondere mit Lernenden ab der Klassenstufe 7 werden seit vielen Jahren empirische Studien zum Führen von Beweisen durchgeführt (Kap. 5). Deshalb werden in diesem Abschnitt die Ergebnisse der mit

Blick auf Teilnehmende der Mathematik-Olympiade gewonnenen Ergebnisse mit denen anderer Studien verglichen. Es wird beschrieben, inwieweit sich in Hinblick auf die fokussierte Zielgruppe ähnliche Tendenzen zeigen wie in anderen Studien. Zudem wird dargelegt, welche neuen Erkenntnisse gewonnen wurden und wo Grenzen dieser Arbeit sind. Dabei wird, angelehnt an die Vorstellung von Ergebnissen anderer empirischer Studien in Kapitel 5, zuerst auf das Gelingen von Beweisversuchen, anschließend auf die Verwendung von Beispielen und dann auf die Darstellungsmittel und Darstellungsformen eingegangen (Abschn. 9.3.1 - 9.3.3).

Im Vergleich zu anderen Studien liegt in der vorliegenden Arbeit ein besonderer Fokus auf Anknüpfungspunkten für eine Weiterentwicklung der Beweisversuche zu einem Beweis. Daher werden abschließend vor allem für Personen, die im Bereich der Förderung von Teilnehmenden der Mathematik-Olympiade aktiv sind, Ansatzpunkte für eine Förderung beschrieben, die durch die erhaltenen Ergebnisse naheliegen (Abschn. 9.3.4). Die Identifikation dieser Ansatzpunkte kann als wesentlicher praktischer Nutzen der vorliegenden Arbeit angesehen werden.

9.3.1 Gelingen von Beweisversuchen

In diesem Abschnitt werden primär die Ergebnisse bezüglich des Nachweises der Gültigkeit der zu zeigenden Aussage eingeordnet und diskutiert. Dabei werden Ergebnisse anderer empirischer Untersuchungen und auch die in den Bildungsstandards und den Lehrplänen für NRW formulierten Kompetenzerwartungen, die in Abschnitt 4.2 vorgestellt wurden, aufgegriffen.

9.3.1.1 Ergebnisse dieser Arbeit im Vergleich zu Ergebnissen anderer Studien

In mehreren empirischen Studien wird nachgewiesen, dass manchen Lernenden in den Sekundarstufen Beweise mathematischer Allaussagen mit unendlichem Gültigkeitsbereich gelingen (Abschn. 5.1.1). Diese Erkenntnis wird durch die Ergebnisse der vorliegenden Arbeit gestützt. In manchen Beweisversuchen aus jeder der Klassenstufen 6 bis 8 kommen **Beweise** vor, jedoch jeweils in unter 3% der analysierten Beweisversuche (Abschn. 8.5.1.2).

Insbesondere in der Klassenstufe 8 hat mit 38% ein bedeutender Anteil der Lernenden zwar keinen Beweis, aber einen **Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge** abgegeben (Abschn. 8.5.1.2). Vor allem in einer Untersuchung mit leistungsstarken Neuntklässlerinnen und Neuntklässlern in Taiwan deutet sich eine ähnliche Tendenz an. Während der Anteil der Beweise in dieser Studie jeweils bei 6% liegt, gehören bei der aus dem Schulunterricht bekannten Allaussage rund 31% und bei der unbekannten Allaussage rund 14% der Beweisversuche in die Kategorie „[p]artial proof, including all information needed but some reasoning omitted“ (Chin & Lin, 2009, S. 62; Abschn. 5.1).

In einigen Studien hervorgehoben wird ein großer Anteil an Bearbeitungen mit empirisch-induktivem Ansatz (Abschn. 5.1.1). Beispielsweise sind in der Studie von Knuth et al. (2002) mit Schülerinnen und Schülern in den Klassenstufen 6 bis 8 in über 50% der Beweisversuche ausschließlich ein oder mehrere Beispiele angeführt. (Abschn. 5.1.1, 5.2.1). In der Untersuchung im Rahmen dieser Arbeit liegt der Anteil an **Beweisversuchen mit prüfendem Ansatz** in der Klassenstufe 5 bei knapp 25%, in den Klassenstufen 6 bis 8 jeweils zwischen 10% und 20%. In

jeder der Klassenstufen 5 bis 8 haben mehr als die Hälfte der Teilnehmenden, die sich erkennbar mit dem Aufgabenteil auseinandergesetzt haben, einen **Beweisversuch mit begründendem Ansatz** formuliert. Im Vergleich zu manchen anderen Untersuchungen wurde in der Studie im Rahmen dieser Arbeit explizit und ohne zusätzlichen Entdeckungsauftrag zum Beweisen aufgefordert (Abschn. 5.1, 5.4, 6.2.2). Forschungsergebnisse von Gerloff (2021) deuten an, dass dies zu dem großen Anteil an Beweisversuchen mit begründendem Ansatz beigetragen haben kann (Abschn. 5.1). Zusätzlich steht auf allen Aufgabenblättern der Mathematik-Olympiade, dass immer der Lösungsweg samt Begründungen und Nebenrechnungen auszuführen ist (Abschn. 2.3.2). Auch dies hat eventuell zu dem großen Anteil an Beweisversuchen mit begründendem Ansatz beigetragen. Außerdem kann die Verwendung des Adverbs „immer“ in der Aufgabenstellung und in den Klassenstufen 5, 6 die Formulierung „begründe, warum“ die Formulierung von Beweisversuchen mit begründendem Ansatz begünstigt haben (Abschn. 6.2.2, 9.1.2). Nichtsdestotrotz erscheint der Anteil der Beweisversuche mit begründendem Ansatz groß, der Anteil derjenigen mit prüfendem Ansatz klein. Schließlich handelt es sich auch in der Untersuchung von Knuth et al. (2002) um Beweisaufgaben, in denen zudem das Adverb „immer“ vorkommt. Und die in der Untersuchung von Knuth et al. (2002) vorgegebenen Aussagen sind einfacher zu beweisen als die in dieser Arbeit fokussierte Aussage.

Die Ergebnisse dieser Arbeit weisen darauf hin, dass die Beweisversuche mit Blick auf den Nachweis der Gültigkeit im Mittel je höher die **Klassenstufe**, desto tragfähiger sind. Dies passt zu den Ergebnissen manch anderer Querschnittsstudien (Abschn. 5.1.1). Entsprechend den Erläuterungen in Abschnitt 9.1.1 ist jedoch bei der Verallgemeinerung der Ergebnisse zu Unterschieden zwischen Beweisversuchen von Teilnehmenden der Mathematik-Olympiade in verschiedenen Klassenstufen auf andere Zielgruppen Vorsicht geboten.

In Einklang mit den Ergebnissen vieler anderer Studien lassen sich deutliche Leistungsunterschiede zwischen Lernenden in der gleichen Klassenstufe festhalten (insb. Abschn. 5.1.1, 8.5.1.2). Dabei deuten die Ergebnisse dieser Arbeit vor allem mit Blick auf Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis an, dass die Beweisversuche aus der Klassenstufe 8 besonders heterogen sind, im Vergleich dazu die Beweisversuche aus der Klassenstufe 5 am homogensten (Abschn. 8.5.3.2). Es stellt sich die Frage, inwieweit Aspekte der vorkommenden Vielfalt an Beweisversuchen von Teilnehmenden der Mathematik-Olympiade mit einem besonderen Erfolg im Wettbewerb oder einer besonderen mathematischen Förderung zusammenhängen. Zudem wäre es interessant zu wissen, inwieweit bestimmte Ausprägungen der Beweisversuche mit Blick auf unterschiedliche Teilgruppen der Teilnehmenden der Mathematik-Olympiade mehr oder weniger relevant sind. Diese Fragen ließen sich auf Grundlage der im Rahmen dieser Arbeit erhobenen Daten bearbeiten (Abschn. 10.1, 10.2, 10.4.2).

Des Weiteren deuten die Ergebnisse dieser Arbeit an, dass das Gelingen von Beweisversuchen in einem gewissen Maße mit der primären Beweisidee zusammenhängt. Zum Beispiel unterscheidet sich die Korrektheit und Vollständigkeit von Beweisversuchen mit begründendem Ansatz je nach primärer Beweisidee deutlich (Abschn. 8.2.6, 8.3.2.2). Nebenbei zeigen die Ergebnisse, dass in fast jeder der Klassenstufen 5 bis 8 alle unterschiedenen **Beweisideen** zu finden

sind. Dabei gehört die Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* in den Klassenstufen 7, 8 zu den häufigsten Beweisideen. Die Beweisidee 3 *Vierfaches einer mittleren Zahl oder Durchschnittsbildung* hingegen gehört in allen Klassenstufen zu den am seltensten verwendeten Beweisideen. Unter anderem hinsichtlich der vorkommenden Beweisideen sind einige statistisch signifikante Unterschiede zwischen den Klassenstufen 5, 6 und den Klassenstufen 7, 8 zu erkennen (Abschn. 8.2.2, 8.3.2.2).

In fast keiner der vorhandenen Studien, die sich mit schriftlichen Bearbeitungen von Aufgaben mit Beweisnotwendigkeit von Schülerinnen und Schülern beschäftigen, werden die Beweisideen berücksichtigt. Eine Ausnahme bildet die Untersuchung von Brunner et al. (2022) mit Lernenden vierter bis sechster Klassen. Bezogen auf eine Aufgabe zur Summe von drei aufeinanderfolgenden Zahlen berichten Brunner et al. (2022), dass alle von ihnen unterschiedenen inhaltlichen Ideen in den analysierten Aufgabenbearbeitungen vorkommen. Dabei ist in allen Klassenstufen die Idee, dass der jeweils kleinste Summand durch die nächstgrößere Zahl ersetzt wird, eher häufig. Die Idee, dass es sich um eine Verdreifachung des mittleren Summanden handelt, kommt hingegen besonders selten vor. Hinsichtlich der vorkommenden inhaltlichen Ideen sind statistisch signifikante Unterschiede zwischen den einzelnen Klassenstufen vier bis sechs zu erkennen (Abschn. 5.1.1).

Was die Beweisideen betrifft passen die Ergebnisse dieser Arbeit zu denen von Brunner et al. (2022). In Ergänzung dazu werden in der vorliegenden Arbeit Zusammenhänge zwischen der primären Beweisidee und anderen Eigenschaften der Beweisversuche deutlich.

Auf weitere Zusammenhänge zwischen verschiedenen Eigenschaften von Beweisversuchen wird in den Abschnitten 9.3.2 und 9.3.3 eingegangen.

9.3.1.2 Ergebnisse dieser Arbeit im Vergleich zu curricularen Vorgaben

Nach den Bildungsstandards und den Lehrplänen für Nordrhein-Westfalen (NRW) können leistungsstarke Schülerinnen und Schüler aller Klassenstufen 5 bis 8 in der Auseinandersetzung mit mathematischen Allaussagen **Beispiele angeben** (Abschn. 4.2). Diesbezüglich zeigen die Ergebnisse dieser Arbeit folgendes Bild: In jeder der Klassenstufen 5 bis 8 enthalten etwa ein Fünftel bis zu einem Drittel aller Beweisversuche, die nicht fehlend oder durchgestrichen sind, ein korrekt und vollständig ausgeführtes Beispiel zu der zu zeigenden Aussage, d. h. eines, das die zu zeigende Aussage bestätigt. Zusätzlich kommt je Klassenstufe in einem mehr als nochmals so großen Anteil der Beweisversuche ein unvollständig oder fehlerhaft ausgeführtes Beispiel zu der zu zeigenden Aussage vor. In manchen Beispielen kommen Rechenfehler vor oder wurde das Gleichheitszeichen prozedural verwendet, weshalb Beispiele fehlerhaft sind. Zudem ist es denkbar, dass einige Lernende die Beispiele unvollständig aufgeschrieben haben, weil ihnen die Teilbarkeit durch vier bzw. acht einer Summe, wie z. B. 48, als offensichtlich erschien. Außerdem ist es möglich, dass Lernende Beispiele unvollständig aufgeschrieben haben, weil sie mithilfe der Beispiele weniger die zu zeigende Aussage bestätigen oder überprüfen als eine Beweisidee generieren oder veranschaulichen wollten.

Insgesamt haben je Klassenstufe von den Lernenden, die sich erkennbar mit der Aufgabe beschäftigt haben, rund 60% bis 80% irgendein Beispiel angegeben. Vor dem Hintergrund der

Bildungsstandards und der Kernlehrpläne für NRW überrascht es dabei, dass dieser Anteil in der Klassenstufe 5 am kleinsten ist (Abschn. 4.2, 8.5.2). Zudem zeigen die Ergebnisse, dass in der Klassenstufe 5 in den Beweisversuchen mit Beispiel häufiger nur genau ein Beispiel angegeben ist als in jeder der höheren Klassenstufen 6 bis 8 (Abschn. 8.5.2.3). Außerdem deuten die Ergebnisse darauf hin, dass die Auswahl der Beispiele und deren Nutzung in der Klassenstufe 5 weniger optimal erfolgt ist als dies vor allem in den Klassenstufen 7, 8 der Fall war (Abschn. 8.5.3.2). Die Beispielnutzung scheint bei Teilnehmenden der Mathematik-Olympiade in der Klassenstufe 5 weniger weit entwickelt zu sein insbesondere als bei Teilnehmenden in den Klassenstufen 7 und 8.

Weiterhin ist nach den Bildungsstandards und den Lehrplänen für Nordrhein-Westfalen (NRW) zu erwarten, dass leistungsstarke Schülerinnen und Schüler aller Klassenstufen 5 bis 8 durchaus über die Angabe von Beispielen hinausgehen, **Invarianten identifizieren und erste allgemeingültige Beweisideen skizzieren** (Abschn. 4.2).

In jeder der Klassenstufen 5 bis 8 gibt es Beweisversuche mit prüfendem Ansatz und einem weiterentwickelbaren Argumentgefüge. Dass in einem Beweisversuch mit prüfendem Ansatz eine Beweisidee erkennbar ist, trifft je Klassenstufe 5 bis 8 in ca. 7% bis 27% der Beweisversuche prüfendem Ansatz zu. Bezogen auf alle Beweisversuche handelt es sich je Klassenstufe um unter 5%.

Wesentlich häufiger sind in allen Klassenstufen 5 bis 8 Beweisversuche mit begründendem Ansatz. Dabei impliziert die Existenz eines begründenden Ansatzes insbesondere, dass eine Invariante identifiziert und als Beweisidee skizziert wurde. Diese Beweisidee kann jedoch falsch, sehr nah an der zu zeigenden Aussage oder nur als einschrittiges Argumentgefüge ausgeführt sein. Beweisversuche mit begründendem Ansatz, aber allenfalls einem Ansatz für ein weiterentwickelbares Argumentgefüge machen in der Klassenstufe 5 rund 38% aller Beweisversuche aus, in der Klassenstufe 6 rund 50%, in der Klassenstufe 7 rund 34% und in der Klassenstufe 8 rund 25%.

Der Anteil der Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge an allen Beweisversuchen beträgt in der Klassenstufe 5 rund 8%, in der Klassenstufe 6 rund 21%, in der Klassenstufe 7 rund 41% und in der Klassenstufe 8 rund 53%. Diese Beweisversuche gehen tendenziell über die Kompetenzerwartungen an Schülerinnen und Schüler am Ende der vierten Klassenstufe hinaus.

Die Aufgabenbeispiele in den Bildungsstandards Mathematik für die Grundschule und die im Lehrplan Mathematik für die Grundschule in NRW explizierten Kompetenzerwartungen lassen vermuten, dass vor allem Schülerinnen und Schüler zu Beginn der Sekundarstufe I die Invarianten ausgehend von der Arbeit mit Beispielen formulieren. Nach den curricularen Vorgaben lässt sich zudem erwarten, dass der Anteil der Beweisversuche mit begründendem Ansatz und insbesondere der Anteil derjenigen mit **deduktiver Schlussweise** je höher die Klassenstufe, desto größer ist (Abschn. 4.2).

Diese Erwartungen bestätigen die erhaltenen Ergebnisse nicht. Je Klassenstufe handelt es sich bei etwa 30% bis 40% aller Beweisversuche um solche mit deduktiver Schlussweise (Abschn. 8.5.2). Im Falle eines Beweisversuchs mit begründendem Ansatz ist die Schlussweise in der Klassenstufe 5 mit etwa 65% am häufigsten deduktiv, statistisch hoch signifikant häufiger als in den Klassenstufen 7 und 8, in denen der Anteil jeweils unter 50% liegt (Abschn. 8.2.1.2). Es ist denkbar, dass viele der Teilnehmenden in der Klassenstufe 5, die einen Beweisversuch mit begründendem Ansatz abgegeben haben, ausschließlich eine eventuell naheliegende Invariante als Beweisidee skizziert haben, ohne mit Beispielen zu arbeiten. Die große Mehrheit der Beweisversuche mit deduktiver Schlussweise aus der Klassenstufe 5 enthalten kein weiterentwickelbares Argumentgefüge (Abschn. 8.2.1.3). Außerdem ist das Forschungsergebnis von Gerloff (2021) aufzugreifen, wonach viele der Lernenden aus dritten und vierten Klassen in verallgemeinerter Form argumentieren (Abschn. 5.1.1). Dabei folgert Krumsdorf (2017) aus seinen Beobachtungen, dass eine allgemein formulierte Aussage nicht unbedingt mit einer subjektiven Realisierung des Allgemeingültigen einhergeht, sondern eventuell nur eine Sprechweise ist, um geführte Erwartungen zu erfüllen (Abschn. 5.2.2).

Was den großen Anteil an Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise in den Klassenstufen 7 und 8 betrifft, ist aufzugreifen, dass die Erhebung der Beweisversuche im ersten Schulhalbjahr stattgefunden hat (Abschn. 9.1.2). Zudem deuten die empirischen Ergebnisse an, dass die Verwendung der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* mit dem großen Anteil an Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise zusammenhängt. Beweisversuche mit der primären Beweisidee 1 machen in den Klassenstufen 7 und 8 ca. 30% aller Beweisversuche aus (Abschn. 6.2.2, 8.2.2.2, 8.2.6, 8.5.2.3). Es stellt sich u. a. die Frage, inwieweit sich bei den Beweisversuchen von Teilnehmenden der Mathematik-Olympiade ab der Klassenstufe 9, das heißt bei denjenigen, die mit dem Umgang mit Variablen und dem Beweisen von Allaussagen aus dem Schulunterricht vertrauter sind, ein anderes Bild zeigen würde (Abschn. 10.1, 10.4.1). Schließlich sollen Schülerinnen und Schüler nach dem Kernlehrplan für das Gymnasium in NRW am Ende der Klassenstufe 8 unter Verwendung ihres mathematischen Wissens mehrschrittige Argumentgefüge entwickeln und formulieren können (MSW NRW, 2007). Am Ende der Klassenstufe neun wird erwartet, dass sie „mathematisches Wissen und mathematische Symbole für Begründungen und Argumentationsketten“ (MSW NRW, 2007, S. 29) nutzen.

9.3.2 Verwendung von Beispielen

In Abschnitt 9.3.1 wurden bereits mehrere Ergebnisse bezüglich der Beispielnutzung vor dem Hintergrund curricularer Vorgaben und Ergebnissen anderer Studien betrachtet. In Ergänzung zu diesen Ergebnissen wird im Folgenden fokussiert auf Zusammenhänge zwischen der Beispielnutzung und der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges eingegangen (Abschn. 9.3.2.1). Danach werden weitere Aspekte der Beispielnutzung separat mit Blick auf Beweisversuche mit prüfendem Ansatz und mit Blick auf Beweisversuche mit begründendem Ansatz thematisiert (Abschn. 9.3.2.2-9.3.2.3). Dabei wird unter anderem

die Vielfalt der Beweisversuche mit Beispiel skizziert. So wie es vorhandene Forschungsergebnisse aufzeigen, ist das Spektrum der Beweisversuche mit Beispiel auch in der Studie im Rahmen dieser Arbeit sehr facettenreich.

9.3.2.1 Beispielnutzung und die Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges

Die Ergebnisse der vorliegenden Arbeit liefern empirische Hinweise darauf, dass Aspekte der Beispielnutzung mit der **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** zusammenhängen (insb. Abschn. 8.2.6.5, 8.5.2.3). Dies passt zu den Ergebnissen des Forschungsprojekts „The Role and Use of Examples in learning to Prove“ (Lynn & Lockwood, 2000; Ozgur et al., 2019; Zaslavsky & Knuth, 2019; Abschn. 3.3.5, 5.2.3).

Konkret sind folgende Ergebnisse dieser Arbeit hervorzuheben: Beweisversuche mit begründendem Ansatz und einem Beispiel zu der zu zeigenden Aussage, solche mit einem Beispiel innerhalb des (übrigen) Argumentgefüges sowie solche mit Reihenfolgebeispielen und dem kleinsten Beispiel sind hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges eher stark. Beweisversuche mit begründendem Ansatz ohne Beispiel oder ausschließlich mit sonstigen Beispielen, solche mit einem großen Beispiel und solche mit allen Beispielen oberhalb des (übrigen) Argumentgefüges sind hinsichtlich der Korrektheit und Vollständigkeit der Beweisversuche im Durchschnitt eher schwach. Dabei scheint es jedoch auf den Umgang mit den Beispielen anzukommen (Abschn. 8.2.6.5). In den Beweisversuchen mit prüfendem Ansatz und Reihenfolgebeispielen ist statistisch hoch signifikant häufiger ein weitgehend korrektes Argumentgefüge zu erkennen als in den anderen Beweisversuchen mit prüfendem Ansatz. Demgegenüber ist in keinem der analysierten Beweisversuche mit prüfendem Ansatz und einem großen Beispiel eine Beweisidee zu erkennen (Abschn. 8.3.2.2).

Dass Beweisversuche mit Reihenfolgebeispielen, und insbesondere solche mit Reihenfolgebeispielen und dem kleinsten Beispiel, vergleichsweise häufig ein weitgehend korrektes Argumentgefüge enthalten, erscheint erwartungsgemäß. Schließlich lässt sich anhand von Reihenfolgebeispielen besonders gut die Invariante der Vergrößerung der Summe um acht erkennen und darstellen (Abschn. 6.2.2). Das Ergebnis von Krumsdorf (2017), dass die Betrachtung schwieriger oder nicht gut vorstellbarer Beispiele, wie etwa großer Beispiele, die Entwicklung eines beispielgebundenen Beweises begünstigt, lässt sich mit den Ergebnissen dieser Studie nicht stützen.

Die Analyse der Beweisversuche im Rahmen dieser Arbeit verdeutlicht ebenfalls, dass die Arbeit mit Beispielen trotz allen Potentials keinen Selbstläufer hin zu einem korrekten und vollständigen Argumentgefüge darstellt, auch nicht bei Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 7 und 8. Beispielsweise deuten manche Beweisversuche an, dass Teilnehmende bei der Arbeit mit Beispielen falsche Verallgemeinerungen vorgenommen haben, so wie es auch Krumsdorf (2017) beobachtet hat (Abschn. 5.2.2, 8.2.3.3). Für die Entwicklung und Formulierung eines Beweises mithilfe von Beispielen scheint eine tiefgehende, kritisch-reflexive Auseinandersetzung mit Beispielen zumindest hilfreich.

9.3.2.2 Beispiele in Beweisversuchen mit prüfendem Ansatz

Passend zu den Ergebnissen anderer empirischer Studien kommen in jeder der Klassenstufen 5 bis 8 **sehr unterschiedliche Beweisversuche mit prüfendem Ansatz** vor. Beispielweise gibt es jeweils Beweisversuche mit prüfendem Ansatz, in denen die angegebenen Beispiele nicht zu der zu zeigenden Aussage passen. Ebenfalls gibt es jeweils Beweisversuche mit prüfendem Ansatz, in denen mindestens ein korrekt und vollständig ausgeführtes eigenes Beispiel vorkommt, sowie solche, in denen ein weitgehend korrektes Argumentgefüge ausgeführt ist. Darüber hinaus erscheinen die Beweisversuche mit prüfendem Ansatz auch mit Blick auf die Schlussweise vielfältig (Abschn. 5.1.1, 5.2, 8.3).

Fokussiert auf Beweisversuche mit prüfendem Ansatz zeigen die Ergebnisse der vorliegenden Arbeit, dass, anders als in der Untersuchung von Küchemann und Hoyles (2002), in der Mehrheit der Fälle mindestens zwei Beispiele angegeben sind. In jeder der Klassenstufen 5 bis 8 enthalten mehr als die Hälfte der Beweisversuche mit prüfendem Ansatz und mindestens zwei Beispielen auch systematische Beispiele (Abschn. 8.3.2.2). Insgesamt sind je Klassenstufe in mehr als der Hälfte der Beweisversuche mit prüfendem Ansatz zumindest **sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges** zu erkennen. Diese Ergebnisse sind mit Blick auf Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche tendenziell günstig.

So wie in der Untersuchung von Küchemann und Hoyles (2002) kommt im Datensatz dieser Arbeit insbesondere auch ein Beweisversuch mit prüfendem Ansatz vor, in dem die Unzulänglichkeit von Beispielen als Beweis expliziert ist. Auch dies stellt ggf. einen guten Anknüpfungspunkt für eine Weiterentwicklung eines Beweisversuchs mit prüfendem Ansatz dar.

Zu der in mehreren Studien beachteten Frage, **warum Lernende einen Beweisversuch mit prüfendem Ansatz formuliert haben**, können anhand von schriftlichen Beweisversuchen in der Regel nur Vermutungen aufgestellt werden. Nach der Analyse der Schlussweise sind in den Klassenstufen 5, 7, 8 jeweils in mehr als einem Viertel der Beweisversuche mit prüfendem Ansatz Beispiele explizit als Grund für die Gültigkeit dargestellt. Dass die Ausführungen klar auf einen induktiven Schluss hindeuten, trifft aber je Klassenstufe bei weniger als einem Zehntel aller Beweisversuche mit prüfendem Ansatz zu (Abschn. 8.3.2.3). Die weitere Auswertung der erhobenen Daten, das heißt insbesondere der Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche, würde Einblicke geben, inwieweit die Teilnehmenden einen empirisch-induktiven Beweisversuch als einen Beweis ansehen (Abschn. 6.2, 10.1, 10.2).

9.3.2.3 Beispiele in Beweisversuchen mit begründendem Ansatz

Ähnlich wie bei Beweisversuchen mit prüfendem Ansatz lassen sich bei Beweisversuchen mit begründendem Ansatz i. d. R. keine sicheren Aussagen treffen, welche **Funktion die angegebenen Beispiele** aus der Sicht der Schülerin bzw. des Schülers im Beweisversuch haben. Es ist aber möglich, zu beschreiben, welche Funktion Beispiele aus der Perspektive von Lesenden einnehmen. Aus der Sicht von Lesenden können Beispiele in Beweisversuchen mit begründendem Ansatz z. B. darauf hinweisen, wie die zu zeigende Aussage interpretiert wurde, allgemein

gehaltene Ausführungen veranschaulichen, ergänzen oder präzisieren und Anknüpfungspunkte bieten, um inhaltliche Lücken im Argumentgefüge zu schließen oder Fehler zu korrigieren (Abschn. 8.2.3.3).

So wie es z. B. die von Küchemann (2008) abgebildeten Beweisversuche erwarten lassen, sind die Beweisversuche mit begründendem Ansatz und Beispiel vielfältig, was die **Anzahl**, die **Art** und den **Ort der Beispiele** betrifft. Zusätzlich zu dieser Vielfalt beschreiben die Ergebnisse im Rahmen dieser Arbeit auch Häufigkeitsverteilungen der Eigenschaften von Beispielen in Beweisversuchen mit begründendem Ansatz. Beispielsweise sind in den Beweisversuchen mit begründendem Ansatz und Beispiel meistens zwei bis fünf Beispiele ausgeführt. Je Klassenstufe kommt in gut einem Drittel bis zu gut der Hälfte der Beweisversuche mit begründendem Ansatz und Beispiel das kleinste Beispiel vor (Abschn. 8.2.3.2). Die recht häufige Angabe des kleinsten Beispiels und das eher seltene Vorkommen großer Beispiele passt dabei zu Ergebnissen von Lynch und Lockwood (2019). Zudem steht das Ergebnis, dass in der Mehrheit der Beweisversuche mit mindestens zwei Beispiele systematische Beispiele vorkommen und dass die Beweisversuche mit systematischen Beispielen teils, aber nicht alle ein weitgehend korrektes Argumentgefüge enthalten, tendenziell in Einklang mit Erkenntnissen von Ozgur et al. (2019).

9.3.3 Darstellungsmittel und Darstellungsformen

In diesem Abschnitt werden die Ergebnisse bezüglich vorkommender Darstellungsmittel und Darstellungsformen vor dem Hintergrund curricularer Vorgaben sowie der Ergebnisse anderer empirischer Untersuchungen eingeordnet und diskutiert.[162] Dabei wird nacheinander auf symbolische Darstellungsmittel und Darstellungsformen mit symbolischem Anteil (Abschn. 9.3.3.1), auf verbale Darstellungsmittel und die Darstellungsformen numerisch-verbal, ikonisch-verbal, verbal (Abschn. 9.3.3.2) sowie auf ikonische Darstellungsmittel und Darstellungsformen mit ikonisch Anteil (Abschn. 9.3.3.3) eingegangen. Es wird deutlich, dass die im Rahmen dieser Arbeit erhaltenen Ergebnisse in mehrerlei Hinsicht in Einklang mit den bereits vorhandenen Erkenntnissen stehen und über den bisherigen Forschungsstand hinausgehen.

9.3.3.1 Symbolische Darstellungsmittel und Darstellungsformen mit symbolischem Anteil

Bezüglich der Verwendung symbolischer Darstellungsmittel fällt auf, dass nicht in der Klassenstufe 5, aber in den **Klassenstufen 6 und 7** einzelne Teilnehmende einen Beweisversuch mit einer Darstellungsform mit symbolischem Anteil formuliert haben. Dabei sind Buchstabenvariablen Lernenden am Anfang dieser Klassenstufen normalerweise nicht oder kaum bekannt. Die Ergebnisse dieser Arbeit lassen offen, inwieweit die Teilnehmenden aus den Klassenstufen 6 und 7, die einen Beweisversuch mit einer Darstellungsform mit symbolischem Anteil formuliert

[162] Die Ergebnisse der Analyse der Darstellungsmittel sind in Abschnitt 8.5.2.4 gebündelt und in den Abschnitten 8.2.4, 8.2.6.5, 8.3.2.4, 8.4.2 differenziert für Beweisversuche mit begründendem, prüfendem und weder begründendem noch prüfendem Ansatz dargestellt. Curriculare Vorgaben sind in Abschnitt 4.2 beschrieben und die relevanten Ergebnisse aus anderen empirischen Studien in Abschnitt 5.3. Im Sinne einer guten Lesbarkeit wird im Folgenden nicht erneut auf diese Abschnitte verwiesen.

haben, mathematisch gefördert wurden (Abschn. 10.1, 10.4.2). Auch kann man fragen, inwieweit diese Teilnehmenden besonders gut bei der Regionalrunde der Mathematik-Olympiade abgeschnitten haben. Schließlich handelt es sich in allen analysierten Fällen um Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge (Abschn. 6.2.2, 8.5.2.4, 10.1, 10.4.2). Zudem fällt auf, dass alle Beweisversuche mit einer Darstellungsform mit symbolischem Anteil aus den Klassenstufen 6 und 7 auf der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* basieren. Es wäre interessant zu wissen, inwieweit die Idee der Verwendung von Buchstabenvariablen zur Nutzung dieser primären Beweisidee beigetragen hat. In beiden Klassenstufen 6 und 7 kommen auch Beweisversuche mit der primären Beweisidee 2 ohne symbolische Darstellungsmittel vor.

Lernenden am Anfang der **Klassenstufe 8** sind Buchstabenvariablen normalerweise aus dem Schulunterricht bekannt, wobei sie eventuell noch nicht allzu gewohnt daran und geübt darin sind, Algebra zu verwenden. Demnach erscheint es durchaus erwartungsgemäß, dass der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil in der Klassenstufe 8 statistisch hoch signifikant größer ist als in jeder der Klassenstufen 5 bis 7 und bei rund 13% liegt. Es bleibt an dieser Stelle offen, welches Bild sich diesbezüglich bei Teilnehmenden der Mathematik-Olympiade ab der Klassenstufe 9 ergäbe. In verschiedenen anderen Studien deutet sich an, dass die Verwendung von Algebra in Bearbeitungen von Aufgaben mit Beweisnotwendigkeit auch bei Schülerinnen und Schülern aus höheren Klassenstufen nicht so stark ausgeprägt ist, wobei es u. a. Unterschiede zwischen Lernenden aus verschiedenen Ländern zu geben scheint.

Des Weiteren ist festzuhalten, dass die **Schlussweise** in ca. drei Viertel der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil deduktiv ist, meistens deduktiv ohne zusätzliches Beispiel. Ein möglicher Erklärungsansatz besteht darin, dass die Verwendung von Algebra die Formulierung deduktiver Argumentgefüge begünstigt hat, weil sie die sprachlichen Anforderungen beim Verfassen eines solchen Argumentgefüges verlagert.

Die Ergebnisse anderer Studien weisen noch darauf hin, dass einigen Lernenden bei der Verwendung von Algebra bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit **Fehler** passieren. In der vorliegenden Studie handelt es sich bei zwei Drittel der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil um Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge. In weniger als einem Zehntel der Beweisversuche mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil ist ein inhaltlicher oder grundlegender Fehler zu erkennen. Dabei gilt es zu bedenken, dass bei der Analyse der Korrektheit nicht berücksichtigt wurde, inwieweit die Variablen angemessen definiert sind (Abschn. 3.3.4, 3.3.11). Zudem ist zu berücksichtigen, dass bei der in dieser Arbeit fokussierten Aussage elementare Kompetenzen im Umgang mit Algebra ausreichen (Abschn. 6.2.2) und dass vielleicht viele der Lernenden, die einen Beweisversuch mit einer Darstellungsform mit symbolischem Anteil formuliert haben, besonders leistungsstark sind oder mathematisch gefördert wurden (Abschn. 10.1, 10.4.2).

9.3.3.2 Verbale Darstellungsmittel und die Darstellungsformen numerisch-verbal, ikonisch-verbal, verbal

Verbale Darstellungsmittel kommen in fast jedem analysierten Beweisversuch vor, der nicht fehlend oder durchgestrichen ist. Zudem haben in jeder der Klassenstufen 5 bis 8 die Beweisversuche meistens eine der Darstellungsformen numerisch-verbal und verbal. Insbesondere ist auch bei den meisten Beweisversuchen mit begründendem Ansatz in jeder der Klassenstufen die Darstellungsform meistens verbal oder numerisch-verbal, was tendenziell in Einklang mit den Ergebnissen von Healy und Hoyles (2000) steht.

Auffällig ist vor allem, dass die Beweisversuche mit begründendem Ansatz und einer der Darstellungsformen verbal, ikonisch-verbal hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges vergleichsweise schwach sind. Ein denkbarer Erklärungsansatz besteht darin, dass in diesen Beweisversuchen oftmals ausschließlich eine vielleicht naheliegende Beweisidee angeführt wurde, ohne dass die Tragfähigkeit der Überlegungen hinterfragt und Beispiele betrachtet wurden. Zudem kann es sein, dass im Falle dieser Beweisversuche vergleichsweise häufig eine erkannte Invariante ausschließlich kurz skizziert und nicht, etwa mithilfe von Beispielen, weiter ausgeführt wurde.

9.3.3.3 Ikonische Darstellungsmittel und Darstellungsformen mit ikonischem Anteil

Ergebnisse anderer Studien weisen darauf hin, dass Beweisversuche von Schülerinnen und Schülern manchmal primär ikonisch dargestellt sind. In der Untersuchung im Rahmen dieser Arbeit kommen in jeder der Klassenstufen 5 bis 8 Beweisversuche mit einer **Darstellungsform mit ikonischem Anteil** vor. Dabei ist die Darstellungsform nie rein ikonisch, so wie es auch in der Studie von Gerloff (2021) der Fall ist. Die Darstellungsform ist fast immer ikonisch-numerisch-verbal. Zudem deutet sich an, dass Beweisversuche mit einer Darstellungsform mit ikonischem Anteil meistens zusammen mit bestimmten primären Beweisideen und Schlussweisen vorkommen. Dabei sind bei der fokussierten Allaussage prinzipiell vielfältige Beweisversuche mit einer Darstellungsform mit ikonischem Anteil denkbar (Abschn. 6.2.2, 8.2.6.5, 8.3.3.4, 8.5.2.4).

Weiterhin ist der Anteil der Beweisversuche mit einer Darstellungsform mit ikonischem Anteil in den Klassenstufen 5 und 6 besonders klein. Auch fokussiert auf die Beweisversuche mit begründendem Ansatz liegt er deutlich unter 10%. Dies ist bemerkenswert, weil die Verwendung ikonischer Darstellungsmittel aus mathematikdidaktischer Sicht gerade für diese jungen Lernenden hilfreich sein kann, um Begründungen zu entwickeln und zu kommunizieren.

Bezüglich der **Art der ikonischen Darstellungsmittel** fällt auf, dass figurierte Zahlen ausschließlich in einzelnen Beweisversuchen vorkommen, obwohl sie inzwischen ein fester Bestandteil in der Ausbildung von Mathematiklehrerinnen und -lehrern sind (Abschn. 8.2.4, 8.3.2.4, 8.4.2; Jahnke & Ufer, 2015). In einigen Beweisversuchen, in denen ikonische Darstellungsmittel vorkommen, sind ausschließlich einzelne Begriffe, wie z. B. „immer“, durch eine Unterstreichung hervorgehoben (Abschn. 8.2.4).

Insgesamt scheint die Verwendung ikonischer Darstellungsmittel zur Entwicklung und Darstellung von Beweisideen nicht so stark ausgeprägt zu sein, insbesondere nicht bei Teilnehmenden der Mathematik-Olympiade am Beginn der Sekundarstufe I.

9.3.4 Ansatzpunkte für Förderung

Den Anstoß für die vorliegende Dissertation gab das Anliegen, **Teilnehmende der Mathematik-Olympiade**, und zwar insbesondere auch solche zu **Beginn der Sekundarstufe I**, im Führen von Beweisen zu fördern (Abschn. 1.1). Bei der Mathematik-Olympiade werden ab der Klassenstufe 5 Beweisaufgaben gestellt (Abschn. 2.3). Dabei scheint es besonders wichtig, auf frühe Ansatzpunkte präalgebraischen Argumentierens zu fokussieren, die zu gültigen Beweisen ausgebaut werden können. Im daher ausgewählten **Themenbereich der Teilbarkeit ganzer Zahlen** ist es gut möglich, auf sehr elementarem Niveau tragfähige Vorformen von Beweisen bis hin zu Beweisen zu formulieren. In der Studie im Rahmen dieser Arbeit haben sich in jeder der Klassenstufen 5 bis 8 über 75% der Teilnehmenden erkennbar mit dem Aufgabenteil auseinandergesetzt (Abschn. 8.1). In jeder der Klassenstufen haben mehr als die Hälfte dieser Teilnehmenden einen Beweisversuch mit begründendem Ansatz abgegeben und in jeder der Klassenstufen sind Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge bei begründendem Ansatz entstanden (Abschn. 8.1, 8.2.1). Auch diese empirischen Ergebnisse sprechen dafür, bei Teilnehmenden der Mathematik-Olympiade mit einer Förderung im Führen von Beweisen zu Allaussage am Anfang der Sekundarstufe I zu beginnen. Vor allem für Personen, die im Bereich der Förderung von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 aktiv sind, werden im Folgenden ausgehend von den Ergebnissen dieser Arbeit Ansatzpunkte für eine Förderung im Führen von Beweisen zusammengestellt.

Die Ergebnisse zeigen, dass in den Klassenstufen 5 und 6 bei mehr als der Hälfte der Beweisversuche offen ist, ob sie sich auf eine ganz korrekte **Interpretation der zu zeigenden Aussage** beziehen. Zudem ist in manchen Beweisversuchen ein Hinweis zu erkennen, dass bei der Interpretation der vorgegebenen Aussage ein Fehler passiert ist (Abschn. 8.2.1, 8.3.2.1, 8.5.2). Die korrekte Interpretation der zu zeigenden Aussage ist eine Voraussetzung für den Beweis derselben. Darüberhinausgehend deuten Ergebnisse von Kirsten (2021) darauf hin, dass ein reichhaltiges Verständnis der Aufgabenstellung eine erfolgreiche Beweiskonstruktion begünstigt. Aus diesen Gründen erscheint es sinnvoll, im Rahmen einer Förderung im Führen von Beweisen einen besonderen Fokus auf das Verstehen der Aufgabenstellung legen. Aktivitäten, die eine korrekte Interpretation der zu zeigenden Aussage und ein darüberhinausgehendes reichhaltiges Verständnis derselben fördern, sind u. a. Paraphrasierungen, Visualisierungen, Beispielbetrachtungen und Plausibilitätsüberprüfungen. Außerdem kann es sehr hilfreich sein, die logische Struktur der zu zeigenden Aussage, die gegebenen Voraussetzungen und die Behauptung herauszuarbeiten. Zudem kann es helfen, die Relevanz von Details der zu zeigenden Aussage zu hinterfragen, Vorwissen, wie z. B. Definitionen, zum zu beweisenden Sachverhalt zu aktivieren und mit der zu zeigenden Aussage zu verknüpfen sowie erste Bezüge zu ähnlichen, bereits bearbeiteten Aufgaben herzustellen (Kirsten, 2021).

Ein weiterer Fokus könnte im Rahmen einer Förderung im Führen von Beweisen auf **der sinnvollen Nutzung von Beispielen** liegen. Beispielsweise gibt es in jeder der Klassenstufen 5 bis 8 Beweisversuche mit prüfendem Ansatz, in denen definitionsgemäß mindestens ein Beispiel mit einem Bezug zu der zu zeigenden Aussage vorkommt, aber kein besonderer Anknüpfungspunkt für die Entwicklung eines korrekten und vollständigen Argumentgefüges zu erkennen ist. Auch gibt es in jeder der Klassenstufen Beweisversuche mit prüfendem Ansatz und einem in den Beispielen angelegten weitgehend korrekten Argumentgefüge, welches jedoch nicht erkannt oder nicht ausgeführt wurde. Weiterhin kommen in jeder der Klassenstufen Beweisversuche mit der Schlussweise deduktiv mit zusätzlichem Beispiel und allenfalls einem Ansatz für ein weiterentwickelbares Argumentgefüge vor, bei denen zwar ein Beispiel angegeben, jedoch vermutlich weder für die Entwicklung noch für die Darstellung einer Beweisidee genutzt wurde (Abschn. 8.2.6, 8.3.4, 8.5.3). Alleine die Beweisversuche mit den genannten Ausprägungen machen in jeder der Klassenstufen 5 bis 7 rund 20% bis 23% aller Beweisversuche aus und in der Klassenstufe 8 rund 17% (Abschn. 8.2.6.4, 8.5.3.2). Die Ergebnisse deuten also darauf hin, dass in jeder der Klassenstufen 5 bis 8 einige Teilnehmende ein Beispiel angegeben haben, ohne das Potential der Arbeit mit Beispielen für die Entwicklung und Formulierung von Beweisen voll auszuschöpfen. Zugleich liefern die Ergebnisse dieser Arbeit, in Einklang mit der mathematikdidaktischen Theorie und ersten anderen Forschungsergebnissen, Hinweise darauf, dass die Existenz, Auswahl und Nutzung von Beispielen unter anderem mit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in Beweisversuchen zusammenhängt (Abschn. 3.3.5, 5.2, 8.2.6.4, 8.3.2.2, 8.5.2, 8.5.3.2). Vor diesem Hintergrund erscheint es lohnend, im Rahmen einer Förderung im Führen von Beweisen einen Fokus auf eine sinnvolle Nutzung von Beispielen zu legen. Es könnte darum gehen, dass Lernende ein Beispiel und insbesondere mehrere sinnvoll ausgewählte Beispiele, wie z. B. systematische Beispiele, extreme Beispiele, nicht allzu spezielle Beispiele und ggf. Klassen von Beispielen, betrachten und diese gezielt nutzen, um Invarianten zu erkennen, Beweisideen zu generieren und zu kommunizieren sowie Überlegungen zu kontrollieren.

Was die Darstellungsmittel betrifft, lässt sich in allen Klassenstufen primär an die Nutzung verbaler und numerischer Darstellungsmittel anknüpfen. Nach den Ergebnissen dieser Arbeit ist der produktive Einsatz ikonischer Darstellungsmittel bei jungen Lernenden zu Beginn der Sekundarstufe I besonders schwach ausgeprägt und erfolgt in allen Klassenstufen 5 bis 8 eher punktuell. Die Möglichkeiten, die die Verwendung ikonischer Darstellungsmittel bietet, werden selten ausgeschöpft (Abschn. 8.2.4, 8.2.6.5, 8.3.2.4, 8.4.2, 8.5.3). Die Verwendung ikonischer Darstellungsmittel erscheint aus theoretischer Sicht vor allem für junge Lernende als Chance, um an Beispielen Beweisideen herauszuarbeiten und darzustellen (Abschn. 3.3.11, 6.2.2). Von den analysierten Beweisversuchen mit einer Darstellungsform mit ikonisch-numerischem Anteil enthalten mehr als drei Viertel ein weiterentwickelbares Argumentgefüge, wenngleich es sich nicht unbedingt um einen Kausalzusammenhang handelt (Abschn. 8.2.6.5, 8.3.2.4). Um Teilnehmende der Mathematik-Olympiade im Führen von Beweisen zu fördern, könnte es durchaus sinnvoll sein, einen Fokus auf die **Verwendung ikonischer Darstellungsmittel** zu legen. Als Grundlage dafür erscheint es vor dem Hintergrund der Ergebnisse dieser Arbeit erfor-

derlich, einen Pool an Ikonisierungen, die im Themenbereich der Teilbarkeit ganzer Zahlen hilfreich sein können, zusammenzustellen. Es ist gut möglich, dass viele Lernende z. B. figurierte Zahlen noch nie gesehen haben, geschweige denn selbstständig und produktiv etwa für die Visualisierung der zu zeigenden Aussage und die Entwicklung einer Beweisidee nutzen können.

Ein Ergebnis im Rahmen dieser Arbeit ist es, dass in jeder der Klassenstufen 5 bis 8 ca. 45% bis 80% der Teilnehmenden einen Beweisversuch mit begründendem Ansatz abgegeben haben. Viele Teilnehmende haben folglich nach allgemeingültigen Argumenten gesucht, zumindest irgendeine Invariante identifiziert und formuliert. Daran lässt sich im Rahmen einer Förderung anknüpfen. Zugleich ist oftmals die Tragfähigkeit des abgegebenen Argumentgefüges oder dessen Vollständigkeit zu hinterfragen. Schließlich machen in jeder der Klassenstufen Beweisversuche mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge über 20% aller Beweisversuche aus. Der Anteil derjenigen mit einem weitgehend korrekten, aber keinem korrekten und vollständigen Argumentgefüge liegt in den Klassenstufen 6, 7 über 10%, in der Klassenstufe 8 über 30% (Abschn. 8.2.1, 8.5.3). Demnach sollte im Rahmen einer Förderung im Führen von Beweisen ein Fokus auf der **Haltung eines inneren Skeptikers** liegen, die es begünstigt, dass Beweisversuche auf inhaltlicher, struktureller und sprachlicher Ebene hinterfragt, kontrolliert und optimiert werden. Dazu wären Anforderungen an mathematische Beweise zu thematisieren, um einen angemessenen und verständnisorientierten Umgang mit Beweisversuchen zu fördern. Auf elementarem Niveau könnte es dann etwa darum gehen, die Passung der Ausführungen zu der zu zeigenden Aussage zu kontrollieren sowie sachinhaltlich falsche Invarianten und Invarianten, die für einen Beweis einer zu zeigenden Aussage nicht hinreichend sind, zu identifizieren. Auf einem fortgeschrittenen Niveau könnte das Ziel darin bestehen, inhaltliche Lücken in Argumentgefügen bis hin zu kleinen Ungenauigkeiten in der Formulierung zu erkennen und zu beheben. Die Förderung der Haltung eines inneren Skeptikers bedeutet in den Termini von Kirsten (2021), der Phase des Validierens in Bearbeitungsprozessen von Beweisaufgaben mehr Gewicht zu geben und die Qualität von Aktivitäten des Überprüfens von Beweisversuchen und die Identifikation von Bearbeitungsbedarf zu verbessern. An dieser Stelle ist zudem eine Verbindung zur mathematikdidaktischen Theorie und Forschung zum Problemlösen hervorzuheben. In mehreren Prozessmodellen zum Problemlösen ist eine Phase der Verifikation (Kontrolle, Rückschau) ausgewiesen, wobei empirischen Studien zufolge Schülerinnen und Schüler in den Klassenstufen 5, 6 ihre Ergebnisse kaum überprüfen (De Corte & Somers, 2014; Pólya, 1949; Rott, 2014; Schoenfeld, 1985).

Die Ergebnisse der Analyse der Beweisversuche im Rahmen dieser Arbeit verdeutlichen, dass in der Klassenstufe 5 bei fast allen Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge die Schlussweise zwischen induktiv und deduktiv liegt, in den Klassenstufen 6 bis 8 jeweils in über 60% der Fälle (Abschn. 8.2.1). Dabei stellt sich vor allem mit Blick auf Teilnehmende zu Beginn der Sekundarstufe I die Frage, inwieweit die **Formulierung deduktiver Argumentgefüge** angestrebt werden sollte oder Beweisversuche mit einem weitgehend korrekten Argumentgefüge und zwischen induktiver und deduktiver Schlussweise als sehr gute Beweisversuche angesehen werden sollten (Abschn. 8.5.3.2). Es könnte lohnend sein, Schülerinnen und Schüler an die Formulierung deduktiver Argumentgefüge ohne Variablen früh heranzuführen, damit sie den Anforderungen an einen mathematischen Beweis

gerecht werden können und damit sie auf diese Option zurückgreifen, auch wenn sie mit dem Umgang mit elementarer Algebra vertraut sind. Wenn man im Rahmen einer Förderung im Führen von Beweisen die Formulierung deduktiver Argumentgefüge anstrebt, ist zu bedenken, dass den Lernenden die Wichtigkeit deduktiven Schließens vielleicht nicht bekannt ist. Außerdem ist zu berücksichtigen, dass der präzise Umgang mit Wortvariablen und weiterer Fachsprache anspruchsvoll sein kann und eventuell guter Unterstützung bedarf, so dass eine Förderung der Formulierung deduktiver Argumentgefüge im Bereich des Beweisens und insbesondere auch im Bereich der Sprachförderung einzuordnen ist.

Schließlich ist mit Blick auf Ansatzpunkte für eine Förderung hervorzuheben, dass in jeder der Klassenstufen 6 bis 8 einzelne Teilnehmende die zu zeigende Aussage bewiesen und weitere gut 5% bis 20% der Teilnehmenden einen Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge formuliert haben. Insbesondere mit Fokus auf der Klassenstufe 8 ist davon auszugehen, dass manche Teilnehmende der Mathematik-Olympiade im Bereich des Führens von Beweisen so weit sind, dass sie einfache mathematische Allaussagen beweisen können. Ihnen würde etwa die Förderung von Monitoring-Aktivitäten in Bezug auf die Korrektheit und Vollständigkeit ihres Argumentgefüges sicherlich keineswegs schaden. Zugleich scheinen sie **größere Herausforderungen**, wie z. B. anspruchsvollere Aufgaben, zu brauchen.

Die genannten Ansatzpunkte etwa in Bezug auf die Nutzung von Beispielen und die Verwendung ikonischer Darstellungsmittel sind nicht unabhängig voneinander, sondern können sich gegenseitig beeinflussen. Im Rahmen einer Förderung wäre es möglich, diese in Schwerpunktsetzungen zu thematisieren. Eine Herangehensweise bestünde darin, die genannten Punkte, wie z. B. die Nutzung von Beispielen und die Formulierung deduktiver Argumentgefüge, bausteinartig in verschiedenen Stationen zu adressieren.

Wie bereits angedeutet, erscheinen mit Blick auf die **Klassenstufen 5, 6** eine Unterstützung in Bezug auf das Verstehen von Aufgabenstellungen, eine Fokussierung auf die Nutzung von Beispielen, eine Einführung in die Verwendung ikonischer Darstellungsmittel und eine Förderung von Monitoring-Aktivitäten insbesondere in Bezug auf die Korrektheit und Verallgemeinerbarkeit von Argumentgefügen als sinnvoll. In den **Klassenstufen 7, 8** sind diese Ansatzpunkte ebenfalls von Bedeutung und für einige Teilnehmende gleichermaßen relevant wie für viele der Teilnehmenden aus der Klassenstufe 5. Zugleich erscheint es vor allem in der **Klassenstufe 8** möglich und sinnvoll, einen besonderen Fokus erstens auf die Haltung eines inneren Skeptikers und zweitens auf die Formulierung deduktiver Argumentgefüge zu legen.

In Anbetracht des großen Spektrums an Beweisversuchen innerhalb jeder der Klassenstufen 5 bis 8 erscheint es wichtig, eine Förderung im Führen von Beweisen **binnendifferenziert** anzulegen, um auch an die vorhandenen Kompetenzen schwächerer Teilnehmender anzuknüpfen und zugleich allen Herausforderungen anzubieten. Dabei lassen sich u. a. durch das Angebot unterschiedlich schwieriger Aufgaben, die z. B. hinsichtlich ihrer Abstraktheit, dem erforderlichen Bereichswissen, im Bereich der Heurismen, der Komplexität des notwendigen Argumentgefüges und der Offenheit der Aufgabenstellung differieren, verschiedene Ausgangslagen be-

rücksichtigen. Förderbausteine und darüberhinausgehend ein Förderangebot für Teilnehmende der Mathematik-Olympiade auszuarbeiten, durchzuführen, zu evaluieren und zu optimieren wäre ein eigenes, lohnendes Forschungsprojekt. Dieses könnte nun auf Grundlage der Ergebnisse dieser Arbeit fundiert angegangen werden.

9.4 Zusammenfassung des neunten Kapitels

Diskussion

In der vorliegenden Arbeit werden über 1000 authentische Beweisversuche zu einer Allaussage aus dem Themenbereich der Teilbarkeit ganzer Zahlen untersucht, die von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 formuliert wurden. Die vorliegende Arbeit verbindet die etablierte **mathematikdidaktische Forschung zum Führen von Beweisen** mit der bisher kaum betrachteten Gruppe der **Teilnehmenden der Mathematik-Olympiade**. Mit Blick auf diese Zielgruppe und insbesondere auf die Klassenstufen 5 und 6 ergänzt diese Arbeit die bisherigen Erkenntnisse zu Beweisversuchen von Schülerinnen und Schülern. Zudem konnten u. a. hinsichtlich Zusammenhängen zwischen verschiedenen Eigenschaften von Beweisversuchen neue Erkenntnisse gewonnen werden.

Wesentliche **Limitationen der vorliegenden Arbeit** resultieren aus der alleinigen Fokussierung auf Beweisversuche und auf eine einzige Aufgabe. Bezüglich der konstruierten, für den Bereich der Teilbarkeit ganzer Zahlen in mehrerlei Hinsicht charakteristischen Aufgabe ist zu erwähnen, dass das angegebene Beispiel und die leichten Unterschiede in der Aufgabenstellung für die verschiedenen Klassenstufen eventuell manche Ergebnisse beeinflusst haben. Außerdem kann es sein, dass eine verschieden starke Positivauswahl der Teilnehmenden der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 zu Unterschieden zwischen den Beweisversuchen aus den verschiedenen Klassenstufen beigetragen hat.

Die Ergebnisse der vorliegenden Arbeit lassen sich nicht unmittelbar auf Teilnehmende der Mathematik-Olympiade in der gesamten Bundesrepublik übertragen oder etwa auf alle Gymnasiastinnen und Gymnasiasten ausweiten. Zugleich ist davon auszugehen, dass zumindest ein Teil der Ausprägungen der Beweisversuche von Teilnehmenden der Mathematik-Olympiade auch im Falle anderer Untersuchungspopulationen vorkämen.

Das im Rahmen der Anwendung des Verfahrens der qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) entwickelte Kategoriensystem und die erarbeiteten Klassifikationen ermöglichen es, Beweisversuche zu der zu zeigenden Aussage aus verschiedenen Perspektiven differenziert zu beschreiben und zu vergleichen. Zudem ist es möglich, Zusammenhänge zwischen verschiedenen Eigenschaften dieser Beweisversuche zu analysieren. Das Kategoriensystem inklusive des damit einhergehenden Begriffsnetzes und die beiden Klassifikationen können als die wesentlichen **Beiträge der vorliegenden Dissertation zur mathematikdidaktischen Theoriebildung und Diskussion** angesehen werden.

Die **empirischen Ergebnisse** der durchgeführten Analyse von Beweisversuchen beschreiben entsprechend den Forschungsfragen, inwieweit es Teilnehmerinnen und Teilnehmern der

Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 8 gelungen ist, die vorgegebene mathematische Allaussage zu beweisen. Zudem legen sie dar, welche Charakteristika deren Beweisversuche auszeichnen und welche Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind. Dabei zeigen sie auf, welche Gemeinsamkeiten und Unterschiede mit Blick auf die verschiedenen Klassenstufen bestehen. Im Vergleich zu den curricularen Vorgaben in den Bildungsstandards und den Lehrplänen in NRW überraschen u. a. die Ergebnisse bezüglich der Beispielnutzung insbesondere in Beweisversuchen aus der Klassenstufe 5. Auch waren der große Anteil an Beweisversuchen mit deduktiver Schlussweise in der Klassenstufe 5 und der große Anteil der Beweisversuche mit zwischen induktiver und deduktiver Schlussweise in der Klassenstufe 8 nicht unbedingt zu erwarten. Zudem ist im Vergleich zu Ergebnissen anderer Studien hervorzuheben, dass der Anteil der Beweisversuche mit prüfendem Ansatz in jeder der Klassenstufen 5 bis 8 unter 25% liegt und viele Teilnehmende einen Beweisversuch mit begründendem Ansatz formuliert haben. Die im Rahmen dieser Arbeit erhaltenen Ergebnisse stehen in mehrerlei, aber nicht in jeder Hinsicht in Einklang mit den Erkenntnissen aus anderen empirischen Untersuchungen und gehen insbesondere im Bereich der Beweisideen und in Bezug auf Beweisversuche von Lernenden in den Klassenstufen 5 bis 6 über den bisherigen Forschungsstand hinaus.

Vor dem Hintergrund der erhaltenen Ergebnisse erscheint es sinnvoll, eine Förderung von Teilnehmenden der Mathematik-Olympiade im Führen von Beweisen möglichst binnendifferenziert und eventuell bausteinartig anzulegen. Dabei ergibt sich aus den empirischen Ergebnissen ein klassenstufenbezogener Fokus auf das Verstehen von Aufgabenstellungen, eine sinnvolle Nutzung von Beispielen, die Verwendung ikonischer Darstellungsmittel, die Haltung eines inneren Skeptikers und die Formulierung deduktiver Argumentgefüge. Die vorliegende Arbeit liefert somit Ansatzpunkte für eine Förderung von Teilnehmenden der Mathematik-Olympiade im Führen von Beweisen. Diese Ansatzpunkte stellen den wesentlichen **praktischen Nutzen** der vorliegenden Arbeit dar. Außerdem können das Kategoriensystem und die Klassifikationen eine Grundlage für qualifizierte Rückmeldungen an Lernende und eine gezielte Unterstützung von Schülerinnen und Schülern bieten. Zudem ist es möglich, die authentischen Beweisversuche, die in der vorliegenden Arbeit abgebildet sind, in der Aus- und Fortbildung von Lehrkräften einzusetzen.

10 Ausblick

Anknüpfend an die Diskussion der im Rahmen dieser Arbeit durchgeführten empirischen Studie, werden in diesem Kapitel mögliche Untersuchungsaspekte, die im Anschluss an diese Arbeit bearbeitet werden könnten, gebündelt und strukturiert. Dabei wird dargelegt, welche dieser Untersuchungsaspekte auf Grundlage der im Rahmen des Forschungsprojekts bereits erhobenen Daten bearbeitet werden könnten (Abschn. 10.1). Dieses Forschungsprojekt zum Führen von Beweisen bei der Mathematik-Olympiade in Deutschland wird anschließend näher beschrieben. Es werden Einblicke in die Anlage und weitere Auswertung dieses Projekts gegeben (Abschn. 10.2-10.5).

10.1 Anregungen für zukünftige Forschungsarbeiten

Um Anregungen für zukünftige Forschungsarbeiten zu geben, werden in diesem Abschnitt mögliche Untersuchungsaspekte in strukturierter und gebündelter Form beschrieben. Dabei wird angegeben, wenn ein Untersuchungsaspekt auf Grundlage bereits erhobener Daten bearbeitet werden könnte oder bereits teilweise bearbeitet wurde.

Die nachfolgenden Ausführungen zu möglichen Untersuchungsaspekten sind im Wesentlichen in drei Teile gegliedert: Grundlegend stellt sich die Frage, welche Anforderungen bei der Mathematik-Olympiade im Bereich des Führens von Beweisen gestellt werden, um die Ergebnisse dieser Arbeit im Vergleich zu den regulären Wettbewerbsanforderungen einordnen zu können (Abschn. 10.1.1). Zudem wäre es lohnend, Beweisversuche unterschiedlicher Zielgruppen zu vergleichen sowie Beweisversuche zu unterschiedlichen Aussagen auszuwerten. Auf diese Weise ließen sich weitere Einblicke in Ansatzpunkte für eine Förderung im Führen von Beweisen und Einblicke in die Verallgemeinerbarkeit der Ergebnisse dieser Arbeit generieren (Abschn. 10.1.2-10.1.4). Drittens wäre es möglich, Beweisvorstellungen oder Beweisprozesse von Teilnehmenden der Mathematik-Olympiade zu untersuchen, um die Ergebnisse der Analyse von Beweisversuchen mit Blick auf Ansatzpunkte für eine Förderung zu ergänzen (Abschn. 10.1.5).

10.1.1 Anforderungen bei der Mathematik-Olympiade

In der Diskussion wurden die Erkenntnisse zu Beweisversuchen von Teilnehmenden der Mathematik-Olympiade vor dem Hintergrund von curricularen Vorgaben und Ergebnissen anderer empirischer Studien eingeordnet. Sie wurden jedoch nicht im Vergleich zu den regulären Wettbewerbsanforderungen diskutiert. Dazu wäre es notwendig, zuerst die Anforderungen, die bei der Mathematik-Olympiade im Bereich des Führens von Beweisen gestellt werden, zu analysieren.

Im Rahmen des dieser Arbeit übergeordneten Forschungsprojekts wurde bereits eine solche Anforderungsanalyse durchgeführt (Abschn. 10.2, 10.3). Diese Anforderungsanalyse könnte im Anschluss an die vorliegende Arbeit vollständig verschriftlicht werden. Darauf aufbauend wäre

es möglich, die Ergebnisse dieser Arbeit vor dem Hintergrund der regulären Wettbewerbsanforderungen zu diskutieren.

10.1.2 Beweisversuche unterschiedlicher Teilgruppen der Teilnehmenden der Mathematik-Olympiade

In der Analyse von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade im Rahmen dieser Arbeit standen die Klassenstufen 5 bis 8 im Fokus. Die Klassenstufen 5 bis 8 wurden ausgewählt, um Beweisversuche von Lernenden mit unterschiedlichen Vorerfahrungen im Führen von Beweisen und unterschiedlichen Darstellungsmöglichkeiten zu vergleichen (Abschn. 1.1, 6.1). Bei den Ergebnissen der Analyse von Beweisversuchen fällt auf, dass in jeder der Klassenstufen 5 bis 8 der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil unter 15% beträgt. Zudem fällt auf, dass in den Klassenstufen 7, 8 in fast der Hälfte aller Beweisversuche die Schlussweise zwischen induktiv und deduktiv liegt (Abschn. 8.5.2). Es stellt sich die Frage, inwieweit sich diesbezüglich bei Beweisversuchen von Teilnehmenden der Mathematik-Olympiade **ab der Klassenstufe 9**, d. h. bei Teilnehmenden, die mit dem Umgang mit Variablen und dem Beweisen von Allaussagen aus dem Schulunterricht vertrauter sind, ein anderes Bild zeigen würde. Diese Frage ließe sich auf Grundlage bereits erhobener Daten beantworten (Abschn. 10.2, 10.4.1).

Weiterhin hat die Analyse der Beweisversuche gezeigt, dass in jeder der Klassenstufen 5 bis 8 ein großes Spektrum unterschiedlicher Charakteristika vorkommt (Kap. 8). Daher erscheint die Frage interessant, inwieweit Aspekte dieser Vielfalt an Beweisversuchen mit einem besonderen **Erfolg im Wettbewerb** oder einer besonderen **mathematischen Förderung** einhergehen. Zudem wäre es bedeutsam zu wissen, inwieweit sich Beweisversuche von Teilnehmenden der Mathematik-Olympiade je nach Wettbewerbserfolg und mathematischer Förderung unterscheiden. Die Beantwortung dieser Frage könnte Hinweise darauf liefern, inwieweit die verschiedenen Charakteristika bei unterschiedlichen Teilgruppen der Teilnehmenden der Mathematik-Olympiade zu erwarten sind, was mit Blick auf Ansatzpunkte für eine Förderung von Relevanz ist. Eine Analyse der Beweisversuche zusammen mit dem Erfolg der Schülerinnen und Schüler bei der Mathematik-Olympiade und deren Teilnahme an mathematischen Förderangeboten wäre auf Grundlage bereits erhobener Daten möglich (Abschn. 10.2, 10.4.2).

Ein weiterer Untersuchungsaspekt könnte darin bestehen, die Beweisversuche hinsichtlich **Geschlechterunterschieden** zu analysieren. Geschlechterunterschiede werden im Kontext mathematischer Schülerwettbewerbe immer wieder diskutiert (z. B. Applebaum et al., 2013; Leder & Taylor, 2010). Um verstärkt Mädchen in Mathematik zu fördern, wurde u. a. die European Girls' Mathematical Olympiad (EGMO) eingeführt und, fokussiert auf Deutschland, z. B. im November 2022 ein Mathe-Mädchen-Mint-Seminar durchgeführt (Bildung und Begabung gGmbH, 2023; Smith, 2012). In Deutschland beträgt der Mädchenanteil bei der Regionalrunde der Mathematik-Olympiade für Schülerinnen und Schüler ab der Klassenstufe 5 ungefähr 43%. Der Mädchenanteil wird von der Tendenz her geringer, je fortgeschrittener die Wettbewerbsrunde

und je höher die Klassenstufe.[163] Dazu passend deuten manche Studien darauf hin, dass Jungen mathematische Problemlöseaufgaben etwas besser bearbeiten als Mädchen und dass die Geschlechterunterschiede bei mathematischen Kompetenzen in höheren Klassenstufen deutlicher ausgeprägt sind als etwa am Anfang der Schulzeit (Hyde et al., 1990, 2008; Kesting, 2005; Prenzel et al., 2005). Was Beweisversuche von Schülerinnen und Schülern ab der Klassenstufe 7 betrifft, deuten mehrere Studien darauf hin, dass insbesondere bei Bearbeitungen geometrischer Beweisaufgaben keine statistisch signifikanten Geschlechterunterschiede bestehen (Healy & Hoyles, 2000; Hoyles & Healy, 2007; Heinze et al., 2007; Senk, 1982; Abschn. 5.1.2). Beobachtungen von Benölken (2008) beim Bearbeiten mathematischer Problemaufgaben durch Grundschulkinder deuten Unterschiede hinsichtlich der Darstellung der Bearbeitungen an. Der bisherige Forschungsstand zeigt folglich keine klare Tendenz. Insbesondere wurde noch nicht untersucht, inwieweit sich Beweisversuche von Mädchen und Jungen zu Beginn der Sekundarstufe I etwa hinsichtlich der Darstellungsform oder der primären Beweisidee unterscheiden. Einblicke in Gemeinsamkeiten und Unterschiede zwischen Beweisversuchen von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade könnten Hinweise für eine geschlechtssensible Förderung im Führen von Beweisen geben.

Schließlich wäre es interessant, Beweisversuche von Teilnehmenden der Mathematik-Olympiade in Zusammenhang mit dem **Studienwunsch** der Lernenden auszuwerten. Dies würde Einblicke geben, wie insbesondere Beweisversuche von Teilnehmenden, die eventuell Mathematik studieren möchten, ausgeprägt sind. Hierzu könnten bereits erhobene Daten herangezogen werden. Insbesondere wäre es auch möglich, Beweisversuche von Teilnehmenden der Mathematik-Olympiade in der Sekundarstufe II diesbezüglich auszuwerten (Abschn. 10.2).

10.1.3 Beweisversuche unterschiedlicher Zielgruppen

In der vorliegenden Arbeit sowie im übergeordneten Forschungsprojekt stehen Teilnehmerinnen und Teilnehmer der Mathematik-Olympiade im Fokus. Dabei handelt es sich um eine besondere Untersuchungspopulation. Es ist zu vermuten, dass zumindest einige der Eigenschaften der Beweisversuche der Teilnehmenden der Mathematik-Olympiade auch in vergleichenden Untersuchungen mit **anderen Schülerinnen und Schülern** vorkämen, wahrscheinlich mit anderen Häufigkeitsverteilungen (Abschn. 9.1.1). Um diesbezüglich nähere Einblicke zu erhalten, könnten z. B. von Lernenden fünfter bis achter Klassen eines Gymnasiums Beweisversuche zu der im Rahmen dieser Arbeit fokussierten Allaussage erhoben werden. Dabei wäre es u. a. spannend zu sehen, inwieweit diese Schülerinnen und Schüler Beweisversuche mit begründendem Ansatz formulieren, weil der im Rahmen der vorliegenden Arbeit festgestellte Anteil an Beweisversuchen mit begründendem Ansatz größer ist als in anderen Studien (Abschn. 9.3.1). Darüber hinaus wäre es interessant, welche Charakteristika die Beweisversuche von Gymnasiastinnen und Gymnasiasten insgesamt auszeichnen und inwieweit Anknüpfungspunkte für eine Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind. Dazu könnte das im Rahmen dieser Arbeit entwickelte Kategoriensystem genutzt werden. Die Ergebnisse

[163] Informationen zu den Teilnehmerzahlen bei der Mathematik-Olympiade können bei der Geschäftsstelle der Mathematik-Olympiade angefragt werden.

ließen sich mit Blick auf eine Förderung im Führen von Beweisen auch von Schülerinnen und Schülern, die nicht an der Mathematik-Olympiade teilnehmen, diskutieren.

Dabei stellt sich ergänzend die Frage, inwieweit zwischen Beweisversuchen aus verschiedenen Klassenstufen eines Gymnasiums ähnliche Unterschiede bestehen wie bei den Beweisversuchen von Teilnehmenden der Mathematik-Olympiade. Dies ist nicht unbedingt klar, weil eine unterschiedliche „Selektion“ zu den mit Blick auf Teilnehmende der Mathematik-Olympiade festgestellten Unterschieden beigetragen haben kann (Abschn. 9.1.1). Zudem deuten zwar auch einzelne andere **Querschnittsstudien** darauf hin, dass Schülerinnen und Schüler höherer Klassenstufen Aufgaben mit Beweisnotwendigkeit im Mittel ähnlich gut oder besser bearbeiten, Längsschnittstudien mit Lernenden siebter bis zehnter Klassen zeigen allerdings keine klare positive Tendenz (Abschn. 5.1). In Ergänzung zu einer Querschnittsstudie wäre es möglich, eine **Längsschnittstudie** durchzuführen, um Einblicke in intraindividuelle Unterschiede zwischen Beweisversuchen von Schülerinnen und Schülern im Verlauf der Klassenstufen 5 bis 8 zu erhalten. Solche Einblicke würden den Forschungsstand erweitern und könnten zu einer Diskussion der derzeit üblichen Praxis hinsichtlich der Förderung der Teilkompetenz des Führens von Beweisen beitragen.

10.1.4 Beweisversuche zu unterschiedlichen Aufgabenstellungen

In der Diskussion dieser Arbeit ist die Frage aufgekommen, inwieweit verschiedene **Merkmale von Aufgaben mit Beweisnotwendigkeit**, wie z. B. der gewählte Operator und das Aufgabenformat, mit Unterschieden in den Bearbeitungen dieser Aufgaben einhergehen (Abschn. 9.1.3). Diesbezügliche Einblicke könnten für die Formulierung von Aufgaben, etwa im Rahmen einer Förderung, wertvolle Erkenntnisse liefern.

Mit Blick auf die Frage nach der Verallgemeinerbarkeit der Ergebnisse dieser Arbeit wäre es außerdem lohnend, Beweisversuche von Teilnehmenden der Mathematik-Olympiade zu **verschiedenen Allaussagen mit unendlichem Gültigkeitsbereich** zu analysieren. Dabei würde auch die Eignung und Anpassungsfähigkeit des entwickelten Kategoriensystems erprobt. Zudem ließe sich in diesem Zuge untersuchen, inwieweit sich die im Rahmen dieser Arbeit festgestellten Zusammenhänge zwischen verschiedenen Eigenschaften von Beweisversuchen auch in Beweisversuchen zu anderen Allaussagen zeigen. Des Weiteren wäre es möglich, Beweisversuche zu anderen Allaussagen ebenfalls mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu klassifizieren. Dies könnte ermöglichen, übergreifende Muster zu sehen und von aufgabenspezifischen Klassifikationen z. B. zu einer bereichsspezifischen Klassifikation zu kommen (Abschn. 9.1.3, 9.2).

Zusätzlich zu der Analyse von Beweisversuchen zu verschiedenen Allaussagen mit unendlichem Gültigkeitsbereich stellt sich die Frage, inwieweit Beweisversuche von Teilnehmenden der Mathematik-Olympiade zu **Allaussagen mit endlichem Gültigkeitsbereich** und solche zu **Existenzaussagen** ähnliche Eigenschaften aufweisen. Dabei wären sowohl inter- als auch intraindividuelle Vergleiche von Interesse, um weitere Einblicke in Ansatzpunkte für eine Förderung im Führen von Beweisen zu generieren (Abschn. 9.1.3). Eine Analyse von Beweisversuchen von Teil-

nehmenden der Mathematik-Olympiade zu Allaussagen mit endlichem Gültigkeitsbereich erscheint zudem vor dem Hintergrund der regulären Wettbewerbsanforderungen sehr relevant. Die ersten Einblicke in eine Analyse von Anforderungen bei der Mathematik-Olympiade in Abschnitt 10.3 zeigen, dass Aussagen mit endlichem Gültigkeitsbereich in den Klassenstufen 5 bis 7 häufig vorkommen und somit auch im Rahmen einer Förderung im Führen von Beweisen von Bedeutung sein könnten und sollten. Eine Analyse von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade zu Existenzaussagen wäre auch vor dem Hintergrund der Vorgaben in den Bildungsstandards und den Lehrplänen für Nordrhein-Westfalen interessant, weil es demnach allen Lernenden nach der Grundschulzeit gelingen sollte, Beispiele und Gegenbeispiele anzugeben (Abschn. 4.2, 9.1.3). Von den Teilnehmenden der Mathematik-Olympiade, deren Beweisversuche in dieser Arbeit analysiert wurden, liegen ebenfalls Beweisversuche zu einer Allaussage mit endlichem Gültigkeitsbereich sowie Beweisversuche zu zwei Existenzaussagen vor. Diese ließen sich auswerten und miteinander vergleichen (Abschn. 6.2, 10.2).

10.1.5 Beweisvorstellungen und Beweisprozesse

Die analysierten Beweisversuche lassen insbesondere mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung einen Interpretationsspielraum, der durch das rein schriftliche Vorliegen bedingt ist und sich auch nur auf das Führen von Beweisen bezieht (Abschn. 9.1.4). Um eine angemessene Interpretation der Ergebnisse der Analysen von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade zu unterstützen, wurden die Probandinnen und Probanden auch zum Validieren und Bewerten fiktiver Beweisversuche aufgefordert (Abschn. 6.2, 9.1.5, 10.2). Die Bearbeitungen der Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche könnten im Anschluss an diese Arbeit ausgewertet werden. Dies würde z. B. Einblicke geben, inwieweit Schülerinnen und Schüler, die einen Beweisversuch mit prüfendem Ansatz beim eigenständigen Führen von Beweisen formuliert haben, einen solchen Beweisversuch als Beweis einer vorgegebenen Aussage ansehen. Weiter gefasst könnte eine Auswertung der Bearbeitungen der Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche Hinweise liefern, **was Lernende als Beweis ansehen**. Dies ist mit Blick auf Ansatzpunkte für eine Förderung im Führen von Beweisen relevant, da eine angemessene Vorstellung davon, was einen mathematischen Beweis ausmacht, beim Führen von Beweisen zwar nicht unbedingt notwendig, aber zumindest hilfreich erscheint. Eine Untersuchung von Ufer et al. (2009) liefert empirische Hinweise darauf, dass ein Zusammenhang zwischen Bearbeitungen von Beweisaufgaben und Bearbeitungen von Aufgaben zum Validieren und Bewerten fiktiver Beweisversuche besteht (Abschn. 5.1.2.3).

Um die Ergebnisse der Analyse von Beweisversuchen mit Blick auf Ansatzpunkte für eine Förderung zu ergänzen, wäre es außerdem sinnvoll, **Beweisprozesse** zu beforschen. Auf diese Weise ließen sich z. B. vertiefte Einblicke gewinnen, warum Bearbeitungen fehlen und welche Funktion Beispielen in den Beweisprozessen zukommt. Zudem könnte z. B. untersucht werden, in welchem Umfang und in welcher Qualität Phasen des Validierens erfolgen. Die Analyse von Beweisprozessen könnte auch Einblicke in die Beweisvorstellungen der Schülerinnen und Schüler geben.

10.2 Überblick über ein Forschungsprojekt zum Führen von Beweisen bei der Mathematik-Olympiade

Von den in Abschnitt 10.1 angeführten Untersuchungsaspekten ließen sich mehrere auf Grundlage bereits erhobener Daten bearbeiten. Die vorliegende Arbeit beschreibt einen Ausschnitt eines umfangreicheren Forschungsprojekts zum Führen von Beweisen bei der Mathematik-Olympiade in Deutschland. In dem umfangreicheren Projekt werden Anforderungen, die bei der Mathematik-Olympiade in den Klassenstufen 5 bis 13 im Themenbereich der Teilbarkeit ganzer Zahlen gestellt werden, sowie Kompetenzen von Teilnehmenden dieses Wettbewerbs untersucht und die Ergebnisse mit Blick auf Ansatzpunkte für eine Förderung im Führen von Beweisen interpretiert (Abb. 176).

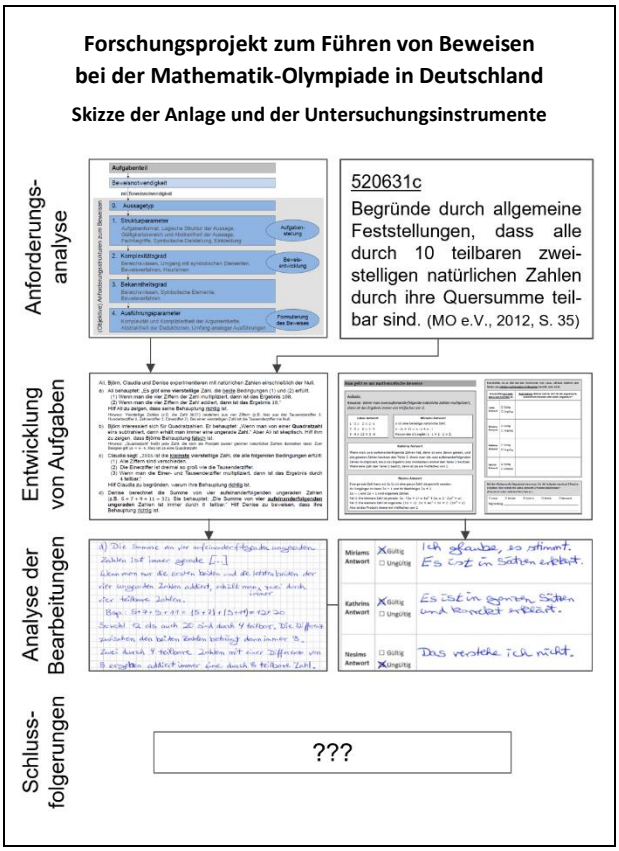


Abbildung 176: Forschungsprojekt zum Führen von Beweisen bei der Mathematik-Olympiade in Deutschland: Skizze der Anlage und der Untersuchungsinstrumente

Im Anschluss an die vorliegende Arbeit könnten die Auswertung und Verschriftlichung dieses Forschungsprojekts fortgeführt werden. Um dies genauer darzulegen, wird im Folgenden das umfangreichere Projekt beschrieben.

10.2.1 Anforderungen bei der Mathematik-Olympiade

In dem umfangreicheren Forschungsprojekt zur Mathematik-Olympiade in Deutschland geht es um das Führen von Beweisen im Themenfeld der ganzen Zahlen. Diesbezüglich werden in einem ersten Schritt **objektive Anforderungsstrukturen von Aufgaben** aus Mathematik-Olympiaden ab der Klassenstufe 5 herausgearbeitet. Dabei soll der Begriff „objektiv“ in Anlehnung an Bruder (1981) hervorheben, dass die Analyse nicht auf die individuellen Schwierigkeiten von Schülerinnen und Schülern abzielt, sondern auf die schwierigkeitsgenerierenden Merkmale der Aufgaben. Ziel ist es, die Anforderungen, die strukturell in den Aufgaben angelegt sind und in den Lösungen von Seiten der Verantwortlichen implizit angedacht wurden, herauszuarbeiten. Deshalb basiert die Analyse auf den Aufgaben der Mathematik-Olympiade und den dazu vom Aufgabenausschuss herausgegebenen Lösungen. Konkret werden alle Aufgabenteile aus dem Themenbereich der Teilbarkeit ganzer Zahlen, die vom Aufgabenausschuss für die 46. bis 55. Mathematik-Olympiade für die Klassenstufen 5 bis 12/13 herausgegeben wurden, untersucht.

Um diese Aufgabenteile hinsichtlich der objektiven Anforderungsstrukturen zum Beweisen zu untersuchen, wurde deduktiv-induktiv ein Analysemodell entwickelt und eine qualitativ-strukturierende Inhaltsanalyse nach Schreier (2012, 2014) durchgeführt. Die Analyseergebnisse beschreiben aus qualitativer und quantitativer Perspektive, welche Anforderungsstrukturen in den Aufgabenteilen zu finden sind. Sie verdeutlichen, welcher Art und auf welchem Niveau die Anforderungen sind, und zeigen auf, welche Unterschiede diesbezüglich zwischen den Klassenstufen, Wettbewerbsrunden und innerhalb von Wettbewerbsrunden bestehen. Für interessierte Leserinnen und Leser werden im nachfolgenden Abschnitt 10.3 Ausschnitte dieser Anforderungsanalyse, wie z. B. das Analysemodell, präsentiert.

10.2.2 Vorhandene Kompetenzen von Teilnehmenden der Mathematik-Olympiade

Die empirischen Ergebnisse der Anforderungsanalyse und das Analysemodell wurden im Rahmen des Forschungsprojekts genutzt, um Beweisaufgaben für eine Erhebung von **Beweisversuchen von Teilnehmenden der Mathematik-Olympiade** im Rahmen der Regionalrunde der 56. Mathematik-Olympiade zu konstruieren. Im zweiten Teil des Forschungsprojekts geht es darum, inwieweit es Teilnehmenden der Mathematik-Olympiade gelingt, mathematische Aussagen aus dem Themenbereich der Teilbarkeit ganzer Zahlen zu beweisen. Zudem wird untersucht, welche Charakteristika deren Beweisversuche auszeichnen und welche Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis zu erkennen sind. Dabei sollen die Ergebnisse sowohl Einblicke in das Spektrum an vorkommenden Ausprägungen geben als auch in deren Häufigkeitsverteilungen bei unterschiedlichen Teilgruppen der Teilnehmenden der Mathematik-Olympiade, wie z. B. Teilnehmenden in verschiedenen Klassenstufen und Teilnehmenden mit einer unterschiedlichen mathematischen Förderung. Mit diesem Ziel wurde bei der Regionalrunde der Mathematik-Olympiade in Nordrhein-Westfalen im November 2016 in jeder der Klassenstufen 5 bis 12/13 eine spezifisch konstruierte Beweisaufgabe in die Wettbewerbsklausur integriert. Die Aufgaben für die Klassenstufen 5 bis 8 sind in Abschnitt 6.2.2 präsentiert. In Ergänzung dazu zeigt die Abbildung 177 die Aufgabe, die ab der Klassenstufe 9 verwendet wurde. Die Aufgabenteile a), b) wurden im Wesentlichen aus der Aufgabe für die Klassenstufen 7 und 8 übernommen. Dies soll ermöglichen, Beweisversuche aus den verschiedenen Klassenstufen vergleichend zu beschreiben. Darüber hinaus kann bezogen auf Teilnehmende der Mathematik-Olympiade ab der Klassenstufe 9 untersucht werden,

Eingesetzte Beweisaufgabe für die Klassenstufen 9 bis 12/13

a) Widerlegen Sie: Wenn man von einer **Quadratzahl** eins subtrahiert, dann erhält man immer eine ungerade Zahl. (*Hinweis:* „Quadratzahl" heißt jede Zahl, die sich als Produkt zweier gleicher ganzer Zahlen darstellen lässt. Zum Beispiel gilt $16 = 4 \cdot 4$. Also ist 16 eine Quadratzahl.)

b) Beweisen Sie: Die Summe von vier **aufeinanderfolgenden ungeraden** positiven Zahlen ist immer durch 8 teilbar. (*Hinweis:* Beispielsweise gilt $5 + 7 + 9 + 11 = 32$ und $32 : 8 = 4$.)

c) Beweisen Sie: **Es gibt keine** natürlichen Zahlen x und y so, dass gilt: $4x^2 - 4y^2 = 56$.

d) Beweisen Sie: **Es gibt unendlich viele** natürliche Zahlen a mit der Eigenschaft, dass $3a^2 + 5$ durch 10 teilbar ist.

Abbildung 177: Eingesetzte Beweisaufgabe für die Klassenstufen 9 bis 12/13

inwieweit es ihnen gelingt, eine Allaussage zu widerlegen und eine Allaussage zu beweisen, die sich ohne symbolische Darstellungsmittel beweisen lässt. Beim Aufgabenteil c) handelt es sich um eine mit Blick auf die regulären Wettbewerbsanforderungen besonders charakteristische Beweisaufgabe. Es ist eine Allaussage mit unendlichem Gültigkeitsbereich zu beweisen, wobei eine Gleichung mit zwei Variablen vorgegeben ist. Der Beweis erfordert u. a. eine Reformulierung in der Darstellung, bei der die dritte binomische Formel hilfreich ist. Im Aufgabenteil d) ist im Vergleich zu den Aufgaben für die Klassenstufen 5 bis 8 keine spezielle Existenzaussage, sondern eine allgemeine Existenzaussage zu beweisen. Auch solche Aufgaben werden regulär in der Mathematik-Olympiade gestellt, die Anforderungen wurden an dieser Stelle an die Klassenstufen 9 bis 12/13 angepasst.

Zusätzlich zur Implementierung der spezifisch konstruierten Beweisaufgabe in die reguläre Wettbewerbsklausur der Regionalrunde der 56. Mathematik-Olympiade haben die Teilnehmenden direkt im Anschluss an die Klausur einen Fragebogen erhalten (Abschn. 6.2.1). Der **Fragebogen** umfasst Angaben zum Geschlecht, zur Teilnahme an mathematischen Förderangeboten und an der Mathematik-Olympiade sowie zum Studienwunsch der Schülerinnen und Schüler, um die angestrebte differenzierte Analyse der Beweisversuche zusammen mit weiteren Merkmalen der Teilnehmerinnen und Teilnehmer zu ermöglichen.

Außerdem enthält der Fragebogen **Aufgaben zum Validieren und Bewerten fiktiver Beweisversuche** (Abb. 178-179), um eine angemessene Interpretation der Beweisversuche der Teilnehmenden zu unterstützen und Einblicke in deren Beweisvorstellungen zu erhalten. Von diesen Aufgaben wurden zwei Versionen erstellt.

Die eine Version fokussiert auf Beweisversuche zu einer wahren Allaussage, die andere auf Beweisversuche zu einer falschen Allaussage. Die Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche zu einer wahren Allaussage (Abb. 178) wurde anknüpfend an ein Untersuchungsinstrument von Healy und Hoyles (2000) entwickelt. In der entwickelten Aufgabe sind vier fiktive Beweisversuche zu einer wahren Allaussage aus dem Themenbereich der Teilbarkeit ganzer Zahlen vorgegeben. Die zu beweisende wahre Allaussage lautet: „Wenn man zwei aufeinanderfolgende natürliche Zahlen multipliziert, dann ist das Ergebnis immer ein Vielfaches von zwei“ (Abb. 178). Im Vergleich zu der Untersuchung von Healy und Hoyles (2000) wurde eine etwas einfachere Aussage gewählt, so dass zumindest der fiktive Beweisversuch mit prüfendem Ansatz und der verbal dargestellte Beweis auch für Schülerinnen und Schüler in der Klassenstufe 5 verständlich sein sollte. Lukas Antwort ist ein Beweisversuch mit prüfendem Ansatz. Bei Miriams Antwort handelt es sich um einen Beweisversuch mit begründendem Ansatz, jedoch nicht um einen Beweis. Es handelt sich nicht um einen gültigen Beweis, weil die als „Kürzen der n's“ bezeichnete Termumformung unzulässig ist. Alternativ lässt sich die Ungültigkeit damit begründen, dass das Argumentgefüge nach dem Term „$n \cdot n + n \cdot 1$“ abbricht, die Teilbarkeit dieses Terms durch zwei ausschließlich für den Spezialfall $n = 1$ gezeigt wird. Die übrigen beiden Beweisversuche von Kathrin und Nesim können als Beweis der vorgegebenen Aussage bezeichnet werden. Bei Kathrins Beweisversuch ist die Darstellungsform verbal, bei Nesims Beweisversuch symbolisch-verbal.

Aufgabenstellung zum Validieren und Bewerten fiktiver Beweisversuche zu einer wahren Allaussage

Nun geht es um mathematische Beweise

Aufgabe

Beweise: *Wenn man zwei aufeinanderfolgende natürliche Zahlen multipliziert, dann ist das Ergebnis immer ein Vielfaches von 2.*

Lukas Antwort

$1 \cdot 2 = \ \ 2 = 2 \cdot 1$

$2 \cdot 3 = \ \ 6 = 2 \cdot 3$

$3 \cdot 4 = 12 = 2 \cdot 6$

Miriams Antwort

n ist eine beliebige natürliche Zahl.

$n \cdot (n+1) = n \cdot n + n \cdot 1$

Kürzen der n's ergibt: $1 \cdot 1 + 1 \cdot 1 = 2$.

Kathrins Antwort

Wenn man zwei aufeinanderfolgende Zahlen hat, dann ist eine davon gerade, und alle geraden Zahlen besitzen den Teiler 2. Wenn man die zwei aufeinanderfolgenden Zahlen multipliziert, muss das Ergebnis also mindestens einmal den Teiler 2 besitzen. Wenn eine Zahl den Teiler 2 besitzt, dann ist sie ein Vielfaches von 2.

Nesims Antwort

Eine gerade Zahl kann als $2a$ (a ist eine ganze Zahl) dargestellt werden.
Ihr Vorgänger ist dann $2a - 1$ und ihr Nachfolger $2a + 1$.
$2a - 1$ und $2a + 1$ sind ungerade Zahlen.
Fall 1: Die kleinere Zahl ist gerade. $2a \cdot (2a + 1) = 4a^2 + 2a = 2 \cdot (2a^2 + a)$
Fall 2: Die kleinere Zahl ist ungerade. $(2a - 1) \cdot 2a = 4a^2 - 2a = 2 \cdot (2a^2 - a)$
Also ist das Produkt immer ein Vielfaches von 2.

Entscheide, ob es sich bei den Antworten von Lukas, Miriam, Kathrin und Nesim um gültige mathematische Beweise handelt oder nicht.

Kreuze bitte pro Zeile genau ein Kästchen an.		Begründung: Warum hast du dich für die angekreuzte Antwort entschieden oder nichts angekreuzt?
Lukas Antwort	☐ Gültig ☐ Ungültig	
Miriams Antwort	☐ Gültig ☐ Ungültig	
Kathrins Antwort	☐ Gültig ☐ Ungültig	
Nesims Antwort	☐ Gültig ☐ Ungültig	

Bei der Mathematik-Olympiade kann man für die Aufgabe maximal 2 Punkte erhalten. Wer würde für seine Antwort 2 Punkte bekommen?
(Kreuze ein oder mehrere Kästchen an.)

☐ Lukas ☐ Miriam ☐ Kathrin ☐ Nesim ☐ Niemand

Begründung: _____

_____

Abbildung 178: Aufgabenstellung zum Validieren und Bewerten fiktiver Beweisversuche zu einer wahren Allaussage

Aufgabenstellung zum Validieren und Bewerten fiktiver Beweisversuche zu einer falschen Allaussage

Nun geht es um mathematische Beweise

Aufgabe

Fynn behauptet: *Wenn eine dreistellige Zahl durch 2 teilbar ist, dann ist die Summe ihrer Ziffern immer durch 2 teilbar.* Beweise, dass Fynns Aussage falsch ist.

Riekes Antwort

$418 : 2 = 209$
$4 + 1 + 8 = 13$
13 ist nicht durch 2 teilbar.
Also ist die Aussage falsch.

Saras Antwort

Wenn eine dreistellige Zahl durch 2 teilbar ist, dann ist die Summe ihrer Ziffern ungerade, also nicht durch 2 teilbar.
Deshalb stimmt die Aussage nie.

Toms Antwort

$200 : 2 = 100$	$456 : 2 = 228$	$738 : 2 = 369$
$2 + 0 + 0 = 2$	$4 + 5 + 6 = 15$	$7 + 3 + 8 = 18$
2 ist durch 2 teilbar.	$15 : 2 = 7\ Rest\ 1$	$18 : 2 = 9$

Fynns Aussage ist doch richtig!

Umans Antwort

Wenn eine (dreistellige) Zahl durch 2 teilbar ist, dann ist ihre Einerziffer gerade.
Wenn die Zehner- und die Hunderterziffer beide gerade oder beide ungerade sind, dann ist die Summe aller drei Ziffern gerade, also durch 2 teilbar.
Ansonsten ist die Summe aller drei Ziffern ungerade, also nicht durch 2 teilbar.
Die Summe der Ziffern ist nicht immer durch 2 teilbar. Die Aussage ist also falsch.

Entscheide, ob es sich bei den Antworten von Rieke, Sara, Tom und Uman um gültige mathematische Beweise handelt oder nicht.

Kreuze bitte pro Zeile genau ein Kästchen an.		Begründung: Warum hast du dich für die angekreuzte Antwort entschieden oder nichts angekreuzt?
Riekes Antwort	☐ Gültig ☐ Ungültig	
Saras Antwort	☐ Gültig ☐ Ungültig	
Toms Antwort	☐ Gültig ☐ Ungültig	
Umans Antwort	☐ Gültig ☐ Ungültig	

Bei der Mathematik-Olympiade kann man für die Aufgabe maximal 2 Punkte erhalten. Wer würde für seine Antwort 2 Punkte bekommen?
(Kreuze ein oder mehrere Kästchen an.)

☐ Rieke ☐ Sara ☐ Tom ☐ Uman ☐ Niemand

Begründung: _____

_____

Abbildung 179: Aufgabenstellung zum Validieren und Bewerten fiktiver Beweisversuche zu einer falschen Allaussage

In der zweiten Version der Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche ist die falsche Allaussage „Wenn eine dreistellige Zahl durch zwei teilbar ist, dann ist die Summe ihrer Ziffern immer durch zwei teilbar“ (Abb. 179) vorgegeben. Die vier fiktiven Beweisversuche wurden unter Berücksichtigung von Ergebnissen empirischer Studien zum Widerlegen von Allaussagen entwickelt.[164] Beispielsweise ist anknüpfend an die Untersuchungsergebnisse von Küchemann und Hoyles (2002) in Umans Antwort allgemeingültig dargelegt, wann die Quersumme einer dreistelligen, durch zwei teilbaren Zahl ungerade ist.

In beiden Versionen der Aufgaben zum Validieren und Bewerten fiktiver Beweisversuche wird, ähnlich wie in der Untersuchung von Ufer et al. (2009), dazu aufgefordert, zu jedem fiktiven Beweisversuch anzugeben, ob es sich um einen gültigen mathematischen Beweis handelt und die Angabe zu begründen. Zusätzlich soll angegeben und begründet werden, für welche der fiktiven Beweisversuche bei der Mathematik-Olympiade die volle Punktzahl vergeben würde. Auf diese Weise wird berücksichtigt, dass die Vorstellungen von Schülerinnen und Schüler, was ein mathematischer Beweis ist, von deren Vorstellungen, was bei der Mathematik-Olympiade erwartet wird, abweichen können. Eigenschaften von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade hängen eventuell stärker mit deren Vorstellungen zu Wettbewerbsanforderungen zusammen als mit deren Vorstellungen dazu, was ein gültiger Beweis ist.

Wie in Abschnitt 6.2.1 erläutert, wurde bei der Datenerhebung in jeden Fragebogen genau eine der beiden Versionen aufgenommen, so dass ca. die Hälfte der Teilnehmenden die eine Version und etwa die Hälfte der Teilnehmenden die andere Version erhalten.

Die beschriebene Datenerhebung mit der spezifisch fokussierten Beweisaufgabe und dem Fragebogen als Untersuchungsinstrumenten wurde in 44 der 55 Regionen Nordrhein-Westfalens

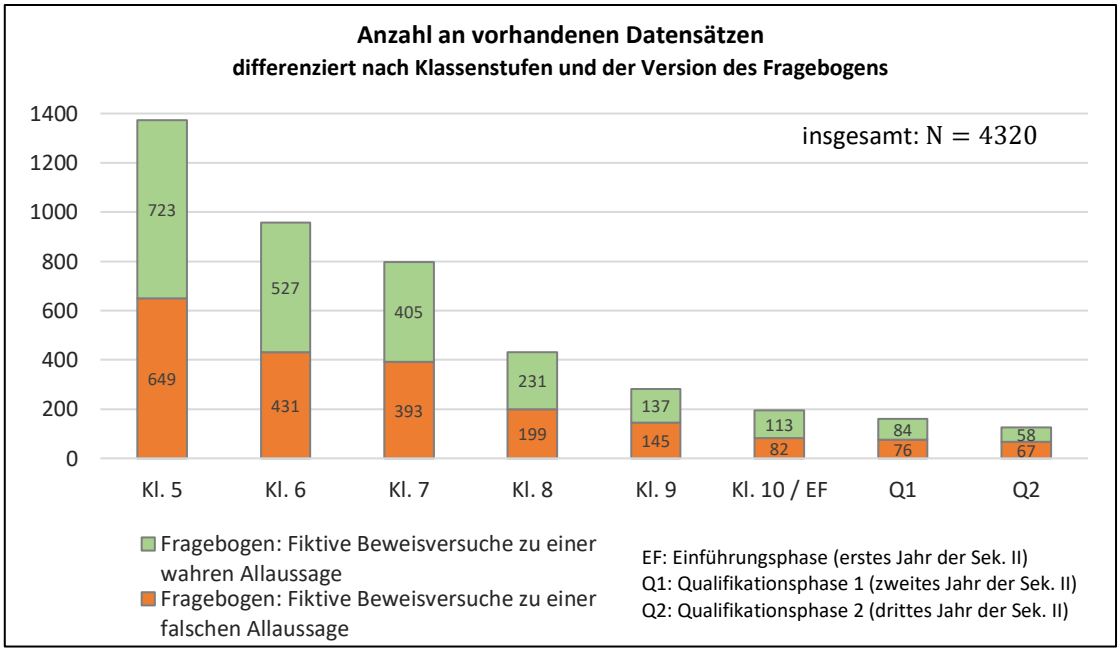


Abbildung 180: Anzahl an vorhandenen Datensätzen differenziert nach Klassenstufen und der Version des Fragebogens

[164] Zu den berücksichtigten Publikationen gehören die von Galbraith (1981), Grundey (2010), Küchemann & Hoyles (2002), Kuntze (2004), Porteous (1990), Stylianides und Al-Murani (2010).

im Rahmen der Regionalrunde der 56. Mathematik-Olympiade durchgeführt (Abschn. 6.2). Von insgesamt 4320 Schülerinnen und Schülern ab der Klassenstufe 5 liegen Beweisversuche und ein bearbeiteter Fragebogen vor (Abb. 180). Dies entspricht rund 75% aller 5775 Schülerinnen und Schüler ab der Klassenstufe 5, die an der Regionalrunde der 56. Mathematik-Olympiade in NRW teilgenommen haben. Die je nach Klassenstufe unterschiedlich hohen Anzahlen an **vorliegenden Datensätzen** sind darauf zurückzuführen, dass regulär an der Regionalrunde der Mathematik-Olympiade in NRW in den verschiedenen Klassenstufen unterschiedlich viele Schülerinnen und Schüler teilnehmen. Je Klassenstufe haben etwa die Hälfte der Teilnehmenden die eine Version der Aufgaben zum Validieren und Bewerten fiktiver Beweisversuche erhalten und die andere Teilnehmenden die andere Version.

Es liegen also von jedem Teilnehmenden in der Regel vier Beweisversuche vor, zwei zu wahren Allaussagen, einer zu einer falschen Allaussage und einer zu einer Existenzaussage. Die Beweisversuche können jeweils hinsichtlich des Spektrums und der Häufigkeitsverteilungen bestimmter Eigenschaften ausgewertet werden. Zudem ist es möglich, Zusammenhänge zwischen verschiedenen Eigenschaften der Beweisversuche zu analysieren und die Beweisversuche etwa mit Blick auf den Nachweis der Gültigkeit zu klassifizieren.

Zusätzlich zu den Beweisversuchen liegt von jedem Teilnehmenden in der Regel eine Bearbeitung einer Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche vor. Diese Bearbeitungen lassen sich sowohl für sich genommen auswerten als auch in Zusammenhang mit Beweisversuchen der Teilnehmenden analysieren.

Bei allen **Analysen** ist es möglich, die Ergebnisse in Abhängigkeit von der Klassenstufe der Teilnehmenden, deren Geschlecht, deren Wettbewerbserfolg, deren Teilnahme an mathematischen Förderangeboten und deren Studienwunsch zu differenzieren. Ebenso kann andersherum untersucht werden, inwieweit bestimmte Eigenschaften der Beweisversuche mit besonderen Ausprägungen dieser personenbezogenen Merkmale einhergehen. Auf diese Weise lässt sich ein differenziertes Bild der Beweisversuche von Teilnehmenden der Mathematik-Olympiade und deren Bearbeitungen der Aufgabe zum Validieren und Bewerten fiktiver Beweisversuche generieren.

Die vorliegende Arbeit, in der Beweisversuche von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 zu einer wahren Allaussage analysiert wurden, beschreibt einen Ausschnitt der möglichen Auswertung der vorgestellten Datenerhebung. Diese Arbeit ist somit im Wesentlichen im zweiten Teil des Forschungsprojekts zu verorten.

10.2.3 Ansatzpunkte für eine Förderung von Teilnehmenden der Mathematik-Olympiade

In einem dritten Schritt können die Ergebnisse der Anforderungsanalyse (Teil 1) und der Analyse vorhandener Beweiskompetenzen (Teil 2) mit Blick auf Ansatzpunkte für eine Förderung im Führen von Beweisen interpretiert werden. Unter Berücksichtigung dieser Ansatzpunkte ein Förderangebot für Teilnehmende der Mathematik-Olympiade auszuarbeiten, durchzuführen, zu evaluieren und zu optimieren ist kein Bestandteil des Forschungsprojekts, sondern wäre ein eigenes komplexes Vorhaben.

10.3 Anforderungen bei der Mathematik-Olympiade

Nachdem in Abschnitt 10.2 das Forschungsprojekt zum Führen von Beweisen bei der Mathematik-Olympiade überblicksartig vorgestellt wurde, gibt dieser Abschnitt konkrete Einblicke in die Analyse objektiver Anforderungsstrukturen von Aufgaben der Mathematik-Olympiade, die im Rahmen dieses Projekts erfolgt ist. Diese Anforderungsanalyse könnte im Anschluss an die vorliegende Arbeit vollständig verschriftlicht werden. Ihre Ergebnisse ermöglichen es, die im Rahmen dieser Arbeit gewonnen Erkenntnisse zu Beweisversuchen von Teilnehmenden der Mathematik-Olympiade vor dem Hintergrund der regulären Wettbewerbsanforderungen einzuordnen und geben Anhaltspunkte für sinnvolle Zielsetzungen im Rahmen einer Förderung von Teilnehmenden der Mathematik-Olympiade im Führen von Beweisen. Im Folgenden werden die Anlage und Datengrundlage der erfolgten Anforderungsanalyse beschrieben, ein Modell zur Analyse objektiver Anforderungsstrukturen zum Beweisen skizziert und Einblicke in Ergebnisse der Anforderungsanalyse gegeben.

10.3.1 Anlage und Datengrundlage der Anforderungsanalyse

In der Anforderungsanalyse werden die objektiven Anforderungsstrukturen von Aufgaben, die ab der Klassenstufe 5 in der Mathematik-Olympiade im Bereich der Teilbarkeit ganzer Zahlen vorkommen, untersucht. Dabei soll der Begriff „objektiv“ in Anlehnung an Bruder (1981) hervorheben, dass die Analyse nicht auf die individuellen Schwierigkeiten von Schülerinnen und Schülern abzielt, sondern auf die schwierigkeitsgenerierenden Merkmale der Aufgaben. Ziel ist es, die Anforderungen, die strukturell in den Aufgaben angelegt sind und in den Lösungen von Seiten der Verantwortlichen implizit angedacht wurden, herauszuarbeiten. Deshalb basiert die Analyse auf den Aufgaben der Mathematik-Olympiade und den dazu vom Aufgabenausschuss herausgegebenen Lösungen.

Untersucht werden die Aufgaben aus dem Themenbereich der Teilbarkeit ganzer Zahlen[165], die vom Aufgabenausschuss für die 46. bis 55. Mathematik-Olympiade herausgegeben wurden, d. h. für die zehn Schuljahre vor der Datenerhebung im Rahmen dieser Arbeit. Um die Auswahl der Aufgaben, die dem Bereich der Teilbarkeit ganzer Zahlen zuzuordnen sind, konsistent und transparent zu gestalten, wird eine qualitative Inhaltsanalyse durchgeführt und mithilfe eines Entscheidungsgraphen das Vorgehen bei der Aufgabenauswahl definiert. Konkret wird die Idee der Entscheidungsbäume mit definierenden Leitfragen von Ott (2016) aufgegriffen. Dabei wird der Codierleitfaden als eine Folge von Entscheidungsfragen dargestellt, deren Beantwortung zu der passenden Codierung führt. Zur Vereinfachung wird im Rahmen dieser Arbeit in Anlehnung an Schorcht (2018) der Begriff „Entscheidungsgraph“ verwendet. Der Entscheidungsgraph, mit dem aus den herausgegebenen Aufgabenteilen alle für die Anforderungsanalyse relevanten Aufgabenteile ausgewählt werden, ist in Abbildung 181 zu sehen. Den Ausgangspunkt bildet stets ein einzelner Aufgabenteil, der die grau unterlegten Entscheidungsfragen dann von oben nach unten durchläuft. Als Aufgabenteil wird dabei jeder Arbeitsauftrag bezeichnet, der

[165] Eine Beschreibung und Eingrenzung des Themenbereichs der Teilbarkeit ganzer Zahlen für den Rahmen dieser Arbeit ist in Abschnitt 1.1 zu finden.

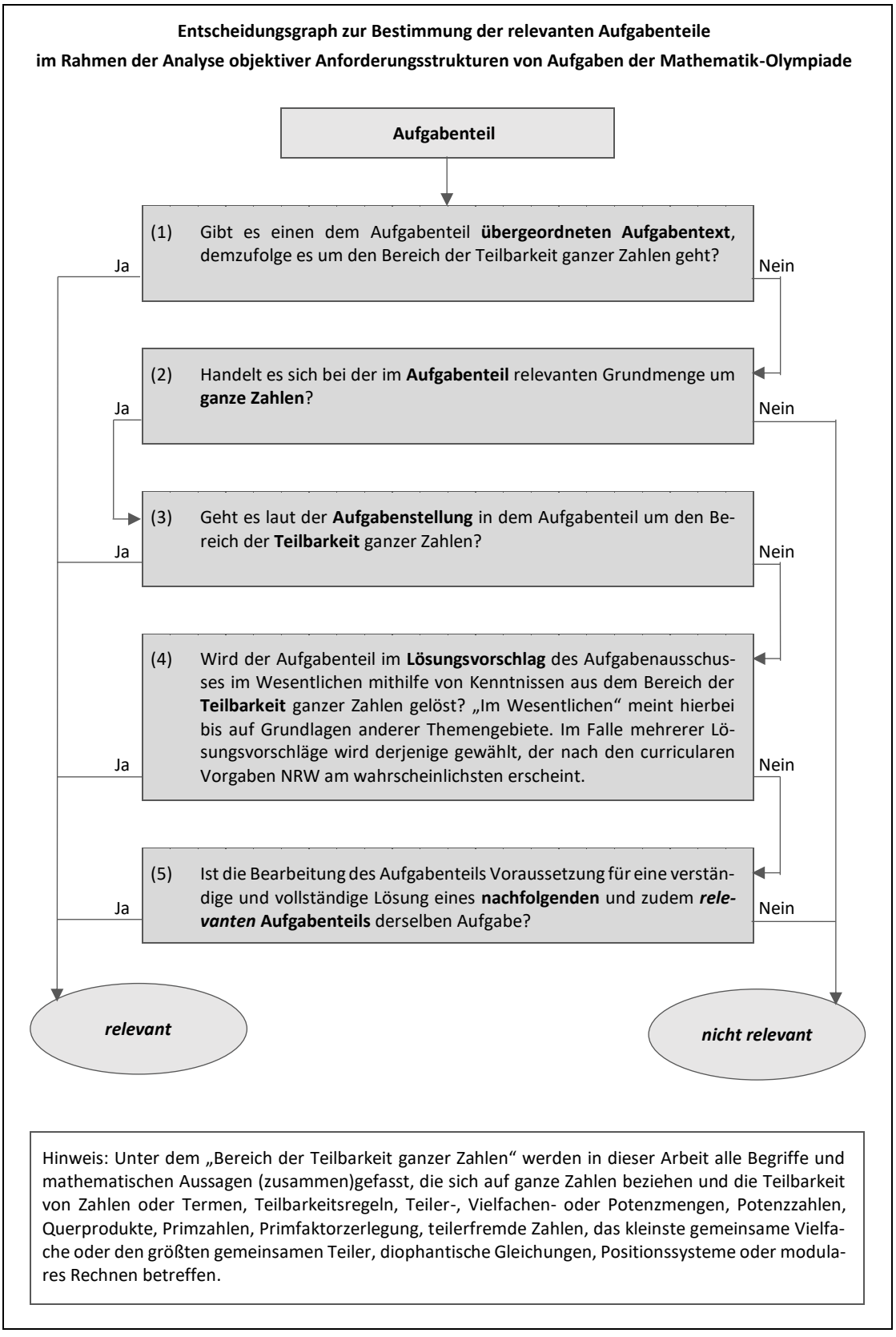


Abbildung 181: Entscheidungsgraph zur Bestimmung der relevanten Aufgabenteile im Rahmen der Analyse objektiver Anforderungsstrukturen von Aufgaben der Mathematik-Olympiade

mithilfe einer Nummerierung a), b), c) usw. separat angegeben ist. Falls eine Aufgabe nicht in mehrere Aufgabenteile gegliedert ist, wird die gesamte Aufgabe als ein Aufgabenteil angesehen. Aufbauend auf dieser Definition bildet bei der Aufgabenauswahl jeder Aufgabenteil, der für Schülerinnen und Schüler ab der Klassenstufe 5 für die 46. bis 55. Mathematik-Olympiade herausgegeben wurde, eine Codiereinheit und die gesamte Aufgabe, der ein Aufgabenteil entstammt, die Kontexteinheit. Mit Fokus auf jeden einzelnen zu analysierenden Aufgabenteil werden die Entscheidungsfragen entsprechend der Pfeilführung so lange von oben nach unten durchlaufen bis der Aufgabenteil einer der Kategorien *relevant* oder *nicht relevant* zuzuordnen ist. Ein Aufgabenteil wird als relevant bezeichnet, wenn er in die Anforderungsanalyse aufgenommen wird. Er ist nicht relevant, wenn er nicht in die Anforderungsanalyse aufgenommen wird. Da bei der Beantwortung der Entscheidungsfragen übergeordnete Aufgabentexte miteinzubeziehen sind, bietet es sich an, alle Aufgabenteile direkt nacheinander zu codieren. Da bei der fünften Entscheidungsfrage zudem die nachfolgenden Aufgabenteile derselben Aufgabe zu berücksichtigen sind, wird mit dem letzten Aufgabenteil einer Aufgabe begonnen. Die Anwendung des Entscheidungsgraphen wird im Folgenden exemplarisch anhand einer Mathematik-Olympiaden-Aufgabe veranschaulicht. Der bei der Untersuchung der Intercoderreliabilität erhaltene Wert für Cohens Kappa $\kappa = 0{,}81$ liegt nach den von Wirtz und Caspar (2002) angelegten Wertegrenzen im sehr guten Bereich (Abschn. 7.2.2 zur Berechnung und Interpretation von Cohens Kappa).

Die Aufgabe 550622 aus Abbildung 182 wurde bei der 55. Mathematik-Olympiade als zweite Aufgabe der Regionalrunde in der Klassenstufe 6 vorgeschlagen und umfasst drei Aufgabenteile. Bei allen Aufgabenteilen ist die erste Frage, ob es einen dem Aufgabenteil übergeordneten Aufgabentext gibt, demzufolge es um den Bereich der Teilbarkeit ganzer Zahlen geht, mit Nein zu beantworten. Aus dem einleitenden Satz „Jens hilft seinem Vater beim Bauen“ lässt sich nicht auf den Bereich der ganzen Zahlen schließen. Es ist also eine genaue Betrachtung jedes einzelnen Aufgabenteils erforderlich. Um bei Bedarf die fünfte Frage im Entscheidungsgraphen, die die nachfolgenden Aufgabenteile miteinbezieht, direkt beantworten zu können, wird rückwärts vorgegangen.

Weiter geht es also mit dem Aufgabenteil c) und der zweiten Frage: Handelt es sich bei der im Aufgabenteil relevanten Grundmenge um ganze Zahlen? Die Lernenden sollen in Teil c) die Menge an Mörtelpulver und an Wasser bestimmen, die für $25\ \mathrm{kg}$ Mörtel benötigt wird. Da Gewichtsangaben nicht ganzzahlig sein müssen, sondern alle (positiven) rationalen Zahlen sein können, wird die Frage mit Nein beantwortet. Daraus folgt der Ausschluss des Aufgabenteils aus der Anforderungsanalyse, der Aufgabenteil c) wird als nicht relevant codiert. In Abbildung 183 ist der zugehörige Graph rot markiert.

Als nächstes liegt der Fokus auf dem Aufgabenteil b). Auch hier ist zu prüfen, ob es sich bei der im Aufgabenteil relevanten Grundmenge um ganze Zahlen handelt. In diesem Aufgabenteil geht es um eine Anzahl an Gehwegplatten, also um natürliche bzw. ganze Zahlen. Die zweite Frage ist somit zu bejahen. Darauf aufbauend wird drittens ermittelt, ob die Aufgabenstellung in den Bereich der Teilbarkeit eingeordnet werden kann. Jedoch wird in der Aufgabenstellung

Klasse 6, Regionalrunde, Aufgabe 550622

Jens hilft seinem Vater beim Bauen.

a) Es sind Bretter ins Haus zu tragen. Jens trägt immer drei Bretter auf einmal. Hätte er vier Bretter auf einmal getragen, hätte er 12-mal weniger laufen müssen. Wie viele Bretter hat Jens ins Haus getragen? Führe eine Probe durch.

b) Beim Mittagessen erzählt Jens seiner Mutter: „Ich wollte alle Gehwegplatten gleichmäßig aufstapeln Als ich jeweils 4 Platten übereinandergelegt habe, blieb eine übrig. Als ich jeweils 6 Platten übereinandergelegt habe, blieb auch eine übrig. Wenn ich jeweils 8 Platten übereinandergelegt hätte, wäre auch wieder eine Platte übrig gewesen."

Die Mutter fragt, wie viele Platten aufzustapeln waren. Jens antwortet: „Mal sehen, ob du das selbst herausbekommst. Es sind übrigens mehr als 50, aber weniger als 100 Platten."
Welche Anzahlen für die Gehwegplatten sind möglich?

c) Der Mörtel wird aus Mörtelpulver und Wasser zusammengemischt. Das Mischungsverhältnis von Pulver zu Wasser beträgt hinsichtlich des Gewichts $2 : 3$. Es sollen 25 kg Mörtel hergestellt werden.

Wie viel Mörtelpulver und wie viel Wasser werden benötigt?

(Mathematik-Olympiaden e. V., 2016, S. 36)

Lösung zu 550622:

a) In den zwölf eingesparten Läufen hätte Jens $(12 \cdot 3 =)\ 36$ Bretter getragen. Diese 36 Bretter hätte er mit den übrigen Läufen transportieren müssen – und zwar je Lauf ein Brett.
Folglich wäre er mit je 4 Brettern genau 36-mal gelaufen und hätte $(36 \cdot 4 =)\ 144$ Bretter getragen.

Probe:
Wenn Jens jeweils 3 Bretter trägt, dann muss er $(144 : 3 =)\ 48$-mal laufen.
Wenn Jens jeweils 4 Bretter trägt, dann muss er $(144 : 4 =)\ 36$-mal laufen.
Wegen $48 - 36 = 12$ hätte er tatsächlich 12-mal weniger laufen müssen.

b) Die stets übrig bleibende Platte bleibt zunächst außer Acht. Dann muss die Anzahl der verbliebenen Platten durch 4, 6 und 8 teilbar sein. Das trifft für genau die Zahlen zu, die Vielfache von 24 sind. Zwischen 50 und 1000 gibt es die Vielfachen $(24 \cdot 3 =)\ 72$ und $(24 \cdot 4 =)\ 96$.
Nun wird die Platte dazugezählt, die stets übrig bleibt.
Folglich sind 73 und 97 als Anzahlen für die gehwegplatten möglich.

[…] (Mathematik-Olympiaden e. V., 2016, S. 110)

Abbildung 182: Aufgabe 550622 der Mathematik-Olympiade und Lösung der Aufgabenteile a, b

ausschließlich das Stapeln der Gehwegplatten beschrieben, Inhalte wie Teilermengen, Teilbarkeitsregeln, Prim- oder Quadratzahlen sind nicht explizit erwähnt. Aus diesem Grund ist die dritte Frage zu verneinen und somit die Lösung dieses Aufgabenteils b) auf den Bereich der Teilbarkeit hin zu untersuchen. Entscheidend für die Beantwortung der Frage ist die vom Aufgabenausschuss herausgegebene Lösung. Diese kann Abbildung 182 entnommen werden. Die Anzahl an Gehwegplatten wird in der Lösung mithilfe des kleinsten gemeinsamen Vielfachen und der Bestimmung von Vielfachen des kleinsten gemeinsamen Vielfachen hergeleitet. Abgesehen von der Durchführung elementarer Rechenoperationen werden keine Kenntnisse aus anderen mathematischen Themengebieten angewendet. Die vierte Frage, ob der Aufgabenteil im Lösungsvorschlag des Aufgabenausschusses im Wesentlichen mithilfe von Kenntnissen aus dem Bereich der Teilbarkeit ganzer Zahlen gelöst wird, ist deshalb mit Ja zu beantworten. Daraus folgt, dass der Aufgabenteil b) in die Anforderungsanalyse aufgenommen wird. Der zugehörige Entscheidungsgraph ist in Abbildung 183 in grün gekennzeichnet.

Beim Aufgabenteil a) geht es um eine Anzahl an Brettern, also um ganze Zahlen. Allerdings wird im Aufgabentext keine Facette des Bereichs der Teilbarkeit angesprochen, weshalb die dritte Frage verneint und auch hier die Lösung hinzugezogen werden muss. Der Aufgabenteil wird durch inhaltliche Überlegungen und die Durchführung elementarer Rechenoperationen gelöst. Der Aussage, dass der Aufgabenteil im Wesentlichen mithilfe von Kenntnissen aus dem Bereich der Teilbarkeit ganzer Zahlen gelöst wird, kann nicht zugestimmt werden. Folglich ist die fünfte Frage zu beantworten: Ist die Bearbeitung des Aufgabenteils Voraussetzung für die verständige und vollständige Lösung eines nachfolgenden und zudem relevanten Aufgabenteils? Der einzige relevante nachfolgende Aufgabenteil ist Teil b). Bei der Bearbeitung des Aufgabenteils b) sind die Überlegungen aus Teil a) nicht notwendig. Folglich wird die fünfte Frage verneint und der Aufgabenteil a) nicht berücksichtigt. Der entsprechende Entscheidungsgraph ist in Abbildung 183 in orange dargestellt.

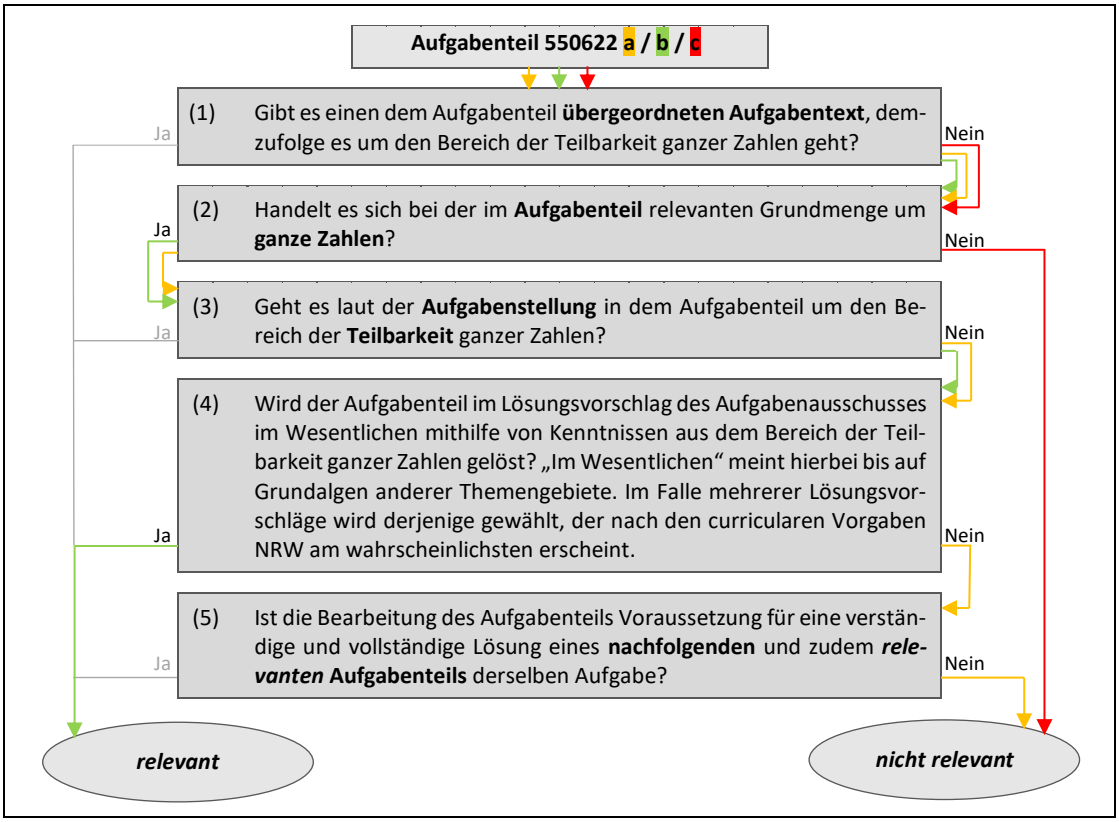


Abbildung 183: Exemplarische Bestimmung der Relevanz der Aufgabenteile 550622 a, b, c

Mit Blick auf die in dieser Arbeit betrachteten Klassenstufen 5 bis 8 wurden 1377 Aufgabenteile für die 46. bis 55. Mathematik-Olympiade vom Aufgabenausschuss herausgegeben. Von diesen 1377 Aufgabenteilen wurden 270 als relevant codiert. Dabei zeigen die Ergebnisse über die Klassenstufen hinweg ein weitgehend homogenes Bild. Beispielsweise sind in jeder der Klassenstufen 17% bis 22% der analysierten Aufgabenteile in die Anforderungsanalyse aufzunehmen. Bezogen auf die Aufgaben, von denen mindestens ein Aufgabenteil relevant ist, sind die Anteile mit 19% bis 23% ebenfalls ähnlich. Die Ergebnisse der Aufgabenauswahl bestätigen also nebenbei zumindest für die Klassenstufen 5 bis 8, dass Aufgaben aus dem Themenbereich der Teilbarkeit ganzer Zahlen in allen Klassenstufen vorkommen, was einen Grund

für die Auswahl dieses Themenbereichs darstellte (Abschn. 1.1, 6.1, 9.3.4).

10.3.2 Ein Modell zur Analyse objektiver Anforderungsstrukturen von Aufgaben mit besonderem Fokus auf dem Führen von Beweisen

Um bei allen relevanten Aufgabenteilen die objektiven Anforderungsstrukturen in Bezug auf das Führen von Beweisen zu untersuchen, wurde deduktiv-induktiv ein Modell entwickelt und eine qualitativ-strukturierende Inhaltsanalyse nach Schreier (2012, 2014) durchgeführt. Die Grundlagen für die Entwicklung des Modells bildeten die Forschungsfragen, die Verfahren zur Bestimmung und zum Vergleich objektiver Anforderungsstrukturen von Bruder (1981, 1985), weitere mathematikdidaktische Publikationen und einige der zu analysierenden Aufgabenteile.[166] Die Abbildung 184 gibt einen Überblick über das entwickelte **Analysemodell**. Dieses wird im Folgenden erklärt und zur Veranschaulichung an einzelnen Stellen en détail ausgeführt.

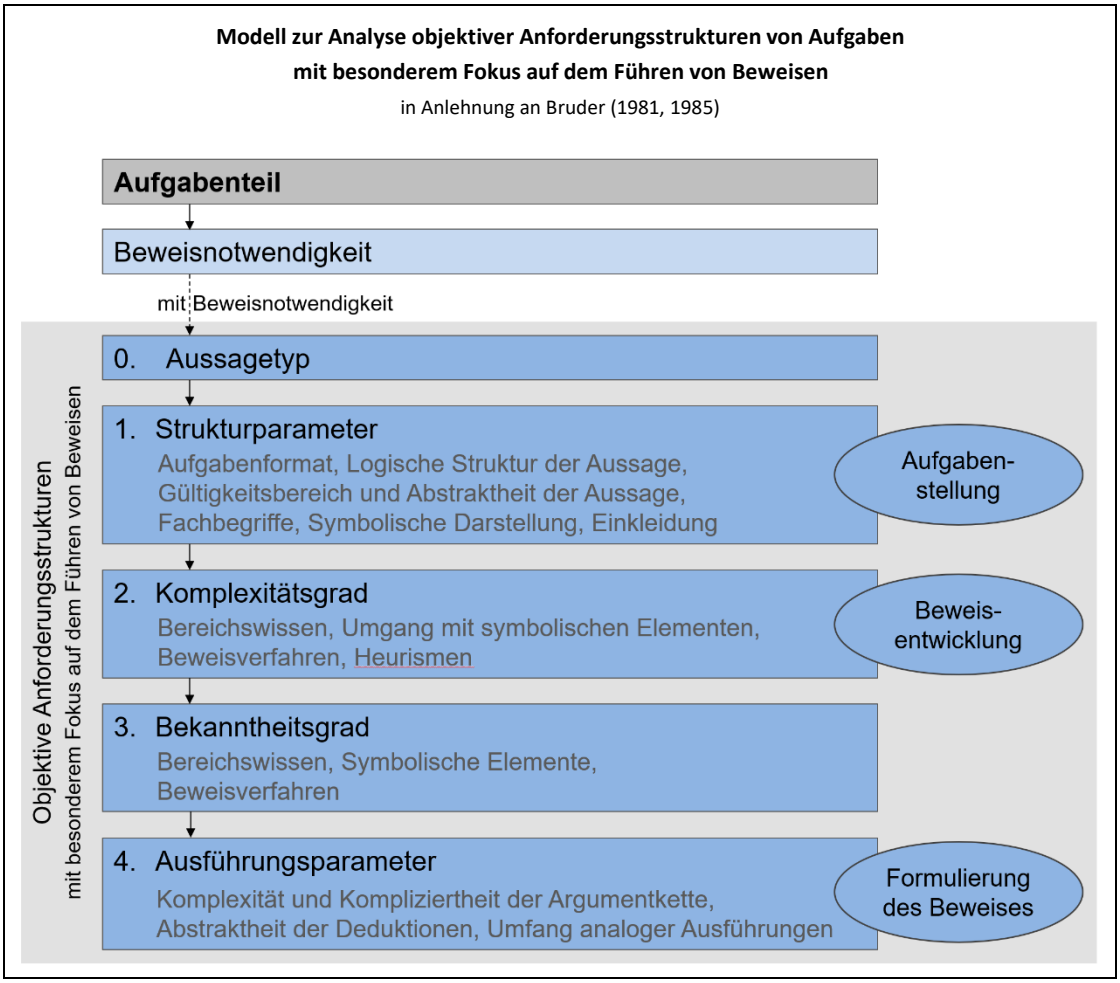


Abbildung 184: Modell zur Analyse objektiver Anforderungsstrukturen von Aufgaben mit besonderem Fokus auf dem Führen von Beweisen

[166] Zu der verwendeten Literatur gehören u. a. die Publikationen von Bruder und Collet (2011), Cohors-Fresenborg et al. (2004), Engel (1998), Epp (2003), Fischer und Malle (1985), Grieser (2013), Maier (1999), Pólya (1949, 1966), Prediger (2013), Reiss et al. (2002), Reiss und Ufer (2009), Sewerin (1979), Schoenfeld (1985), Stein (1984) und Ufer et al. (2008) sowie die Bildungsstandards und Lehrpläne für NRW (KMK, 2003, 2004, 2012; MSW NRW, 2007, 2008, 2014). Es wurden die relevanten Aufgabenteile der 47., der 49. und der 52. Mathematik-Olympiade in die Entwicklung des Modells miteinbezogen.

Oben im Analysemodell ist eine Unterscheidung zwischen Aufgabenteilen mit und solchen ohne Beweisnotwendigkeit zu sehen. Da es um die Anforderungen im Bereich des Führens von Beweisen geht, aber bei der Mathematik-Olympiade nicht in allen Aufgabenteilen eine Aussage bewiesen werden muss, wird zuerst grundlegend zwischen **Aufgabenteilen mit** und **Aufgabenteilen ohne Beweisnotwendigkeit** unterschieden. Diese sind im Rahmen des Forschungsprojekts wie folgt definiert:

- Aufgabenteil mit Beweisnotwendigkeit
  Die korrekte und vollständige Bearbeitung des Aufgabenteils erfordert den Beweis einer mathematischen Aussage. Dabei sind die erforderlichen Transformationen nicht angegeben. Es ist notwendig, selbstständig eine Beweisidee zu generieren. Die zu zeigende Aussage kann in der Aufgabenstellung vorgegeben sein, muss es aber nicht.
- Aufgabenteil ohne Beweisnotwendigkeit
  Alle Aufgabenteile, auf die die vorhergehende Definition nicht zutrifft, sind ohne Beweisnotwendigkeit. Zu Aufgabenteilen ohne Beweisnotwendigkeit gehören solche, in denen ausschließlich eine Behauptung aufgestellt, eine vorgegebene Aussage geprüft, eine Vorschrift auf Beispiele angewendet oder eine Lösung angegeben werden muss.

Die Abbildung 185 zeigt exemplarisch mehrere Aufgabenteile der Mathematik-Olympiade. In den Aufgabenteilen 520724a-c geht es um eine Teilbarkeitsregel für die Zahl 7. Die in der Aufgabenstellung postulierte Regel soll im Aufgabenteil a) auf die Zahl 364 angewendet und im Aufgabenteil b) am Beispiel der Zahl $3\,645\,068$ überprüft werden. Diese beiden Aufgabenteile a), b) sind ohne Beweisnotwendigkeit. Im dritten Aufgabenteil c) ist die Teilbarkeitsregel zu beweisen, es handelt sich also um einen Aufgabenteil mit Beweisnotwendigkeit. Im Aufgabenteil 470613 ist die Aussage, dass es genau 60 fünfstellige Zahlen gibt, die beide angegebenen Bedingungen erfüllen, zu finden und zu beweisen. Dieser Aufgabenteil ist ebenfalls mit Beweisnotwendigkeit.

Klassenstufe 7, Regionalrunde, Aufgabe 520724

Eine positive ganze Zahl soll auf Teilbarkeit durch 7 geprüft werden. Betrachtet wird die folgende Regel.

Die Einerziffer der Zahl wird gestrichen und dann wird das Doppelte der Einerziffer abgezogen. Wenn die so erhaltene Zahl durch 7 *teilbar ist, dann ist auch die Ausgangszahl durch* 7 *teilbar.*

Beispiel: Aus der zu prüfenden Zahl 539 *entsteht mit der Zwischenrechnung* $539 - 2 \cdot 9 = 53 - 18$ *die Zahl* 35*. Da* 35 *durch* 7 *teilbar ist, ist nach der Regel auch* 539 *durch* 7 *teilbar.*

a) Zeige mit Hilfe dieser Regel, dass die Zahl 364 durch 7 teilbar ist.

b) Zeige durch viermalige Anwendung dieser Regel, das $3\,645\,068$ durch 7 teilbar ist. Notiere hierzu die Zahlen, deren Teilbarkeit durch 7 zu prüfen ist. Berechne dann auch den Wert des Quotienten $3\,645\,068 : 7$ durch schriftliche Division.

c) Beweise die Regel. (Mathematik-Olympiaden e. V., 2013, S. 38, H. i. O.)

Klassenstufe 6, Schulrunde, Aufgabe 470613

Wie viele fünfstellige Zahlen gibt es, die die beiden folgenden Bedingungen erfüllen?

(1) Die gesuchten Zahlen sind durch 12 teilbar.

(2) Jede der gesuchten Zahlen wird aus fünf aufeinander folgenden Ziffern der Folge $1, 2, 3, 4, 5, 6, 7, 8, 9$ gebildet, wobei die Ziffern aber in einer beliebigen Reihenfolge verwendet werden dürfen.

(Mathematik-Olympiaden e. V., 2008, S. 27)

Abbildung 185: Aufgaben 520724, 470613 der Mathematik-Olympiade

Die beschriebene Unterscheidung zwischen Aufgabenteilen mit Beweisnotwendigkeit und Aufgabenteilen ohne Beweisnotwendigkeit ermöglicht eine erste Beurteilung der Relevanz des Führens von Beweisen im Wettbewerb. Zudem dient diese Unterscheidung der Auswahl der Aufgabenteile, bei denen eine detaillierte Analyse objektiver Anforderungsstrukturen in Bezug auf das Führen von Beweisen sinnvoll ist.

Alle **Aufgabenteile mit Beweisnotwendigkeit** werden detailliert hinsichtlich ihrer objektiven Anforderungsstrukturen im Bereich des Führens von Beweisen analysiert. Um diese Anforderungsstrukturen möglichst genau zu erfassen, wurden die Verfahren zur Bestimmung und zum Vergleich objektiver Anforderungsstrukturen von Regina Bruder (1981, 1985) aufgegriffen und weiterentwickelt. Dabei wurden schwierigkeitsgenerierende Merkmale von Aufgaben herausgearbeitet, isoliert und strukturiert. Zudem wurden zu jedem schwierigkeitsgenerierenden Merkmal in einem analytischen Vorgehen drei Anforderungsniveaus ausgemacht. Die Einteilung der Niveaus erfolgte theoriebasiert, sie ist begründet, aber nicht empirisch überprüft.

Bei den Aufgabenteilen mit Beweisnotwendigkeit hängt es grundlegend vom **Aussagetyp** ab, wie die zu zeigende Aussage zu beweisen ist und welche Vorstellungen zum Beweisen wichtig sind (u. a. Epp, 2003). Deshalb wird als erstes schwierigkeitsgenerierendes Merkmal der Aussagetyp erfasst (Abb. 184).

Danach werden unter dem Oberbegriff „**Strukturparameter**“ schwierigkeitsgenerierende Merkmale untersucht, die durch die logische Struktur und Abstraktheit der zu zeigenden Aus-

Modell zur Analyse objektiver Anforderungsstrukturen von Aufgaben mit besonderem Fokus auf dem Führen von Beweisen **Definitionen der Anforderungsniveaus zur Beschreibung des Gültigkeitsbereichs und der Abstraktheit der zu zeigenden Aussage bei Aufgabenteilen mit Beweisnotwendigkeit**
Niveau GB1: Endlicher Gültigkeitsbereich, im Wesentlichen natürliche Zahlen bis 100 Die zu zeigende Aussage hat einen endgültigen Gültigkeitsbereich und bezieht sich im Wesentlichen auf natürlichen Zahlen bis 100. Im „Wesentlichen“ meint dabei, dass der Gültigkeitsbereich nicht deutlich über den Zahlenraum bis 100 hinausgeht. Der Gültigkeitsbereich ist so klein, dass es prinzipiell möglich erscheint, alle relevanten Elemente der Grundmenge einzeln zu betrachten, sei es, um die zu zeigende Aussage zu veranschaulichen, um sie zu beweisen oder um eine Beweisidee zu generieren. Klassenstufe 6, Landesrunde, Aufgabenteil 520631c „Begründe durch allgemeine Feststellungen, dass alle durch 10 teilbaren zweistelligen natürlichen Zahlen durch ihre Quersumme teilbar sind.“ (Mathematik-Olympiaden e. V., 2013, S. 35)
Niveau GB2: Endlicher Gültigkeitsbereich, keine Fokussierung auf natürliche Zahlen bis 100 Die zu zeigende Aussage hat einen endlichen Gültigkeitsbereich, der deutlich über den Zahlenraum bis 100 hinausgeht. Eine Betrachtung aller Einzelfälle erscheint ohne technische Unterstützung kaum umsetzbar. Klassenstufe 8, Landesrunde, Aufgabenteil 510835 „Ermittle alle geordneten Tripel (p_1, p_2, p_3)“ von Primzahlen p_1, p_2 und p_3, welche die Gleichung $p_1 \cdot (p_2 + p_3) = 1683$ erfüllen und für die $p_2 < p_3$ gilt.“ (Mathematik-Olympiaden e. V., 2012a, S. 56)
Niveau GB3: Unendlicher Gültigkeitsbereich Die zu zeigende Aussage hat einen unendlichen Gültigkeitsbereich. Klassenstufe 8, Landesrunde, Aufgabenteil 520833c „Betrachtet wird folgende Regel: Es seien p und q ganze Zahlen. Dann ist $1000\ p + q$ genau dann durch 7 teilbar, wenn $p - q$ durch 7 teilbar ist. [...] Beweise diese Regel.“ (Mathematik-Olympiaden e. V., 2012a, S. 56)

Tabelle 19: Modell zur Analyse objektiver Anforderungsstrukturen von Aufgaben mit besonderem Fokus auf dem Führen von Beweisen: Definitionen der Anforderungsniveaus zur Beschreibung des Gültigkeitsbereichs und der Abstraktheit der Aussage bei Aufgabenteilen mit Beweisnotwendigkeit

sage oder durch die Formulierung der Aufgabe bedingt sind. Eines dieser schwierigkeitsgenerierenden Merkmale ist der Gültigkeitsbereich und die Abstraktheit der zu zeigenden Aussage. Dazu sind die unterschiedenen drei Anforderungsniveaus in Tabelle 19 definiert. Das grundlegende Anforderungsniveau 1 lautet: *Endlicher Gültigkeitsbereich mit Fokus auf natürliche Zahlen bis Hundert*. Beispielsweise bezieht sich der Aufgabenteil 520631c aus der Klassenstufe 6 ausschließlich auf zweistellige natürliche Zahlen und ist deshalb auf dem Niveau 1 zu verorten. Bezüglich des Gültigkeitsbereichs und der Abstraktheit der Aussage anspruchsvoller sind Aufgabenteile, die sich zwar auf einen endlichen Gültigkeitsbereich beziehen, der jedoch nicht auf natürliche Zahlen bis Hundert fokussiert. Das ist z. B. beim Aufgabenteil 510835 für die Klassenstufe 8 der Fall. In diesem Aufgabenteil geht es um Primzahlen, die eventuell größer als 100, aber sicher kleiner als 1683 sind. Auf dem dritten, höchsten Anforderungsniveau werden Aufgabenteile verortet, wenn sich die zu zeigende Aussage auf einen unendlichen Gültigkeitsbereich bezieht. Beispielsweise gehört der Aufgabenteil 530833c, in dem eine Aussage für beliebige ganze Zahlen p, q zu zeigen ist, in diese Kategorie.

Während der Strukturparameter auf objektive Anforderungsstrukturen fokussiert, die im Wesentlichen direkt aus der Aufgabenstellung resultieren, hängen die Ausprägungen der schwierigkeitsgenerierenden Merkmale, die anschließend unter den Oberbegriffen „**Komplexitätsgrad**“, „**Bekanntheitsgrad**“ und „**Ausführungsparameter**“ erfasst werden, zum Teil vom Lösungsweg ab. Dementsprechend werden bei der Analyse der Aufgabenteile mit Beweisnotwendigkeit bezüglich des Komplexitätsgrads, des Bekanntheitsgrads und des Ausführungsparameters die vom Aufgabenausschuss herausgegebenen Lösungen miteinbezogen. Unter dem Oberbegriff „Komplexitätsgrad“ werden die Art und der Umfang des Bereichswissens, der symbolischen Elemente, der Beweisverfahren und der Heurismen, die für die vorgeschlagene Lösung erforderlich sind, beschrieben. In Ergänzung dazu wird mit dem Bekanntheitsgrad berücksichtigt, dass die Schwierigkeit eines Aufgabenteils auch davon abhängt, inwieweit z. B. das erforderliche Bereichswissen und die symbolischen Elemente den Schülerinnen und Schülern, die diesen Aufgabenteil bearbeiten sollen, bekannt sind. Die Aufgabenteile werden diesbezüglich auf Grundlage der Vorgaben in den Bildungsstandards und den Lehrplänen für Nordrhein-Westfalen analysiert. Unter dem Oberbegriff „Ausführungsparameter“ sind schwierigkeitsgenerierende Merkmale, die sich auf das Formulieren des Beweises beziehen, subsummiert. Konkret werden die Komplexität, der Umfang und die Abstraktheit des zu formulierenden Argumentgefüges analysiert.

10.3.3 Ergebnisse der Anforderungsanalyse

Jeder der relevanten Aufgabenteile aus dem Bereich der Teilbarkeit ganzer Zahlen wird entsprechend dem Analysemodell sukzessive analysiert. Im Folgenden werden exemplarisch mit Blick auf die Klassenstufen 5 bis 8, die in dieser Arbeit im Fokus stehen, Einblicke in die Ergebnisse dieser Analyse gegeben. Die exemplarisch angeführten Ergebnisse veranschaulichen, wie sich das Analysemodell zum Beschreiben von Anforderungsstrukturen in einzelnen Aufgabenteilen und zum Vergleich der Anforderungsstrukturen verschiedener Aufgabenteile eignet. Zudem geben sie konkrete Einblicke, wie die Anforderungsstrukturen zum Beweisen bei der Mathematik-Olympiade ausgeprägt sind.

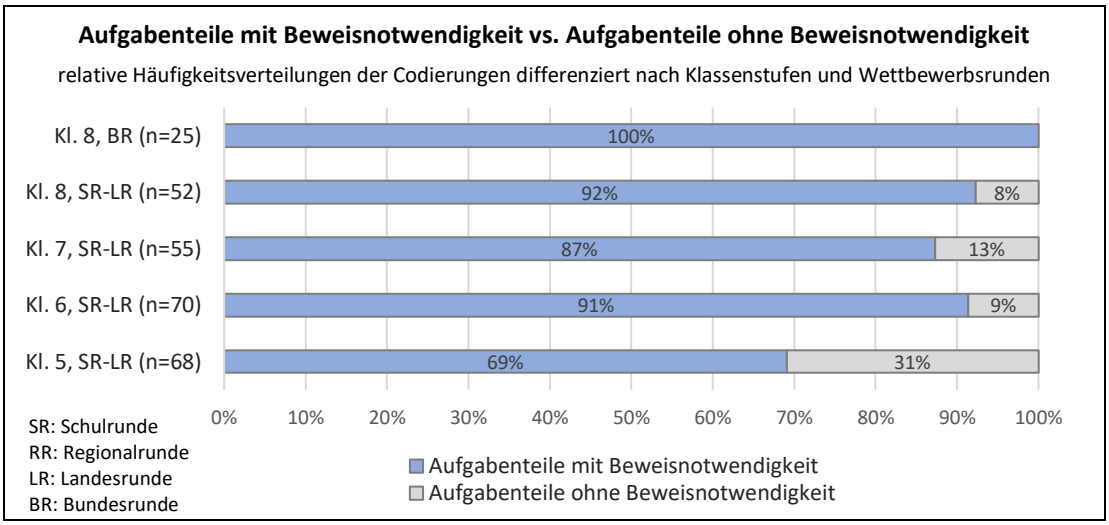
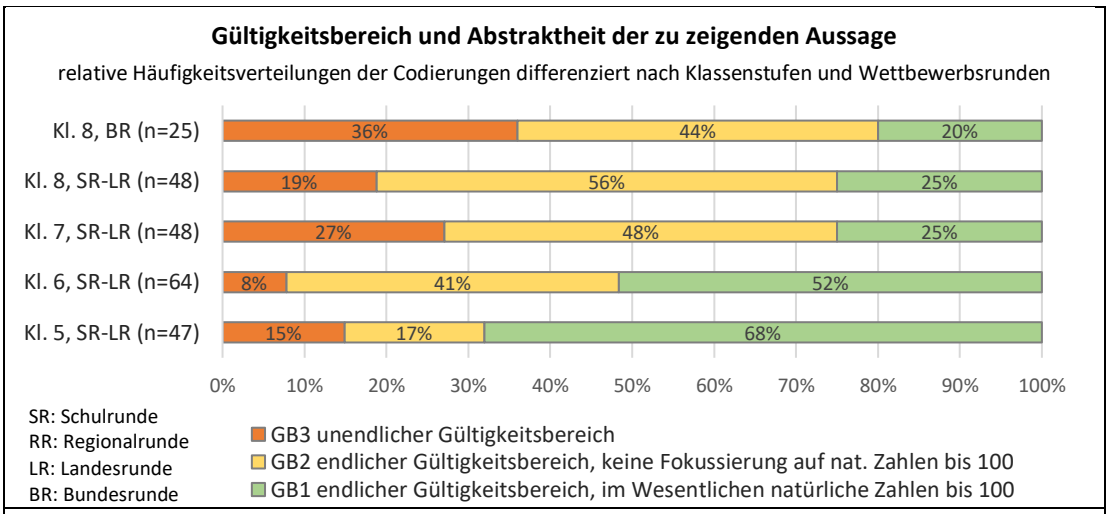


Abbildung 186: Analyse von Aufgabenteilen der Mathematik-Olympiade: Aufgabenteile mit Beweisnotwendigkeit vs. Aufgabenteile ohne Beweisnotwendigkeit. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen und Wettbewerbsrunden

Die Abbildung 186 zeigt Ergebnisse der Unterscheidung zwischen Aufgabenteilen mit und solchen ohne Beweisnotwendigkeit. Bezüglich der grundlegenden Frage, inwieweit in der Mathematik-Olympiade im Bereich der Teilbarkeit ganzer Zahlen Aufgabenteile mit Beweisnotwendigkeit gestellt werden, sind folgende Tendenzen zu erkennen: In der Klassenstufe 5 handelt es sich bei rund 70% der analysierten Aufgabenteile um Aufgabenteile mit Beweisnotwendigkeit. Dabei ist der Anteil der Aufgabenteile mit Beweisnotwendigkeit in der Schulrunde mit etwa 50% statistisch hoch signifikant kleiner als in den fortgeschritteneren Wettbewerbsrunden mit jeweils etwa 90% ($p < .001^{N-1}$, $V = .463$). In den Klassenstufen 6 bis 8 liegt der Anteil der Aufgabenteile mit Beweisnotwendigkeit in der Schul- bis Landesrunde durchgehend bei etwa 90%. Mit Blick auf die Bundesrunde in der Klassenstufe 8 beträgt der Anteil 100%.



p-Werte von Signifikanztests und dazugehörige Effektstärken

	Alle einzelnen Kategorien GB1-3	GB1 vs. GB2-3
Kl. 7 vs. Kl. 5	$p < .001^{U}$, $r = .377$	$p < .001^{N-1}$, $V = .432$
Kl. 7 vs. Kl. 6	$p = .001^{U}$, $r = .314$	$p = .005^{N-1}$, $V = .268$
Kl. 8 vs. Kl. 5	$p = .001^{U}$, $r = .355$	$p < .001^{N-1}$, $V = .432$
Kl. 8 vs. Kl. 6	$p = .003^{U}$, $r = .279$	$p = .005^{N-1}$, $V = .268$

Abbildung 187: Analyse von Aufgabenteilen der Mathematik-Olympiade. Fokus: Gültigkeitsbereich und Abstraktheit der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen und Wettbewerbsrunden

Weiterhin zeigt die Abbildung 187 Ergebnisse der Analyse des **Gültigkeitsbereichs und der Abstraktheit der zu zeigenden Aussage** bei den Aufgabenteilen mit Beweisnotwendigkeit. Demnach bezieht sich in den **Klassenstufen 5 und 6** jeweils die Mehrheit dieser Aufgabenteile auf den Zahlenraum bis Hundert. In den meisten anderen Aufgabenteilen geht es um Aussagen mit endlichem Gültigkeitsbereich, aber ohne Fokussierung auf natürliche Zahlen bis Hundert. Zugleich kommen in beiden Klassenstufen 5 und 6, und insbesondere auch in allen Wettbewerbsrunden in den Klassenstufen 5 und 6, Aussagen mit unendlichem Gültigkeitsbereich vor. In keinem der analysierten Aufgabenteile, in denen eine Aussage mit unendlichem Gültigkeitsbereich zu zeigen ist, handelt es sich um einen Aufgabenteil a) oder um den einzigen Aufgabenteil, aus dem eine Aufgabe besteht. Mit Blick auf den Gültigkeitsbereich und die Abstraktheit der zu zeigenden Aussage und die Klassenstufen 5 und 6 lässt sich somit festhalten, dass der in dieser Arbeit fokussierte Aufgabenteil d), in dem die Teilbarkeit durch vier der Summe von vier aufeinanderfolgenden ungeraden Zahlen zu beweisen ist, durchaus zu den regulären Wettbewerbsanforderungen passt und zugleich vergleichsweise anspruchsvoll ist. Vor dem Hintergrund der regulären Wettbewerbsanforderungen erscheint es als sehr sinnvoll, dass dieser Aufgabenteil nicht als Teil a) der spezifisch fokussierten Aufgabe gestellt wurde (Abschn. 6.2).

In den **Schul- bis Landesrunden in den Klassenstufen 7 und 8** sind die Aufgabenteile hinsichtlich des Gültigkeitsbereichs und der Abstraktheit der zu zeigenden Aussage tendenziell anspruchsvoller (Abb. 187). Insbesondere beträgt der Anteil der Aufgabenteile, die sich im Wesentlichen auf natürliche Zahlen bis Hundert beziehen, in den Klassenstufen 7, 8 rund 25% im Vergleich zu über 50% in den Klassenstufen 5, 6 und ist statistisch hoch signifikant geringer. In den Klassenstufen 7 und 8 sind unter den analysierten Aufgabenteilen mit Beweisnotwendigkeit solche mit einem endlichen Gültigkeitsbereich, aber keiner Fokussierung auf den Zahlenraum bis Hundert besonders häufig. Die Aufgabenteile, in denen eine Aussage mit unendlichem Gültigkeitsbereich zu beweisen ist, machen mit Blick auf die Schul- bis Landesrunden etwa 20% aus. Sie kommen auch als Aufgabenteil a) sowie als einziger Aufgabenteil bzw. als vollständige Aufgabe vor. Der im Rahmen dieser Arbeit in einer Regionalrunde gestellte Aufgabenteil d) ist mit Blick auf die Klassenstufen 7, 8 und den Gültigkeitsbereich und die Abstraktheit dieser Aussage durchaus als gewöhnlich einzuordnen.

10.4 Beweisversuche unterschiedlicher Teilgruppen der Teilnehmenden der Mathematik-Olympiade

Bei der Diskussion der Ergebnisse der Analyse der Beweisversuche aus den Klassenstufen 5 bis 8 ist mehrmals die Frage nach Eigenschaften von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade **ab der Klassenstufe 9** aufgekommen, u. a. weil diese mit elementarer Algebra besser vertraut sind als Lernende am Anfang der Klassenstufen 5 bis 8 (Abschn. 9.1.1, 9.3.1, 9.3.3). In der Datenerhebung bei der Regionalrunde der Mathematik-Olympiade 2016/17 wurden auch die Teilnehmenden ab der Klassenstufe 9 bis zum Abitur aufgefordert zu beweisen, dass die Summe von vier aufeinanderfolgenden ungeraden Zahlen durch acht teilbar ist (Abschn. 10.2.2). Als erste Einblicke in die Beweisversuche dieser Teilnehmenden werden in Abschnitt 10.4.1 Beweisversuche aus der Klassenstufe 9 beschrieben.

Zusätzlich ist in der Diskussion der Ergebnisse dieser Arbeit die Frage aufgekommen, inwieweit besondere Ausprägungen von Beweisversuchen aus den Klassenstufen 5 bis 8, wie z. B. die Existenz einer Darstellungsform mit symbolischem Anteil und die Existenz eines korrekten und vollständigen Argumentgefüges, mit einer spezifischen mathematischen **Förderung** oder einem besonderen **Wettbewerbserfolg** einhergehen. Außerdem stellt sich mit Blick auf Ansatzpunkte für eine Förderung die Frage, inwieweit bei unterschiedlichen Teilgruppen der Teilnehmenden der Mathematik-Olympiade, wie etwa denjenigen, die sich für die Landesrunde qualifiziert haben, und denen, die sich nicht für die Landesrunde qualifiziert haben, bestimmte Ausprägungen der Beweisversuche besonders häufig sind oder etwa nicht vorkommen (Abschn. 9.3, 10.1.4). In Abschnitt 10.4.2 werden diesbezüglich Einblicke gegeben.

10.4.1 Beweisversuche von Teilnehmenden in der Klassenstufe 9

Bei der Regionalrunde der 56. Mathematik-Olympiade in Nordrhein-Westfalen wurden Teilnehmende in der Klassenstufe 9, ähnlich wie diejenigen in den Klassenstufen 7 und 8, aufgefordert, die Teilbarkeit der Summe von vier aufeinanderfolgenden ungeraden Zahlen durch acht zu beweisen. Von insgesamt 282 Schülerinnen und Schülern aus der Klassenstufe 9 liegen Beweisversuche vor (Abschn. 10.2.2). Um Einblicke in deren Beweisversuche zu geben, wird eine Zufallsstichprobe von $\mathrm{N} = 72$ Beweisversuchen mit dem entwickelten Kategoriensystem analysiert.[167] Im Folgenden werden die wesentlichen Ergebnisse dieser Analyse dargestellt. Dabei werden die Beweisversuche nacheinander mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage, mit Blick auf Charakteristika der Beweisversuche und mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis beschrieben (Abschn. 10.4.1.1-10.4.1.3). Die Ergebnisse werden anschließend resümiert und diskutiert (Abschn. 10.4.1.4).

10.4.1.1 Klassenstufe 9: Nachweis der Gültigkeit

Mit Blick auf den Nachweis der Gültigkeit ist primär festzuhalten, dass in der Klassenstufe 9 rund 14% der Beweisversuche als Beweis der zu zeigenden Aussage bezeichnet werden können. Insgesamt 33% der Beweisversuche enthalten ein weitgehend korrektes und vollständiges, deduktives Argumentgefüge, über 50% enthalten ein weitgehend korrektes und vollständiges Argumentgefüge bei zwischen induktiver und deduktiver oder deduktiver Schlussweise bzw. bei begründendem Ansatz. Über 90% der Beweisversuche aus der Klassenstufe 9 sind Beweisversuche mit begründendem Ansatz. Alle genannten Anteile sind statistisch signifikant größer als in jeder der Klassenstufen 5 bis 8 (Tab. 20).

Der Gesamtanteil der Beweisversuche mit einem weiterentwickelbaren Argumentgefüge beträgt in der Klassenstufe 9 rund 75% und ist ebenfalls statistisch hoch signifikant größer als in jeder der Klassenstufen 5 bis 8. Beweisversuche, in denen weder ein Beispiel zu der zu zeigenden Aussage noch ein weiterentwickelbares Argumentgefüge vorkommt, machen in der Klassenstufe 9 rund 13% aus, während dieser Anteil in jeder der Klassenstufen 5 bis 8 bei mindestens 25% liegt und teils statistisch signifikant größer ist (Tab. 20; Abschn. 8.5.1).

[167] Analog zu den Klassenstufen 5 bis 8 werden 25% der Beweisversuche gezogen (Abschn. 7.3.3).

Insgesamt scheint die Tendenz, dass die Beweisversuche der Teilnehmenden der Mathematik-Olympiade mit Blick auf den Nachweis der Gültigkeit je höher die Klassenstufe, desto tragfähiger sind, auch unter Berücksichtigung der Klassenstufe 9 zuzutreffen.

Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage
Fokus: Klassenstufe 9 vs. Klassenstufe 5-8; p-Werte von Signifikanztests und dazugehörige Effektstärken

	NwG3.6 vs. ¬NwG3.6	NwG3.5-3.6 vs. ¬NwG3.5-3.6	(NwG3.5-3.6,Sw3-4) vs. ¬(NwG3.5-3.6,Sw3-4)
Kl. 9 vs. Kl. 8	$p = .004^{ex}$, $V = .233$	$p = .016^{N-1}$, $V = .182$	$p < .001^{ex}$, $V = .524$
Kl. 9 vs. Kl. 7	$p < .001^{ex}$, $V = .275$	$p < .001^{N-1}$, $V = .369$	$p < .001^{N-1}$, $V = .380$
Kl. 9 vs. Kl. 6	$p < .001^{ex}$, $V = .264$	$p < .001^{N-1}$, $V = .390$	$p < .001^{N-1}$, $V = .452$
Kl. 9 vs. Kl. 5	$p < .001^{ex}$, $V = .342$	$p < .001^{N-1}$, $V = .523$	$p < .001^{N-1}$, $V = .279$
	NwG3.1-3.6 vs. ¬NwG3.1-3.6	NwG2.3-2.4,3.3-3.6 vs. ¬NwG2.3-2.4,3.3-3.6	NwG99,1.0-2.1.1,3.1.1 vs. ¬NwG99,1.0-2.1.1,3.1.1
Kl. 9 vs. Kl. 8	$p = .010^{N-1}$, $V = .196$	$p = .008^{N-1}$, $V = .201$	$p = .053^{N-1}$, $V = .146$
Kl. 9 vs. Kl. 7	$p = .001^{N-1}$, $V = .205$	$p < .001^{N-1}$, $V = .292$	$p = .034^{N-1}$, $V = .131$
Kl. 9 vs. Kl. 6	$p < .001^{N-1}$, $V = .225$	$p < .001^{N-1}$, $V = .468$	$p < .001^{N-1}$, $V = .265$
Kl. 9 vs. Kl. 5	$p < .001^{N-1}$, $V = .361$	$p < .001^{N-1}$, $V = .638$	$p < .001^{N-1}$, $V = .354$

NwG3.6 *begründender Ansatz, korrektes und vollständiges, deduktives Argumentgefüge*
NwG3.5-3.6 *begründender Ansatz, weitgehend korrektes und vollständiges Argumentgefüge*
NwG3.5-3.6,Sw3-4 *begründender Ansatz, weitgehend korrektes und vollständiges, deduktives Argumentgefüge*
NwG3.1-3.6 *begründender Ansatz*
NwG2.3-2.4,3.3-3.6 *weiterentwickelbares Argumentgefüge*
NwG99,1.0-2.1.1,3.1.1 *kein weiterentwickelbares Argumentgefüge und kein Beispiel zu der zu zeigenden Aussage*

Tabelle 20: Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Fokus: Klassenstufe 9 vs. Klassenstufen 5-8; p-Werte von Signifikanztests und dazugehörige Effektstärken

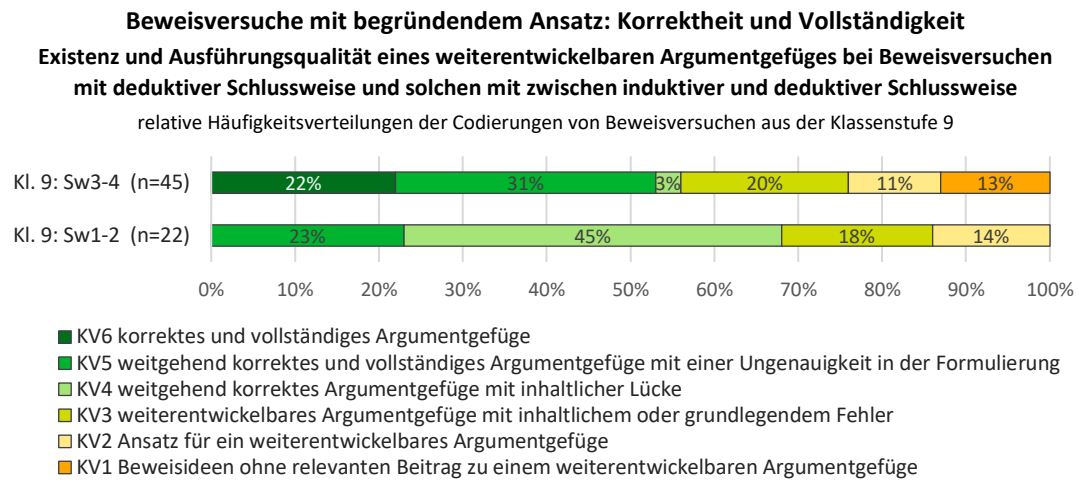
10.4.1.2 Klassenstufe 9: Charakteristika der Beweisversuche

In diesem Abschnitt wird beschrieben, inwieweit verschiedene Eigenschaften bei den analysierten Beweisversuchen aus der Klassenstufe 9 besonders häufig oder eher selten sind und welche Gemeinsamkeiten und Unterschiede diesbezüglich im Vergleich zu den Beweisversuchen aus den Klassenstufen 5 bis 8 bestehen. In Analogie zum Abschnitt 8.5.2 werden dazu nacheinander der Ansatz, die Korrektheit und Vollständigkeit und die Schlussweise, die Beweisideen, die Verwendung von Beispielen und die Darstellungsformen fokussiert.

Ansatz, Korrektheit, Vollständigkeit und Schlussweise

Entsprechend der vorhergehenden Beschreibung der Beweisversuche mit Blick auf den Nachweis der Gültigkeit erscheint die **Existenz eines begründenden Ansatzes und eines weitgehend korrekten und vollständigen Argumentgefüges** für Beweisversuche aus der Klassenstufe 9 als charakteristisch. Zudem fällt auf, dass aus rund 85% der Beweisversuche aus der Klassenstufe 9 der Bezug auf die zu zeigende Aussage hervorgeht.

Die **Schlussweise** ist bei fast zwei Drittel aller Beweisversuche aus der Klassenstufe 9 deduktiv. Der Anteil der Beweisversuche mit deduktiver Schlussweise ist in der Klassenstufe 9 statistisch signifikant größer als in den Klassenstufen 7 und 8 (Kl. 7 vs. Kl. 9: $p < .001^{N-1}$, $V = .261$; Kl. 8 vs. Kl. 9: $p = .014^{N-1}$, $V = .203$). Die Schlussweise ist sowohl bei mehr als zwei Drittel der Beweisversuche mit begründendem Ansatz ohne weiterentwickelbares Argumentgefüge (KV1-2) als auch bei mehr als zwei Drittel der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) deduktiv. Bei den Beweisversuchen mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit



Sw1-2 vs. Sw3-4: p-Werte von Signifikanztests und dazugehörige Effektstärken

alle Kategorien	KV1 vs. ¬KV1	KV4 vs. ¬KV4	KV1-4 vs. KV5-6	KV1-5 vs. KV6
$p = .394^{U}$, $r = .104$	$p = .167^{ex}$, $V = .219$	$p < .001^{ex}$, $V = .548$	$p = .018^{N-1}$, $V = .290$	$p = .024^{ex}$, $V = .293$

Abbildung 188: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges bei Beweisversuchen mit deduktiver Schlussweise und solchen mit zwischen induktiver und deduktiver Schlussweise. Relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen aus der Klassenstufe 9 sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

inhaltlicher Lücke (KV4) ist die Schlussweise in ca. 90% der Fälle zwischen induktiv und deduktiv. Anders als in den Klassenstufen 5 bis 8 sind in der Klassenstufe 9 die Beweisversuche mit deduktiver Schlussweise hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges im Mittel nicht eindeutig stärker oder schwächer als diejenigen mit zwischen induktiver und deduktiver Schlussweise (Abb. 188; Abschn. 8.2.1.3).

Beweisideen

In rund 94% der analysierten Beweisversuche aus der Klassenstufe 9 ist mindestens eine **Beweisidee** zu erkennen. Es handelt sich um alle Beweisversuche mit begründendem Ansatz, einen Beweisversuch mit prüfendem Ansatz und einen Beweisversuch mit nicht eindeutig kategorisierbarem Ansatz.

Von den Beweisversuchen mit begründendem Ansatz enthalten, ähnlich wie in der Klassenstufe 8, rund 70% eine der zentralen Beweisideen, rund 16% keine zentrale, aber eine aus-

Beweisversuche mit begründendem Ansatz: Beweisideen
Fokus: Klassenstufe 9 vs. Klassenstufen 5-8; p-Werte von Signifikanztests und dazugehörige Effektstärken

	BIziel	$BIziel3^2$ vs. $\neg BIziel3^2$	$BIziel2\text{-}3^2$ vs. $\neg BIziel2\text{-}3^2$
Kl. 9 vs. Kl. 8	$p = .352^{\chi^2\text{-}1}$, $V = .119$	$p = .002^{N-1}$, $V = .260$	$p < .001^{N-1}$, $V = .369$
Kl. 9 vs. Kl. 7	$p = .007^{\chi^2\text{-}1}$, $V = .218$	$p < .001^{N-1}$, $V = .283$	$p < .001^{N-1}$, $V = .470$
Kl. 9 vs. Kl. 6	$p < .001^{\chi^2\text{-}1}$, $V = .422$	$p < .001^{N-1}$, $V = .389$	$p < .001^{N-1}$, $V = .466$
Kl. 9 vs. Kl. 5	$p < .001^{\chi^2\text{-}1}$, $V = .609$	$p < .001^{N-1}$, $V = .408$	$p < .001^{N-1}$, $V = .596$

BIziel zielführendste Beweisidee; $BIziel3^2$ mehrere zentrale Beweisideen; $BIziel2\text{-}3^2$ mehrere ausbaubare Beweisideen

	$BIp2,8$ vs. $\neg BIp2,8$	$BIp3,5$ vs. $\neg BIp3,5$	$BIp1$ vs. $\neg BIp1$
Kl. 9 vs. Kl. 8	$p < .001^{N-1}$, $V = .329$	$p = .122^{N-1}$, $V = .128$	$p < .001^{N-1}$, $V = .328$
Kl. 9 vs. Kl. 7	$p < .001^{N-1}$, $V = .475$	$p = .028^{ex}$, $V = .163$	$p < .001^{N-1}$, $V = .251$
Kl. 9 vs. Kl. 6	$p < .001^{N-1}$, $V = .431$	$p = .007^{ex}$, $V = .193$	$p = .782^{N-1}$, $V = .018$
Kl. 9 vs. Kl. 5	$p < .001^{N-1}$, $V = .564$	$p = .004^{ex}$, $V = .210$	$p = .237^{N-1}$, $V = .080$

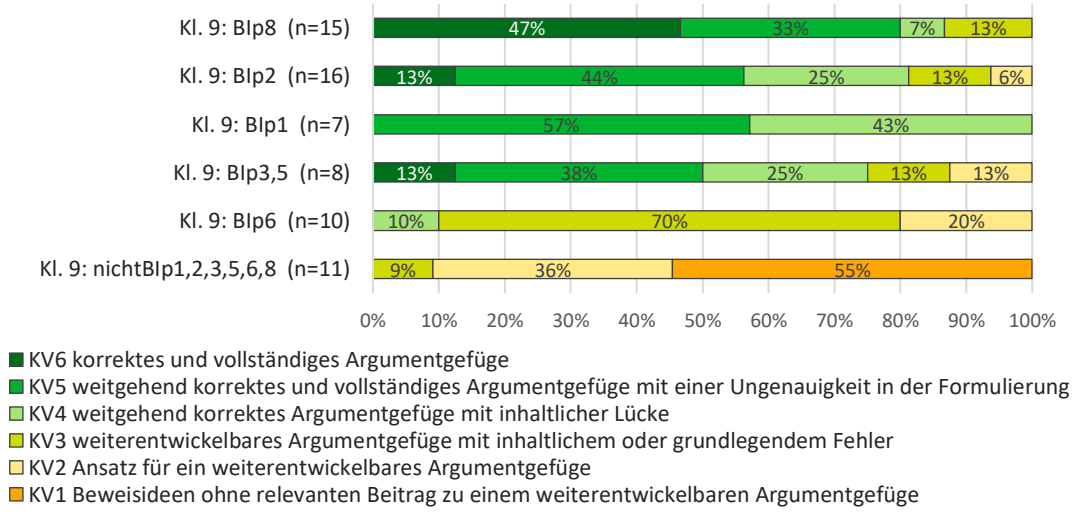
Tabelle 21: Beweisversuche mit begründendem Ansatz: Beweisideen. Fokus: Klassenstufe 9 vs. Klassenstufen 5-8; p-Werte von Signifikanztests und dazugehörige Effektstärken

baubare Beweisidee und rund 13% keine ausbaubare Beweisidee. Dabei fällt auf, dass in der Klassenstufe 9 in über 20% der Beweisversuche mit begründendem Ansatz **mehrere zentrale Beweisideen** vorkommen, während in den Klassenstufen 5 bis 8 jeweils in maximal 5% und damit in einem statistisch hoch signifikant kleineren Anteil der Beweisversuche mit begründendem Ansatz mehrere zentrale Beweisideen ausgeführt sind (Tab. 21). Der Anteil der Beweisversuche mit **mehreren ausbaubaren Beweisideen** an allen Beweisversuchen mit begründendem Ansatz liegt in der Klassenstufe 9 über 50%, in jeder der Klassenstufen 5 bis 8 unter 20%, wobei die Unterschiede ebenfalls statistisch hoch signifikant sind (Tab. 21). Die Existenz mehrerer ausbaubarer Beweisideen kann also als ein Charakteristikum der Beweisversuche mit begründendem Ansatz aus der Klassenstufe 9 angesehen werden.

In Einklang mit den vorhergehenden Ausführungen ist in fast der Hälfte der Beweisversuche mit begründendem Ansatz aus der Klassenstufe 9 eine der Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* die **primäre Beweisidee**, während dieser Anteil in jeder der Klassenstufen 5 bis 8 unter 20% liegt und jeweils statistisch hoch signifikant kleiner ist (Tab. 21). Die Existenz einer dieser primären Beweisideen erscheint somit für Beweisversuche mit begründendem Ansatz aus der Klassenstufe 9 charakteristisch. Außerdem fällt auf, dass die primären Beweisideen 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung* und 5 *Reste bei Division einzelner Summanden durch vier oder acht* zusammen genommen rund 12% der Beweisversuche mit begründendem Ansatz und somit einen mehr als doppelt so großen, teils statistisch signifikant größeren Anteil als in jeder der Klassenstufen 5 bis 8 ausmachen (Tab. 21). Die

Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges differenziert nach primären Beweisideen

relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen aus der Klassenstufe 9



p-Werte von Signifikanztests und dazugehörige Effektstärken

	alle Kategorien	KV1-3 vs. KV4-6	KV1-4 vs. KV5-6
BIp1,2,3,5,8 vs. ¬BIp1,2,3,5,8	$p < .001^{U}$, $r = .723$	$p < .001^{N-1}$, $V = .757$	$p < .001^{N-1}$, $V = .590$
BIp2 vs. BIp6	$p = .001^{U}$, $r = .400$	$p = .001^{ex}$, $V = .695$	$p = .004^{ex}$, $V = .575$

Abbildung 189: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges differenziert nach primären Beweisideen. Relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen aus der Klassenstufe 9 sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

primäre Beweisidee 1 *Vergrößerung einer Summe mit Erkennen einer Invariante* kommt in der Klassenstufe 9 mit rund 10% in den Beweisversuchen mit begründendem Ansatz statistisch hoch signifikant seltener vor als in den Klassenstufen 7, 8 mit etwa 35% (Tab. 21).

Im Falle einiger der **primären Beweisideen** unterscheiden sich die Beweisversuche mit begründendem Ansatz hinsichtlich der **Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges** deutlich (Abb. 189). Beispielsweise enthalten in der Klassenstufe 9 etwa 80% der Beweisversuche mit der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* ein weitgehend korrektes Argumentgefüge, hingegen 90% der Beweisversuche mit der primären Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* kein weitgehend korrektes Argumentgefüge. Bei den Beweisversuchen mit begründendem Ansatz und einer der primären Beweisideen 1, 2, 3, 5, 8 ist die Wahrscheinlichkeit, eine Zuordnung zu einer höheren der Kategorien zu erhalten, statistisch hoch signifikant größer als im Falle einer anderen primären Beweisidee. Zudem lässt sich festhalten, dass für die Beweisversuche aus der Klassenstufe 9 mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge (KV5-6) die primären Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* mit einem relativen Anteil von über 70% charakteristisch sind. Insgesamt betrachtet passen die erkennbaren Tendenzen bezüglich der Zusammenhänge zwischen primären Beweisideen und der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges zu denen in den Klassenstufen 6 bis 8 (Abb. 189; Abschn. 8.2.6.3).

Weiterhin sind in der Klassenstufe 9, ähnlich wie in den Klassenstufen 5 bis 8, im Falle einzelner **primärer Beweisideen** Auffälligkeiten hinsichtlich der **Schlussweise** zu sehen. Beispielsweise ist bei den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* der Anteil derjenigen mit zwischen induktiver und deduktiver Schlussweise mit rund 86% statistisch hoch signifikant größer als bei den übrigen Beweisversuchen mit begründendem Ansatz ($p = .004^{ex}$, $V = .385$). Bei den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* ist die Schlussweise mit rund 93% statistisch hoch signifikant häufiger deduktiv als bei den übrigen Beweisversuchen mit begründendem Ansatz ($p = .004^{ex}$, $V = .299$). Andersherum betrachtet basieren von den Beweisversuchen mit zwischen induktiver und deduktiver Schlussweise mehr als die Hälfte auf einer der primären Beweisideen 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* und 6 *Teilbarkeit der Summen zweier Zahlenpaare*. Von den Beweisversuchen mit deduktiver Schlussweise basieren gut 50% auf einer der primären Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*.

Beispiele

Ein Beispiel kommt in der Klassenstufe 9 in rund 54% der Beweisversuche, die nicht fehlend oder durchgestrichen sind, vor. Dieser Anteil der **Beweisversuche mit Beispiel** ist in der Klassenstufe 9 kleiner als in jeder der Klassenstufen 5 bis 8. Der Unterschied zu jeder der Klassenstufen 6 bis 8 ist statistisch hoch signifikant (jeweils: $p < 0{,}01^{N-1}$; $0{,}16 < V < 0{,}26$).

In der Klassenstufe 9 erscheinen die **Beweisversuche mit begründendem Ansatz und einem Beispiel zu der zu zeigenden Aussage** hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges im Mittel weder stärker noch schwächer als die Beweisversuche mit begründendem Ansatz ohne Beispiel. Insbesondere enthalten von den Beweisversuchen mit begründendem Ansatz ohne Beispiel, denen mit einem Beispiel, aber ausschließlich sonstigen Beispielen und denen mit einem Beispiel zu der zu zeigenden Aussage jeweils die Hälfte bis zu zwei Drittel ein weitgehend korrektes Argumentgefüge.

Eine detaillierte Betrachtung verdeutlicht, dass sich vor allem die **Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge** aus der Klassenstufe 9 von denen in den Klassenstufen 6 bis 8 in Bezug auf die Beispielnutzung unterscheiden. Während in beiden Klassenstufen 7, 8 von den Beweisversuchen mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge über 70% mindestens ein Beispiel, über 60% mehrere Beispiele und mehr als 50% Reihenfolgebeispiele enthalten, sind es in der 9. Klassenstufe 48%, 35% bzw. 20%. Alle drei Anteile sind in der Klassenstufe 9 statistisch hoch signifikant kleiner als in beiden Klassenstufen 7, 8 (Tab. 22). Eine ergänzende Differenzierung nach primären Beweisideen deutet an, dass diese Unterschiede mit den unterschiedlichen Häufigkeitsverteilungen der primären Beweisideen zusammenhängen. Beispielsweise kommen in jeder der Klassenstufen 6 bis 9 in mehr als 75% der Beweisversuche mit begründendem Ansatz, einem weitgehend korrekten Argumentgefüge und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* Reihenfolgebeispiele vor. Demgegenüber enthalten in jeder der Klassenstufen 6 bis 9 maximal 20% der Beweisversuche mit begründendem Ansatz, einem weitgehend korrekten Argumentgefüge und einer der primären Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* Reihenfolgebeispiele. Dabei ist die primäre Beweisidee 1 *sukzessive Vergrößerung einer Summe in Abhängigkeit des kleinsten Summanden* in der Klassenstufe 9 statistisch hoch signifikant seltener als in den Klassenstufen 7, 8 (Tab. 21). Beweisversuche mit einer der primären Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* kommen in der Klassenstufe 9 statistisch hoch signifikant häufiger vor als in jeder der Klassenstufen 5 bis 8 (Tab. 21).

Neben den beschriebenen Unterschieden sind bezüglich der Verwendung von Beispielen in der Klassenstufe 9 ähnliche Tendenzen zu erkennen wie in den Klassenstufen 5 bis 8. Beispielsweise sind in allen Klassenstufen 5 bis 9 in weniger als 10% der Beweisversuche mit Beispiel **mehr als fünf Beispiele** zu finden. Ein **großes Beispiel** kommt ebenfalls in unter 10% der Beweisversuche mit Beispiel vor. In keinem der Beweisversuche mit einem großen Beispiel ist ein

Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge
Fokus: Klassenstufe 9 vs. Klassenstufen 6 - 8; p-Werte von Signifikanztests und dazugehörige Effektstärken

	mit/ohne Beispiel	mit/ohne mehrere Beispiele	mit/ohne Reihenfolgebeispiele
Kl. 9 vs. Kl. 8	$p = .004^{N-1}$, $V = .322$	$p = .002^{N-1}$, $V = .343$	$p = .001^{N-1}$, $V = .381$
Kl. 9 vs. Kl. 7	$p < .001^{N-1}$, $V = .623$	$p < .001^{N-1}$, $V = .561$	$p < .001^{N-1}$, $V = .462$
Kl. 9 vs. Kl. 6	$p = .030^{N-1}$, $V = .251$	$p = .007^{N-1}$, $V = .312$	$p = .119^{N-1}$, $V = .180$

Tabelle 22: Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge. Fokus: Klassenstufe 9 vs. Klassenstufen 6-8; p-Werte von Signifikanztests und dazugehörige Effektstärken

weitgehend korrektes und vollständiges Argumentgefüge zu erkennen. Das **kleinste Beispiel** ist in etwa der Hälfte der Beweisversuche mit Beispiel angegeben, in weiteren knapp 15% der Beweisversuche mit Beispiel ist ein anderes besonderes Beispiel vorhanden. Außerdem sind in allen Klassenstufen 6 bis 9 **Zusammenhänge zwischen der Beispielnutzung und der primären Beweisidee** zu erkennen. Was die Klassenstufe 9 betrifft, ist diesbezüglich hervorzuheben, dass in 13% der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* ein Beispiel vorkommt, der Anteil ist im Vergleich zu den anderen Beweisversuchen mit begründendem Ansatz statistisch signifikant kleiner ($\mathrm{p} = .001^{\mathrm{N\text{-}1}}$, $\mathrm{V} = .418$). Demgegenüber enthalten rund 90% der Beweisversuche mit begründendem Ansatz und der primären Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* ein Beispiel, der Anteil ist im Vergleich zu den anderen Beweisversuchen mit begründendem Ansatz statistisch signifikant größer ($\mathrm{p} = .014^{\mathrm{ex}}$, $\mathrm{V} = .317$). Von den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* enthalten in der Klassenstufe 9 etwa 85% Reihenfolgebeispiele und etwa 85% das kleinste Beispiel. Beide Anteile sind im Vergleich zu den anderen Beweisversuchen mit begründendem Ansatz statistisch hoch signifikant größer (jeweils: $\mathrm{p} < .001^{\mathrm{ex}}$, $0{,}49 < \mathrm{V} < 0{,}68$; Abschn. 8.2.3.2, 8.2.6.4, 8.3.2.2, 8.5.2.3 zu den Klassenstufen 5 bis 8).

Darstellungsform

Von den Beweisversuchen aus der Klassenstufe 9, die nicht fehlend oder durchgestrichen sind[168], enthalten 97% verbale Darstellungsmittel und 96% numerische Darstellungsmittel, ähnlich wie in den Klassenstufen 5 bis 8. Die Darstellungsform ist jedoch ausschließlich in rund einem Drittel der Beweisversuche aus der Klassenstufe 9 **numerisch-verbal** und damit in einem statistisch hoch signifikant kleineren Anteil als in jeder der Klassenstufen 5 bis 8 (jeweils: $\mathrm{p} \leq .003^{\mathrm{N\text{-}1}}$; $0{,}16 < \mathrm{V} < 0{,}27$). Die Darstellungsformen **verbal, ikonisch-verbal** kennzeichnen etwa ein Achtel der Beweisversuche aus der Klassenstufe 9 und damit einen statistisch hoch signifikant kleineren Anteil als in den Klassenstufen 5, 6 (jeweils: $\mathrm{p} < .001^{\mathrm{N\text{-}1}}$; $0{,}16 < \mathrm{V} < 0{,}24$).

Am häufigsten sind in der Klassenstufe 9 Beweisversuche mit einer **Darstellungsform mit symbolischem Anteil**. Sie machen etwa die Hälfte aller Beweisversuche aus, die nicht fehlend oder durchgestrichen sind, und damit einen statistisch hoch signifikant größeren Anteil dieser Beweisversuche als in jeder der Klassenstufen 5 bis 8 (jeweils: $\mathrm{p} < .001^{\mathrm{N\text{-}1}}$; $0{,}38 \leq \mathrm{V} < 0{,}66$). Zudem fällt auf, dass in gut einem Drittel der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil mathematische Strukturierungselemente vorkommen.

In der Klassenstufe 9 ist bei den Beweisversuchen mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil im Vergleich zu den anderen Beweisversuchen mit begründendem Ansatz nach dem Mann-Whitney-U-Test die Wahrscheinlichkeit größer, bezüglich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in eine höhere der Kategorien zu gehören (Abb. 190). Etwa 70% der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil enthalten ein weitgehend korrektes und vollständiges

[168] Die nachfolgenden relativen Häufigkeitsangaben zu Darstellungsmitteln und Darstellungsformen beziehen sich, ohne ständige Erwähnung, durchgehend auf die Beweisversuche, die nicht fehlend oder durchgestrichen sind.

Argumentgefüge, knapp 80% ein weitgehend korrektes Argumentgefüge. Andersherum betrachtet hat die Darstellungsform insbesondere bei allen Beweisversuchen, die als Beweis bezeichnet werden können, einen symbolischen Anteil (Abb. 190).

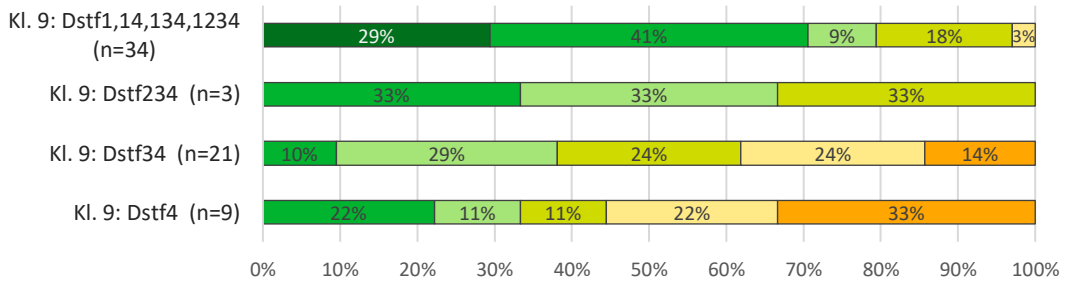
Bei rund zwei Drittel der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil aus der Klassenstufe 9 ist die Schlussweise deduktiv ohne zusätzliches Beispiel. Der Anteil der Beweisversuche mit Beispiel ist im Fall einer Darstellungsform mit symbolischem Anteil statistisch hoch signifikant kleiner als bei den anderen Beweisversuchen aus der Klassenstufe 9 ($p = .002^{N-1}$, $V = .370$).

Die analysierten Beweisversuche mit einer Darstellungsform mit symbolischem Anteil aus der Klassenstufe 9 basieren auf fünf verschiedenen primären Beweisideen. Dabei ist in ca. 80% dieser Beweisversuche eine der Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen* die primäre Beweisidee. Der Anteil der primären Beweisideen 2, 8 ist bei den Beweisversuchen mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil statistisch hoch signifikant größer als bei den anderen Beweisversuchen mit begründendem Ansatz ($p < .001^{N-1}$, $V = .675$). Andersherum betrachtet haben von den Beweisversuchen mit einer der primären Beweisideen 2, 8 über 85% eine Darstellungsform mit symbolischem Anteil. Von den Beweisversuchen mit begründendem Ansatz und der primären Beweisidee 1 sind es ca. 40%. Von denen mit einer der ausbaubaren primären Beweisideen 3 bis 7 sind es ca. 15%.

Beweisversuche mit begründendem Ansatz

Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges je nach Darstellungsform

relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen aus der Klassenstufe 9



KV6 korrektes und vollständiges Argumentgefüge
KV5 weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung
KV4 weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke
KV3 weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler
KV2 Ansatz für ein weiterentwickelbares Argumentgefüge
KV1 Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge

Dstf1,14,134,1234 *Darstellungsform mit symbolischem Anteil*; Dstf4 *Darstellungsform verbal*
Dstf234 *Darstellungsform ikonisch-numerisch-verbal*; Dstf34 *Darstellungsform numerisch-verbal*
Andere als diese Darstellungsformen kommen in den analysierten Beweisversuchen aus der Klassenstufe 9 nicht vor.

p-Werte von Signifikanztests und dazugehörige Effektstärken

	Dstf1,14,134,1234 vs ¬Dstf1,14,134,1234	Dstf1,14,134,1234 vs. Dstf4	Dstf1,14,134,1234 vs. Dstf34
Kl. 9	$p < .001^{U}$, $r = .589$	$p < .001^{U}$, $r = .537$	$p < .001^{U}$, $r = .674$

Abbildung 190: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges je nach Darstellungsform. Relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen aus der Klassenstufe 9 sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

In der Klassenstufe 9 gibt es unter den analysierten Beweisversuchen mit begründendem Ansatz drei (4%), bei denen die Darstellungsform keinen symbolischen, aber **ikonisch-numerischen Anteil** hat. In diesen Beweisversuchen ist ein weiterentwickelbares Argumentgefüge vorhanden. Zwei der drei Beweisversuche basieren auf der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Weiterhin hat einer der analysierten Beweisversuche aus der Klassenstufe 9 die Darstellungsform *symbolisch-ikonisch-numerisch-verbal*. Die primäre Beweisidee ist bei diesem Beweisversuch die Idee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*. Bei den Beweisversuchen mit begründendem Ansatz aus der Klassenstufe 9 ist im Fall der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* mit einer relativen Häufigkeit von 43% der Anteil der Darstellungsformen mit ikonisch-numerischem Anteil am größten. Dies steht in Einklang mit den Ergebnissen zu den Klassenstufen 5 bis 8 (Abschn. 8.2.6.5).

10.4.1.3 Klassenstufe 9: Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche zu einem Beweis

Um die Beweisversuche aus der Klassenstufe 9 in einem dritten Schritt mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis zu beschreiben, wird die anhand der Beweisversuche aus den Klassenstufen 5 bis 8 entwickelte Klassifikation verwendet (insb. Abschn. 8.5.3). Die entwickelte Klassifikation, d. h. vor allem die Differenzierung von Beweisversuchen mit unterschiedlichen primären Beweisideen, erscheint auch mit Blick auf die Klassenstufe 9 als angemessen. Beispielsweise ist, ähnlich wie in den Klassenstufen 5 bis 8, im Falle der primären Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* die Existenz von Reihenfolgebeispielen besonders häufig und der Anteil der Beweisversuche mit einer Darstellungsform mit ikonisch-numerischem Anteil am größten (Abschn. 8.5.2, 10.4.1.2). Wie groß die jeweiligen Anteile der Beweisversuche aus der Klassenstufe 9 in Bezug auf die in der Klassifikation unterschiedenen Gruppen sind, ist in Abbildung 191 dargestellt.[169]

Mit Blick auf Anknüpfungspunkte für die Weiterentwicklung der Beweisversuche aus der Klassenstufe 9 zu einem Beweis ist grundlegend festzuhalten, dass die Gruppe **We9900** ***fehlend*** kaum vorkommt. Es lässt sich somit davon ausgehen, dass in der Regel irgendeine Auseinandersetzung mit der Aufgabenstellung erfolgt ist. Zudem ist in fast 95% der Beweisversuche zumindest irgendeine **Beweisidee** zu erkennen, die sich prinzipiell aufgreifen lässt (Abschn. 10.4.1.2). Überdies enthalten über 90% der Beweisversuche einen **begründenden Ansatz** (We3125-3453), was andeutet, dass viele der Teilnehmenden nach allgemeingültigen oder verallgemeinerbaren Argumenten gesucht haben, und ebenfalls einen Anknüpfungspunkt für eine Weiterentwicklung der Beweisversuchen darstellt. Außerdem lässt sich bei der Weiterentwicklung von Beweisversuchen aus der Klassenstufe 9 in der Regel davon ausgehen, dass die Begrifflichkeiten der zu zeigenden Aussage korrekt interpretiert wurden. Zumindest gehört keiner der analysierten Beweisversuche zu einer der Gruppen, die sich primär durch einen **Hinweis auf einen grundlegenden Fehler** charakterisieren (Abb. 191).

[169] Die in Abbildung 191 zu sehenden Angaben beziehen sich genau auf die Beweisversuche aus der Klassenstufe 9, die als Teil der Zufallsstichprobe analysiert wurden. Es ist wahrscheinlich, dass das Spektrum an vorkommenden Gruppen größer ist, als dies aus Abbildung 191 hervorgeht.

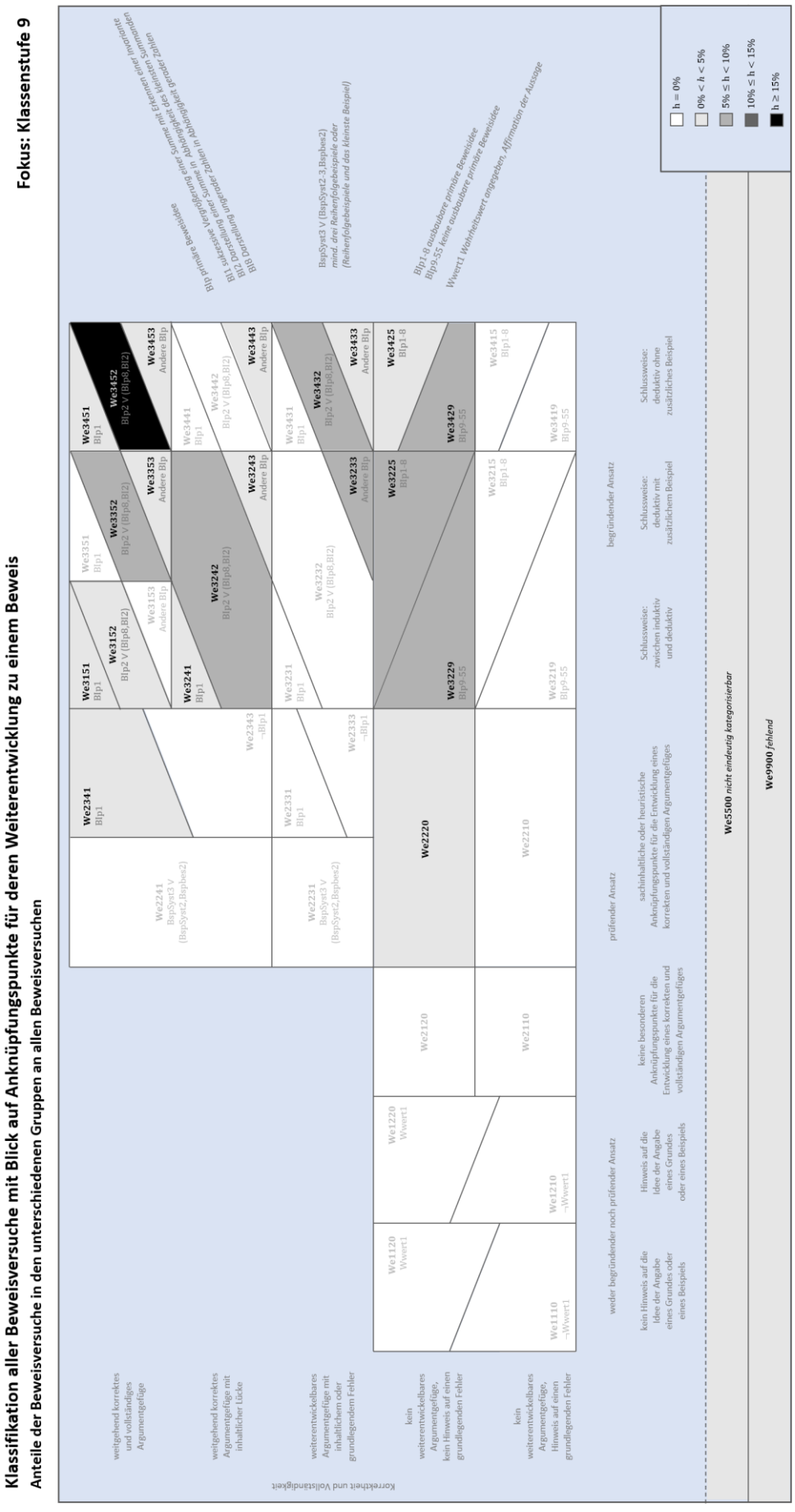


Abbildung 191: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 9)

In der Klassenstufe 9 besonders häufig sind Beweisversuche in der Gruppe We3452 und ergänzend solche in der Gruppe We3352. Der Anteil der Beweisversuche in einer der Gruppen **We3352,3452** liegt über 25%. Die Beweisversuche in diesen Gruppen, wie z. B. die Bearbeitung 10.4.1a in Abbildung 192, enthalten alle ein weitgehend korrektes und vollständiges Argumentgefüge bei deduktiver Schlussweise. Die primäre Beweisidee ist immer eine der Ideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*, was in der Regel eine Reformulierung der zu zeigenden Aussage impliziert. Die Darstellungsform hat in fast allen analysierten Fällen symbolischen Anteil. Zudem kommen in mehr als einem Drittel der analysierten Fälle auch Symbole mit fester Bedeutung oder mathematische Strukturierungselemente vor. Da die Beweisversuche in den Gruppen We3352,3452 allenfalls Formulierungsfehler enthalten, lassen die Ergebnisse vermuten, dass den meisten Teilnehmenden, die einen der Beweisversuche in den Gruppen We3352,3452 formuliert haben, ein gewisser Umgang mit elementarer Algebra gelingt. Zugleich ist festzuhalten, dass in etwa zwei Drittel der Beweisversuche in den Gruppen We3352,3452 Formulierungsfehler oder Formulierungslücken vorkommen, d. h. im Detail auch bei vielen der Beweisversuche in diesen Gruppen Überarbeitungsbedarf zu erkennen ist.

Bearbeitung 10.4.1a: Klassenstufe 9, weiblich (Code: 0910225)

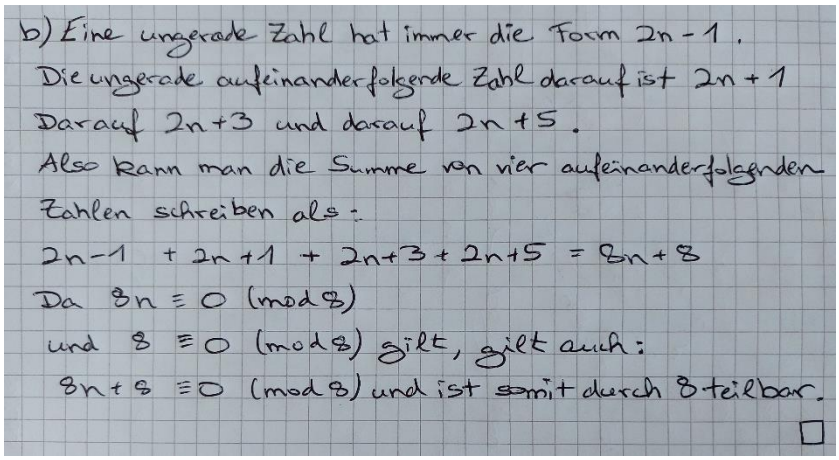
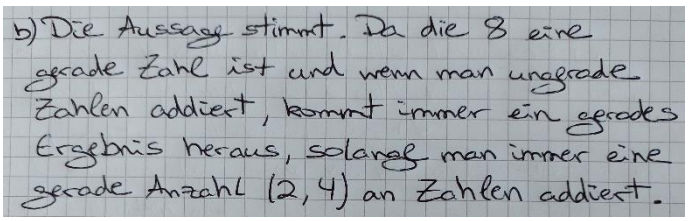


Abbildung 192: Bearbeitung 10.4.1a (We3452 *begründender Ansatz, Schlussweise deduktiv ohne zusätzliches Beispiel, weitgehend korrektes und vollständiges Argumentgefüge, primäre Beweisidee* 2 *oder primäre Beweisidee* 8 *zusammen mit der Beweisidee* 2)

In der Klassenstufe 9 sind ebenfalls **Beweisversuche mit begründendem Ansatz, aber keinem weitgehend korrekten Argumentgefüge** (We3215,3219,3225,3229,3231,3232,3233,3415, 3419,3425,3429,3431,3432,3433) häufig. Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler (We3231,3232,3233,3431,3432,3433) machen rund 20% aller Beweisversuche aus der Klassenstufe 9 aus, Beweisversuche mit begründendem Ansatz und keinem weiterentwickelbaren Argumentgefüge (We3215,3219,3225,3229,3415,3419,3425,3429) weitere 20%. Bei allen diesen Beweisversuchen sind die vorhandenen Ausführungen für einen Beweis der zu zeigenden Aussage zu hinterfragen und deutlich zu verändern. Dabei lässt sich hinsichtlich der Beispielnutzung und der Verwendung von Darstellungsmitteln Folgendes anmerken: Ähnlich

wie in jeder der Klassenstufen 5 bis 8 kommt in der Klassenstufe 9 in rund 60% dieser Beweisversuche kein Beispiel vor. Der Anteil derjenigen mit mehreren Beispielen liegt bei rund 30%, der Anteil derjenigen mit Reihenfolgebeispielen bei ca. 10%. In den Beweisversuchen mit begründendem Ansatz, aber keinem weiterentwickelbaren Argumentgefüge (We3215,3219, 3225,3229,3415,3419,3425,3429) aus der Klassenstufe 9 sind fast immer nur numerische oder verbale Darstellungsmittel und eventuell strukturierende Darstellungsmittel zu finden (Bearbeitung 10.4.1b in Abb. 193). Die Beweisversuche mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit einem inhaltlichen oder grundlegenden Fehler (We3231,3232,3233,3431,3432,3433) aus der Klassenstufe 9 haben zum Teil eine Darstellungsform mit symbolischem Anteil, was aber nicht unbedingt impliziert, dass der inhaltliche oder grundlegende Fehler durch Schwierigkeiten im Umgang mit Algebra bedingt ist (Bearbeitung 10.4.1c in Abb. 193).

Bearbeitung 10.4.1b: Klassenstufe 9, weiblich (Code: 0921808)



Bearbeitung 10.4.1c: Klassenstufe 9, männlich (Code: 0920625)

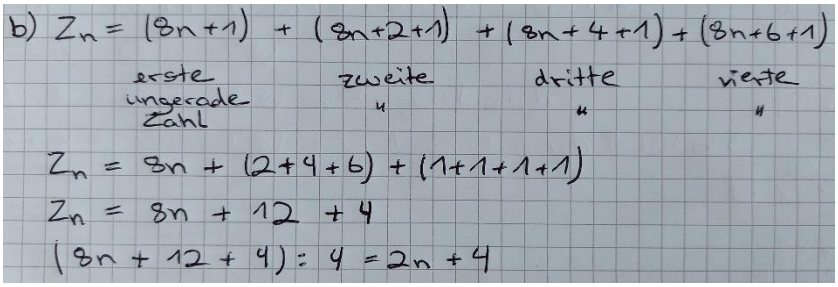


Abbildung 193: Bearbeitung 10.4.1b (We3249 *begründender Ansatz, Schlussweise deduktiv ohne zusätzliches Beispiel, kein weiterentwickelbares Argumentgefüge, kein Hinweis auf einen grundlegenden Fehler, keine ausbaubare primäre Beweisidee*), Bearbeitungen 10.4.1c (We3342 *begründender Ansatz, Schlussweise deduktiv ohne zusätzliches Beispiel, weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler, primäre Beweisidee* 2 *oder primäre Beweisidee* 8 *zusammen mit der Beweisidee* 2)

10.4.1.4 Klassenstufe 9: Zusammenfassung und Diskussion

Die Beweisversuche aus der Klassenstufe 9 erscheinen mit Blick auf den **Nachweis der Gültigkeit** im Mittel tragfähiger als die aus den Klassenstufen 5 bis 8. In der Klassenstufe 9 ist unter anderem der Anteil der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge statistisch signifikant größer als in jeder der Klassenstufen 5 bis 8 (Abschn. 10.4.1.1). Die Ergebnisse passen zu der Idee eines kumulativen Lernprozesses und den Erwartungen an Beweisversuche von Teilnehmenden der Mathematik-Olympiade, die sich ausgehend von den Vorgaben in den Bildungsstandards und den Lehrplänen in NRW formulieren lassen (Abschn. 4.2, 9.3.1.2).

Die Ergebnisse der Analyse der Beweisversuche der Neuntklässlerinnen und Neuntklässler lassen sich mit **Ergebnissen von Healy und Hoyles (2000) sowie Chin und Lin (2009)** vergleichen.

In der britischen Studie von Healy und Hoyles (2000) mit mathematisch leistungsstarken Zehntklässlerinnen und Zehntklässlern beträgt der Anteil der Beweise bei einer nicht aus dem Schulunterricht bekannten Allaussage 3%. Insgesamt 9% der Beweisversuche enthalten eine unvollständige oder vollständige deduktive Begründung. In der taiwanesischen Studie von Chin und Lin (2009) mit Lernenden in der neunten Klassenstufe beträgt der Anteil der Beweise bei der nicht aus dem Schulunterricht bekannten Allaussagen 6%. Insgesamt 20% der Beweisversuche enthalten eine unvollständige oder vollständige deduktive Begründung (Abschn. 5.1.1). In der Studie im Rahmen dieser Arbeit enthalten 14% der Beweisversuche von Neuntklässlerinnen und Neuntklässler einen Beweis, rund 33% ein weitgehend korrektes und vollständiges, deduktives Argumentgefüge (Abschn. 10.4.1.1). Die Ergebnisse dieser Arbeit sind also tendenziell besser als die der beiden anderen Studien. Dabei ist die Fokussierung dieser Arbeit auf Teilnehmende der Mathematik-Olympiade, d. h. tendenziell besonders leistungsstarke Schülerinnen und Schüler zu bedenken. Außerdem ist zu berücksichtigen, dass in den Studien in Großbritannien und im Taiwan eine andere Allaussage verwendet wurde als im Rahmen der vorliegenden Arbeit. Zudem lassen sich die Vergleichbarkeit des verwendeten Beweisbegriffs und der Kategoriensysteme nicht klar beurteilen.

In der Studie von Healy und Hoyles (2000) wird die hauptsächliche Darstellungsform der Beweisversuche zu der nicht aus dem Schulunterricht bekannten Allaussage in 3% der Fälle als „formal (algebraic)“ (Healy & Hoyles, 2000, S. 408) bezeichnet. In der Studie im Taiwan beträgt dieser Anteil 11% (Chin & Lin, 2009; Abschn. 5.3.1). Von den im Rahmen dieser Arbeit analysierten Beweisversuchen aus der Klassenstufe 9 haben 47% eine Darstellungsform mit symbolischem Anteil. Der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil erscheint somit groß. Es kann aber sein, dass die Kategorie *Darstellungsform mit symbolischem Anteil* weiter gefasst ist als die Kategorie „formal (algebraic)“ (Chin & Lin, 2009, S. 61; Healy & Hoyles, 2000, S. 408) in den anderen Untersuchungen. Zudem ist auch an dieser Stelle die Fokussierung dieser Arbeit auf Teilnehmende der Mathematik-Olympiade und die Unterschiedlichkeit der zu zeigenden Aussagen zu bedenken.

Manche Studien liefern Hinweise darauf, dass einigen Lernenden die **Verwendung symbolischer Darstellungsmittel** bei der Bearbeitung von Aufgaben mit Beweisnotwendigkeit Schwierigkeiten bereitet und ihnen dabei Fehler passieren (Healy & Hoyles, 2000; Küchemann, 2008; Veilande et al. (2018)). Dies lässt sich mit Blick auf die analysierten Beweisversuche aus der Klassenstufe 9 nicht klar bestätigen. Die Beweisversuche mit begründendem Ansatz und einer Darstellungsform mit symbolischem Anteil sind hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges, vereinfacht gesagt, besser als die anderen Beweisversuche mit begründendem Ansatz. In rund 80% der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil ist ein weitgehend korrektes Argumentgefüge formuliert. Zusätzlich zur Fokussierung auf Teilnehmende der Mathematik-Olympiade ist dabei insbesondere zu berücksichtigen, dass für einen Beweis der in dieser Arbeit vorgegebenen Allaus-

sage elementare Kompetenzen im Umgang mit Algebra ausreichen (Abschn. 6.2.2). Um ergänzende Einblicke zu erhalten, inwieweit den Teilnehmenden aus der Klassenstufe 9 der Umgang mit Algebra beim Führen von Beweisen gelingt, ließen sich weitere, bereits erhobene Beweisversuche dieser Teilnehmenden auswerten. Im Teil c) der spezifisch fokussierten Aufgabe, die die Teilnehmenden ab der Klassenstufe 9 erhalten haben, ist eine Allaussage vorgegeben, deren Beweis die Verwendung von Algebra und dabei auch anspruchsvollere Termumformungen erfordert (Abschn. 10.2.2).

Die Ergebnisse dieser Arbeit weisen auf **Zusammenhänge zwischen mehreren Eigenschaften der Beweisversuche** aus der Klassenstufe 9 hin. Beispielsweise sind bei den Beweisversuchen mit begründendem Ansatz Zusammenhänge zwischen primären Beweisideen und der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges zu erkennen, ähnliche Zusammenhänge wie in den Klassenstufen 6 bis 8. Im Unterschied zu diesen niedrigeren Klassenstufen sind in der Klassenstufe 9 die Beweisversuche mit zwischen induktiver und deduktiver Schlussweise im Mittel hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges ähnlich gut wie die mit deduktiver Schlussweise. Auch sind in der Klassenstufe 9 die Beweisversuche mit begründendem Ansatz und einem Beispiel zu der zu zeigenden Aussage hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüge im Mittel nicht klar besser oder schlechter als die Beweisversuche mit begründendem Ansatz und keinem Beispiel. Die Ergebnisse deuten an, dass es Teilnehmenden der Mathematik-Olympiade aus der Klassenstufe 9 im Vergleich zu denen aus den Klassenstufen 5 bis 8 besser gelingt, deduktiv zu argumentieren und ohne Beispiel zu arbeiten.

Bei Beweisversuchen mit begründendem Ansatz aus der Klassenstufe 9 scheinen insbesondere das häufige Vorkommen der primären Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*, die Häufung an Beweisversuchen mit einer Darstellungsform mit symbolischem Anteil, die Häufung an Beweisversuchen mit deduktiver Schlussweise und die im Vergleich zu den Klassenstufen 7, 8 etwas andere Beispielnutzung zusammenzuhängen. Es lässt sich vermuten, aber nicht sicher sagen, dass einige Unterschiede zwischen den Beweisversuchen aus der Klassenstufe 9 und denen aus den Klassenstufen 5 bis 8 primär dadurch bedingt sind, dass die Schülerinnen und Schüler in der Klassenstufe 9 mit dem Umgang mit Algebra vertrauter sind.

Mit Blick auf **Ansatzpunkte für eine Förderung im Führen von Beweisen** deuten die Analyseergebnisse unter anderem Folgendes an: Bei der Förderung von etwa der Hälfte der Teilnehmenden aus der Klassenstufe 9 wäre wahrscheinlich die Verwendung anspruchsvollerer Aufgaben als die im Rahmen dieser Arbeit fokussierte Beweisaufgabe möglich und sinnvoll. Im Rahmen einer Förderung sollte eine Haltung eines inneren Skeptikers unterstützt werden. Dabei könnten Aktivitäten des Überprüfens und Identifizierens von Bearbeitungsbedarfen, anknüpfend an die vorhandenen Kompetenzen der Lernenden, auf einem unterschiedlichen Niveau erfolgen (Abschn. 9.3.4).

Der Umgang mit Algebra scheint den meisten Teilnehmenden aus der Klassenstufe 9, die Algebra verwenden, bei einer einfachen Beweisaufgabe zu gelingen. Um diesbezüglich ergänzende Informationen zu erhalten, wäre es sinnvoll, die Beweisversuche zum Aufgabenteil c)

der Beweisaufgabe, die Teilnehmende der Mathematik-Olympiade ab der Klassenstufe 9 im Rahmen des Forschungsprojekts erhalten haben, zu analysieren (Abschn. 10.2.2 zum Aufgabenteil c)). Bei den Beweisversuchen mit begründendem Ansatz, aber keinem weitgehend korrekten Argumentgefüge aus der Klassenstufe 9 zeigt sich hinsichtlich der Darstellungsmittel und der Beispielnutzung ein ähnliches Bild wie bei denen aus den Klassenstufen 5 bis 8. Außerdem kommen auch in der Klassenstufe 9 Beweisversuche mit prüfendem Ansatz vor. Vor diesem Hintergrund lässt sich diskutieren, inwieweit bei einer Förderung von Teilnehmenden der Mathematik-Olympiade in der Klassenstufe 9 eine sinnvolle Nutzung von Beispielen einen Schwerpunkt darstellen sollte.

10.4.2 Beweisversuche, Wettbewerbserfolg und Förderung

Im vorliegenden Abschnitt werden die im Rahmen dieser Arbeit analysierten Beweisversuche der Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 9 zusammen mit deren Wettbewerbserfolg und Teilnahme an mathematischen Förderangeboten betrachtet. Es werden Einblicke gegeben, inwieweit besondere Eigenschaften von Beweisversuchen mit einem Wettbewerbserfolg oder einer spezifischen mathematischen Förderung einhergehen. Zudem wird betrachtet, welche Eigenschaften die Beweisversuche der Teilnehmenden mit einem besonderem Wettbewerbserfolg oder einer besonderen mathematischen Förderung zeigen.

Dabei steht zuerst der Wettbewerbserfolg im Fokus. Unter dem Begriff „Wettbewerbserfolg“ wird berücksichtigt, ob sich die Teilnehmenden der Regionalrunde für die anschließende Landesrunde qualifizieren konnten und ob sie an der Regionalrunde im vorhergehenden Schuljahr teilgenommen haben (Abschn. 10.4.2.1-10.4.2.2). Danach liegt der Fokus auf der Teilnahme an mathematischen Förderangeboten. Diesbezüglich wird unter anderem zwischen Lernenden, mit einer intensiven oder langjährigen, aktuellen Förderung und Lernenden mit allenfalls lange zurückliegender Förderung unterschieden (Abschn. 10.4.2.3). Abschließend werden wesentliche Ergebnisse bezüglich des Wettbewerbserfolgs und der Teilnahme an mathematischen Förderangeboten zusammen betrachtet und resümiert (Abschn. 10.4.2.4).

10.4.2.1 Mathematik-Olympiade: Qualifikation für die Landesrunde

Von den Beweisversuchen, die als Teil der Zufallsstichprobe analysiert wurden, stammen je Klassenstufe 2% bis 10% von Teilnehmenden, die sich für die anschließende Landesrunde qualifiziert haben (im Folgenden kurz: „Landesrundenteilnehmenden“). Zudem wurden, entsprechend den Erläuterungen in Abschnitt 7.3.3, in der bewusst gezogenen Stichprobe verstärkt Beweisversuche von Landesrundenteilnehmenden berücksichtigt.

Die im Rahmen dieser Arbeit analysierten Beweisversuche werden im Folgenden genutzt, um Einblicke zu geben, inwieweit die Beweisversuche der Landesrundenteilnehmenden besondere Eigenschaften auszeichnen und inwieweit besondere Eigenschaften der Beweisversuche mit einer Qualifikation für die Landesrunde einhergehen. Im Anschluss an die vorliegende Arbeit wäre es möglich, die Beweisversuche aller Landesrundenteilnehmenden auszuwerten, um diesbezüglich weitere und noch aussagekräftigere Ergebnisse zu erhalten.

Klassenstufen 8 und 9

In den Klassenstufen 8, 9 sind die Beweisversuche der **Landesrundenteilnehmenden** weitgehend homogen. Es handelt sich durchgehend um Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge. Folglich passen in allen Beweisversuchen von Landesrundenteilnehmenden aus den Klassenstufen 8, 9 die Ausführungen zu der zu zeigenden Aussage. Zudem impliziert die Existenz eines weitgehend korrekten und vollständigen Argumentgefüges, dass die Beweisversuche alle eine ausbaubare primäre Beweisidee und eine zentrale Beweisidee enthalten. Oftmals ist die zentrale Beweisidee auch die primäre Beweisidee, was mit Blick auf die Komplexität des Argumentgefüges günstig ist und zugleich impliziert, dass maximal noch eine weitere der klar über die zu zeigende Aussage hinausgehenden Beweisideen 1 bis 8 für einen Beweis erforderlich ist. In mehr als der Hälfte der Beweisversuche von Landesrundenteilnehmenden aus den Klassenstufen 8 und 9 ist die Schlussweise deduktiv ohne zusätzliches Beispiel. In allen diesen Beweisversuchen kommen Variablen vor, so dass die Darstellungsform symbolischen Anteil hat. Es gibt auch Beweisversuche von Landesrundenteilnehmenden in der Klassenstufe 8, in denen die Schlussweise zwischen induktiv und deduktiv liegt. Teils kommen auch in diesen Beweisversuchen Variablen vor. Die Ergebnisse deuten also an, dass Teilnehmende der Landesrunde der Mathematik-Olympiade aus den Klassenstufen 8 und 9 zu einer solchen Allaussage mit unendlichem Gültigkeitsbereich zumindest ein weitgehend korrektes und vollständiges Argumentgefüge bei begründendem Ansatz entwickeln können, und dass ein großer Anteil dieser Lernenden dabei Variablen verwendet.

Zugleich ist festzuhalten, dass sich in beiden Klassenstufen 8, 9 deutlich mehr als die Hälfte der Teilnehmenden, die einen **Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge** formuliert haben, nicht für die Landesrunde qualifiziert haben. Dies ist plausibel, weil die Bewertung des im Rahmen dieser Arbeit analysierten Beweisversuchs in jeder der Klassenstufen maximal ein Zehntel der bei der Regionalrunde erreichbaren Punktzahl ausmacht. Zudem liegt der Anteil der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge in den Klassenstufen 8, 9 über 20%, während sich in NRW normalerweise je Klassenstufe nicht mehr als 10% der Teilnehmenden für die Landesrunde qualifizieren.

Klassenstufen 5 bis 7

Mit Blick auf die Beweisversuche der Landesrundenteilnehmenden in den Klassenstufen 5 bis 7 fällt auf, dass es sich in jeder der Klassenstufen in über 80% der Fälle um Beweisversuche mit **begründendem Ansatz** handelt. Allerdings enthalten diese Beweisversuche teils kein weiterentwickelbares Argumentgefüge und vereinzelt ist ein Hinweis auf einen grundlegenden Fehler zu erkennen. Außerdem haben sich in jeder der Klassenstufen 5 bis 7 auch Schülerinnen und Schüler für die Landesrunde qualifiziert, die weder einen Beweisversuch mit begründendem Ansatz noch einen Beweisversuch mit prüfendem Ansatz und einem weiterentwickelbaren Argumentgefüge formuliert haben. Mit Blick auf Anknüpfungspunkte für eine Förderung

lässt sich also die Tendenz festhalten, dass in den Klassenstufen 5 bis 7 auch Landesrundenteilnehmenden nicht unbedingt ein weitgehend korrektes Argumentgefüge zu der fokussierten Allaussage gelingt.

In den Klassenstufen 5 bis 7 haben sich von den Teilnehmenden, die einen Beweisversuch mit begründendem Ansatz und einem **weitgehend korrekten und vollständigen Argumentgefüge** formuliert haben, jeweils einzelne, aber weniger als ein Fünftel für die Landesrunde qualifiziert. Von den Teilnehmenden aus den Klassenstufe 6 und 7, die einen Beweisversuche mit begründendem Ansatz und einer **Darstellungsform mit symbolischem Anteil** formuliert haben, hat in etwa die Hälfte die Qualifikation für die Landesrunde geschafft. Die Existenz eines weitgehend korrekten und vollständigen Argumentgefüges bei begründendem Ansatz oder einer Darstellungsform mit symbolischem Anteil impliziert nicht, dass sich die Schülerin bzw. der Schüler für die Landesrunde qualifiziert hat.

Was die **primäre Beweisidee** betrifft, ist bezogen auf die Klassenstufen 5 bis 7 Folgendes zu sehen: Der Anteil der Beweisversuche mit einer ausbaubaren primären Beweisidee ist bei den Landesrundenteilnehmenden mit 61% statistisch hoch signifikant größer als bei den anderen Schülerinnen und Schülern mit 28% ($\mathrm{p} < .001^{\mathrm{N\text{-}1}}$, $\mathrm{V} = .155$). Die Beweisversuche der Landesrundenteilnehmenden basieren zu knapp 30% auf der primären Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare*. Der Anteil der Beweisversuche mit dieser primären Beweisidee ist bei den Landesrundenteilnehmenden statistisch signifikant größer als bei den anderen Schülerinnen und Schülern ($\mathrm{p} = .010^{\mathrm{ex}}$, $\mathrm{V} = .155$). Der Anteil der Beweisversuche basierend auf der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* hingegen ist bei den Landesrundenteilnehmenden und den Nicht-Landesrundenteilnehmenden ähnlich. Gleiches gilt bezogen auf jede der primären Beweisideen, außer der Beweisidee 6.

Bei den Beweisversuchen mit der primären Beweisidee 6 ist der Anteil der Landesrundenteilnehmenden aber nicht auffallend groß. Auch haben sich von den Teilnehmenden, die einen Beweisversuch mit der primären Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* oder der primären Beweisidee 5 *Reste bei Division einzelner Summanden durch vier oder acht* formuliert haben, weniger als ein Viertel für die Landesrunde qualifiziert. Abgesehen vom Gesamtanteil an Beweisversuchen mit einer ausbaubaren primären Beweisidee scheinen in den Klassenstufen 5 bis 7 keine klaren Zusammenhänge zwischen der primären Beweisidee und der (Nicht-)Teilnahme an der Landesrunde zu bestehen.

Die **Ergebnisse** zu Beweisversuchen von Landesrundenteilnehmenden aus den Klassenstufen 5 bis 7 überraschen **vor dem Hintergrund der Konzeption und Durchführungspraxis der Mathematik-Olympiade in NRW** nicht allzu sehr. Wie bereits beschrieben, hat die Bearbeitung des in dieser Arbeit fokussierten Aufgabenteils im Wettbewerb maximal ein Zehntel der erreichbaren Punktzahl ausgemacht. Zudem deuten die in den Abschnitten 2.3.2, 10.3 beschriebenen Einblicke in die regulären Anforderungen bei der Mathematik-Olympiade an, dass der im Rahmen dieser Arbeit fokussierte Aufgabenteil im Vergleich zu den anderen Aufgabenteilen nicht der leichteste war. Es kann sein, dass der im Rahmen dieser Arbeit fokussierte Aufgabenteil nicht optimal bearbeitet wurde, aber der Rest der Klausur sehr gut, so dass die Qualifikation für die

Landesrunde trotzdem gelungen ist. Des Weiteren ist die Frage, wie viele Punkte für einen Beweisversuch vergeben wurden, ein Stück weit von den verantwortlichen Lehrkräften in den einzelnen Regionen abhängig. Es ist prinzipiell möglich, dass ein Beweisversuch mit begründendem Ansatz und einem Ansatz für ein weiterentwickelbares Argumentgefüge hoch bepunktet wurde und für einen Beweisversuch mit einem weitgehend korrekten und vollständigen Argumentgefüge bei zwischen induktiver und deduktiver Schlussweise wenige Punkte vergeben wurden. Schließlich werden in einigen Regionen in NRW unabhängig von der Klassenstufe die Teilnehmenden mit der höchsten Gesamtpunktzahl zur Landesrunde geschickt, so dass es z. B. vorkommen kann, dass sich niemand aus der Klassenstufe 7 für die Landesrunde qualifiziert. In anderen Regionen werden die besten einer jeden Klassenstufe zur Landesrunde geschickt, so dass sich eventuell auch Teilnehmende, die deutlich weniger als die volle Punktzahl erzielen, für die Landesrunde qualifizieren. Insgesamt spielen neben der konkreten Bearbeitung des in dieser Arbeit fokussierten Aufgabenteils mit Blick auf die Frage, ob sich jemand für die Landesrunde qualifiziert, mehrere weitere Faktoren eine Rolle, so dass die erhaltenen Ergebnisse plausibel sind.

10.4.2.2 Mathematik-Olympiade: Wiederholte Teilnahme

In diesem Abschnitt stehen die Beweisversuche von Teilnehmenden, die auch bei der Regionalrunde der Mathematik-Olympiade im vorhergehenden Schuljahr dabei waren, im Fokus. Im Folgenden werden diese Teilnehmenden vereinfachend als „Mehrfachteilnehmende“ und alle anderen als „Erstteilnehmende“ bezeichnet. Der vorliegende Abschnitt gibt Einblicke bezüglich der Frage, inwieweit sich die Beweisversuche von Mehrfachteilnehmenden durch besondere Eigenschaften auszeichnen und inwieweit Beweisversuche mit bestimmten Eigenschaften sehr häufig von Mehrfachteilnehmenden stammen. Diese Einblicke basieren auf den Daten aus der Zufallsstichprobe, die im Rahmen dieser Arbeit gezogen wurde. Bezogen auf die Zufallsstichprobe beträgt der Anteil der Mehrfachteilnehmenden an allen Teilnehmenden in der Klassenstufe 5 rund 20%, in der Klassenstufe 6 rund 48%, in den Klassenstufen 7 und 8 ca. 55% und in der Klassenstufe 9 rund 65%.[170]

In jeder der Klassenstufen 5 bis 9 ist bei den Mehrfachteilnehmenden der Anteil der **Beweisversuche mit begründendem Ansatz** größer als bei den Erstteilnehmenden. Dieser Unterschied zwischen Erst- und Mehrfachteilnehmenden ist bezogen auf die Klassenstufen 5, 7 statistisch signifikant (jeweils: $p \leq .020^{N-1}$; $0{,}13 < V < 0{,}18$). In der Klassenstufe 5 beträgt der Anteil der Beweisversuche mit begründendem Ansatz bei den Erstteilnehmenden 44% und bei den Mehrfachteilnehmenden 61%. In der Klassenstufe 7 sind es 65% und 81%. In der Klassenstufe 8 ist der Anteil der Beweisversuche mit einer der Schlussweisen $Sw2\text{-}4$ *deduktiv oder deduktiv mit induktivem Anteil* bei den Mehrfachteilnehmenden mit 67% im Vergleich zu den Erstteilnehmenden mit 47% statistisch signifikant größer ($p = .049^{N-1}$, $V = .195$).

[170] Bei diesen Angaben und den nachfolgenden Ausführungen sind ausschließlich die Lernenden berücksichtigt, bei denen bezüglich der Teilnahme an der Regionalrunde im vorhergehenden Schuljahr Informationen vorliegen. Dies ist in der Klassenstufe 5 bei rund 85%, in den Klassenstufen 6 bis 9 bei jeweils über 95% der Lernenden, deren Beweisversuche als Teil der Zufallsstichprobe analysiert wurden, der Fall. Die folgenden Ausführungen fokussieren auf die Beweisversuche, bei denen bekannt ist, ob es sich um den Beweisversuch eines Mehrfachteilnehmenden handelt.

Bezogen auf jede der **Klassenstufen 5 bis 7** ist auch der Anteil der Beweisversuche mit einer ausbaubaren primären Beweisidee bei den Mehrfachteilnehmenden statistisch signifikant größer als bei den Erstteilnehmenden (jeweils $p \leq 0{,}031$, $0{,}14 < V < 0{,}17$). Dieser Unterschied ist allerdings in Zusammenhang mit den unterschiedlichen Anteilen an Beweisversuchen mit begründendem Ansatz zu sehen. Fokussiert auf die Beweisversuche mit begründendem Ansatz ist in keiner der Klassenstufen 5 bis 7 ein statistisch signifikanter Unterschied zu erkennen.

In den **Klassenstufen 8 und 9** ist zu sehen, dass fast alle Beweisversuche mit begründendem Ansatz und einem korrekten und vollständigen Argumentgefüge von Mehrfachteilnehmenden stammen. In der Klassenstufe 9 beträgt der Anteil der Beweisversuche mit einem korrekten und vollständigen Argumentgefüge an allen Beweisversuchen mit begründendem Ansatz bei der Mehrfachteilnehmenden 20%, bei den Erstteilnehmenden 5%. Der Anteil der Beweisversuche ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge an allen Beweisversuchen mit begründendem Ansatz ist in der Klassenstufe 9 bei den Mehrfachteilnehmenden mit 4% im Vergleich zu 18% bei den Erstteilnehmenden kleiner. Auch ist der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil an allen Beweisversuchen mit begründendem Ansatz in der Klassenstufe 9 bei den Mehrfachteilnehmenden mit 58% im Vergleich zu den Erstteilnehmenden mit 36% größer. Der Anteil der Beweisversuche mit der Darstellungsform *verbal* an allen Beweisversuchen mit begründendem Ansatz ist in der Klassenstufe 9 bei den Mehrfachteilnehmenden mit 9% im Vergleich zu 23% bei den Erstteilnehmenden kleiner. Diese Auffälligkeiten sind jedoch keine statistisch signifikanten Unterschiede.

Abgesehen vom Anteil an Beweisversuchen mit begründendem Ansatz und dem Anteil der Beweisversuche mit einer ausbaubaren primären Beweisidee sind in keiner der **Klassenstufen 5 bis 9** statistisch signifikante Unterschiede zwischen den Beweisversuchen von Mehrfach- und Erstteilnehmenden zu erkennen. Auch in Bezug auf die prozentualen Anteile sind die Ergebnisse neben den genannten Punkten unauffällig.

Zusammenfassend weisen die Daten darauf hin, dass sich die Beweisversuche von Mehrfachteilnehmenden nicht sehr deutlich von denen von Erstteilnehmenden unterscheiden. Insbesondere implizieren besondere Eigenschaften der Beweisversuche, wie z. B. die Existenz eines weitgehend korrekten und vollständigen Argumentgefüges, nicht, dass die Beweisversuche von Mehrfachteilnehmenden verfasst wurden.

Die unterschiedlichen Anteile an Beweisversuchen mit begründendem Ansatz sind durchaus plausibel. Gleiches würde für weitere **Unterschiede zwischen Beweisversuchen von Mehrfachteilnehmenden und von Erstteilnehmenden** gelten. Schließlich ist es denkbar, dass Mehrfachteilnehmende mit den Anforderungen bei der Mathematik-Olympiade besser vertraut sind als Erstteilnehmende und daher z. B. versuchen, einen Beweisversuch mit begründendem Ansatz zu formulieren. Auch ist es denkbar, dass vor allem Schülerinnen und Schüler erneut teilnehmen, die in der vorhergehenden Regionalrunde gut abgeschnitten haben. Zudem kann es sein, dass Lernende Dank ihres guten Abschneidens bei der Mathematik-Olympiade an mathematischen Förderangeboten teilnehmen dürfen und dadurch auf die Mathematik-Olympiade im nachfolgenden Jahr besser vorbereitet sind. Nebenbei lassen Ergebnisse von Veilande et al

(2018) Unterschiede zwischen Beweisversuchen von Mehrfachteilnehmenden und Erstteilnehmenden vermuten. Veilande et al. (2018) berichten, dass bei der 42. Open Mathematical Olympiad of Latvia die Teilnehmenden aus der Klassenstufe 9, die ebenfalls in der sechsten und in der achten Klassenstufe an diesem Wettbewerb teilgenommen haben, im Mittel eine höhere Gesamtpunktzahl erreicht haben als die anderen Teilnehmenden aus der Klassenstufe 9.

Dass im Rahmen dieser Arbeit zwischen den Beweisversuchen von Erst- und Mehrfachteilnehmenden kaum statistisch signifikanten Unterschiede zu erkennen sind, kann ein Stück weit damit zusammenhängen, dass die „Erstteilnehmenden“ eventuell nicht im vorhergehenden Schuljahr, aber etwa vor zwei oder drei Jahren bereits an der Regionalrunde der Mathematik-Olympiade teilgenommen haben. Zudem ist es möglich, dass die mittels schriftlicher Selbstauskunft erhobenen Daten bei einzelnen Teilnehmenden nicht korrekt sind und die Ergebnisse dadurch geringfügig verzerrt sind. Im Anschluss an die vorliegende Arbeit wäre es möglich, eine noch größere Stichprobe der Beweisversuche von Teilnehmenden der Mathematik-Olympiade zu analysieren, um z. B. klarer herauszuarbeiten, wie deutlich die Beweisversuche von Mehrfachteilnehmenden und Erstteilnehmenden in der Klassenstufe 9 differieren.

10.4.2.3 Teilnahme an mathematischen Förderangeboten

Dieser Abschnitt gibt Einblicke, inwieweit bestimmte Eigenschaften der Beweisversuche mit einer besonderen Förderung der Lernenden einhergehen und inwieweit sich die Beweisversuche von Teilnehmenden der Mathematik-Olympiade in Abhängigkeit ihrer Teilnahme an mathematischen Förderangeboten unterscheiden. Verschiedene mathematische Förderangebote sind in Abschnitt 2.2.3 kurz beschrieben. Entsprechend Abbildung 194 werden im Folgenden fünf Gruppen von Teilnehmenden unterschieden:

- Foerd5 *Förderung durch JuMa oder ein Mathematikstudium*
  Die Lernenden nehmen am Förderprogramm „Jugend trainiert Mathematik“ (JuMa) teil, haben innerhalb der letzten zwei Jahre an JuMa teilgenommen oder haben eine Lehrveranstaltung eines Hochschulstudiums Mathematik absolviert.
- Foerd4 *sonstige intensive oder langjährige, aktuelle Förderung*
  Die Lernenden nehmen seit mehr als zwei Schuljahren an einem Förderangebot teil oder haben innerhalb der vorhergehenden zwei Schuljahre an mehreren verschiedenen Angeboten teilgenommen.
- Foerd3 *geringe und kurzzeitige, aktuelle Förderung*
  Die Lernenden haben im laufenden oder vorhergehenden Schuljahr an einem mathematischen Förderangebot teilgenommen, jedoch nicht an mehreren Angeboten und auch nicht über deutlich mehr als zwei Schuljahre hinweg.
- Foerd2 *frühere Förderung*
  Die Lernenden haben etwa zwei bis drei Schuljahre vor der Datenerhebung an einem mathematischen Förderangebot teilgenommen, jedoch nicht im laufenden oder vorhergehenden Schuljahr.
- Foerd1 *allenfalls lange zurückliegende Förderung*
  Die Lernenden haben an keinem mathematischen Förderangebot teilgenommen oder die Teilnahme liegt mehr als drei Schuljahre zurück.

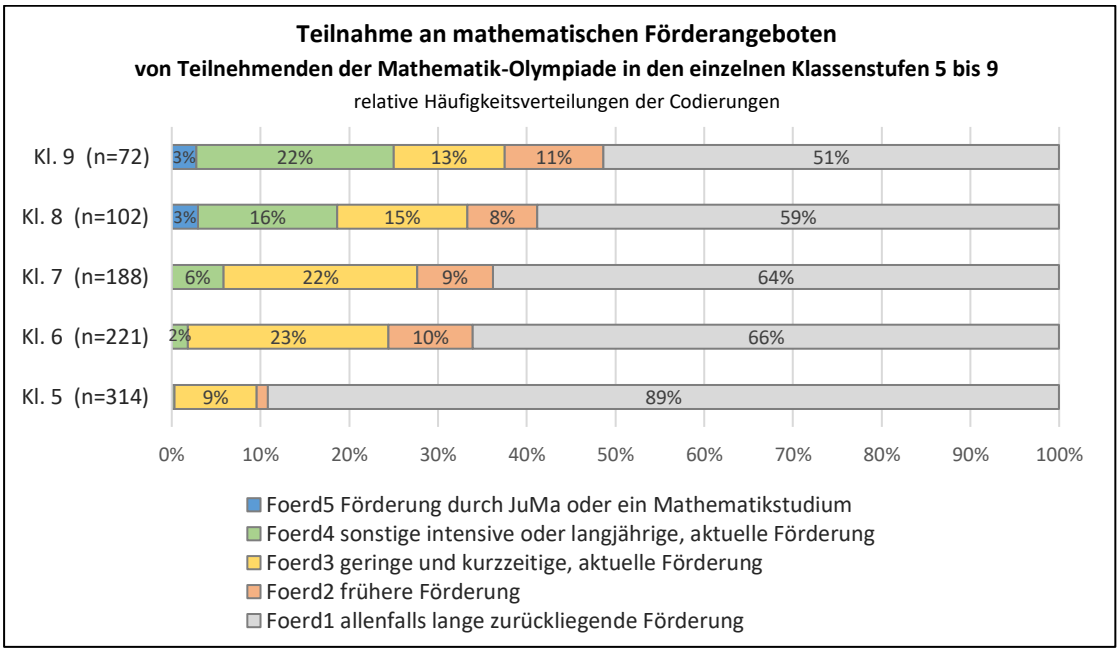


Abbildung 194: Teilnahme an mathematischen Förderangeboten von Teilnehmenden der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 9. Relative Häufigkeitsverteilungen der Codierungen

Die Abbildung 194 zeigt, ein wie großer Anteil der Teilnehmenden der Mathematik-Olympiade, deren Beweisversuche als Teil der Zufallsstichprobe analysiert wurden, an mathematischen Förderangeboten teilgenommen haben.[171] Darüber hinaus wurden Schülerinnen und Schüler mit einer unterschiedlichen Teilnahme an mathematischen Förderangeboten in der bewusst gezogenen Stichprobe systematisch berücksichtigt (Abschn. 7.3.3). Auf dieser Datengrundlage werden im Folgenden Eigenschaften der Beweisversuche von Teilnehmenden der Mathematik-Olympiade in Zusammenhang mit deren Teilnahme an mathematischen Förderangeboten beschrieben.

Klassenstufen 5 bis 7

In den Klassenstufen 5 bis 7 sind jeweils fast keine statistisch signifikanten Unterschiede zwischen **Beweisversuchen von Teilnehmenden, die zu verschiedenen der Gruppen Foerd1-5 gehören**, zu erkennen. Was die **Klassenstufe 5** betrifft, ist kein einziger statistisch signifikanter Unterschied zu sehen. In der Zufallsstichprobe aus der Klassenstufe 5 kommt ein Schüler vor, der bezüglich der Teilnahme an mathematischen Förderangeboten in die Gruppe Foerd4 *sonstige intensive oder langjährige, aktuelle Teilnahme* gehört. Es fällt auf, dass dieser Schüler basierend auf der primären Beweisidee 6 *Teilbarkeit zweier Summen durch vier oder acht* einen Beweisversuch mit begründendem Ansatz, deduktiver Schlussweise ohne zusätzliches Beispiel und einem weitgehend korrekten und vollständigen Argumentgefüge, d. h. einen der besten Beweisversuche aus der Klassenstufe 5, formuliert hat. In der bewusst gezogenen Stichprobe sind zwei weitere Teilnehmende berücksichtigt, die zur Gruppe Foerd4 gehören. Einer von

[171] Die Angaben basieren auf der schriftlichen Selbstauskunft der Teilnehmenden. Je Klassenstufe liegen von etwa 97% der Teilnehmenden, deren Beweisversuche als Teil der Zufallsstichprobe analysiert wurden, Angabe bezüglich der Teilnahme an Förderangeboten vor. In Abbildung 194und im folgenden Fließtext sind ausschließlich die Teilnehmenden, von denen eine solche Angabe vorliegt, berücksichtigt.

ihnen hat einen Beweisversuch mit prüfendem Ansatz ohne weiterentwickelbares Argumentgefüge abgegeben. Der andere hat einen Beweisversuch mit begründendem Ansatz und einem weiterentwickelbaren Argumentgefüge mit inhaltlichem oder grundlegendem Fehler abgegeben. In der **Klassenstufe 6** ist als einziger statistisch signifikanter Unterschied zu sehen, dass bei den Teilnehmenden mit einer aktuellen Förderung (Foerd3-5) der Anteil der Beweisversuche mit prüfendem Ansatz mit 7% im Vergleich zu 21% bei den anderen Teilnehmenden (Foerd1-2) statistisch hoch signifikant kleiner ist ($p = .006^{N-1}$, $V = .187$). In der **Klassenstufe 7** ist als einziger statistisch signifikanter Unterschied zu erkennen, dass die Beweisversuche mit begründendem Ansatz von Teilnehmenden in der Gruppe Foerd4 *sonstige intensive oder langfristige, aktuelle Förderung* mit 38% im Vergleich zu 9% bei denen in der Gruppe Foerd1 *allenfalls lange zurückliegende Förderung* zu einem statistisch signifikant größeren Anteil ein weitgehend korrektes und vollständiges Argumentgefüge enthalten ($p = .049^{ex}$, $V = .245$). Zugleich ist festzuhalten, dass in der Klassenstufe 7 manche Teilnehmende, die zu der Gruppe Foerd4 gehören, einen Beweisversuch mit prüfendem Ansatz oder einen Beweisversuch mit begründendem Ansatz ohne weiterentwickelbares Argumentgefüge formuliert haben. Insgesamt deuten die Ergebnisse darauf hin, dass in den **Klassenstufen 5 bis 7** die Beweisversuche nicht deutlich zwischen Teilnehmenden mit einer unterschiedlichen mathematischen Förderung differieren, auch nicht bezogen auf die primäre Beweisidee oder die Darstellungsform. In allen Klassenstufen 5 bis 7 sind die Beweisversuche zumindest mit Blick auf jede der Gruppen Foer1-3 vielfältig.

Andersherum, d. h. von den Beweisversuchen ausgehend, betrachtet, ist festzuhalten: Sämtliche Eigenschaften der Beweisversuche, die in den einzelnen Klassenstufen 5 bis 7 vorkommen, können von Lernenden mit einer aktuellen Förderung (Foerd3-5), mit einer früheren Förderung (Foerd2) oder mit allenfalls lange zurückliegender Förderung (Foerd1) sein. Beispielsweise stammen die Beweisversuche mit begründendem Ansatz in jeder der Klassenstufen 5 bis 7 mehrheitlich von Teilnehmenden mit allenfalls einer lange zurückliegenden Förderung, zu mindestens 10% von Teilnehmenden mit einer früheren Förderung und zu mindestens 10% von Teilnehmenden mit einer aktuellen Förderung. Auch die Beweisversuche mit einer Darstellungsform mit symbolischem Anteil sind nicht unbedingt von Schülerinnen und Schülern, die an einem mathematischen Förderangebot teilnehmen oder teilgenommen haben. Beispielsweise stammt in der Klassenstufe 6 die Hälfte der Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge und einer Darstellungsform mit symbolischem Anteil von Lernenden in der Gruppe Foerd1 *allenfalls lange zurückliegende Förderung*.

Klassenstufen 8 und 9

In den Klassenstufen 8, 9 sind insbesondere die Beweisversuche von Teilnehmenden in der Gruppe **Foerd5** ***Förderung durch JuMa oder ein Mathematikstudium*** auffällig. Diese haben alle einen Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge, wenn auch nicht unbedingt ein korrektes und vollständiges, deduktives Argumentgefüge formuliert. In der Klassenstufe 8 ist der Anteil der Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge an den Beweisversuchen mit begründendem Ansatz bei den Teilnehmenden in der Gruppe Foerd5 *Förderung durch*

JuMa oder ein Mathematikstudium mit 100% statistisch signifikant größer als bei den anderen Teilnehmenden mit 27% ($p = .025^{ex}$, $V = .302$).

Ansonsten sind in der **Klassenstufe 8** kaum statistisch signifikante Unterschiede zwischen den Beweisversuchen der Teilnehmenden in den verschiedenen Gruppen $Foerd1\text{-}5$ zu sehen. Ähnlich wie in den Klassenstufen 5 bis 7 weisen die Daten darauf hin, dass die Beweisversuche der Teilnehmenden in jeder der Gruppen $Foerd1\text{-}4$ vielfältig sind. Sämtliche Ausprägungen der Beweisversuche, die in der Klassenstufe 8 vorkommen, können tendenziell von Lernenden mit einer aktuellen Förderung, einer früheren Förderung oder mit allenfalls lange zurückliegender Förderung stammen.

In der **Klassenstufe 9** haben fast alle Teilnehmende mit einer früheren oder aktuellen Förderung ($Foerd2\text{-}5$) einen Beweisversuch mit begründendem Ansatz verfasst. Andersherum betrachtet stammen, mit Ausnahme einer fehlenden Bearbeitung, alle analysierten Beweisversuche ohne begründenden Ansatz von Teilnehmenden mit allenfalls lange zurückliegender Förderung ($Foerd1$).

Zudem sind in der Klassenstufe 9 mehrere Unterschiede zwischen Beweisversuchen von Teilnehmenden mit einer aktuellen Förderung ($Foerd3\text{-}5$) und solchen ohne aktuelle Förderung ($Foerd1\text{-}2$) zu sehen. Der Anteil der Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge an allen Beweisversuchen mit begründendem Ansatz ist bei den Teilnehmenden mit einer aktuellen Förderung mit 69% im Vergleich zu 27% bei denen ohne aktuelle Förderung statistisch hoch signifikant größer ($p = .001^{N\text{-}1}$, $V = .417$). Der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil an allen Beweisversuchen ist bei den Teilnehmenden mit einer aktuellen Förderung mit 62% im Vergleich zu 38% bei denen ohne aktuelle Förderung statistisch hoch signifikant größer ($p < .001^{N\text{-}1}$, $V = .474$). Der Anteil der Beweisversuche mit einer ausbaubaren primären Beweisidee ist bei den Teilnehmenden mit einer aktuellen Förderung mit 96% im Vergleich zu 73% bei denen ohne aktuelle Förderung statistisch signifikant größer ($p = .024^{ex}$, $V = .289$). Insbesondere ist der Anteil der Beweisversuche mit einer der primären Beweisideen 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* und 8 *Darstellung ungerader in Abhängigkeit gerader Zahlen* bei den Teilnehmenden mit einer aktuellen Förderung mit 67% im Vergleich zu 29% bei denen ohne aktuelle Förderung statistisch hoch signifikant größer ($p = .002^{N\text{-}1}$, $V = .369$). Dabei ist davon auszugehen, dass die genannten Unterschiede miteinander zusammenhängen (Abschn. 10.4.1.2-10.4.1.3). Außerdem implizieren die vorhergehenden Ausführungen zum Beispiel, dass die Beweisversuche mit einer Darstellungsform mit symbolischem Anteil zu einem statistisch hoch signifikant größeren Anteil von Teilnehmenden mit einer aktuellen Förderung stammen als die mit einer anderen Darstellungsform.

Zusammenfassung und Diskussion

Bei der Analyse von Beweisversuchen von Teilnehmenden der Mathematik-Olympiade zusammen mit der Teilnahme an mathematischen Förderangeboten zeigt sich, dass die Teilnehmenden in der Gruppe $Foer5$ *Förderung durch JuMa oder ein Mathematikstudium* alle einen Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen

Argumentgefüge formuliert haben. Dies erscheint erwartungsgemäß, weil sowohl im Spitzenförderprogramm Jugend trainiert Mathematik (JuMa) als auch im Mathematikstudium das Führen von Beweisen eine große Rolle einnimmt. Zudem dürfen an dem Förderprogramm JuMa in der Regel nur Schülerinnen und Schüler teilnehmen, die sich z. B. durch ein sehr gutes Abschneiden bei der Mathematik-Olympiade als mathematisch besonders leistungsstark gezeigt haben. Bei Schülerinnen und Schülern, die an einer Hochschule Mathematik studieren, lässt sich ebenfalls davon ausgehen, dass diese mathematisch leistungsstark sind.

Zugleich ist ein interessantes Ergebnis, dass in jeder der Klassenstufen 5 bis 9 auch einzelnen Schülerinnen und Schülern mit allenfalls lange zurückliegender Förderung ein Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge gelungen ist. Die Ergebnisse deuten an, dass sämtliche Ausprägungen der Beweisversuche, die in den einzelnen Klassenstufen 5 bis 9 vorkommen, von Lernenden mit einer aktuellen Förderung, mit einer früheren Förderung oder mit allenfalls lange zurückliegender Förderung sein können. Die Ergebnisse bestätigen, dass die im Rahmen dieser Arbeit fokussierte Beweisaufgabe so konstruiert ist, dass Lernende aus allen Klassenstufen und insbesondere auch solche mit allenfalls lange zurückliegender Förderung die Aufgabe prinzipiell gut bearbeiten können. Mit Blick auf Ansatzpunkte für eine Förderung im Führen von Beweisen ist, anknüpfend an die Überlegungen in Abschnitt 9.3.4, die Wichtigkeit einer deutlichen Binnendifferenzierung hervorzuheben. Auch im Falle der Arbeit mit bestimmten Teilgruppen der Teilnehmenden der Mathematik-Olympiade erscheint eine deutliche Binnendifferenzierung erforderlich, wenn man an die vorhandenen Kompetenzen der Lernenden sinnvoll anknüpfen möchte.

Ferner regen die Ergebnisse zum Nachdenken an, da u. a. auch die Beweisversuche der Teilnehmenden mit einer aktuellen Förderung vielfältig sind und, abgesehen von der Klassenstufe 9, kaum statistisch signifikante Unterschiede zu den Beweisversuchen von Teilnehmenden mit allenfalls lange zurückliegender Förderung aufweisen. Dabei ist die Aussagekraft der vorliegenden personenbezogenen Angaben, auf denen die Einteilung der Gruppen $\mathrm{Foer1\text{-}5}$ basiert, zu bedenken. Die personenbezogenen Angaben stammen aus den Fragebögen, die die Teilnehmenden im Anschluss an die Wettbewerbsklausur erhalten haben (Abschn. 6.2, 10.2). Da die Angaben mittels schriftlicher Selbstauskunft erhoben wurden, können sie vereinzelt nicht ganz stimmig sein. Zudem ist die Vielfalt der in NRW vorhandenen mathematischen Förderangebote groß. Beispielsweise ist auf Grundlage der Angabe einer erfolgten Teilnahme an einer schulinternen Mathe-AG keine genaue Aussage möglich, wie zeitintensiv die Förderung war und inwieweit Beweise thematisiert wurden. Im Anschluss an diese Arbeit wäre es dennoch möglich, Beweisversuche von Teilnehmenden der Mathematik-Olympiade noch detaillierter in Zusammenhang mit der Teilnahme an bestimmten Angeboten auszuwerten, um diesbezüglich weitere Erkenntnisse zu gewinnen.

10.4.2.4 Zusammenfassende Betrachtung: Beweisversuche, Wettbewerbserfolg und Förderung

In diesem Abschnitt werden die Beweisversuche zusammen mit allen drei zuvor einzeln betrachteten Merkmalen, d. h. zusammen mit der **Teilnahme an der Landesrunde, der Teilnahme an der Regionalrunde im vorhergehenden Schuljahr und der Teilnahme an mathematischen Förderangeboten** betrachtet. Dazu werden die Beweisversuche der Teilnehmenden der Mathematik-Olympiade in Abhängigkeit dieser Aspekte gruppiert (Abb. 195).

		Teilnahme an mathematischen Förderangeboten	
		Foerd1-2 *keine aktuelle Förderung*	Foerd3-5 *aktuelle Förderung*
Wettbewerbserfolg	LR1 *Landesrundenteilnahme*	MoFoerd31	MoFoerd32
	LR0,Mf1 *keine Landesrundenteilnahme, aber Mehrfachteilnahme*	MoFoerd21	MoFoerd22
	LR0,Mf0 *keine Landesrundenteilnahme und keine Mehrfachteilnahme*	MoFoerd11	MoFoerd12

Abbildung 195: Gruppierung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade in Abhängigkeit derer Teilnahme an der Landesrunde, derer Teilnahme an der Regionalrunde im vorhergehenden Schuljahr und derer Teilnahme an mathematischen Förderangeboten

In der Gruppierung wird bei den Landesrundenteilnehmenden nicht zusätzlich zwischen Mehrfachteilnehmenden und Erstteilnehmenden differenziert, weil viele der Landesrundenteilnehmenden zugleich Mehrfachteilnehmende sind. In der Klassenstufe 5 handelt es sich bei etwa 30% der Landesrundenteilnehmenden um Mehrfachteilnehmende, in der Klassenstufe 6 bei etwa 65%, in der Klassenstufe 7 bei 90% und in den Klassenstufen 8 und 9 bei 100%.[172]

Was die Teilnahme an mathematischen Förderangeboten betrifft, wird in der Gruppierung zwischen Lernenden mit einer aktuellen Förderung (Foerd3-5) und solchen ohne aktuelle Förderung (Foerd1-2) unterschieden. Während insbesondere in den Klassenstufen 5 bis 7 kaum Unterschiede zwischen den Beweisversuchen der Schülerinnen und Schüler in Abhängigkeit derer Teilnahme an mathematischen Förderangeboten zu erkennen sind, sind mit Blick auf die Klassenstufe 9 Unterschiede zwischen Beweisversuchen von Teilnehmenden mit einer aktuellen Förderung und solchen ohne aktuelle Förderung deutlich geworden (Abschn. 10.4.2.3).

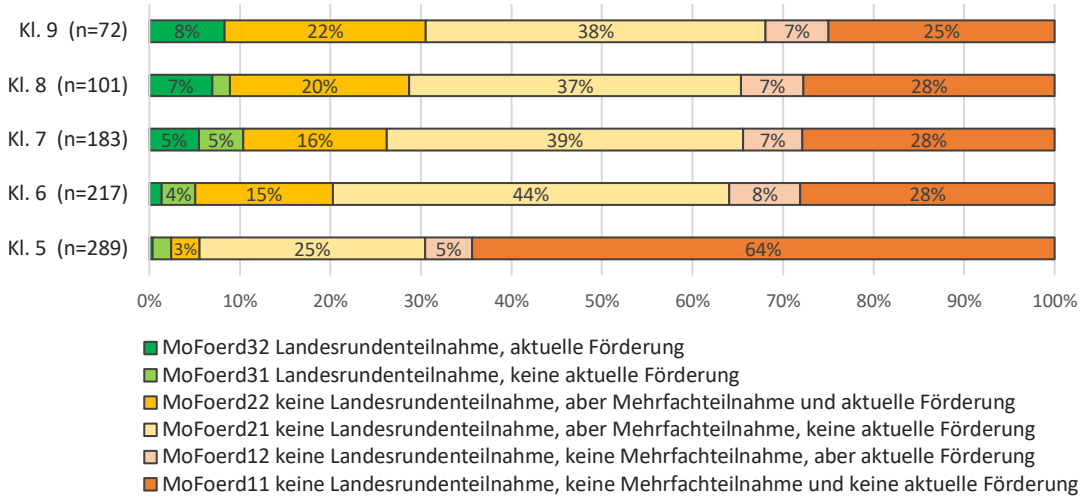
Die Abbildung 196 zeigt die relativen Häufigkeiten der gebildeten **Gruppen MoFoerd11 bis MOFoerd32** differenziert nach Klassenstufen.[173] Es fällt insbesondere auf, dass in der Klas-

[172] Die Daten deuten an, dass die Teilnahme an der Landesrunde, die Teilnahme an der Regionalrunde im vorhergehenden Schuljahr und die Teilnahme an mathematischen Förderangeboten sich gegenseitig bedingen, insbesondere mit Blick auf die Klassenstufen 8 und 9. Von den Landesrundenteilnehmenden in den Klassenstufen 8 und 9 sind jeweils mehr als drei Viertel der Schülerinnen und Schüler solche mit einer intensiven oder langjährigen, aktuellen Förderung (Foerd4-5). Andersherum betrachtet haben sich insbesondere fast alle Teilnehmenden aus den Klassenstufen 8 und 9, die beim Förderprogramm „Jugend trainiert Mathematik“ (JuMa) mitmachen oder Mathematik studieren (Foerd5), für die Landesrunde qualifiziert. Von den Teilnehmenden mit allenfalls einer früheren Förderung (Foerd1-2) haben sich in jeder der Klassenstufen 5 bis 9 weit unter 10% für die Landesrunde qualifiziert. Weiterhin ist in den Klassenstufen 7 und 8 der Anteil derjenigen mit einer aktuellen Förderung bei den Mehrfachteilnehmenden statistisch signifikant größer als bei den Erstteilnehmenden (jeweils: $p < 0{,}045$; $0{,}16 < \mathrm{V} < 0{,}21$).

[173] Die Angaben in Abbildung 196 beziehen sich genau auf die Beweisversuche der Zufallsstichprobe (Abschn. 7.3.3, 10.4.1), die sich eindeutig einer der Gruppen zuordnen lassen. In der Klassenstufe 5 ist dies bei 88% der Beweisversuche aus der Zufallsstichprobe der Fall, in den Klassenstufen 6 bis 9 jeweils bei mindestens 94%.

Gruppierung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade in Abhängigkeit deren Teilnahme an der Landesrunde, deren Teilnahme an der Regionalrunde im vorhergehenden Schuljahr und deren Teilnahme an mathematischen Förderangeboten

relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen differenziert nach Klassenstufen



MoFoerd11 vs. ¬MoFoerd11: p-Werte von Signifikanztests und dazugehörige Effektstärken

Kl. 5 vs. Kl. 6	Kl. 5 vs. Kl. 7	Kl. 5 vs. Kl. 8	Kl. 5 vs. Kl. 9
$p < .001^{N-1}, V = .359$	$p < .001^{N-1}, V = .356$	$p < .001^{N-1}, V = .323$	$p < .001^{N-1}, V = .317$

Abbildung 196: Gruppierung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade in Abhängigkeit derer Teilnahme an der Landesrunde, derer Teilnahme an der Regionalrunde im vorhergehenden Schuljahr und derer Teilnahme an mathematischen Förderangeboten. Relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken

senstufe 5 ca. zwei Drittel der Beweisversuche zu der Gruppe MoFoer11 *keine Landesrundenteilnahme, keine Mehrfachteilnahme und keine aktuelle Förderung* gehören, in den Klassenstufen 6 bis 9 mit unter einem Drittel jeweils statistisch hoch signifikant weniger. Außerdem deuten die Daten an, dass es sich in den Klassenstufen 6 bis 9 häufig um Lernende handelt, die bereits im vorhergehenden Schuljahr bei der Regionalrunde dabei waren, aber an keinem Förderangebot teilnehmen und sich nicht für die Landesrunde qualifizieren (MoFoerd21).

In Abschnitt 10.4.2.1 ist bereits deutlich geworden, dass in jeder der Klassenstufen 5 bis 9 Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge sowohl von Landesrundenteilnehmenden als auch von Nicht-Landesrundenteilnehmenden stammen können. Entsprechend Abschnitt 10.4.2.2 implizieren besondere Eigenschaften der Beweisversuche, wie z. B. die Existenz eines weitgehend korrekten und vollständigen Argumentgefüges oder einer Darstellungsform mit symbolischem Anteil nicht, dass die Beweisversuche von Mehrfachteilnehmenden verfasst wurden. In Abschnitt 10.4.2.3 hat sich gezeigt, dass in den Klassenstufen 5 bis 8 sämtliche Eigenschaften der Beweisversuche, die in diesen Klassenstufen vorkommen, von Teilnehmenden mit einer aktuellen Förderung, mit einer früheren Förderung oder mit allenfalls lange zurückliegender Förderung sein können. Daher wird im Folgenden nicht ausgehend von den Eigenschaften der Beweisversuche geschaut, zu welchen der Gruppen MoFoerd11 bis MoFoerd32 die Beweisversuche gehören. Primär werden andersherum **Eigenschaften der Beweisversuche in Abhängigkeit ihrer Zugehörigkeit zu bestimmten Gruppen** thematisiert.

In den **Klassenstufen 5 bis 7** ist das Spektrum an Beweisversuchen bezogen auf jede der Gruppen $\mathrm{MoFoerd11}$ bis $\mathrm{MoFoerd32}$ groß. Beispielsweise kommen in fast jeder der Gruppen fehlende Beweisversuche oder Beweisversuche mit weder begründendem noch prüfendem Ansatz, Beweisversuche mit prüfendem Ansatz sowie Beweisversuche mit begründendem Ansatz mit und ohne weitgehend korrektes Argumentgefüge vor. Zudem kommen bezogen auf fast jede der Gruppen sehr unterschiedliche primären Beweisideen vor. Unterschiede zwischen den Beweisversuchen aus den einzelnen Gruppen $\mathrm{MoFoerd11}$ bis $\mathrm{MoFoerd32}$ sind in den Klassenstufen 5 bis 7 kaum zu erkennen. In der Klassenstufe 5 fällt im Wesentlichen auf, dass der Anteil der Beweisversuche mit begründendem Ansatz bei den Landesrundenteilnehmenden über 80% liegt, bei den anderen Teilnehmenden mit unter 50% statistisch signifikant kleiner ist ($\mathrm{p} = .010^{\mathrm{N\text{-}1}}$, $\mathrm{V} = .152$). In der Klassenstufe 6 fällt auf, dass der Anteil der Beweisversuche mit prüfendem Ansatz in der Gruppe $\mathrm{MoFoerd11}$ *keine Landesrundenteilnahme, keine Mehrfachteilnahme und keine aktuelle Förderung* mit 30% am größten ist und statistisch hoch signifikant größer als bei den anderen Teilnehmenden mit 12% ($\mathrm{p} = .001^{\mathrm{N\text{-}1}}$, $\mathrm{V} = .217$). In der Klassenstufe 7 ist der Anteil der Beweisversuche mit prüfendem Ansatz in beiden Gruppen $\mathrm{MoFoerd11{,}12}$ eher groß. Die Beweisversuche der Erstteilnehmenden, die sich nicht für die Landesrunde qualifiziert haben ($\mathrm{MoFoerd11{,}12}$) sind im Vergleich zu den anderen Beweisversuchen ($\mathrm{MoFoerd21\text{-}32}$) statistisch signifikant häufiger Beweisversuche mit prüfendem Ansatz und statistisch signifikant seltener Beweisversuche mit begründendem Ansatz (jeweils: $p \leq 0{,}015$; $0{,}19 < V < 0{,}29$).

In der **Klassenstufe 8** sind alle Beweisversuche von Landesrundenteilnehmenden, egal ob mit oder ohne aktuelle Förderung ($\mathrm{MoFoerd31{,}32}$), Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge. In ca. der Hälfte der Fälle hat die Darstellungsform symbolischen Anteil. Außerdem ist der Anteil der Beweisversuche mit begründendem Ansatz in der Gruppe $\mathrm{MoFoerd11}$ mit 64% am kleinsten, statistisch signifikant kleiner als bezogen auf die anderen Beweisversuche ($\mathrm{p} = .037^{\mathrm{N\text{-}1}}$, $\mathrm{V} = .209$). Dabei sind zugleich in jeder der Gruppen $\mathrm{MoFoerd11{,}12{,}21{,}22}$ beispielsweise Beweisversuche mit prüfendem Ansatz, Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge sowie vielfältige primäre Beweisideen zu finden.

In der **Klassenstufe 9** handelt es sich bei allen Beweisversuchen von Landesrundenteilnehmenden ($\mathrm{MoFoerd31{,}32}$) um Beweisversuche mit einem weitgehend korrekten und vollständigen Argumentgefüge bei deduktiver Schlussweise ohne zusätzliches Beispiel. Fast immer ist ein Beweis formuliert. Die Darstellungsform hat in allen Beweisversuchen von Landesrundenteilnehmenden symbolischen Anteil und die primäre Beweisidee ist in fast allen Fällen eine der Ideen 2 *Darstellung einer Summe in Abhängigkeit der kleinsten Summanden* und 8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*. Zudem ist zu sehen, dass alle Beweisversuche ohne begründenden Ansatz von Teilnehmenden ohne aktuelle Förderung stammen ($\mathrm{MoFoerd11{,}21}$) und es sich fast immer um Erstteilnehmende handelt ($\mathrm{MoFoerd11}$). Nebenbei ist in der Gruppe $\mathrm{MoFoerd11}$ der Anteil der Darstellungsformen mit symbolischem Anteil am kleinsten und statistisch signifikant kleiner als bei den anderen Beweisversuchen ($\mathrm{p} = .015^{\mathrm{N\text{-}1}}$, $\mathrm{V} = .289$). Der Anteil der Darstellungsform *verbal* ist in der Gruppe $\mathrm{MoFoerd11}$ am größten und statistisch signifikant größer als bei den anderen Beweisversuchen ($\mathrm{p} = .038^{\mathrm{ex}}$,

$V = .267$). Zugleich ist zu sehen, dass in jeder der Gruppen mit Nicht-Landesrundenteilnehmenden ($\mathrm{MoFoerd}11,12,21,22$) sowohl hinsichtlich der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges als auch hinsichtlich der primären Beweisidee, der Beispielnutzung und der Darstellungsform vielfältige Beweisversuche mit begründem Ansatz vorkommen.

Über alle **Klassenstufen 5 bis 9** hinweg lässt sich festhalten, dass am ehesten die Beweisversuche in den Gruppen $\mathrm{MoFoerd}11$ und $\mathrm{MoFoerd}31,32$ sowie mit Einschränkungen auch die Beweisversuche in der Gruppe $\mathrm{MoFoerd}12$ besondere Eigenschaften aufweisen, während vor allem die Beweisversuche in den Gruppen $\mathrm{MoFoerd}21,22$ vielfältig sind.

Was die **Verallgemeinerbarkeit der Ergebnisse dieser Arbeit auf Gymnasiastinnen und Gymnasiasten** betrifft, sind vermutlich die Ergebnisse zu Beweisversuchen von Teilnehmenden der Mathematik-Olympiade in der Gruppe $\mathrm{MoFoerd}11$ *keine Landesrundenteilnahme, keine Mehrfachteilnahme und keine aktuelle Förderung* am ehesten übertragbar. Die vorhergehenden Ausführungen deuten somit an, dass im Falle einer Untersuchung mit „beliebigen" Gymnasiastinnen und Gymnasiasten aus einer der Klassenstufen 5 bis 9 der Anteil der Beweisversuche mit begründendem Ansatz kleiner und der Anteil der Beweisversuche mit prüfendem Ansatz größer wäre als in der vorliegenden Studie. Außerdem weisen die vorhergehenden Ausführungen darauf hin, dass insbesondere auch in der Klassenstufe 9 der Anteil der Beweisversuche mit einer Darstellungsform mit symbolischem Anteil kleiner wäre. Diese Tendenzen passen zu Ergebnissen anderer Studien mit Schülerinnen und Schülern in den entsprechenden Klassenstufen (Kap. 5; z. B. Healy & Hoyles, 2000; Knuth et al., 2002; Porteous, 1990).

Was die **Förderung von Teilnehmenden der Mathematik-Olympiade im Führen von Beweisen** betrifft, ist ergänzend zum Abschnitt 9.3.4 nochmals die Bedeutung einer Binnendifferenzierung hervorzuheben. Auch fokussiert auf die einzelnen Teilgruppen $\mathrm{MoFoerd}11,12,21,22$ kommen in jeder der Klassenstufen deutlich verschiedene Beweisversuche vor. In den Klassenstufen 5 bis 7 sind auch die Beweisversuche von Landesrundenteilnehmenden ($\mathrm{MoFoerd}31,32$) recht unterschiedlich. Im Spezialfall einer Förderung von Landesrundenteilnehmenden aus den Klassenstufen 8 und 9 lässt sich nach den Ergebnissen dieser Arbeit[174] davon ausgehen, dass den Schülerinnen und Schülern zu der fokussierten Allaussage ein Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge gelingt. Auch deuten die Ergebnisse an, dass viele dieser Lernenden bei einer solchen Aufgabe Algebra verwenden. Bei einer Förderung wäre die Verwendung anspruchsvollerer Aufgaben möglich und sinnvoll. Es erscheint insbesondere lohnend, die Haltung eines inneren Skeptikers zu unterstützen und, was die Klassenstufe 8 betrifft, einen Fokus auf die Formulierung deduktiver Argumentgefüge zu setzen.

[174] Es kann an dieser Stelle nicht beurteilt werden, ob sich ausschließlich Schülerinnen und Schüler, denen zu der fokussierten Allaussage ein Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge gelingt, für die Landesrunde qualifiziert hätten, wenn der Beweis der Allaussage nicht Teil der Wettbewerbsklausur gewesen wäre.

10.5 Zusammenfassung des zehnten Kapitels

Ausblick

Anknüpfend an die Diskussion der vorliegenden Studie lassen sich mehrere Untersuchungsaspekte anführen, um **Anregungen für zukünftige Forschungsarbeiten** zu geben. Die Untersuchungsaspekte beziehen sich u. a. auf die regulären Wettbewerbsanforderungen, Beweisversuche unterschiedlicher Teilgruppen von Teilnehmenden der Mathematik-Olympiade, deren Beweisversuche zu anderen Aussagen und deren Beweisprozesse.

Ein Teil dieser Untersuchungsaspekte ließe sich auf Grundlage bereits erhobener Daten bearbeiten und wurden bereits teilweise bearbeitet. Die vorliegende Arbeit beschreibt einen Ausschnitt eines **Forschungsprojekts zum Führen von Beweisen bei der Mathematik-Olympiade in Deutschland**. In diesem Projekt werden Anforderungen, die bei der Mathematik-Olympiade gestellt werden, sowie Kompetenzen von Teilnehmerinnen und Teilnehmern dieses Wettbewerbs untersucht und die Ergebnisse mit Blick auf Ansatzpunkte für eine Förderung im Führen von Beweisen interpretiert. Dabei werden die Klassenstufen 5 bis 13 und der Themenbereich der Teilbarkeit ganzer Zahlen fokussiert.

Die Analyse von Anforderungen, die bei der Mathematik-Olympiade in Deutschland ab der Klassenstufe 5 im Bereich des Führens von Beweisen gestellt werden, wurde bereits durchgeführt. Die Ergebnisse dieser **Anforderungsanalyse** beschreiben aus qualitativer und quantitativer Sicht, welcher Art und auf welchem Niveau die Anforderungen sind, und zeigen auf, welche Unterschiede diesbezüglich zwischen den Klassenstufen, Wettbewerbsrunden und innerhalb von Wettbewerbsrunden bestehen. Zudem weisen sie z. B. darauf hin, dass die in dieser Arbeit fokussierte Beweisaufgabe hinsichtlich des Gültigkeitsbereichs und der Abstraktheit der zu zeigenden Aussage im Vergleich zu den regulären Anforderungen in den Klassenstufen 5 und 6 als anspruchsvoll, im Vergleich zu denen in den Klassenstufen 7 und 8 als normal einzuordnen ist.

Zusätzlich zu den in dieser Arbeit fokussierten Beweisversuchen von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 5 bis 8 wurden im Rahmen des Forschungsprojekts auch von Teilnehmenden der Mathematik-Olympiade in den Klassenstufen 9 bis 12/13 Beweisversuche zu der in dieser Arbeit fokussierten Allaussage erhoben. Die Auswertung einer Zufallsstichprobe der **Beweisversuche von Neuntklässlerinnen und Neuntklässlern** deutet an, dass deren Beweisversuche mit Blick auf den Nachweis der Gültigkeit im Mittel besser sind als die aus den Klassenstufen 5 bis 8. Zudem fällt auf, dass Darstellungsformen mit symbolischem Anteil in der Klassenstufe 9 statistisch hoch signifikant häufiger vorkommen als in jeder der Klassenstufen 5 bis 8. Was eine Förderung von Teilnehmenden aus der Klassenstufe 9 betrifft, erscheint es lohnend, einen Fokus auf die Unterstützung einer Haltung eines inneren Skeptikers zu legen.

Auf Grundlage im Rahmen des Forschungsprojekts erhobener Daten ist es weiterhin möglich, die **Beweisversuche** zusammen mit dem **Wettbewerbserfolg der Teilnehmenden** und

deren **Teilnahme an mathematischen Förderangeboten** auszuwerten. Bezogen auf die in dieser Arbeit analysierten Beweisversuche sind unter anderem die folgenden zwei Auffälligkeiten zu erkennen: In den Klassenstufen 8 und 9 haben sich ausschließlich Teilnehmende für die Landesrunde qualifiziert, die zu der in dieser Arbeit fokussierten Allaussage einen Beweisversuch mit begründendem Ansatz und einem weitgehend korrekten und vollständigen Argumentgefüge formuliert haben. Sämtliche Ausprägungen der Beweisversuche, die in den einzelnen Klassenstufen 5 bis 9 vorkommen, können tendenziell von Lernenden mit einer aktuellen Förderung, mit einer früheren Förderung oder mit allenfalls lange zurückliegender Förderung stammen.

Im Anschluss an die vorliegende Arbeit wäre es unter anderem möglich, die Anforderungsanalyse zu publizieren, bereits erhobene Beweisversuche unterschiedlicher Teilgruppen der Teilnehmenden der Mathematik-Olympiade weitergehend zu untersuchen, Beweisversuche zu anderen Aussagen zu analysiert sowie Bearbeitungen von Aufgaben zum Validieren und Bewerten fiktiver Beweisversuche auszuwerten. Dadurch ließen sich weitere Einblicke in **Ansatzpunkte für eine Förderung** von Teilnehmenden der Mathematik-Olympiade im Führen von Beweisen generieren.

Schlusswort

„Beweise sind das Herz der Mathematik“

(Ziegler, 2008, S. 409)

Literaturverzeichnis

Abels, N. & Knipping, C. (2020). Argumentationsstrukturen in Beweisprozessen und –produkten. In H.-S. Siller, W. Weigel, J.F. Wörler (Hrsg.), *Beiträge zum Mathematikunterricht 2020*. Münster: WTM-Verlag.

Aberdein, A. (2013). The parallel structure of mathematical reasoning. In A. Aberdein & I. J. Dove (Hrsg.), *The argument of mathematics* (S. 361–380). Dordrecht: Springer.

Alsina, C. & Nelsen, R. B. (2013). *Bezaubernde Beweise: Eine Reise durch die Eleganz der Mathematik*. Berlin: Springer Spektrum.

Applebaum, M., Kondratieva, M. & Freiman, V. (2013). Mathematics Competitions and Gender Issues: a Case of the Virtual Marathon. *Mathematics Competitions*, *26*(1), 23–40.

Aricha-Metzer, I. & Zaslavsky, O. (2019). The nature of students' productive and non-productive example-use for proving. *Journal of Mathematical Behavior*, *53*, 304–322.

Aufgabenausschuss des Mathematik-Olympiaden e. V. (Hrsg.) (2011). *51. Mathematik-Olympiade, 1. Stufe (Schulrunde), Olympiadeklasse 7, Aufgaben*. Online: https://www.mathematik-olympiaden.de/aufgaben/51/1/A51071.pdf (Zugriff: 21.03.2018).

Aufgabenausschuss des Mathematik-Olympiaden e. V. (Hrsg.) (2015a). *55. Mathematik-Olympiade, 1. Stufe (Schulrunde), Olympiadeklasse 5, Aufgaben*. Online: https://www.mathematik-olympiaden.de/aufgaben/55/1/A55051.pdf (Zugriff: 26.11.2019).

Aufgabenausschuss des Mathematik-Olympiaden e. V. (Hrsg.) (2015b). *55. Mathematik-Olympiade, 4. Stufe (Bundesrunde), Olympiadeklasse 8, Aufgaben – 2. Tag*. Online: https://www.mathematik-olympiaden.de/aufgaben/55/4/A55084b.pdf (Zugriff: 26.11.2019).

Balacheff, N. (1988). Aspects of proof in pupils' practice of school mathematics. In D. Pimm (Hrsg.), *Mathematics teachers and children* (S. 316–230). London: Hodder & Stoughton.

Bauermann, P. (2018). Einleitung. In Mathematik-Olympiaden e. V. (Hrsg.), *Die 57. Mathematik-Olympiade 2017/2018: Aufgaben und Lösungen* (S. 4–5). Rostock: Adiant Druck.

Baumert, J., Bos, W., Klieme, E., Lehmann, R., Lehrke, M., Hosenfeld, I., Neubrand, J. & Watermann, R. (Hrsg.) (1999). *Testaufgaben zu TIMSS/III: Mathematisch-naturwissenschaftliche Grundbildung und voruniversitäre Mathematik und Physik der Abschlußklassen der Sekundarstufe II (Population 3)*. Berlin: Max-Planck-Institut für Bildungsforschung.

Bell, A. W. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics*, *7*(1-2), 23–40.

Benesch, T. (2013). *Schlüsselkonzepte zur Statistik: die wichtigsten Methoden, Verteilungen, Tests anschaulich erklärt*. Berlin/Heidelberg: Springer Spektrum.

Bender, P. & Jahnke, H. N. (1992). Intuition and rigor in mathematics instruction. *Zentralblatt für Didaktik der Mathematik*, *24*(7), 259–264.

Benölken, R. (2008). Besonderheiten mathematisch begabter Mädchen im Grundschulalter. In E. Vásárhelyi (Hrsg.), *Beiträge zum Mathematikunterricht 2008* (S. 327–330). Münster: Martin Stein Verlag.

Beutelspacher, A. (2009). *„Das ist o.B.d.A. trivial!" Tipps und Tricks zum Formulieren mathematischer Gedanken* (9. Aufl.). Wiesbaden: Vieweg & Teubner.

Beutelspacher, A. (2016). *Kleines Mathematikum: die 101 wichtigsten Fragen und Antworten zur Mathematik* (4. Aufl.). München: Beck.

Bezold, A. (2009). *Förderung von Argumentationskompetenz durch selbstdifferenzierende Lernangebote. Eine Studie im Mathematikunterricht der Grundschule*. Hamburg: Kovac.

Biehler, R. & Kempen, L. (2013). Students' use of variables and examples in their transition from generic proof to formal proof. In B. Ubuz, C. Haser & M. A. Mariotti (Hrsg.), *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education* (S. 86–95). Ankara: Middle East Technical University.

Biehler, R. & Kempen, L. (2016). Didaktisch orientierte Beweiskonzepte – Eine Analyse zur mathematikdidaktischen Ideenentwicklung. *Journal für Mathematik-Didaktik*, *37*(1), 141–179.

Bildung und Begabung gGmbH (Hrsg.) (o. J.). *Jugend trainiert Mathematik*. Online: https://www.mathe-wettbewerbe.de/juma (Zugriff: 21.03.2018).

Bildung und Begabung gGmbH (Hrsg.) (2018a). *Der Weg ins Team: Auswahlwettbewerb zur IMO*. Online: https://www.mathe-wettbewerbe.de/imo/imo-auswahlwettbewerb (Zugriff: 21.03.2018).

Bildung und Begabung gGmbH (Hrsg.) (2018b). *Mathe auf Weltniveau: Die Internationale Mathematik-Olympiade (IMO)*. Online: https://www.mathe-wettbewerbe.de/imo/imo-allgemeines (Zugriff: 21.03.2018).

Bildung und Begabung gGmbH (Hrsg.) (2018c). *Mathematik-Olympiade in Deutschland. Klassenstufen 5 bis 13*. Online: https://www.mathe-wettbewerbe.de/mo/klassenstufe-5-bis-13 (Zugriff: 21.03.2018).

Bildung und Begabung gGmbH (Hrsg.) (2023). *„Mädchen machen Mathe"*. Online: https://www.mathe-wettbewerbe.de/maedchenfoerderung (Zugriff: 14.04.2023).

Boero, P. (1999). Argumentation and mathematical proof: A complex, productive, unavoidable relationship in mathematics and mathematics education. *International Newsletter on the Teaching and Learning of Mathematical Proof*, *7*/8.

Bortz, J. & Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* (3. Aufl.). Heidelberg: Springer.

Bortz, J., Lienert, G. A. & Boehnke, K. (2008). *Verteilungsfreie Methoden in der Biostatistik* (3. Aufl.). Heidelberg: Springer.

Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin/Heidelberg: Springer.

Bourier, G. (2018). *Wahrscheinlichkeitsrechnung und schließende Statistik: Praxisorientierte Einführung – Mit Aufgaben und Lösungen* (9. Aufl.). Wiesbaden: Springer Gabler.

Branford, B. (1913). *Betrachtungen über mathematische Erziehung vom Kindergarten bis zur Universität*. Leipzig/Berlin: Teubner.

Bruder, R. (1981). Zur quantitativen Bestimmung und zum Vergleich objektiver Anforderungsstrukturen von Bestimmungsaufgaben im Mathematikunterricht. *Wissenschaftliche Zeitschrift der Pädagogischen Hochschule „Karl Liebknecht" Potsdam*, *25*(1), 173–178.

Bruder, R. (1985). Zur Bestimmung objektiver Anforderungsstrukturen von Begründungs- und Beweisaufgaben im Mathematikunterricht. *Wissenschaftliche Zeitschrift der Pädagogischen Hochschule „Karl Liebknecht" Potsdam*, *29*(1), 108–111.

Bruder, R. & Collet, C. (2011). *Problemlösen lernen im Mathematikunterricht*. Berlin: Cornelsen Scriptor.

Bruder, R. & Müller, H. (1983). Zur Entwicklung des Könnens im Lösen von Begründungs- und Beweisaufgaben. *Mathematik in der Schule*, *21*, 886–894.

Bruder, R. & Pinkernell, G. (2011). Die richtigen Argumente finden. *mathematik lehren*, *168*, 2–7.

Brunner, E. (2013). *Innermathematisches Beweisen und Argumentieren in der Sekundarstufe 1: Mögliche Erklärungen für systematische Bearbeitungsunterschiede und leistungsförderliche Aspekte*. Münster: Waxmann.

Brunner, E. (2014). *Mathematisches Argumentieren, Begründen und Beweisen: Grundlagen, Befunde und Konzepte*. Berlin/Heidelberg: Springer Spektrum.

Brunner, E., Lampart, J., & Jullier, R. (2022). Schriftliches mathematisches Argumentieren in zwei unterschiedlichen Inhaltsbereichen in den Jahrgangsstufen 4–6. *Journal für Mathematik-Didaktik*, *43*(1)

Busing, F. M. T. A., Weaver, B. & Dubois, S. (2015). 2×2 *Tables: a note on Campbell's recommendation*. Online: https://onlinelibrary.wiley.com/doi/epdf/10.1002/sim.6808 (Zugriff: 30.10.2015).

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, *24*(4), 359–387.

Chin, E. & Lin, F. (2009). A comparative study on junior high school students' proof conceptions in algebra between Taiwan and the UK. *Journal of Mathematics Education*, *2*(2), 52–67.

Coe, R. & Ruthven, K. (1994). Proof practice and constructs. *British Educational Research Journal*, *20*, 41-53.

Cohors-Fresenborg, E., Sjuts, J. & Sommer, N. (2004). Komplexität von Denkvorgängen und Formalisierung von Wissen. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA 2000* (S. 109–144). Wiesbaden: VS Verlag für Sozialwissenschaften.

Davis, P. J. (1986). The Nature of Proof. In M. Carss (Hrsg.), *Proceedings of the Fifth International Congress on Mathematical Education* (S. 352–358). Online: https://doi.org/10.1007/978-1-4757-4238-1_31 (Zugriff: 17.01.2016).

Davis, P. J. & Hersh, R. (1985). *Erfahrung Mathematik* (2. Aufl.). Basel/Bosten/Stuttgart: Birkenhäuser.

Dawkins, P. C. & Weber, K. (2017). Values and norms of proof for mathematicians and students. *Educational Studies in Mathematics*, *95*(2), 123–142.

de Corte, E. & Somers, R. (1982). Estimating the outcome of a task as a heuristic strategy in arithmetic problem solving: a teaching experiment with sixth grades. *Human Learning: A Journal of Practical Research and Applications*, 1, 105–121.

de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras*, *24*, 17–24.

Douek, N. (1998). Some remarks about argumentation and mathematical proof and their educational implications. In I. Schwank (Hrsg.), *Proceedings of the first congress of the European Society for Research in Mathematics Education* (Bd. 1, S. 125–139). Online: https://digilander.libero.it/leo723/materiali/algebra/cerme1-proceedings-1-v1-0-2.pdf#page=125 (Zugriff: 19.01.2016).

Dreyfus, T., Nardi, E. & Leikin, R. (2012). Forms of proof and proving in the classroom. In G. Hanna & M. de Villiers (Hrsg.), *Proof and Proving in Mathematics Education: The 19th ICMI Study* (S. 191–214). Heidelberg [u. a.]: Springer Science + Business Media.

Durand-Guerrier, V. (2008). Truth versus validity in mathematical proof. *Zentralblatt für Didaktik der Mathematik*, *40*(3), 373–384.

Durand-Guerrier, V., Boero, P., Douek, N., Epp, S. & Tanguay, D. (2012). Argumentation and proof in the mathematics classroom. In G. Hanna & M. de Villiers (Hrsg.), *Proof and Proving in Mathematics Education: The 19th ICMI Study* (S. 349–368). Heidelberg [u. a.]: Springer Science + Business Media.

Duval, R. (1995). *Sémiosis et pensée humaine : Registres sémiotiques et apprentissages intellectuels*. Bern: Peter Lang.

Ellis, A. B., Lockwood, E., Williams, C., Dogan, M. F. & Knuth, E. (2012). Middle school students' example use in conjecture exploration and justification. In L. R. Van Cadenst, J. J. Lo & J. L. Kratky (Hrsg.), *Proceedings of the 34th annual meeting of the North American chapter of the international group for the psychology of mathematics education* (S. 135–142). Online: https://www.researchgate.net/publication/261547913_Middle_school_students'_example_use_in_conjecture_exploration_and_justification (Zugriff: 28.11.2020).

Engel, A. (1998). *Problem-Solving Strategies*. New York [u. a.]: Springer.

Engel, W. (2011). Die Geschichte der Mathematik-Olympiaden: Meine Erinnerungen zur 50. Deutschen Mathematikolympiade an mathematische Schülerwettbewerbe und die Förderung mathematisch begabter Jugendlicher in den Jahren bis 1990. In Mathematik-Olympiaden e. V. (Hrsg.), *Festschrift: 50 Jahre Mathematik-Olympiaden 1961 – 2011* (S. 11–37). Rostock: Adiant Druck.

Epp, S. (2003). The Role of Logic in Teaching Proof. *The American Mathematical Monthly*, *110*(10), 886–889.

Ernst-Brandt, G. (2011). Die Mathematik-Olympiade in den Bundesländern: Nordrhein-Westfalen. In Mathematik-Olympiaden e. V. (Hrsg.), *Festschrift: 50 Jahre Mathematik-Olympiaden 1961 – 2011* (S. 167–169). Rostock: Adiant Druck.

Fetzer, M. (2011). Wie argumentieren Grundschulkinder im Mathematikunterricht? Eine argumentationstheoretische Perspektive. *Journal für Mathematik-Didaktik, 32*(1), 27–51.

Fischer, A., Heinze, A. & Wagner, D. (2009). Mathematiklernen in der Schule – Mathematiklernen an der Hochschule: die Schwierigkeiten von Lernenden beim Übergang ins Studium. In A. Heinze & M. Grüßing (Hrsg.), *Mathematiklernen vom Kindergarten bis zum Studium* (S. 245–264). Münster: Waxmann.

Fischer, R. & Malle, G. (1985). *Mensch und Mathematik: Eine Einführung in didaktisches Denken und Handeln*. Mannheim/Wien/Zürich: BI Wissenschaftsverlag.

Fishbein, E. (1982). Intuition and proof. *For the Learning of Mathematics, 3*(2), 9-18.

Flade, L. & Pruzina, M. (1990). Zum Beweisen im Mathematikunterricht – alte Fragen neu gestellt. *Mathematik in der Schule*, *28*(2/3), 83–94.

Frischemeier, D., Panse, A. & Pecher, T. (2016). Schwierigkeiten von Studienanfängern bei der Bearbeitung mathematischer Übungsaufgaben. In A. Hoppenbrock, R. Biehler, R. Hochmuth & H.-G. Rück (Hrsg.), *Lehren und Lernen von Mathematik in der Studieneingangsphase: Herausforderungen und Lösungsansätze* (S. 229–242). Wiesbaden: Springer.

Galbraith, P. L. (1981). Aspects of proving: A clinical investigation of process. *Educational Studies in Mathematics*, *12*(1), 1–28.

Gerloff, S. (2021). *Begründen bei Geometrieaufgaben der Grundschule: Eine Untersuchung der schulischen Anforderungen und der Kompetenzen im dritten Schuljahr.* Wiesbaden: Springer.

Girnat, B. (2009). Geometrische Hintergrundtheorien des Beweisens im Schulalltag: Auszüge aus einer qualitativen Studie über Lehreransichten. In M. Ludwig, R. Oldenburg & J. Roth (Hrsg.), *Argumentieren, Beweisen und Standards im Geometrieunterricht – AK Geometrie 2007/08* (S. 189–202). Franzbecker.

Gläser, J. & Laudel, G. (2009). *Experteninterviews und qualitative Inhaltsanalyse* (3. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Grieser, D. (2013). *Mathematisches Problemlösen und Beweisen: Eine Entdeckungsreise in die Mathematik*. Wiesbaden: Springer.

Grieser, D. (2016). Mathematisches Problemlösen und Beweisen: Ein neues Konzept in der Studieneingangsphase. In A. Hoppenbrock, R. Biehler, R. Hochmuth & H.-G. Rück (Hrsg.), *Lehren und Lernen von Mathematik in der Studieneingangsphase* (S. 661-675). Wiesbaden: Springer.

Gronau, H.-D. (2011). Die Geschichte der Mathematik-Olympiaden: Von der Wende bis 2011. In Mathematik-Olympiaden e. V. (Hrsg.), *Festschrift: 50 Jahre Mathematik-Olympiaden 1961 – 2011* (S. 38–42). Rostock: Adiant Druck.

Gronau, H.-D., Grünwald, N. & Prestin, J. (2011). Der Aufgabenausschuss. In Mathematik-Olympiaden e. V. (Hrsg.), *Festschrift: 50 Jahre Mathematik-Olympiaden 1961 – 2011* (S. 43–47). Rostock: Adiant Druck.

Grundey, S. (2010). Eigenständige Beweisaktivitäten im Mathematikunterricht – Schülervorstellungen und Kompetenzen. In A. Lindmeier & S. Ufer (Hrsg.), *Beiträge zum Mathematikunterricht 2010* (S. 361-364). Münster: WTM-Verlag.

Grundey, S. (2015). *Beweisvorstellungen und eigenständiges Beweisen: Entwicklung und vergleichende empirische Untersuchung eines Unterrichtskonzepts am Ende der Sekundarstufe*. Wiesbaden: Springer Spektrum.

Häder, M. (2015). *Empirische Sozialforschung: Eine Einführung* (3. Aufl.). Wiesbaden: Springer.

Halmos, P. R. (1970). How to write mathematics. *L'Enseignement Mathématique*, *16*(2), 123–152.

Hanna, G. (1983). *Rigorous proof in mathematics education* (Bd. 48). Toronto: OISE.

Hanna, G. (1989). More than formal proof, *For the Learning of Mathematics*, *9*(1), 20–23.

Hanna, G. & Barbeau, E. (2002). What is proof? In B. Baigrie (Hrsg.), *History of modern science and mathematics* (4. Aufl., Bd. 1, S. 36–48). New York: Charles Scribner's Sons.

Hanna, G., de Villiers, M., Arzarello, F., Dreyfus, T., Durand-Guerrier, V., Jahnke, H. N., Lin, F.-L., Selden, A., Tall, D. & Yecdokimov, O. (2012). ICMI study 19: Proof and proving in mathematics education: Discussion document. In G. Hanna & M. de Villiers (Hrsg.), *Proof and proving in mathematics education: The 19th ICMI study* (S. 443–452). Dordrecht: Springer.

Hanna, G. & Jahnke, J. N. (1996). Proof and proving. In A. J. Bishop (Hrsg.), *International handbook of mathematics education* (S. 877–908). New York: Kluwer Academic Publishers.

Hartkens, J. (2018). *Mathematische Reflexion in argumentativ geprägten Unterrichtsgesprächen: Eine empirisch-interpretative Untersuchung im 3. und 4. Grundschuljahr*. Wiesbaden: Springer Spektrum.

Healy, L. & Hoyles, C. (2000). A Study of Proof Conceptions in Algebra. *Journal for Research in Mathematics Education*, *31*(4), 396–428.

Hein, K. (2021). *Logische Strukturen beim Beweisen und ihre Verbalisierung: Eine sprachintegrative Entwicklungsforschungsstudie zum fachlichen Lernen*. Wiesbaden. Springer.

Hein, K. & Prediger, S. (2017). Fostering and investigating students' pathways to formal reasoning: A design research project on structural scaffolding for 9th graders. In T. Dooley & G. Guedet (Hrsg.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education* (S. 163-170). Dublin: DCU/ERME.

Heintz, B. (2000). *Die Innenwelt der Mathematik: Zur Kultur und Praxis einer beweisenden Disziplin*. Wien/New-York: Springer.

Heinze, A. (2010). Mathematicians' Individual Criteria for Accepting Theorems and Proofs: An Empirical Approach. In G. Hanna, H. N. Jahnke, & H. Pulte (Hrsg.), *Explanation and Proof in Mathematics: Philosophical and Educational Perspectives* (S. 101–111). Berlin/Heidelberg/New York: Springer.

Heinze, A. & Reiss, K. (2003). Reasoning and Proof: Methodological Knowledge as a Component of Proof Competence. *International Newsletter on the Teaching and Learning of Mathematical Proof*, 4–6.

Heinze, A., Kessler, S., Kuntze, S., Lindmeier, A., Moorman, M., Reiss, K., Rudoph-Albert, F. & Zöttl, L. (2007). Kann Paul besser argumentieren als Marie? Betrachtungen zur Beweiskompetenz von Mädchen und Jungen aus differentieller Perspektive: Eine Reanalyse von vier empirischen Untersuchungen. *Journal für Mathematik-Didaktik*, *28*(2), 148–167.

Heinze, A. & Reiss, K. (2004). The teaching of proof at lower secondary level – a video study. *ZDM - International Journal on Mathematics Education, 36*(3), 98-104.

Heinze, A., Reiss, K. & Rudolph, F. (2005). Mathematics achievement in mathematics from a differential perspective. *Zentralblatt für Didaktik der Mathematik*, *37*(3), 212–220.

Hemmi, K. (2006). Approaching proof in a community of mathematical practice. Dissertation. Stockholm: Stockholm University. Online: http://www.diva-portal.org/smash/get/diva2:189608/FULLTEXT01.pdf (Zugriff: 16.03.2016).

Herrmann, A. (2012). Beweisstrategien in der Linearen Algebra – eine Fallstudie zum Thema Unterraum. In M. Ludwig & M. Kleine (Hrsg.), *Beiträge zum Mathematikunterricht 2012* (S. 373–376). Münster: WTM-Verlag.

Hersh, R. (1993). Proving is Convincing and Explaining. *Educational Studies in Mathematics*, *24*(4), 389–399.

Hoyles, C. & Healy, L. (2007). Curriculum Change and Geometrical Reasoning. In P. Boero (Hrsg.), *Theorems in school: From history, epistemology and cognition to classroom practice* (S. 81–115). Rotterdam: Sense Publishers.

Huawei, Z. (2002). Abilities Tested by Mathematics Olympiads. *Mathematics Competitions*, *15*(2), 66–82.

Hussy, W. (2013). Quantitative Forschungsmethoden. In W. Hussy, M. Schreier & G. Echterhoff (Hrsg.), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (S. 115–164. Berlin/Heidelberg: Springer.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B. & Williams, C. C. (2008). Gender Similarities Characterize Math Performance. *Science*, *321*, 494–495.

Hyde, J. S., Fennema, E. & Lamon, S. J. (1990). Gender Differences in Mathematics Performance: A Meta-Analysis. *Psychological Bulletin*, *107*, 139–155.

Jahnke, H. N. (2007). Beweisen und hypothetisch-deduktives Denken. *Der Mathematikunterricht*, *53*(5), 10–21.

Jahnke, H. N. (2009). Hypothesen und ihre Konsequenzen – ein anderer Blick auf die Winkelsummensätze. *Praxis der Mathematik*, *51*(30), 26–30.

Jahnke, H. N. & Ufer, S. (2015). Argumentieren und Beweisen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 331–355). Berlin/Heidelberg: Springer.

Kanel-Belov, A. & Yavich, R. (2014). Mathematical Competitions and Olympiads – Mathematic or sport. In G. Howell, L. Sheffield & R. Leikin (Hrsg.), *The 8th Conference of MCG: Proceedings* (S. 83–89). Online: https://www.igmcg.org/conferences (Zugriff: 20.01.2016).

Kautschitsch, H. (1989). Wie kann ein Bild das Allgemeingültige vermitteln? In H. Kautschitsch, & W. Metzler (Hrsg.), *Anschauliches Beweisen* (S. 177–186). Wien/Stuttgart: B. G. Teubner.

Kempen, L. (2019). *Begründen und Beweisen im Übergang von der Schule zur Hochschule: Theoretische Begründung, Weiterentwicklung und Evaluation einer universitären Erstsemesterveranstaltung unter der Perspektive der doppelten Diskontinuität.* Wiesbaden: Springer Spektrum.

Kempen, L. & Biehler, R. (2014). The quality of argumentations of first-year pre-service teachers. In S. Oesterle, P. Liljedahl, C. Nicol, & D. Allen (Hrsg.), *Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education* (Bd. 3, S. 425–432). Vancouver: PME.

Kempen, L., Krieger, M. & Tebaartz, P. C. (2016). Über die Auswirkungen von Operatoren in Beweisaufgaben. In Institut für Mathematik und Information der Pädagogischen Hochschule Heidelberg (Hrsg.), *Beiträge zum Mathematikunterricht 2016* (Bd. 1, S. 521–524). Münster: WTM-Verlag.

Kempen, L., Tebaartz, P. C. & Krieger, M. (2019). On the effect of using different phrasings in proving tasks. In U. T. Jankvist, M. van den Heuvel-Panhuizen & M. Veldhuis (Hrsg.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education* (S. 225-232). Online: www.mathematik.tu-dortmund.de/~prediger/ERME/CERME11_Proceedings_2019.pdf (Zugriff: 17.09.2020).

Kesting, F. (2005). *Mathematisches Vorwissen zu Schuljahresbeginn bei Grundschülern der ersten drei Schuljahre – eine empirische Untersuchung*. Hildesheim [u.a.]: Franzbecker.

Kirsten, K. (2017). Identifizierung von Phasen und Aktivitäten im Beweisprozess von Studienanfänger/innen. In Institut für Mathematik der Universität Potsdam (Hrsg.), *Beiträge zum Mathematikunterricht 2017* (S. 1129–1132). Münster: WTM-Verlag.

Kirsten, K. (2018). Theoretical and Empirical Description of Phases in the Proving Processes of Undergraduates. In V. Durand-Guerrier, R. Hochmuth, S. Goodchild & N. M. Hogstad (Hrsg.), *Proceedings of the Second Conference of the International Network for Didactic Research in University Mathematics* (S. 326-335). Kristiansand: University of Agder and INDRUM.

Kirsten, K. (2021). *Beweisprozesse von Studierenden: Ergebnisse einer empirischen Untersuchung zu Prozessverläufen und phasenspezifischen Aktivitäten*. Wiesbaden: Springer Spektrum.

Klieme, E. & Leutner, D. (2006). *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Antrag an die Deutsche Forschungsgemeinschaft auf Einrichtung eines Schwerpunktprogramms*. Online: https://www.pedocs.de/volltexte/2011/4493/pdf/ZfPaed_2006_Klieme_Leutner_Kompetenzmodelle_Erfassung_Lernergebnisse_D_A.pdf (Zugriff: 19.12.2019).

Klimke, D., Lautmann, R., Stäheli, U., Weischer, C. & Wienold, H. (Hrsg.) (2020). Lexikon zur Soziologie (6. Aufl., S. 389). Wiesbaden: Springer.

Kluge, S. (1999). *Empirisch begründete Typenbildung: Zur Konstruktion von Typen und Typologien in der qualitativen Sozialforschung*. Opladen: Leske + Budrich.

Kluge, S. (2000). Empirisch begründete Typenbildung in der qualitativen Sozialforschung. *Forum Qualitative Sozialforschung*, *1*(1), 1–11.

Kelle, U. & Kluge, S. (2010). *Vom Einzelfall zum Typus: Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung* (2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.

KMK – Kultusministerkonferenz (Hrsg.) (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss: Beschluss vom 4.12.2003*. München: Luchterland.

KMK – Kultusministerkonferenz (Hrsg.) (2004). *Bildungsstandards im Fach Mathematik für den Primarbereich: Beschluss vom 15.10.2004*. München: Luchterland.

KMK – Kultusministerkonferenz (Hrsg.) (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife: Beschluss vom 18.10.2011*. München: Luchterland.

KMK – Kultusministerkonferenz (Hrsg.) (2015). *Förderstrategie für leistungsstarke Schülerinnen und Schüler*. Kultusministerkonferenz: Förderung von leistungsstarken Schülerinnen und Schülern. Online: www.kmk.org/fileadmin/pdf/350-KMK-TOP-011-Fu-Leistungsstarke_-_neu.pdf (Zugriff: 21.03.2018).

Knipping, C. (2003). *Beweisprozesse in der Unterrichtspraxis: Vergleichende Analysen von Mathematikunterricht in Deutschland und Frankreich*. Hildesheim: Franzbecker.

Knipping, C. & Reid, D. (2015). Reconstructing Argumentation Structures: A Perspective on Proving Processes in Secondary Mathematics Classroom Interactions. In A. Bikner-Ahsbahs, C. Knipping & N. Presmeg (Hrsg.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (S. 75–101). Dordrecht: Springer.

Knuth, E. J., Choppin, J., Slaughter, M. & Sutherland, J. (2002). *Mapping the Conceptual Terrain of Middle School Students' Competencies in Justifying and Proving*. Online: https://labweb.education.wisc.edu/knuth/mathproject/papers/Knuth_PME02.pdf (Zugriff: 30.08.2018).

Krummheuer, G. (2001). Narratives Argumentieren im Mathematikunterricht der Grundschule. In H.-G. Roßbach, K. Nölle / K. Czerwenka (Hrsg.), *Forschungen zu Lehr- und Lernkonzepten für die Grundschule* (S. 167–173). Opladen: Leske & Budrich.

Krummheuer, G. (2003). Argumentationsanalyse in der mathematikdidaktischen Unterrichtsforschung. *ZDM – Zentralblatt für Didaktik der Mathematik, 35*(6), 247–256.

Krumsdorf, J. (2009a). Beispielgebundenes Beweisen. In M. Neubrand (Hrsg.), *Beiträge zum Mathematikunterricht 2009* (S. 711–714). Münster: WTM-Verlag.

Krumsdorf, J. (2009b). Beweisen am Beispiel: Beispielgebundenes Beweisen zwischen induktivem Prüfen und formalem Beweisen. *Praxis der Mathematik in der Schule, 51*, 8–13.

Krumsdorf, J. (2011). Sprachliche Aspekte beispielgebundenen Beweisens. In R. Haug & L. Holzäpfel (Hrsg.), *Beiträge zum Mathematikunterricht 2011* (S. 499–502). Münster: WTM-Verlag.

Krumsdorf, J. (2017). *Beispielgebundenes Beweisen*. Münster: WTM-Verlag.

Küchemann, D. (2008). *Looking for Structure. A report of the Proof Materials Project. January 2004 – December 2005*. London: Institute of Education. Online: http://www.mathsmedicine.co.uk/ioe-proof/PMPintro2.html (Zugriff: 01.06.2019).

Küchemann, D. & Hoyles, C. (2002). Students' Understanding of a Logical Implication and its Converse. In A. D. Cockburn & E. Nardi (Hrsg.), *Proceedings of the 26th Conference of the International group for the Psychology of Mathematics Education* (Bd. 3, S. 241-248). Online: https://www.igpme.org/publications/current-proceedings/ (Zugriff: 14.12.2015).

Kuckartz, U. (2012). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Weinheim/Basel: Beltz Juventa.

Kuckartz, U., Rädiker, S., Ebert, T. & Schehl, J. (2013). *Statistik: Eine verständliche Einführung* (2. Aufl.). Wiesbaden: Springer VS.

Kugel, M. (o. J.). *Aufgaben + Lösungen*. Online: https://www.olympiade-mathematik.de/index.php?option=com_content&view=article&id=2&Itemid=102 (Zugriff: 16.11.2019).

Kuntze, S. (2004). Wissenschaftliches Denken von Schülerinnen und Schülern bei der Beurteilung gegebener Beweisbeispiele aus der Geometrie: Ergebnisse einer Untersuchung textlicher Eigenproduktionen von Schülerinnen und Schülern der 8. Jahrgangsstufe des Gymnasiums. *Journal für Mathematik-Didaktik, 25*(3/4), 245-268.

Kuntze, S. (2006). *Themenstudienarbeit: Konzeption einer Lernumgebung für den gymnasialen Mathematikunterricht und Evaluation einer Themenstudienarbeit zum mathematischen Beweisen und Argumentieren*. München: Verlag Dr. Hut.

Kuntze, S. (2009). Geometrische Beweiskompetenz fördern durch Reflexions- und Schreibanlässe zu beweisbezogenem Metawissen. In M. Ludwig, R. Oldenburg & J. Roth (Hrsg.), *Argumentieren, Beweisen und Standards im Geometrieunterricht – AK Geometrie 2007/08* (S. 219–237). Hildesheim/Berlin: Franzbecker.

Lai, Y., Weber, K. & Mejía-Ramos, J. P. (2012). Mathematicians' perspectives on features of a good pedagogical proof. *Cognition and Instruction*, *30*(2), 146–169.

Langmann, H.-H. (2016). Einleitung. In Mathematik-Olympiaden e. V. (Hrsg.), *Die 55. Mathematik-Olympiade 2015/16* (S. 4–7). Hamburg: Hereus.

Leder, G. C. & Taylor, P. J. (2010). Are Raelene, Marjorie and Betty still in the race? *The Australian Mathematics Teacher, 66*(4), 17–24.

Leiß, D. & Blum, W. (2006). Beschreibung zentraler mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung, & O. Köller (Hrsg.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen* (S. 33–50). Berlin: Cornelsen.

Leuders, T. (2010). *Erlebnis Arithmetik zum aktiven Entdecken und selbständigen Erarbeiten*. Berlin/Heidelberg: Springer.

Lin, F.-L. (2005). Modelling student's learning on mathematical proof and refutation. In H. L. Chick & J. L. Vincent (Hrsg.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 1, S. 3–18). Melbourne: PME.

Linn, M. C. & Petersen, A.C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*(6), 1479–1498.

Lynch, A. G. & Lockwood, E. (2019). A comparision between mathematicians' and students' use of examples for conjecturing and proving. *Journal of Mathematical Behavior*, *53*, 323–338.

Lockwood, E., Ellis, A. B. & Lynch, A. G. (2016). Mathematicians' example-related activity when exploring and proving conjectures. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 165–196.

Maier, H. (1999). Wieviel Fachsprache brauchen die Schüler im Mathematikunterricht? In M. Neubrand (Hrsg.), *Beiträge zum Mathematikunterricht 1999* (S. 19–26). Hildesheim/Berlin: Franzbecker.

Malle, G. (2002). Begründen. Eine vernachlässigte Tätigkeit im Mathematikunterricht. *mathematik lehren*, *110*, 4–8.

Manin, J. I. (1977). *A course in mathematical logic*. New York [u. a.]: Springer.

Mariotti, M. A. (2006). Proof and proving in mathematics education. In A. Gutierrez & P. Boero (Hrsg.), *Handbook of research on the psychology of mathematics education: Past, present and future* (S. 173–204). Rotterdam: Sense Publishers.

Marrades, R. & Gutiérrez, A. (2000). Proofs produced by secondary school students learning geometry in a dynamic computer environment. *Educational Studies in Mathematics, 44*(1/2), 87–125.

Mason, J. & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics*, *15*(3), 277–289.

Mathematik-Olympiaden e. V. (Hrsg.) (o. J.). *Mathematik-Olympiade in Deutschland*. Online https://www.mathematik-olympiaden.de/moev/index.php/olympiaden (Zugriff: 04.05.2016).

Mathematik-Olympiaden e. V. (Hrsg.) (2005). *Die 44. Mathematik-Olympiade 2006/2007: Aufgaben und Lösungen*. Hamburg: Hereus.

Mathematik-Olympiaden e. V. (Hrsg.) (2006). *Die 45. Mathematik-Olympiade 2006/2007: Aufgaben und Lösungen*. Hamburg: Hereus.

Mathematik-Olympiaden e. V. (Hrsg.) (2007). *Die 46. Mathematik-Olympiade 2006/2007: Aufgaben und Lösungen*. Hamburg: Hereus.

Mathematik-Olympiaden e. V. (Hrsg.) (2008). *Die 47. Mathematik-Olympiade 2007/2008: Aufgaben und Lösungen*. Hamburg: Hereus.

Mathematik-Olympiaden e. V. (Hrsg.) (2012a). *Die 51. Mathematik-Olympiade 2011/2012: Aufgaben und Lösungen*. Hamburg: Hereus.

Mathematik-Olympiaden e. V. (Hrsg.) (2012b). *Durchführungsbeschlüsse der Jury zum Reglement für die Bundesrunden der Mathematik-Olympiade*. Online: https://www.mathe-wettbewerbe.de/download/mo-br-jury (Zugriff: 25.03.2018).

Mathematik-Olympiaden e. V. (Hrsg.) (2012c). *Reglement für die Bundesrunde der Mathematik-Olympiaden*. Online: https://www.mathe-wettbewerbe.de/download/mo-br-reglement (Zugriff: 25.03.2018).

Mathematik-Olympiaden e. V. (Hrsg.) (2013). *Die 52. Mathematik-Olympiade 2012/2013: Aufgaben und Lösungen*. Hamburg: Hereus.

Mathematik-Olympiaden e. V. (Hrsg.) (2016). *Die 55. Mathematik-Olympiade 2015/2016: Aufgaben und Lösungen*. Hamburg: Hereus.

Mathematik-Olympiaden e. V. (Hrsg.) (2018). *Mathematik-Olympiade in Deutschland. Empfehlungen zur Durchführung*. Online: https://www.mathe-wettbewerbe.de/download/durchfuehrungsempfehlungen-5-13 (Zugriff: 21.03.2018).

Mathematik-Olympiaden e. V. (Hrsg.) (2020). *Die 59. Mathematik-Olympiade 2019/2020: Aufgaben und Lösungen*. Rostock: Adiant Druck.

Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11. Aufl.). Weinheim/Basel: Beltz.

Meyer, M. (2007). *Entdecken und Begründen im Mathematikunterricht: Von der Abduktion zum Argument*. Hildesheim: Franzbecker.

Meyer, M. & Voigt, J. (2008). Entdecken mit latenter Beweisidee: Analyse von Schulbuchseiten. *Journal für Mathematik-Didaktik*, *29*(2), 124–151.

Meyer, M. & Voigt, J. (2009). Entdecken, Prüfen und Begründen: Gestaltung von Aufgaben zur Erarbeitung mathematischer Sätze. *mathematica didactica*, *32*, 31–66.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics, 27*, 249–266.

Moya, J. A., Gutiérrez, A. & Jaime, A. (2015). Discriminating proof abilities of secondary school students with different mathematical talent. In K. Krainer & N. Vondrová (Hrsg.), *Proceedings of the 9th Congress of the European Society for Research in Mathematics Education* (S. 171-177). Online: https://www.uv.es/angel.gutierrez/archivos1/textospdf/MoyGutJai15.pdf (Zugriff: 18.11.2016).

MSW NRW – Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen (Hrsg.) (2004a). *Kernlehrplan für die Gesamtschule – Sekundarstufe I in Nordrhein-Westfalen, Mathematik*. Online: https://www.schulentwicklung.nrw.de/lehrplaene/upload/lehrplaene_download/gesamtschule/gs_mathematik.pdf (Zugriff: 18.12.2019).

MSW NRW – Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen (Hrsg.) (2004b). *Kernlehrplan für die Realschule in Nordrhein-Westfalen, Mathematik*. Online: https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/45/rs_mathematik.pdf (Zugriff: 18.12.2019).

MSW NRW – Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.) (2007). *Kernlehrplan für das Gymnasium – Sekundarstufe I (G8) in Nordrhein-Westfalen, Mathematik*. Online: https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/46/gym8_mathematik.pdf (Zugriff: 19.01.2016).

MSW NRW – Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.) (2008a). *Kernlehrplan für die Sekundarstufe II Gymnasium/Gesamtschule in Nordrhein-Westfalen, Mathematik*. Online: https://www.schulentwicklung.nrw.de/lehrplaene/lehrplannavigator-s-ii/gymnasiale-oberstufe/mathematik/index.html (Zugriff: 19.01.2016).

MSW NRW – Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.) (2008b. *Lehrplan Mathematik für die Grundschulen des Landes Nordrhein-Westfalen*. Lehrplannavigator. Online: https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_gs/GS_LP_M.pdf (Zugriff: 19.01.2016).

Müller, H. (1991). Grundtypen des Begründens im Mathematikunterricht. *Mathematik in der Schule*, *29*(11), 737–745.

NCTM – National Council of Teachers of Mathematics (Hrsg.) (2000). *Principles and standards for school mathematics*. Reston: Author.

Nelsen, R. B. (2000). *Proofs without words 2: more exercises in visual thinking*. Washington, DC: The Mathematical Association of America.

Neubrand, M. & Möller, M. (1990). *Einführung in die Arithmetik: Ein Arbeitsbuch für Studierende des Lehramts der Primarstufe*. Bad Salzdetfurth: Franzbecker.

Neumann, A., Beier, F., Ruwisch, S. (2014). Schriftliches Begründen im Mathematikunterricht. *Zeitschrift für Grundschulforschung, 7*(1), 113–125.

Ostsieker, L. & Biehler, R. (2012). Analyse von Beweisprozessen von Studienanfänger/innen bei der Bearbeitung von Aufgaben zur Konvergenz von Folgen. In M. Ludwig & M. Kleine (Hrsg.), *Beiträge zum Mathematikunterricht 2012* (S. 641–644). Münster: WTM-Verlag.

Ott, B. (2016). *Textaufgaben grafisch darstellen: Entwicklung eines Analyseinstruments und Evaluation einer Interventionsmaßnahme*. Münster: Waxmann.

Ozgur, Z., Ellis, A. B., Vinsonhaier, R., Fatih Dogan, M. & Knuth, E. (2019). From examples to proof : Purposes, strategies, and affordances of example use. *Journal of Mathematical Behavior, 53*, 284–303.

Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics*, *66*(1), 23–41.

Pólya, G. (1949). *Schule des Denkens: Vom Lösen mathematischer Probleme*. Bern: Francke.

Pólya, G. (1966). *Vom Lösen mathematischer Aufgaben: Einsicht und Entdeckung, Lernen und Lehren*. Basel [u. a.]: Birkhäuser.

PONS Langenscheidt GmbH (Hrsg.) (o. J.). *Klassifikation*. PONS Online-Wörterbuch. Online: https://de.pons.com/übersetzung/deutsche-rechtschreibung/Klassifikation (Zugriff: 27.04.2022).

Porteous, K. (1990). What do children really believe? *Educational Studies in Mathematics, 21*(6), 589–598.

Prediger, S. (2013). Sprachmittel für mathematische Verstehensprozesse. In A. Pallack (Hrsg.), *Impulse für eine zeitgemäße Mathematiklehrer-Ausbildung. MNU-Dokumentation der 16. Fachleitertagung Mathematik* (S. 26–36). Neuss: Seeberger.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J. & Schiefele, U. (Hrsg.) (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* Münster [u. a.]: Waxmann.

Prestin, J. (2017). Internationale Mathematik-Olympiade: Bericht über die 58. Internationale Mathematik-Olympiade (IMO). *Mitteilungen des Mathematik-Olympiaden e. V.*, *43*, 23–35.

QUA-LiS NRW – QualitätsAgentur – Landesinstitut für Schule (Hrsg.) (2018). *Mathematische Angebote für interessierte Schülerinnen und Schüler*. Online: https://www.schulentwicklung.nrw.de/sinus/front_content.php?idcat=2895&lang=9 (Zugriff: 10.06.2018).

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, *7*(1), 5–41.

Recio, A. M. & Godino, J. D. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics, 48*, 83–99.

Reid, D. A. & Knipping, C. (2010). *Proof in mathematics education: Research, learning and teaching*. Rotterdam: Sense Publisher.

Reiss, K. (o. J.). *Argumentieren, Begründen, Beweisen im Mathematikunterricht*. Online: http://blk.mat.uni-bayreuth.de/material/db/59/beweis.pdf (Zugriff: 19.12.2019).

Reiss, K. (2009). Wege zum Beweisen: Einen „Habit of Mind" im Mathematikunterricht etablieren. *mathematik lehren*, *155*, 4–11.

Reiss, K., Heinze, A., Kuntze, S., Kessler, S., Rudolph-Albert, F., & Renkl, A. (2006). Mathematiklernen mit heuristischen Lösungsbeispielen. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms* (S. 194–208). Münster: Waxmann.

Reiss, K., Hellmich, F. & Thomas, J. (2002). Individuelle und schulische Bedingungsfaktoren für Argumentationen und Beweise im Mathematikunterricht. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen. 45. Beiheft der Zeitschrift für Pädagogik* (S. 51–64). Online: https://www.academia.edu/54234991/Individuelle_und_schulische_Bedingungsfaktoren_f%C3%BCr_Argumentationen_und_Beweise_im_Mathematikunterrricht (Zugriff: 27.11.2015).

Reiss, K. & Ufer, S. (2009). Was macht mathematisches Arbeiten aus? Empirische Ergebnisse zum Lernen von Argumentationen, Begründungen und Beweisen. *Jahresbericht der Deutschen Mathematiker-Vereinigung, 111*(4), 155–177.

Reyhani, E., Hamidi, F. & Kolahdouz, F. (2011). A study on algebraic proof conception of high school second graders. *Procedia – Social and Behavioral Sciences, 31*, 236-241.

Rott, B. (2014). Mathematische Problembearbeitungsprozesse von Fünftklässlern – Entwicklung eines deskriptiven Phasenmodells. *Journal für Mathematik-Didaktik*, *35*, 251–282.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*. Newbury Park: Sage.

Rustemeyer, R. (1992). *Praktisch-methodische Schritte der Inhaltsanalyse*. Münster: Aschendorff.

Rychen, D. S. & Salganik, L. H. (Hrsg.) (2003). *Key Competencies for a Successful Life and a Well-Functioning Society*. Göttingen: Hogrefe & Huber.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando: Academic Press.

Schorcht, H. M. S. (2018). *Typisierung mathematikhistorischer Beispiele in deutschen Mathematikschulbüchern der Klassenstufen 1 bis 7*. Münster: WTM-Verlag.

Schreier, M. (2012). *Qualitative Content Analysis in Practice*. London: SAGE.

Schreier, M. (2013). Qualitative Forschungsmethoden. In W. Hussy, M. Schreier & G. Echterhoff (Hrsg.), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (S. 189–221). Berlin/Heidelberg: Springer.

Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Forum Qualitative Sozialforschung*, *15*(1), 1–27.

Schwarzkopf, R. (2000). *Argumentationsprozesse im Mathematikunterricht: Theoretische Grundlagen und Fallstudien*. Hildesheim: Franzbecker.

Selden, A. (2012). Transitions and Proof and Proving at Tertiary Level. In G. Hanna & M. de Villiers (Hrsg.), *Proof and Proving in Mathematics Education: The 19th ICMI Study* (S. 391–420). Heidelberg [u. a.]: Springer Science + Business Media.

Selden, A. & Selden, J. (2003). Validation of Proofs Considered as Texts: Can Undergraduates Tell Whether an Argument Proves a Theorem? *Journal for Research in Mathematics Education*, *34*(1), 4–36.

Selden, A. & Selden, J. (2013). The Genre of Proof. In M. Fried & T. Dreyfus (Hrsg.), *Mathematicians and mathematics educators: Searching for common ground* (S. 237–260). Dordrecht: Springer.

Senk, S. (1982). *Achievement in Writing Geometry Proofs*. Online: https://www.academia.edu/49068937/Achievement_in_Writing_Geometry_Proofs (Zugriff: 23.10.2020).

Senk, S. & Usiskin, Z. (1983). Geometry Proof Writing: A New View of Sex Differences in Mathematics Ability. *American Journal of Education, 91*(2), 187–201.

Sewerin, H. (1979). *Mathematische Schülerwettbewerbe: Beschreibungen, Analysen, Aufgaben, Trainingsmethoden; mit Ergebnissen einer Umfrage zum Bundeswettbewerb Mathematik*. München: Manz.

Smith, G. (2012). *EGMO: History*. Online: https://www.egmo.org/history/ (Zugriff: 01.04.2023).

Smith, E. & Henderson, K. (1959). Proof. In H. Fawcett, A. Hach, C. Junge, H. Syer, H. van Engen & P. Jones (Hrsg.), *The growth of mathematical ideas K-12 twenty-fourth yearbook* (S. 111–181). Washington, DC: NCTM.

Stangl, W. (2023). *Klassifikation*. Online Lexikon für Psychologie und Pädagogik. Online: https://lexikon.stangl.eu/10274/klassifikation (Zugriff: 27.04.2022).

Steinbring, H. (2009). Children's Ways of Mathematical Argumentation in the Classroom Environment. *Mediterranean Journal for Research in Mathematics Education*, *8*(1), 59–71.

Straehler-Pohl, H. & Gellert, U. (2012). Klassifikation. Facetten eines für die Unterrichtsforschung zentralen Begriffs. In U. Gellert & M. Sertl (Hrsg.), *Zur Soziologie des Unterrichts. Arbeiten mit Basil Bernsteins Theorie des pädagogischen Diskurses* (S. 89–118). Weinheim: Beltz Juventa.

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, *38*(3), 289–321.

Stylianides, A. J. & Al-Murani, T. (2010). Can a proof and a counterexample coexist? Students' conceptions about the relationship between proof and refutation. *Research in Mathematics Education, 12*(1), 21–36.

Tebaartz, P. C. (2013). *Beweiskompetenz von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade – eine Studie*. Unveröffentlichte wissenschaftliche Hausarbeit, Justus-Liebig-Universität Gießen.

Tebaartz, P. C. (2015). Eigenproduktionen zu Beweisaufgaben von Teilnehmerinnen und Teilnehmern der Mathematik-Olympiade. In F. Caluori, H. Linneweber-Lammerskitten & C. Streit (Hrsg.), *Beiträge zum Mathematikunterricht 2015* (S. 908 – 911). Münster: WTM-Verlag.

Tebaartz, P. C. & Lengnink, K. (2015). Was heißt mathematischer Beweis? – Realisierungen in Schülerproduktionen. In A. Budke, A. Creyaufmüller, M. Kuckuck, M. Meyer, F. Schäbitz, K. Schüter, & G. Weiss (Hrsg.), *Fachlich argumentieren lernen* (S. 105–120). Münster: Waxmann.

Thurston, W. P. (1994). On proof and progress in mathematics. *Bulletin of the American Mathematical Society*, *30*(2), 161–177.

Toulmin, S. E. (1958). *The Uses of Arguments*. Cambridge: University Press.

Ufer, S. (2008). Entwicklung geometrischer Beweiskompetenz in der Sekundarstufe I. In E. Vásárhelyi (Hrsg.), *Beiträge zum Mathematikunterricht 2008* (S. 771–774). Münster: WTM-Verlag.

Ufer, S., Heinze, A. & Reiss, K. (2008). Individual predictors of geometrical proof competence. In O. Figueras & A. Sepélveda (Hrsg.), *Proceedings of the Joint Meeting of PME 32 and PME-NA XXX* (Bd. 4, S. 361–368). Mexico: Cinvestac-UMSNH.

Ufer, S., Heinze, A., Kuntze, S. & Rudolph-Albert, F. (2009). Beweisen und Begründen im Mathematikunterricht: Die Rolle von Methodenwissen für das Beweisen in der Geometrie. *Journal für Mathematikdidaktik*, *30*(1), 30–54.

Veilande, I., Ramana, L. & Krauze, S. (2018). Repeated Participation at the Mathematical Olympiads: A Comparative Study of the Solutions of Selected Problems. In F. M. Singer (Hrsg.), *Mathematical Creativity and Mathematical Giftedness*. Cham: Springer.

Vogt, M. (2021). *Schriftliche Begründungskompetenzen in stochastischen Aufgabenkontexten: Eine Evaluations- und Interventionsstudie im 4. Schuljahr.* Wiesbaden: Springer Spektrum.

Walsch, W. (1975). *Zum Beweisen im Mathematikunterricht* (2. Aufl.). Berlin: Volk und Wissen.

Weber, K. (2001). Student Difficulty in Constructing Proofs: The Need for Strategic Knowledge. *Educational Studies in Mathematics*, *48*, 101–119.

Weinert, F. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 17–31). Weinheim: Beltz.

Winter, H. (1991). *Entdeckendes Lernen im Mathematikunterricht: Einblicke in die Ideengeschichte und ihre Bedeutung für die Pädagogik* (2. Aufl.). Braunschweig: Vieweg.

Wirtz, M. A. (Hrsg.) (2021). *Klassifikation*. Dorsch – Lexikon der Psychologie. Online: https://dorsch.hogrefe.com/stichwort/klassifikation#search=f9467a6029f9117e4734bcdd87b9ef3a&offset=0 (Zugriff: 27.04.2022).

Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Bern: Hogrefe Verlag für Psychologie.

Wittmann, E. (2009). Beweisen und Argumentieren. In H.-G. Weigand, A. Filler, R. Hölzl, S. Kuntze, M. Ludwig, J. Roth, B. Schmidt-Thieme, & G. Wittmann (Hrsg.), *Didaktik der Geometrie für die Sekundarstufe I* (S. 35–54). Heidelberg: Springer Spektrum.

Wittmann, E. & Müller, G. (1988). Wann ist ein Beweis ein Beweis? In P. Bender (Hrsg.), *Mathematikdidaktik: Theorie und Praxis* (S. 237–257). Berlin: Cornelsen.

Wittmann, E. C. & Ziegenbalg, J. (2004). Sich Zahl um Zahl hochangeln. In G. N. Müller, H. Steinbring, & E. C. Wittmann (Hrsg.), *Arithmetik als Prozess* (S. 35–53). Seelze: Kallmeyer.

Zaslavsky, O. & Knuth, E. (2019). The complex interplay between examples and proving: Where are we and where should we head? *Journal of Mathematical Behavior, 53*, 242–244.

Ziegler, G. (2008). Über das Buch der Beweise. Was ist Mathematik? – Versuch einer Antwort in vier Thesen. *Der mathematische und naturwissenschaftliche Unterricht*, *61*(7), 407–413.

Abbildungsverzeichnis

Abbildung 1: Zwei Beweisversuche von Fünftklässlern aus dem Datensatz dieser Arbeit ..1

Abbildung 2: Bundesweite Konzeption der Mathematik-Olympiade für Schülerinnen und Schüler in den Sekundarstufen ..12

Abbildung 3: Aufgabe 520631 der Mathematik-Olympiade und dazugehörige Lösung ..20

Abbildung 4: Aufgabe 590523 der Mathematik-Olympiade und dazugehörige Lösung ..21

Abbildung 5: Aufgabe 510835 der Mathematik-Olympiade und dazugehörig Lösung ..22

Abbildung 6: Modell der parallelen Struktur mathematischer Beweise von Aberdein (2013, S. 363) ..32

Abbildung 7: Grundstruktur des Modells zur Analyse von Beweisversuchen zu Allaussagen ..38

Abbildung 8: Analysemodell. Fokus: Begründender, prüfender oder sonstiges Ansatz ..39

Abbildung 9: Vereinfachte Übersicht über relevante Aspekte berücksichtigter mathematikdidaktischer Konzepte sowie Zuordnung von Beweisversuchen mit begründendem bzw. prüfendem Ansatz zu den Begrifflichkeiten dieser Konzepte ..41

Abbildung 10: Bearbeitung 3.3.2a (Ans3 *begründender Ansatz*) ..50

Abbildung 11: Bearbeitung 3.3.2b (Ans2 *prüfender Ansatz*) ..51

Abbildung 12: Bearbeitung 3.3.2c (Ans1 *weder begründender noch prüfender Ansatz*) ..51

Abbildung 13: Bearbeitung 3.3.2d (Ans55 *nicht eindeutig kategorisierbarer Ansatz*) ..52

Abbildung 14: Analysemodell. Fokus: Begründender Ansatz, Beweisideen ..54

Abbildung 15: Bearbeitung 3.3.3b (BI1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*) ..56

Abbildung 16: Bearbeitung 3.3.3b (BI2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden*) ..57

Abbildung 17: Bearbeitung 3.3.3c (BI3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*) ..57

Abbildung 18: Bearbeitung 3.3.3d (BI4 *Zahldarstellung im Dezimalsystem*) ..58

Abbildung 19: Bearbeitung 3.3.3e (BI5 *Reste bei Division einzelner Summanden durch vier oder acht*) ..59

Abbildung 20: Bearbeitung 3.3.3f (BI6 *Teilbarkeit der Summen zweier Zahlenpaare*) ..60

Abbildung 21: Bearbeitung 3.3.3g (BI7 *Teilbarkeit von vier aufeinanderfolgenden geraden Zahlen*) ..60

Abbildung 22: Bearbeitung 3.3.3h (BI8 *Darstellung ungerader Zahlen in Abhängigkeit gerader Zahlen*) ..61

Abbildung 23: Bearbeitung 3.3.3i (BI9 *Parität von Summen ungerader oder gerader Zahlen*) ..61

Abbildung 24: Bearbeitung 3.3.3j (BI10 *Abstände ungerader Zahlen*) ..62

Abbildung 25: Bearbeitung 3.3.3k (BI11 *vier Zahlen*) ..62

Abbildung 26: Bearbeitung 3.3.3l (BI12 *Vierer- oder Achterreihe*) 63
Abbildung 27: Bearbeitung 3.3.3m (BI13 *Regeln zur Teilbarkeit von Produkten oder Summen durch vier oder acht*) 63
Abbildung 28: Bearbeitung 3.3.3n (BI55 *nicht eindeutig kategorisierbar*) 64
Abbildung 29: Bearbeitungen 3.3.3p-q (Beweisversuche mit begründendem Ansatz: Analyse der Beweisideen) 65
Abbildung 30: Analysemodell. Fokus: Begründender Ansatz, Korrektheit und Vollständigkeit 69
Abbildung 31: Bearbeitung 3.3.4a (PzzA1 *Hinweis auf einen grundlegenden Fehler*) 74
Abbildung 32: Bearbeitung 3.3.4b (PzzA2 *Bezug auf die zz. Aussage unklar*) 74
Abbildung 33: Bearbeitung 3.3.4c (PzzA3 *Passung zu der zz. Aussage*) 75
Abbildung 34: Bearbeitung 3.3.4d (KV1 *Beweisideen ohne relevanten Beitrag zu einem weiterentwickelbaren Argumentgefüge*) 75
Abbildung 35: Bearbeitung 3.3.4e (KV2 *Ansatz für ein weiterentwickelbares Argumentgefüge*) 76
Abbildung 36: Bearbeitung 3.3.4f (KV3 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler*) 77
Abbildung 37: Exemplarischer Erwartungshorizont (BIp1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*, Variante 2.1.1 *Betrachtung der jeweils nächstgrößeren Summe, Teilbarkeit durch acht*) 78
Abbildung 38: Bearbeitung 3.3.4g (KV4 *weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke*) 78
Abbildung 39: Bearbeitung 3.3.4h (KV5 *weitgehend korrektes und vollständiges Argumentgefüge mit einer Ungenauigkeit in der Formulierung*) 79
Abbildung 40: Bearbeitung 3.3.4i (KV6 *korrektes und vollständiges Argumentgefüge*) 81
Abbildung 41: Analysemodell. Fokus: Begründender Ansatz, Beispiele 83
Abbildung 42: Beweisversuche mit begründendem Ansatz: Vorgehensweise bei der Untersuchung der Verwendung von Beispielen 86
Abbildung 43: Bearbeitungen 3.3.5a-b (Beweisversuche mit begründendem Ansatz: Analyse der Verwendung von Beispielen) 90
Abbildung 44: Bearbeitungen 3.3.5c-e (Beweisversuche mit begründendem Ansatz: Analyse der Beispiele oberhalb des (übrigen) Argumentgefüges und Analyse der Beispiele innerhalb des (übrigen) Argumentgefüges) 92
Abbildung 45: Analysemodell. Fokus: Begründender Ansatz, Schlussweise 95
Abbildung 46: Bearbeitung 3.3.6a (Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit*) 97
Abbildung 47: Bearbeitung 3.3.6b (Sw2 *deduktiv mit induktivem Anteil*) 98
Abbildung 48: Bearbeitung 3.3.6c (Sw3 *deduktiv mit Angabe eines Beispiels*) 98
Abbildung 49: Bearbeitung 3.3.6d (Sw4 *deduktiv ohne zusätzliches Beispiel*) 99
Abbildung 50: Analysemodell. Fokus: Prüfender Ansatz, Korrektheit, Beweisideen, Beispiele und Angabe eines Wahrheitswerts 101

Abbildung 51: Bearbeitungen 3.3.7a-c (Beweisversuche mit prüfendem Ansatz: Analyse der Existenz, Passung, Anzahl und Art von Beispielen, der Angabe eines Wahrheitswerts, der Korrektheit und der Beweisideen) 104

Abbildung 52: Analysemodell. Fokus: Prüfender Ansatz, Schlussweise.....110

Abbildung 53: Kategorien zur Beschreibung der Schlussweise bei Beweisversuchen mit prüfendem Ansatz und einer Affirmation der Aussage 112

Abbildung 54: Bearbeitung 3.3.8a (Swpr1 *manifeste Beweisidee*) 112

Abbildung 55: Bearbeitung 3.3.8b (Swpr2.1 *Wahrheitswert explizit mit Beispielen begründet, evtl. subjektiv realisierte Beweisidee, Hinweis auf Allgemeingültigkeit*), Bearbeitung 3.3.8c (Swpr2.2 *Wahrheitswert explizit mit Beispielen begründet, evtl. subjektiv realisierte Beweisidee*), Bearbeitung 3.3.8d (Swpr2.3 *Wahrheitswert explizit mit Beispielen begründet, Hinweis auf Allgemeingültigkeit*), Bearbeitung 3.3.8e (Swpr2.4 *Wahrheitswert explizit mit Beispielen begründet*).....113

Abbildung 56: Bearbeitung 3.3.8f (Swpr3.3 *Wahrheitswert nach allen Beispielen expliziert, Hinweis auf Allgemeingültigkeit*), Bearbeitung 3.3.38g (Swpr4.4 *Wahrheitswert vor allen Beispielen expliziert*) 114

Abbildung 57: Bearbeitung 3.3.8h (Swpr55 *nicht eindeutig kategorisierbar*).....114

Abbildung 58: Analysemodell. Fokus: Weder begründender noch prüfender Ansatz, Angabe eines Wahrheitswerts und Art der Bearbeitung.....116

Abbildung 59: Bearbeitung 3.3.9a (WNart4.1 *Berufung auf die Aussage, Allgemeingültigkeit der Aussage als Grund*), Bearbeitung 3.3.9b (WNart4.2 *Berufung auf die Aussage, Gültigkeit der Aussage als Grund*).....118

Abbildung 60: Bearbeitung 3.3.9c (WNart3.1 *Angabe der Aussage oder eines Wahrheitswerts, Affirmation mit Betonung der Allgemeingültigkeit*), Bearbeitung 3.3.9d (WNart3.2 *Angabe der Aussage oder eines Wahrheitswerts, Affirmation ohne Betonung der Allgemeingültigkeit*), Bearbeitung 3.3.9e (WNart3.3 *Angabe der Aussage oder eines Wahrheitswerts, keine klare Affirmation*) 118

Abbildung 61: Bearbeitung 3.3.9f (WNart2.2 *abgebrochene Bearbeitung, Abbruch nach Angabe der Gültigkeit*), Bearbeitung 3.3.9g (WNart2.3 *abgebrochene Bearbeitung, durchgestrichene Ausführungen*), Bearbeitung 3.3.9h (WNart2.4 *abgebrochene Bearbeitung, durchgestrichene Ausführungen*).....119

Abbildung 62: Bearbeitung 3.3.9i-j (WNart1 *sonstiges*).....120

Abbildung 63: Beweisversuche mit weder begründendem noch prüfendem Ansatz, Arten der Bearbeitung und Anknüpfungspunkte für die Entwicklung eines Beweises 121

Abbildung 64: Analysemodell. Fokus: Nicht eindeutig kategorisierbarer Ansatz, Korrektheit, Beweisideen, Beispiele und Angabe eines Wahrheitswerts ...123

Abbildung 65: Bearbeitungen 3.3.10a-c (Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz: Analyse der Passung der Ausführungen zu der zu zeigenden Aussage, der Erkennbarkeit einer Beweisidee und ggf. der primären Beweisidee, der Existenz, Passung, Anzahl und Art von Beispielen sowie der Existenz und Art eines Wahrheitswerts) 125

Abbildung 66: Analysemodell. Fokus: Darstellungsmittel und Darstellungsform 126

Abbildung 67: Bearbeitungen 3.3.11a-b (Beweisversuche mit symbolischen Darstellungsmitteln: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform) 130

Abbildung 68: Bearbeitungen 3.3.11c-d (Beweisversuche mit ikonischen Darstellungsmitteln: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform), Bearbeitung 3.3.11e (Beweisversuch ohne numerische Darstellungsmittel: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform) 131

Abbildung 69: Bearbeitung 3.3.11f (Beweisversuch ohne verbale Darstellungsmittel: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform), Bearbeitung 3.3.11g (Beweisversuch mit verschiedenartigen strukturierenden Darstellungsmitteln: Analyse der vorkommenden Darstellungsmittel und der Darstellungsform) 134

Abbildung 70: Grundstruktur des Modells zur Analyse von Beweisversuchen zu Allaussagen 139

Abbildung 71: Ausdifferenzierung der Kompetenz „Mathematisch argumentieren“ in drei Anforderungsbereiche (KMK, 2003, S. 13-14, H. i. O.) 148

Abbildung 72: Für die Studie konzipierte Beweisaufgabe für die Klassenstufen 5 bis 8 ... 186

Abbildung 73: Darstellung der Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* (oben), der Beweisidee 2 *Darstellung einer Summe in Abhängigkeit des kleinsten Summanden* (unten links) und der Beweisidee 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung* (unten rechts) insbesondere mithilfe figurierter Zahlen .. 196

Abbildung 74: Darstellung der Beweisidee 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante* (links) und der Beweisidee 6 *Teilbarkeit der Summen zweier Zahlenpaare* (rechts) mithilfe verschiedener ikonischer Darstellungsmittel 197

Abbildung 75: Beweis der zu zeigenden Aussage in symbolischer Darstellungsform 197

Abbildung 76: Schritte einer qualitativ-strukturierenden Inhaltsanalyse nach Schreier (2012, 2014) 203

Abbildung 77: Kreuztabelle mit beobachteten Häufigkeiten sowie dazugehörige Kreuztabelle mit erwarteten Häufigkeiten bei empirischer Unabhängigkeit zweier Merkmale X und Y mit p bzw. q ($p, q \in \mathbb{N} \geq 2$) möglichen Ausprägungen 213

Abbildung 78: Schritte der qualitativ-strukturierenden Inhaltsanalyse im Rahmen dieser Arbeit 219

Abbildung 79: Forschungsfragen und ihre Konkretisierungen 229

Abbildung 80: Beweisversuche mit begründendem, prüfendem, weder begründendem noch prüfendem und nicht eindeutig kategorisierbarem Ansatz sowie fehlende Bearbeitungen: Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 232

Abbildung 81: Beweisversuche mit begründendem Ansatz: Passung zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 235

Abbildung 82: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 236

Abbildung 83: Beweisversuche mit begründendem Ansatz: Passung zu der zu zeigenden Aussage in Abhängigkeit der Existenz eines weitgehend korrekten Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 238

Abbildung 84: Beweisversuche mit begründendem Ansatz: Schlussweise. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 239

Abbildung 85: Beweisversuche mit begründendem Ansatz: Gruppierung der Beweisversuche bezüglich derer Korrektheit und Vollständigkeit sowie der Schlussweise und dazugehörige Häufigkeitsverteilungen 241

Abbildung 86: Beweisversuche mit begründendem Ansatz: Gruppen an Beweisversuchen in Abhängigkeit der Korrektheit, Vollständigkeit und Schlussweise. Besonders häufige Gruppen und deren Anteile differenziert nach Klassenstufen 242

Abbildung 87: Beweisversuche mit begründendem Ansatz: Schlussweise in Abhängigkeit der Existenz eines weitgehend korrekten Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 243

Abbildung 88: Beweisversuche mit begründendem Ansatz: Existenz eines weitgehend korrekten Argumentgefüges in Abhängigkeit der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 244

Abbildung 89: Beweisversuche mit begründendem Ansatz aus den Klassenstufen 5, 8 mit BI1,10,11 (oben) bzw. BI7,8,11 (unten) 246

Abbildung 90: Beweisversuche aus den Klassenstufen 5, 8 mit BIp9, BI55 (oben) und Beweisversuche aus den Klassenstufen 6, 8 mit BIp2 (unten)....247

Abbildung 91: Relative Häufigkeiten des Vorkommens der unterschiedenen Beweisideen in den Beweisversuchen mit begründendem Ansatz differenziert nach Klassenstufen 249

Abbildung 92: Relative Häufigkeiten des Vorkommens der unterschiedenen Beweisideen in den Beweisversuchen mit begründendem Ansatz differenziert nach Klassenstufen: p-Werte von Signifikanztests und dazugehörige Effektstärken 250

Abbildung 93: Beweisversuche mit begründendem Ansatz: Mehrere Beweisideen vs. genau eine Beweisidee. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests, dazugehörige Effektstärken 251

Abbildung 94: Beweisversuche mit begründendem Ansatz: Zielführendsten Beweisidee. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 252

Abbildung 95: Beweisversuche mit begründendem Ansatz: Zielführendste Beweisidee in Abhängigkeit der primären Beweisidee. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 253

Abbildung 96: Beweisversuchen mit begründendem Ansatz: Existenz und Passung eines Beispiels zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 256

Abbildung 97: Beweisversuche mit begründendem Ansatz und Beispiel: Gesamtanzahl an Beispielen. Relative Häufigkeitsverteilungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 258

Abbildung 98: Beweisversuchen mit begründendem Ansatz und Beispiel: Systematik angegebener Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests, dazugehörige Effektstärken 258

Abbildung 99: Beweisversuche mit begründendem Ansatz und Beispiel: Orte angegebener Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 260

Abbildung 100: Beweisversuche mit begründendem Ansatz und Beispiel: Orte aller angegebenen Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 260

Abbildung 101: Beweisversuche mit begründendem Ansatz und einem Beispiel oberhalb: Zusammenfassende Betrachtung der Art der oberhalb angegebenen Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 261

Abbildung 102: Bearbeitungen 8.2.3a (Beweisversuch mit begründendem Ansatz und einem korrekt und vollständig ausgeführten Beispiel zu der zz. Aussage, aber keinem weiterentwickelbaren Argumentgefüge) 264

Abbildung 103: Bearbeitung 8.2.3b (Beweisversuch mit begründendem Ansatz und Beispielen, die u. a. vermuten lassen, dass die zu zeigende Aussage korrekt interpretiert wurde und sich die Ausführungen auf die zu zeigende Aussage beziehen) 264

Abbildung 104: Bearbeitung 8.2.3c (Beweisversuch mit begründendem Ansatz, Beispielen und der Schlussweise Sw4 *deduktiv ohne zusätzliches Beispiel*)....265

Abbildung 105: Bearbeitung 8.2.3d (Beweisversuch mit begründendem Ansatz und der Schlussweise Sw3 *deduktiv mit zusätzlichem Beispiel*) 266

Abbildung 106: Bearbeitungen 8.2.3e-f (Beweisversuche mit begründendem Ansatz, der Schlussweise Sw2 *deduktiv mit induktivem Anteil* und einem Beispiel unterhalb des übrigen Argumentgefüges) 267

Abbildung 107: Bearbeitungen 8.2.3g-h (Beweisversuche mit begründendem Ansatz und der Schlussweise Sw2 *deduktiv mit induktivem Anteil*, die zu einem bedeutsamen Anteil mithilfe von Beispielen entwickelt oder dargestellt wurden) 269

Abbildung 108: Bearbeitungen 8.2.3i-j (Beweisversuche mit begründendem Ansatz und der Schlussweise Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit*, die vermuten lassen, dass durch die Betrachtung von Reihenfolgebeispielen eine Beweisidee gefunden wurde).....270

Abbildung 109: Bearbeitung 8.2.3k (Beweisversuch mit begründendem Ansatz, der Schlussweise Sw1 *induktiv mit Andeutung der Verallgemeinerbarkeit* und einem weitgehend korrekten und vollständigen Argumentgefüge basierend auf der Beweisidee 3 *Vielfaches einer mittleren Zahl oder Durchschnittsbildung*, der von einem Fünftklässler verfasst wurde).....271

Abbildung 110: Beweisversuche mit begründendem Ansatz: Vorkommende Darstellungsmittel. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 273

Abbildung 111: Bearbeitungen 8.2.4a-c (Beweisversuche mit begründendem Ansatz und Buchstabenvariablen, die am Beispiel der Klassenstufe 6 die Unterschiedlichkeit der Variablennutzung veranschaulichen) 274

Abbildung 112: Bearbeitung 8.2.4d (Beweisversuch mit begründendem Ansatz, in dem ausschließlich numerische und ikonische Darstellungsmittel vorkommen).....275

Abbildung 113: Beweisversuche mit begründendem Ansatz: Darstellungsform. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 277

Abbildung 114: Beweisversuche mit begründendem Ansatz: Nachweis der Gültigkeit. Klassifikation der Beweisversuche 279

Abbildung 115: Beweisversuche mit begründendem Ansatz: Nachweis der Gültigkeit. Relative Häufigkeitsverteilungen der Gruppen an Beweisversuchen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 282

Abbildung 116: Beweisversuche mit begründendem Ansatz: Passung zu der zu zeigenden Aussage in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 285

Abbildung 117: Beweisversuche mit begründendem Ansatz: Primäre Beweisideen in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise. Spektrum an vorkommenden Ausprägungen und relative Häufigkeitsverteilungen der Codierungen 290

Abbildung 118: Beweisversuche mit begründendem Ansatz: Existenz und Passung eines Beispiels zu der zu zeigenden Aussage in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen 297

Abbildung 119: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in Abhängigkeit der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 298

Abbildung 120: Beweisversuche mit begründendem Ansatz: Existenz und Art angegebener Beispiele in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen 300

Abbildung 121: Beweisversuche mit begründendem Ansatz: Existenz eines weitgehend korrekten Argumentgefüges und der primären Beweisidee 1 in Abhängigkeit der Existenz von Reihenfolgebeispielen und des kleinsten Beispiels. Relative Häufigkeitsverteilungen der Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 301

Abbildung 122: Beweisversuche mit begründendem Ansatz: Existenz und Ort angegebener Beispiele in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges sowie der Schlussweise. Relative Häufigkeitsverteilungen der Codierungen 304

Abbildung 123: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges in Abhängigkeit der Existenz eines Beispiels innerhalb des (übrigen) Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 307

Abbildung 124: Beweisversuche mit begründendem Ansatz: Existenz, Art und Ort von Beispielen differenziert nach primären Beweisideen. Relative Häufigkeitsverteilungen der Codierungen sowie p-Werte von Signifikanztests und dazugehörie Effektstärken 309

Abbildung 125: Beweisversuche mit begründendem Ansatz: Existenz von Reihenfolgebeispielen und des kleinsten Beispiels (Fokus: Primäre Beweisideen 4, 12). Relative Häufigkeitsverteilungen von Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 310

Abbildung 126: Beweisversuche mit begründendem Ansatz: Darstellungsformen in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges und der Schlussweise. Relative Häufigkeitsverteilungen von Codierungen 313

Abbildung 127: Beweisversuche mit begründendem Ansatz: Darstellungsformen in Abhängigkeit von Schlussweisen. Relative Häufigkeitsverteilungen von Codierungen differenziert nach Kl. sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 314

Abbildung 128: Beweisversuche mit begründendem Ansatz: Darstellungsformen in Abhängigkeit der Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Relative Häufigkeitsverteilungen von Codierungen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 315

Abbildung 129: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis.....322

Abbildung 130: Bearbeitung 8.2.6a (We3429 *begründender Ansatz, Schlussweise deduktiv ohne zusätzliches Beispiel, kein weiterentwickelbares Argumentgefüge, kein Hinweis auf einen grundlegenden Fehler, keine ausbaubare primäre Beweisidee*).....329

Abbildung 131: Bearbeitung 8.2.6b (We3229 *begründender Ansatz, Schlussweise induktiv mit Andeutung der Verallgemeinerbarkeit, deduktiv mit induktivem Anteil oder deduktiv mit zusätzlichem Beispiel, kein weiterentwickelbares Argumentgefüge, kein Hinweis auf einen grundlegenden Fehler, keine ausbaubare primäre Beweisidee*) 330

Abbildung 132: Bearbeitung 8.2.6c (We3225 *begründender Ansatz, Schlussweise induktiv mit Andeutung der Verallgemeinerbarkeit, deduktiv mit induktivem Anteil oder deduktiv mit zusätzlichem Beispiel, kein weiterentwickelbares Argumentgefüge, kein Hinweis auf einen grundlegenden Fehler, ausbaubare primäre Beweisidee*) 331

Abbildung 133: Bearbeitung 8.2.6d (We3241 *begründender Ansatz, Schlussweise induktiv mit Andeutung der Verallgemeinerbarkeit, deduktiv mit induktivem Anteil oder deduktiv mit zusätzlichem Beispiel, weitgehend korrektes Argumentgefüge mit inhaltlicher Lücke, primäre Beweisidee* 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*).....333

Abbildung 134: Bearbeitung 8.2.6e (We3151 *begründender Ansatz, Schlussweise zwischen induktiv und deduktiv, weitgehend korrektes und vollständiges Argumentgefüge, primäre Beweisidee* 1 *sukzessive Vergrößerung einer Summe mit Erkennen einer Invariante*) 334

Abbildung 135: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Häufigkeitsverteilung der Beweisversuche (Fokus: Klassenstufe 5) 336

Abbildung 136: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Häufigkeitsverteilung der Beweisversuche (Fokus: Klassenstufe 6) 336

Abbildung 137: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Häufigkeitsverteilung der Beweisversuche (Fokus: Klassenstufe 7) 337

Abbildung 138: Klassifikation der Beweisversuche mit begründendem Ansatz mit besonderem Fokus auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Häufigkeitsverteilung der Beweisversuche (Fokus: Klassenstufe 8) 337

Abbildung 139: Beweisversuche mit prüfendem Ansatz: Passung eines Beispiels zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen 342

Abbildung 140: Beweisversuche mit prüfendem Ansatz: Passung der Ausführungen zu der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen 343

Abbildung 141: Beweisversuche mit prüfendem Ansatz: Existenz und Art eines angegebenen Wahrheitswerts. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen 344

Abbildung 142: Gruppierung der Beweisversuche mit prüfendem Ansatz mit Blick auf die Interpretation der zu zeigenden Aussage. Gebildete Gruppen und deren Häufigkeiten 345

Abbildung 143: Bearbeitung 8.3.2a (IntzzA3 *Andeutung einer korrekten Interpretation der zu zeigenden Aussage*) 346

Abbildung 144: Bearbeitung 8.3.2b (IntzzA2 *Passung der Ausführungen zu der zu zeigenden Aussage bei teilweise unklarer Interpretation*) 346

Abbildung 145: Bearbeitung 8.3.2c (IntzzA1 *Hinweis auf eine nicht ganz korrekte Interpretation der zz. Aussage oder einen inhaltlichen Fehler*) 346

Abbildung 146: Beweisversuche mit prüfendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 347

Abbildung 147: Beweisversuche mit prüfendem Ansatz: Systematik angegebener Beispiele. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken....348

Abbildung 148: Beweisversuche mit prüfendem Ansatz: Existenz und Art eines besonderen Beispiels. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Wert eines Signifikanztests und dazugehörige Effektstärke 350

Abbildung 149: Gruppierung der Beweisversuche mit prüfendem Ansatz mit Blick auf die Entwicklung eines korrekten und vollständigen Argumentgefüges. Gebildete Gruppen und deren Häufigkeiten 351

Abbildung 150: Bearbeitung 8.3.2d (EntwAg41 *weitgehend korrektes Argumentgefüge basierend auf der primären Beweisidee* 1) 352

Abbildung 151: Bearbeitung 8.3.2e (EntwAg33 *weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler basierend auf einer anderen als der primären Beweisidee* 1) 353

Abbildung 152: Bearbeitung 8.3.2f (EntwAg21 *Anlage eines weiterentwickelbaren Argumentgefüges*)....353

Abbildung 153: Bearbeitungen 8.3.2g (EntwAg20 *sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges*) 354

Abbildung 154: Bearbeitungen 8.3.2h (EntwAg20 *sachinhaltliche oder heuristische Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges*) 354

Abbildung 155: Bearbeitung 8.3.2i (EntwAg10 *keine besonderen Anknüpfungspunkte für die Entwicklung eines korrekten und vollständigen Argumentgefüges*)354

Abbildung 156: Beweisversuche mit prüfendem Ansatz: Schlussweise. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 356

Abbildung 157: Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage 359

Abbildung 158: Beweisversuche mit prüfendem Ansatz: Nachweis der Gültigkeit. Relative Häufigkeitsverteilungen der Gruppen differenziert nach Klassenstufen 360

Abbildung 159: Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis 362

Abbildung 160: Klassifikation der Beweisversuche mit prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den einzelnen Gruppen an allen Beweisversuchen mit prüfendem Ansatz sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 364

Abbildung 161: Klassifikation der Beweisversuche mit weder begründendem noch prüfendem Ansatz mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis 368

Abbildung 162: Bearbeitung 8.4.4a (We1220 *Affirmation der Aussage, Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*) 369

Abbildung 163: Bearbeitung 8.4.4b (We1120 *Affirmation der Aussage, kein Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*) 369

Abbildung 164: Bearbeitung 8.4.4c (We1120 *keine Affirmation der Aussage, Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*), Bearbeitung 8.4.4d (We1110 *keine Affirmation der Aussage, kein Hinweis auf die Idee der Angabe eines Grundes oder eines Beispiels*) 369

Abbildung 165: Bearbeitung 8.4.4e (We5500 *nicht eindeutig kategorisierbar*)..... 370

Abbildung 166: Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage 373

Abbildung 167: Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Gruppen mit Fokus auf (1) der Existenz abgegebener Ausführungen und eines begründenden, prüfenden oder weder begründenden noch prüfenden Ansatzes sowie (2) der Existenz abgegebener Ausführungen und der Korrektheit, Vollständigkeit dieser Ausführungen 375

Abbildung 168: Klassifikation aller Beweisversuche mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Gruppen differenziert nach Klassenstufen 376

Abbildung 169: Alle Beweisversuche: Gesamtanzahl an Beispielen. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 383

Abbildung 170: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis..... 388

Abbildung 171: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 5) 392

Abbildung 172: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 6) 393

Abbildung 173: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 7) 394

Abbildung 174: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 8) 395

Abbildung 175: Formulierung des Aufgabenteils d) ohne Angabe eines Beispiels zu der zu zeigenden Aussage 411

Abbildung 176: Forschungsprojekt zum Führen von Beweisen bei der Mathematik-Olympiade in Deutschland: Skizze der Anlage und der Untersuchungsinstrumente....440

Abbildung 177: Eingesetzte Beweisaufgabe für die Klassenstufen 9 bis 12/13 441

Abbildung 178: Aufgabenstellung zum Validieren und Bewerten fiktiver Beweisversuche zu einer wahren Allaussage....443

Abbildung 179: Aufgabenstellung zum Validieren und Bewerten fiktiver Beweisversuche zu einer falschen Allaussage....444

Abbildung 180: Anzahl an vorhandenen Datensätzen differenziert nach Klassenstufen und der Version des Fragebogens 445

Abbildung 181: Entscheidungsgraph zur Bestimmung der relevanten Aufgabenteile im Rahmen der Analyse objektiver Anforderungsstrukturen von Aufgaben der Mathematik-Olympiade 448

Abbildung 182: Aufgabe 550622 der Mathematik-Olympiade und Lösung der Aufgabenteile a, b 450

Abbildung 183: Exemplarische Bestimmung der Relevanz der Aufgabenteile 550622 a, b, c 451

Abbildung 184: Modell zur Analyse objektiver Anforderungsstrukturen von Aufgaben mit besonderem Fokus auf dem Führen von Beweisen 452

Abbildung 185: Aufgaben 520724, 470613 der Mathematik-Olympiade 453

Abbildung 186: Analyse von Aufgabenteilen der Mathematik-Olympiade: Aufgabenteile mit Beweisnotwendigkeit vs. Aufgabenteile ohne Beweisnotwendigkeit. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen und Wettbewerbsrunden....456

Abbildung 187: Analyse von Aufgabenteilen der Mathematik-Olympiade. Fokus: Gültigkeitsbereich und Abstraktheit der zu zeigenden Aussage. Relative Häufigkeitsverteilungen der Codierungen differenziert nach Klassenstufen und Wettbewerbsrunden 456

Abbildung 188: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges bei Beweisversuchen mit deduktiver Schlussweise und solchen mit zwischen induktiver und deduktiver Schlussweise. Relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen aus der Klassenstufe 9 sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 460

Abbildung 189: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges differenziert nach primären Beweisideen. Relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen aus der Klassenstufe 9 sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 461

Abbildung 190: Beweisversuche mit begründendem Ansatz: Existenz und Ausführungsqualität eines weiterentwickelbaren Argumentgefüges je nach Darstellungsform. Relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen aus der Klassenstufe 9 sowie p-Werte von Signifikanztests und dazugehörige Effektstärken 465

Abbildung 191: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Anteile der Beweisversuche in den unterschiedenen Gruppen an allen Beweisversuchen (Fokus: Klassenstufe 9) 467

Abbildung 192: Bearbeitung 10.4.1a (We3452 *begründender Ansatz, Schlussweise deduktiv ohne zusätzliches Beispiel, weitgehend korrektes und vollständiges Argumentgefüge, primäre Beweisidee* 2 *oder primäre Beweisidee* 8 *zusammen mit der Beweisidee* 2) 468

Abbildung 193: Bearbeitung 10.4.1b (We3249 *begründender Ansatz, Schlussweise deduktiv ohne zusätzliches Beispiel, kein weiterentwickelbares Argumentgefüge, kein Hinweis auf einen grundlegenden Fehler, keine ausbaubare primäre Beweisidee*), Bearbeitungen 10.4.1c (We3342 *begründender Ansatz, Schlussweise deduktiv ohne zusätzliches Beispiel, weiterentwickelbares Argumentgefüge mit inhaltlichem oder grundlegendem Fehler, primäre Beweisidee* 2 *oder primäre Beweisidee* 8 *zusammen mit der Beweisidee* 2)..... 469

Abbildung 194: Teilnahme an mathematischen Förderangeboten von Teilnehmenden der Mathematik-Olympiade in den einzelnen Klassenstufen 5 bis 9. Relative Häufigkeitsverteilungen der Codierungen 478

Abbildung 195: Gruppierung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade in Abhängigkeit derer Teilnahme an der Landesrunde, derer Teilnahme an der Regionalrunde im vorhergehenden Schuljahr und derer Teilnahme an mathematischen Förderangeboten 482

Abbildung 196: Gruppierung der Beweisversuche von Teilnehmenden der Mathematik-Olympiade in Abhängigkeit derer Teilnahme an der Landesrunde, derer Teilnahme an der Regionalrunde im vorhergehenden Schuljahr und derer Teilnahme an mathematischen Förderangeboten. Relative Häufigkeitsverteilungen der Codierungen von Beweisversuchen differenziert nach Klassenstufen sowie p-Werte von Signifikanztests und dazugehörige Effektstärken..... 483

Tabellenverzeichnis

Tabelle 1: Kategorien zur Beschreibung der Existenz und Passung eines Beispiels zu der zu zeigenden Aussage insbesondere bei Beweisversuchen mit begründendem Ansatz 89

Tabelle 2: Kategorien zur Beschreibung, inwieweit in Beweisversuchen mit prüfendem Ansatz vollständig und weitgehend korrekt ausgeführte Beispiele zu der zu zeigenden Aussage vorkommen 105

Tabelle 3: Kategorien zur Beschreibung, ob einer und ggf. welcher Wahrheitswert in Beweisversuchen angegeben ist 106

Tabelle 4: Wertebereiche des Kappa-Koeffizienten für die Einschätzung der Güte der Übereinstimmung in Anlehnung an Wirtz und Caspar (2002, S. 59) 212

Tabelle 5: Übersicht über die Verwendung von Signifikanztests zur Überprüfung von Unterschieds- und Zusammenhangshypothesen sowie über die Berechnung von Cramérs V im Rahmen dieser Arbeit im Falle nominal skalierter Merkmale und falls die ordinale Struktur von Merkmalen nicht in Tests eingehen soll 216

Tabelle 6: Zufallsstichprobe: Anzahlen an vorhandenen und auszuwählenden Datensätzen....221

Tabelle 7: Intercoderreliabilität (Cohens Kappa κ). Fokus: Ansatz, Beweisideen in Beweisversuchen mit begründendem Ansatz 224

Tabelle 8: Intercoderreliabilität (Cohens Kappa κ). Fokus: Beweisversuche mit begründendem Ansatz: Korrektheit und Vollständigkeit, Beispiele sowie Schlussweise 225

Tabelle 9: Intercoderreliabilität (Cohens Kappa κ). Fokus: Beweisversuche mit prüfendem Ansatz....226

Tabelle 10: Intercoderreliabilität (Cohens Kappa κ). Fokus: Beweisversuche mit weder begründendem noch prüfendem Ansatz 226

Tabelle 11: Intercoderreliabilität (Cohens Kappa κ). Fokus: Beweisversuche mit nicht eindeutig kategorisierbarem Ansatz 226

Tabelle 12: Intercoderreliabilität (Cohens Kappa κ). Fokus: Darstellungsmittel und Darstellungsform 226

Tabelle 13: Beweisversuche mit begründendem Ansatz: Häufigste primäre Beweisideen differenziert nach Klassenstufen....254

Tabelle 14: Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Fokus: Klassenstufe 5; p-Werte von Signifikanztests und dazugehörige Effektstärken 378

Tabelle 15: Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Fokus: Klassenstufen 5, 6 vs. Klassenstufen 7, 8; p-Werte von Signifikanztests und dazugehörige Effektstärken....379

Tabelle 16: Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Fokus: Klassenstufe 8; p-Werte von Signifikanztests und dazugehörige Effektstärken 379

Tabelle 17: Anteile der Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge: p-Werte von Signifikanztests und dazugehörige Effektstärken 380

Tabelle 18: Klassifikation aller Beweisversuche mit Blick auf Anknüpfungspunkte für deren Weiterentwicklung zu einem Beweis. Relative Häufigkeiten der Beweisversuche in den gebildeten Gruppen differenziert nach Klassenstufen..... 391

Tabelle 19: Modell zur Analyse objektiver Anforderungsstrukturen von Aufgaben mit besonderem Fokus auf dem Führen von Beweisen: Definitionen der Anforderungsniveaus zur Beschreibung des Gültigkeitsbereichs und der Abstraktheit der Aussage bei Aufgabenteilen mit Beweisnotwendigkeit 454

Tabelle 20: Gruppen an Beweisversuchen mit Blick auf den Nachweis der Gültigkeit der zu zeigenden Aussage. Fokus: Klassenstufe 9 vs. Klassenstufen 5-8; p-Werte von Signifikanztests und dazugehörige Effektstärken..... 459

Tabelle 21: Beweisversuche mit begründendem Ansatz: Beweisideen. Fokus: Klassenstufe 9 vs. Klassenstufen 5-8; p-Werte von Signifikanztests und dazugehörige Effektstärken..... 460

Tabelle 22: Beweisversuche mit begründendem Ansatz und einem weitgehend korrekten Argumentgefüge. Fokus: Klassenstufe 9 vs. Klassenstufen 6-8; p-Werte von Signifikanztests und dazugehörige Effektstärken..... 463

Eidesstattliche Erklärung

Ich erkläre:

Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.