

# Leseentwicklung in der Grundschule: Eine Längsschnittstudie unter besonderer Betrachtung der förderdiagnostischen Kompetenz der Lehrkräfte

Dissertation  
zur Erlangung des Doktorgrades (Dr. phil.)  
des Fachbereiches Sozial- und Kulturwissenschaften  
der Justus-Liebig-Universität Gießen

vorgelegt von

Dipl.-Psych. Nadine Henrichs, geb. Losert

Hamburg, Juni 2021

Erstgutachterin: Prof. Dr. Michaela Timberlake

Zweitgutachterin: Prof. Dr. Christina Schwenck

Datum der Disputation: 28.10.2022

## **Zusammenfassung**

Lesekompetenz wird als bedeutsame, wenn nicht wichtigste, Kompetenz angesehen, die Schülerinnen und Schüler im Verlauf der Grundschulzeit erwerben müssen. Unzureichende Lesefähigkeiten erschweren eine selbstständige Teilhabe am gesellschaftlichen Leben (Hurrelmann, 2013, S. 163; Spinner, 2010, S. 48). Zudem können Schwierigkeiten im Lesen zu niedrigeren Schulabschlüssen und in der Folge zu schlechteren Chancen am Arbeitsmarkt führen (Esser, Wyschkon & Schmidt, 2002, S. 241). Aus diesem Grund kommt den förderdiagnostischen Kompetenzen von Grundschullehrkräften ein besonderes Augenmerk zu, da diese die Schwierigkeiten frühzeitig erkennen und mit geeigneten Unterstützungsmaßnahmen eingreifen müssen.

Zentrale Fragen dieser Studie waren: Können Lehrkräfte leseschwache Kinder ausmachen? Entwickeln sich die Kinder, die den Lehrkräften als leseschwach aufgefallen sind, positiver als Kinder, die nicht aufgefallen sind? Ergreifen die Grundschullehrkräfte unterstützende Maßnahmen und führen diese zu einer verbesserten Leseentwicklung?

Es wurde zur Beantwortung der Fragen eine Beobachtungsstudie durchgeführt, die zeitgleich in Deutschland, Südtirol und Finnland stattfand, um einen Vergleich der Länder zu ermöglichen. Die vorliegende Studie berücksichtigt nur die Daten aus Deutschland, die in 7 Grundschulen mit 13 Klassen und insgesamt 281 teilnehmenden Grundschulkindern erhoben wurden. Die Daten der Kinder wurden über Fragebögen und schulische Leistungstests zu Rechtschreib- und Lesefähigkeiten erhoben. Insgesamt gab es sieben Messzeitpunkte, wovon die ersten sechs jeweils zu Beginn, zur Mitte und zum Ende des ersten bzw. zweiten Schuljahres stattfanden. Die förderdiagnostische Kompetenz der Lehrkräfte wurde erfasst, indem sie in Interviews drei offene Fragen zu ihren Schülerinnen und Schülern beantworteten. Sie sollten zu jedem Messzeitpunkt die Kinder nennen, die ihnen auffällig geworden waren. Danach wurden sie zu jedem dieser Kinder gefragt, was ihnen aufgefallen war und schließlich, was getan wird. Die offenen Antworten der Lehrkräfte wurden durch zwei Beurteilerinnen kategorisiert. Die Antwortkategorien mit ausreichender Beurteilerübereinstimmung wurden in die Datenauswertung einbezogen. Zur statistischen Auswertung der Fragestellungen wurde teilweise auf statistisches Matching zurückgegriffen, da keine randomisierte Gruppenzuteilung erfolgen konnte. Ebenso liegt durch die Klassenverbände eine geschachtelte Datenstruktur vor, die mit dem statistischen Matching entsprechend berücksichtigt werden konnte. Fehlende Daten, wie sie in einer Langzeitstudie häufig vorkommen, wurden durch multiple Imputation über den chained equations-Ansatz (MICE) ergänzt (Enders, 2017; Rubin, 1987). Die Auswertung erfolgte mit den Programmen SPSS und R sowie RStudio.

Entgegen der Annahme fielen den Lehrkräften Mitte der ersten Klasse 94 Prozent und Ende der ersten Klasse 74 Prozent der leseschwachen Kinder nicht als solche auf. Insgesamt liegt dieser Anteil signifikant über dem theoretisch angenommenen Anteil von 33 Prozent, wie aufgrund der Ergebnisse der IGLU-Studie 2006 angenommen wurde (Valtin et al., 2010, S. 76). Das Geschlecht der Kinder konnte als ein Einflussfaktor auf die Urteilsakkuratheit der

Lehrkräfte ausgemacht werden. Während sich in der ersten Klasse noch kein Zusammenhang zeigte, wurden in der zweiten Klasse dagegen signifikant weniger leseschwache Mädchen leseauffällig als leseschwache Jungen.

Ein weiteres wesentliches Ergebnis ist, dass die Lesefähigkeiten der Kinder nicht dadurch beeinflusst wurden, ob die Lehrkräfte die Leseschwierigkeiten erkannt hatten. Zwischen ihnen und Kontrollgruppen, die über statistisches Matching bestimmt worden waren, zeigte sich kein signifikanter Unterschied in der Leseentwicklung und der Leseleistung zum Ende des vierten Schuljahres. Kinder, die den Lehrkräften allgemein negativ auffällig geworden sind, beispielsweise durch fehlende Lernmotivation oder auffälliges Verhalten, entwickelten sich sogar negativer in hierarchiehöheren Ebenen des Lesens als eine über statistisches Matching ermittelte Kontrollgruppe.

Analog zu Ergebnissen der IGLU-Studien verblieben auch in dieser Studie etwa zwei Drittel der Kinder ohne schulinterne Leseförderung (Bremerich-Vos, Stahns, Hußmann & Schurig, 2017, S. 281; Valtin et al., 2010, S. 76). Der Erfolg der genannten Unterstützungsmaßnahmen wurde ebenfalls betrachtet. Es konnte kein Zusammenhang gefunden werden zwischen dem Zeitpunkt der ersten Nennung einer Förderung und der Lesekompetenz zum Ende der vierten Klasse. Spezifische Förderprogramme wurden nur für zwei bis drei Kinder genannt, sodass der Erfolg dieser Maßnahmen nicht statistisch geprüft werden konnte.

Zudem konnte die Studie aufzeigen, dass eine Teilnahme an schulinterner Deutschförderung nicht zu besseren Lesefähigkeiten zum Ende der vierten Klasse führte. Leseschwache Kinder, die keine schulinterne Deutschförderung erhalten haben, zeigten zum Ende der vierten Klasse sogar eine bessere Leistung im sinnentnehmenden Lesen von Sätzen als Kinder, die eine solche Förderung erhalten haben. Der Grund für diesen Unterschied kann mit den vorliegenden Daten nicht ermittelt werden. Insgesamt nannten die Lehrkräfte allerdings nur sehr wenig spezifische Fördermethoden und keine Verfahren, die in der gesamten Klasse durchgeführt werden können, wie beispielsweise Lautleseverfahren.

Die Studie zeigt die Relevanz für regelmäßige und klassenweite Lesescreenings, um frühzeitig und zuverlässig leseschwache Kinder ermitteln zu können. Zudem sollten häufiger lese-spezifische Fördermaßnahmen ergriffen werden, da die Teilnahme an schulinterner Deutschförderung zu keiner besseren Leseentwicklung führte.

Schlüsselwörter: Lesekompetenz, Leseentwicklung, Förderdiagnostische Kompetenzen von Lehrkräften, Urteilsakkuratheit, Leseförderung, Grundschul Kinder

## **Abstract**

Literacy is seen as an important, if not the most important, skill that students must acquire by the end of elementary school. Studies have found insufficient reading abilities hinder independent participation in social life (Hurrelmann, 2013, p. 163; Spinner, 2010, p. 48). Furthermore, reading disabilities can lead to lower academic qualifications and thus, to a lack of opportunities in the work force (Esser, Wyschkon & Schmidt, 2002, p. 241). Therefore, the ability for elementary school teachers to recognize and provide early diagnosis and intervention measures to students with reading difficulties is essential.

The purpose of this study was to investigate the reading development of children and the diagnostic ability of their elementary school teachers. Research questions central to this study were: Are teachers able to correctly identify children with reading difficulties? Once a student is identified as having reading difficulties, do they show any improvements over those who were not identified as such? Do elementary school teachers intervene to bolster the reading skills of those who have been identified as having reading difficulties? And lastly, do these interventions provide a measurable improvement in the reading development of these children?

To investigate these questions, an observational study was conducted simultaneously in Germany, South Tyrol, and Finland. The study presented here focuses on the German data that comes from 7 schools and 13 classes equalling 281 participating elementary school students. The children filled in questionnaires and school performance tests were carried out to measure their reading and writing skills. There were seven measurement dates, the first six taking place at the beginning, the middle, and the end of grade one and grade two. The seventh and last measurement took place at the end of grade four. The diagnostic skills of elementary school teachers in regard to recognizing reading difficulties as well as to what measures they implemented to help their students were assessed through interviews with three open-ended questions. At each measurement date, without specifying a subject or skill, they were asked to name students who they noticed during their studies. After naming a student, they were asked what they had noticed and subsequently, what they were doing about it. The open answers from the teachers were categorized by two raters. Categories with acceptable inter-rater reliability were included in the data analyses. Some of the analyses made use of statistical matching, as this kind of study cannot be done with randomized control groups. Additionally, having the students organized in classes led to nested data structure which could also be taken into account by statistical matching. Missing data which appear often in longitudinal studies were handled by the chained equations approach for multiple imputation – MICE (Enders, 2017; Rubin, 1987). The statistical analyses were performed by the software programmes SPSS and R together with RStudio.

Contrary to previously held assumptions, 94 percent of children with reading difficulties in the middle of grade 1 and 74 percent at the end of grade 1 were not recognized as such by their teachers. In regard to the results of the German PIRLS 2006, it was assumed that 33 percent

of children with reading difficulties would not be recognized (Valtin et al., 2010, p. 76). This turned out to be a significant difference. One key finding in this study was the gender of the children. In grade 1, there was no relationship between teachers' judgement accuracy and the children's gender. In grade 2, however, significantly less girls who were weak in reading were recognized by their teachers as opposed to boys with the same difficulties.

Another key finding was that children's reading abilities were not influenced by teachers being aware of their reading difficulties. There was no significant difference between them and control groups that were created through statistical matching in their reading development and their reading ability at the end of grade 4. Children who stood out negatively in general, e. g., due to a low learning motivation or behavioural issues, developed even more negatively in hierarchically higher text comprehension processes than a control group determined through statistical matching.

In keeping with results of the German PIRLS in recent years, it could be found that two thirds of children with reading difficulties did not receive any measures to promote reading within school (Bremerich-Vos, Stahns, Hußmann & Schurig, 2017, p. 281; Valtin et al., 2010, p. 76). Another research question concerned the success of support measures given in school. There was no relationship between the point in time a teacher first mentioned a child receiving any support in school to foster their reading skills and their reading competencies at the end of grade 4. Specific training material to foster reading was only mentioned for two to three children. Thus, the success of using these could not be examined statistically.

Additionally, this study found that receiving measures to foster writing and reading in school did not lead to better reading abilities at the end of grade 4. Children with reading difficulties who had not participated in such courses actually showed a better performance in reading comprehension at the end of grade 4 than children who had participated. The reason for this difference cannot be determined with the present data. Altogether, teachers only named very few specific training programmes to foster reading and no programmes that can be carried out with the whole class such as oral reading instructional programmes.

The take-aways of this study include the importance of regular and class-wide screenings to reliably identify children with reading difficulties. Additionally, training programmes specific to fostering reading abilities should be used more often, as receiving measures to foster writing and reading in school did not lead to better reading abilities.

Keywords: literacy, reading development, reading assessment, teacher professionalization, judgment accuracy, reading intervention, elementary school children

## Inhalt

<b>1</b>	<b>Einleitung</b>	<b>11</b>
<b>2</b>	<b>Diagnostik im schulischen Setting</b>	<b>14</b>
2.1	Begriffe und Definitionen	14
2.2	Anlässe und Funktionen von pädagogisch-psychologischer Diagnostik	16
2.3	Methoden und Gütekriterien der Diagnostik	17
2.3.1	Gütekriterien	18
2.3.1.1	Objektivität	18
2.3.1.2	Reliabilität	19
2.3.1.3	Validität	20
2.3.1.4	Normierung	21
2.3.2	Weitere Gütekriterien	22
2.4	Anforderungen an Lehrkräfte	22
2.5	Kurze Darstellung der Diagnostikausbildung von Lehrkräften	23
2.6	Zwischenfazit	25
<b>3</b>	<b>Diagnostische Kompetenzen von Lehrkräften</b>	<b>26</b>
3.1	Zum Begriff der diagnostischen Kompetenz	26
3.2	Historischer Abriss zu den Anfängen des Forschungsgebiets	28
3.3	Aktueller Stand der Forschung	29
3.3.1	Akkuratheit von Lehrkräfteurteilen	29
3.3.2	Einflussfaktoren auf die Urteile von Lehrkräften	33
3.3.2.1	Schülermerkmale	35
3.3.2.2	Lehrkraftmerkmale	38
3.3.2.3	Kontaktdauer	39
3.3.2.4	Referenzgruppeneffekt	40
3.3.3	Auswirkungen der diagnostischen Urteile auf Schülerinnen und Schüler	41
3.4	Zwischenfazit und offene Fragen	45
<b>4</b>	<b>Über die Lesekompetenz von Grundschulkindern</b>	<b>47</b>
4.1	Definitionen und Begriffsbestimmung	47
4.2	Relevanz von Lesekompetenz	51
4.3	Zwischenfazit	52
<b>5</b>	<b>Der ungestörte Leseerwerb</b>	<b>53</b>
5.1	Grundprinzipien der deutschen Sprache im Schriftspracherwerb	53
5.2	Prozessmodelle des Leseerwerbs	55
5.2.1	Zwei-Wege-Modell des (lauten) Wortlesens	55
5.2.2	Konnektionistische Modelle der Worterkennung	56
5.2.3	Simple View of Reading	56
5.2.4	Prozessmodell des Lesenlernens	57
5.3	Entwicklungsmodelle des Leseerwerbs	58
5.3.1	Entwicklungsmodell nach Frith und Weiterentwicklung	58
5.3.2	Kompetenzentwicklungsmodell des Lesens	60
5.4	Leseunterricht in der Grundschule	61
5.4.1	Anfangsunterricht im Lesen	61
5.4.2	Leseerwerb im weiteren Verlauf der Grundschulzeit	63
5.4.3	Lesekompetenz zum Ende der Grundschulzeit	64
5.5	Zwischenfazit	65
<b>6</b>	<b>Schwierigkeiten beim Leseerwerb</b>	<b>67</b>
6.1	Lese-Rechtschreibstörung nach ICD-10 und DSM-5	67
6.2	Weitere Begriffe und Definitionen	69
6.3	Prävalenzen	72
6.4	Komorbiditäten und Auswirkungen	75
6.4.1	Psychische Auffälligkeiten	76

6.4.2	Verlauf von Lese-Rechtschreibschwierigkeiten.....	82
6.4.3	Schullaufbahn und Berufserfolg.....	83
6.5	Zwischenfazit .....	85
<b>7</b>	<b>Voraussetzungen und Vorläuferfertigkeiten des Leseerwerbs .....</b>	<b>87</b>
7.1	Phonologische Bewusstheit.....	87
7.2	Benennungsgeschwindigkeit .....	91
7.3	Zwischenfazit .....	94
<b>8</b>	<b>Ursachen und Risikofaktoren von Leseschwierigkeiten .....</b>	<b>95</b>
8.1	Emotionale, behaviorale und kognitive Faktoren .....	95
8.1.1	Neurobiologische und genetische Faktoren.....	96
8.1.2	Kognitive Fähigkeiten .....	97
8.1.3	Exekutive Funktionen .....	98
8.1.4	Aufmerksamkeit und Konzentration .....	101
8.1.5	Spracherwerbsstörungen .....	102
8.1.6	Geschlecht und affektive Faktoren .....	103
8.2	Soziale und kulturelle Herkunft .....	106
8.3	Ursachen durch am Leseprozess beteiligte Teilfertigkeiten .....	109
8.4	Zwischenfazit .....	111
<b>9</b>	<b>Förderdiagnostische Aspekte beim Leseerwerb in der Grundschule .....</b>	<b>113</b>
9.1	Diagnostik von Störungen beim Leseerwerb .....	113
9.1.1	Indikatoren für Schwierigkeiten beim Leseerwerb.....	113
9.1.2	Semiformelle Diagnostik.....	115
9.1.3	Formelle Diagnostik.....	116
9.1.3.1	Ein-Minuten-Lese Flüssigkeitstest des SLRT-II .....	116
9.1.3.2	Würzburger Leise Leseprobe (WLLP-R).....	117
9.1.3.3	Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6).....	117
9.2	Rahmenbedingungen zur Leseförderung.....	118
9.3	Inhaltliche Aspekte der Leseförderung .....	124
9.3.1	Förderung der basalen Lesefertigkeiten .....	124
9.3.2	Förderung des Leseverständnisses.....	127
9.4	Schlussfolgernd: Anforderungen an die Lehrkraft.....	128
<b>10</b>	<b>Ziel der Studie und Fragestellungen .....</b>	<b>130</b>
10.1	Ziel der Studie .....	130
10.2	Methodisches Vorgehen.....	131
10.3	Fragestellungen und Hypothesen.....	133
10.3.1	Fragestellung 1: Erkennen von Leseschwierigkeiten .....	133
10.3.2	Fragestellung 2: Einfluss der Diagnose auf die Leseentwicklung.....	135
10.3.3	Fragestellung 3: Eingeleitete (Förder-)Maßnahmen bei Auffälligkeiten .....	137
10.3.4	Fragestellung 4: Erfolg der durchgeführten (Förder-)Maßnahmen .....	138
<b>11</b>	<b>Methodik .....</b>	<b>140</b>
11.1	Stichprobenakquise .....	140
11.2	Ablauf der Studie.....	140
11.3	Soziodemographische Daten.....	141
11.4	Testverfahren bei den Grundschulkindern .....	141
11.4.1	Vorgaben zu den Itemanalysen für die selbst erstellten Verfahren .....	142
11.4.2	Vor- und Teilleistungen des Schriftspracherwerbs.....	144
11.4.2.1	Silbensegmentation mit Bildern .....	144
11.4.2.2	Silbensegmentation mit Wörtern.....	145
11.4.2.3	Satzsegmentation.....	146
11.4.2.4	Buchstabenkenntnis .....	146
11.4.2.5	Laut-Buchstaben-Zuordnung .....	147
11.4.3	Schriftspracherwerb .....	147
11.4.3.1	Schreiben nach Bildern .....	147

11.4.3.2	Rechtschreibleistung .....	149
11.4.3.3	Lesekompetenz .....	149
11.4.4	Schulisches Selbstkonzept und Lernmotivation .....	150
11.4.5	Klassenklima und Fähigkeitsselbstkonzept.....	151
11.4.6	Erfassung der kognitiven Fähigkeiten .....	152
11.4.6.1	Logisches Denken .....	152
11.4.6.2	Benennungsgeschwindigkeit .....	153
11.4.6.3	Arbeitsgedächtnis .....	153
11.5	Befragung der Lehrkräfte.....	154
11.5.1	Kodierung zu Frage 1: „Was fällt auf?“ .....	154
11.5.2	Kodierung zu Frage 2: „Was wird getan?“ .....	155
11.5.3	Ablauf der Kodierung.....	157
11.5.4	Beurteilerübereinstimmungen.....	158
11.5.4.1	Beurteilerübereinstimmung zu Frage 1: „Was fällt auf?“ .....	159
11.5.4.2	Lehrkrafturteil leseauffällig.....	162
11.5.4.3	Beurteilerübereinstimmung zu Frage 2: „Was wird getan?“ .....	162
11.5.4.4	Fazit zur Beurteilerübereinstimmung .....	164
11.6	Allgemeine Aspekte zu den Auswertungsmethoden .....	165
11.6.1	Bestimmung der Leseschwäche .....	166
11.6.2	Geschachtelte Datenstruktur .....	167
11.6.3	Umgang mit fehlenden Daten .....	168
11.6.4	Statistisches Matching zur Bestimmung der Kontrollgruppen .....	171
<b>12</b>	<b>Ergebnisse.....</b>	<b>175</b>
12.1	Lehrkräfte .....	175
12.2	Schülerinnen und Schüler .....	175
12.2.1	Familiensprache .....	175
12.2.2	Sozioökonomischer Hintergrund.....	176
12.2.3	Leseschwache Kinder zu den einzelnen Messzeitpunkten .....	177
12.2.4	Zu Beginn der Grundschulzeit leseschwache Kinder .....	182
12.2.5	Drop-Outs von Schülerinnen und Schülern.....	184
12.3	Fragestellung 1: Erkennen von Leseschwierigkeiten .....	185
12.3.1	Genannte Auffälligkeiten der Lehrkräfte .....	185
12.3.1.1	Auffälligkeiten allgemein.....	185
12.3.1.2	Positive Auffälligkeiten.....	188
12.3.1.3	Leseauffälligkeit.....	188
12.3.2	Hypothese 1.1 .....	189
12.3.3	Hypothese 1.2 .....	192
12.3.4	Hypothese 1.3 .....	195
12.3.5	Hypothese 1.4 .....	197
12.3.6	Hypothese 1.5 .....	199
12.4	Fragestellung 2: Einfluss der Diagnose auf die Leseentwicklung.....	200
12.4.1	Bestimmen des Lese-Lernzuwachses .....	201
12.4.2	Statistisches Matching über Propensity Scores .....	202
12.4.2.1	Kontrollgruppen für T2-T6 .....	204
12.4.2.2	Kontrollgruppen für T2-T7 .....	207
12.4.2.3	Kontrollgruppen für T3-T6 .....	211
12.4.2.4	Kontrollgruppen für T3-T7 .....	214
12.4.3	Hypothese 2.1 .....	217
12.4.4	Hypothese 2.2 .....	218
12.4.5	Hypothese 2.3 .....	220
12.4.6	Hypothese 2.4 .....	221
12.4.7	Hypothese 2.5 .....	222
12.4.8	Hypothese 2.6 .....	223
12.5	Fragestellung 3: Eingeleitete (Förder-)Maßnahmen bei Auffälligkeiten .....	224

12.5.1	Genannte Maßnahmen.....	224
12.5.1.1	Lesespezifische Maßnahmen.....	225
12.5.1.2	Genannte Maßnahmen für allgemein auffällige Kinder.....	226
12.5.1.3	Genannte Maßnahmen für leseauffällige Kinder.....	226
12.5.1.4	Genannte Maßnahmen für leseschwache Kinder.....	228
12.5.1.5	Schulinterne Fördermaßnahmen.....	229
12.5.2	Hypothese 3.1.....	230
12.5.3	Hypothese 3.2.....	232
12.6	Fragestellung 4: Erfolg der durchgeführten (Förder-)Maßnahmen.....	232
12.6.1	Hypothese 4.1.....	232
12.6.1.1	Erfolg schulinterner Deutschförderung.....	235
12.6.1.2	Erfolg von Förderunterricht Deutsch.....	237
12.6.1.3	Schulinterne Deutschförderung nach statistischem Matching.....	239
12.6.1.4	Erfolg der durchgeführten Maßnahmen.....	241
12.6.2	Hypothese 4.2.....	242
12.6.3	Hypothese 4.3.....	245
12.7	Zusätzliche Analysen.....	246
12.7.1	Urteilsakkuratheit der Lehrkräfte.....	246
12.7.1.1	Kontaktdauer.....	247
12.7.1.2	Ausprägung der Leseschwäche.....	248
12.7.1.3	Rechtschreibleistung.....	249
12.7.2	Mädchen, Jungen und Lesen.....	250
12.7.3	Häufigkeit von außerschulischen Hilfen.....	251
12.7.4	Die Gruppe der ehemals leseschwachen Kinder.....	252
<b>13</b>	<b>Diskussion.....</b>	<b>256</b>
13.1	Erkennen von Leseschwierigkeiten.....	256
13.2	Durch Auffälligkeiten beeinflusste Leseentwicklung.....	258
13.3	Eingeleitete (Förder-)Maßnahmen bei Auffälligkeiten.....	260
13.4	Erfolg der durchgeführten (Förder-)Maßnahmen.....	261
13.5	Limitationen und Anregungen zu weiterer Forschung.....	263
13.6	Praktische Implikationen.....	265
13.7	Fazit und Ausblick.....	268
	<b>Literaturverzeichnis.....</b>	<b>269</b>
	<b>Anhang A: Itemanalysen der Testbatterie von Koch und Georg (2010).....</b>	<b>290</b>
	I. Silbensegmentation.....	290
	II. Satzsegmentation.....	291
	III. Buchstabenkenntnis.....	292
	IV. Laut-Buchstaben-Zuordnung.....	292
	V. Schreiben nach Bildern.....	293
	<b>Anhang B: Beurteilerübereinstimmungen.....</b>	<b>294</b>
	I. Absolute Häufigkeiten.....	294
	II. Beurteilerübereinstimmungen für alle Inhaltskategorien.....	295
	<b>Anhang C: Berechnungen mit GPower.....</b>	<b>298</b>

# 1 Einleitung

„Lesen ist eine universelle Kulturtechnik, die eine Teilhabe am sozialen und kulturellen Leben ermöglicht“ (Spörer, Brunstein & Arbeiter, 2007, S. 298).

Lesen als solches wird in der Grundschule zunächst neben Rechnen und Schreiben explizit eingeübt. Im Verlauf der Sekundarstufe, aber auch darüber hinaus, z. B. in der beruflichen Ausbildung, wird dann häufig vorausgesetzt, dass Schülerinnen und Schüler über eine ausreichende Lesekompetenz verfügen. Ist dies nicht der Fall, werden sie sehr wahrscheinlich Schwierigkeiten haben, die an sie gestellten Anforderungen erfüllen zu können, da unbekannte Texte zunehmend flüssig und sinnentnehmend gelesen werden müssen (Bos, Buddeberg, Bremerich-Vos & Schwippert, 2012, S. 13; Christophel, Baadte, Heyne & Schnotz, 2017; Valtin et al., 2010, S. 78). Ohne ausreichende Lesekompetenz verringern sich somit die Aussichten auf eine gute schulische wie berufliche Ausbildung sowie allgemein auf eine aktive gesellschaftliche Teilhabe (Hurrelmann, 2013, S. 163; Spinner, 2010, S. 48). Dies verdeutlicht die Bedeutsamkeit der Lesekompetenz und weist diese als womöglich bedeutendste Kompetenz aus, die Schülerinnen und Schüler im Laufe der Grundschule in ausreichendem Maße erwerben müssen, um im weiteren Verlauf ihres Lebens selbstgesteuert lernen zu können.

In der IGLU-Studie 2016 wurde aufgezeigt, dass 19 Prozent der Schülerinnen und Schüler im vierten Schuljahr über keine ausreichende Lesekompetenz verfügen (Bos, Valtin, Hußmann, Wendt & Goy, 2017, S. 15). Erste Schwierigkeiten im Leseerwerb zeigen sich bei einem Teil der Kinder oft schon innerhalb der ersten Wochen nach Schuleintritt. Nicht bei allen Schülerinnen und Schülern bleiben diese anfänglichen Schwierigkeiten bestehen, aber wenn dies der Fall ist, besteht die Gefahr, dass sich diese manifestieren. Dies zeigt auch die Relevanz auf, anfängliche Probleme bei Grundschulkindern möglichst sicher zu erkennen – in der Literatur wird hier teilweise von ca. drei Monaten nach Schulbeginn als frühestmöglichem Zeitraum berichtet – und in der Folge mit entsprechenden Unterstützungsangeboten aufzufangen (Fischer & Gasteiger-Klicpera, 2013, S. 63; Klicpera, Schabmann & Gasteiger-Klicpera, 2013, S. 145). Geschieht dies nicht, können sich die anfänglichen Schwierigkeiten im weiteren Verlauf zu Leseschwächen und Lesestörungen entwickeln (Klicpera et al., 2013, S. 149). Die Relevanz einer guten Diagnostik und anschließenden Förderung in der Grundschulzeit zeigt sich daran, dass vorhandene Defizite in der Lesekompetenz im Anschluss an die Grundschulzeit eine hohe Stabilität aufweisen und anscheinend nicht mehr ausgeglichen werden können (Valtin et al., 2010, S. 43). Dies ist allgemeiner auch als „Matthäus-Effekt“ bekannt (Nation, 2005, S. 264; Richter & Müller, in Druck, S. 15). Eine frühzeitige noch „intensiver[e] und systematischer[e]“ Förderung kann dies eventuell verhindern (Fischer & Gasteiger-Klicpera, 2013, S. 63).

In der Grundschulzeit werden folglich die Grundlagen für die spätere Laufbahn der Kinder gelegt und somit sind vor allem die Grundschullehrkräfte – und hier insbesondere die

Lehrkräfte, welche das Schulfach Deutsch unterrichten – besonders gefordert: da sich die Schwierigkeiten im Leseerwerb bereits im Anfangsunterricht zeigen, müssen Grundschullehrkräfte über die Fähigkeit verfügen, solche Schwierigkeiten frühzeitig zu erkennen und darauf folgend mit geeigneten Fördermaßnahmen reagieren. Forschungsergebnissen zufolge fällt ein großer Teil an leseschwachen Grundschulkindern den Lehrkräften aber gar nicht erst auf (Valtin et al., 2010, S. 75–76). Nur ein Teil der leseschwachen Kinder, die auch korrekterweise als leseschwach identifiziert worden waren, erhielt eine Leseförderung in der Schule (Bremerich-Vos et al., 2017, S. 281; Valtin et al., 2010, S. 76).

Es herrscht allerdings weiterhin Forschungsbedarf in diesem Bereich, da es kaum Studien gibt, die systematischer auf die Diagnosefähigkeiten von Lehrkräften in Bezug auf die Lesekompetenz ihrer Schüler und Schülerinnen blicken (Schmidt, 2015, S. 89–90).

Das Ziel dieser Arbeit ist es, die förderdiagnostische Kompetenz von Grundschullehrkräften in Bezug auf die Leseentwicklung von Grundschulkindern zu betrachten. Es soll dadurch herausgearbeitet werden, wie gut Grundschullehrkräfte leseschwache Kinder erkennen und inwiefern die eingeleiteten Unterstützungsmaßnahmen Kinder in ihrer Leseentwicklung fördern.

Die Arbeit besteht aus einem theoretischen und einem empirischen Teil. Zunächst wird im Theorieteil der aktuelle Forschungsstand zu den förderdiagnostischen Kompetenzen der Lehrkräfte herausgearbeitet werden. Hierfür wird in Kapitel 2 Diagnostik im Allgemeinen und förderpädagogische Diagnostik im Speziellen betrachtet, wobei zwischen formaler und informeller Diagnostik unterschieden wird. Ebenso wird der Begriff der semiformalen Diagnostik vorgestellt, die von Grundschullehrkräften häufig durchgeführt wird. In Kapitel 3 wird der Forschungsstand über diagnostische Kompetenzen von Lehrkräften herausgearbeitet. Besonderes Augenmerk liegt auf der Urteilsakkuratheit sowie Faktoren, die diese beeinflussen können.

Im Anschluss wird in Kapitel 4 der Begriff der Lesekompetenz definiert und in Kapitel 5 der ungestörte Leseerwerb bei Grundschulkindern sowie der Leseunterricht in der Grundschule vorgestellt, um darauf aufbauend in Kapitel 6 mögliche Schwierigkeiten hierbei darzustellen. In diesem Zuge werden die Begrifflichkeiten, die für Leseschwierigkeiten verwendet werden, erläutert. Kapitel 7 befasst sich mit den Voraussetzungen und den Vorläuferfertigkeiten des Leseerwerbs, insbesondere der Rolle der phonologischen Bewusstheit und der Benennungsgeschwindigkeit. Ursachen für eine Leseschwäche und mögliche Einflussfaktoren werden in Kapitel 8 behandelt, bevor dann abschließend in Kapitel 9 auf spezifische förderdiagnostische Aspekte und Möglichkeiten zur Förderung von leseschwachen Kindern eingegangen wird.

Der empirische Teil beginnt mit Kapitel 10, in dem das Ziel der Studie sowie die betrachteten Fragestellungen im Detail vorgestellt werden. Kapitel 11 stellt das methodische Vorgehen dar. Es handelt sich um eine Langzeitstudie mit insgesamt sieben Messzeitpunkten, wobei jeweils drei davon in Klasse eins und zwei stattfanden und der siebte Messzeitpunkt als Follow-Up-Erhebung zum Ende der vierten Klasse. Insgesamt haben 281 Grundschul Kinder

in 13 Klassen von 7 Grundschulen an der Erhebung teilgenommen. Da es sich um eine Beobachtungsstudie handelt, konnte keine Randomisierung durchgeführt werden. Zur Auswertung der Fragestellung wurde auf statistisches Matching zurückgegriffen, wobei hierdurch auch die geschachtelte Datenstruktur durch die Datenerhebung im Klassenverband berücksichtigt werden konnte. Die Daten der Kinder wurden über Fragebögen und schulische Leistungstests erhoben. Die förderdiagnostische Kompetenz wurde durch offene Fragen in Interviews mit den Lehrkräften erfasst. Das methodische Vorgehen wird in Kapitel 11 beschrieben. Die Ergebnisdarstellung erfolgt in Kapitel 12. Im letzten Kapitel werden die zentralen Befunde unter Berücksichtigung des aktuellen Forschungsstands zusammengefasst und ein Ausblick auf weitere Forschungsmöglichkeiten sowie praktische Implikationen für Lehrkräfte gegeben.

## 2 Diagnostik im schulischen Setting

Diagnostik gilt als ein wesentlicher Bestandteil der Lehrtätigkeit, insbesondere im Bereich der Grundschule. Die Primarstufe steht vor allem deswegen besonders im Fokus, weil hier die Grundlagen für das Lernen in der weiterführenden Schule, aber auch für lebenslanges Lernen generell gelegt werden sollen: In den ersten Grundschuljahren lernen Kinder lesen, schreiben und rechnen. Diese Fähigkeiten werden ab der Sekundarstufe häufig als ausreichend erworben vorausgesetzt und dort oftmals nicht mehr weiter eingeübt, sodass diejenigen, welche diese Kompetenzen in der Grundschule nicht ausreichend erworben haben, spätestens in der Sekundarstufe Schwierigkeiten haben werden, dem Unterricht ausreichend folgen zu können (Klicpera et al., 2013, S. 149). Grundschullehrkräfte müssen demnach in der Lage sein, diagnostisch Schwierigkeiten im Erwerb dieser Fähigkeiten zu erkennen.

Es sollen hier nun zunächst allgemein die Begrifflichkeiten erörtert werden. Im Anschluss wird auf Anlässe zur Diagnostik eingegangen werden, bevor abschließend auf die diagnostische Ausbildung von Grundschullehrkräften geblickt und weiterführend im nächsten Kapitel die diagnostische Kompetenz von Grundschullehrkräften betrachtet werden wird.

### 2.1 Begriffe und Definitionen

Der Begriff „Diagnostik“ stammt aus dem Griechischen und bedeutet, „etwas gründlich kennenzulernen, zu entscheiden, zu beschließen“ (Mienert & Pitcher, 2011, S. 132).

Da pädagogisch-psychologische Diagnostik Theorien und Methoden der psychologischen Diagnostik beinhaltet, wird zunächst kurz auf diesen Begriff eingegangen (Hesse & Latzko, 2017, S. 59). Schmidt-Atzert und Amelang sehen die psychologische Diagnostik als „Teildisziplin der Psychologie“, welche die Anwendung im Bereich der pädagogischen Psychologie einschließt und definieren sie wie folgt:

Psychologische Diagnostik [...] dient der Beantwortung von Fragestellungen, die sich auf die Beschreibung, Klassifikation, Erklärung oder Vorhersage menschlichen Verhaltens und Erlebens beziehen. Sie schließt die gezielte Erhebung von Informationen über das Verhalten und Erleben eines oder mehrerer Menschen sowie deren relevanter Bedingungen ein. Die erhobenen Informationen werden für die Beantwortung der Fragestellung interpretiert. Das diagnostische Handeln wird von psychologischem Wissen geleitet. Zur Erhebung von Informationen werden Methoden verwendet, die wissenschaftlichen Standards genügen.

(Schmidt-Atzert & Amelang, 2012, S. 4)

Dieser Definition entsprechend liegt zunächst eine Fragestellung vor, die mithilfe von Diagnostik und der hierüber erhaltenen Informationen beantwortet werden soll. Es findet eine „gezielte“ – und somit bewusst durchgeführte – Erhebung statt. Ebenso soll diese mit Methoden durchgeführt werden, die „wissenschaftlichen Standards genügen“; hierauf wird in einem

späteren Abschnitt eingegangen werden. Die psychologische Diagnostik ist allerdings nicht direkt gleichzusetzen mit der pädagogisch-psychologischen Diagnostik. Kritisch diskutiert wird, ob die pädagogisch-psychologische Diagnostik sich aus der psychologischen Diagnostik heraus entwickelt hat oder von Anfang an stets eigenständig gewesen ist (Wilhelm & Kunina-Habenicht, 2015, S. 306). Es zeigen sich allerdings, wie bereits genannt, große Ähnlichkeiten in Bezug auf die im Fokus stehenden Fragestellungen sowie die zugrundeliegenden Theorien und die verwendeten Methoden (Hesse & Latzko, 2017, S. 59; Wilhelm & Kunina-Habenicht, 2015, S. 306).

Ziel der pädagogisch-psychologischen Diagnostik ist „die möglichst feinkörnige Erfassung des Wissens eines einzelnen Schülers oder einer einzelnen Schülerin“ (Maier, 2015, S. 36). Maier (2015, S. 11) hält weiterhin fest, dass die pädagogisch-psychologische Diagnostik stets in einen Kontext eingeordnet werden muss und dabei nie Selbstzweck ist. Ein Merkmal der pädagogisch-psychologischen Diagnostik, welches von Wilhelm und Kunina-Habenicht hervorgehoben wird, ist der starke Fokus auf die „Frage der Veränderbarkeit“ (Wilhelm & Kunina-Habenicht, 2015, S. 306). Es handelt es sich auch hier um einen diagnostischen Prozess, der zur Voraussetzung hat, dass es einen konkreten Anlass gibt. Das Vorhandensein eines Problems, zum Beispiel die Vermutung, dass eine Schülerin oder ein Schüler Schwierigkeiten im Leseerwerb hat, soll überprüft werden und wird somit zur Fragestellung, die es zu beantworten gilt. Weiterhin werden Hypothesen über mögliche Ursachen aufgestellt, welche ebenfalls überprüft werden sollen. Es müssen Methoden ausgewählt werden, die zur Beantwortung der Fragestellung geeignet sind. Hierzu zählen beispielsweise Tests, Gespräche und Beobachtungen. An die Hypothesenprüfung schließt sich das diagnostische Urteil an, welches die Fragestellung beantwortet. Während die psychologische Diagnostik zum Ziel hat, „Entscheidungen sowie Prognosen und deren Evaluation“ vorzubereiten (Mienert & Pitcher, 2011, S. 132), ist eine Diagnose in der pädagogisch-psychologischen Diagnostik nicht gleichzusetzen mit dem Endpunkt der Tätigkeiten (Hesse & Latzko, 2017, S. 65). In der Folge fließen die Ergebnisse dieser Diagnostik in die Handlungen der Lehrkraft ein, wie beispielsweise in das Durchführen von Interventionen zur Förderung des Lesens (Hesse & Latzko, 2017, S. 64–65). Das diagnostische Urteil ist entsprechend in den größeren Kontext der Tätigkeiten einer Lehrkraft eingebettet. Hesse und Latzko (2017, S. 65) sehen auch an dieser Stelle den diagnostischen Prozess noch nicht als beendet an, da die gewählte Intervention auf ihre Effektivität hin überwacht werden muss und es so zu weiterer Diagnostik kommen muss, um den Erfolg der gewählten Intervention sicherzustellen.

Der psychologischen sowie der pädagogisch-psychologischen Diagnostik liegt als Ausgangspunkt demnach eine Fragestellung zugrunde, die es zu beantworten gilt. In der pädagogisch-psychologischen Diagnostik liegt der Fokus allerdings neben der Beantwortung der Fragestellung auch auf den daraus folgenden Handlungen der Lehrkraft. Auf mögliche

Fragestellungen und somit Anlässe zur pädagogisch-psychologischen Diagnostik sowie die Ziele dieser wird im Folgenden eingegangen.

## 2.2 Anlässe und Funktionen von pädagogisch-psychologischer Diagnostik

Die Anlässe und damit verbunden auch die Ziele, die mit den diagnostischen Tätigkeiten verfolgt werden, können sehr unterschiedlich sein. Schrader (2013, S. 155) hält hierzu fest, dass sich in den letzten Jahren eine Wandlung im Fokus der Diagnostetätigkeiten ergeben hat: Während früher vor allem die Einschätzung von Lernergebnissen und damit die Bewertung der Leistungen der Schülerinnen und Schüler im Zentrum der Diagnostetätigkeiten standen, sind nun „die Unterrichtsgestaltung, die Steuerung des Lehr-Lern-Prozesses und die Unterrichtsentwicklung“ im Fokus. Ersteres spielt im Bereich der *Selektionsdiagnostik* bzw. der *Leistungsdiagnostik* eine tragende Rolle. Diese ist dann relevant, wenn es zum Beispiel um die Entscheidung geht, welchen Schulzweig eine Schülerin oder ein Schüler im Anschluss an die Grundschule besuchen soll. Hier wird darauf geachtet, welche Leistungen eine Schülerin oder ein Schüler erbringt und ob diese als ausreichend für den entsprechenden Schulzweig eingeschätzt werden (Hascher, 2008, S. 76; Maier, 2015, S. 41). Maier (2015) verwendet hierfür auch den Begriff der *summativen Leistungsdiagnostik*. Als summativ wird hier die „abschließende Beurteilung des Erfolgs von Lehr- und Lernprozessen“ beschrieben (Maier, 2015, S. 112). Neben dem bereits genannten Beispiel der Übergangsempfehlung im Anschluss an die Grundschule gehören hier auch Abschlusszeugnisse und die Diagnose von Entwicklungsstörungen sowie die Ermittlung, ob Anspruch auf sonderpädagogische Förderung besteht, dazu. Neben der Selektion erfährt hier das Abschlusszeugnis auch die Aufgabe der Zertifizierung für den weiteren beruflichen Werdegang (Maier, 2015, S. 113).

Im Gegensatz dazu steht die *Förderdiagnostik*, welche den Blick auf die einzelne Person und ihren gegenwärtigen Lernstand legt mit dem Ziel, diese optimal zu fordern und zu fördern. Entsprechend Schraders Ausführungen werden auch hier die aktuellen Leistungen von Schülerinnen und Schülern betrachtet. Das Endergebnis ist allerdings keine Beurteilung z. B. in Form einer Schulnote (Maier, 2015, S. 41). Ziel ist es stattdessen, den Leistungsstand von Schülerinnen und Schülern zu erfassen und mit Soll-Zuständen, welche sich z. B. aus curricularen Vorgaben ergeben, zu vergleichen, um die Diskrepanz zu eliminieren oder zu minimieren (Schrader, 2013, S. 155). Der Lernprozess als solcher steht im Mittelpunkt und die erbrachte Leistung soll dem eigentlichen Potential einer Schülerin oder eines Schülers angeglichen werden (Hascher, 2008, S. 76). Hier sollten Lehrkräfte möglichst frühzeitig auftretende Schwierigkeiten bemerken und mit angemessenen Unterstützungsangeboten oder unter Umständen auch einer Anpassung des eigenen Unterrichts reagieren. Im Fokus stehen also nicht nur die schulischen Leistungen, sondern auch daraus folgende mögliche Konsequenzen wie beispielsweise die Anpassung des Unterrichts oder die Nutzung von weiteren Fördermöglichkeiten. Schmidt (2015, S. 92) gibt zum Bereich der förderdiagnostischen Tätigkeiten von Lehrkräften folgendes an: „Lehrkräfte sollen, im Idealfall, prozessorientiert die

Fähigkeiten ihrer Lernenden diagnostizieren und daraus Konsequenzen für die Auswahl von Lehrangeboten ziehen“. Maier (2015) verwendet hier den Begriff der *formativen Leistungsdiagnostik*. Im deutschsprachigen Raum wird dieser Begriff noch nicht sehr lange verwendet, im englischsprachigen Bereich ist er gebräuchlicher (Maier, 2015, S. 121). Dennoch ist diese Art der Diagnostik nicht neu. Maier (2015, S. 122) verweist hier auf die internationale, sonderpädagogische Forschung und die Notwendigkeit einer fortlaufenden Diagnostik, um Lernfortschritte zu überprüfen und bei Bedarf Lehr- und Lernprozesse zu optimieren.

Im Falle der formativen bzw. Förderdiagnostik können sowohl formelle wie auch informelle Diagnostikmethoden Anwendung finden (Maier, 2015, S. 147). Auf diese wird im nächsten Abschnitt eingegangen werden.

### 2.3 Methoden und Gütekriterien der Diagnostik

Eine weitere Unterteilung – neben der in Selektions- und Förderdiagnostik – stellt Hascher (2008, S. 74–75) in Bezug auf den Formalisierungsgrad der gewählten diagnostischen Methoden auf. Sie unterscheidet zwischen formeller, informeller und semiformeller Diagnostik, welcher sich Lehrkräfte bedienen können. Die formelle Diagnostik nutzt „gezielt und systematisch“ empirisch überprüfte Methoden (Hascher, 2008, S. 74). Hierzu gehören beispielsweise standardisierte normierte Leistungstests zur Erfassung von Lese-, Rechtschreib- oder mathematischen Fähigkeiten. Auf der anderen Seite steht die informelle Diagnostik, die häufig aus beiläufigen Beobachtungen sowie Erfahrungen der Lehrkraft besteht (Hascher, 2008, S. 74). Oft sind der Lehrkraft die Vorgänge hierbei nicht bewusst und es liegt demzufolge kein vorangehender Anlass vor. Im ersten Schritt kann eine „grobe Einschätzung“ sicher ausreichend sein, wenn hierdurch generell Schwierigkeiten einer Schülerin oder eines Schülers im Lernprozess auffallen (Hesse & Latzko, 2017, S. 31). Das Problem an dieser Form der Diagnostik liegt aber darin, dass sich verzerrende Faktoren recht schnell negativ auf die Urteilsakkuratheit auswirken können (Hascher, 2008, S. 75).

Schrader und Helmke (2014, S. 46) verwenden eine etwas andere Unterteilung. Die beiden Autoren unterscheiden zwischen expliziten und impliziten Urteilen, wobei die expliziten Urteile nach Erhebung geeigneter Daten und Vergleich mit einer Norm gefällt werden. Als Beispiel ist hier die Klassenarbeit zu nennen, welche als kriteriale Norm das korrekte Lösen einer Mindestanzahl an Aufgaben vorsieht. Bei impliziten Urteilen kommt es dagegen auch zu intuitiven und subjektiven Einschätzungen der Leistung. Diese Art von Urteilen findet man vor allem im Verlauf einer Unterrichtsstunde und sie unterliegt nicht immer der bewussten Kontrolle einer Lehrkraft (Schrader & Helmke, 2014, S. 46).

Hascher (2008, S. 75) bezeichnet die Fehleranfälligkeit bei informeller Diagnostik als „immens“, kritisiert allerdings, dass die Einteilung in formell und informell „zu kurz“ greift. Von ihr stammt die Bezeichnung der dritten Form, der semiformellen Diagnostik (Hascher, 2008, S. 75). Sie definiert diese als eine Zwischenform, zu der jegliche diagnostische Tätigkeiten zählen, die nicht der formellen Diagnostik zugerechnet werden können, die aber „nicht nur zu

impliziten Urteilen“ führt (Hascher, 2008, S. 75). Dies trägt der Tatsache Rechnung, dass Lehrkräfte z. B. im Unterricht durchaus gezielt Schülerinnen und Schüler beobachten, dies aber nicht unter Berücksichtigung formeller Kriterien und Vorgaben tun.

### 2.3.1 Gütekriterien

Jürgens und Lissmann (2015, S. 28) führen an, dass implizite Urteile ihre Legitimation haben, da diese der Lehrkraft erste Hinweise geben und somit zur Einleitung von expliziter Diagnostik führen können. Sie merken aber auch an, dass „prinzipiell ein Anspruch“ besteht, Leistungsbeurteilungen in Schulen „optimal zu gestalten“ (Jürgens & Lissmann, 2015, S. 127). Wenn man sich die erforderlichen Gütekriterien ansieht, wird deutlich, wie formelle Tests das Risiko von Urteilsfehlern minimieren und damit eine möglichst hohe Qualität der durchgeführten Beurteilungen gewährleisten können. Problematisch wird es daher lediglich dann, wenn die implizite bzw. informelle Diagnostik die einzige diagnostische Tätigkeit darstellt oder die Erkenntnisse hieraus über die Ergebnisse der expliziten und formellen Diagnostik gestellt werden (Jürgens & Lissmann, 2015, S. 28).

Gütekriterien sind wichtige Merkmale, die etwas über die Qualität eines diagnostischen Verfahrens aussagen (Schmidt-Atzert & Amelang, 2012, S. 129–130). Weiterhin führen Schmidt-Atzert und Amelang an, dass die Qualität eines diagnostischen Verfahrens nicht nur am Ergebnis, sondern auch am Prozess der Erstellung des Verfahrens gemessen werden kann. Gütekriterien sind folglich grundlegend, um eine gute Diagnostik gewährleisten zu können, weswegen auch hier kurz auf dieses Thema eingegangen wird.

Objektivität, Reliabilität und Validität zählen zu den Hauptgütekriterien, während die Normierung – ein weiteres wichtiges Kriterium – häufig zu den Nebengütekriterien gezählt wird (Jürgens & Lissmann, 2015, S. 128). Schmidt-Atzert und Amelang führen diese vier Kriterien ohne weitere Unterteilung nebeneinander als „die wichtigsten Gütekriterien“ auf (Schmidt-Atzert & Amelang, 2012, S. 131). Daher sollen diese im Folgenden vorgestellt werden.

#### 2.3.1.1 Objektivität

Unter Objektivität versteht man, dass das Testergebnis nicht von der durchführenden Testleitung beeinflusst wird. Objektivität wird, im Gegensatz zu anderen Gütekriterien, meist nicht numerisch bestimmt. Stattdessen werden Maßnahmen vorgeschlagen, welche die Objektivität erhöhen und die im Manual des Verfahrens dokumentiert sein sollen (Schmidt-Atzert & Amelang, 2012, S. 133).

Zum einen beziehen sich diese auf die Durchführung des Verfahrens: Um die sogenannte Durchführungsobjektivität zu gewährleisten, soll es eine sehr genaue Anweisung über die Durchführung geben, welche nicht nur den Ablauf des Verfahrens beschreibt, sondern auch Angaben darüber macht, wie auf Rückfragen zu reagieren ist, welches Material verwendet wird etc. Die Aufgaben sind ebenfalls exakt vorgegeben und das zu verwendende

Testmaterial ist enthalten (Jürgens & Lissmann, 2015, S. 128; Schmidt-Atzert & Amelang, 2012, S. 133–134).

Zudem bezieht sich der Begriff auch auf die Auswertung eines Verfahrens. Um eine möglichst hohe Auswertungsobjektivität zu erreichen, muss genau vorgegeben sein, wie Antworten oder das Verhalten bewertet werden (Jürgens & Lissmann, 2015, S. 128; Schmidt-Atzert & Amelang, 2012, S. 134–135). Bei freier Beantwortung von Fragen kann keine exakte Vorgabe erfolgen, welche Antworten hier als richtig oder falsch bewertet werden sollen. Schmidt-Atzert und Amelang geben an, dass für solche Fragen neben Bewertungsrichtlinien auch Beispiele für korrekte und falsche Antworten genannt werden sollten. Ebenso sollten im Manual Angaben zu finden sein, wie mit Auslassungen umzugehen ist oder wie die getesteten Personen eine gegebene Antwort korrigieren können (Schmidt-Atzert & Amelang, 2012, S. 135).

Nach Auswertung eines Verfahrens erhält man oft Rohwerte. Diese sind für sich genommen nicht interpretierbar. Objektivität bezieht sich entsprechend auch auf die Interpretation der Ergebnisse, die Interpretationsobjektivität, sodass den erzielten Ergebnissen unabhängig von der auswertenden Person die gleiche Bedeutung zugewiesen wird (Jürgens & Lissmann, 2015, S. 128–129; Schmidt-Atzert & Amelang, 2012, S. 136). Das getestete Merkmal, beispielsweise „Leseverständnis“, sollte im Manual klar beschrieben werden. Die Ausprägung des Merkmals sollte anhand von Normen festgestellt werden können (Schmidt-Atzert & Amelang, 2012, S. 136). Auf die Normierung wird in Kapitel 2.3.1.4 eingegangen.

#### 2.3.1.2 Reliabilität

Unter Reliabilität versteht man die Genauigkeit, mit der ein Test misst, wobei hier noch nicht berücksichtigt wird, ob das getestete Merkmal auch das Merkmal ist, welches getestet werden soll (Jürgens & Lissmann, 2015, S. 129; Schmidt-Atzert & Amelang, 2012, S. 137). Es wird unterschieden zwischen verschiedenen Arten der Reliabilität, wobei einige vor allem während der Testkonstruktion relevant sind (Jürgens & Lissmann, 2015, S. 129).

Die Paralleltestreliabilität gibt an, wie ähnlich zwei äquivalente Tests bzw. deren Aufgaben sind. Bei wiederholter Durchführung eines Tests können Erinnerungseffekte nicht ausgeschlossen werden. Daher wird häufig eine zweite Version des Tests mit ähnlichen, aber nicht identischen Aufgaben bevorzugt. Dieser Test sollte dennoch zum gleichen Ergebnis über die Ausprägung des getesteten Merkmals kommen wie der erste Test mit anderen Aufgaben (Jürgens & Lissmann, 2015, S. 129). Da der Konstruktionsaufwand sehr hoch ist, finden sich vergleichsweise wenig diagnostische Verfahren, die über eine Parallelversion verfügen (Schmidt-Atzert & Amelang, 2012, S. 139).

Mit der Retestreliabilität bezeichnet man die Genauigkeit, mit der ein Test zum gleichen Ergebnis kommt, wenn dieser von derselben Person zum zweiten Mal durchgeführt wird. Hier liegt zu beiden Zeitpunkten der gleiche Test vor. Wichtig ist es, hierbei den Zeitabstand zwischen beiden Messungen zu beachten. Ein großer Abstand bedeutet, dass in der Zwischenzeit andere Faktoren wie zum Beispiel Unterricht eine Verbesserung und somit ein besseres

Abschneiden im Test hervorrufen können, während ein zu kurzer zeitlicher Abstand dazu führen kann, dass die Aufgaben erinnert werden können und die Leistung durch diesen Erinnerungseffekt besser wird (Jürgens & Lissmann, 2015, S. 129). Die Retestreliabilität wird über die Korrelation der Testergebnisse einer Stichprobe, welche denselben Test wiederholt durchgeführt hat, errechnet. Da die Stabilität des getesteten Merkmals die Retestreliabilität beeinflusst, ist ein niedriger Wert – abhängig von eben dieser Stabilität – nicht unbedingt Zeichen für einen qualitativ schlechten Test (Schmidt-Atzert & Amelang, 2012, S. 137).

Zwei weitere Reliabilitäten sind nur im Prozess der Testkonstruktion relevant, weshalb diese hier nur kurz angerissen werden: die Split-Half-Reliabilität und die Konsistenzanalyse. Um die Split-Half-Reliabilität zu bestimmen, werden die Testaufgaben halbiert. Die Ergebnisse in beiden Testhälften sollten jeweils gleich sein. Bei der Konsistenzanalyse wird ein Test in so viele Teile geteilt, wie er Aufgaben hat. Hieraus wird dann die Übereinstimmung berechnet (Jürgens & Lissmann, 2015, S. 129–130).

### 2.3.1.3 Validität

Mit der Validität wird angegeben, ob der Test das misst, was er messen soll. Ein Test, der das Leseverständnis eines Kindes messen soll, sollte nicht stattdessen die Konzentrationsfähigkeit eines Kindes erfassen. Schmidt-Atzert und Amelang (2012) bezeichnen die Validität als vergleichbar „mit dem Wirkungsnachweis bei einem Medikament“ (Schmidt-Atzert & Amelang, 2012, S. 143). Da in den Testungen meist Konstrukte erfasst werden sollen, die nur indirekt zugänglich sind, wie z. B. Leseverständnis, aber auch Konzentrationsfähigkeit oder Intelligenz, kann die Validität auf keinem direkten Weg bestimmt werden (Jürgens & Lissmann, 2015, S. 130).

Nach Schmidt-Atzert und Amelang (2012, S. 145) ist ein Verfahren inhaltsvalid, wenn die Items des Verfahrens repräsentativ sind für das Merkmal, welches erfasst werden soll. Die Inhaltsvalidität wird erhoben, indem Experten das diagnostische Verfahren begutachten und daraufhin prüfen (Jürgens & Lissmann, 2015, S. 130). Auch eine Beschreibung über das Vorgehen beim Erstellen des diagnostischen Verfahrens und eine Darstellung des Konzepts, nach dem vorgegangen wurde, kann die Inhaltsvalidität belegen (Schmidt-Atzert & Amelang, 2012, S. 145).

Eine weitere Bestimmung erfolgt über die Kriteriumsvalidität. Hierunter versteht man den „Zusammenhang zwischen Testergebnissen und konkreten Leistungen oder Verhaltensweisen außerhalb der Testsituation“, wobei das Kriterium „für den vorgesehenen Einsatzbereich des Tests relevant sein [muss]“ (Schmidt-Atzert & Amelang, 2012, S. 147). Die prognostische Validität bezeichnet die Vorhersagegültigkeit eines Tests, gemessen anhand eines konkreten Kriteriums: Wie gut kann der Test eine zukünftige Leistung vorhersagen? Im schulischen Bereich kann dies zum Beispiel ein Schuleingangstest sein, der die Schulfähigkeit des Kindes erfassen und entsprechend voraussagen soll, ob das Kind den kommenden schulischen Herausforderungen gewachsen ist. Diese Form der Kriteriumsvalidität ist nur erforderlich, wenn

eine zukünftige Leistung oder ein zukünftiges Verhalten vorausgesagt werden soll (Schmidt-Atzert & Amelang, 2012, S. 147).

Die Konstruktvalidität sagt etwas darüber aus, wie gut ein Test mit einer dahinterstehenden Theorie übereinstimmt (Jürgens & Lissmann, 2015, S. 131). Mit der Konstruktvalidität möchte man entsprechend belegen, dass das Konstrukt, welches erfasst werden soll, auch erfasst wird (Schmidt-Atzert & Amelang, 2012, S. 148). Jürgens und Lissmann führen als Beispiel die Theorie an, dass prüfungsängstliche Schülerinnen und Schüler schlechtere Leistungen erzielen. Ein Test zur Erhebung der Prüfungsangst sollte demnach eine negative Korrelation mit einem Leistungstest erbringen. Wenn dem so ist, liegt eine hohe Konstruktvalidität vor (Jürgens & Lissmann, 2015, S. 131). Im Zusammenhang hierzu stehen die konvergente und die diskriminante Validität. Die Ergebnisse eines Tests, der beispielsweise Leseverständnis messen soll, werden mit einem anderen Test, der das gleiche Konstrukt erfasst, oder auch mit einem Außenkriterium wie der Schulnote in Deutsch korreliert. Das Ergebnis hier ist die konvergente Validität, die möglichst hoch ausfallen sollte, da beide Tests das gleiche Konstrukt erfassen möchten (Jürgens & Lissmann, 2015, S. 130–131). Hier sollte aber darauf geachtet werden, dass die verwendeten Verfahren nicht inhaltsgleiche oder -ähnliche Items enthalten, welche die konvergente Validität allein durch diesen Umstand hoch ausfallen lassen (Schmidt-Atzert & Amelang, 2012, S. 150). Zudem wird der Test mit einem Verfahren korreliert, welches ein anderes Konstrukt, das nicht mit dem des Tests zusammenhängen sollte, beispielsweise ein Mathematiktest, erfasst. Das Ergebnis hier ist die diskriminante Validität, ebenfalls berechnet als Korrelation, welche nach Erwartung null sein sollte (Jürgens & Lissmann, 2015, S. 130–131). Schmidt-Atzert und Amelang fordern hierzu, dass die konvergente Validität „deutlich höher“ ausfallen muss als die diskriminante Validität (Schmidt-Atzert & Amelang, 2012, S. 149).

#### 2.3.1.4 Normierung

Erzielt eine Schülerin beispielsweise 38 von 50 möglichen Punkten in einem Test zur Erfassung des Leseverständnisses, so sagt dieser Wert nicht viel aus. Ob diese Leistung als gut, durchschnittlich oder schwach einzuschätzen ist, lässt sich nicht feststellen. Erst im Vergleich kann eine Lehrkraft das Ergebnis entsprechend interpretieren. Daher sollten für die Individualdiagnostik Normierungen vorliegen (Schmidt-Atzert & Amelang, 2012, S. 164). Wenn anhand einer Normstichprobe festgestellt wird, dass 86 Prozent aller gleichaltrigen Schülerinnen und Schüler ein gleich gutes oder ein schlechteres Ergebnis erzielen, hat die Schülerin hier entsprechend einen Prozentrang von 86 erreicht und somit ein überdurchschnittliches Ergebnis. Ein Prozentrang von 50 bezeichnet den Durchschnittswert. Der Prozentrang ist rangskaliert und hat daher den entscheidenden Nachteil, dass sich keine Rechenoperationen mit ihm durchführen lassen (Jürgens & Lissmann, 2015, S. 132). Alternativ kann man Standardnormen verwenden, welche sich auf die Gaußsche Normalverteilung beziehen (Jürgens & Lissmann, 2015, S. 132; Schmidt-Atzert & Amelang, 2012, S. 165). Es gibt keine Vorgabe,

welche Normwerte verwendet werden sollen, sodass zahlreiche Normen parallel vorliegen (Schmidt-Atzert & Amelang, 2012, S. 165–166). Die z-Skala reicht von -4 bis +4 bei einem Mittelwert von 0 und einer Standardabweichung von 1. Diese Skala ist allerdings wenig anschaulich, weswegen sie eher selten Verwendung findet (Jürgens & Lissmann, 2015, S. 132). Für Intelligenztests wird häufig die IQ-Skala angegeben mit einem Mittelwert von 100 und einer Standardabweichung von 15. Dies geht auf Wechslers „Abweichungs-Intelligenz-Quotienten“ zurück (Wechsler, 1958, S. 36). T-Werte mit einem Mittelwert von 50 und einer Standardabweichung von 10 finden sich häufig in Persönlichkeitstests, aber auch in schulischen Leistungstests. Diese Normwerte können leicht von einer in eine andere Skala übertragen werden (Schmidt-Atzert & Amelang, 2012, S. 166).

### 2.3.2 Weitere Gütekriterien

Neben diesen zuvor genannten Gütekriterien gibt es noch zahlreiche weitere, die bei der Konstruktion, aber auch bei der Auswahl vor der Durchführung Beachtung finden sollten. So sollte ein Test beispielsweise fair sein und keine Gruppe (z. B. Mädchen oder Jungen, Kinder mit Migrationshintergrund) benachteiligen. Auch sollte das diagnostische Verfahren möglichst ökonomisch sein. Die Durchführung und Auswertung sollten nicht unverhältnismäßig lange dauern oder sehr kompliziert sein (Jürgens & Lissmann, 2015, S. 133).

## 2.4 Anforderungen an Lehrkräfte

Um Diagnostik durchführen zu können, muss eine Lehrkraft folglich in der Lage sein, abhängig vom Untersuchungsanlass und der zu beantwortenden Fragestellung, den diagnostischen Prozess zu planen und im Zuge dessen geeignete diagnostische Verfahren auszuwählen. Hierzu benötigt sie neben dem fachlichen Wissen auch entsprechend diagnostisches Wissen. Standardisierte Verfahren verlangen von der durchführenden Person entsprechend fundierte Kenntnisse, um die Eignung eines diagnostischen Verfahrens einschätzen und anschließend auch anwenden zu können (Hascher, 2008, S. 75; Hesse & Latzko, 2017, S. 52). Hascher (2008, S. 75) führt ebenfalls an, dass die Schule entsprechende Rahmenbedingungen gewähren muss, damit formelle Diagnostik stattfinden kann und kommt zu dem Schluss, dass beide Voraussetzungen im Alltag von Lehrkräften nicht unbedingt gegeben sind, während die informelle Diagnostik dagegen einfacher umzusetzen ist.

Eine Lehrkraft muss ebenfalls prüfen können, ob für den gegebenen Anlass eine standardisierte Testung zielführend ist, da diese sich häufig auf übergreifende Kompetenzen beziehen, oder ob nicht ein anderes, eventuell auch selbst erstelltes Verfahren geeigneter ist. Dies ist zum Beispiel der Fall, wenn eine Lehrkraft den Wissensstand am Ende einer unterrichteten Lerneinheit überprüfen möchte (Maier, 2015, S. 64). Weiterhin hält Maier (2015, S. 65) fest, dass häufig adaptive Verfahren benötigt werden, da es sich bei Bildungsprozessen nicht um die Erfassung eines stabilen Merkmals handelt, sondern um Wissenszuwachs und somit ein „fest definiertes Set aus einzelnen Testaufgaben“ nicht passend ist.

Es müssen stets auch die Rahmenbedingungen, innerhalb derer Diagnostik stattfinden kann, betrachtet werden. Da Unterricht in Gruppen stattfindet, hat die Lehrkraft selten Zeit, ausführliche Individualdiagnostik zu betreiben. Entsprechende in Klassensystemen durchzuführende Gruppentests sind hier eine Alternative (Maier, 2015, S. 36).

Eine Lehrkraft muss auch über fachdidaktisches Wissen in Bezug auf die zu diagnostizierenden Inhalte verfügen (Maier, 2015, S. 19–20). Maier führt als Beispiel die Lesediagnostik an. Hier müssen die Stufen des Schriftspracherwerbs bekannt sein. Ebenso muss eine Lehrkraft auch wissen, „welche Teilfertigkeiten beim Lesen ineinandergreifen“, um ein diagnostisches Verfahren sinnvoll nutzen zu können (Maier, 2015, S. 20).

## 2.5 Kurze Darstellung der Diagnostikausbildung von Lehrkräften

Wie in den vorangegangenen Abschnitten deutlich wurde, benötigt eine Lehrkraft neben ihrem Fachwissen auch ein umfangreiches diagnostisches Grundlagenwissen, um geeignete diagnostische Verfahren auswählen, anwenden und interpretieren zu können. Daher wird nun ein Blick auf die Ausbildung von Lehrkräften geworfen.

Die Ausbildung von Lehrkräften unterscheidet sich je nach Bundesland, es gibt aber übergreifende bzw. grundlegende Regelungen von der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. 2004 wurden von der Kultusministerkonferenz (KMK) Standards für die Lehrerbildung festgelegt, welche 2014 im Zuge der Inklusion aktualisiert wurden (KMK, 2014a, 2014b). Diese Standards geben Kompetenzen an, welche im Verlauf der Ausbildung sowie der Fort- und Weiterbildung von den (angehenden) Lehrkräften erworben werden sollen. Somit bilden die Standards die Grundlage für die Inhalte der Ausbildung von Lehrkräften (KMK, 2014a, S. 2). Hier werden diagnostische Kompetenzen als eine Fähigkeit aufgeführt, über die Lehrkräfte verfügen müssen, um erfolgreich unterrichten zu können (KMK, 2014a, S. 3).

Zunächst muss in der Lehrkräfteausbildung zwischen der ersten Phase, der universitären Ausbildung, und der zweiten Phase, dem Vorbereitungsdienst, unterschieden werden. Laut den Standards für die Lehrerbildung enthalten beide Phasen theoretische und praktische Anteile mit jeweils unterschiedlicher Gewichtung (KMK, 2014a, S. 4), wobei der ersten Phase ein stärkerer Anteil an Theorie und der zweiten Phase der Schwerpunkt in der Praxis zugrunde liegt. Zusätzlich werden nach Abschluss der Ausbildung Fort- und Weiterbildungen als dritte Phase bezeichnet. Diese werden nicht weiter explizit angesprochen, sollen aber ebenso berücksichtigt sein, da die Kompetenzen als „Ziele des lebenslangen Lernens im Lehrerberuf“ angesehen werden (KMK, 2014a, S. 4). Insgesamt werden in den Standards Kompetenzen genannt, welche in vier Kompetenzbereiche mit insgesamt elf Kompetenzen unterteilt sind: Unterrichten, Erziehen, Beurteilen und Innovieren (KMK, 2014a, S. 7–14).

Der dritte Kompetenzbereich „Beurteilen“ fokussiert auf diagnostische und beratende Tätigkeiten und unterteilt sich in zwei Kompetenzen. Die erste Kompetenz in diesem Bereich wird wie folgt umschrieben: „Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und

Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern“ (KMK, 2014a, S. 11). In den dazugehörigen Standards werden die „Grundlagen der Lernprozessdiagnostik“ als im Rahmen der theoretischen Ausbildung zu erwerbende Kenntnisse explizit genannt. Ebenso explizit erwähnt wird die Kooperation im Zuge der Diagnostik und Förderung sowie der Beratung. Als Kooperationspartner werden Kolleginnen und Kollegen sowie externe Einrichtungen und andere Professionen genannt (KMK, 2014a, S. 11). Die zweite Kompetenz in diesem Bereich fokussiert stärker auf die Leistungsbeurteilung: „Lehrerinnen und Lehrer erfassen die Leistungsentwicklung von Schülerinnen und Schülern und beurteilen Lernen und Leistungen auf der Grundlage transparenter Beurteilungsmaßstäbe“ (KMK, 2014a, S. 12). In den zu dieser Kompetenz aufgelisteten Standards wird die Kenntnis unterschiedlicher „Formen und Wirkungen der Leistungsbeurteilung“ genannt sowie die Kenntnis verschiedener Bezugssysteme, um Leistungen beurteilen zu können. Ebenso sollen Lehrkräfte die Ergebnisse der Leistungsüberprüfungen auch nutzen, um den eigenen Unterricht anzupassen.

Auch wenn in den Standards für Lehrerbildung explizit von Diagnostik die Rede ist, geben Hesse und Latzko (2017, S. 52) an, dass sie in Lehrerfortbildungen häufig gefragt werden, ob Lehrkräfte explizit testen und diagnostizieren dürfen. Hier scheint es diesbezüglich große Unsicherheiten zu geben. Sie führen allerdings nicht an, ob diese Unsicherheit bei Lehrkräften aus allen Schulzweigen vorkommt. Einzelne Bundesländer sehen allerdings ein Verbot von Intelligenz- und Persönlichkeitsdiagnostik für Lehrkräfte vor (Hesse & Latzko, 2017, S. 53). Schulische Leistungstests sind hiervon jedoch nicht betroffen. Hesse und Latzko (2017, S. 53) führen weiterhin an, dass Förderschullehrkräfte im Studium eine umfangreichere diagnostische Ausbildung erhalten als angehende Lehrkräfte an allgemeinbildenden Schulen. Die Autorinnen kommen zu dem Schluss, dass eine verstärkte diagnostische Tätigkeit von Regelschullehrkräften wünschenswert ist und eine Kooperation zwischen Lehrkräften und ebenso mit Schulpsychologinnen und Schulpsychologen die Auswertungs- und Interpretationsobjektivität zusätzlich erhöhen kann (Hesse & Latzko, 2017, S. 55).

Abs (2006) berichtet von einer Erhebung in Hessen über die Selbsteinschätzung von angehenden Lehrkräften in Bezug auf ihre diagnostischen Kompetenzen. Befragt wurden Referendarinnen und Referendare aus je fünf hessischen Studienseminaren für das Lehramt am Gymnasium und für das Lehramt an Real-, Haupt-, Grund- sowie Förderschulen. Es wurden Aussagen von 1102 Referendarinnen und Referendaren berücksichtigt, welche ihre Kompetenz in fünf Standards zur Diagnostik selbst einschätzen sollten (Abs, 2006, S. 228). Diejenigen, die das Lehramt am Gymnasium studiert hatten, schätzten ihre diagnostische Kompetenz signifikant niedriger ein als Referendarinnen und Referendare für das Lehramt an Real-, Haupt-, Grund- sowie Förderschulen. Eine Unterscheidung zwischen den letztgenannten Gruppen ist nicht möglich, da die Daten hier nicht getrennt erhoben wurden. Auf den Forschungsstand zur diagnostischen Kompetenz von Lehrkräften wird ausführlich im nächsten Kapitel eingegangen werden.

## 2.6 Zwischenfazit

Zum Alltag aller Lehrkräfte gehören viele diagnostische Tätigkeiten. Lernfortschritte müssen überwacht werden und Lernstände müssen festgehalten werden. Hier sind nicht nur die formelle Diagnostik zu nennen, wie sie beispielsweise auch durch hinzugezogene Förderschullehrkräfte durchgeführt werden kann, sondern auch alltägliche informelle und semiformelle diagnostische Tätigkeiten, die unter anderem dazu dienen, den eigenen Unterricht an die Schülerinnen und Schüler anpassen zu können.

Auch wenn in den Bildungsstandards der Erwerb diagnostischer Fähigkeiten von zukünftigen Lehrkräften gefordert und vorausgesetzt wird, unterscheiden sich die Inhalte der universitären Ausbildung zum Teil sehr stark, wobei Förderschullehrkräfte entsprechend ihrer Tätigkeit den größten Teil an diagnostischer Ausbildung erhalten.

Im Folgenden soll nun entsprechend der Frage nachgegangen werden, was bisher aus der Forschung über die diagnostischen Kompetenzen von Lehrkräften – und hier insbesondere von Grundschullehrkräften – bekannt ist und wie akkurat sie ihre Schülerinnen und Schüler hinsichtlich ihrer Stärken und Schwächen einschätzen können.

### 3 Diagnostische Kompetenzen von Lehrkräften

Die Bedeutsamkeit diagnostischer Tätigkeiten in der Grundschule hat zur Folge, dass der Frage nach dem Stand der diagnostischen Kompetenz von Grundschullehrkräften ein ganz besonderes Augenmerk gilt. Die Anzahl an Forschungsarbeiten auf diesem Gebiet ist dagegen lange Zeit relativ überschaubar geblieben (Helmke, 2015, S. 129; Hesse & Latzko, 2009, S. 31). In den letzten Jahren sind die diagnostischen Kompetenzen von Lehrkräften allerdings stärker in den Fokus der Forschung gerückt, da z. B. bei den PISA- aber auch bei den IGLU-Untersuchungen die Relevanz dieser Thematik deutlich wurde, als sich zeigte, dass Lehrkräfte zuweilen nicht bemerken, dass ihre Schülerinnen und Schüler Schwierigkeiten haben – hier vor allem im Bereich des Schriftspracherwerbs (Artelt, Stanat, Schneider & Schiefele, 2001, S. 119–120; Valtin et al., 2010, S. 76).

Im Folgenden wird zunächst eine Begriffsbestimmung der diagnostischen Kompetenz erfolgen, bevor der aktuelle Forschungsstand zu diagnostischen Kompetenzen von Lehrkräften herausgearbeitet und der Einfluss dieser auf schulische Leistungen erörtert wird. Das Kapitel schließt mit offenen Fragen ab.

#### 3.1 Zum Begriff der diagnostischen Kompetenz

Diagnostische Kompetenz wird von Artelt und Gräsel (2009, S. 157) als eine „Schlüsselkompetenz“ von Lehrkräften bezeichnet. Sie definieren diese als „die Kompetenz von Lehrkräften [...], Merkmale ihrer Schülerinnen und Schüler angemessen zu beurteilen und Lern- und Aufgabenanforderungen adäquat einzuschätzen“. Die Fähigkeit zur Einschätzung alleine reicht aber noch nicht aus: Lorenz und Artelt führen weiter an, dass neben der Einschätzung der schulischen Leistungen ebenso die „Auswahl von geeignetem Unterrichtsmaterial und die Qualität von Maßnahmen der Binnendifferenzierung abhängig von der Güte diagnostischer Urteile“ (Lorenz & Artelt, 2009, S. 212) sind. Kaiser und Möller fassen es ähnlich zusammen und nennen die „Urteilsgenauigkeit bezüglich Personen- und Aufgabenmerkmalen“ als Kern der diagnostischen Kompetenz, wobei auch sie die „Anpassung des Unterrichts an die Lernvoraussetzungen der Schülerinnen und Schüler“ als Teil der diagnostischen Kompetenz sehen (Kaiser & Möller, 2017, S. 56). Hier wird also der Blick nicht nur auf die leistungsdiagnostische Komponente gelegt, sondern auch auf die förderdiagnostische, welche die Optimierung des eigenen Unterrichts einschließt.

Nach Karst (2012, S. 25) wird das Konstrukt der „judgment accuracy“ im angloamerikanischen Raum lediglich theoretisch begründet, während im deutschsprachigen Raum eine differenziertere Auseinandersetzung mit der Thematik stattfindet. Hier wird, ebenfalls anders als im angloamerikanischen Raum, auch darauf hingewiesen, dass es nicht „die eine“ diagnostische Kompetenz gibt: Bei Lehrkräften, welche Schülerinnen und Schüler aufgrund mehrerer Merkmale wie Intelligenz, Ängstlichkeit und Leistungsmotivation einschätzen sollten, konnten keine signifikanten Zusammenhänge zwischen den Akkuratheitsindikatoren dieser Einschätzungen gefunden werden (Spinath, 2005, S. 93). Vielmehr handelt es sich hierbei

daher um mehrere Teilkompetenzen (Karst, 2012, S. 25; Ohle & McElvany, 2015, S. 5; Spinath, 2005). Allerdings muss erwähnt werden, dass auch schon Coladarci (1986, S. 144) zu dem Ergebnis kam, dass die Urteilsakkuratheit von Lehrkräften je nach Subtest schwankt, sodass hier bereits vermutet werden konnte, dass es sich nicht um eine einzelne Kompetenz handelt.

Weinert (2000, S. 14–15) nennt vier Basiskompetenzen, die für eine erfolgreiche Lehrkraft notwendig sind: Sachkompetenzen, diagnostische Kompetenzen, didaktische Kompetenzen und Klassenführungskompetenzen. Bereits zu diesem Zeitpunkt findet sich im deutschsprachigen Raum folglich die Verwendung des Plurals für die diagnostischen Kompetenzen. Er definiert diese als „Bündel von Kompetenzen“, welche es der Lehrkraft ermöglichen, „Lernfortschritte und die Leistungsprobleme der einzelnen Schüler sowie die Schwierigkeiten verschiedener Lernaufgaben im Unterricht fortlaufend beurteilen zu können, sodass das didaktische Handeln auf diagnostischen Einsichten aufgebaut werden kann“ (Weinert, 2000, S. 14). Gleichzeitig bezeichnet er diesen Bereich aber auch als „vielfach defizitär“ (Weinert, 2000, S. 14). Es lassen sich zudem nach Karst (2012, S. 97; 212) mehrere voneinander unabhängige diagnostische Kompetenzen finden, wie beispielsweise aufgabenbezogene, personenbezogene oder personenspezifische diagnostische Kompetenzen. Aktuell halten auch Hesse und Latzko (2017, S. 42–43) in einem Fazit fest, dass, ausgehend von Forschungen von Spinath (2005) sowie Weinert und Schrader (1986), „der Begriff ‚diagnostische Kompetenz‘ im Plural verwendet werden [sollte]“.

Weiterhin sind diagnostische Kompetenzen kontext- und somit fachspezifisch (Karst, 2012, S. 86; Lorenz & Artelt, 2009, S. 218). Dies lässt sich dadurch erklären, dass Aufgaben in ihrer Schwierigkeit eingeschätzt werden müssen und den Lehrkräften gleichzeitig bewusst sein sollte, was Schülerinnen und Schüler jeweils an fachspezifischem Wissen und Kompetenzen benötigen, um die Aufgaben erfolgreich bewältigen zu können (Lorenz & Artelt, 2009, S. 212). In ihrer Studie konnten Lorenz und Artelt einen Zusammenhang zwischen der Urteilsakkuratheit von Grundschullehrkräften, erfasst über die Beurteilung der Lesekompetenz, und der Einschätzung des Wortschatzes der Kinder durch diese Lehrkräfte finden. Ein Zusammenhang mit ihrer Urteilsakkuratheit im mathematischen Bereich fand sich dagegen nicht (2009, S. 218). Alle Lehrkräfte unterrichteten in der Primarstufe sowohl Deutsch als auch Mathematik. Lehrkräfte mit hoher diagnostischer Kompetenz in einem Bereich zeigten hier also nicht unbedingt auch eine hohe diagnostische Kompetenz in einem anderen Bereich.

Schrader (2013, S. 154) bezeichnet die diagnostische Kompetenz bei Lehrkräften als „die Fähigkeit, die im Lehrberuf anfallenden diagnostischen Aufgabenstellungen erfolgreich zu bewältigen, und [...] die Qualität der dabei erbrachten Diagnoseleistungen“. Hier finden sich folglich auch zwei Komponenten: die Diagnostik an sich als Aufgabe der Lehrkraft und gleichzeitig die Qualität – also das Ergebnis – dieser Diagnostik.

Abs (2006, S. 222) schlägt vor, die Fähigkeit zur Umsetzung der didaktischen Konzepte als Komponente der diagnostischen Kompetenz im Sinne der Förderdiagnostik aufzunehmen.

Demnach sollte man diagnostische Kompetenzen bei Lehrkräften differenziert betrachten und Unterrichtsfächer oder auch zu erwerbende Fertigkeiten wie Lesen oder Rechnen unterscheiden. Er begründet dies unter anderem damit, dass die aufgabenspezifische Diagnostik, in dem Sinne, dass Aufgaben in ihrer Schwierigkeit für Schülerinnen und Schüler eingeschätzt werden müssen, auch zur diagnostischen Kompetenz gezählt wird.

Folglich sollten diagnostische Kompetenzen inhaltsbezogen betrachtet werden und im Sinne der Förderdiagnostik auch didaktische Konzepte einbeziehen. Im Folgenden soll nun herausgearbeitet werden, inwieweit Lehrkräfte – insbesondere Grundschullehrkräfte – über diagnostische Kompetenzen verfügen und wie sich diese auf die Leistungen von Schülerinnen und Schülern auswirken.

### 3.2 Historischer Abriss zu den Anfängen des Forschungsgebiets

Der Beginn der Forschung zu diagnostischen Urteilen von Lehrkräften lässt sich in den 1980ern und zu Beginn der 90er Jahre ausmachen, als mehrere Forschergruppen sich mit dieser Thematik auseinandersetzten. Für den amerikanischen Raum sind hier vornehmlich Coladarci und Hoge zu nennen (Coladarci, 1986; Hoge, 1983; Hoge & Coladarci, 1989). Im deutschsprachigen Raum haben vor allem Schrader und Helmke früh schon zu diesem Thema geforscht (Schrader & Helmke, 1987). Coladarci (1986, S. 141–142) beispielsweise betrachtete in einer Studie, wie gut Lehrkräfte Vorhersagen über die Leistung ihrer Schülerinnen und Schüler in einzelnen Tests machen konnten. Das Ergebnis zeigte, dass die Vorhersagen generell recht zutreffend waren, dies aber für die guten Schülerinnen und Schüler stärker zutrifft als für die schwachen. Er kritisierte, dass in vorhergehenden Studien oft nach zukünftig zu erwartender Leistung gefragt wurde, was seiner Meinung nach nicht das diagnostische Urteil erfasst, bei welchem nach dem gegenwärtigen Leistungsstand gefragt wird. Weiterhin merkte er an, dass Lehrkräfte die Leistungen ihrer Schülerinnen und Schüler häufig in eine Rangfolge bringen oder ihnen eine Schulnote geben sollten, diese zusammenfassenden Angaben aber nicht den eigentlichen Wissensstand der Lehrkräfte über ihre Schülerinnen und Schüler darstellen (Coladarci, 1986, S. 141–142).

Die von Schrader und Helmke (1987, S. 27) untersuchten Hauptschullehrkräfte überschätzten die Leistungsstreuung in ihren Klassen sowie das Leistungsniveau. Die Autoren unterschieden in ihrer Studie zwischen der Niveau-, der Differenzierungs- und der Vergleichskomponente des diagnostischen Urteils – eine Unterteilung, die sich auch heute noch häufig findet (zum Beispiel Karing, Matthäi & Artelt, 2011; Lorenz & Artelt, 2009; Oerke, McElvany, Ohle, Ullrich & Horz, 2016). Unter der Niveauebene versteht man die „Übereinstimmung zwischen Urteils- und tatsächlichem Leistungsniveau“ (Schrader, Helmke, Hosenfeld, Halt & Hochweber, 2006, S. 267). Dies wird häufig als Urteilsakkuratheit oder Urteilsgenauigkeit bezeichnet. Die Differenzierungskomponente bezieht sich auf die Fähigkeit, die Heterogenität von Aufgaben korrekt einzuschätzen. Die Vergleichskomponente – auch Rangordnungskomponente genannt – bezieht sich dagegen auf die Einschätzung der Aufgabenschwierigkeiten

und die Fähigkeit, diese in einer Rangordnung korrekt einschätzen zu können (Schrader et al., 2006, S. 267–268).

Mit den Ergebnissen der PISA-Studie von 2000 (Baumert, Klieme et al., 2001) zeigte sich die Relevanz des Forschungsbereichs um die diagnostischen Kompetenzen von Lehrkräften vor allem beim Erwerb der Lesekompetenz im deutschen Raum sehr deutlich. So wurde hier im Bereich der Sekundarstufe I eine mangelnde diagnostische Fähigkeit bei den Lehrkräften festgestellt. Diese unterschätzten den Anteil an leseschwachen 15-jährigen Schülerinnen und Schülern ungemein: Knapp 89 Prozent der Schülerinnen und Schüler, auf die dies zutrifft, wurden von den Lehrkräften nicht als leseschwach eingestuft (Artelt et al., 2001, S. 119–120). Die IGLU-Studie 2006 zeigte im Anschluss daran, dass die Probleme in der Einschätzung der Leseleistung auch bei Grundschullehrkräften zu beobachten sind (Valtin et al., 2010, S. 76–77). Auf diese Ergebnisse wird in Kapitel 3.3.1 genauer eingegangen werden.

### 3.3 Aktueller Stand der Forschung

Im Forschungsbereich um die diagnostischen Kompetenzen von Lehrkräften wurde sehr häufig die Akkuratheit der Lehrerurteile untersucht (Coladarci, 1986; Machts, Kaiser, Schmidt & Möller, 2016; Südkamp, Kaiser & Möller, 2012). Diese bildet die Grundlage, um Schwierigkeiten von Schülerinnen und Schülern adäquat erkennen zu können. Deswegen wird im Folgenden auch zunächst dieser Forschungsbereich eingehend betrachtet, bevor Einflussfaktoren auf die Urteilsakkuratheit sowie der Einfluss von diagnostischen Kompetenzen auf die Leistungen der Schülerinnen und Schüler betrachtet werden.

#### 3.3.1 Akkuratheit von Lehrkräfteurteilen

Sehr häufig findet sich in der deutschsprachigen Literatur der Begriff der Urteilsgenauigkeit, während im englischsprachigen Raum von „judgment accuracy“ die Rede ist. Praetorius und Südkamp (2017) kritisieren dabei den deutschen Begriff, da „Genauigkeit“ häufig mit Reliabilität gleichgesetzt wird, was an dieser Stelle irreführend ist. Daher bevorzugen sie den Begriff der Urteilsakkuratheit (Praetorius & Südkamp, 2017, S. 13). Urteilsakkuratheit bedeutet hierbei eine Übereinstimmung zwischen Lehrkrafturteil und dem zu beurteilenden Merkmal (erhoben mit Tests und/oder Fragebögen) auf Seite der Schülerinnen und Schüler (Praetorius & Südkamp, 2017, S. 13).

In der in Kapitel 3.2 bereits kurz dargestellten Studie von Coladarci (1986, S. 144) gab es eine Übereinstimmung von 75 Prozent zwischen den Lehrerurteilen mit der erbrachten schulischen Leistung. Er vergleicht dieses Ergebnis mit einer Studie von Leinhardt (1983), welche eine etwas niedrigere Urteilsakkuratheit von 64 Prozent erhalten hatte. Leinhardt hatte ihre Untersuchung allerdings im Bereich der „special education needs“, also mit lernschwachen Schülerinnen und Schülern, durchgeführt, was Coladarci wiederum für konsistent mit seinen Ergebnissen hält, da er auch zu dem Ergebnis kam, dass die Urteilsakkuratheit von Lehrkräften bei schwachen Schülerinnen und Schülern sinkt (Coladarci, 1986, S. 144). In einer

neueren Studie zeigte sich das gleiche Ergebnis: Feinberg und Shapiro (2009) ließen 74 Grundschullehrkräfte (in den Klassen zwei bis fünf in den USA) die Leseflüssigkeit von durchschnittlichen sowie von schwachen Schülerinnen und Schülern vorhersagen. Aus jeder Klasse wurden hierfür zwei Schülerinnen und Schüler (je ein Kind war durchschnittlich im Lesen und eins war unterdurchschnittlich) ausgewählt. Die Lehrkräfte wurden über das Auswahlverfahren nicht informiert. Schülerinnen und Schüler mit unzureichenden Englischkenntnissen oder festgestelltem Förderbedarf wurden nicht in die Studie einbezogen (Feinberg & Shapiro, 2009, S. 455). Die Urteilsakkuratheit der Lehrkräfte war moderat. Insgesamt wurden die Leseleistungen der Schülerinnen und Schüler – sowohl der durchschnittlichen als auch der unterdurchschnittlichen – überschätzt (Feinberg & Shapiro, 2009, S. 459). Weiterhin zeigte sich, dass die Akkuratheit für die Einschätzungen der durchschnittlichen Schülerinnen und Schüler höher war als für die unterdurchschnittlichen (Feinberg & Shapiro, 2009, S. 460). Machts, Kaiser, Schmidt und Möller (2016) führten eine Metaanalyse zur Urteilsakkuratheit von Lehrkräften in Bezug auf kognitive Fähigkeiten ihrer Schülerinnen und Schüler wie beispielsweise Intelligenz oder auch Hochbegabung durch. In die Metaanalyse flossen Daten aus 33 Studien aus Deutschland, den USA und weiteren Ländern ein. Es zeigten sich in den Ergebnissen keine signifikanten Unterschiede zwischen den Ländern. Insgesamt konnten sie eine mittelhohe Urteilsakkuratheit von Lehrkräften nachweisen, wobei diese in Bezug auf Intelligenz etwas höher ausfiel als in den Bereichen Kreativität und Hochbegabung (Machts et al., 2016, S. 99). Hier wurden allerdings keine Urteile zu schulischen Leistungen erfasst. Eine Metaanalyse von Südkamp et al. (2012) konzentrierte sich dagegen auf diesen Bereich, insbesondere auf die mathematischen und schriftsprachlichen Leistungen. Untersuchungen, die sich auf andere Eigenschaften wie z. B. Motivation konzentrierten, wurden ausgeschlossen. Kriterium zum Einbezug einer Untersuchung war unter anderem, dass es eine englischsprachige Zusammenfassung gab, da die Studien mit englischsprachigen Begriffen gesucht wurden. In die Auswertung gingen schließlich 73 Studien aus verschiedenen Ländern ein. Auch hier gab es keine signifikanten Unterschiede zwischen den Ländern. Es wurde eine Effektstärke von  $r = 0.63$  in Bezug auf die Urteilsakkuratheit für schulische Leistungen ermittelt, welche als relativ hoch bezeichnet wird. Es konnte weiterhin kein Unterschied in Bezug zur Höhe der Urteilsakkuratheit für den mathematischen und den schriftsprachlichen Bereich gefunden werden (Südkamp et al., 2012, S. 755).

Diese beiden Metaanalysen zeigen, dass Lehrkräfte über diagnostische Kompetenzen verfügen und schulische Leistungen sowie kognitive Fähigkeiten ihrer Schülerinnen und Schüler recht gut einschätzen können, die Akkuratheit ihrer Einschätzungen allerdings je nach Bereich auch unterschiedlich hoch ausfallen kann. Weiterhin muss einschränkend festgehalten werden, dass es zum Teil sehr große Schwankungen zwischen den Lehrkräften in der Urteilsakkuratheit gibt (Leuders, Leuders & Philipp, 2014, S. 732; Lorenz & Artelt, 2009, S. 216). Leuders et al. (2014, S. 732) nennen Streuungen zwischen 0.3 und 0.9 als typisch.

Wenn diagnostische Kompetenz in dem Sinne erhoben wird, dass Lehrkräfte ihren Schülerinnen und Schülern eine Schulnote für deren Leistung geben sollen und diese mit der in einem Leistungstest gezeigten Leistung verglichen wird, kommt es zu einer guten Übereinstimmung. In einer frühen Metaanalyse von Hoge und Coladarci (1989), in die 16 Studien eingeschlossen wurden, konnte eine Übereinstimmung von  $r = 0.66$  gefunden werden. Bei Einschätzungen im Sinne einer Rangfolge zeigen Lehrkräfte ebenfalls sehr gute diagnostische Urteile (Schrader, 2013, S. 158). Gleichzeitig merken hier Südkamp et al. (2012, S. 757) kritisch an, dass dies auch dann zutrifft, wenn Lehrkräfte ihre Schülerinnen und Schüler systematisch über- oder unterschätzen, sodass diese Angabe entsprechend vorsichtig bewertet werden muss.

Die meisten Studien fokussieren auf den Bereich der kognitiven Merkmale. Es liegen weit weniger Studien vor, welche die Urteilsakkuratheit von Lehrkräften bei nicht-kognitiven Merkmalen erforschen (Bilz, Steger & Fischer, 2016, S. 4; Stang & Urhahne, 2016a, S. 206). Die Ergebnisse deuten darauf hin, dass die Urteilsakkuratheit hier relativ gering zu sein scheint, sehr sicher aber geringer als bei kognitiven Merkmalen ausfällt (Stang & Urhahne, 2016a, S. 206–207). In einer Studie wurde beispielsweise die Urteilsakkuratheit bei Mobbing unter Schülerinnen und Schülern untersucht (Bilz et al., 2016). Hier wurde aufgezeigt, dass Lehrkräfte die Zahl der Mobbing-Täterinnen bzw. -Täter sowie derer, die sowohl als Täterin bzw. Täter wie auch als Mobbing-Opfer in Erscheinung treten, überschätzen. Gleichzeitig zeigte sich hier keine Unterschätzung der Mobbing-Opfer, was ein von internationalen Forschungen abweichendes Ergebnis darstellt (Bilz et al., 2016, S. 10). Bilz et al. kommen zu dem Schluss, dass Lehrkräfteurteile in diesem Bereich ungenauer ausfallen als im Leistungsbereich, was einen Hinweis auf unterschiedliche zugrundeliegende Kompetenzen liefert. Sie vermuten, dass Lehrkräfte stärker auf den Leistungsbereich von Schülerinnen und Schülern fokussieren und somit Bereiche, welche die Persönlichkeit betreffen, eher unbemerkt bleiben (Bilz et al., 2016, S. 11).

Seyda (2018, S. 8) untersuchte in seiner Studie die Akkuratheit von Sportlehrkräften in deren Beurteilungen physischer Fähigkeitsselbstwahrnehmungen ihrer Schülerinnen und Schüler. Hierzu wurden 35 Sportlehrkräfte zu ihren insgesamt 509 Grundschulkindern (dritte und vierte Klassen) befragt. Im Ergebnis zeigte sich, dass die Sportlehrkräfte die physische Fähigkeitsselbstwahrnehmung signifikant unter- und die Streuung der Ausprägung in ihrer Klasse signifikant überschätzten. Ob die Sportlehrkräfte Sport studiert hatten oder nicht, beeinflusste dagegen nicht die Akkuratheit der Einschätzungen. Als mögliche Erklärung führt Seyda (2018, S. 12–13) an, dass die Leistung der Kinder den Blick auf die Fähigkeitsselbstwahrnehmung beeinflussen könnte – unter anderem auch, weil sportliche Leistung eine einfacher zu erfassende Information ist.

In einer experimentellen Studie von Kaiser, Möller, Helm und Kunter (2015) kam es zu einer signifikanten Überschätzung der Mathematikleistung. Die Studie fand mit Lehrkräften im Vorbereitungsdienst in einem fiktiven Setting mit ebenfalls fiktiven Schülerinnen und Schülern

einer siebten Klasse einer Gesamtschule statt. Gleichzeitig wurde die Streuung der Leistungen durch die Lehrkräfte signifikant unterschätzt (Kaiser et al., 2015, S. 293). Auf die Studie wird im Abschnitt 3.3.2 noch einmal eingegangen werden, da hier die Frage nach Einflussfaktoren auf die Urteilsakkuratheit im Fokus stand.

Für den Bereich des Lesens wurde vor allem durch die PISA- und die IGLU-Studien deutlich, dass die Urteilsakkuratheit häufig nicht zufriedenstellend ausfällt.

Ziel von PISA (Programme for International Student Assessment) ist es, den teilnehmenden Staaten mit international standardisierten Testinstrumenten Daten über den Erfolg ihrer Bildungssysteme – auch im internationalen Vergleich – mitzuteilen. Den Schwerpunkt der Untersuchungen bilden hier im wechselnden Rhythmus die Erhebung der Lesekompetenz sowie der mathematischen und der naturwissenschaftlichen Grundbildung von 15-jährigen Schülerinnen und Schülern (Baumert, Stanat & Demmrich, 2001, S. 15). Für den Erhebungszeitraum im Jahr 2000 kam heraus, dass die Lehrkräfte der Sekundarstufe I leseschwache Schülerinnen und Schüler häufig nicht erkennen. Von den Jugendlichen, die hier im Lesetest als Risikoschülerinnen und -schüler identifiziert wurden, wurden knappe 89 Prozent nicht von den Lehrkräften als solche benannt (Artelt et al., 2001, S. 119–120).

In den IGLU-Studien (Internationale Grundschul-Lese-Untersuchung; im Original PIRLS – Progress in International Reading Literacy Study) stehen Grundschulkindern im Fokus. Hier wird sich im Gegensatz zu PISA nur auf den Bereich der Lesekompetenz konzentriert, aber auch diese Untersuchung ist international angelegt. Es sollen langfristige Entwicklungen in Bezug auf die Bildungssysteme betrachtet werden. Die Untersuchungen finden in Deutschland und weiteren teilnehmenden Nationen seit 2001 alle fünf Jahre statt. Getestet wird die Lesefähigkeit von Kindern in der vierten Klasse (Hußmann, Wendt, Kasper, Bos & Goy, 2017, S. 29). Im Rahmen der IGLU-Studie 2006 wurde außerdem überprüft, wie viele schwache Leserinnen und Leser von den Grundschullehrkräften als solche benannt wurden, indem die Grundschullehrkräfte darum gebeten worden waren, den Kindern in ihrer Klasse für ihre Leseleistung eine Schulnote zu geben (Valtin et al., 2010, S. 75–76). Ein Drittel der leseschwachen Schülerinnen und Schüler hatte hier die Schulnote 3 („befriedigend“) oder sogar besser erhalten. Umgekehrt hatte ca. ein Drittel der lesestarken Schülerinnen und Schüler ebenfalls die Schulnote 3 (und teilweise sogar schlechter als 3) erhalten. Von den Schülerinnen und Schülern mit Schwierigkeiten im Lesen, die als solche erkannt worden waren, erhielten außerdem nur 38 Prozent eine Förderung in der Schule (Valtin et al., 2010, S. 76). Im Rahmen der IGLU-Studie 2016 schätzten die Lehrkräfte den Anteil der leseschwachen Schülerinnen und Schüler auf ca. 17 Prozent, während knapp 19 Prozent der Schülerinnen und Schüler in der Studie als leseschwach identifiziert wurden (Bremerich-Vos et al., 2017, S. 281). Den Anteil konnten die Lehrkräfte hier daher recht genau einschätzen. Tieferegehende Analysen, zum Beispiel ob die Lehrkräfte die jeweils leseschwachen Schülerinnen und Schüler auch korrekt erkennen, sind mit diesen Ergebnissen aber nicht möglich, da die Lehrkräfte nicht weiter befragt wurden und nur den Anteil schätzen sollten.

Es zeigt sich demnach, dass Lehrkräfte insgesamt die Leistung ihrer Schülerinnen und Schüler moderat bis gut einschätzen können, wobei dies im Bereich Lesen in den großen PISA- und stellenweise auch in den IGLU-Studien nicht immer der Fall gewesen ist und hier zum Teil niedrige Übereinstimmungen gefunden wurden. Die Rangreihenfolge können Lehrkräfte dagegen recht gut bilden, während es insgesamt eher zu einer Überschätzung der Leistungen der Schülerinnen und Schüler kommt. Für schwächere Schülerinnen und Schüler scheint die Einschätzung zusätzlich schwerer zu fallen als für stärkere. Die Streubreite zwischen den einzelnen Lehrkräften ist allerdings recht hoch. Es ist daher wichtig, auf mögliche Einflussfaktoren zu blicken.

### 3.3.2 Einflussfaktoren auf die Urteile von Lehrkräften

Neben der Urteilsakkuratheit beschäftigen sich viele Studien auch mit der Frage nach möglichen Einflussfaktoren auf die Akkuratheit dieser Lehrerurteile. So können Urteile schnell verzerrt sein, z. B. durch Voreinstellungen und Erwartungen (Schrader & Helmke, 2014, S. 46–47; Stang & Urhahne, 2016a, S. 205).

Schrader und Helmke (2014, S. 47) stellten ein Modell über den Zusammenhang von Leistungserwartungen, Unterricht, weiteren Einflussfaktoren wie die Klassenzusammensetzung und über das daraus resultierende diagnostische Urteil auf. Dieses Modell verdeutlicht das komplexe Zusammenspiel mehrerer Faktoren, wie der Diagnosekompetenz und Klassen- bzw. Schülermerkmalen, Erwartungen, Unterrichtsgestaltung und Beobachtungen der Lehrkraft. Sie führen aber ebenfalls an, dass die Interpretation gemachter Beobachtungen – abhängig von den Erwartungen – das diagnostische Urteil einer Lehrkraft beeinflussen kann: Wird z. B. eine unerwartet gute Leistung einer Schülerin oder eines Schülers auf Glück oder auf Anstrengung bezogen? Hierüber kann es zu Urteilsverzerrungen und Fehleinschätzungen kommen (Schrader & Helmke, 2014, S. 49).

In der von Südkamp et al. (2012) durchgeführten Metaanalyse wurde neben der Analyse der Urteilsakkuratheit an sich auch ein Blick auf Einflussfaktoren geworfen. Auf der einen Seite stehen hier die Urteilsmerkmale, welche die Akkuratheit beeinflussen können: die Frage nach der Anzahl gelöster Antworten in einem Test oder ob die Lehrkräfte Schülerinnen und Schüler in einer spezifischen Fähigkeit oder allgemein in ihrer Leistung bewerten sollen (Südkamp et al., 2012, S. 745). Auf der anderen Seite kann auch die Art des durchgeführten Leistungstests einen Einfluss ausüben. Ebenso halten sie fest, dass die Urteilsakkuratheit unterschiedlich ausfällt, je nachdem, welche Zeitspanne zwischen den Urteilen auf Lehrkräfteseite und den durchgeführten Tests liegt (Südkamp et al., 2012, S. 746). Ein dritter wichtiger Punkt sind Merkmale auf Schüler- und Lehrerseite. Da verschiedene Studien hohe interindividuelle Unterschiede in Bezug auf die Urteilsakkuratheit finden konnten, ist die Frage, welche Faktoren hier auf Lehrkräfteseite zu finden sind, die diese Akkuratheit beeinflussen. Da allerdings zu wenig Studien vorgelegen haben, die sich mit den Einflussfaktoren auf Lehrkräfteseite beschäftigen, konnten diese Faktoren nicht in der hier durchgeführten Metaanalyse

berücksichtigt werden (Südkamp et al., 2012, S. 746). Auf Seiten der Schülerinnen und Schüler gab es Hinweise darauf, dass negative Verhaltensweisen einen negativen Einfluss auf die Urteile der Lehrkräfte in Bezug auf akademische Leistungen hatten – aber auch diese Einflussfaktoren konnten aufgrund einer zu niedrigen Anzahl an geeigneten Studien in dieser Metaanalyse nicht berücksichtigt werden (Südkamp et al., 2012, S. 747). Das Publikationsjahr der Studien hatte keinen Einfluss auf die Urteilsakkuratheit, sodass nicht davon ausgegangen werden kann, dass sich die Fähigkeit, Schülerinnen und Schüler zu beurteilen, über die Jahre verändert hat (Südkamp et al., 2012, S. 752). Es konnten insgesamt zwei Faktoren ausgemacht werden, welche die Urteilsakkuratheit beeinflussten: informierte vs. uninformierte Urteile sowie Kongruenz vs. Inkongruenz in Bezug auf den Leistungstest und das Lehrkräfteurteil (Südkamp et al., 2012, S. 755). Informierte Lehrkräfte wussten, welcher Test als Vergleich für ihr Urteil dienen würde. Die Urteilsakkuratheit war in diesen Fällen insgesamt höher als bei uninformierten Lehrkräften. Wenn zwischen dem zu tätigendem Urteil und der im Test abgefragten Leistung Kongruenz herrschte, fiel die Übereinstimmung ebenfalls höher aus als bei Inkongruenz (Südkamp et al., 2012, S. 755–756). Als Ergebnis der Metaanalyse entstand ein heuristisches Modell zu Moderatoren der Urteilsgenauigkeit, welches zum Teil auf theoretischen Implikationen und zum Teil auf den erhaltenen empirischen Ergebnissen beruht (Südkamp et al., 2012, S. 756). Es berücksichtigt auf der einen Seite das Lehrkräfteurteil, welches durch Merkmale der Lehrkraft und Urteilsmerkmale beeinflusst wird, sowie auf der anderen Seite die schulischen Leistungen von Schülerinnen und Schülern, welche von Schülermerkmalen sowie Testmerkmalen beeinflusst werden. Die Übereinstimmung der schulischen Leistung mit dem von den Lehrkräften gefällten Urteil bezeichnet dann die Urteilsakkuratheit. Weiterhin beeinflussen sich auch alle Bereiche gegenseitig – so beeinflusst zum Beispiel die Kongruenz zwischen dem durchgeführten Leistungstest und den Merkmalen des Lehrkräfteurteils die Akkuratheit. Ebenso ist auch möglich, dass Schülermerkmale die Urteilsakkuratheit beeinflussen können (Südkamp et al., 2012, S. 757; Südkamp, Kaiser & Möller, 2017, S. 34). Empirisch überprüft werden konnten in dieser Metaanalyse aber nicht alle Faktoren. Einen in der Metaanalyse empirisch belegten signifikanten Einfluss hatten hier, wie oben berichtet, die Kongruenz zwischen dem durchgeführten Leistungstest und der Art des Lehrkräfteurteils sowie die Informiertheit der Lehrkräfte über den durchgeführten schulischen Leistungstest. Als Kritikpunkt wird angeführt, dass dieses Modell in vielen Bereichen noch spekulativ ist (Südkamp et al., 2012, S. 757).

Lehrkräfte sollten sich folglich über mögliche Ursachen bewusst sein, um darauf achten zu können, sich in ihrem Urteil nicht ablenken zu lassen oder womöglich schwache Schülerinnen und Schüler nicht zu erkennen, weil andere Faktoren die Schwierigkeiten überdecken. In den folgenden Abschnitten wird noch einmal auf einzelne mögliche Ursachen für Urteilsfehler und -verzerrungen eingegangen werden.

### 3.3.2.1 Schülermerkmale

Oft werden bei Studien über den Einfluss von Schülermerkmalen auf die Urteilsakkuratheit mehrere Merkmale parallel untersucht, sodass der Forschungsstand im Folgenden nicht strikt nach einzelnen Faktoren getrennt dargestellt wird, sondern stattdessen einzelne Studien diskutiert werden, die gemeinsam den Forschungsstand abbilden.

Eine bildungsferne Herkunftsfamilie oder auch das Geschlecht des Kindes können mit niedrigeren schulischen Leistungen im Zusammenhang stehen. Dennoch sollte das Merkmal für sich genommen nicht ausschlaggebend dafür sein, dass eine Schülerin oder ein Schüler unabhängig von seiner oder ihrer individuellen Leistung schlechter bewertet wird. Karing, Matthäi et al. (2011) gehen in ihrer Studie unter anderem der Frage nach, ob Geschlecht und sozioökonomischer Status mit dem Lehrkräfteurteil über die Lesekompetenz von Schülerinnen und Schülern zusammenhängen. Ausgehend von bisherigen Forschungsergebnissen postulieren sie, dass diese beiden Faktoren keinen Einfluss auf die Urteilsakkuratheit haben sollten, wobei sie einschränkend festhalten, dass viele Studien aus dem mathematischen Bereich stammen und sich daher nicht unbedingt auf das Fach Deutsch übertragen lassen. In dieser Studie wurden 64 Hauptschul-, Realschul- und Gymnasiallehrkräfte, die das Fach Deutsch unterrichten, zum Ende des fünften Schuljahres zu ihren Schülerinnen und Schülern befragt (Karing, Matthäi, et al., 2011, S. 162–163). Weder für globale noch für aufgabenspezifische Lehrkräfteurteile in Bezug auf die Lesekompetenz konnte hier ein Einfluss des Geschlechts oder des sozioökonomischen Status gefunden werden (Karing, Matthäi, et al., 2011, S. 168). Dagegen führen (Kaiser et al., 2015, S. 284) mehrere Untersuchungen an, in denen sowohl die soziale Herkunft als auch das Geschlecht die Urteile von Lehrkräften – auch nach Kontrolle der schulischen Leistung – beeinflussten: So untersuchten Ready und Wright (2011) in ihrer Studie die Urteilsakkuratheit von 1822 Kindergarten-Lehrkräften in den USA. Die Lehrkräfte sollten einmal im Herbst, also zu Beginn des Kindergartenjahres, und im darauffolgenden Frühjahr die sprachlichen und schriftsprachlichen Fähigkeiten von insgesamt 9493 Kindern einschätzen. Die Autoren kommen zu dem Schluss, dass die Hälfte der Abweichungen in den Leistungseinschätzungen durch tatsächliche Leistungsunterschiede der Kinder entstanden war, die übrige Hälfte aber aufgrund von systematischen Fehlern in der Wahrnehmung der Lehrkräfte zustande gekommen sein muss (Ready & Wright, 2011, S. 348). Mädchen wurden hierbei leicht stärker eingeschätzt als Jungen. Die höchste Leistungsunterschätzung lag allerdings bei Kindern mit niedrigem sozioökonomischen Status vor. Dieser Effekt war in Klassen, die insgesamt einen niedrigeren sozioökonomischen Status aufwiesen, noch stärker (Ready & Wright, 2011, S. 355). Maaz et al. (2008) nahmen dagegen den Übergang von der Grundschule in die Sekundarstufe in den Blick. Die Studie fand in Berlin statt, wo die meisten Schülerinnen und Schüler erst nach der sechsten Klasse in die Sekundarstufe wechseln. Lehrkräfte sollten hier zum Ende der fünften Klasse die Lernkompetenz ihrer Schülerinnen und Schüler einschätzen und die Schulnote vorhersagen. Sowohl das Geschlecht als auch der sozioökonomische Status waren Prädiktoren für die

Einschätzung der Lernkompetenz durch die Lehrkräfte – auch unter Kontrolle der Testleistung (Maaz et al., 2008, S. 536). Die Einschätzung fiel für Schülerinnen und Schüler mit einem höheren sozioökonomischen Status sowie für Mädchen im Vergleich zu Jungen positiver aus. Bei vergleichbarem sozioökonomischen Hintergrund fiel die Leistungseinschätzung für Schülerinnen und Schüler in insgesamt sozial stärkeren Klassen schlechter aus als für Schülerinnen und Schüler in sozial schwächeren Klassen (Maaz et al., 2008, S. 536). Es wird angemerkt, dass der Grund der Verzerrung aus den Daten nicht erkennbar ist. Einerseits könnte es sein, dass ein günstigerer sozioökonomischer Hintergrund als Ressource für einen höheren Bildungsgang gesehen wird. Andererseits wird vermutet, dass z. B. unterschiedliches Arbeits- oder Lernverhalten in die Bewertung mit einfließt (sowohl bezüglich des sozioökonomischen Hintergrunds wie auch des Geschlechts), was in dieser Studie allerdings nicht erhoben wurde (Maaz et al., 2008, S. 541). In ihrer eigenen Studie konnten Kaiser et al. (2015) ebenfalls das Geschlecht als Einflussfaktor ausmachen. Sie konzentrierten sich auf die Urteilsakkuratheit bei der Einschätzung von Mathematikleistungen. 251 Lehrkräfte im Vorbereitungsdienst sollten 16 fiktive Schülerinnen und Schüler einer 7. Klasse einer Gesamtschule hinsichtlich ihrer Mathematikleistung einschätzen. Hiervon bildeten 186 die Experimentalgruppe und die übrigen die Kontrollgruppe. Die Experimentalgruppe erhielt sowohl urteilsrelevante Informationen (die mündliche Leistung sowie die Note der letzten Mathematikarbeit) wie auch urteilsirrelevante Informationen. Letztere waren hier die Deutschleistung, Intelligenz, schulisches Selbstkonzept und das kulturelle Kapital sowie das Geschlecht (Kaiser et al., 2015, S. 289–290). Die Kontrollgruppe erhielt nur die relevanten Informationen und die Angabe des Geschlechts (Kaiser et al., 2015, S. 278–288). Es zeigte sich, dass die Deutschleistung der Schülerinnen und Schüler, deren Intelligenz sowie das Geschlecht die Einschätzung der Lehrkräfte beeinflussten: Eine bessere Deutschnote oder eine höhere Intelligenz führten zu einer akkurateren Einschätzung der Mathematikleistung. Ebenso erhielten Mädchen eine akkuratere Einschätzung als Jungen. Das schulische Selbstkonzept und das kulturelle Kapital hatten dagegen keinen Einfluss (Kaiser et al., 2015, S. 294–295).

In einer Studie von Roos et al. (2016) wurden Alter und Geschlecht von Kindern und Jugendlichen als mögliche Einflussfaktoren auf Lehrkräfteurteile untersucht. Die Einschätzung des Sozial- und Lernverhaltens stand hier im Fokus. Die Studie wurde über knapp zwei Jahre im norddeutschen Raum durchgeführt. Teilgenommen haben 30 Lehrkräfte sowie 189 Schülerinnen und Schüler, welche die Grund-, Haupt-, Realschule oder das Gymnasium besuchten. Das Alter der Schülerinnen und Schüler lag zwischen 9 und 18 Jahren (Roos et al., 2016, S. 202). Es zeigte sich bei Jungen und Mädchen in Abhängigkeit vom Alter ein gegenläufiger Effekt: Bei den Mädchen wurde der Zusammenhang zwischen Selbsteinschätzung des Sozialverhaltens und Fremdeinschätzung durch die Lehrkraft stärker, je älter die Mädchen waren – die Akkuratheit stieg hier also an. Bei den Jungen wurde der Zusammenhang zwischen Selbst- und Fremdeinschätzung des Sozialverhaltens dagegen mit dem Alter immer geringer und war im oberen Altersbereich (ab ca. 11.5 Jahren) nicht mehr vorhanden (Roos et al.,

2016, S. 202) – die Akkuratheit sank hier also. Die Einschätzung des Lernverhaltens durch die Lehrkräfte stand in keinem Zusammenhang mit den Angaben der Schülerinnen und Schüler. Nach Geschlecht getrennt zeigte sich hier allerdings für die Gruppe der Mädchen ein Zusammenhang mit der Einschätzung durch die Lehrkräfte, nicht aber für die Gruppe der Jungen (Roos et al., 2016, S. 205). Für Mädchen konnten die Lehrkräfte also akkuratere Urteile bilden, allerdings teilweise abhängig vom Alter.

Hinnant, O'Brian und Ghazarian (2009) führten eine Langzeitstudie durch, in der Kinder von der Geburt bis zur Klasse 5 begleitet wurden. Über diesen Zeitraum nahmen 955 Kinder an der Studie teil. Die Lehrkräfte sollten die teilnehmenden Kinder bezüglich ihrer Leistungen in Mathematik und Deutsch in der ersten, der dritten und der fünften Klasse einschätzen. Es zeigte sich, dass die Lehrkräfte die Leistungen von Mädchen im Lesen überschätzten. Das Geschlecht in Bezug auf die Mathematikleistung war nur in der fünften Klasse ein Prädiktor und auch hier wurde die Leistung der Mädchen überschätzt. Die sozialen Fähigkeiten der Kinder waren durchgehend positiv korreliert mit der Einschätzung der Lehrkräfte im Lesen und Mathematik. Weiterhin ergaben sich für Jungen, die einer ethnischen Minderheit angehörten, die niedrigsten Leistungen im Lesen, wenn sie unterschätzt wurden und den größten Zuwachs im Lesen, wenn sie überschätzt wurden (Hinnant et al., 2009, S. 669).

Kuhl und Hannover (2012) beschäftigen sich in ihrer Studie mit einem von mehreren Erklärungsansätzen zu dem Phänomen, dass Mädchen selbst bei statistischer Kontrolle der Leistung bessere Noten erhalten als Jungen. Die Studie fand mit 1533 Jungen und Mädchen aus insgesamt 78 vierten Klassen an Grundschulen statt. Ihr Fokus lag auf dem Erklärungsansatz, dass Lehrkräfte bei Mädchen vermehrt selbstgesteuertes Lernen wahrnehmen, was in der Folge zu einer besseren Benotung führt (Kuhl & Hannover, 2012, S. 153–154). Es wurden die letzten Halbjahresnoten in Deutsch und Mathematik erhoben. Zusätzlich wurden standardisierte Tests durchgeführt, um das Leseverständnis und die mathematischen Kompetenzen zu erfassen. Die Lehrkräfte schätzten für jedes ihrer Kinder das selbstgesteuerte Lernen ein. Es zeigte sich, dass die Mädchen eine durchschnittlich höhere Leistung im Leseverständnistest zeigten sowie in der Benotung im Fach Deutsch. Der Unterschied in der Benotung blieb jedoch auch nach statistischer Kontrolle der Leistung bestehen (Kuhl & Hannover, 2012, S. 159). In der Wahrnehmung des selbstgesteuerten Lernens ergaben sich Unterschiede: Erwartungsgemäß erhielten Mädchen hier höhere Zuschreibungen als Jungen. In weitergehenden Analysen zeigte sich, dass die Wahrnehmung des selbstgesteuerten Lernens einen Einfluss auf die Benotung hatte. Ein höheres Maß an selbstgesteuertem Lernen führte zu einer besseren Benotung. Allerdings blieb die bessere Benotung auch nach statistischer Kontrolle der Lesefähigkeit und des selbstgesteuerten Lernens bestehen (Kuhl & Hannover, 2012, S. 158). In Mathematik gab es dagegen keine Unterschiede in der Benotung. Es konnte allerdings eine Interaktion zwischen Benotung und Geschlecht gefunden werden: Jungen erhielten bei gleicher Leistung im Mathematiktest und bei gleicher Wahrnehmung des selbstgesteuerten Lernens eine bessere Benotung als Mädchen. Die erhaltenen Ergebnisse werden

dahingehend interpretiert, dass „geschlechtsstereotype Erwartungen oder Beurteilungstendenzen“ auf Seiten der Lehrkräfte vorhanden sein können (Kuhl & Hannover, 2012, S. 159). Der sozioökonomische Status scheint teilweise ein Einflussfaktor zu sein auf die Urteilsakkuratheit von Lehrkräften, aber nicht alle Studien konnten dies nachweisen. Die oben genannten Studien zeigen für Kindergartenkinder in den USA für den schriftsprachlichen Erwerb einen Einfluss, da Lehrkräfte hier Kinder mit niedrigerem sozioökonomischen Status am stärksten unterschätzt hatten (Ready & Wright, 2011). Auch für die Einschätzung der Lernkompetenz Ende der sechsten Klasse konnte solch ein Einfluss gezeigt werden (Maaz et al., 2008). Keinen Einfluss dagegen konnten Karing, Matthäi et al. (2011) für den Bereich Deutsch in der Sekundarstufe und Kaiser et al. (2015) für den Bereich Mathematik in der Sekundarstufe und für Lehrkräfte im Vorbereitungsdienst nachweisen.

Insgesamt lässt sich auch für das Geschlecht als Einflussfaktor festhalten, dass es keine eindeutigen Ergebnisse gibt, es sich aber abzeichnen scheint, dass es hier einen Einfluss auf die Urteile der Lehrkräfte gibt. Während Karing, Matthäi et al. (2011) keinen Einfluss nachweisen konnten, haben andere Studien diesen aufzeigen können. In allen hier vorgestellten Studien, in denen Unterschiede ausgemacht wurden, wurden Mädchen positiver bewertet. Dies gilt sowohl für den Bereich der Mathematik als auch für Deutsch, wobei Hinnant et al. (2009) dies für Mathematik nur für die fünfte Klasse, aber nicht für die niedrigeren Klassenstufen aufzeigen konnten. Kuhl und Hannover (2012) nahmen die Wahrnehmung des selbstgesteuerten Lernens als Moderator in den Blick. Sie konnten hier einen Zusammenhang finden: Mädchen wurden als selbstgesteuerter im Lernen wahrgenommen und hierüber ließ sich auch zum Teil die bessere Einschätzung in Deutsch erklären.

### 3.3.2.2 Lehrkraftmerkmale

Das heuristische Modell der Akkuratheit diagnostischer Urteile von Lehrkräften von Südkamp et al. (2012) berücksichtigt neben Schülermerkmalen auch Merkmale auf Seiten der Lehrkräfte, um fehlerhafte Urteile über schulische Leistungen von Schülerinnen und Schülern zu erklären. In diesem Modell wird davon ausgegangen, dass Merkmale der Lehrkraft – hier die Berufserfahrung als Beispiel – die Akkuratheit beeinflussen (Südkamp et al., 2017, S. 33). Zur Berufserfahrung von Lehrkräften wurden entsprechend mehrere Studien durchgeführt. Hierbei konnte allerdings oft kein Zusammenhang gefunden werden (Christophel et al., 2017, S. 277; Hesse & Latzko, 2017, S. 27; Ohle & McElvany, 2015, S. 6). An der Studie von Christophel et al. (2017) nahmen neben Lehrkräften, die zwischen 0 und 17 Jahren Berufserfahrung hatten, auch Lehramtsstudierende teil, die im 1. bis 12. Fachsemester waren und 2 bis 23 Wochen Praktikumserfahrung hatten (Christophel et al., 2017, S. 273). Es zeigte sich, dass die Berufserfahrung keinen Einfluss auf die Fähigkeit zur Einschätzung der Aufgabenschwierigkeit hatte (Christophel et al., 2017, S. 277). Auch die Autorinnen und der Autor des heuristischen Modells führen abschließend kritisch an, dass Belege für die Berufserfahrung als Einfluss auf die Akkuratheit bisher nicht vorhanden sind (Südkamp et al., 2017, S. 35).

Stang und Urhahne (2016a) untersuchten, wie akkurat Lehrkräfte die Mathematikleistung ihrer Schülerinnen und Schüler sowie weitere Faktoren wie Konzentration und Arbeitsverhalten einschätzen können. Im Fokus lag aber vor allem die Frage, ob soziodemographische Merkmale der Lehrkräfte (Alter, Geschlecht und Berufserfahrung) ihre Urteile beeinflussen. Insgesamt nahmen 17 Lehrkräfte an der Studie teil. Es flossen Daten von 357 Fünftklässlerinnen und -klässlern ein (Stang & Urhahne, 2016a, S. 209). Dabei konnte eine mäßige Übereinstimmung zwischen den Lehrkräfteurteilen auf der einen und den schulischen Matheleistungen, der Konzentration, der Arbeitsweise sowie dem Sozialverhalten auf der anderen Seite gefunden werden, wobei die nicht-leistungsbezogenen Eigenschaften noch einmal schlechter eingeschätzt werden konnten als die Mathematikleistung (Stang & Urhahne, 2016a, S. 215). Für Berufserfahrung, Geschlecht oder Alter der Lehrkräfte und ihre Urteile konnte kein Zusammenhang festgestellt werden (Stang & Urhahne, 2016a, S. 212).

Wenn die diagnostische Kompetenz sich also durch den beruflichen Alltag nicht von selbst zu verbessern scheint, liegt es nahe, die Aus- sowie Fortbildung von Lehrkräften in den Blick zu nehmen. Kaiser und Möller (2017) untersuchten mithilfe des Simulierten Klassenraums, wie sich die diagnostische Kompetenz von Lehramtsstudierenden über einen Zeitraum von drei Jahren entwickelte. Als Indikator diente die Urteilsakkuratheit. Eine Kontrollgruppe bildeten Pädagogikstudierende. Im Simulierten Klassenraum gaben fiktive Schülerinnen und Schüler auf ausgewählte Fragen korrekte oder falsche Antworten. Ob eine Schülerin oder ein Schüler sich melden würde und ob die Antwort korrekt sein würde, wurde aufgrund von voreingestellten Fähigkeits- und Motivationsparametern bestimmt. Über den Verlauf von drei Jahren zeigte sich keine Verbesserung der Urteilsakkuratheit. Dennoch wird von der Trainierbarkeit der diagnostischen Kompetenz ausgegangen (Kaiser & Möller, 2017, S. 71). Kaiser und Möller beziehen sich dabei unter anderem auf eine Studie von Klug (2011, S. 127), die zeigen konnte, dass Gymnasiallehrkräfte mithilfe eines Trainings über diagnostische Prozesse bei der Einschätzung des Lernverhaltens von Schülerinnen und Schülern ihre diagnostische Kompetenz sowie ihr Wissen über diagnostische Prozesse steigern konnten.

### 3.3.2.3 Kontaktdauer

Es wird vermehrt darauf hingewiesen, dass eine zu kurze Kontaktdauer – also eine verhältnismäßig kurze Zeit, die die Lehrkraft mit den Schülerinnen und Schülern verbracht hat – die Urteilsakkuratheit negativ beeinflussen kann. Bailey und Drummond (2006) führten ihre Studie daher erst zum Ende des Schulhalbjahres durch, um den Lehrkräften mindestens ein halbes Jahr mit den Kindern zu geben, bevor sie deren Lesefähigkeiten einschätzen sollten. Auch Stang und Urhahne (2016a, S. 215) nannten als möglichen Einflussfaktor in ihrer Studie, dass die Lehrkräfte die Schülerinnen und Schüler zum Zeitpunkt der Untersuchung erst etwa drei Monate gekannt hatten und dies eventuell ein zu kurzer Zeitraum sei, um diese akkurat einschätzen zu können.

Der Frage, ob eine höhere Kontaktdauer einen positiven Einfluss auf die Urteilsakkuratheit hat, gingen Oerke et al. (2016) in ihrer Studie nach. Sie untersuchten die Akkuratheit der Urteile nach einem halben Jahr und nach eineinhalb Jahren Kontaktdauer, wobei sich die Urteilsakkuratheit hier auf die Fähigkeit zur Bild-Text-Integration beim Lesen bezog (Oerke et al., 2016, S. 35). Während in der Querschnittsuntersuchung zum ersten Messzeitpunkt keine Unterschiede in der Urteilsakkuratheit zwischen Lehrkräften, die ihre (fünfte) Klasse ein halbes Jahr kannten und Lehrkräften, die ihre (sechste) Klasse eineinhalb Jahre kannten, gefunden werden konnten, zeigte sich eine Verringerung der Urteilsfehler bei Betrachtung der Längsschnittdaten (Oerke et al., 2016, S. 43). Diese Verringerung bezieht sich allerdings nur auf die Einschätzung in einer konkreten Aufgabe, die aus mehreren Teilen bestand, nicht aber auf Aussagen zu den Leistungen der Schülerinnen und Schüler im Gesamtttest. Insgesamt zeigen sich demnach keine Unterschiede in der Urteilsakkuratheit bei einem und eineinhalb Jahren Kontaktdauer (Oerke et al., 2016, S. 43). Allerdings war hier die Dauer von sechs Monaten das Minimum, welche sowohl von Bailey und Drummond (2006, S. 154) als auch von Stang und Urhahne (2016a, S. 215) nicht als zu kurz bezeichnet wurde. Ready und Wright (2011, S. 344–346) konnten feststellen, dass die Urteilsakkuratheit nach längerer Kontaktdauer anstieg: Die Spannweite der Leistung der untersuchten Kindergartenkinder wurde im Frühjahr von den Lehrkräften unterschätzt. Im Frühjahr darauf waren die Einschätzungen dagegen akkurater.

#### 3.3.2.4 Referenzgruppeneffekt

Die Leistungsfähigkeit einer Klasse kann die Urteilsakkuratheit einer Lehrkraft ebenfalls beeinflussen: Die Urteile von Lehrkräften sind nicht unabhängig von der durchschnittlichen Leistung einer Klasse (Trautwein & Baeriswyl, 2007, S. 122). Dieser Einflussfaktor ist auch als Referenzgruppeneffekt bekannt. Trautwein und Baeriswyl berichten über ihre Studie zum Referenzgruppeneffekt bei Übertrittsempfehlungen in die Sekundarstufe I zum Ende der sechsten Klasse im schweizerischen Kanton Freiburg, wobei hier nur deutschsprachige Schülerinnen und Schüler betrachtet wurden. Wenn eine Klasse leistungsstark war, erhielt eine Schülerin oder ein Schüler eine schlechtere Übertrittsempfehlung durch die Lehrkraft, als eine Schülerin oder ein Schüler mit der gleichen Leistung (ermittelt mit einem standardisierten schulischen Leistungstest), die bzw. der in einer weniger leistungsstarken Klasse war (Trautwein & Baeriswyl, 2007, S. 128). Südkamp und Möller untersuchten den Referenzgruppeneffekt mithilfe des Simulierten Klassenraumes und 42 Lehramtsstudierenden, wobei diese entweder Lehramt an Gymnasien oder Lehramt an Realschulen studierten. Andere Lehrämter waren nicht vertreten (Südkamp & Möller, 2009, S. 165). Die Lehramtsstudierenden sollten in diesem Simulierten Klassenraum die Leistung von simulierten Schülerinnen und Schülern einschätzen, wobei diese (über eine randomisierte Zuteilung) entweder eine Klasse mit hohem oder mit niedrigem durchschnittlichen Leistungsniveau vor sich hatten. Die Akkuratheit der Leistungseinschätzung der simulierten Schülerinnen und Schüler wird als „im

mittleren Bereich“ und somit zufriedenstellend bezeichnet (Südkamp & Möller, 2009, S. 171). In Bezug auf die Vergabe von Schulnoten für die Leistungen der simulierten Schülerinnen und Schüler durch die Lehramtsstudierenden konnte ein Referenzgruppeneffekt für das mittlere Leistungsniveau der Klasse ermittelt werden. In einer Klasse mit einem niedrigen Leistungsniveau vergaben die Lehramtsstudierenden bessere Noten für Schülerinnen und Schüler als für Schülerinnen und Schüler mit dem gleichen Anteil korrekter Antworten, die in der durchschnittlich leistungsfähigeren Klasse waren. Während bei der Notenvergabe die Leistungen der übrigen Schülerinnen und Schüler die Wahl der Lehramtsstudierenden beeinflusste, geschah dies allerdings nicht, wenn sie die Leistung direkt über den Anteil korrekter gegebener Antworten der jeweiligen Schülerin bzw. des jeweiligen Schülers einschätzen sollten (Südkamp & Möller, 2009, S. 171–172).

Maaz et al. (2008) fanden in ihrer Studie ebenfalls einen Referenzgruppeneffekt. Bei gleicher Testleistung bekamen Schülerinnen und Schüler in leistungsstarken Klassen schlechtere Noten und niedrigere Lernkompetenzeinschätzungen von den Lehrkräften als Schülerinnen und Schüler in leistungsschwächeren Klassen (Maaz et al., 2008, S. 536; 542).

Becker und Birkelbach (2013, S. 216–217) halten fest, dass auch ein umgekehrter Effekt auftreten könnte im Sinne des Halo-Effektes, den die Autoren als „positiven Referenzgruppeneffekt“ bezeichnen: Eine insgesamt leistungsstarke Klasse strahlt auch auf schwächere Schülerinnen und Schüler ab und so werden alle positiver beurteilt als Schülerinnen und Schüler in weniger leistungsstarken Klassen. Auch schwächere Schülerinnen und Schüler könnten ihrer Vermutung nach hier vom Halo-Effekt profitieren, wenn sie sich weniger am Unterricht beteiligen und die Lehrkraft sie dann entsprechend weniger gut einschätzen kann (Becker & Birkelbach, 2013, S. 217). In einer eigenen Untersuchung, in der Lehrkräfte angeben sollten, welchen Schülerinnen und Schülern sie ein Hochschulstudium zutrauen, konnten sie den positiven Referenzgruppeneffekt sowohl für die Durchschnittsnoten der Klasse als auch für die durchschnittlichen kognitiven Fähigkeiten der Klasse nachweisen (Becker & Birkelbach, 2013, S. 226).

### 3.3.3 Auswirkungen der diagnostischen Urteile auf Schülerinnen und Schüler

Die Akkuratheit der von Lehrkräften gefällten Urteile, über die mittlerweile vergleichsweise häufig geforscht wird (Kaiser & Möller, 2017, S. 57), ist sehr relevant und grundlegend für Lehrkräfte. Diaper (1990, S. 266) hält beispielsweise fest, dass die Aufmerksamkeit der Lehrkraft auf eine Schülerin oder einen Schüler schon zu einer Verbesserung der Leistung führen kann. Die Urteile sind daher der erste notwendige Schritt.

Es wird angenommen, dass eine gute Urteilsfähigkeit über den Leistungsstand der Schülerinnen und Schüler sowie über die Aufgabenschwierigkeit auch einen positiven Einfluss auf die Wirksamkeit des Unterrichts haben sollte (Anders, Kunter, Brunner, Krauss & Baumert, 2010, S. 178; 180). Kaiser und Möller (2017, S. 71–72) halten ebenso fest, dass aus guten Urteilsfähigkeiten auch positive Auswirkungen auf die Gestaltung des Unterrichts resultieren

sollten. Machts et al. (2016) merken allerdings kritisch an, dass die von Hoge und Coladarci (1989) sowie von Südkamp et al. (2012) in ihren Metaanalysen gefundenen Effektstärken für die Urteilsakkuratheit von respektive  $r = 0.66$  und  $r = 0.63$  zwar „recht hoch“ sind, aber nur 40 Prozent der Varianz der schulischen Leistung erklären (Machts et al., 2016, S. 86). Die diagnostische Akkuratheit allein führt demnach nicht zu einem höheren Lernerfolg, auch der Aspekt der Förderung muss mit eingeschlossen werden (Hesse & Latzko, 2017, S. 35). Brühwiler (2017) bezeichnet die „enge Verknüpfung diagnostischer und didaktischer Kompetenz“ als den „Kern adaptiver Lehrkompetenz“, welcher als kognitiver Prozess verstanden wird, der die Grundlage für die Planung wie auch für die Durchführung von adaptivem Unterricht bildet (Brühwiler, 2017, S. 124). Über „Wirkmechanismen“ der diagnostischen Kompetenzen im Unterricht bzw. welchen Einfluss auf schulische Leistungen diese Lehrkräfteurteile zeigen, ist allerdings kaum etwas bekannt (Leuders et al., 2014, S. 731). Hierzu gibt es bisher nur wenige Studien, welche zudem zu unterschiedlichen Ergebnissen gekommen sind (Anders et al., 2010, S. 178). Nach Schrader gibt es einen Zusammenhang zwischen Lernerfolg der Schülerinnen und Schüler auf der einen und Urteilsakkuratheit der Lehrkräfte auf der anderen Seite, wobei auch er festhält, dass dieser bisher nicht ausreichend erforscht wurde (Schrader, 2013, S. 161).

Es konnte allerdings gezeigt werden, dass die Urteilsakkuratheit einen Einfluss auf die Selbsteinschätzung und die Lernmotivation von Schülerinnen und Schülern haben kann (Urhahne et al., 2010, S. 280–281). Urhahne et al. (2010) gingen der Frage nach, ob sich von einer Lehrkraft unter- und überschätzte Schülerinnen und Schüler voneinander unterscheiden. Als Kriterium für eine deutliche Über- bzw. Unterschätzung wurde eine Abweichung in der Höhe von einer halben Standardabweichung vom Mittelwert des Leistungstests angesetzt. Es zeigte sich in der durchgeführten Studie, dass die Lehrkräfte Schülerinnen und Schüler (hier waren es Kinder der vierten Klasse) deutlich häufiger über- als unterschätzten: 81 von 144 Schülerinnen und Schülern wurden deutlich über- und 20 deutlich unterschätzt. Deutlich unterschätzte Schülerinnen und Schüler unterschieden sich nicht signifikant in ihrer Leistung von den deutlich überschätzten Schülerinnen und Schülern. Sie zeigten allerdings eine niedrigere Erfolgserwartung und ein niedrigeres Fähigkeitsselfkonzept sowie eine höhere Leistungsängstlichkeit bei vergleichbarer Leistungsmotivation und einem ebenfalls vergleichbaren Leistungsanspruch an sich selbst (Urhahne et al., 2010, S. 280–281). Urhahne et al. (2010) vermuten, dass die Meinung, die Lehrkräfte über ihre Schülerinnen und Schüler haben, vermutlich die von den Schülerinnen und Schülern gezeigte Leistung sowie ihre Selbsteinschätzung beeinflusst.

Brühwiler (2017) untersuchte den Zusammenhang der adaptiven Lehrkompetenz sowie der Urteilsakkuratheit mit der von Schülerinnen und Schülern wahrgenommenen Unterrichtsqualität und deren Lernzuwachs in einer Unterrichtseinheit zur Keimung von Samen. An der Studie nahmen 49 Lehrkräfte mit insgesamt 898 Schülerinnen und Schülern aus der deutschsprachigen Schweiz teil. Ein Teil der Lehrkräfte unterrichtete in der 7. bzw. 8. Klasse der

Sekundarstufe I, ein anderer Teil unterrichtete in der Primarstufe in der 4. bzw. 5. Klasse (Brühwiler, 2017, S. 126). Als Kern der Lehradaptivität standen die diagnostische bzw. didaktische Planungs- und Handlungskompetenz im Fokus. Diese wurden unabhängig von den von den Lehrkräften unterrichteten Schülerinnen und Schülern mithilfe eines Vignetten- und Videotests erhoben. Zudem wurde die Urteilsakkuratheit als weiterer Faktor betrachtet, welche durch Einschätzung der schulischen Leistung der eigenen Schülerinnen und Schüler zum Abschluss der Unterrichtseinheit erhoben wurde (Brühwiler, 2017, S. 127–128). Die diagnostisch-didaktischen Handlungs- und Planungskompetenzen, nicht aber die Urteilsakkuratheit, erwiesen sich als signifikante Prädiktoren für den Leistungszuwachs der Schülerinnen und Schüler. Die Urteilsakkuratheit hat demnach keinen direkten Effekt auf die schulischen Leistungen, stand dagegen allerdings stärker mit der wahrgenommenen Unterrichtsqualität im Zusammenhang (Brühwiler, 2017, S. 133). Auch die diagnostische Planungskompetenz und die didaktische Handlungskompetenz hängen positiv mit der Unterrichtsqualität zusammen. Keine Zusammenhänge konnten für die didaktische Planungskompetenz und die diagnostische Handlungskompetenz mit der Unterrichtsqualität gefunden werden (Brühwiler, 2017, S. 131).

Anders et al. (2010) konnten ebenfalls einen positiven Zusammenhang der diagnostischen Kompetenzen von Lehrkräften mit der Unterrichtsqualität in den Klassen aufzeigen, hier bezogen auf den Mathematikunterricht. Die Studie untersuchte 3483 Schülerinnen und Schüler aus 155 Klassen zum Ende der 9. Klasse und in der 10. Klasse. Die diagnostischen Kompetenzen der Lehrkräfte wurden über die Einschätzung der Aufgabenschwierigkeit und über die Leistungseinschätzung erhoben. Betrachtet wurde unter anderem die nachfolgende kognitive Aktivität der Schülerinnen und Schüler, welche die Lernprozesse bezeichnet, die „Schülerinnen und Schüler durch die aktive und selbstständige Auseinandersetzung mit dem Lerngegenstand stimulieren“ (Anders et al., 2010, S. 179). Es zeigte sich ein positiver Zusammenhang zwischen der Fähigkeit, die Aufgabenschwierigkeit einzuschätzen, und der Unterrichtsqualität im Sinne einer Erhöhung der kognitiven Aktivität der Schülerinnen und Schüler. Dieser Zusammenhang fand sich allerdings nicht zwischen der Einschätzung der schulischen Leistungen und der kognitiven Aktivität (Anders et al., 2010, S. 190). Die Autorinnen und Autoren vermuten, dass die Komponente der Einschätzung der schulischen Leistungen andere positive Auswirkungen haben könnte, was in dieser Studie aber nicht überprüft werden konnte (Anders et al., 2010, S. 190–191). Auch wenn die Studie erste Hinweise gibt, so ist zu fragen, ob diese Ergebnisse auf den Bereich der Grundschule und der Grundschullehrkräfte übertragen werden können. Da diagnostische Kompetenzen zudem fachspezifisch sind und es sich um eine Studie handelt, welche das Fach Mathematik im Fokus hat, ist auch hier die Frage zu stellen, inwiefern sich die Erkenntnisse auf das Fach Deutsch und insbesondere auf den Leseerwerb übertragen lassen.

Karing, Pfof und Artelt (2011, S. 123) halten fest, dass sich die wenigen empirischen Studien zu dieser Thematik – wie zuvor schon festhalten – vor allem auf den mathematischen Bereich

konzentrieren und es insgesamt keine eindeutigen Befunde gibt. Sie untersuchten daher den Zusammenhang zwischen der diagnostischen Kompetenz von Lehrkräften und der Lese- sowie der mathematischen Kompetenz von Schülerinnen und Schülern am Ende der fünften und am Ende der sechsten Klasse. Zusätzlich betrachteten sie, ob die beiden Unterrichtsvariablen Individualisierung und Strukturierungshilfen Moderatoren darstellen (Karing, Pfof, et al., 2011, S. 124). 40 Deutsch- sowie 29 Mathematiklehrkräfte von insgesamt 29 Schulen sollten je 7 randomisiert ausgewählte Schülerinnen und Schüler ihrer Klasse danach beurteilen, welche Aufgaben aus den durchgeführten Lese- bzw. Rechentests diese korrekt beantworten würden (Karing, Pfof, et al., 2011, S. 128). Sowohl Haupt- wie auch Realschulen und Gymnasien nahmen an der Studie teil. Es wurden zwei Komponenten der diagnostischen Kompetenz betrachtet: die aufgabenspezifischen Treffer (wie viele Aufgaben für die jeweiligen Schülerinnen und Schüler als gelöst eingeschätzt wurden in Bezug zu den tatsächlich korrekt gelösten Aufgaben) und die Rangordnungskomponente. Der Grad an Individualisierung und die Gabe von Strukturierungshilfen wurden als mögliche Moderatoren für den Zusammenhang zwischen der diagnostischen Kompetenz und der Lesekompetenzentwicklung bzw. der Entwicklung mathematischer Kompetenz im sechsten Schuljahr betrachtet. Für den Bereich der Mathematik gab es keine Zusammenhänge zwischen der diagnostischen Kompetenz, der Individualisierung bzw. der Gabe von Strukturierungshilfen und der Entwicklung der mathematischen Kompetenz. Ob dies daran lag, dass die beiden Komponenten der diagnostischen Kompetenz zu niedrig ausgeprägt waren oder die Mathematiklehrkräfte ihre diagnostischen Kompetenzen nicht zur Adaptation ihres Unterrichts verwendeten, kann mit dieser Studie nicht geklärt werden (Karing, Pfof, et al., 2011, S. 142). Für den Deutschunterricht zeigt sich dagegen ein anderes Bild: Eine hohe Trefferquote hängt hier positiv mit der Lesekompetenzentwicklung der Schülerinnen und Schüler zusammen. Weiterhin wird dieser Einfluss durch den Grad der Individualisierung moderiert, nicht aber durch den Grad an gegebenen Strukturierungshilfen. Dies wird so gedeutet, dass schwächere Schülerinnen und Schüler Strukturierungshilfen benötigen und hiervon profitieren, nicht aber stärkere, die stattdessen eventuell von offenem Unterricht und selbstgesteuertem Lernen deutlicher profitieren (Karing, Pfof, et al., 2011, S. 141). Für die Rangordnungskomponente ergaben sich keine Zusammenhänge. Die Rangordnung war in dieser Studie aber vergleichsweise schwer zu bestimmen, da die Aufgaben nicht stark genug zwischen den Schülerinnen und Schülern differenzierten (Karing, Pfof, et al., 2011, S. 140).

Eine weitere Studie, die sich mit schulischen Leistungen im Deutschunterricht beschäftigt, kommt von Behrmann und Souvignier (2013). Hier wurde über ein Jahr lang das Förderprogramm „Wir werden Textdetektive“ und in Wiederholungsstunden „Wir sind Textdetektive“ in 75 Klassen durchgeführt. An der Studie nahmen entsprechend 75 Lehrkräfte sowie 1865 Schülerinnen und Schüler der fünften bis siebten Klasse teil (Behrmann & Souvignier, 2013, S. 286). Im Fokus lagen die Auswirkungen der Klassenführungskompetenz, der Urteilsakku­ratheit sowie der Rückmeldehäufigkeit der Lehrkräfte auf die Entwicklung der Lesekompetenz

und des Lesestrategiewissens der Schülerinnen und Schüler. In der Auswertung zeigte sich kein signifikanter Zusammenhang zwischen der Klassenführungscompetenz, der Urteilsakkuratheit und der Rückmeldehäufigkeit. Weiterhin hatten alle drei Faktoren keinen signifikanten Einfluss auf das Leseverständnis (Behrmann & Souvignier, 2013, S. 288). Dagegen lag eine mittelgroße signifikante Interaktion zwischen Urteilsakkuratheit und Rückmeldehäufigkeit auf das Leseverständnis vor: Das Leseverständnis erhöhte sich, wenn eine Lehrkraft sowohl eine hohe Urteilsakkuratheit hatte als auch überdurchschnittlich häufig Rückmeldungen über die erbrachte Leistung gab. Bei hoher Urteilsakkuratheit, aber wenig Leistungsrückmeldungen, zeigte sich dieser Effekt dagegen nicht (Behrmann & Souvignier, 2013, S. 288). Bei niedriger Urteilsakkuratheit war die Häufigkeit der Rückmeldungen irrelevant (Behrmann & Souvignier, 2013, S. 288–289). Das Lesestrategiewissen verbesserte sich durch häufigere Rückmeldungen und durch eine höhere Klassenführungscompetenz, nicht aber durch die Urteilsakkuratheit (Behrmann & Souvignier, 2013, S. 289). Die Autoren vermuten, dass Strategiewissen einfacher zu erwerben ist als Leseverständnis und daher Rückmeldungen über die gezeigte Leistung ausreichend sind und alle Lehrkräfte hier wahrscheinlich mithilfe des Förderprogramms fähig waren, adäquates Feedback über die Lesestrategien zu geben. Zudem wurde die Urteilsakkuratheit über das Leseverständnis erhoben, nicht aber über das Lesestrategiewissen (Behrmann & Souvignier, 2013, S. 290).

### 3.4 Zwischenfazit und offene Fragen

In diesem Kapitel wurden bisherige Forschungsergebnisse zu diagnostischen Kompetenzen von Lehrkräften berichtet. Es gibt nicht *die eine* diagnostische Kompetenz, sondern es liegen verschiedene diagnostische Kompetenzen vor, die sich je nach Kontext oder auch Unterrichtsfach voneinander unterscheiden.

Der Forschungsstand zeigt, dass Lehrkräfte zum einen über diagnostische Kompetenzen verfügen sollen und dies auch in unterschiedlich starken Ausprägungen generell tun, zuweilen aber auch ein großer Anteil an Fehleinschätzungen vorkommt und hier starker Verbesserungsbedarf herrscht. Vor allem im Bereich der Lesekompetenz zeigte sich in der PISA 2000-Studie sowie in der IGLU 2006-Studie, dass Lehrkräfte leseschwache Schülerinnen und Schüler zum Teil nicht korrekt als solche identifizierten und ihnen somit die benötigte Unterstützung nicht zukam. Insgesamt konnte die bisherige Forschung allerdings zeigen, dass Lehrkräfte relativ akkurate Aussagen über die Leistungen ihrer Schülerinnen und Schüler treffen können. Dies gilt für starke und durchschnittliche Schülerinnen und Schüler stärker als für schwache. Die Akkuratheit der Lehrkräfte ist vor allem bei Aussagen über die Rangreihenfolge der Leistungen innerhalb der Klasse gegeben. Insgesamt zeigt sich auch, dass die Urteilsakkuratheit für kognitive Merkmale höher ausfällt als für nicht-kognitive Merkmale.

Zwischen einzelnen Lehrkräften herrschen allerdings auch große Diskrepanzen in ihren Fähigkeiten, Schülerinnen und Schüler korrekt einschätzen zu können, weswegen Einflussfaktoren auf Lehrkräfteurteile auch zunehmend in den Blick von Forscherinnen und

Forschern gerückt sind. Auf der einen Seite finden sich verschiedene Schülermerkmale, die die Akkuratheit der Lehrkräfte beeinflussen. Hier sind vor allem das Geschlecht und der sozioökonomische Hintergrund zu nennen. Mehrere Studien konnten zeigen, dass Mädchen bei gleicher Leistung positiver bewertet werden als Jungen und dass Kinder mit einem niedrigeren sozioökonomischen Hintergrund schlechter bewertet werden als Kinder mit gleicher Leistung, aber einem höheren sozioökonomischen Hintergrund.

Auch auf Seiten der Lehrkräfte werden Einflussfaktoren untersucht. Die Berufserfahrung scheint allerdings keinen Einfluss auf die Urteilsakkuratheit zu haben. Lehrkräfte müssen dagegen eine bestimmte Mindestzeit mit den von ihnen einzuschätzenden Schülerinnen und Schülern verbracht haben, um diese akkurat einschätzen zu können. In Studien wurde die Kontaktdauer von einem halben bis eineinhalb Jahren betrachtet, wobei keine Unterschiede mehr gefunden werden konnten. Man kann also davon ausgehen, dass Lehrkräfte nach einem halben Jahr genug Informationen über ihre Schülerinnen und Schüler haben, um diese einschätzen zu können. Für kürzere Zeiten sollte dies kritisch betrachtet werden, da nicht ausgeschlossen werden kann, dass diese Zeit zu gering ist.

Studien konnten ebenfalls Referenzgruppeneffekte der Beurteilung durch Lehrkräfte aufzeigen. So kommt es vor, dass schwächere Schülerinnen und Schüler, die in einer leistungsstarken Klasse sind, schlechter benotet werden als gleich starke Schülerinnen und Schüler, die in einer insgesamt leistungsschwächeren Klasse sind. Umgekehrt kann es auch zum Halo-Effekt kommen, sodass auch schwächere Schülerinnen und Schüler in einer leistungsstarken Klasse stärker eingeschätzt werden, vor allem, wenn sie sich selbst selten beteiligen, da hier die positive Einschätzung über die ganze Klasse auch auf die schwächeren Schülerinnen und Schüler abstrahlt.

Bei der Frage nach den Auswirkungen der diagnostischen Urteile auf die Leistung von Schülerinnen und Schüler ist der Forschungsstand noch lückenhaft. Es gibt Hinweise darauf, dass eine gute Urteilsfähigkeit auch Auswirkungen auf die Planung und Gestaltung des Unterrichts und in der Folge auch auf die Schülerinnen und Schüler haben kann. Speziell für den Bereich des Leseunterrichts konnte gezeigt werden, dass eine hohe Urteilsakkuratheit zusammen mit häufigen Rückmeldungen zu einem erhöhten Leseverständnis führen kann. Ebenso kann eine hohe Urteilsakkuratheit zusammen mit einer Individualisierung des Unterrichts zu einer Leistungssteigerung in der Lesekompetenz führen. Die Studien, die dies aufzeigten, wurden allerdings in der Sekundarstufe I durchgeführt. Ob diese Ergebnisse auch für den Bereich der Grundschule übertragbar sind, ist nicht sicher.

## 4 Über die Lesekompetenz von Grundschulkindern

Lesekompetenz ist eine Fähigkeit, die Schülerinnen und Schüler erwerben müssen, um erfolgreich und ein aktiver selbstständiger Teil der Gesellschaft sein zu können (Hurrelmann, 2013, S. 163; Spinner, 2010, S. 48). Teilweise wird auch darüber diskutiert, ob die Lesekompetenz nicht sogar die mit Abstand wichtigste zu erwerbende Kompetenz ist, da sie nicht nur in den sprachlichen Fächern, sondern auch auf Leistungen im mathematischen Bereich Einfluss hat (Mayer, 2016, S. 46). Ein frühzeitiges Erkennen und Eingreifen bei Schwierigkeiten im Erwerb der Lesekompetenz ist daher auf jeden Fall unabdingbar, sodass hier der diagnostischen Kompetenz einer Lehrkraft besondere Beachtung zukommen sollte (Fischer & Gasteiger-Klicpera, 2013, S. 78; Klicpera et al., 2013, S. 137).

Im Folgenden wird das Konzept der Lesekompetenz im Detail vorgestellt und die Relevanz dieser Kompetenz weiter herausgearbeitet.

### 4.1 Definitionen und Begriffsbestimmung

Nach Philipp (2012, S. 38) findet sich oft keine einheitliche Verwendung der Begriffe *Lesekompetenz* und *Leseverständnis*, weswegen es nicht einfach ist, diese eindeutig zu definieren. Klicpera et al. (2013, S. 69) bezeichnen den Versuch, Leseverständnis zu definieren, ebenfalls als „relativ schwieriges Unterfangen“. Erschwerend kommt noch hinzu, dass die Definition der Lesekompetenz einem historischen Wandel unterzogen ist: Die Bedeutung des Begriffs ist abhängig von „lesebezogenen Bildungsnormen“, die sich mit der Zeit verändern (Garbe, Holle & Jesch, 2009, S. 14). Sicher ist lediglich, dass Lesekompetenz das Leseverständnis im „Kern“ hat, aber über den Begriff des Leseverständnisses noch hinausgeht (Philipp, 2012, S. 38).

Lesekompetenz bezieht sich nach (Artelt et al., 2007, S. 11) auf „eine Disposition, die Personen befähigt, bestimmte Arten von text- und lesebezogenen Anforderungen erfolgreich zu bewältigen“. Wie der Begriff „bewältigen“ bereits impliziert, handelt es sich hierbei nicht um eine passive Fähigkeit. Das Lesen ist vielmehr ein aktiver Prozess, bei dem die lesende Person sich den Lesestoff aneignet und aufbauend auf bereits vorhandenem Welt- wie auch Vorwissen neues Wissen konstruiert (Artelt et al., 2007, S. 11). Dass eine Leserin oder ein Leser sich diese in Texten enthaltenen Wissensinhalte selbst aneignen kann, ist nach Garbe et al. (2009) unabdingbar für lebenslanges und eigenständiges Lernen. Sie referieren über die Entwicklung in der internationalen Bildungspolitik der letzten zwei Jahrzehnte, in deren Zuge der Erwerb von Lesekompetenz als Fähigkeit, sich selbst Wissensinhalte anzueignen, als notwendig und essenziell vor allem im Zusammenhang mit „lebenslangem Lernen“ angesehen wird, wobei die Art der Texte, die zum Erwerb der Kompetenz gelesen werden, von niedriger Relevanz ist (Garbe et al., 2009, S. 19). Auch Groeben (2009, S. 12) differenziert bei der Begriffsbestimmung zwischen der Kulturtechnik *Lesen* und der Lesekompetenz als solcher, welche eben mehr ist als das reine Lesen, wie es im Leseunterricht der ersten Schuljahre erlernt wird. Gleichzeitig sieht er eine Unterscheidung zwischen gelesenen Textarten

(in literarische und nicht-literarische) als nicht sinnvoll an, da es bei der Lesekompetenz um „das Lesen generell“ geht (Groeben, 2009, S. 12). Hier stehen also ganz deutlich die Kompetenz und damit verbunden die Fähigkeit, diese in unterschiedlichen Situationen einzusetzen, im Mittelpunkt, nicht jedoch die gelesenen Texte an sich. Dies zeigt sich auch darin, dass in deutschen Curricula hierzu zunehmend Kompetenzen aufgelistet sind, die erworben werden sollen und weniger häufig konkret zu lernende Bildungsinhalte (Garbe et al., 2009, S. 19–20). Allerdings beantwortet auch dies nur unzureichend, was unter Lesekompetenz genau zu verstehen ist.

In diesem Jahrtausend wurde die Lesekompetenz von Schülerinnen und Schülern in drei großen und zum Teil internationalen Studien untersucht: IGLU und PISA als internationale Studien sowie DESI für den deutschen Raum.

In den IGLU-Studien, in welchen Grundschulkindern im Fokus stehen, wird von einem aktiven Modell des Leseverständnisses ausgegangen, in dem die Leserin oder der Leser mit dem zu lesenden Text und dem Kontext interagiert (Garbe et al., 2009, S. 24). Auch in den PISA-Studien, bei denen die Lesekompetenz von 15-jährigen Schülerinnen und Schülern untersucht wird, findet sich eine ähnliche Definition, wobei sich sowohl PISA als auch IGLU am angloamerikanischen Begriff „reading literacy“ orientieren. Die Lesekompetenz wird in den IGLU-Studien beschrieben als „die Fähigkeit [...], in unterschiedlichen, für die Lebensbewältigung praktisch bedeutsamen Verwendungssituationen lesen zu können“ (Hußmann, Wendt, et al., 2017, S. 37). Ferner sind dort drei Teilbereiche der Lesekompetenz aufgeführt, welche im Fokus der Erhebungen stehen: Leseintentionen, Leseverstehensprozesse sowie Leseselbstkonzept, -motivation und -verhalten (Hußmann, Wendt, et al., 2017, S. 37). In den PISA-Studien steht Lesekompetenz für die Fähigkeit, „geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potenzial weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen“ (Baumert, Klieme, et al., 2001, S. 80). Hier beschränkt sich die Definition auf den kognitiven Aspekt, in dem Lesekompetenz als Fähigkeit gesehen wird, sich aktiv mit Texten auseinanderzusetzen zu können und abhängig vom eigenen Vorwissen die Bedeutung des Textes zu konstruieren (Garbe et al., 2009, S. 21–22). Die Lesekompetenz der Jugendlichen wird in fünf Kompetenzstufen unterteilt. Kompetenzstufe I wird beispielsweise erreicht, wenn Schülerinnen und Schüler in einem Text „eine oder mehrere unabhängige, aber ausdrücklich angegebene Informationen“ (Artelt et al., 2001, S. 89) finden können, Hauptgedanken oder Absicht der Autorin oder des Autors, welche wiederholt oder deutlich formuliert sind, erkennen oder wenn einfache Verknüpfungen zwischen Gelesenem und eigenem Alltagswissen hergestellt werden können. Die höchste Stufe, Kompetenzstufe V, zeichnet sich unter anderem dadurch aus, dass Leserinnen und Leser „tief eingebettete Informationen“ in unbekanntem Texten finden können, auch bei unbekanntem Format das Gelesene im Detail verstehen, kritisch bewerten sowie Hypothesen dazu bilden können (Artelt et al., 2001, S. 89).

DESI („Deutsch-Englisch-Schülerleistungen-International“) ist eine ausschließlich deutsche Studie, bei der 11.000 Schülerinnen und Schüler der neunten Klassen in allen Schularten sowie deren Lehrkräfte, Eltern und Schulleitungen befragt bzw. untersucht wurden (Klieme, 2008, S. 1). Neben der Erfassung des Leistungsstands sollten auch Anregungen zur Optimierung und Anpassung von Curricula und Unterrichtsmaterialien sowie Verbesserungsmöglichkeiten in der Lehrkräfteausbildung herausgearbeitet werden (Klieme, 2008, S. 1). Es standen also verstärkt didaktische Fragen im Vordergrund (Garbe et al., 2009, S. 26f). Der Begriff der Lesekompetenz stützt sich bei dieser Studie auf ein Prozessmodell, womit sich DESI sowohl von den IGLU-Studien als auch von den PISA-Studien unterscheidet. Gestützt auf empirische Ergebnisse konnten hier vier Niveauelemente ausgemacht werden (Gailberger & Willenberg, 2008, S. 61). Auf der untersten Niveauelemente A kann eine Leserin oder ein Leser einem Text basale Informationen mithilfe einzelner Wörter entnehmen. Auf Niveaustufe B gelingt es ihr oder ihm, Inferenzen zu bilden und somit z. B. zwischen einzelnen Sätzen eine sinnhafte Verknüpfung zu bilden. Zudem beinhaltet diese Niveaustufe auch die Fähigkeit, den Fokus beim Lesen auf einzelne relevante Textstellen zu richten. Auf der dritten Stufe, Niveaustufe C, erfolgen ebenfalls Verknüpfungen beim Lesen. Hier geschieht dies aber nicht mehr nur mit Sätzen oder Informationen im gleichen Text, sondern mit bereits vorhandenem Vor- und Weltwissen. Außerdem können auch Beziehungen zwischen Inhalten aus verschiedenen Absätzen hergestellt werden, während die Verknüpfungen auf Niveaustufe B lediglich innerhalb des gleichen Absatzes gemacht werden können (Gailberger & Willenberg, 2008, S. 61). Auf der obersten Stufe, Niveaustufe D, kann eine Leserin oder ein Leser mentale Modelle von gelesenen Texten entwickeln. Diese werden als „konstruktive Zusammenfassung des Textes durch den Leser [sic!]“ definiert (Gailberger & Willenberg, 2008, S. 61–62). Auch hier spielt wieder das Vor- und Weltwissen eine große Rolle. Im Unterschied zur Niveaustufe C werden Verknüpfungen für den gesamten Text erstellt (Gailberger & Willenberg, 2008, S. 62).

Bemängelt wird bei allen diesen Studien, dass sich die Konzepte zu stark auf den kognitiven Aspekt des Lesens beschränken. Hurrelmann (2013, S. 162) kritisiert an dieser Stelle, dass die Definition der Lesekompetenz in den PISA-Studien nicht weitreichend genug ist und andere Bereiche, wie die Lesemotivation, unberücksichtigt bleiben. Sie erweitert die Lesekompetenz entsprechend um den Bereich des Kompetenzerwerbs sowie zusätzlich zur in den PISA-Studien vorhandenen kognitiven Komponente des Leseverständnisses um motivationale, emotionale und interaktionsbezogene Komponenten sowie deren normative Zusammenhänge. Das Konstrukt der Lesekompetenz fokussiert nach ihr daher zum einen auf die kognitive Komponente, zum anderen aber auch auf Bereiche wie Lesemotivation, Leselust und Gespräche über Gelesenes.

Auch Rosebrock und Nix (2015, S. 13–14) sind der Meinung, dass die in den PISA-Studien verwendete Definition zu kurz greift, da diese nicht auf die Erwerbswege oder auf die sozialen Faktoren des Leseerwerbs eingeht. Ihr Ziel war es, eine Definition zu finden, die für

didaktische Zwecke nutzbar ist. Dadurch soll es Lehrkräften ermöglicht werden, aufgrund dieser Definition sowohl Schwierigkeiten im Lesen zu diagnostizieren als auch geeignete Interventionsmaßnahmen auszuwählen. Rosebrock und Nix unterteilen das Lesen hierfür in drei Ebenen. Im Zentrum stehen die kognitiven Tätigkeiten auf der Prozessebene. Hierzu zählen zum Beispiel die Identifikation von Wörtern und Sätzen sowie die Bildung von lokalen wie auch globalen Kohärenzen (Rosebrock & Nix, 2015, S. 15). Auf der Subjektebene finden sich Faktoren, welche sich auf Eigenschaften innerhalb der Leserin oder des Lesers beziehen, wie zum Beispiel deren lesebezogenes Selbstkonzept oder auch die Lesemotivation. Die dritte Ebene beinhaltet soziale Faktoren, zu welcher sowohl die Familie wie auch Peers und die Schule gehören. Hier steht die Anschlusskommunikation, also Gespräche über das Gelesene, im Fokus. Dieses Modell ähnelt dem von Hurrelmann, da beide Modelle auch nicht-kognitive Faktoren einbeziehen. Auf der anderen Seite kritisiert Müller allerdings, dass es sich bei der Lesekompetenz um eine kognitive Fähigkeit handelt, weswegen Komponenten wie Lesemotivation nicht Bestandteil sein sollten, da sie „zu einer Aufweichung des Begriffs“ (Müller, 2014, S. 26) führen würden.

Häufig findet sich dagegen in den Definitionen eine Unterteilung in zwei Ebenen: die basale Lesefertigkeit als niedrigere und das Leseverständnis als höhere Ebene (Klicpera et al., 2013, S. 68–70; Munser-Kiefer & Kirschhock, 2009, S. 166). Kamzela (2015, S. 60) hält fest, dass basale Lesefertigkeiten zunächst in ausreichendem Maß erworben werden müssen. Als hierarchieniedrige Prozesse setzen sich diese aus der Worterkennung über Rekodieren und Dekodieren, dem Verstehen von einzelnen Sätzen und der Bildung lokaler Kohärenzen zusammen. Unter Dekodieren wird die Fähigkeit verstanden, Wörter ohne Verlesungen zu erkennen bzw. Verlesungen direkt korrigieren zu können (Rosebrock & Nix, 2015, S. 37). Lokale Kohärenzen müssen gebildet werden können, um z. B. die Grammatik richtig zu verstehen oder auch Rückverweise (als Beispiel: Die Kinder waren im Zoo. Sie hatten Spaß.) korrekt zu interpretieren (Kamzela, 2015, S. 55). Eine weitere Fertigkeit auf der niedrigeren Ebene ist die Leseflüssigkeit, welche ebenfalls die Dekodierung von Wörtern beinhaltet (sowohl die grundlegende Fähigkeit hierzu als auch die folgende Automatisierung), zusätzlich aber auch die Fähigkeit, Texte in einer angemessenen Geschwindigkeit lesen und sinngemäß betonen zu können (Kamzela, 2015, S. 55; Rosebrock & Nix, 2015, S. 37).

Das Leseverständnis als höhere Ebene bildet nicht nur eine Komponente von Lesekompetenz, sondern sie ist selbst ebenfalls ein mehrdimensionales Konstrukt, auf welche die basale Lesefertigkeit als Einflussfaktor einwirkt (Klicpera et al., 2013, S. 70). Weitere Komponenten des Leseverständnisses sind nach Klicpera et al. (2013) das Wort-, Satz- und Textverständnis, Inferenzbildung, Differenzierung zwischen zentralen Inhalten und Detailinhalten, Verständnis für Textstrukturen und Diskursformen, Vorwissen, Vorerfahrung und Interesse, metakognitives Bewusstsein sowie das Überwachen des eigenen Verständnisses (Klicpera et al., 2013, S. 71).

Insgesamt lässt sich festhalten, dass Lesekompetenz als Begriff nicht einheitlich definiert ist, aber übereinstimmend als aus basalen Fertigkeiten wie dem Dekodieren von Wörtern und der Lesegeschwindigkeit sowie dem Leseverständnis bestehend interpretiert wird. Teilweise enthalten die Definitionen der Lesekompetenz zudem neben kognitiven auch soziale und motivationale Komponenten.

## 4.2 Relevanz von Lesekompetenz

Wie eingangs bereits erwähnt, kommt der Lesekompetenz ein hoher Stellenwert zu, weswegen diese Relevanz noch einmal explizit erörtert werden soll.

Hurrelmann hält fest, dass „die Lesekompetenz eine wichtige Bedingung für die Aneignung weiterer Kompetenzen darstellt und deshalb eine Schlüsselstellung in der Tradierung kulturellen Kapitals hat“ (2013, S. 163).

Kinder mit niedriger Lesekompetenz werden spätestens in der Sekundarstufe I Schwierigkeiten haben, erfolgreich am Unterricht in allen Unterrichtsfächern, in denen die Schriftsprache eine Rolle spielt, teilnehmen zu können (Tarelli, Valtin, Bos, Bremerich-Vos & Schwippert, 2012, S. 13; Valtin et al., 2010, S. 78). Mayer (2016, S. 46) sieht die Notwendigkeit, über eine ausreichende Lesefähigkeit zu verfügen, ab spätestens der dritten Klasse und dies ebenfalls für alle diese Unterrichtsfächer. Den ersten beiden Schuljahren kommt daher eine wichtige Rolle im Erwerb der Lesefähigkeit zu.

Paetsch, Radmann, Felbrich, Lehmann und Stanat (2016) gingen in ihrer Studie der Frage nach, inwiefern die Lesekompetenz sich auf die mathematischen Kompetenzen von Schülerinnen und Schülern auswirkt. Hierfür wurden Kinder in einer Längsschnittstudie von der vierten bis zur sechsten Klasse betrachtet. Auch bei Konstanthaltung des sozioökonomischen Status und der kognitiven Grundfähigkeiten zeigte sich ein signifikanter Zusammenhang zwischen der Lesekompetenz in der vierten Klasse und den mathematischen Leistungen in derselben Klassenstufe, aber auch ein positiver Zusammenhang zwischen der Lesekompetenz der vierten Klasse und den Lernzuwachsraten in Mathematik, sodass die Lesekompetenz einen starken Prädiktor darstellt. Dieser blieb auch bestehen, wenn neben dem sozioökonomischen Status und den kognitiven Grundfähigkeiten zusätzlich die Familiensprache kontrolliert wurde (Paetsch et al., 2016, S. 35–36).

Die Notwendigkeit, Texte sinnentnehmend lesen zu können, beschränkt sich allerdings nicht nur auf die Schule. Auch im Berufsalltag wird die Fähigkeit des lebenslangen Lernens zunehmend wichtiger, da erwartet wird, dass zum Beispiel regelmäßig Fort- und Weiterbildungen absolviert werden. Insgesamt wird dem Lesen in unserer Gesellschaft daher eine Schlüsselkompetenz für ein erfolgreiches und selbstständiges Leben zugeschrieben, auch für Berufsfelder, in denen diese früher nicht relevant war, wie beispielsweise der Landwirtschaft (Spinner, 2010, S. 48). Beck, Jäpel und Becker (2010) betrachteten in ihrer Studie den Bildungserfolg von Kindern mit Migrationshintergrund in der Schweiz. Lesekompetenz zeigt einen negativen Zusammenhang mit einer fehlenden Ausbildung: Eine hohe Lesekompetenz ging mit

einer niedrigeren Ausbildungslosigkeit einher (Beck et al., 2010, S. 332). Eine internationale Studie der OECD, an der auch Deutschland teilgenommen hat, konnte zeigen, dass die Lesekompetenz im Erwachsenenalter unter anderem mit einem höheren Einkommen und einer höheren Wahrscheinlichkeit, erwerbstätig zu sein, zusammenhängt (Organisation for economic co-operation and development, 2000, S. 84).

Baumert und Maaz (2010, S. 163) halten zusammenfassend fest, dass Schülerinnen und Schüler, die laut dem Lesetest der PISA-Studie nicht über Kompetenzstufe I im Lesen hinaus kamen, in der Folge auch in Bezug auf ihre berufliche Ausbildung eine Risikogruppe darstellen. In der PISA-Studie 2000 zeigte sich bei den 15-jährigen Schülerinnen und Schülern ein Anteil von 21 Prozent, der die Kompetenzstufe I nicht überschritten hatte. Sie schlussfolgern, dass diese Schülerinnen und Schüler in der Folge u. a. Schwierigkeiten beim Übergang in die berufliche Erstausbildung haben würden. Im Jahr 2006 waren 17 Prozent der Altersgruppe der 20- bis 30-Jährigen ohne abgeschlossene Berufsausbildung und ohne Teilnahme an einer Bildungsmaßnahme. Die Ergebnisse bezeichnen sie als „konsistent“ zueinander (Baumert & Maaz, 2010, S. 163).

#### 4.3 Zwischenfazit

Insgesamt lässt sich festhalten, dass eine ausreichende Lesekompetenz eine selbstständige Teilhabe am gesellschaftlichen Leben ermöglicht (Hurrelmann, 2013, S. 163; Spinner, 2010, S. 48). Auch wenn Lesekompetenz nicht einheitlich definiert ist, wird doch aus den bisherigen Erläuterungen deutlich, dass die Fähigkeit, adäquat lesen zu können und den Texten Sinn entnehmen zu können, unabdingbar ist für den Schulerfolg. Es wird erwartet, dass Kinder ab der dritten Klasse, spätestens aber zum Ende der Grundschule, die Fähigkeit, flüssig zu lesen und Texten Sinn entnehmen zu können, ausreichend erworben haben. Daher kommt der Grundschule bzw. den Grundschullehrkräften eine entscheidende Bedeutung zu. Sie müssen dafür sorgen, dass möglichst alle Schülerinnen und Schüler über eine ausreichende Lesekompetenz verfügen.

## 5 Der ungestörte Leseerwerb

Nachdem das Konzept der Lesekompetenz dargestellt wurde, ist es noch notwendig, den ungestörten Erwerb dieser Fähigkeit zu betrachten, um in den folgenden Abschnitten mögliche Schwierigkeiten beim Erwerb zu beleuchten und herauszuarbeiten, woran Lehrkräfte diese ausmachen können. Nach Groeben (2009, S. 13) enthält die Lesekompetenz keine erbliche Komponente. Dementsprechend befinden sich im Gehirn auch keine Areale, die spezifisch für das Lesen zuständig sind (Steinbrink & Lachmann, 2014, S. 2). Es handelt sich hierbei also nicht um eine Fertigkeit, die sich automatisch im Laufe der Entwicklung eines Kindes herausbildet und selbstständig aufbaut, sondern sie wird im Laufe der Sozialisation erworben. Daraus ergibt sich, dass es gezieltem Training und langer Übung bedarf, um über eine ausreichend gute Lesekompetenz zu verfügen.

In der Forschung und in der Literatur findet sich eine Vielzahl an theoretischen Modellen zum Leseerwerb. Diese Modelle lassen sich in zwei Arten unterteilen: Auf der einen Seite gibt es Entwicklungsmodelle, mit denen der Ablauf des Schriftspracherwerbs dargestellt werden soll (z. B. Frith, 1986; Günther, 1986; Klicpera, Schabmann, Gasteiger-Klicpera & Schmidt, 2017) und auf der anderen Seite finden sich Prozessmodelle, welche die Vorgänge beschreiben, die während des Lesens stattfinden (z. B. Coltheart, 2006; Seidenberg, 2005). Für Lehrkräfte haben diese theoretischen Modelle vor allem den Nutzen, dass diese ihnen eine Orientierung geben können, auf welchem Lernstand ihre Schülerinnen und Schüler jeweils sind und welche Fähigkeiten sie als nächstes erwerben sollten (Klicpera et al., 2017, S. 59).

Aus Gründen der Überschaubarkeit werden im Folgenden nur einige ausgewählte und einflussreiche Modelle genauer betrachtet. Lesen und Schreiben entwickeln sich zudem nicht unabhängig voneinander (Schilcher, 2012, S. 25). Aufgrund der dieser Untersuchung zugrundeliegenden Thematik der Lesekompetenz wird sich in den folgenden Modellen aber weitgehend auf den Bereich des Leseerwerbs konzentriert und der Rechtschreiberwerb wird nur insofern thematisiert werden, wie er für das Modell im Zusammenhang mit dem Leseerwerb relevant ist.

### 5.1 Grundprinzipien der deutschen Sprache im Schriftspracherwerb

Oftmals werden Studien aus dem englischsprachigen Raum übernommen und auf den Schriftspracherwerb von deutschsprachigen Kindern übertragen (Klicpera et al., 2017, S. 89). Wenn man die für die englische Sprache erlangten Erkenntnisse aber auf den deutschsprachigen Raum übertragen möchte, muss man stets bedenken, dass es nicht nur Gemeinsamkeiten gibt, sondern dass sich die englische und die deutsche Sprache in einigen Bereichen stark voneinander unterscheiden, beispielsweise in der Ausprägung der Phonem-Graphem-Korrespondenz (Klicpera et al., 2017, S. 47). Daher ist es sinnvoll, zunächst die Grundprinzipien der deutschen Sprache zu beleuchten, um in der Folge die Ergebnisse aus der Forschung, aber auch die Modelle des Leseerwerbs entsprechend werten zu können.

Die deutsche Schriftsprache entspricht einem phonologischen System (Schröder-Lenzen, 2013, S. 16). Die Verschriftlichung orientiert sich in solch einem System an der lautlichen Struktur der Sprache und nicht an der Bedeutung des Wortes (Mayer, 2016, S. 14;16). Dies steht im Gegensatz zu logographischen Systemen, wie es die Kanji im japanischen Schriftsystem (木, 白), aber auch die arabischen Zahlen (1, 2) beispielsweise sind. Diese Schriftzeichen stellen bildlich die Bedeutung der Wörter dar, stehen aber nicht im Zusammenhang mit der Aussprache (Mayer, 2016, S. 14–15).

Wörter in der gesprochenen Sprache lassen sich in Phoneme unterteilen. Ein Phonem ist die kleinste bedeutungstragende Einheit der Lautsprache, beispielsweise /h/ in Hut und /w/ in Wut. Verschriftlicht werden Wörter in phonologischen Systemen mit sogenannten Graphemen, den kleinsten bedeutungstragenden Einheiten der Schriftsprache (im eben genannten Beispiel also <h> bei Hut und <w> bei Wut sowie jeweils das <u> und das <t>). Im hier verwendeten Beispiel stimmen Phoneme und Grapheme miteinander überein. In der deutschen Sprache gibt es nicht immer eine solche 1:1 Zuordnung zwischen Buchstaben und Lauten (Schröder-Lenzen, 2013, S. 16), jedoch finden sich insgesamt sehr viele davon, sodass die deutsche Sprache als relativ transparent betrachtet wird (Marx, 2007, S. 109; Mayer, 2016, S. 17; Richter & Müller, in Druck, S. 9). So gibt es in der deutschen Lautsprache ungefähr 40 Phoneme, die durch insgesamt 31 Grapheme dargestellt werden (Mayer, 2016, S. 20). Bei der Verschriftlichung der gesprochenen Wörter werden Phoneme in entsprechende Grapheme umgewandelt. Man spricht von der Phonem-Graphem-Korrespondenz, um auszudrücken, wie genau diese Zuordnung in einer Sprache stattfinden kann. Es kann unterschiedliche Buchstaben und Buchstabenkombinationen geben, die den gleichen Laut schriftlich wiedergeben (zum Beispiel <sch> in schreiben, aber <s> in sprechen). Eine hohe Phonem-Graphem-Korrespondenz, wie sie im Deutschen vorkommt, bedeutet, dass ein Phonem durch genau ein Graphem oder nur in Ausfällen durch mehr als ein Graphem verschriftlicht wird. Ein Beispiel hierfür gibt Schröder-Lenzen: Das lang gesprochene [i:] wird in der Regel durch <ie> verschriftlicht (z. B. fliegen), selten aber durch <ieh> (ziehen), <ih> (ihn) oder gar <i> (Tiger) (Schröder-Lenzen, 2013, S. 21).

Als Graphem-Phonem-Korrespondenz wird die Zuordnung von geschriebenen Buchstaben bzw. Graphemen zu ihrer Aussprache bezeichnet. Dies macht sich oft bei Fremdwörtern bemerkbar. So werden die Wörter Psyche, Yards und Nylon mit <y> geschrieben, aber jeweils unterschiedlich ausgesprochen. Ein weiteres Beispiel ist der Buchstabe v, welcher keinen eigenen Laut hat und als /f/ (Vogel) oder als /w/ (Vase) ausgesprochen wird (Schröder-Lenzen, 2013, S. 25–26).

Die Graphem-Phonem-Korrespondenz ist im deutschen Schriftsystem höher als die Phonem-Graphem-Korrespondenz (Mayer, 2016, S. 17). Die Zuordnungen sind insgesamt aber auch in der deutschen Sprache relativ komplex. Im Vergleich zu anderen Sprachen zeigen sie jedoch eine relativ hohe Transparenz und sind weitgehend regelhaft (Mayer, 2016, S. 17; Schröder-Lenzen, 2013, S. 179). Es spricht auch für sich, dass das erste Beispiel zur

Graphem-Phonem-Korrespondenz (<y>) unter anderem mit englischen Fremdwörtern erfolgte, da diese weitaus schwieriger und vielfältiger ist als im Deutschen (Schrüder-Lenzen, 2013, S. 26).

## 5.2 Prozessmodelle des Leseerwerbs

Prozessmodelle versuchen das abzubilden, was während des Lesens geschieht. Im Gegensatz zu Entwicklungsmodellen fokussieren sie oft nicht auf Schülerinnen und Schüler, die gerade das Lesen erlernen. Im Falle des Zwei-Wege-Modells des (lauten) Wortlesens liegt der Fokus beispielsweise auf erfahrenen Leserinnen und Lesern. Für dieses Modell wurden zudem Personen betrachtet, die aufgrund von Hirnschädigungen Schwierigkeiten in Teilbereichen des Lesens haben (Coltheart, 2006, S. 9–10), um hierüber einen tieferen Einblick in die Prozesse des Lesens zu erhalten.

### 5.2.1 Zwei-Wege-Modell des (lauten) Wortlesens

Das Zwei-Wege-Modell des (lauten) Wortlesens ist nicht für den Prozess des Leseerwerbs entwickelt worden, sondern verdeutlicht stattdessen den Prozess, der beim (lauten) Lesen selbst abläuft. Es wurde ursprünglich in den 1970er Jahren von Coltheart und Kollegen erstellt (Coltheart, 1978), seitdem aber erweitert (Coltheart, 2005a, S. 9–13). Das Modell konzentriert sich auf die Worterkennung, da Erkenntnisse über die Abläufe der Worterkennung nach Coltheart (2006, S. 6) auch einen Beitrag zum Verständnis über das Satz- und Leseverständnis leisten.

Das Zwei-Wege-Modell geht davon aus, dass ein Wort auf zwei Arten gelesen werden kann: Entweder geschieht dies direkt über das lexikalische Lesen oder über die nicht-lexikalische Route. Wenn ein Wort über die lexikalische Route abgerufen werden kann, bedeutet das, dass es im mentalen Lexikon der Leserin oder des Lesers abgespeichert ist (Coltheart, 2006, S. 8–9). An dieser Stelle unterscheidet sich das weiterentwickelte von dem ursprünglichen Modell. Anfänglich sah das Modell nur die Route über „das mentale Lexikon“ vor, unterteilte dieses aber nicht weiter (Coltheart, 2006, S. 8). Nach dem erweiterten Modell liegen im mentalen Lexikon drei Arten an Informationen über das Wort vor: Informationen über die Schreibweise (orthographisches Lexikon), Informationen über die Bedeutung (semantisches System) sowie Informationen über die Aussprache des Wortes (phonologisches Lexikon). Wenn ein Wort dagegen nicht im mentalen Lexikon abgespeichert ist, wird auf die nicht-lexikalische Route zurückgegriffen. Auch hier unterscheidet sich das erweiterte vom ursprünglichen Modell und gibt nun explizit an, wie dieses nicht-lexikalische Lesen stattfindet: Das Erlesen geschieht in diesem Fall über die gelernten Graphem-Phonem-Zuordnungen (Coltheart, 2006, S. 10). Es kommt hier folglich zum phonologischen Rekodieren, also zum synthetisierenden Lesen eines Wortes. Diese Art des Lesens zeigt sich bei unbekanntem Wörtern, da diese zwangsläufig nicht Bestandteil des mentalen Lexikons sein können (Klicpera et al., 2017, S. 46–47).

Die zwei Wege – lexikalisches und nicht-lexikalisches Lesen – konnten bei Personen mit Hirnschädigungen nachgewiesen werden, die in Folge dieser Hirnschädigungen an erworbener Dyslexie leiden, welche sich bei Fehlern im lauten Wortlesen in eben dieser lexikalischen oder nicht-lexikalischen Route zeigen (Coltheart, 2005b, S. 10).

### 5.2.2 Konnektionistische Modelle der Worterkennung

Die konnektionistischen Modelle der Worterkennung zählen ebenfalls zu den Prozessmodellen und entstanden als Kritik zum Zwei-Wege-Modell (Mayer, 2016, S. 32). Es wird in diesen Modellen davon ausgegangen, dass das Lesen mit nur einem Verarbeitungssystem abgebildet werden kann und nicht wie bei Colthearts Modell mit zwei getrennten Systemen, der lexikalischen und der nicht-lexikalischen Route (Klicpera et al., 2017, S. 50). Das konnektionistische Modell von Seidenberg (2005) besteht aus drei Ebenen, welche jeweils eine andere Art an Informationen über ein Wort enkodieren: Semantik, Orthographie und Phonologie. Diese drei Ebenen werden beim Lesen aktiviert und interagieren miteinander, damit es schließlich zum korrekten Erlesen des Wortes kommt, wobei alle drei Ebenen jeweils mit unterschiedlichen Gewichtungen in das Erlesen einfließen und somit manche Verbindungen wahrscheinlicher sind als andere (Seidenberg, 2005, S. 239). Das Modell geht weiterhin davon aus, dass die Aktivierung der semantischen Einheiten simultan sowohl durch visuelle wie auch durch phonologische Prozesse erreicht wird (Seidenberg, 2005, S. 241). Mayer (2016, S. 33) gibt ein Beispiel für die Interaktion bei der Aktivierung für den deutschsprachigen Raum: Das Graphem <v> wird sehr häufig als /f/ (Vogel, Veilchen), aber nur vergleichsweise selten als /w/ (Vase) ausgesprochen. Daher ist die erste Verbindung wahrscheinlicher als die zweite und ist somit enger verknüpft. Je häufiger diese beiden Komponenten gemeinsam auftreten, umso schneller wird diese Verbindung auch in Zukunft aktiviert (Mayer, 2016, S. 34).

Nach Richter und Müller (in Druck, S. 5) entsprechen die aus diesen Modellen entstehenden Vorhersagen oft denen aus den Zwei-Wege-Modellen.

### 5.2.3 Simple View of Reading

Das Modell des Simple View of Reading von Gough und Tunmer (1986) stellt kein eigenständiges Modell dar, sondern ergänzt die Zwei-Wege-Modelle sowie die konnektionistischen Modelle und berücksichtigt neben der Worterkennung auch das Leseverständnis auf Satz- und Textebene (Mayer, 2016, S. 35). Die Autoren führen außerdem an, dass es ihnen vor allem um die Rolle des Dekodierens im Leseprozess geht (Gough & Tunmer, 1986, S. 6).

Es handelt sich hierbei um ein sehr bekanntes und – wie der Name schon suggeriert – um ein leicht nachzuvollziehendes Modell. Es beschreibt, dass sowohl die Dekodierfähigkeit wie auch das Sprachverständnis benötigt werden, um einen Text lesen zu können. Dargestellt wird dies durch die Formel:  $R = D \times C$ , wobei R für reading steht, D für decoding und C für comprehension (im Sinne des Sprachverständnisses). Beide Faktoren D und C stehen in

einem multiplikativen Zusammenhang und sind somit zum einen nicht unabhängig voneinander, beschreiben aber zum anderen sowohl eine Voraussetzung als auch eine Komponente der Lesefähigkeit (Gough & Tunmer, 1986, S. 7). Nach Mayer (2016, S. 37) können die in diesem Modell gemachten zentralen Annahmen als empirisch bestätigt angesehen werden.

#### 5.2.4 Prozessmodell des Lesenlernens

Das Prozessmodell des Lesenlernens von Marx (2007) versteht sich als Integrationsmodell, welches sowohl den Simple View of Reading-Ansatz wie auch die Stufenmodelle, welche im nächsten Abschnitt vorgestellt werden, und die zuvor genannten konnektionistischen Modelle berücksichtigt. Im Fokus stehen zudem die spezifischen Teilfertigkeiten des Leseprozesses. Marx (2007, S. 115) merkt an, dass der Leseerwerb nicht erst mit Schuleintritt beginnt, was ebenfalls bei den Stufenmodellen deutlich wird. Daher hat dieses Modell seinen Ausgangspunkt bei unspezifischen Merkmalen wie der sozialen Herkunft, kognitiven und sprachlichen Fähigkeiten oder dem Arbeitsgedächtnis. Diese Merkmale wirken unter anderem auf die phonologischen Fertigkeiten der Kinder, welche hier als „(bereichs-)spezifische Vorläuferfertigkeiten der Worterkennung und des Wortverstehens“ bezeichnet werden (Marx, 2007, S. 115–116). Marx sieht diese Fertigkeiten als fast ausschließlich relevant für den Erwerb und Gebrauch der Schriftsprache. Die phonologischen Fertigkeiten und die Fertigkeiten zur Rekodierung bewirken den Aufbau der Cipher-Bewusstheit. Schröder-Lenzen (2013, S. 54) bezeichnet die Cipher-Bewusstheit als den Sichtwortschatz, den Leserinnen und Leser mit Übung aufbauen: Ausgehend von den Graphem-Phonem-Korrespondenzen eines Wortes bildet sich eine Gedächtnisspur, die auch über diese Buchstaben-Laut-Ebene hinausgeht. Diese Gedächtnisspur wird auch als „orthographic cipher“ bezeichnet. Hiermit können einzelne Teile eines Wortes oder das Wort als Ganzes abgespeichert werden (Marx, 2007, S. 113–114). Marx weist darauf hin, dass mit diesem Begriff nicht die visuelle Wahrnehmung bezeichnet wird, sondern die Abspeicherung über einen Verbund von „orthographischen Informationen, phonologischer Rekodierung, Aussprache und Bedeutung“ (Marx, 2007, S. 114). Der Aufbau dieser Cipher-Bewusstheit sorgt für eine Steigerung der Lesegeschwindigkeit. Relevant für dieses Modell ist ebenfalls das Hörverstehen, welches sich auf das Leseverständnis auswirkt. Marx sieht hier einen multiplikativen Zusammenhang zwischen Hörverstehen und der Fähigkeit des Rekodierens (2007, S. 117). Das Hörverstehen wird durch andere im Modell genannte Faktoren bedingt wie zum Beispiel dem Arbeitsgedächtnis, der Sprachentwicklung, aber auch der Cipher-Bewusstheit (Marx, 2007, S. 115). Als Extrembeispiel nennt Marx das Lesen von Wörtern einer fremden Sprache: Bei ausreichenden Rekodierfähigkeiten gelingt das Erlesen, das resultierende Leseverständnis wird aber bei null Prozent liegen. Entsprechend führt ein Einüben der Worterkennung nicht automatisch zu einem verbesserten Leseverständnis auf Satz- oder Textebene, wenn das Hörverstehen sich nicht ebenfalls entsprechend weiterentwickelt (Marx, 2007, S. 117).

### 5.3 Entwicklungsmodelle des Leseerwerbs

Entwicklungsmodelle versuchen, den Ablauf des Schriftspracherwerbs zu modellieren. Diese Modelle bestehen aus mehreren Phasen, die sich meist qualitativ voneinander unterscheiden und nacheinander durchlaufen werden. Zudem berücksichtigen die meisten Modelle, dass der Schriftspracherwerb nicht erst mit Eintritt in die Schule beginnt und setzen daher entsprechend früher an (Marx, 2007, S. 105).

Für die einzelnen Entwicklungsstufen werden meist charakteristische vorherrschende Strategien beschrieben, die die Lernenden zum Lesen oder Schreiben einsetzen (Marx, 2007, S. 105). Zudem werden häufige bzw. für die jeweilige Entwicklungsstufe relevante Fehler genannt. Dies hat den Vorteil, dass Fehler direkt als Indiz genutzt werden können, auf welcher Stufe des Schriftspracherwerbsprozesses sich ein Kind befindet. Es wird aber auch darauf hingewiesen, dass nicht jedes Kind strikt diese Phasen durchläuft. Die einzelnen Entwicklungsstufen können sich auch überlappen und teilweise parallel verlaufen oder von einem Kind viel schneller oder langsamer durchlaufen werden als andere Phasen (Mayer, 2016, S. 38; Schröder-Lenzen, 2013, S. 75).

Auch muss beachtet werden, dass viele Modelle aus dem englischsprachigen Raum kommen und damit nur beispielhaft für den deutschsprachigen Raum verwendet werden sollten. Auf jeden Fall sind die Modelle dazu dienlich, den Prozess bildhafter darzustellen.

#### 5.3.1 Entwicklungsmodell nach Frith und Weiterentwicklung

Das Modell von Frith (1986) wurde entwickelt, als es zum damaligen Zeitpunkt kaum Modelle gab, die sich mit dem regulären Schriftspracherwerb beschäftigten. Es besteht aus drei Phasen jeweils für den Erwerb des Lesens und des Schreibens (Frith, 1986, S. 72). Lesen und Schreiben beeinflussen sich dabei gegenseitig. Eine Weiterentwicklung beim Erwerb des Schreibens hilft dem Kind demnach dabei, auch beim Lesen die nächste Stufe zu erreichen (Frith, 1986, S. 76–78). Friths Modell besteht aus den Phasen „logografisches Lesen/Schreiben“, „alphabetisches Lesen/Schreiben“ und „orthografisches Lesen/Schreiben“. An dieser Stelle wird nur der Bereich des Leseerwerbs näher betrachtet.

Auf der ersten Stufe, dem logografischen Lesen, kommt es noch nicht zu dem Lesen, wie geübte Leserinnen und Leser es kennen. Stattdessen orientieren sich Kinder an bestimmten Symbolen, wie z. B. am Logo einer bekannten Firma (Frith, 1986, S. 72). Einzelne Buchstaben werden hier vielleicht (wieder-)erkannt, aber nicht als solche und auch nicht im Zusammenhang mit weiteren Buchstaben „gelesen“, sondern sie dienen lediglich der Orientierung (so kann das <x> am Ende des Wortes Max als Orientierung für den eigenen Namen dienen, wobei auch Fax oder ein Fantasiewort wie lax als Max „gelesen“ werden). Dennoch hat das Kind schon eine bestimmte Leistung erbracht: Es hat erkannt, dass Symbole eine Funktion haben und Bedeutung vermitteln können. Die Verbindung zwischen Symbol und bestimmten Lauten wird hier jedoch noch nicht hergestellt.

An dieser Stelle soll direkt schon auf die Kritik bzw. die Skepsis in Bezug auf die Existenz der logographischen Phase im Anfangsunterricht in deutschsprachigen Ländern hingewiesen werden (Klicpera et al., 2017, S. 28; Mayer, 2016, S. 40). Diese Phase wird mittlerweile nicht mehr unbedingt als Teil der regulären Entwicklung, sondern eher als Besonderheit des Anfangsunterrichts im englischsprachigen Raum angesehen (Mayer, 2016, S. 40). Frith war sich allerdings über mögliche Einschränkungen ihres Modells in anderen Sprachen als der englischen bewusst (Frith, 1986, S. 71). Wie in Kapitel 5.1 bereits kurz erwähnt, zeigt die englische Sprache eine Orthographie, die weniger Regelmäßigkeiten als die deutsche Sprache aufweist (Schründer-Lenzen, 2013, S. 70). Daraus ergeben sich auch weniger einheitliche Verknüpfungen von Graphemen zu Phonemen, sodass hier im Anfangsunterricht vermehrt wichtige Wörter dadurch vermittelt werden, indem auf visuelle Besonderheiten hingewiesen wird. Für den deutschsprachigen Raum lässt sich diese Phase nur im Vorschulalter finden (Mayer, 2016, S. 40–41).

Die Verbindung zwischen Symbolen (vor allem von Buchstaben) und Lauten geschieht erst auf der nächsten Stufe, der des alphabetischen Lesens (Frith, 1986, S. 72). Das Kind hat gelernt, dass Buchstaben und Buchstabenfolgen bestimmte Laute wiedergeben. Die Einsicht in das phonetisch-phonologische Prinzip wird hier erworben (Schründer-Lenzen, 2013, S. 69). Kinder lesen in dieser Phase zunächst noch Buchstabe für Buchstabe. Die Synthese einzelner Buchstaben zu einem Laut geschieht noch recht langsam. Dennoch können in dieser Phase nun auch unbekannte lautgetreue Wörter erlesen werden (Schründer-Lenzen, 2013, S. 69).

Wenn das Kind die dritte und in diesem Modell letzte Stufe, die des orthographischen Lesens, erreicht, orientiert es sich nicht mehr an einzelnen Buchstaben, sondern an Kombinationen von Buchstaben wie beispielsweise Silben und Morphemen, den kleinsten bedeutungstragenden Einheiten eines Wortes (Frith, 1986, S. 72–73). Hier kommt es nun auch zu einem einfachen Erlesen bei bekannten Buchstabenkombinationen und zum Aufbau des sogenannten Sichtwortschatzes. Das Kind hat ein inneres Lexikon aufgebaut, auf welches es beim Lesen direkt zugreifen kann, ohne ein Wort Buchstabe für Buchstabe erlesen zu müssen (Schründer-Lenzen, 2013, S. 72). Erst wenn Kinder auf dieser Stufe angekommen sind, können sie eine ausreichende Lesegeschwindigkeit erreichen, die dann in der Folge das sinnentnehmende Lesen ermöglicht, da weniger kognitive Ressourcen für das Erlesen einzelner Buchstaben oder Wörter benötigt werden und diese somit zum Verstehen des gelesenen Textes verwendet werden können.

In Anlehnung an das Modell von Frith entstanden mehrere weitere Modelle. Günther (1986) übernahm das Modell von Frith für den deutschsprachigen Raum und erweiterte dieses um zwei Stufen. Es setzt früher an als das ursprüngliche Modell und beginnt mit der präliteral-symbolischen Phase, welche bei Frith lediglich angedeutet wird (Günther, 1986, S. 33). Die oben genannte Einsicht, dass Symbole Bedeutung vermitteln können, ist ausschlaggebend für diese Phase. Erst durch diese Erkenntnis sind Kinder dafür bereit, Lesen und Schreiben

erlernen zu können (Günther, 1986, S. 34–35). Im Anschluss an die drei Stufen aus Friths Modell, welche Günther übernommen hat, folgt in diesem Modell die integrativ-automatisierte Stufe. Einzelne Teilprozesse werden hier integriert. Diese zusätzliche Phase soll vor allem verdeutlichen, dass eine Automatisierung des Lesens und Schreibens nicht sofort erfolgt, sondern ein länger andauernder Prozess ist (Günther, 1986, S. 43; Mayer, 2016, S. 43). Günther weist auch noch einmal explizit darauf hin, dass die Stufen nicht einheitlich und strukturiert hintereinander ablaufen, sondern sich auch gegenseitig überlappen oder sich verschieben können (Günther, 1986, S. 43).

### 5.3.2 Kompetenzentwicklungsmodell des Lesens

An den oben genannten Modellen von Frith und Günther wird kritisiert, dass diese ihren Ursprung im englischsprachigen Raum haben und sich die Annahmen daher nicht direkt auf die deutsche Orthografie übertragen lassen. Vor allem die logographische Phase steht in der Kritik, weil diese im deutschsprachigen Raum beim Leseerwerb in der Grundschule kaum zu beobachten ist (Klicpera et al., 2017, S. 28; 31). Zudem geben diese Modelle eine vorgegebene Abfolge der einzelnen Stufen vor. Ein anderes Modell wurde aus den Ergebnissen deutschsprachiger Forschung abgeleitet. Das Kompetenzentwicklungsmodell des Lesens konzentriert sich auf den Beginn des Leseerwerbs. Auch in diesem Modell finden sich verschiedene Entwicklungsphasen, welche aber nicht hintereinander, sondern parallel ablaufen. Es legt den Fokus stärker auf wesentliche Lesekompetenzen, die im Verlauf des Leseerwerbs erlangt werden müssen (Klicpera et al., 2017, S. 28; 30). Weiterhin berücksichtigt das Modell auch Einflussfaktoren, beispielsweise die Instruktion im Unterricht oder Vorläuferfertigkeiten wie die phonologische Bewusstheit oder die Benennungsgeschwindigkeit. Das Ausmaß des Einflusses kann sich über den Verlauf der Zeit ändern und auch abhängig von der Art des Erstleseunterrichts unterschiedlich hoch ausfallen (Klicpera et al., 2017, S. 30–31).

Das Modell setzt mit dem Erstleseunterricht an, da Kinder im deutschsprachigen Raum vor Schuleintritt im Allgemeinen über wenig Kenntnisse in der Schriftsprache verfügen (Klicpera et al., 2017, S. 28). Das Modell beschreibt, analog zu den zuvor vorgestellten Prozessmodellen, zwei Arten des Zugriffs: die phonologische Rekodierung und das mentale Lexikon. Beide Kompetenzen müssen erworben werden, um gut lesen zu können. Bei erfahrenen Leserinnen und Lesern geschieht dies generell über das mentale Lexikon, aber auch hier muss stellenweise auf die phonologische Rekodierung zurückgegriffen werden, wenn unbekannte Wörter erlesen werden sollen (Klicpera et al., 2017, S. 29). Ein wichtiger Unterschied im Vergleich zu anderen Entwicklungsmodellen ist, dass sich beide Zugriffswege parallel entwickeln (Klicpera et al., 2017, S. 30). Die phonologische Rekodierung muss allerdings zunächst grundlegend erworben werden, um die ersten Wörter im mentalen Lexikon abspeichern zu können. Ab diesem Zeitpunkt erfolgt dann aber eine parallele Entwicklung beider Zugriffswege (Klicpera et al., 2017, S. 30).

Die phonologische Rekodierung erfolgt anfänglich durch das Erlesen von kleinen Einheiten, also einzelnen Buchstaben und Graphemen. Im weiteren Verlauf des Leseerwerbs werden die erlesenen Einheiten zunehmend größer: Es werden Silben oder auch Morpheme erlesen (Klicpera et al., 2017, S. 30). Wörter werden im mentalen Lexikon über deren Eigenschaften, also über das Schriftbild, die Aussprache und die Bedeutung, abgespeichert (Klicpera et al., 2017, S. 29). Das Lesen erfolgt zunehmend automatisiert. Auch ist der lexikalische Zugriff schneller als die phonologische Rekodierung, sodass sich mit dem Ausbau des mentalen Lexikons und der Automatisierung der phonologischen Rekodierung die Lesegeschwindigkeit zunehmend erhöht und sich die Fehler beim Lesen verringern (Klicpera et al., 2017, S. 30). Klicpera et al. (2017, S. 31) gehen auch noch einmal auf Unterschiede in den Sprachen ein: Sprachen mit einer relativ hohen Graphem-Phonem-Korrespondenz, wie es in der deutschen, aber auch der griechischen oder italienischen Sprache der Fall ist, ermöglichen den Kindern ein schnelleres Erlernen des Lesens – sowohl von regulären wie auch von Pseudowörtern. Das Erlesen einzelner Grapheme führt im Deutschen beispielsweise fast immer zur korrekten Aussprache, während dies in der englischen Sprache nicht der Fall ist. Hier müssen größere Einheiten erlesen werden, um zur korrekten Aussprache des Wortes zu gelangen. Das Modell berücksichtigt weiterhin Kompetenzen, welche für das Lesen benötigt werden, die sich aber erst im Verlauf des Unterrichts entwickeln (Klicpera et al., 2017, S. 31). Daher sind diese Faktoren als Folge und weniger als Prädiktor für den erfolgreichen Leseerwerb zu sehen. Ein hier genanntes Beispiel sind „phonologische Kompetenzen auf Phonemebene“ und die „Sensitivität für Silben“, welche sich ab der zweiten bis zur vierten Klasse entwickeln (Klicpera et al., 2017, S. 32). Diese Kompetenzen erleichtern den Leseerwerb, was aber umgekehrt nicht bedeutet, dass der Leseerwerb bei einem Mangel automatisch misslingt. Gute Leserinnen und Leser entwickeln diese Kompetenzen allerdings früher als schwache (Klicpera et al., 2017, S. 32).

## 5.4 Leseunterricht in der Grundschule

Bisher wurde deutlich, dass der Leseerwerb ein langwieriger Prozess ist, welcher zum einen nicht erst mit Eintritt in die Grundschule beginnt, aber mit Übergang in die Sekundarstufe auch noch nicht beendet ist (Richter & Müller, in Druck, S. 15). In diesem Abschnitt wird sich, entsprechend der dieser Untersuchung vorliegenden Fragestellung, auf den Leseunterricht in der Grundschule bezogen. Dieser wandelte sich in den letzten Jahrzehnten zum Teil sehr stark.

### 5.4.1 Anfangsunterricht im Lesen

Schründer-Lenzen (2013) stellt ausführlich die Entwicklung des Anfangsunterrichts im Lesen seit Valentin Ickelsamer zu Beginn des 16. Jahrhunderts dar, welcher eine Anlautmethode zum Erlernen des Lesens vorschlug (Schründer-Lenzen, 2013, S. 150). Zu dieser Zeit bestand das Lesen im Anfangsunterricht darin, „Buchstabennamen und ihre Schreibweise

mechanisch einzuprägen“ (Schründer-Lenzen, 2013, S. 148). Dies wird auch als „Buchstabiermethode“ bezeichnet, welche im 19. Jahrhundert in Preußen schließlich verboten wurde (Schründer-Lenzen, 2013, S. 150).

Im 17. Jahrhundert, also ca. hundert Jahre nach Ickelsamer, wurde auch von Comenius ein lautsynthetisches Verfahren vorgeschlagen (Schründer-Lenzen, 2013, S. 149). Die Lautiermethoden unterscheiden sich darin, ob mit Anlauten „als eigenem Sinnträger“ gearbeitet wird oder mit der Nachahmung von Geräuschen, wie beispielsweise das Gackern einer Gans („ga ga“) für den Buchstaben /g/ (Schründer-Lenzen, 2013, S. 149–150). Die Anlautmethode, welche Bestandteil der synthetischen Lesemethoden ist, ist auch heute noch relevant. Bei den synthetischen Lesemethoden geht man von den einzelnen Lauten aus. Ein Kind erliest entsprechend schrittweise ein Wort und muss die einzelnen Laute nach und nach zusammenfügen. Einzelne Grapheme müssen hierfür in Phoneme übertragen und auf diese Weise nacheinander zum gesamten Wort zusammengesetzt werden. Hier zeigt sich eine große Schwierigkeit, da durch ein schnelles Sprechen der Laute nicht automatisch das korrekte Wort gebildet wird (Mahling, 2015, S. 550; Schründer-Lenzen, 2013, S. 151–152). Zudem wurde sich auf die Verbindung der Buchstaben zu Silben und zu Wörtern konzentriert, die Bedeutung der Wörter war weniger relevant (Klicpera et al., 2017, S. 91).

Demgegenüber steht die analytische Methode, die mit ganzen Wörtern (Ganzwortmethode) oder auch ganzen Sätzen (Ganzsatzmethode) beginnt (Schründer-Lenzen, 2013, S. 152). Die Kinder sollen sich hier komplette Wörter oder teilweise auch vollständige (kurze) Sätze einprägen. Die einzelnen Buchstaben der Wörter werden dagegen erst später eingeführt und gelernt (Mahling, 2015, S. 550; Schründer-Lenzen, 2013, S. 151–152).

In den 1950er und 1960er Jahren sorgten diese beiden konträren Vorgehensweisen für eine kontroverse Diskussion, da man sich uneinig war, welche Methode besser geeignet war, den Kindern das Lesen beizubringen (Schründer-Lenzen, 2013, S. 155). Es zeigte sich, dass die synthetische Methode in den ersten Schuljahren zu besseren Ergebnissen führt, zum Ende der Grundschulzeit beide Methoden aber gleichauf sind. Gleichzeitig haben schwächere Kinder mit der analytischen Methode stärkere Schwierigkeiten als mit der synthetischen Methode (Klicpera et al., 2017, S. 90–91; Schabmann, Schmidt, Klicpera, Gasteiger-Klicpera & Klingebiel, 2009, S. 332–333; Schründer-Lenzen, 2013, S. 155). Schründer-Lenzen (2013, S. 156) führt an, dass sich schließlich eine Kombination der beiden Methoden entwickelte. (Klicpera et al., 2017, S. 92) bezeichnen die heutige Vorgehensweise als analytisch-synthetische Methode. Aus dem Namen wird direkt auch die Verbindung der beiden Methoden deutlich. Sie berichten weiterhin, dass die aktuellen synthetischen bzw. analytischen Vorgehensweisen viel geringere Gegensätze zueinander aufweisen als in den 1950er und 1960er Jahren, aber dass sich auch hier Unterschiede in der Wirkungsweise finden lassen (Klicpera et al., 2017, S. 92). Nach drei Monaten zeigt sich, dass Kinder, die mit der synthetischen Vorgehensweise unterrichtet wurden, den Kindern, die mit der analytischen Methode unterrichtet wurden, im Lesen voraus sind. Sie konnten unbekannte Wörter besser lesen und auch die

Buchstabenkenntnis war hier besser. Allerdings verschwanden diese Unterschiede zum Ende der zweiten Klasse und waren auch in höheren Klassenstufen nicht mehr zu finden. Es sind vor allem schwächere Schülerinnen und Schüler, die vom synthetischen Ansatz stärker profitieren (Klicpera et al., 2017, S. 93). Die synthetische Methode scheint daher in den ersten Phasen der ganzheitlichen überlegen zu sein und vor allem bei schwächeren Kindern für bessere Lernergebnisse zu sorgen. Insgesamt geben Klicpera et al. (2017, S. 94) diesem Ansatz daher den Vorzug, wobei sie auch hinzufügen, dass die Reinformen im deutschsprachigen Raum kaum zu finden sind und meist Mischformen angewendet werden. Sie merken zudem an, dass die Unterschiede zwischen einzelnen Lehrkräften höher sind als die Unterschiede der Unterrichtsformen und schlussendlich die Unterrichtsqualität als wichtiger zu erachten ist. In diesem Zuge nennen sie drei Faktoren, welche ausschlaggebend sind für einen guten Leseunterricht: die Zeit, welche die Kinder zum Lesen aufwenden, die Strukturiertheit des Vorgehens der Lehrkraft sowie die Fähigkeit der Lehrkraft, auf individuelle Lernprozesse einzugehen und diese zu begleiten (Klicpera et al., 2017, S. 94). Marx (2007, S. 94) hält resümierend fest, dass unabhängig von der Lehrmethode (synthetisch oder analytisch) fast jedes Kind die Schriftsprache erlernt und dabei „seinen eigenen Weg“ finden muss, dass es aber auch immer Kinder gibt, die Schwierigkeiten haben. Dazu kommt, dass die unterschiedlichen Herangehensweisen jeweils unterschiedliche Fertigkeiten im Anfangsunterricht auf Seiten der Kinder verlangen und die gewählte Methode je nach Vorerfahrung unterschiedlich effektiv ist. Er kommt allerdings zu dem Schluss, dass empirische Untersuchungen notwendig sind, um hierüber abschließend urteilen zu können (Marx, 2007, S. 94).

Schilcher (2012, S. 19–20) kritisiert, dass in der Ausbildung der Lehrkräfte zwischen 1980 und bis zum Ende der neunziger Jahre literatur- oder lesedidaktische Inhalte für den Deutschunterricht nicht im Fokus standen und Lehrkräfte Methoden anwendeten und ausprobierten, zu denen es keine theoretischen Grundlagen oder empirischen Belege hinsichtlich ihrer Effektivität gab, und sich dies erst durch die Ergebnisse der PISA-Studie 2000 änderte. Auch Mahling (2015, S. 563) hält kritisch fest, dass vor allem Längsschnittstudien fehlen, die Methoden auf ihre Wirksamkeit hin untersuchen. Klicpera et al. (2017, S. 97–98) merken in diesem Zuge auch an, dass über die Frage nach der Geschwindigkeit, mit der im Anfangsunterricht vorgegangen wird, wie beispielsweise die Einführung aller Buchstaben, bisher kaum empirische Forschung stattgefunden hat. Sie berichten von Ergebnissen ihrer Längsschnittuntersuchung, wonach die Klassen, die schneller alle Buchstaben kannten, über das Ende der ersten Klasse hinaus bessere Leistungen erzielten und dass die Lehrkräfte, die mit einer höheren Geschwindigkeit vorgegangen waren, im zweiten Schuljahr mehr Zeit für das Üben von Lesen und Schreiben aufwenden konnten (Klicpera et al., 2017, S. 98).

#### 5.4.2 Leseerwerb im weiteren Verlauf der Grundschulzeit

Klicpera et al. (2017, S. 98) sehen das zweite Schuljahr als eine „Übergangsklasse“. Kinder, die noch Schwierigkeiten mit der Phonem-Graphem-Zuordnung haben, können hier noch

aufholen. Gleichzeitig geht es ab der dritten Klasse nicht mehr darum, das Lesen zu lernen, sondern das Lesen zum Lernen zu nutzen. Daher sollte der Fokus in der zweiten Klasse im Lesen üben liegen und somit auch viel Zeit mit dem Lesen verbracht werden. Hierfür ist es notwendig, die Kinder zum Lesen zu motivieren und ihnen passende Literatur zur Verfügung zu stellen, welche sie nicht überfordert, die aber an ihren Interessen anknüpft (Klicpera et al., 2017, S. 98). Sie sehen zudem ein wiederholtes Lesen der gleichen Texte als Möglichkeit, die Lesegeschwindigkeit zu erhöhen und beziehen sich hier auf eine Metaanalyse von Therrien (Klicpera et al., 2017, S. 99). Insgesamt wurden in diese Metaanalyse 18 experimentelle quantitative Studien mit Schülerinnen und Schülern im Alter von 5 bis 18 Jahren eingeschlossen (Therrien, 2004, S. 253). Sowohl bei Schülerinnen und Schülern ohne Schwierigkeiten wie auch bei denen, die Lernschwierigkeiten hatten, führte das wiederholte Lesen zu einer Steigerung der Leseflüssigkeit und des Leseverständnisses, welche sich auch auf unbekannte Texte übertrug (Therrien, 2004, S. 257). Dieses Verfahren kann daher im regulären Unterricht eingesetzt werden, um die Lesefähigkeit aller Schülerinnen und Schüler zu steigern.

Laut der IGLU-Studie 2016 verbringen deutsche Grundschul Kinder der vierten Klasse im Durchschnitt 136 Minuten pro Woche (mit einer Streuung von 126 Minuten) mit Leseunterricht oder anderen Leseaktivitäten, was auf das Jahr umgerechnet ungefähr 90 Stunden sind (Bremerich-Vos et al., 2017, S. 280). Der international ermittelte Durchschnitt liegt allerdings um einiges höher bei fast 160 Stunden. Für deutsche Grundschul Kinder ergibt sich, unter Berücksichtigung der Unterrichtszeit, ein Anteil von 11 Prozent für Leseunterricht, was unter dem internationalen Mittelwert von 18 Prozent liegt. Bremerich-Vos et al. (2017, S. 281) resümieren, dass der Leseunterricht in Deutschland vergleichsweise sehr gering ausfällt. Sie führen auch an, dass in der vierten Klasse das Leseverständnis im Fokus der Vermittlung steht, es aber Schülerinnen und Schüler gibt, die noch Schwierigkeiten mit den hierarchieniedrigen Fähigkeiten wie der Leseflüssigkeit haben und hier entsprechend gezielt gefördert werden müssen (Bremerich-Vos et al., 2017, S. 279). Die explizite Vermittlung und das Einüben von Strategien, um Texte sinnentnehmend lesen zu können, wird allerdings erst seit relativ kurzer Zeit durchgeführt (Klicpera et al., 2017, S. 105). Richter und Müller (in Druck, S. 17) warnen, dass sich Lesestrategietrainings sogar negativ auf das Leseverständnis auswirken können, wenn die Worterkennung noch nicht ausreichend erworben wurde. Somit stellt diese Fähigkeit die erforderliche Basis des Leseerwerbs dar und es zeigt sich erneut die Notwendigkeit, dass Lehrkräfte den Lernstand ihrer Schülerinnen und Schüler korrekt einschätzen können.

#### 5.4.3 Lesekompetenz zum Ende der Grundschulzeit

Im Jahr 2004 haben die Bundesländer sich auf Bildungsstandards für die vierte Klasse in den Fächern Deutsch und Mathematik geeinigt. Aufgrund der föderalen Struktur des Bildungswesens in der Bundesrepublik Deutschland obliegt es den einzelnen Bundesländern, diese

Vorgaben entsprechend in ihren Lehrplänen umzusetzen (KMK, 2005, S. 3; Schröder-Lenzen, 2013, S. 236). Speziell für den Bereich Lesen sehen die Bildungsstandards vor, dass Schülerinnen und Schüler lernen, mit Texten und Medien umgehen zu können. Dies bedeutet, dass sie über Lesefähigkeiten, aber auch über Leseerfahrungen verfügen sollen, um sich Texte erschließen sowie diese präsentieren zu können (KMK, 2005, S. 7). Die Grundschule soll „zum genießenden, informierenden, selektiven, interpretierenden und kritischen Lesen“ hinführen (KMK, 2005, S. 9). Damit soll „eine tragfähige Grundlage für weiteres Lernen, für weiteres selbstbestimmtes Lesen und eine bewusste Auswahl geeigneter Medien“ gelegt werden (KMK, 2005, S. 9). Es sollen sowohl Sach- wie auch literarische Texte berücksichtigt werden, wobei das Ziel beim Lesen literarischer Texte sein soll, sich „identifizierend und abgrenzend mit literarischen Figuren [auseinanderzusetzen]“ (KMK, 2005, S. 9). Ebenso wird berücksichtigt, dass Kinder sich in der „Medienwelt orientieren“ können sollen (KMK, 2005, S. 9).

Das Lesetempo als Zeichen für die Leseflüssigkeit sollte gegen Ende der Grundschulzeit etwa der gesprochenen Sprache entsprechen, was mehr als 220 Wörtern pro Minute bei leisem Lesen und mehr als 200 Wörtern pro Minute bei lautem Vorlesen entspricht (Garbe et al., 2009, S. 147). Studien zeigen, dass zum Ende der vierten Klasse die überwiegende Mehrzahl der Kinder die hierarchieniedrigen Leseprozesse beherrscht. Hierzu zählen das Erkennen häufiger Wörter sowie die phonologische Rekodierung von Wörtern beim Lesen (Richter & Müller, in Druck, S. 14–15). Wenn dies nicht der Fall ist, kommt es noch zu vielen Fehlern in der Worterkennung, was das sinnentnehmende Lesen stark beeinträchtigt. Richter und Müller (in Druck, S. 15) fügen allerdings hinzu, dass auch die Entwicklung des Sichtwortschatzes und der Worterkennungsprozesse zum Ende der vierten Klasse noch nicht abgeschlossen sind. Die IGLU-Studie 2016 kommt zu dem Schluss, dass ca. 19 Prozent der Viertklässlerinnen und Viertklässler Kompetenzstufe III nicht erreicht haben, was bedeutet, dass hier mit „erheblichen Schwierigkeiten beim Lernen in allen Fächern“ gerechnet werden muss (Bos et al., 2017, S. 15).

## 5.5 Zwischenfazit

Es erscheint relevant, noch einmal deutlich hervorzuheben, dass die einzelnen Stufen oder Phasen in den Entwicklungsmodellen nur als grobe Orientierung dienen. Nicht alle Stufen werden von jedem Kind durchschritten. Dennoch dienen diese Modelle als gute Orientierung für die Lehrkraft, an welcher Stelle des Leselernprozesses sich eine Schülerin oder ein Schüler ungefähr befindet. Klicpera et al. (2017, S. 59–60) kommen zu dem Schluss, dass sich gerade die Entwicklungsmodelle auf den ungestörten Schriftspracherwerb konzentrieren und zudem im deutschsprachigen Raum vermutlich auch nur eingeschränkte Gültigkeit haben. Diese Modelle zeigen auf, dass sich zu Beginn des Leseerwerbs auf bestimmte Merkmale in Wörtern bzw. Buchstaben konzentriert wird, sich dies aber in späteren Phasen abbauen soll, sodass Kinder auf die Graphem-Phonem-Verbindung zurückgreifen und zudem einen

Sichtwortschatz aufgebaut haben. Wenn dies nicht geschieht, ist das ein guter Hinweis auf Schwierigkeiten (Klicpera et al., 2017, S. 59–60). Ebenso helfen Prozessmodelle dabei, einen Einblick in die kognitiven Abläufe zu erhalten und hierüber Aufschluss darüber zu bekommen, mit welchen Prozessen Kinder vielleicht Probleme haben und wo eine Lehrkraft ansetzen kann, um das Kind entsprechend zu fördern.

Auch Mayer (2016, S. 43) hält fest, dass Modelle zum Erwerb des Lesens helfen können, Schwierigkeiten zu erkennen, und in der Folge gegebenenfalls zusätzliche Unterstützungsangebote vom Übergang aus der alphabetischen in die orthographische Phase gegeben werden sollten. Er berücksichtigt dabei auch das konnektionistische Modell, welches Assoziationen zwischen orthographischen und phonologischen Einheiten als relevant für den Leseerwerb in den Fokus rückt. Als drittes führt er noch das „Simple View of Reading“-Modell an, welches nicht nur die schriftsprachlichen Kompetenzen, sondern auch „semantische, grammatische und pragmatische Fähigkeiten“ als wesentlich erachtet (Mayer, 2016, S. 44). Er kommt zu dem Schluss, dass eine einzelfallorientierte Diagnostik durchgeführt werden muss, um festzustellen, welche Förderung und Unterstützung ein Kind benötigt (Mayer, 2016, S. 44).

## 6 Schwierigkeiten beim Leseerwerb

In Kapitel 5 wurden Modelle des ungestörten Leseerwerbs vorgestellt. In Studien ist allerdings deutlich geworden, dass nicht alle Kinder ohne Schwierigkeiten das Lesen erlernen (Fischer & Gasteiger-Klicpera, 2013, S. 63; Klicpera et al., 2017, S. 152). Diese Schwierigkeiten können sich in einigen Fällen auch verfestigen und es kann zur Entwicklung von Leseschwächen bis hin zu einer Lesestörung kommen (Fischer & Gasteiger-Klicpera, 2013, S. 63; Klicpera et al., 2017, S. 156). Es gibt zahlreiche Begriffe, die für diese Schwierigkeiten im Leseerwerb verwendet werden wie beispielsweise Lese-Rechtschreibstörung, Legasthenie, Lesestörung oder auch Leseschwäche. Ausgehend von der verwendeten Begrifflichkeit unterscheidet sich die Gruppe der „schwachen Leserinnen und Leser“ zumindest in Details (Valtin et al., 2010, S. 43).

In diesem Kapitel soll nun zunächst auf die Begriffe in Bezug auf Schwierigkeiten im Leseerwerb eingegangen werden, um anschließend die Prävalenzen zu betrachten, welche je nach Definition stark schwanken können. Zum Schluss soll auch auf mögliche Folgen und Komorbiditäten eines gestörten Leseerwerbs eingegangen werden. Den Ursachen wird in einem gesonderten Kapitel nachgegangen, da diese Grundlage für präventive und unterstützende Maßnahmen der Lehrkraft darstellen und ihnen daher eine besondere Bedeutung in der Grundschule zukommt. Wichtig ist an dieser Stelle bereits hervorzuheben, dass die Ursachen keineswegs nur in den betroffenen Kindern selbst zu verorten sind, sondern dass beispielsweise auch Umweltfaktoren eine wichtige Rolle spielen können (McElvany, Becker & Lütke, 2009, S. 130; McElvany & Ohle, 2015, S. 47; Tiedemann & Billmann-Mahecha, 2007, S. 46).

### 6.1 Lese-Rechtschreibstörung nach ICD-10 und DSM-5

Auch wenn es sich um kein medizinisches, sondern um ein vorrangig pädagogisches und schulisches Problem handelt und die Klassifikationen nach ICD-10 und DSM-5 daher nur eingeschränkte Relevanz haben, soll dennoch zu Beginn hierauf Bezug genommen werden, da diese beiden Systeme sehr häufig als Grundlage von Definitionen verwendet werden (Klicpera et al., 2017, S. 135). Steinbrink und Lachmann (2014, S. 3) merken an, dass es sich hier um die einzigen „relativ klar formulierten“, zugleich aber nicht unumstrittenen diagnostischen Kriterien handelt, die international anerkannt sind.

Die internationale statistische Klassifikation der Krankheiten ICD-10 (Deutsches Institut für Medizinische Dokumentation und Information, 2018) führt die Lese-Rechtschreibstörung (LRS) im Kapitel *F80-89 Entwicklungsstörungen* an. Als Oberbezeichnung finden sich hier unter *F81 Umschriebene Entwicklungsstörungen schulischer Fertigkeiten*. Eine Intelligenzminderung oder erworbene Erkrankungen wie eine Hirnschädigung gelten als Ausschluss für diese Entwicklungsstörungen (Deutsches Institut für Medizinische Dokumentation und Information, 2018, S. 325). Weiter teilt die ICD-10 die Entwicklungsstörungen ein in Lese- und Rechtschreibstörung, isolierte Rechtschreibstörung, Rechenstörung, kombinierte Störungen schulischer Fertigkeiten sowie sonstige und nicht näher bezeichnete Entwicklungsstörungen

schulischer Fertigkeiten. Die isolierte Rechtschreibstörung schließt hierbei explizit eine Lesestörung – ob vorher bestehend oder gleichzeitig – aus.

Die Lese- und Rechtschreibstörung wird wie folgt definiert:

#### F81.0 Lese- und Rechtschreibstörung

Das Hauptmerkmal ist eine umschriebene und bedeutsame Beeinträchtigung in der Entwicklung der Lesefertigkeiten, die nicht allein durch das Entwicklungsalter, Visusprobleme oder unangemessene Beschulung erklärbar ist. Das Leseverständnis, die Fähigkeit, gelesene Worte wieder zu erkennen, vorzulesen und Leistungen, für welche Lesefähigkeit nötig ist, können sämtlich betroffen sein. Bei umschriebenen Lesestörungen sind Rechtschreibstörungen häufig und persistieren oft bis in die Adoleszenz, auch wenn einige Fortschritte im Lesen gemacht werden. Umschriebenen Entwicklungsstörungen des Lesens gehen Entwicklungsstörungen des Sprechens oder der Sprache voraus. Während der Schulzeit sind begleitende Störungen im emotionalen und Verhaltensbereich häufig.

(Deutsches Institut für Medizinische Dokumentation und Information, 2018, S. 325)

Explizit ausgeschlossen werden körperliche Ursachen wie ein vermindertes Sehvermögen oder externe Faktoren wie eine unangemessene Beschulung. Eine isolierte Lesestörung existiert in der ICD-10 nicht. Es wird in dieser Definition lediglich darauf verwiesen, dass Lesestörungen „häufig“ gemeinsam mit Rechtschreibstörungen auftreten.

Das DSM-5 (Diagnostic and Statistical Manual of Mental Disorders) führt unter dem Kapitel *Spezifische Lernstörung* Störungen in der Lese- und Rechtschreibentwicklung auf (Falkai & American Psychiatric Association, 2015). Speziell genannt werden die Beeinträchtigung beim Lesen, die Beeinträchtigung beim schriftlichen Ausdruck und die Beeinträchtigung beim Rechnen. Die Beeinträchtigung beim Lesen wird weiterhin unterteilt in Beeinträchtigungen im Bereich Lesegenauigkeit, Lesegeschwindigkeit oder -flüssigkeit sowie Leseverständnis (Falkai & American Psychiatric Association, 2015, S. 47). Lese- und Rechtschreibstörungen sind hier getrennt aufgeführt. Dies bedeutet, dass nach dem DSM-5 eine isolierte Lesestörung auftreten kann. Ebenso können Beeinträchtigungen im Lesen und Schreiben aber auch in Kombination vorkommen.

Eine Uneinigkeit in den diagnostischen Kriterien nach ICD-10 und DSM-5 herrscht also in der Frage, ob eine Lesestörung auch isoliert, also unabhängig von einer Rechtschreibstörung, auftreten kann. Die Kriterien nach ICD-10 sehen dies nicht vor, während das DSM-5 die isolierte Lesestörung explizit vorgibt. Die Existenz einer isolierten Lesestörung wird klinisch jedoch nicht angezweifelt (Schulte-Körne & Remschmidt, 2003, S. A 396).

Im Vergleich zur Vorgängerversion wie auch zur ICD-10 ist das IQ-Diskrepanzkriterium im DSM-5 aufgegeben worden. Der Begriff der Dyslexie wurde im Vergleich zum DSM-IV

eingeführt und kann als alternative Bezeichnung verwendet werden, wenn Schwierigkeiten im basalen Lesen, also bei der Worterkennung und der Graphem-Phonem-Zuordnung, und in der Rechtschreibung auftreten (Schulte-Körne, 2014, S. 370). Klicpera et al. (2017, S. 136–137) betrachten ebenfalls die Änderungen in der Definition vom DSM-IV zum DSM-5 und fassen zusammen, dass das Leseverständnis explizit aufgenommen wurde und die Forderung nach der Diskrepanz zu den intellektuellen Fähigkeiten abgeschwächt wurde, während die Diskrepanz zur Altersgruppe im standardisierten Leistungstest mindestens 1.5 Standardabweichungen betragen soll. Es kann davon aber abgewichen werden und es können stattdessen Diskrepanzen von 1 bis 2.5 Standardabweichungen als Kriterium angelegt werden (Schulte-Körne, 2014, S. 370).

Wie oben bereits angemerkt, sehen Klicpera et al. (2017, S. 135) in den Definitionen der ICD-10 und DSM-5 nur begrenzte Relevanz und verstehen diese daher nur als „einen Vorschlag unter mehreren“. Sie kritisieren die Festlegung eines Diskrepanzkriteriums von zwei Standardabweichungen im Leistungstest zur durchschnittlichen Leistung der Vergleichsgruppe als „zu einschränkend“ (Klicpera et al., 2017, S. 136), da bereits bei einer Diskrepanz von einer Standardabweichung eine Beeinträchtigung vorliegen kann und halten diese daher für ausreichend, um eine Lese- oder Rechtschreibstörung zu diagnostizieren. Für die Diagnose von Schwierigkeiten im Leseverständnis fordern sie eine Diskrepanz von einer Standardabweichung zwischen der Leistung in der Lesegeschwindigkeit auf der einen Seite und dem Leseverständnis auf der anderen Seite (Klicpera et al., 2017, S. 137).

## 6.2 Weitere Begriffe und Definitionen

Es existieren im deutschsprachigen Raum zahlreiche Begriffe, die teilweise synonym verwendet werden. Nach Steinbrink und Lachmann (2014, S. 4–5) gibt es in der Bundesrepublik Deutschland die amtlichen Begriffe „besondere Schwierigkeiten im Lesen und/oder Rechtschreiben“, „Legasthenie“, „Lese-Rechtschreibstörung“ und „Lese-Rechtschreibschwäche“, die nicht einheitlich verwendet werden. Weiterhin kritisieren sie, dass hinter dem gleichen Begriff teilweise unterschiedliche Definitionen mit unterschiedlichen Entstehungsursachen und Verläufen stehen. Sie führen an, dass die Lese-Rechtschreibstörung als Begriff im Alltag häufig verwendet wird, wenn die Symptomatik „schwach“ ausgeprägt ist und die Diskrepanz zwischen der Intelligenz und der Leistung im Lesen und Schreiben geringer ausfällt, während der Begriff Legasthenie für eine starke Ausprägung verwendet wird, halten diese Einteilung aber nicht für sinnvoll (Steinbrink & Lachmann, 2014, S. 5; 51). Strehlow und Haffner (2007, S. 396) halten fest, dass eine fehlende einheitliche Sichtweise bzw. Definition der Lese-Rechtschreibschwierigkeiten nachvollziehbar ist, da diese im Grenzbereich verschiedener Forschungsgebiete angesiedelt sind und somit auch aus unterschiedlichen Perspektiven, wie beispielsweise der Psychologie, der Kinder- und Jugendpsychiatrie, aber auch der Linguistik und der Pädagogik betrachtet werden. Einig sind sich die Vertreterinnen und Vertreter der Fachgebiete dabei, dass es sich in allen Fällen um Kinder handelt, die besondere

Schwierigkeiten beim Schriftspracherwerb haben (Strehlow & Haffner, 2007, S. 397). Auch hier steht aber wieder die Frage im Raum, ob die Intelligenz berücksichtigt werden soll.

Steinbrink und Lachmann bezeichnen die Lese-Rechtschreibstörung als „dauerhafte Probleme beim Erlernen des Lesens und Schreibens“, wobei die Betroffenen „diese Fertigkeiten statistisch gesehen nie in dem Grade erwerben wie die breite Masse der Bevölkerung“ (2014, S. 2). Als Ausschlussfaktoren werden auch hier körperliche Beeinträchtigungen wie Sehschwäche genannt, aber auch allgemeine Faktoren des Lernens und Lehrens, wie fehlende Motivation, mangelnde Deutschkenntnisse oder ein längerer Krankenhausaufenthalt des Kindes. Sie grenzen diese Entwicklungsstörung von der erworbenen Störung des Schriftsprachgebrauchs ab, die Erwachsene zum Beispiel nach einer Hirnschädigung erleiden können. Weiterhin führen sie an, dass keine Einigkeit darüber herrscht, ob eine Diskrepanz in der Lese-Rechtschreibleistung zur kognitiven Leistungsfähigkeit vorliegen soll. Sofern diese Diskrepanz gefordert wird, ist es ihrer Meinung nach allerdings fraglich, ob die allgemeine Intelligenz hierfür das geeignete Maß ist (Steinbrink & Lachmann, 2014, S. 3). Dummer-Smoch hält dagegen am IQ-Diskrepanzkriterium fest und definiert eine „Diskrepanz-Legasthenie“ als „die spezifische, mit Diskrepanzen zwischen Intelligenz und Lese- bzw. Rechtschreibleistungen sowie weiteren Diskrepanzen in Schulleistungen verbundene Lese-Rechtschreibschwäche“ (2007, S. 378). Sie gibt an, dass das IQ-Diskrepanzkriterium umstritten ist, sieht aber „als Kern des Erscheinungsbildes einer Legasthenie eine diskrepante Begabungsstruktur“ (Dummer-Smoch, 2007, S. 378). Analog zu Steinbrink und Lachmann (2014) merkt auch sie an, dass der Gesamt-IQ als Kriterium infrage gestellt werden sollte, da dieser Wert je nach verwendetem Intelligenztest und Anteil an sprachlichen und nichtsprachlichen Aufgaben unterschiedlich hoch ausfällt (Dummer-Smoch, 2007, S. 379). Sie kritisiert entsprechend die Reduzierung des „komplexen qualitativen Erscheinungsbildes auf zwei Werte“ (Dummer-Smoch, 2007, S. 379), nämlich auf den Intelligenzquotienten und den Prozentrang im schulischen Leistungstest.

Mayer (2016, S. 45) legt seiner Definition der Lese-Rechtschreibstörung die Beschreibungen von ICD-10 und DSM-5 zugrunde, gibt allerdings auch explizit an, die Begriffe Lese-Rechtschreibstörung, Legasthenie, besondere Schwierigkeiten im Erlernen des Lesens und Rechtschreibens, Lese-Rechtschreibschwierigkeiten, (Entwicklungs-)Dyslexie und Schriftspracherwerbsstörungen synonym zu verwenden. Als Lese-Rechtschreibstörung definiert er eine Lernstörung,

„die sich durch Probleme beim Erwerb und der Anwendung der indirekten Lesestrategie (= phonologisches Rekodieren) und/oder der automatisierten Worterkennung sowie beeinträchtigter Rechtschreibung charakterisieren lässt. Sie kann aus Defiziten in der phonologischen Informationsverarbeitung infolge neurobiologischer Fehlentwicklungen resultieren und geht oft mit Spracherwerbsstörungen einher. Die Lernstörung tritt unabhängig von kognitiven Fähigkeiten auf und ist nicht die Folge unangemessenen

Unterrichts. Sie kann sich negativ auf das Leseverständnis, die kognitive, die sprachliche sowie die sozio-emotionale Entwicklung auswirken.“ (Mayer, 2016, S. 45)

Grundlegendes Charakteristikum sind hier Schwierigkeiten im basalen Lesen sowie beeinträchtigte Rechtschreibfähigkeiten. Schwierigkeiten im Leseverständnis werden als eine mögliche Folge betrachtet. In dieser Definition liegt zudem der Fokus sehr stark auf der phonologischen Informationsverarbeitung. Mayer hebt außerdem hervor, dass der Lese- und Rechtschreiberwerb eine Entwicklungsaufgabe für Kinder darstellt, welche in die „gesamte sprachlich-kognitive Entwicklung des Kindes eingebettet ist“ (Mayer, 2016, S. 45). Da mit Schrift die „Sprache visuell abgebildet wird“ (Mayer, 2016, S. 45), sieht er Kinder mit sprachlichen Beeinträchtigungen als besondere Risikogruppe. Es wird bewusst auf die Bezeichnung „Spezifische Lernstörung“ verzichtet, da betroffene Kinder spätestens ab der dritten Klasse in sämtlichen Unterrichtsfächern, in denen die Schriftsprache Verwendung findet, potenziell Schwierigkeiten bekommen und sich die Störung negativ auf ihre schulischen Leistungen auch in diesen Fächern auswirken kann (Mayer, 2016, S. 46). Mayer verzichtet außerdem analog zum DSM-5 auf das IQ-Diskrepanzkriterium. Er begründet dies damit, dass das Diskrepanzkriterium nur eine „geringe praktische Relevanz“ hat, da sich weder die Schwierigkeiten noch die Förderung der beiden Gruppen grundlegend voneinander unterscheiden (Mayer, 2016, S. 46–47).

Fischbach et al. (2013, S. 66) kritisieren ebenfalls das IQ-Diskrepanzkriterium, da sich keine Unterschiede zwischen Lernschwächen und Lernstörungen in Bezug auf Symptome, zugrundeliegende kognitive Defizite, Erfolge durch lerntherapeutische Interventionen oder dem Leidensdruck der betroffenen Kinder finden. Sie merken auch kritisch an, dass die Unterscheidung in Lernschwäche und Lernstörung im schulischen Setting dazu führt, dass Kinder mit Lernstörungen Hilfen erhalten, wie einen Nachteilsausgleich oder Kostenerstattung für Lerntherapie, während dies bei Kindern mit einer Lernschwäche nicht der Fall ist.

Während sich die meisten Definitionen auf Lese- und Rechtschreibschwierigkeiten konzentrieren, beschreibt Schulte-Körne (2010, S. 719) konkret die auftretende Problematik einer Lesestörung. Hiernach zeigen Kinder vor allem eine sehr niedrige Lesegeschwindigkeit. Dadurch, dass sie wesentlich länger brauchen zum Erlesen von Sätzen, können sie sich nicht auf den Inhalt des Gelesenen fokussieren, was das Leseverständnis beeinträchtigt, und benötigen auch deutlich mehr Zeit als Kinder ohne Lesestörung. Dummer-Smoch (2007, S. 375) hält speziell zum Lesen schließlich noch fest, dass Abweichungen vom ungestörten und normalen Leseerwerb als „Leselernschwierigkeiten“ bezeichnet werden können und nennt hierfür unterschiedliche Ursachen wie beispielsweise ein Kind, welches zu früh eingeschult wurde oder einen Jungen, der aufgrund des Berufes des Vaters im ersten Schuljahr schon drei verschiedene Grundschulen besuchen musste und entsprechend nicht „Fuß fassen konnte“ (Dummer-Smoch, 2007, S. 375). Diese Schwierigkeiten entsprechen allerdings nicht den

Kriterien einer Lesestörung oder -schwäche. Dennoch ist eine Förderung unerlässlich, wobei die Art der Förderung abhängig von den Ursachen für die Schwierigkeiten ist.

### 6.3 Prävalenzen

Die Bestimmung der Häufigkeit von Lese-Rechtschreibstörungen gestaltet sich schwierig. Für den Erstleseunterricht sind die Prävalenzen generell nur schwer zu ermitteln, da Lehrkräfte unterschiedlich schnell vorgehen und daher der Maßstab, was Kinder bereits kennen und können sollen, zu sehr vom Vorgehen der Lehrkraft abhängt. Ein Vergleich der Leistungen mit normierten Leistungstests ist daher erst ab Ende der ersten Klasse bzw. Anfang der zweiten Klasse sinnvoll (Klicpera et al., 2017, S. 137). Zudem ist die Prävalenz auch stark abhängig von der zugrunde gelegten Definition der Lese-Rechtschreibstörung. So konstatieren Schulte-Körne und Remschmidt (2003, S. A 397) die Schwankungen in den Häufigkeitsangaben in Abhängigkeit der diagnostischen Kriterien wie auch der gewählten Untersuchungsmethodik und geben Zahlen im Bereich von 3 bis 20 Prozent an. Im internationalen Vergleich von Studien kommen sie zu einer Prävalenz von 4 bis 5 Prozent. Strehlow und Haffner (2007, S. 398) führen Prävalenzraten von 1 bis 40 Prozent an. Hieran sieht man auch sehr deutlich, wie groß die Schwankungen sein können, wenn unterschiedliche Definitionen und Kriterien zugrunde gelegt werden.

Werden Forschungskriterien der ICD-10 herangezogen, welche zwei Standardabweichungen unter dem von der Intelligenz vorhergesagten Wert im schulischen Leistungstest fordern, liegt die Prävalenz von Lese-Rechtschreibstörungen bei zwei bis vier Prozent (Klicpera et al., 2017, S. 138). Am häufigsten wird hingegen eine Standardabweichung von 1.5 empfohlen (Klicpera et al., 2017, S. 139). Allerdings wird angemerkt, dass das gewählte Ausmaß an Abweichung die Prävalenz bestimmt und die Zahlen daher willkürlich sind, da zwei Standardabweichungen unterhalb des Durchschnitts entsprechend der Normalverteilung einer Häufigkeit von unter zwei Prozent entspricht (Klicpera et al., 2017, S. 139; Schmidt-Atzert & Amelang, 2012, S. 166). Steinbrink und Lachmann (2014, S. 75) fassen zusammen, dass die meisten Prävalenzangaben zwischen vier und acht Prozent liegen. Sie halten ebenso die Schwierigkeit von Prävalenzangaben fest, solange keine Einigkeit über die diagnostischen Kriterien vorherrscht: Bei einem Kriterium von einer Standardabweichung unterhalb des Durchschnitts als Diskrepanz sind knapp 16 Prozent der Schülerinnen und Schüler betroffen und bei einer Diskrepanz von zwei Standardabweichungen weniger als 3 Prozent. Diese Prävalenz verringert sich allerdings noch einmal, da Ausschlusskriterien ebenfalls einbezogen werden müssen. Unter Berücksichtigung der Intelligenz (also dem doppelten Diskrepanzkriterium), reduziert sich die Prävalenz auf ein Prozent, was von Steinbrink und Lachmann allerdings als „unverhältnismäßig klein“ bezeichnet wird (2014, S. 75). Sie kommen zu dem Schluss, dass die Kriterien inhaltlich festgelegt werden müssen, da die Grenzwerte sonst rein „statistisch-theoretischer Natur“ ohne praktische Relevanz sind (Steinbrink & Lachmann, 2014, S. 76). Um dieses Dilemma zu umgehen, werden kriteriumsorientierte Tests

vorgeschlagen (Klicpera et al., 2017, S. 139). Dieses Vorgehen entspricht den internationalen Vergleichsstudien, wo zusätzlich zu den standardisierten Leistungstests noch „Minimalanforderungen an die Lesefähigkeit“ aufgestellt wurden (Klicpera et al., 2017, S. 139).

Fischbach et al. (2013, S. 67) setzten sich in ihrer Studie mit der Problematik der Prävalenzen und des doppelten Diskrepanzkriteriums auseinander. Sie kritisieren, dass in anderen Untersuchungen die Prävalenzen für isolierte Lernstörungen und -schwächen überschätzt werden, da nicht kontrolliert wird, ob bei diesen Kindern noch andere Lernstörungen vorliegen, welche dann ebenfalls in den Prävalenzangaben zu kombinierten Lernstörungen genannt werden (Fischbach et al., 2013, S. 68). Ihre Untersuchung fand mit einer unausgelesenen Stichprobe von 2195 Kindern Ende der zweiten bzw. Anfang der dritten Klasse statt. Sie hatte zum Ziel, die Häufigkeit von isolierten sowie kombinierten Lernschwächen und Lernstörungen zu erfassen, das Geschlechterverhältnis bei diesen zu erheben und zu überprüfen, inwiefern sich die Prävalenzen verändern, wenn Kinder mit kombinierten Lernstörungen und -schwächen auch bei den isolierten Lernstörungen und -schwächen erfasst werden (Fischbach et al., 2013, S. 68). Hier definierten sie, dass eine Lernschwäche vorhanden ist, wenn eine Teilleistung (im Lesen, Schreiben oder im Rechnen) bei einem T-Wert unter 40 liegt, während die Intelligenz einen Wert von mindestens 85 auf der IQ-Skala hat. Eine Lernstörung liegt dagegen dann vor, wenn zusätzlich die Diskrepanz zwischen der Intelligenz und der jeweiligen Minderleistung mindestens 1.2 Standardabweichungen beträgt. Eine isolierte Lernschwäche bzw. Lernstörung wird dann diagnostiziert, wenn nur eine Teilleistung (Rechnen, Lesen oder Schreiben) unter einem T-Wert von 40 liegt. Eine kombinierte Lernstörung liegt dagegen vor, wenn die Diskrepanz zwischen der Intelligenz und der Minderleistung in allen unterdurchschnittlichen Teilleistungen vorhanden ist. Ist dies nicht der Fall, diagnostizieren sie stattdessen eine kombinierte Lernschwäche (Fischbach et al., 2013, S. 69). Es zeigt sich, dass isolierte Lernschwächen jeweils eine Prävalenz von 5 bis 6 Prozent aufweisen, wobei Schwierigkeiten im Rechtschreiben am häufigsten auftreten. Die isolierte Leseschwäche steht mit 5 Prozent auf Rang 3. Eine kombinierte Lese-Rechtschreibschwäche ist mit 4 Prozent noch weniger häufig. Ein fast identisches Bild zeigt sich bei den Lernstörungen, wobei die kombinierten Lernstörungen am seltensten auftreten (Fischbach et al., 2013, S. 70). Entsprechend der ursprünglich formulierten Annahme erhöht sich die Prävalenz der isolierten Lernstörungen und -schwächen beträchtlich, wenn das Vorliegen einer kombinierten Lernstörung bzw. -schwäche nicht entsprechend kontrolliert wird (Fischbach et al., 2013, S. 71). Die Ergebnisse dieser Studie zeigen daher deutlich, dass stets alle drei Teilleistungen betrachtet werden müssen, wenn Lernstörungen bzw. -schwächen diagnostiziert werden sollen und dass ohne diese Kontrolle eine Angabe der korrekten Prävalenz nicht möglich ist. Zudem zeigt sich hier auch, dass Lernschwierigkeiten nicht selten sind, da insgesamt ca. ein Drittel der Kinder von basalen Lernschwierigkeiten betroffen war, wobei 13 Prozent der Kinder eine Lernstörung im Sinne des doppelten Diskrepanzkriteriums aufwiesen. Es ergibt sich aus diesen Ergebnissen aber auch, dass 43 Prozent der Kinder mit einer Lernschwäche, was 10 Prozent aller

Schülerinnen und Schüler entspricht, gleichzeitig aufgrund des fehlenden doppelten Diskrepanzkriteriums keine Lernstörung hatten und ihnen entsprechend die Möglichkeit entgeht, eine Förderung zu erhalten (Fischbach et al., 2013, S. 71). Die Autorinnen und Autoren der Studie plädieren zudem dafür, die Lesestörung bzw. Leseschwäche getrennt von der Lese-Rechtschreibstörung bzw. -schwäche zu diagnostizieren, da diese isoliert häufiger vorkommen als in der kombinierten Form (Fischbach et al., 2013, S. 72).

Moll, Kunze, Neuhoff, Bruder und Schulte-Körne (2014, S. 2) hatten ihrer Untersuchung sehr ähnliche Fragestellungen zugrunde gelegt. Sie untersuchten 1633 Kinder in der dritten und vierten Klasse. Auch hier standen die Frage der Prävalenz der isolierten Lese-, Rechtschreib- und Rechenstörung sowie Kombinationen davon und die Frage nach der Geschlechterverteilung im Fokus. Sie verzichteten allerdings, analog zu den diagnostischen Kriterien des DSM-5, auf das doppelte Diskrepanzkriterium. Sie legten drei unterschiedliche Diskrepanzkriterien für die schulische Leistung an: die Diskrepanz von einer Standardabweichung, von 1.25 und von 1.5 Standardabweichungen (Moll et al., 2014, S. 4). Es zeigte sich, dass die Hälfte der Kinder mit Lernschwierigkeiten eine isolierte Lernstörung aufwies, die andere Hälfte hatte eine kombinierte. Dies ist aber abhängig von der Höhe des verwendeten Diskrepanzkriteriums. Bei einer Standardabweichung von 1 kamen kombinierte Lernstörungen signifikant häufiger vor als isolierte. Isolierte Lernstörungen hatten dagegen eine gleich hohe oder auch höhere Prävalenz als kombinierte Lernstörungen, wenn man eine Standardabweichung von 1.5 annahm (Moll et al., 2014, S. 5). Sie gingen ebenfalls der Frage nach, wie sich die Häufigkeit verändert, wenn die Kinder in der Analyse nicht berücksichtigt werden, die in einer Teilleistung (zum Beispiel im Lesen) unterdurchschnittlich abschneiden und in einer anderen (zum Beispiel im Rechtschreiben) knapp über dem Cutoff-Wert liegen. Die Häufigkeit des Auftretens der isolierten Lernstörung ändert sich hierdurch allerdings nicht. Es zeigten sich sogar teilweise sehr große Unterschiede in den Leistungen in den beiden Teilfertigkeiten, sodass Moll et al. die Hypothese aufstellen, dass die beteiligten kognitiven Prozesse für die Leseflüssigkeit und die Rechtschreibung unterschiedliche sein könnten (2014, S. 7). Sehr häufig wird auch mitgeteilt, ob sich Jungen und Mädchen in der Prävalenz von Lernstörungen unterscheiden. Schulte-Körne und Remschmidt (2003, S. A 397) geben hierzu an, dass Jungen etwa doppelt bis dreifach so häufig von einer Lese-Rechtschreibstörung betroffen sind. Klicpera et al. (2017, S. 141) fassen zusammen, dass fast immer mehr Jungen als Mädchen betroffen zu sein scheinen, dass dies aber nur für klinische Stichproben und Untersuchungen in sonderpädagogischen Einrichtungen in großem Ausmaß zutrifft. Für unselektierte Stichproben ist der Unterschied meist weniger gravierend. Sie geben hier ein Verhältnis von 2:1 an sowie ein Verhältnis von 3:2 bis 3:1 für klinische und sonderpädagogische Bereiche. Wenn man Lesen und Rechtschreiben getrennt betrachtet, werden für den Bereich der Rechtschreibung etwas höhere Prävalenzen für Jungen angegeben, für Lesestörungen im Bereich der Leseflüssigkeit kaum Geschlechtsunterschiede und für das Leseverständnis durchgehend bessere Leistungen bei den Mädchen als bei den Jungen (Klicpera et al., 2017,

S. 142). Wyschkon et al. (2018, S. 119) berichten ebenfalls über eine doppelt so hohe Prävalenz von Lese-Rechtschreibschwierigkeiten bei Jungen im Vergleich zu Mädchen.

Steinbrink und Lachmann (2014, S. 79) geben an, dass auch international betrachtet bei Männern bzw. Jungen häufiger Lese-Rechtschreibstörungen diagnostiziert werden als bei Frauen bzw. Mädchen. Das genaue Verhältnis schwankt allerdings häufig. Es wird aber meist von einer doppelt so hohen Prävalenz bei Jungen ausgegangen. Biologische Ursachen für diese Unterschiede werden heutzutage angezweifelt (Steinbrink & Lachmann, 2014, S. 79). Die Studie von Fischbach et al. (2013) betrachtete neben den allgemeinen Prävalenzen für Lernstörungen und -schwächen auch das Geschlechterverhältnis. Bezogen auf Schwierigkeiten im Lesen zeigt sich hier, dass die isolierte Leseschwäche und die isolierte Lesestörung bei Jungen ca. eineinhalbmals so häufig auftritt wie bei Mädchen. Diese Unterschiede sind allerdings nicht signifikant (Fischbach et al., 2013, S. 71). Das gleiche Bild zeigt sich auch für Schwierigkeiten im Rechtschreiberwerb, wobei der Unterschied in der isolierten Rechtschreibschwäche signifikant ist. Bei einer kombinierten Lese-Rechtschreibstörung bzw. einer kombinierten Lese-Rechtschreibschwäche zeigt sich sogar eine dreimal so hohe Prävalenz für Jungen wie für Mädchen. Bei einer kombinierten Lernstörung zeigen sich dagegen keine Unterschiede zwischen Jungen und Mädchen in der Prävalenz.

Moll et al. (2014, S. 5) haben ebenfalls das Geschlechterverhältnis in Bezug zur Prävalenz betrachtet. Hier hatten Jungen häufiger als Mädchen eine isolierte Rechtschreibstörung sowie eine kombinierte Lese-Rechtschreibstörung. Keine Unterschiede gab es dagegen bei der isolierten Lesestörung sowie bei der Lernstörung in allen drei Teilleistungsbereichen, Lesen, Rechnen und Schreiben. Ebenso wie bei Fischbach et al. (2013) gab es auch hier keine Geschlechtsunterschiede in der Häufigkeit der kombinierten Lernstörung.

Insgesamt zeigt sich also, dass die Lesestörung durchaus isoliert auftreten kann und das auch häufiger als die kombinierte Lese-Rechtschreibstörung, auch wenn sie nach der ICD-10 nicht einzeln diagnostiziert wird. Des Weiteren schwanken die Prävalenzen je nachdem, ob man das einfache oder das doppelte Diskrepanzkriterium anlegt und wie hoch die Diskrepanz jeweils ausfallen soll. Jungen scheinen etwas häufiger betroffen zu sein als Mädchen, allerdings nicht bei der kombinierten Lernstörung.

#### 6.4 Komorbiditäten und Auswirkungen

Der lange Zeit vorherrschende Irrglaube, dass Kinder, die schlecht im Lesen und Schreiben sind, gleichzeitig gut im Rechnen sind, ist empirisch widerlegt (Schulte-Körne, 2010, S. 719). Wie bereits in Kapitel 6.3 zu den Prävalenzen genannt, können die Lernstörungen in den Bereichen Lesen, Schreiben sowie Rechnen auch kombiniert auftreten. Fischbach et al. (2013, S. 70) fanden in ihrer Studie eine Prävalenz von zwei Prozent für das Auftreten einer kombinierten Lernstörung und von vier Prozent für eine kombinierte Lernschwäche. Schulte-Körne (2010, S. 719) gibt die Komorbidität von Rechenstörungen mit 20 bis 40 Prozent an. Da die Lesefähigkeit allerdings eine wichtige Rolle spielt beim Lernen in anderen

Unterrichtsfächern, ist davon auszugehen, dass sich eine Lesestörung auch negativ auf die Leistungen in anderen Unterrichtsfächern auswirkt (Klicpera et al., 2017, S. 143–44; Mayer, 2016, S. 46). Zusätzlich wird in den Kriterien der ICD-10 auch auf mögliche „begleitende Störungen im emotionalen und Verhaltensbereich“ hingewiesen (Deutsches Institut für Medizinische Dokumentation und Information, 2018, S. 325). Insgesamt deutet sich daher an, dass sich Lese-Rechtschreibstörungen auf verschiedene Lebensbereiche auswirken können und auch weitreichende Folgen haben. Hierauf wird in diesem Abschnitt genauer eingegangen.

#### 6.4.1 Psychische Auffälligkeiten

Zahlreiche Studien haben sich mit dem Auftreten von Komorbiditäten und Folgeerkrankungen bzw. Auswirkungen auf die psychische Gesundheit der betroffenen Kinder beschäftigt. Insgesamt konnte aufgezeigt werden, dass Kinder mit Lernschwächen und -störungen im Vergleich zu Kindern ohne solche Schwierigkeiten in verschiedenen Bereichen Auffälligkeiten zeigen. Bender et al. (2017) nennen zusammenfassend als Komorbiditäten ADHS, Sprachentwicklungsstörungen sowie Auffälligkeiten in der Grob- und Feinmotorik und als mögliche Folge psychische Störungen wie „Misserfolgserlebnisse, Selbstzweifel, Mobbing und Überforderung“ bis hin zu „Depressionen, Angststörungen und Störungen des Sozialverhaltens“ sowie „Konzentrationschwächen, motorische Unruhe, Motivationsverlust oder auch eine generelle Schulangst und -unlust“ (Bender et al., 2017, S. 66). Mayer (2016, S. 51–52) hält fest, dass sich Lese-Rechtschreibprobleme negativ auf die sprachliche, die kognitive und die sozio-emotionale Entwicklung auswirken können. Vor allem Schwierigkeiten beim Lesen führen dazu, dass Kinder Bücher weder für den eigenen Wissenszuwachs noch für die Wortschatzerweiterung nutzen können, sodass hieraus weitere Schwierigkeiten beim Lernen entstehen können. Im Bereich der sozio-emotionalen Entwicklung sieht er vor allem Probleme im Jugendalter, da die aktive Teilnahme in sozialen Netzwerken ausreichende Lese- und Rechtschreibkenntnisse voraussetzen und somit ein heutzutage wesentlicher Bestandteil der Interaktion mit anderen gefährdet sein kann.

Klicpera et al. (2017, S. 145) führen an, dass für das Auftreten von emotionalen und Verhaltensproblemen gemischte Befunde vorliegen. Zum einen zeigen sich keine langfristigen negativen Folgen auf das Selbstwertgefühl oder in der Häufigkeit psychischer Störungen bei Männern mit Lese-Rechtschreibstörungen. Bei Frauen zeigt sich dagegen eine geringere berufliche Qualifikation. Außerdem gehen sie früher feste Bindungen ein und haben in der Folge häufiger Partnerschaftsprobleme (Klicpera et al., 2017, S. 146). Strehlow und Haffner (2007, S. 407) kommen in Bezug zu emotionalen Störungen ebenfalls zu widersprüchlichen Ergebnissen. Nach Sichtung von Studien und aufgrund einer eigenen Untersuchung halten sie fest, dass es keine andauernden gravierenden emotionalen Störungen als Folge zu geben scheint. Zwischen der Lese-Rechtschreibschwäche und hyperkinetischen Störungen sowie dissozialen Verhaltensweisen zeigen sich allerdings starke Zusammenhänge (Strehlow & Haffner,

2007, S. 408). Externale Störungen stehen dagegen im Zusammenhang mit Lese-Rechtschreibstörungen. Es deutet sich zwar an, dass diese eine Folge der Lese-Rechtschreibstörungen sind, dies ist aber aufgrund der korrelativen Studien nicht sicher zu sagen. Es könnten auch andere Faktoren moderierend einwirken (Strehlow & Haffner, 2007, S. 409).

Mugnaini, Lassi, La Malfa und Albertini (2009) arbeiten in ihrem Übersichtsartikel den Stand der Forschung zum Auftreten von internalisierenden Störungen bei Kindern mit Lese-Rechtschreibstörungen heraus. Sie kommen zu dem Schluss, dass Lesestörungen mit depressiven und Angst-Symptomen zusammenhängen und nennen weitere Risikofaktoren, die einen schlechteren Verlauf begünstigen, wie beispielsweise den Grad der Lesestörung, das Vorhandensein einer Rechenstörung oder weiterer Schwierigkeiten, ADHS, eine niedrige Intelligenz oder das späte Erkennen der Leseschwierigkeiten (Mugnaini et al., 2009, S. 260). Sie fanden auch heraus, dass die Mütter von Kindern mit Lesestörungen höhere Stress-Level und mehr depressive Symptome zeigten als Mütter von Kindern ohne Lesestörungen. Auch hier muss berücksichtigt werden, dass es sich um Korrelationsstudien handelt und so nur von einem Zusammenhang, nicht aber von Ursache und Wirkung gesprochen werden kann (Mugnaini et al., 2009, S. 261).

Esser, Wyschkon und Schmidt (2002, S. 235) bezeichnen den Entwicklungsverlauf bei Lese-Rechtschreibstörungen insgesamt als ungünstig und stützen sich dabei auf ihre Langzeitstudie. Im Fokus stand neben dem Verlauf auch der Schulerfolg der betroffenen Schülerinnen und Schüler, die berufliche Laufbahn sowie mögliche psychische Auffälligkeiten, die sich in der Folge entwickeln könnten (Esser et al., 2002, S. 236). Verglichen wurden Kinder mit Lese-Rechtschreibstörung (definiert über das doppelte Diskrepanzkriterium), Kinder mit anderen umschriebenen Entwicklungsstörungen und als Kontrollgruppe Kinder ohne solche Störungen. Insgesamt fanden vier Erhebungen statt: Die erste, als die Kinder acht Jahre alt waren, die nächste im Alter von 13, die dritte im Alter von 18 und die letzte Erhebung im Alter von 25 Jahren (Esser et al., 2002, S. 236). Psychiatrische Auffälligkeiten wurden über Interviews mit den Betroffenen und mit deren Eltern erhoben. Insgesamt wurden verschiedene Symptomgruppen erfasst: Dissoziale Symptome, Sucht-Symptome, internalisierende Symptome, hyperkinetische Symptome, andere Symptome (z. B. Nägelkauen, Tics), psychotische Symptome sowie Sprach- und Sprechstörungen (Esser et al., 2002, S. 236–237). Der höchste Schulabschluss sowie die hier erzielte Durchschnittsnote wurden als Indikator für den Schulerfolg in die Analyse einbezogen. Auch der Berufsabschluss und das Selbstbild wurden erhoben. Auf diese Ergebnisse wird in Abschnitt 6.4.3 eingegangen. Nicht mehr erfasst wurden die Lese- und Rechtschreibleistungen der Betroffenen im Erwachsenenalter (Esser et al., 2002, S. 237). Es zeigte sich, dass Kinder mit Lese-Rechtschreibstörungen sowohl im Kindesalter wie auch im weiteren Verlauf signifikant häufiger dissoziale und hyperkinetische sowie Sucht-Symptome zeigen als die Kontrollgruppe, wobei die Sucht-Symptome vor allem auf vermehrten Alkohol- und Nikotinkonsum zurückzuführen sind (Esser et al., 2002, S. 239). Im Vergleich mit der Gruppe der Kinder mit anderen umschriebenen Entwicklungsstörungen

fanden sich nur tendenzielle, aber keine signifikanten Unterschiede für die hier genannten Symptome (Esser et al., 2002, S. 241).

Fischbach, Schuchardt, Mähler und Hasselhorn (2010, S. 202) kritisieren, dass die Forschung über Folgen für Kinder mit Lernstörungen und Lernschwächen lange vernachlässigt wurde. In ihrer Studie beschäftigen sie sich entsprechend mit der Frage nach sozio-emotionalen Auffälligkeiten bei Kindern mit Lernschwierigkeiten, wobei sie zwischen Kindern mit einer Lernstörung und Kindern mit einer Lernschwäche unterscheiden. Sie sehen das mögliche Auftreten externalisierender Störungen wie auch internalisierender Störungen bei Lernstörungen in den Bereichen Lesen, Schreiben und Rechnen als von Studien recht eindeutig nachgewiesen an, während sie die Ergebnisse zu emotionalen Störungen und zum Sozialverhalten als widersprüchlich bezeichnen (Fischbach et al., 2010, S. 202–203). In ihre Untersuchung eingegangen sind Daten von Kindern aus den Jahren 2000 bis 2009. Ein Ausschlusskriterium waren komorbide Störungen des Sozialverhaltens. Eine Lernstörung wurde hier mit dem doppelten Diskrepanzkriterium von 1.2 Standardabweichungen definiert. Eine Lernschwäche wurde diagnostiziert, wenn die Minderleistung und die Intelligenz diese Diskrepanz nicht erfüllten. Es wurde nicht zwischen Schreiben und Lesen differenziert, sodass dieser Gruppe alle Kinder zugeteilt wurden, die im Schreiben, im Lesen oder in beiden Bereichen Minderleistungen zeigten (Fischbach et al., 2010, S. 203). Erhoben wurden emotionale Auffälligkeiten, schulische Motivation und Selbstkonzept, Verhaltensauffälligkeiten sowie soziales Erleben und Verhalten. Diese Auffälligkeiten wurden über Befragungen der Kinder, deren Eltern sowie einer Lehrkraft des jeweiligen Kindes erfasst (Fischbach et al., 2010, S. 204). Im Bereich des sozialen Erlebens ergaben sich keine Unterschiede zu Kindern der Normierungsstichproben. Kinder mit Lernstörungen und -schwächen fühlen sich also nicht weniger oder stärker sozial integriert (Fischbach et al., 2010, S. 206). In allen anderen Bereichen weichen die Werte der Kinder mit Lernstörungen und -schwächen vom Mittelwert der Normierungsstichproben ab (Fischbach et al., 2010, S. 204). Aus Lehrer- und Elternsicht zeigen sich vermehrt Schwierigkeiten in der sozialen Interaktion (Fischbach et al., 2010, S. 208). Eine niedrigere Anstrengungsbereitschaft und ein niedrigeres Selbstkonzept bei Kindern mit Lernstörungen und -schwächen werden neben den Auffälligkeiten im emotionalen Bereich von Fischbach et al. (2010, S. 208) besonders hervorgehoben. Über die Ursachen der vermehrt auftretenden Schwierigkeiten kann die Studie aufgrund der korrelativen Daten keine Aussage machen. Da hier eine Inanspruchnahme-Stichprobe vorliegt von Kindern, die mit ihren Eltern eine Beratungsstelle für Lernschwierigkeiten aufgesucht haben, sind die Ergebnisse nicht unbedingt auf unselektierte Stichproben übertragbar (Fischbach et al., 2010, S. 208).

Schwierigkeiten im Rechnen oder Schwierigkeiten im Lesen und/oder Schreiben gehen nicht mit einem unterschiedlichen Ausmaß an sozio-emotionalen Auffälligkeiten einher (Fischbach et al., 2010, S. 208). Wenn Lernschwierigkeiten dagegen kombiniert, also im Schreiben und/oder Lesen sowie im Rechnen, auftreten, zeigen sich stärkere Auffälligkeiten als bei

Kindern, die nur in einem Teilbereich Lernschwierigkeiten haben (Fischbach et al., 2010, S. 209). Zuletzt wurde auch darauf geblickt, ob Lernschwächen und Lernstörungen mit unterschiedlich starken Ausprägungen an Auffälligkeiten einhergehen. Für das emotionale Erleben und die Motivation zeigten sich keine Unterschiede, wohl aber für externalisierendes Verhalten, soziale Probleme sowie Symptome einer ADHS. Hier hatten Kinder mit Lernstörungen stärkere Ausprägungen. Fischbach et al. kommen zu dem Schluss, dass das stärkere Ausmaß an Verhaltensschwierigkeiten im Rahmen von entsprechender Förderung Beachtung finden sollte, aber ansonsten kaum Unterschiede zwischen Kindern mit einer Lernstörung und Kindern mit einer Lernschwäche vorhanden sind (Fischbach et al., 2010, S. 209).

In der Studie von Kohn, Wyschkon und Esser (2013) standen Komorbiditäten bei Teilleistungsstörungen im Bereich Lesen und Schreiben sowie im Rechnen und eventuelle Unterschiede zwischen diesen beiden Gruppen im Fokus. Ausgegangen wurde von der Vermutung, dass Kinder mit Rechenstörungen häufiger internalisierende und Kinder mit Lese-Rechtschreibstörungen häufiger externalisierende Störungen zeigen. Ebenso wurden die gesundheitsbezogene Lebensqualität sowie das Selbstwertgefühl der betroffenen Kinder berücksichtigt. Die Lese-Rechtschreibstörung wurde mithilfe des doppelten Diskrepanzkriteriums definiert, wobei als Maß die nonverbale Intelligenz berücksichtigt wurde (Kohn, Wyschkon & Esser, 2013, S. 9). Vier Symptomgruppen wurden über Fremdbeurteilung durch die Eltern erhoben: Symptome der hyperkinetischen Störung, der Störung des Sozialverhaltens, der emotionalen Störung sowie somatische Symptome (Kohn, Wyschkon & Esser, 2013, S. 10). Es zeigte sich bei Kindern mit einer Lese-Rechtschreibstörung häufiger hyperkinetische Symptome als bei Kindern mit Rechenstörungen oder mit durchschnittlichen schulischen Leistungen im Lesen, Schreiben und Rechnen. Für somatische Symptome ergaben sich nur bei Mädchen, nicht aber bei Jungen Unterschiede. Bei Mädchen mit Lese-Rechtschreibstörung traten dabei allerdings weniger Symptome auf als bei Mädchen mit Rechenstörung oder mit durchschnittlichen Leistungen. Für Hyperaktivität zeigte sich dagegen ein anderes Bild. Hier gab es signifikante Unterschiede nur für Jungen. Sowohl Jungen mit Lese-Rechtschreibstörungen wie auch Jungen mit Rechenstörungen zeigten mehr Symptome als Jungen in der Kontrollgruppe (Kohn, Wyschkon & Esser, 2013, S. 11). In Bezug auf die Lebensqualität gab es höhere Problemwerte bei Kindern mit einer Lese-Rechtschreibstörung. Diese Kinder zeigten eine niedrigere Zufriedenheit mit der Schule als Kinder der Kontrollgruppe (Kohn, Wyschkon & Esser, 2013, S. 14). Ebenso wurde ein signifikanter Unterschied in der Anstrengungsbereitschaft gefunden. Kinder mit Lese-Rechtschreibschwäche oder mit Rechenschwäche hatten signifikant niedrigere Werte als Kinder der Kontrollgruppe. Es wurden in dieser Studie dagegen keine Unterschiede für Symptome der Störung des Sozialverhaltens gefunden. Da hierfür aber nur wenige Symptome angegeben waren, könnte ein Bodeneffekt vorliegen (Kohn, Wyschkon & Esser, 2013, S. 13). Insgesamt zeigt sich in der Studie, dass es bezüglich der Komorbiditäten von psychischen Auffälligkeiten keine Unterschiede zwischen Kindern mit

einer Lese-Rechtschreibstörung und Kindern mit einer Rechenstörung gibt (Kohn, Wyschkon & Esser, 2013, S. 14).

Goldston et al. (2007, S. 26) konzentrierten sich auf Jugendliche mit Lesestörung und gingen der Frage nach, ob diese im Vergleich zu Jugendlichen mit durchschnittlichen Lesefähigkeiten häufiger psychische Auffälligkeiten zeigen. Die teilnehmenden Jugendlichen wurden über einen Zeitraum von 4.5 Jahren befragt. Es zeigte sich, dass Jugendliche mit einer Lesestörung signifikant häufiger von ADHS, affektiven Störungen und Angststörungen, hier vor allem von sozialer Phobie und generalisierter Angststörung, betroffen sind als Jugendliche ohne Lesestörung (Goldston et al., 2007, S. 29). Für Störungen des Sozialverhaltens konnte kein Unterschied zwischen Jugendlichen mit und ohne Lesestörung gefunden werden (Goldston et al., 2007, S. 30). Bei Kontrolle des sozioökonomischen Hintergrunds zeigten sich weiterhin ADHS und Angststörungen mit einer höheren Auftretenshäufigkeit bei Jugendlichen mit Lesestörungen. Je älter die Jugendlichen wurden, umso mehr trat in der Gruppe der Jugendlichen mit Lesestörung Substanzmissbrauch auf (Goldston et al., 2007, S. 29). Die übrigen psychischen Auffälligkeiten zeigten keine Veränderung über die Zeit (Goldston et al., 2007, S. 30).

Endlich, Dummert, Schneider und Schwenck (2014) befassten sich mit dem Ausmaß an Verhaltensproblemen bei Kindern mit umschriebenen Entwicklungsstörungen. Ein besonderer Fokus lag auf möglichen Unterschieden zwischen isolierten und kombinierten Lernstörungen, da ein additiver Effekt der Schwierigkeiten von isolierten zu kombinierten Lernstörungen angenommen wurde (Endlich et al., 2014, S. 63). Ebenso sollte untersucht werden, ob eine längere Dauer der Beeinträchtigung auch zu stärkeren Verhaltensproblemen führt. Eine Lernschwäche lag dann vor, wenn das Kind mindestens durchschnittliche Intelligenz ( $IQ > 85$ ) aufwies, bei einer Minderleistung im Bereich Rechnen, Lesen oder Schreiben ( $PR < 25$ ), sowie eine Diskrepanz von mindestens einer Standardabweichung zwischen schulischer Leistung und Intelligenzquotient vorlag. Es handelt sich hier um eine Langzeitstudie mit der ersten Erhebung zu Beginn der zweiten Klasse und der letzten Erhebung in der Mitte der vierten Klasse (Endlich et al., 2014, S. 63). Die Verhaltensprobleme der Kinder wurden über einen Elternfragebogen erfasst (Endlich et al., 2014, S. 64). Kinder, die sowohl im Rechnen wie auch im schriftsprachlichen Bereich Lernschwächen haben, nicht aber Kinder, die nur in einem Teilbereich eine Lernschwäche haben, zeigen hiernach häufiger internalisierende und externalisierende Verhaltensprobleme gegenüber der Kontrollgruppe, die in allen Bereichen mindestens durchschnittliche Leistungen erbracht hat. Ebenso berichteten die Eltern im Vergleich zu den Eltern der Kinder aus der Kontrollgruppe häufiger von einer ADHS-Diagnose (Endlich et al., 2014, S. 67). Kinder, die zu beiden Messzeitpunkten, also in der zweiten und der vierten Klasse, eine Lernschwäche aufwiesen, zeigten ein stärkeres Ausmaß an Aufmerksamkeitsproblemen als Kinder, die nur zu einem Messzeitpunkt betroffen waren (Endlich et al., 2014, S. 68). Die Dauer der Lernschwierigkeiten scheint demnach die psychosoziale Belastung zu erhöhen.

Huck und Schröder (2016) analysierten Daten aus 210 Diagnosegesprächen. Es liegt somit eine Inanspruchnahme-Stichprobe vor, da alle Diagnosegespräche im Rahmen einer Beratung vor Beginn einer Lerntherapie stattgefunden haben (Huck & Schröder, 2016, S. 158). Im Zentrum stand die Frage nach dem Auftreten von Symptomen einer somatoformen Störung bei Kindern mit Lernschwierigkeiten. Ebenso berücksichtigt wurden Mobbing Erfahrungen (Huck & Schröder, 2016, S. 158). Die untersuchten Kinder und Jugendlichen befanden sich zum Zeitpunkt der Diagnose in der ersten bis zwölften Klasse mit einem Median in der vierten Klassenstufe. Es zeigte sich bei knapp 70 Prozent der Kinder und Jugendlichen mindestens eine psychosoziale Auffälligkeit. Somatoforme Störungen, internalisierende Störungen, Konzentrationsschwierigkeiten sowie Mobbing Erfahrungen traten jeweils bei über 20 Prozent der Stichprobe auf. Schulvermeidung bzw. -unlust und Schwierigkeiten in der sozialen Integration traten seltener auf (Huck & Schröder, 2016, S. 159). Kinder und Jugendliche ab der sechsten Klasse sprachen häufiger von Belastungen als jüngere. Mädchen berichteten häufiger von internalisierenden Störungen und von Symptomen einer somatoformen Störung, Jungen häufiger von ADHS (Huck & Schröder, 2016, S. 160). Die Übertragbarkeit der Ergebnisse ist allerdings eingeschränkt, da es sich um eine Inanspruchnahme-Stichprobe handelt.

Anhand der vorgestellten Studien zeigt sich, dass psychische Auffälligkeiten bei Kindern mit Lese-Rechtschreibschwierigkeiten in stärkerem Maß auftreten als bei Kindern ohne solche Schwierigkeiten. Das Vorhandensein hyperkinetischer Störungen ist dabei häufig untersucht worden. Diese treten sowohl bei Schwächen wie auch bei Störungen im Bereich Lesen und Schreiben vermehrt auf. Die Studie von Kohn, Wyschkon und Esser (2013) konnte dies allerdings nur für Jungen, nicht aber für Mädchen feststellen, während Huck und Schröder (2016, S. 160) von einer Geschlechterdifferenz berichten, wonach Jungen mit Lese-Rechtschreibschwierigkeiten häufiger unter ADHS leiden als Mädchen. Für Störungen des Sozialverhaltens zeigen sich gemischte Befunde, da neben Studien, welche für betroffene Kinder und Jugendliche ein vermehrtes Auftreten aufzeigen konnten, auch zwei Studien vorliegen, die keine Häufung bei Kindern mit Lese-Rechtschreibschwierigkeiten finden konnten (Goldston et al., 2007, S. 30; Kohn, Wyschkon & Esser, 2013, S. 13). Internalisierende Störungen treten dagegen vermehrt auf (Goldston et al., 2007, S. 29; Huck & Schröder, 2016, S. 159; Mugnaini et al., 2009, S. 260). Auf motivationaler Ebene konnte unter anderem eine niedrigere Anstrengungsbereitschaft bei Kindern und Jugendlichen mit Lese-Rechtschreibschwierigkeiten gefunden werden, was wiederum die vorhandenen Schwierigkeiten verstärken könnte (Fischbach et al., 2010, S. 208; Kohn, Wyschkon & Esser, 2013, S. 16). Es wird auch deutlich, dass kombinierte Lernstörungen und -schwächen sowie eine längere Dauer zu einer stärkeren Belastung führen (Endlich et al., 2014, S. 67–68; Fischbach et al., 2010, S. 209).

#### 6.4.2 Verlauf von Lese-Rechtschreibschwierigkeiten

Wie im vorigen Abschnitt gezeigt wurde, kommt es mit anhaltender Dauer der Lese-Rechtschreibschwierigkeiten zu einer stärkeren Belastung der betroffenen Personen (Endlich et al., 2014, S. 68). Zudem negiert die Studienlage ein „Herauswachsen“ der Lese-Rechtschreibstörung. So verfügen vier bis sechs Prozent der Erwachsenen in Deutschland über eine Lese- und Rechtschreibkompetenz, welche maximal auf dem Niveau der vierten Klasse liegt (Schulte-Körne & Remschmidt, 2003, S. A 397). Auch Klicpera et al. (2017, S. 145) führen an, dass zwar etwa ein Drittel bis die Hälfte der Kinder, die im Erstleseunterricht zunächst Schwierigkeiten haben, den Rückstand wieder aufholen kann, dass aber diejenigen, die in der zweiten Klasse immer noch Schwierigkeiten haben, diese nicht von alleine überwinden können. Als möglichen Grund nennen sie, dass diese Kinder aufgrund der Lese-schwierigkeiten häufig weniger lesen als Kinder ohne Schwierigkeiten und sich dies negativ auf das Leseverständnis, aber auch auf das „Wissen um die spezifische Schreibweise von Wörtern“ auswirkt (Klicpera et al., 2017, S. 145). Kinder mit isolierten Leseschwierigkeiten lesen außerdem zwar relativ fehlerfrei, dafür aber sehr langsam, was bis zum Ende der Schulzeit oft nicht überwunden wird. Isolierte Rechtschreibschwierigkeiten können dagegen von einem Drittel der Kinder bis zum Ende der Schulzeit überwunden werden (Klicpera et al., 2017, S. 146). Ein weiteres Drittel kann dies nicht und das letzte Drittel fällt zudem im Verlauf auch im Lesen zurück, sodass sich in der Folge sowohl Lese- wie auch Rechtschreibschwierigkeiten finden. Insgesamt halten Klicpera et al. (2017, S. 145) fest, dass die schlechte Prognose vor allem dann besteht, wenn Schwierigkeiten im Lesen vorhanden sind oder wenig gelesen wird.

Kohn, Wyschkon, Ballaschk, Ihle und Esser (2013, S. 78) halten fest, dass die hohe Stabilität von Lese-Rechtschreibstörungen in Studien bereits aufgezeigt wurde. In ihrer eigenen Studie lag der Fokus auf der Frage, wie der Verlauf von umschriebenen Entwicklungsstörungen in einer Zeitspanne von zweieinhalb Jahren aussieht, aber auch, wie viele Kinder sich in diesem Zeitraum so in ihrer Leistung steigerten, dass die Diagnose einer umschriebenen Entwicklungsstörung nicht mehr vorlag. Ebenso betrachteten sie die Häufigkeit von Klassenwiederholung und den Besuch weiterführender Schulen (Kohn, Wyschkon, Ballaschk, et al., 2013, S. 79). Die Studie fand im Potsdamer Raum statt, wo die Grundschule nach der sechsten Klasse endet und erst dann eine weiterführende Schule besucht wird. Zu Beginn der Studie waren die teilnehmenden Kinder im Schnitt 8;9 Jahre alt und befanden sich in Klasse eins bis fünf. Zur Diagnose von umschriebenen Entwicklungsstörungen wurde das doppelte Diskrepanzkriterium mit je 1.5 Standardabweichungen angesetzt. Lernschwächen werden diagnostiziert, wenn in den Teilleistungen eine niedrigere Diskrepanz von mindestens einer Standardabweichung zum Normierungsmittel vorliegt oder das doppelte Diskrepanzkriterium nicht erfüllt wird. Die Störungen, die hier berücksichtigt wurden, beziehen sich auf Lesen, Schreiben, Rechnen sowie die expressive Sprache, wobei an dieser Stelle nur die Ergebnisse für den Bereich Lesen und Schreiben wiedergegeben werden. Minderleistungen im Lesen und

Schreiben wurden zunächst der gemeinsamen Gruppe „Schülerinnen und Schüler mit LRS“ zugeordnet (Kohn, Wyschkon, Ballaschk, et al., 2013, S. 80). Diese Gruppe war im Verlauf der zweieinhalb Jahre recht stabil in ihren Leistungen. Ungefähr drei Viertel der Kinder erfüllten nach dieser Zeit weiterhin die diagnostischen Kriterien einer Lese-Rechtschreibstörung. Die Kinder, die sich verbessert hatten, erzielten zum zweiten Messzeitpunkt mindestens durchschnittliche Leistungen im Lesen und im Schreiben (Kohn, Wyschkon, Ballaschk, et al., 2013, S. 82). Es erfolgte eine genauere Analyse, in welcher die Bereiche Lesen und Schreiben getrennt betrachtet wurden. Es zeigt sich, dass die Kinder, die sich verbesserten, zum ersten Messzeitpunkt eine bessere Leistung im Lesen erzielt hatten. Für die Rechtschreibung gab es zum ersten Messzeitpunkt keinen Unterschied zwischen dieser Gruppe und der Gruppe der Kinder, die sich nicht verbessert haben. Weiterhin zeigt sich, dass die isolierte Rechtschreibstörung weniger andauernd ist als die isolierte Lesestörung. Für die isolierte Rechtschreibstörung erfüllten nur 14 Prozent zum zweiten Messzeitpunkt weiterhin die Diagnosekriterien. 43 Prozent erzielten sogar durchschnittliche Leistungen. Für die isolierte Lesestörung zeigt sich dagegen, dass 30 Prozent weiterhin die Diagnosekriterien für eine Lesestörung erfüllten und nur 17 Prozent durchschnittliche Leistungen zeigten. Die Hälfte erfüllte die Kriterien einer Lese-Rechtschreibschwäche (Kohn, Wyschkon, Ballaschk, et al., 2013, S. 82). Auch Wyschkon et al. (2018, S. 114) konnten in ihrer Studie bei einem knappen Drittel der Kinder mit Lese-Rechtschreibstörung (N = 17) zum ersten Messzeitpunkt über einen Verlauf von 20 Monaten eine Verbesserung der Leistung in dem Sinne auffinden, dass die Kinder nach 20 Monaten eine mindestens durchschnittliche Lese-Rechtschreibleistung aufwiesen. Knapp die Hälfte der übrigen Kinder (N = 37) wies weiterhin eine Lese-Rechtschreibstörung auf. 20 Prozent hatten zu diesem Zeitpunkt eine Lese-Rechtschreibschwäche.

#### 6.4.3 Schullaufbahn und Berufserfolg

Da sich vor allem die Lesefähigkeit nicht nur auf die Leistung im Unterrichtsfach Deutsch, sondern auch auf andere Unterrichtsfächer auswirkt, ist es wichtig, einen Blick auf die Folgen für die Schullaufbahn und im weiteren Verlauf auch für den Berufserfolg zu werfen.

Bei Kindern mit einer umschriebenen Entwicklungsstörung kann es häufiger zu Klassenwiederholung kommen, wobei dies vor allem für die ersten beiden Klassenstufen zutrifft (Kohn, Wyschkon, Ballaschk, et al., 2013, S. 85). In der Studie von Kohn, Wyschkon, Ballaschk et al. (2013, S. 85) zeigten sich ab der vierten Klasse allerdings keine Unterschiede mehr. Es wird vermutet, dass dies der flexiblen Schuleingangsphase geschuldet ist, die vorsieht, dass der Wechsel in die dritte Klasse flexibel nach einem, zwei oder drei Schuljahren geschehen kann. Für den Übergang ins Gymnasium zeigten sich signifikante Unterschiede. Nur knapp ein Viertel der Kinder mit einer umschriebenen Entwicklungsstörung wechselte auf diesen Schulzweig, während es knapp 70 Prozent der Kinder ohne umschriebene Entwicklungsstörung waren (Kohn, Wyschkon, Ballaschk, et al., 2013, S. 85).

Die von Esser et al. (2002) durchgeführte Langzeitstudie erhob, wie bereits erwähnt, auch den Schulerfolg sowie den Berufsabschluss und das Selbstbild der lese-rechtschreibschwachen Studienteilnehmerinnen und -teilnehmer. Trotz vergleichbarer kognitiver Fähigkeiten liegt der Schulerfolg der Betroffenen, gemessen als der höchste erzielte Schulabschluss sowie der hier erreichten Durchschnittsnote, unter dem der Nicht-Betroffenen und der Personen mit anderen umschriebenen Entwicklungsstörungen (Esser et al., 2002, S. 238). Im direkten Vergleich hatten sie auch häufiger gar keinen Berufsabschluss und waren öfter arbeitslos als die anderen beiden Gruppen (Esser et al., 2002, S. 241).

Wyschkon et al. (2018, S. 120) konnten in ihrer Studie aufzeigen, dass Kinder mit Lese-Rechtschreibstörung schlechtere Deutschnoten hatten sowie eher Schultypen mit einem niedrigeren Bildungsniveau besuchten als Kinder mit durchschnittlicher Intelligenz, aber keiner Lese-Rechtschreibstörung. Hinsichtlich der Mathematiknote zeigten sich dagegen keine signifikanten Unterschiede zwischen den Gruppen.

Daniel et al. (2006, S. 508) gingen der Frage nach, ob Jugendliche mit einer Lesestörung im Vergleich zu Jugendlichen ohne Lesestörung eine erhöhte Schulabbruchrate und damit eventuell verbunden auch ein erhöhtes Selbstmordrisiko aufzeigen. Hierfür wurden 188 Jugendliche im Alter von 15 Jahren aus dem Südosten der USA untersucht (Daniel et al., 2006, S. 509). Die Suizidalität wurde über Interviews erhoben, wobei suizidale Verhaltensweisen im Sinne von Selbstmordgedanken und Selbstmordversuchen erfragt wurden (Daniel et al., 2006, S. 510). Jugendliche mit einer Lesestörung brachen, auch unter Berücksichtigung des sozioökonomischen Hintergrunds, signifikant häufiger die Schule ab und wiesen eine höhere Suizidalität auf als Jugendliche ohne Lesestörung. Es zeigte sich ebenfalls ein Zusammenhang zwischen der Schulabbruchrate und der Suizidalität (Daniel et al., 2006, S. 512).

Schulte-Körne, Deimel, Jungermann und Remschmidt (2003) führten eine Nachuntersuchung von ehemaligen Internatsschülerinnen und -schülern mit einer Lese-Rechtschreibstörung durch. Es handelt sich hierbei um ein Internat, welches als Legastheniezentrum die Förderung von Kindern mit Lese-Rechtschreibstörung im Fokus hat. In der Nachuntersuchung nach über 20 Jahren wurde neben der Rechtschreibleistung und der Intelligenz auch der Beruf der ehemaligen Schülerinnen und Schüler sowie der Beruf der Eltern erfasst. Ebenso hatten Schulte-Körne et al. Zugriff auf die Testleistungen der damaligen Kinder. Über die Leseentwicklung kann nichts gesagt werden, da für Erwachsene kein standardisierter Lesetest zur Verfügung stand (Schulte-Körne et al., 2003, S. 269). In der Nachuntersuchung erfüllten jedoch nur noch 38 Prozent der Teilnehmenden das Kriterium einer Rechtschreibstörung. Hinsichtlich des Schulabschlusses zeigte sich, dass diejenigen, die an der Studie teilnahmen, die Schule mit einem höheren Schulabschluss verlassen hatten als die ehemaligen Schülerinnen und Schüler, die an der Nachuntersuchung nicht mehr teilnahmen (Schulte-Körne et al., 2003, S. 270). 41 Prozent hatten Abitur gemacht und nur drei Prozent einen Hauptschulabschluss (Schulte-Körne et al., 2003, S. 273). In Bezug auf das Berufsprestige zeigte sich eine Überlegenheit der Stichprobe im Vergleich zur allgemeinen Bevölkerung.

Diese Überlegenheit fällt bei den Vätern der untersuchten Personen noch deutlicher aus, was allerdings auch zeigt, dass die Kinder das Niveau der Väter nicht erreichen konnten (Schulte-Körne et al., 2003, S. 272). Die Ergebnisse sind allerdings nur eingeschränkt valide, da die ehemaligen Schülerinnen und Schüler insgesamt überdurchschnittlich intelligent sind und auch die Schulform (ein Internat mit Fokus auf Förderung bei Lese-Rechtschreibstörungen) nicht repräsentativ für deutsche Schulen ist (Schulte-Körne et al., 2003, S. 274). Ebenso lagen Schulte-Körne et al. (2003, S. 273) keine Informationen über die Art der Förderung am Internat vor, sodass über deren Einfluss keine Aussagen gemacht werden konnten. Sie kommen aber zu dem Schluss, dass ihre Studie zeigt, dass eine allgemeingültige Vorhersage über den Verlauf nicht möglich ist, da dieser stark von den persönlichen wie auch schulischen Rahmenbedingungen abhängig sein kann (Schulte-Körne et al., 2003, S. 274).

## 6.5 Zwischenfazit

Die Begriffsbestimmung von Lese- und Rechtschreibschwierigkeiten gestaltet sich schwierig. Viele orientieren sich bei der Definition an der ICD-10 und dem DSM-5. Kontrovers diskutiert wird vor allem die Frage, ob das doppelte Diskrepanzkriterium berücksichtigt und welches Maß der Intelligenz verwendet werden soll. Weiterhin schwanken die Angaben der zur Diagnose von Lese-Rechtschreibstörungen und -schwächen verwendeten Standardabweichung, um die Minderleistung entsprechend festzustellen. Allen Definitionen gemein ist dagegen, dass von ausreichenden kognitiven Fähigkeiten zum Erwerb der Lesekompetenz ausgegangen werden muss und physische Ursachen, wie beispielsweise ein nicht ausreichendes Sehvermögen, ausgeschlossen werden müssen.

Das Intelligenz-Diskrepanzkriterium sollte in der Praxis für Lehrkräfte dagegen keine oder nur eine untergeordnete Rolle spielen, da alle betroffenen Kinder einem Leidensdruck unterliegen sind und gleichermaßen von Förderung profitieren können.

Abhängig von der verwendeten Definition ändern sich auch die Prävalenzen der Lernstörungen und -schwächen. So finden sich Angaben zur Häufigkeit zwischen 1 und 40 Prozent. Es wird daher stellenweise vorgeschlagen, kriteriumsorientierte Tests zu verwenden, um die Häufigkeit zu bestimmen.

Jungen sind häufiger von Lese-Rechtschreibschwierigkeiten betroffen, über das genaue Verhältnis von Jungen zu Mädchen herrscht dagegen keine Einigkeit. Es kann von einem 2:1-Verhältnis in unselektierten Stichproben ausgegangen werden.

Vor allem die Lesestörung, aber auch die Rechtschreibstörung, sind relativ stabil und betroffene Kinder und Jugendliche benötigen daher gezielt Förderung. Mit längerer Dauer und wenn kombinierte Lernschwierigkeiten vorliegen, erhöhen sich insbesondere die begleitenden psychosozialen Belastungen bzw. es können sich in der Folge solche entwickeln. Hier treten vor allem hyperkinetische Störungen, aber auch internalisierende Störungen vermehrt auf. Die Vergleichbarkeit der Studien ist allerdings erschwert, da sie im Detail unterschiedliche Definitionen für die Lernschwierigkeiten verwenden und andere Tests zur Erhebung von

schulischen Leistungen, der Intelligenz oder der psychischen Auffälligkeiten zum Einsatz kommen.

Betroffene erhalten auch häufiger niedrigere Schulabschlüsse und haben ein höheres Risiko, die Schule ohne Abschluss zu verlassen. Insgesamt ergeben sich so für die Betroffenen schlechtere Chancen auf dem Arbeitsmarkt.

Es zeigt sich deutlich die Relevanz von gezielter und intensiver, früh ansetzender Förderung, um dieser negativen Gesamtentwicklung entgegenzuwirken.

## 7 Voraussetzungen und Vorläuferfertigkeiten des Leseerwerbs

Die oben genannten Modelle setzen zum Teil beim Beginn des Leseerwerbs in der Grundschule an. Wie anhand dieser Modelle dennoch ersichtlich wurde, beginnt das Lesen lernen nicht erst mit Eintritt in die Grundschule (Marx, 2007, S. 105). Vielmehr müssen zuvor zahlreiche Entwicklungsschritte von den Kindern durchlaufen werden, wie zum Beispiel die grundlegende Erkenntnis, dass Bilder und Symbole Bedeutung enthalten können (Günther, 1986, S. 34–35). Aber auch spezifischere Fähigkeiten entwickeln sich in der Zeit vor dem Grundschuleintritt. So erkennen Kinder langsam die Bedeutung einzelner Bestandteile eines Wortes und können auch Reime bilden (Hose und Rose zum Beispiel).

Brem und Maurer (2015, S. 131) führen als relevante Vorläuferfertigkeiten die Buchstabenkenntnis, die phonologische Bewusstheit und die Benennungsgeschwindigkeit an. Klicpera et al. (2017, S. 25) nennen als Vorläuferfertigkeit zwar ebenfalls die Buchstabenkenntnis, halten aber fest, dass für Österreich und Deutschland gezeigt werden konnte, dass Kinder vor Eintritt in die Schule nur sehr wenige Buchstaben nennen konnten. Diese Fertigkeit wird daher zumindest in diesen Ländern eher dem Schriftspracherwerb in der Grundschule zugerechnet. Yamada et al. (2011) hatten in ihrer Studie Vorläuferfertigkeiten bei Kindern im Vorschulalter im Blick. Betrachtet wurden die Geschwindigkeiten, mit der jeweils Buchstaben sowie Anlaute genannt werden können (Yamada et al., 2011, S. 705). Die Hälfte der vierzehn an der Studie teilnehmenden Kinder wurde als „at risk“ in Bezug auf die Entwicklung von Leseschwierigkeiten identifiziert (unterhalb des 35. Perzentils in einer der beiden Vorläuferfertigkeiten), während die andere Hälfte „on track“ war (zwischen dem 50. und dem 75. Perzentil). Die Studie konnte zeigen, dass Kinder im Vorschulalter, die über mindestens ausreichend entwickelte Vorläuferfertigkeiten verfügen, auch eine höhere Hirnaktivität in den für das Lesen relevanten phonologischen Arealen aufweisen als Kinder mit niedrigen Vorläuferfertigkeiten. Zusätzliche Leseförderung der „at risk“-Kinder führte dazu, dass diese die „on track“-Kinder zum Teil einholen konnten (Yamada et al., 2011, S. 711). Dies zeigt, dass Vorläuferfertigkeiten eine wichtige Rolle beim Leseerwerb, aber auch beim Erkennen von eventuellen Schwierigkeiten zukommen (Yamada et al., 2011, S. 712).

Im Folgenden werden die zwei häufig thematisierten und für den Schriftspracherwerb relevanten Vorläuferfertigkeiten, die phonologische Bewusstheit sowie die Benennungsgeschwindigkeit, noch einmal näher beleuchtet und ihre Bedeutung für den Leseerwerb im deutschsprachigen Raum herausgearbeitet. Die Buchstabenkenntnis wird als Einflussfaktor erst im nächsten Kapitel behandelt werden, da diese im deutschsprachigen Raum für den vorschulischen Bereich noch keine Rolle spielt.

### 7.1 Phonologische Bewusstheit

Skrowonek und Marx (1989) definierten für ihr Screening zur Früherkennung von Lese-Rechtsschreibschwierigkeiten die phonologische Bewusstheit im weiteren und im engeren Sinne wie folgt: Phonologische Bewusstheit im weiteren Sinne bezeichnet die Fähigkeit, Aufgaben

lösen zu können, die das „Umgehen mit den lautlichen Aspekten der Sprechsprache verlangen“ (Skrowonek & Marx, 1989, S. 42). Hierzu zählen sie „Reimpaare finden und Silben segmentieren“ (Wörter in ihre Silben zerlegen können) und halten fest, dass es sich hierbei um Aufgaben handelt, die in „konkreten, dem Kind vertrauten Spielhandlungen enthalten sind“ (Skrowonek & Marx, 1989, S. 42). Zur phonologischen Bewusstheit im engeren Sinne zählen sie die Fähigkeit, Sprache in ihre „lautlichen Strukturen“ zu zerlegen, wobei diese „weder semantische noch sprechrhythmische Bezüge aufweisen“ (Skrowonek & Marx, 1989, S. 42). Hierzu gehören Lautvergleiche - „Hörst du ein /i/ in ‚Igel‘? Hörst du ein /i/ in ‚Auto‘?“ - und das Zusammenfügen von Lauten zu Worten - „Zange als /ts/-/ange/“ (Skrowonek & Marx, 1989, S. 42–43). Schröder-Lenzen (2013, S. 88) bezeichnet die phonologische Bewusstheit im weiteren Sinne als die Fähigkeit, „Reime zu erkennen, Silben zu segmentieren und zusammensetzen“. Unter der phonologischen Bewusstheit im engeren Sinne versteht sie die Fähigkeit, „das gesamte Wort auf seine lautlichen Bestandteile hin abhören zu können“ (Schröder-Lenzen, 2013, S. 88). Wolf, Schroeders und Kriegbaum (2016, S. 9) beschreiben die phonologische Bewusstheit im weiteren Sinne als die Fähigkeit, größere sprachliche Einheiten (Wörter, Silben und Reime) wahrnehmen und analysieren zu können. Die phonologische Bewusstheit im engeren Sinne bezieht sich im Gegensatz dazu auf die kleinsten sprachlichen Einheiten, also auf die Phoneme (Wolf et al., 2016, S. 9).

Schröder-Lenzen (2013, S. 88) merkt hierzu an, dass „im Detail“ keine Einigkeit darüber herrscht, was unter phonologischer Bewusstheit verstanden wird, dass aber im pädagogischen Bereich zwischen der phonologischen Bewusstheit im weiteren und im engeren Sinne unterschieden wird.

Die Art und Höhe des Einflusses der phonologischen Bewusstheit unterscheidet sich mutmaßlich für die verschiedenen Bereiche des Schriftspracherwerbs und auch für unterschiedliche Sprachen. So spielt die phonologische Bewusstheit insgesamt im englischsprachigen Raum eine größere Rolle als im deutschsprachigen (Fricke, Szczerbinski, Fox-Boyer & Stackhouse, 2015, S. 45). Weiterhin beeinflusst sie sowohl den Erwerb des Schreibens wie auch den des Lesens, hängt allerdings im deutschsprachigen Raum stärker mit dem Schreiben als mit dem Lesen zusammen (Fricke et al., 2015, S. 31).

Gorecki und Landerl (2015, S. 140) kritisieren, dass bei Studien, welche den Einfluss der phonologischen Bewusstheit auf den Leseerwerb untersuchen wollten, sehr häufig die Lesefähigkeit zu Beginn nicht erhoben oder nicht berücksichtigt wurde. Die beiden Autorinnen führen in diesem Zuge mehrere Studien an, die darauf hinweisen, dass im deutschsprachigen Raum die phonologische Bewusstheit bezogen auf das Lesen nur eine geringe Vorhersagekraft für das erste Schuljahr zu haben scheint, aber nicht darüber hinaus (Gorecki & Landerl, 2015, S. 141). Sie untersuchten in einer eigenen Studie, welchen Einfluss die phonologische Bewusstheit auf die Lesefähigkeit von deutschsprachigen Grundschulkindern hat. Der Studie liegen die Daten von 343 Kindern zugrunde, welche einmal zu Beginn der ersten Klasse und ein zweites Mal zum Ende der ersten Klasse getestet wurden. Erhoben wurden die Intelligenz

sowie die phonologische Bewusstheit und die Lesefähigkeit. Die phonologische Bewusstheit wurde über 24 Wörter und 24 Pseudowörter erfasst, wobei die Kinder aufgefordert wurden, Silben und einzelne Phoneme auszulassen, zum Beispiel /mast/ ohne /m/ auszusprechen (Gorecki & Landerl, 2015, S. 141–142). Ausgewertet wurden die Daten mithilfe von Strukturgleichungsmodellen. Die Autorinnen haben einmal die Vorhersagekraft der phonologischen Bewusstheit auf die spätere Leseleistung ohne Kontrolle der Leseleistung zu Beginn der ersten Klasse und einmal mit Kontrolle dieser Leseleistung betrachtet. Unter Kontrolle der frühen Leseleistung zeigte sich keine Vorhersagekraft mehr für die phonologische Bewusstheit auf die Leseleistung zum Ende der ersten Klasse. Weiterhin zeigte sich auch keine Vorhersagekraft der Leseleistung zu Beginn der ersten Klasse auf die phonologische Bewusstheit zum Ende der ersten Klasse. Die Autorinnen schlussfolgern aus ihren Ergebnissen, dass sich die phonologische Bewusstheit und die Leseleistung parallel entwickeln. Kinder, deren phonologische Bewusstheit stärker ausgeprägt ist, verfügen auch über eine bessere Leseleistung. Gorecki und Landerl gehen davon aus, dass diese Kinder dementsprechend keine Schwierigkeiten beim Leseerwerb zeigen werden. Umgekehrt ließ sich aber nicht feststellen, dass Kinder mit einer niedrigen phonologischen Bewusstheit Risikokinder beim Erwerb des Lesens waren, sodass sich eine niedrige phonologische Bewusstheit nicht zur Vorhersage von Schwierigkeiten im Lesen eignet (Gorecki & Landerl, 2015, S. 144).

Pfost hält fest, dass mehrere Metaanalysen die Rolle der phonologischen Bewusstheit als Prädiktor für die Lesekompetenz aufzeigen konnten (Pfost, 2015, S. 125). Da sich diese Studien aber nur auf den englischsprachigen Raum stützen und sich die englische und die deutsche Sprache im Hinblick auf die Phonem-Graphem- sowie die Graphem-Phonem-Korrespondenz jeweils voneinander unterscheiden, kann nicht direkt von einer Übertragbarkeit auf den deutschsprachigen Raum ausgegangen werden (Pfost, 2015, S. 126). Daher führte Pfost (2015) eine Metaanalyse zur phonologischen Bewusstheit im deutschsprachigen Raum durch, in die 19 Arbeiten über insgesamt 21 Studien aus den Jahren 1985 bis 2013 eingegangen sind. Da die Übertragbarkeit der Ergebnisse aus dem englischsprachigen Raum angezweifelt wird, bezog die Metaanalyse von Pfost entsprechend nur Studien über den möglichen Einfluss der phonologischen Bewusstheit im deutschsprachigen Raum ein. Weiterhin wurden nur Studien mit mindestens zwei Messzeitpunkten berücksichtigt, in denen die phonologische Bewusstheit zeitlich früher als die Lesefähigkeit erhoben wurde und zwar entweder im Kindergarten oder in der ersten Grundschulklasse. Die Lesefertigkeit musste nach Beginn des Leseunterrichts in der Grundschule erfasst werden und die Daten mussten in Deutschland, Österreich oder der deutschsprachigen Schweiz und ausschließlich mit deutschen Tests erhoben worden sein (Pfost, 2015, S. 127). Insgesamt zeigte sich, dass ein Zehntel der Varianz in den späteren Lese- und Rechtschreibleistungen durch die phonologische Bewusstheit, welche zeitlich früher erhoben wurde, vorausgesagt wird. Pfost hält fest, dass der gefundene Zusammenhang zwischen phonologischer Bewusstheit und der späteren Leseleistung gering ist, da ein Großteil der Varianz unaufgeklärt bleibt (Pfost, 2015, S. 134).

Die von Gorecki und Landerl (2015) angebrachte Kritik, dass die Leseleistung zum Zeitpunkt der Erhebung der phonologischen Bewusstheit nicht kontrolliert wurde, ist auch hier relevant, da Pfost dies in seiner Metaanalyse nicht berücksichtigt hat. Insgesamt sollte die phonologische Bewusstheit allerdings kritisch betrachtet werden, da die deutsche Sprache eine relativ transparente Orthographie aufweist und die phonologische Bewusstheit eine andere Rolle zu spielen scheint mit einer weniger klaren Relevanz für den Leseerwerb, als dies im englischsprachigen Raum der Fall ist.

Wolf et al. (2016, S. 11) führen an, dass im internationalen Vergleich auch die Unterrichtsgestaltung einbezogen werden muss, wenn es um den Effekt der phonologischen Bewusstheit geht. Dadurch, dass die deutsche Schriftsprache eine moderate Transparenz zeigt, wird häufig mit phonetischen bzw. synthetischen Methoden unterrichtet, während dies z. B. im englischsprachigen Raum aufgrund der Inkonsistenz nicht möglich ist. Somit ist es schwierig, die Effektivität einer gezielten Förderung der phonologischen Bewusstheit zu erfassen, da auch die Kontrollgruppen durch den Unterricht an sich ebenfalls eine Förderung in diesem Bereich erhalten. In ihrer Metaanalyse sind die Autorinnen und Autoren dennoch der Frage nachgegangen, welche Effekte sich für die Förderung der phonologischen Bewusstheit auf die phonologische Bewusstheit selbst sowie auf die Entwicklung des Lesens und Schreibens der Kinder finden lassen (Wolf et al., 2016, S. 13). In die Metaanalyse gingen 31 Publikationen mit 27 Primärstudien ein. Voraussetzung war, dass es sich um Trainingsstudien mit mindestens zwei Messzeitpunkten (Prä- und Post-Testung) handelte, welche auf standardisierten Leistungstests aufbauten, eine Kontrollgruppe ohne entsprechendes Training enthielten und in denen Effektstärken entweder berichtet wurden oder sich diese aufgrund der angegebenen Daten berechnen ließen (Wolf et al., 2016, S. 14). Es wurden mehrere Moderatoranalysen vorgenommen, um festzustellen, ob und in welchem Ausmaß verschiedene Merkmale, wie beispielsweise der Zeitpunkt der Förderung, Inhalte der Förderung, die Qualität der Implementation oder Merkmale der Kinder selbst die Förderung der phonologischen Bewusstheit bzw. deren Wirkung auf den Schriftspracherwerb beeinflussen. Weiterhin wurde getrennt zwischen einem Einfluss auf den Schreiberwerb, einem Einfluss auf den Leseerwerb und einem Einfluss auf die Entwicklung der phonologischen Bewusstheit selbst (Wolf et al., 2016, S. 13). Im Ergebnis zeigte sich, dass vorschulische Förderungen der phonologischen Fertigkeiten keinen statistisch signifikanten Einfluss auf den Erwerb der Lesekompetenz hatten. Für den Schreiberwerb zeigten sich geringe Effekte (Wolf et al., 2016, S. 25). Dieses Bild präsentierte sich auch bei schulischer Förderung der phonologischen Bewusstheit. Es gab weder für das Leseverständnis noch für die Dekodierfähigkeit signifikante Effekte. Wolf et al. (2016, S. 27) ziehen einen Vergleich zu internationalen Metaanalysen und halten fest, dass im deutschsprachigen Raum eine Förderung der phonologischen Bewusstheit weniger effektiv ist als für andere Sprachregionen.

Fricke et al. (2015) haben in ihrer Studie die Vorhersagekraft der Benennungsgeschwindigkeit, der Buchstabenkenntnis, der phonologischen Bewusstheit sowie der sprachlichen

Fähigkeiten auf die Lese-Rechtschreibfertigkeiten bei 78 Kindern vor Eintritt in die Grundschule sowie ein und zwei Jahre später geprüft (Fricke et al., 2015, S. 34). Analog zu den genannten Forschungsergebnissen zeigte sich auch hier, dass die phonologische Bewusstheit ein Prädiktor für die Rechtschreibleistung war. Zusätzlich stellte sich heraus, dass die phonologische Bewusstheit nicht die Leseflüssigkeit, wohl aber die Lesegenauigkeit vorherzusagen konnte (Fricke et al., 2015, S. 45). Insgesamt konnten sie in ihrer Studie zeigen, dass die phonologische Bewusstheit eine Rolle spielt, auch wenn diese im deutschsprachigen Raum geringer ausfällt als für Sprachen mit geringerer Transparenz (Fricke et al., 2015, S. 46). Die Ergebnisse zur Benennungsgeschwindigkeit werden im nächsten Abschnitt vorgestellt.

Zusammenfassend lässt sich aus der bisherigen Forschungslage schlussfolgern, dass die phonologische Bewusstheit ein Prädiktor für die spätere Lesekompetenz bzw. einzelne Teilkompetenzen sein kann, dieser Einfluss aber in der deutschen Sprache weit weniger stark ausgeprägt ist als in anderen Sprachen wie beispielsweise Englisch.

## 7.2 Benennungsgeschwindigkeit

Die zweite mit dem Schriftspracherwerb zusammenhängende Fertigkeit ist die Benennungsgeschwindigkeit. Als Benennungsgeschwindigkeit bezeichnet man die Fähigkeit, Objekte, Bilder oder auch Symbole schnell benennen zu können. Sie erweist sich als Prädiktor für die spätere Lesefähigkeit bei Kindern im Vorschulalter (Mayer, 2016, S. 99). Phonologische Bewusstheit und Benennungsgeschwindigkeit sind weitgehend voneinander unabhängig und zeigen nur eine geringe Korrelation miteinander (Klicpera et al., 2017, S. 175; Moll, Wallner & Landerl, 2012, S. 11; Vander Stappen & Reybroeck, 2018, S. 10).

Denckla und Rudel (1976) konnten die Relevanz der Benennungsgeschwindigkeit für die Lesefertigkeit bereits in den 1970ern aufzeigen. In einer ihrer Studien untersuchten sie die Benennungsgeschwindigkeit von geläufigen Farben, Objekten, Zahlen und Buchstaben bei Kindern und Jugendlichen im Alter von sieben bis 13 Jahren. Sie bildeten hierzu drei Gruppen: eine Experimentalgruppe bestehend aus Kindern mit Lernstörungen, eine zweite Experimentalgruppe bestehend aus Kindern mit einer Lesestörung sowie eine Kontrollgruppe, die aus Kindern mit durchschnittlichen Leseleistungen bestand. Die Lesestörung wurde über die Dekodierfähigkeit erhoben (Denckla & Rudel, 1976, S. 472). Die Benennungsgeschwindigkeit wurde mit vier Arten an Items – Farben, Buchstaben, Objekte und Zahlen – überprüft. Fünf verschiedene Varianten einer Item-Art wurden jeweils erstellt. Auf einer Tafel wurden diese in fünf Reihen zu je zehn Items zufällig angeordnet, sodass auf jeder Tafel insgesamt 50 Items einer Art präsentiert wurden. Diese 50 Items sollten der Reihe nach so schnell wie möglich genannt werden. Die Zeit wurde dabei gestoppt. Fehler wurden in allen drei Gruppen nur sehr selten gemacht (Denckla & Rudel, 1976, S. 473). Die Kontrollgruppe konnte die Begriffe allerdings signifikant schneller nennen als die Gruppe der Kinder mit Lernstörungen. Beide Gruppen waren signifikant schneller als die Gruppe der Kinder mit Lesestörung. Die

Unterschiede bestanden jeweils unabhängig vom Alter (Denckla & Rudel, 1976, S. 474). Für alle drei Gruppen zeigte sich außerdem, dass das Benennen von Objekten am meisten Zeit brauchte, gefolgt vom Benennen von Farben. Das Benennen von Zahlen und Buchstaben gelang am schnellsten (Denckla & Rudel, 1976, S. 475).

Moll et al. (2012) konnten in ihrer Studie mit Kindern der zweiten bis vierten Klasse eine Vorhersageleistung der Benennungsgeschwindigkeit auf die Leseflüssigkeit von Wörtern und Pseudowörtern aufzeigen. Analog zu den oben genannten Forschungsergebnissen zur phonologischen Bewusstheit ergab sich diese hier als Prädiktor für die Rechtschreibleistung (Moll et al., 2012, S. 11). Umgekehrt waren die Vorhersageleistung der Benennungsgeschwindigkeit für die Rechtschreibleistung sowie die Vorhersageleistung der phonologischen Bewusstheit für die Leseleistung sehr gering (Moll et al., 2012, S. 12). Abschließend führen Moll et al. (2012, S. 13) noch an, dass diese Ergebnisse auch in „sprachvergleichenden Studien“ gezeigt werden konnten und somit nicht nur für den deutschsprachigen Raum gültig sind. Die Benennungsgeschwindigkeit hat sich demnach in verschiedenen Sprachen als Prädiktor für Lesegeschwindigkeit herausgestellt. Zum Leseverständnis gibt es dagegen keine so deutlichen Befunde (Fricke et al., 2015, S. 31). Daher haben Fricke et al. (2015) in ihrer Studie die Vorhersagekraft der Benennungsgeschwindigkeit sowohl für die Lesegeschwindigkeit und -genauigkeit wie auch für das Leseverständnis betrachtet. Die Benennungsgeschwindigkeit erwies sich als sehr relevant für alle überprüften Lesefertigkeiten und dies unabhängig von der ebenfalls überprüften phonologischen Bewusstheit (Fricke et al., 2015, S. 46). Auch für die Rechtschreibfähigkeiten ist die Benennungsgeschwindigkeit ein Prädiktor gewesen, allerdings mit niedrigerer Vorhersagekraft als die im vorherigen Abschnitt thematisierte phonologische Bewusstheit. Einschränkend halten Fricke et al. (2015, S. 46) fest, dass viele Aufgaben zur Überprüfung der Lesefähigkeiten Speed-Tests waren und sich hierdurch der eigentliche Zusammenhang verstärkt haben kann.

Trotz der zweifelsohne hohen Relevanz dieser Fähigkeit galt die Benennungsgeschwindigkeit bisher, anders als die phonologische Bewusstheit, als nicht trainierbar (Mayer, 2016, S. 112), sodass dieser daher eher eine Rolle als Indikator für mögliche Schwierigkeiten im Leseerwerb zugekommen ist. Vander Stappen und Reybroeck (2018) überprüften in ihrer Studie unter anderem die Trainierbarkeit der Benennungsgeschwindigkeit. Zu ihren Fragestellungen zählten zusätzlich, ob sich eine Wirksamkeit für das Training der phonologischen Bewusstheit zeigt und welchen Einfluss die beiden Fertigkeiten auf den Schriftspracherwerb haben (Vander Stappen & Reybroeck, 2018, S. 4). Die Studie fand in Belgien statt mit 36 französischsprachigen Kindern der zweiten Klasse. Eine Hälfte erhielt ein Training der phonologischen Bewusstheit, während die andere Hälfte ein Training der Benennungsgeschwindigkeit bekam. Die Trainings fanden jeweils zwei Monate lang statt. Es fanden Testungen vor dem Training, direkt danach und in einem Follow-Up sechs Monate später statt (Vander Stappen & Reybroeck, 2018, S. 5). Die Benennungsgeschwindigkeit wurde trainiert, indem Bilder,

welche Nomen der häufigsten Wörter für Kinder in der ersten Klasse zeigten, so schnell wie möglich zu benennen waren. Dies wurde in unterschiedlichen und auch spielerischen Varianten umgesetzt. Die Kinder wurden auch dazu ermutigt, ihren eigenen Geschwindigkeitsrekord zu brechen. Die im Training verwendeten Items waren nicht identisch mit den Items, die in den Tests verwendet wurden (Vander Stappen & Reybroeck, 2018, S. 12).

Beide Gruppen zeigten nach dem Training der anderen Gruppe gegenüber eine Überlegenheit in dem Bereich, in dem sie zuvor gefördert worden waren. Die Gruppe, die die Förderung der phonologischen Bewusstheit erhalten hatte, zeigte außerdem Transfereffekte im Bereich der Rechtschreibung, aber keinen Effekt auf die Lesegenauigkeit oder die Lesegeschwindigkeit (Vander Stappen & Reybroeck, 2018, S. 11). In der Gruppe, die in der Benennungsgeschwindigkeit trainiert worden war, zeigten sich keine Transfereffekte auf die Rechtschreibung, wohl aber auf die Wortlesegeschwindigkeit. In bisherigen Studien konnten dagegen keine Trainingseffekte der Förderung von Benennungsgeschwindigkeit gefunden werden (Vander Stappen & Reybroeck, 2018, S. 12). Die Autorinnen merken an, dass dies an der Art des Trainings gelegen haben mag, da sie sich auf die serielle Benennungsgeschwindigkeit konzentriert hatten, oder daran, dass das Training in anderen Studien mit weiteren Übungen kombiniert wurde. Sie ziehen aus den Ergebnissen ihrer Studie den Schluss, dass die Benennungsgeschwindigkeit eine wichtige Rolle bei der Automatisierung des Leseprozesses spielen könnte (Vander Stappen & Reybroeck, 2018, S. 12). Passend zu den Ergebnissen waren auch Georgiou, Parrila, Cui und Papadopoulos (2013, S. 222) zu dem Schluss gekommen, dass die serielle Benennungsgeschwindigkeit mit dem Lesen im Zusammenhang steht. Serielle Benennungsgeschwindigkeit bezeichnet das Nennen von Objekten nacheinander in einer vorgegebenen Reihenfolge. Dies entspricht der Vorgehensweise bei Denckla und Rudel (1976), bei der zum Beispiel fünf Items (in dieser Studie waren das: Ball, Katze, Baum, Huhn und Apfel) gezeigt wurden, randomisiert präsentiert in fünf Reihen mit je zehn dieser Items (Georgiou et al., 2013, S. 220). Die diskrete Benennungsgeschwindigkeit wird überprüft, indem nacheinander jeweils nur ein Objekt gezeigt wird, welches benannt werden soll. In einer dritten Methode wurde die Benennungsgeschwindigkeit darüber getestet, dass ein bestimmtes Zielitem in einer Reihe von mehreren Items durchgestrichen werden sollte (Georgiou et al., 2013, S. 220). Die vierte Methode verlangte, dass die Kinder dann, wenn ein bestimmtes Zielitem, zum Beispiel die Katze, gezeigt wurde, mit Ja antworten sollten und ansonsten mit Nein. In ihrer Studie stand nur die serielle Benennungsgeschwindigkeit über alle getesteten Klassenstufen hinweg mit der Leseflüssigkeit im Zusammenhang (Georgiou et al., 2013, S. 222). Sie vermuten, dass nicht nur das Erkennen von visuellen Stimuli, sondern auch die mündliche Produktion dem Zusammenhang der Benennungsgeschwindigkeit mit dem Lesen zugrunde liegt (Georgiou et al., 2013, S. 224).

### 7.3 Zwischenfazit

Damit Kinder Lesen lernen können, müssen bestimmte Vorläuferfertigkeiten in ausreichendem Maß vorhanden sein. Während sich im deutschsprachigen Raum für die phonologische Bewusstheit zeigt, dass diese stärker mit dem Erwerb des Schreibens zusammenhängt und einen größeren Prädiktor darstellt als für den Leseerwerb, zeigt sich ein umgekehrtes Bild für die Benennungsgeschwindigkeit. Die Beziehung zwischen dieser und dem Leseverständnis ist bisher zwar seltener untersucht worden, es deutet sich aber ein starker Zusammenhang an. Für die Lesegeschwindigkeit existieren dagegen mehr Studien, welche ebenfalls einen starken Zusammenhang aufzeigen. Auch wenn bisher oft davon ausgegangen war, dass die Benennungsgeschwindigkeit nicht trainierbar ist, konnten Vander Stappen und Reybroeck (2018) zeigen, dass zumindest die serielle Benennungsgeschwindigkeit gefördert werden kann, was sich in der Folge auch auf die Lesefähigkeiten positiv auswirken kann.

In diesem Abschnitt wurden zwei Vorläuferfertigkeiten thematisiert. Wie am Beispiel der Buchstabenkenntnis deutlich wird, sind dies nicht die einzigen Fertigkeiten, die eine Rolle spielen. Neben den Vorläuferfertigkeiten gibt es im weiteren Entwicklungsverlauf Einflussfaktoren, die den Leseerwerb begünstigen oder auch erschweren können. Klicpera et al. halten fest, dass mehrere Teilfertigkeiten am Prozess des Schriftspracherwerbs beteiligt sind, welche „ineinander greifen und sich gegenseitig verstärken“ (Klicpera et al., 2017, S. 150). Daher ist es notwendig, diese Teilfertigkeiten im Detail zu betrachten, um erkennen zu können, in welchen Bereichen Schülerinnen und Schüler eventuell Schwierigkeiten haben, damit eine Förderung möglichst genau an diesen Schwierigkeiten ansetzen kann. Deswegen werden im folgenden Kapitel weitere Fertigkeiten als mögliche Einflussfaktoren auf den Leseerwerb betrachtet.

## 8 Ursachen und Risikofaktoren von Leseschwierigkeiten

Lesen ist ein komplexer Vorgang, bei dem verschiedene kognitive Prozesse eine Rolle spielen. Daher gibt es auch nicht *den einen* Grund, der für eine Leseschwäche oder eine Lese­störung verantwortlich gemacht werden kann, sondern ein ebenso komplexes Gebilde aus verschiedenen Einfluss- und Risikofaktoren, die zu Schwierigkeiten im Leseerwerb führen bzw. die Entstehung einer Lese-Rechtschreibstörung begünstigen können (Klicpera et al., 2017, S. 178; Krinzinger & Günther, 2013, S. 42; Müller, Richter, Križan, Hecht & Ennemoser, 2015, S. 55; Valtin et al., 2010, S. 48). Klicpera et al. (2017, S. 179) gehen von einem „interaktiven Modell der Entwicklung von Lese- und Rechtschreibschwierigkeiten“ aus, wobei die Faktoren sich auch gegenseitig beeinflussen können. Sie unterteilen diese in individuelle Lernvoraussetzungen, Faktoren der familiären Interaktion sowie Faktoren in der Unter­richtung und in der Schule. Steinbrink und Lachmann (2014, S. 88) führen an, dass noch nicht eindeutig geklärt ist, welche Ursachen genau eine Lese-Rechtschreibstörung bedingen. Nicht bei jeder schwachen Leserin bzw. schwachem Leser zeigen sich alle Risikofaktoren und das Auftreten eines Faktors bedeutet umgekehrt auch nicht, dass eine Leseschwäche oder eine Lese­störung unbedingt auftreten muss (Klicpera et al., 2017, S. 178).

Um zu erfahren, wie gute Leserinnen und Leser vorgehen, wenn sie Informationen aus Texten entnehmen, wird häufig auf erfahrene Erwachsene zurückgegriffen, die z. B. mit der Methode des lauten Denkens einen Einblick in ihr Vorgehen geben können. Ebenso können auch neuropsychologische Vorgänge betrachtet werden. Boardman, Klingner, Buckley, Annamma und Lasser (2015, S. 1261) geben in ihrem Artikel einen kurzen Forschungsüberblick. Die Ergebnisse zeigen, dass gute Leserinnen und Leser keine Schwierigkeiten beim Dekodieren oder der Leseflüssigkeit zeigen, über einen größeren Wortschatz verfügen und häufiger sowie auch effizienter Gebrauch von Lesestrategien machen als weniger gute Leserinnen und Leser.

Im Folgenden werden entsprechende Risiko- und Einflussfaktoren erörtert. Es werden zunächst individuelle Faktoren näher betrachtet. Darauf folgen Faktoren im Umfeld, bevor abschließend Faktoren berücksichtigt werden, die ihrerseits als Teilprozesse auch am Lese­prozess selbst beteiligt sind.

### 8.1 Emotionale, behaviorale und kognitive Faktoren

Wie im interaktiven Modell der Entwicklung von Lese- und Rechtschreibschwierigkeiten bereits angeführt, können individuelle Faktoren die Entwicklung von Leseschwierigkeiten verursachen oder begünstigen. Diese können in kognitive, emotionale und behaviorale Faktoren unterteilt werden, welche in diesem Abschnitt betrachtet werden (Klicpera et al., 2017, S. 179).

### 8.1.1 Neurobiologische und genetische Faktoren

Es zeigt sich ein familiär erhöhtes Risiko für Lese-Rechtschreibschwierigkeiten. Kinder mit einer Lese-Rechtschreibstörung haben somit häufiger Elternteile oder Geschwister, die ebenfalls eine Lese-Rechtschreibstörung aufweisen (Lindberg, 2016, S. 53–54). Insgesamt liegt ein um „drei- bis fünffach erhöhtes Risiko“ für die Entwicklung einer Lesestörung vor, wenn Geschwister oder Eltern ebenfalls eine Lesestörung haben (Nagler, Lindberg & Hasselhorn, 2018, S. 37).

Allerdings gibt es keine für das Lesen spezifischen Hirnregionen und es lassen sich auch keine spezifischen Gene finden, die für das Lesen oder das Schreiben verantwortlich sind (Steinbrink & Lachmann, 2014, S. 2). Es finden sich dagegen Gene, die für die Sprache bzw. für kognitive Grundlagen zuständig sind (Steinbrink & Lachmann, 2014, S. 90). Steinbrink und Lachmann (2014) sowie Klicpera et al. (2017) geben hierzu jeweils einen Forschungsüberblick. Beide Quellen führen Studien an, die den bereits genannten genetischen Einfluss mithilfe von Zwillingsstudien nachweisen konnten. Beispielhaft sei hier die Studie von DeFries und Alarcón (1996) hervorgehoben, die über Ergebnisse aus der „Colorado Twin Study of Reading Disability“ berichtet. In dieser Studie wurden 324 mono- und dizygoten Zwillingspaare, von denen mindestens einer bzw. eine Schwierigkeiten im Lesen zeigte, mit Zwillingspaaren verglichen, die keine Schwierigkeiten im Lesen zeigten, sowie mit Kindern, die keinen Zwillingspartner hatten, aber ebenso Leseschwierigkeiten aufwiesen (DeFries & Alarcón, 1996, S. 41–42). Sie kamen zu dem Schluss, dass etwa die Hälfte der auftretenden Leseschwierigkeiten auf genetische Einflüsse zurückzuführen ist. Des Weiteren vermuten sie aufgrund ihrer Ergebnisse eine mögliche Ursache auf Chromosom 6 (DeFries & Alarcón, 1996, S. 46). Auch Klicpera et al. (2017, S. 181) fassen zusammen, dass der momentane Forschungsstand mögliche ursächliche Gene auf den Chromosomen 1, 2, 3, 6, 15, 18 und X vermuten lässt. Es ist somit kein einzelnes Gen ursächlich für eine Lesestörung, sondern es werden mindestens neun Kandidatenregionen und 14 Kandidatengene vermutet (Carrion-Castillo, Franke & Fisher, 2013, S. 216; Nagler et al., 2018, S. 37; Scerri & Schulte-Körne, 2010, S. 192). Demnach gibt es mittlerweile mehrere etablierte Theorien sowie Einigkeit darüber, dass „die Lese-Rechtschreibstörung durch eine genetisch bedingte Prädisposition für eine Störung in der Entwicklung neurobiologischer Strukturen und damit verbundener Prozesse begründet ist“ (Steinbrink & Lachmann, 2014, S. 88). Welche genauen neurobiologischen Strukturen dies sind, ist dagegen noch nicht eindeutig geklärt.

Es konnten auch genetische Einflüsse auf die kognitiven Grundlagen des Schriftspracherwerbs, so beispielsweise auf die phonologische Bewusstheit, gefunden werden (Steinbrink & Lachmann, 2014, S. 90–91). Carrion-Castillo et al. (2017, S. 458) konnten in ihrer Langzeitstudie zwei Einzelnukleotid-Polymorphismen auf dem Gen KIAA0319 ausmachen, welche mit der Benennungsgeschwindigkeit zusammenhängen. Steinbrink und Lachmann (2014, S. 91) halten zusammenfassend fest, dass es Genorte gibt, die eine Anfälligkeit zur Entwicklung einer Lese-Rechtschreibstörung beinhalten, diese an sich aber nicht hinreichend oder

notwendig sind, eine Lese-Rechtschreibstörung auch auszulösen. Die Anfälligkeit als solche führt in der Regel demnach erst in Kombination mit weiteren Faktoren zu Schwierigkeiten beim Leseerwerb.

Dank moderner bildgebender Verfahren ist es möglich, herauszufinden, welche Areale im Gehirn beim Lesen aktiviert werden. Steinbrink und Lachmann (2014, S. 93–94) geben die bisherigen Forschungsergebnisse wieder. Es können sich bei Lese-Rechtschreibstörungen Über- und Unteraktivierungen in Regionen der linken Gehirnhälfte sowie Überaktivierungen in der rechten Gehirnhälfte zeigen (Steinbrink & Lachmann, 2014, S. 93). Sie verweisen an dieser Stelle auf die Metaanalyse von Richlan, Kronbichler und Wimmer (2009). In diese Metaanalyse gingen 17 Studien ein, die entweder über fMRT oder PET die Gehirnaktivität von Personen mit und von Personen ohne Lesestörung beim Lesen von Wörtern oder Pseudowörtern erfasst und verglichen haben (Richlan et al., 2009, S. 3300). Im dorsalen Lesesystem konnten Unteraktivierungen bei Personen mit Lesestörungen ausgemacht werden, während es im anterioren Lesesystem zu Überaktivierungen kam (Richlan et al., 2009, S. 3306). Sie konnten allerdings keine Dysfunktion des Gyrus angularis feststellen, obwohl diese Region schon seit dem 19. Jahrhundert als zentrale Stelle im Gehirn für Lesekompetenz galt (Morgan, 1896, S. 1378). Als letztes berichten sie noch über eine Überaktivierung im linken Gyrus lingualis, was auf eine mögliche längere visuelle Verarbeitung bei Personen mit Lesestörung schließen könnte (Richlan et al., 2009, S. 3307). Klicpera et al. (2017, S. 187–188) merken allerdings zurecht kritisch an, dass diese Ergebnisse nicht darauf schließen lassen, dass die veränderten Aktivierungen im Gehirn die Ursache für Lesestörungen sind, da sie umgekehrt auch eine Folge darstellen können.

### 8.1.2 Kognitive Fähigkeiten

Die Intelligenz wurde bereits in Kapitel 6.2 in Verbindung mit der Definition der Lese-Rechtschreibschwierigkeiten thematisiert, wo sie im Zuge des doppelten Diskrepanzkriteriums eine kontrovers diskutierte Rolle spielt. Das doppelte Diskrepanzkriterium sieht vor, dass die Leistung in der betroffenen Teilfertigkeit unterhalb der durch die Intelligenz anzunehmenden möglichen Leistung liegt. Der Fakt, dass dieses Kriterium existiert, macht bereits deutlich, dass eine niedrige Intelligenz nicht alleinige Ursache für Schwierigkeiten beim Leseerwerb sein kann, da die von einer diagnostizierten Lesestörung betroffenen Kinder im Lesen eine Leistung zeigen, die unterhalb ihres kognitiven Potentials liegt. Vielmehr ist eine Intelligenzminderung laut ICD-10 gleichzeitig Ausschlusskriterium für eine Lesestörung (Deutsches Institut für Medizinische Dokumentation und Information, 2018, S. 325).

Tischler, Daseking und Petermann (2017) untersuchten den Zusammenhang zwischen Intelligenz und Leseleistungen bei Kindern im Alter von sieben bis 16 Jahren. Ziel der Studie war, herauszufinden, inwiefern für den Leseerwerb grundlegende kognitive Faktoren die Lesefähigkeit beschreiben können (Tischler et al., 2017, S. 49). Um die Intelligenz möglichst umfassend zu erheben, wurde der HAWIK IV verwendet, da dieser neben der Berechnung des

Gesamt-IQs noch Indizes für das Sprachverständnis, das wahrnehmungsgebundene logische Denken, das Arbeitsgedächtnis und die Verarbeitungsgeschwindigkeit enthält (Tischler et al., 2017, S. 51). Im Ergebnis zeigte sich, dass einzelne kognitive Fähigkeiten mit der Lesefertigkeit zusammenhängen. Den größten Zusammenhang zeigten das Arbeitsgedächtnis und das Sprachverständnis (Tischler et al., 2017, S. 54). Das wahrnehmungsgebundene logische Denken, welches die fluide Intelligenz abbildet, steht nur moderat im Zusammenhang mit den Lesefertigkeiten. Auch der vierte und letzte Index, die Verarbeitungsgeschwindigkeit, steht im Zusammenhang mit dem Lesen. Tischler et al. (2017, S. 56) kommen zu dem Schluss, dass insgesamt ein moderater Zusammenhang zwischen den Leseleistungen und der Intelligenz besteht und sprechen sich für eine „differenziertere Berücksichtigung“ der Intelligenz in Bezug zur Leseleistung aus, damit die Förderung an die kognitiven Voraussetzungen des Kindes angepasst werden kann.

Des Weiteren gehören auch Schülerinnen und Schüler mit Lernbehinderungen zur Risikogruppe, insofern es sich bei diesen um „inaktive Lernende“ handelt, die ihr Lernen nicht überwachen und kein strategisches Vorgehen aufweisen (Klingner, Urbach, Golos, Brownell & Menon, 2010, S. 60). Eine Lernbehinderung liegt vor, wenn Lernrückstände in mehreren Unterrichtsfächern – insbesondere Deutsch und Mathematik – vorliegen, die mindestens zwei Schuljahre betragen und wenn diese Lernrückstände nicht aufgrund mangelnden Unterrichts sondern aufgrund einer unterdurchschnittlichen Intelligenz – meist wird ein IQ zwischen 55 und 85 angegeben – vorhanden sind (Grünke & Grosche, 2014, S. 78). Für diesen spezifischen Fall wird durch die Definition an sich eine unterdurchschnittliche intellektuelle Leistungsfähigkeit verlangt. Im Gegensatz zu dieser Gruppe verfügen gute Leserinnen und Leser als aktiv Lernende nicht nur über ein entsprechend großes Repertoire an Lesestrategien, sondern haben auch Leseziele und sind sich somit darüber im Klaren, warum sie einen Text lesen und nach welchen Informationen sie im Text suchen (Rosebrock & Nix, 2015, S. 74).

### 8.1.3 Exekutive Funktionen

Exekutive Funktionen sind das Denken und Handeln regulierende höhere kognitive Prozesse (Schuchardt & Mähler, 2016, S. 393). Hierzu werden drei grundlegende Funktionen gezählt: die inhibitorische Kontrolle, das Arbeitsgedächtnis und die kognitive Flexibilität (Diamond, 2014, S. 19). Über das Arbeitsgedächtnis bei Kindern mit Lese-Rechtschreibschwierigkeiten wurden relativ viele Studien durchgeführt, während für die Fähigkeiten der kognitiven Flexibilität und der Inhibition vergleichsweise wenige Studien vorliegen (Schuchardt & Mähler, 2016, S. 395). Daucourt, Schatschneider, Connor, Al-Otaiba und Hart (2018, S. 2) bezeichnen alle drei Komponenten als relevant für das Lesen. Inhibition ist definiert als die Fähigkeit, unbedeutende Informationen zu unterdrücken. Spezifisch für das Lesen bedeutet das, dass irrelevante Wortbedeutungen ignoriert werden müssen oder zum Beispiel am Ende des Absatzes mit dem Lesen pausiert werden muss, wenn nur dieser Absatz gelesen werden soll. Das Arbeitsgedächtnis ist dafür zuständig, durchgehend neue Informationen zu verarbeiten

und zu behalten sowie ältere Informationen ersetzen zu können, wenn diese weniger relevant sind. Beim Lesen müssen entsprechend die Bedeutungen von bekannten Wörtern abgerufen und unbekannte Wörter dekodiert werden. Weiterhin muss der bisher gelesene Text in Erinnerung behalten und nachfolgende Textstellen müssen inhaltlich antizipiert werden. Die kognitive Flexibilität bezeichnet die Fähigkeit, zwischen Prozessen hin und her wechseln zu können. Beim Lesen muss beispielsweise zwischen zeitlichen Ebenen in einem Text oder zwischen bekannten und unbekanntem Wörtern unterschieden werden. Bisherige Studien sind insgesamt zu gemischten Befunden gekommen, was den Zusammenhang zwischen diesen drei Komponenten und der Lesestörung angeht (Daucourt et al., 2018, S. 5). In der Studie von Daucourt et al. (2018) mit 420 Kindern aus den USA stellten sich alle drei Komponenten der exekutiven Funktionen als signifikante Prädiktoren für das Auftreten einer Lesestörung heraus. Im Detail zeigte sich, dass jede der drei Komponenten der exekutiven Funktionen eine gleich starke Vorhersageleistung hatte und bei steigendem Ausmaß an Schwierigkeiten in den exekutiven Funktionen auch die Wahrscheinlichkeit stieg, eine Lesestörung zu haben (Daucourt et al., 2018, S. 11).

Wie bereits genannt, konnten auch Tischler et al. (2017, S. 54) in ihrer Studie zeigen, dass das Arbeitsgedächtnis mit der Lesefertigkeit zusammenhängt. Auch das phonologische Arbeitsgedächtnis, teilweise auch als phonologische Schleife bezeichnet, wurde in Studien berücksichtigt. Es handelt sich um eine Komponente des Arbeitsgedächtnisses und bezieht sich auf die Verarbeitung sprachlicher Informationen (Hasselhorn & Marx, 2000, S. 136). Die Kapazität bzw. Funktion des phonologischen Arbeitsgedächtnisses wird überprüft, indem Items im Sekundenrhythmus akustisch dargeboten werden. Diese Items sollen im Anschluss reproduziert werden. Die Anzahl der korrekt reproduzierten Items wird als Gedächtnisspanne bezeichnet (Hasselhorn & Marx, 2000, S. 136–137). Ein beeinträchtigtes phonologisches Arbeitsgedächtnis wird als eine von mehreren Ursachen für Lese- und Rechtschreibschwierigkeiten angesehen und beeinflusst nicht nur das Dekodieren und Rekodieren negativ, sondern auch das Leseverständnis (Hasselhorn & Marx, 2000, S. 145–146; Schuchardt, Brandenburg, Fischbach & Mähler, 2017, S. 170).

Schuchardt et al. (2017, S. 170) führen an, dass nicht alle Kinder mit Lese-Rechtschreibschwierigkeiten auch Beeinträchtigungen im Arbeitsgedächtnis zeigen. Sie schlussfolgern daraus, dass es sich bei Lese-Rechtschreibstörungen um eine heterogene Gruppe von Störungen handelt und entsprechende Subgruppen vorhanden sind, deren Ursachen folglich ebenfalls heterogen sind. Sie vermuten weiter, dass Komorbiditäten hierfür verantwortlich sind, da Lese-Rechtschreibschwierigkeiten „überzufällig häufig“ zusammen mit weiteren Entwicklungs- oder auch Verhaltensstörungen auftreten, so zum Beispiel mit einer Rechenschwäche, mit Aufmerksamkeitsstörungen, Hyperaktivität und Impulsivität (Schuchardt et al., 2017, S. 171). Daher haben sie in ihrer Studie Kinder mit Lese-Rechtschreibschwierigkeiten hinsichtlich ihrer kognitiven Funktionen verglichen. Es wurden die Daten von 99 Kindern einbezogen, welche drei Gruppen zugeordnet werden konnten. Die erste Gruppe an

Kindern hatte unterdurchschnittliche Leistungen im Lesen und/oder im Schreiben, aber nicht in Mathematik und zeigte ein Aufmerksamkeitsdefizit. Die zweite Gruppe hatte ebenfalls unterdurchschnittliche Leistungen im Lesen und/oder Schreiben bei unbeeinträchtigten Mathematikleistungen, aber kein Aufmerksamkeitsdefizit. Die Kontrollgruppe hatte durchgehend durchschnittliche Leistungen und ebenfalls kein Aufmerksamkeitsdefizit (Schuchardt et al., 2017, S. 172). Es zeigten sich für die Gruppe der Kinder mit Lese-Rechtschreibschwäche Beeinträchtigungen im phonologischen Arbeitsgedächtnis, in der Benennungsgeschwindigkeit sowie in der phonologischen Bewusstheit (Schuchardt et al., 2017, S. 176). Kinder mit einem komorbiden Aufmerksamkeitsdefizit wiesen dagegen vornehmlich Auffälligkeiten in den zentral-exekutiven Funktionen auf. Eine Überlappung der beiden Gruppen trat nur bei der phonologischen Bewusstheit auf, was damit erklärt wird, dass hier das Arbeitsgedächtnis besonders gefordert wird. Schuchardt et al. (2017, S. 176) kommen zu dem Schluss, dass Kinder mit zusätzlichem Aufmerksamkeitsdefizit eine eigene Gruppe darstellen, die aufgrund dieser Problematik, aber nicht aufgrund von Einschränkungen in der phonologischen Informationsverarbeitung Schwierigkeiten im Lesen und Schreiben haben. Insgesamt lassen sich die Kinder relativ gut den beiden Subgruppen zuordnen, aber auch hier weist nicht jedes Kind mit Lese-Rechtschreibschwierigkeiten die typischen Auffälligkeiten auf. Ebenso zeigen auch einige Kinder der Kontrollgruppe die jeweiligen Auffälligkeiten, ohne jedoch eine Lese-Rechtschreibschwäche ausgebildet zu haben (Schuchardt et al., 2017, S. 176–177).

In vielen Studien, so auch in einigen der hier genannten, wurde allerdings nicht zwischen Störungen bzw. Schwächen im Lesen und im Schreiben unterschieden. Brandenburg et al. (2013) trennten in ihrer Studie danach, ob Schwierigkeiten im Lesen, im Schreiben oder kombinierte Schwierigkeiten in beiden Bereichen vorlagen und gingen der Frage nach, ob sich entsprechend unterscheidbare Arbeitsgedächtnisprofile finden lassen (Brandenburg et al., 2013, S. 148). Für die Kinder mit Schwierigkeiten im Lesen ergaben sich keine Defizite für den visuell-räumlichen Notizblock. In der phonologischen Schleife und der zentralen Exekutive zeigten sich Defizite in der Ziffernschleife vorwärts, der Wortschleife rückwärts und der Zählspanne (Brandenburg et al., 2013, S. 152). Kinder mit kombinierten Lese-Rechtschreibschwierigkeiten wiesen Defizite in jeweils drei Subtests für die phonologische Schleife und die zentrale Exekutive auf, nicht aber im visuell-räumlichen Notizblock. Brandenburg et al. (2013, S. 154) interpretieren die Ergebnisse so, dass Kinder mit isolierten Leseschwierigkeiten im phonologischen Arbeitsgedächtnis weitestgehend unbeeinträchtigt sind. Dadurch, dass sich im visuell-räumlichen Notizblock keine Defizite zeigen, wird darauf geschlossen, dass eine beeinträchtigte visuelle Informationsverarbeitung als Ursache für Leseschwierigkeiten ausgeschlossen werden kann. Nur in der zentralen Exekutive ergaben sich entsprechend große Defizite, wobei die in dieser Studie verwendeten Tests einen hohen verbalen Anteil enthielten. Bei Kindern mit kombinierten Schwierigkeiten scheinen sich die Defizite in den Arbeitsgedächtnisfunktionen im Vergleich zu den isolierten Störungen zu addieren (Brandenburg et al., 2013, S. 155).

Zuletzt merken Schuchardt et al. (2017, S. 177) noch kritisch an, dass unklar ist, ob die Auffälligkeiten Ursache oder Folge der Lese-Rechtschreibschwierigkeiten sind, und dass es in Studien Belege für beide Wirkungsrichtungen gibt. Schuchardt und Mähler (2016, S. 400) kommen zu dem Schluss, dass bisher noch Längsschnittstudien fehlen, die Aufschluss über die Wirkrichtung geben können. Gleichzeitig sehen sie die bisherigen Erkenntnisse als hilfreich, da ein Wissen über Beeinträchtigungen in den exekutiven Funktionen die Planung von Fördermaßnahmen entsprechend beeinflussen sollte, um angemessen unterstützen zu können (Schuchardt & Mähler, 2016, S. 401).

#### 8.1.4 Aufmerksamkeit und Konzentration

Yamada et al. (2011, S. 711–712) berichten, dass die Daten mehrerer Kinder in ihrer Studie nicht ausgewertet werden konnten, weil diese nicht die notwendige Zeit über stillsitzen konnten. Sie waren alle Teil der „at risk“-Gruppe gewesen mit niedrigeren Werten bezüglich der Vorläuferfertigkeiten für den Schriftspracherwerb. Eine weitere, von der gleichen Forschergruppe durchgeführte Studie (Stevens et al., 2013), kam ebenfalls zu dem Ergebnis, dass die „at risk“-Kinder im Kindergartenalter über eine niedrigere selektive Aufmerksamkeit verfügen als die Kinder, die niedrigere Ausprägungen in den Vorläuferfertigkeiten aufwiesen. Gleichzeitig zeigte sich aber auch, dass eine entsprechende Förderung zu einer Verbesserung der selektiven Aufmerksamkeit führte (Stevens et al., 2013, S. 81). Nach Schuchardt, Fischbach, Balke-Melcher und Mähler (2015, S. 191) kommt bei 20 Prozent der Kinder mit Lese-Rechtschreibstörung bzw. -schwäche eine komorbide ADHS vor, während nur fünf Prozent der Kinder ohne diese Schwierigkeiten eine ADHS haben. Das gemeinsame Auftreten lässt aber keine Aussagen zur Wirkrichtung zu. ADHS kann allerdings auch keine hinreichende Ursache für eine Lese-Rechtschreibstörung sein, da diese kombiniert mit Lese-Rechtschreibschwierigkeiten „nur“ bei jedem fünften Kind auftreten.

Günther et al. (2016) erhoben in ihrer Studie mehrere Aufmerksamkeitsmaße sowie die Leseleistungen bei Kindern. Dabei unterschieden sie vier Gruppen: Kinder mit diagnostizierter Lese-Rechtschreibstörung, Kinder mit diagnostizierter ADHS, Kinder, welche sowohl eine diagnostizierte Lese-Rechtschreibstörung wie auch ADHS hatten und eine Kontrollgruppe von Kindern ohne Auffälligkeiten (Günther et al., 2016, S. 354). Es wurden zudem die Eltern nach Verhaltenseinschätzungen zu ihren Kindern befragt. Kinder mit Lese-Rechtschreibstörung wiesen häufiger internalisierende Störungen auf als Kinder ohne diese Störung (Günther et al., 2016, S. 358). Erhöhte externalisierende Verhaltensweisen gegenüber der Kontrollgruppe zeigten alle Gruppen an Kindern. Insgesamt traten bei Kindern mit komorbider ADHS und Lese-Rechtschreibstörung die meisten Verhaltensauffälligkeiten auf (Günther et al., 2016, S. 355). In Bezug auf die Leseleistungen ergab sich erwartungsgemäß, dass alle Kinder mit Lese-Rechtschreibstörung insgesamt die niedrigsten Leistungen erzielten (Günther et al., 2016, S. 358). Gleichzeitig wiesen die Kinder mit ADHS ohne Lese-Rechtschreibstörung im Vergleich zur Kontrollgruppe ebenfalls Schwierigkeiten im Dekodieren und Wortverständnis

auf. Es wird vermutet, dass dies durch eine geringere Aufmerksamkeit und Handlungskontrolle zustande kommt. Wäre die Testung ohne zeitliche Limitierungen erfolgt, so vermuten die Autorinnen und Autoren, hätten sich bei den Kindern mit ADHS eventuell keine schlechteren Leseleistungen ergeben (Günther et al., 2016, S. 358). Einschränkend halten Günther et al. (2016, S. 359) noch fest, dass bei Kindern mit einer Lese-Rechtschreibstörung eventuell eine vorhandene Aufmerksamkeitsstörung (ohne Hyperaktivität) unerkannt bleibt, da diese Form der Aufmerksamkeitsstörung generell nicht oft erkannt wird. Ob diese aber auch ursächlich verantwortlich ist für die Lesestörung, ist in diesen Fällen nicht bekannt. Daher ist eine umfassende Diagnostik notwendig, um zugrundeliegende Aufmerksamkeitsstörungen zu erkennen (Günther et al., 2016, S. 359).

Zusammenfassend lässt sich festhalten, dass Aufmerksamkeitsstörungen sehr wahrscheinlich in der Folge zumindest zu Leseschwierigkeiten führen können.

#### 8.1.5 Spracherwerbsstörungen

Die ICD-10-Beschreibung der Lese- und Rechtschreibstörung führt bereits an, dass dieser Störung „Entwicklungsstörungen des Sprechens oder der Sprache“ vorausgehen (Deutsches Institut für Medizinische Dokumentation und Information, 2018, S. 325).

Die Sprachentwicklung von Kindern zeigt eine große Variabilität, wobei im Alter von fünf bis sechs Jahren ein rezeptiver Sprachschatz von 9000 bis 14000 Wörtern und ein aktiver Sprachschatz von 2000 bis 3000 Wörtern erreicht wird (Petermann, 2016, S. 132).

Spracherwerbsstörungen werden in primäre und sekundäre unterschieden, wobei die primäre dann vorliegt, wenn die Faktoren der sekundären Störung ausgeschlossen werden können (Petermann, 2016, S. 132). Sekundäre Sprachstörungen liegen vor, wenn sprachliche Auffälligkeiten aufgrund anderer zugrundeliegender Erkrankungen oder Störungen auftreten. Hierzu zählen neuronale oder Hörschädigungen, genetische Erkrankungen, psychische Störungen oder auch Umgebungsbedingungen wie ein sprach- und anregungsarmes Umfeld. Petermann (2016, S. 132) hält fest, dass Mehrsprachigkeit kein Risikofaktor für die Entwicklung einer solchen Spracherwerbsstörung ist, dass diese Gruppe aber dennoch Unterstützung benötigt, da sie eigene Entwicklungsanforderungen zu bewältigen hat.

Spracherwerbsstörungen zeigen eine große Variabilität und können verschiedene Sprachebenen betreffen. Sowohl das Sprachverständnis wie auch die Sprachproduktion können beeinträchtigt sein (Ronniger, Melzer, Petermann & Reißling, 2016, S. 136). Es kann in der Folge beim Leseerwerb zu Beeinträchtigungen in der Erzählfähigkeit sowie im Textverständnis kommen (Ronniger et al., 2016, S. 137). Catts, Fey, Tomblin und Zhang (2002) untersuchten in einer Langzeitstudie den Einfluss von Sprachbeeinträchtigungen auf die Leseentwicklung von Kindern. Etwa die Hälfte der Kinder mit Sprachbeeinträchtigungen im Kindergartenalter hatte in der zweiten und in der vierten Klasse Lesestörungen entwickelt. Insgesamt zeigte diese Gruppe niedrigere Leseleistungen als eine unbeeinträchtigte Kontrollgruppe (Catts et al., 2002, S. 1152).

In der Studie von Tischler et al. (2017, S. 54) konnte ein Zusammenhang der Lesefertigkeit mit dem Sprachverständnis festgestellt werden. Zusätzlich erklären Sprachentwicklungsstörungen im Vorschulalter einen großen Teil der Varianz der Lesefertigkeiten. Die berücksichtigten Sprachentwicklungsstörungen beinhalten auch Wortschatzdefizite, differenzieren ansonsten allerdings nicht weiter (Tischler et al., 2017, S. 55).

Insgesamt kann festgehalten werden, dass für Kinder mit Spracherwerbsstörungen ein großes Risiko besteht, zumindest Schwierigkeiten beim Erwerb von Lesen und Schreiben aufzuweisen und dass die Wahrscheinlichkeit, eine Lese-Rechtschreibstörung zu entwickeln, erhöht ist.

#### 8.1.6 Geschlecht und affektive Faktoren

Deutschland gehört zu den Nationen, bei denen es einen Unterschied zwischen den Geschlechtern in Bezug zur Lesekompetenz gibt. In allen Ländern, in denen sich die Leseleistungen von Jungen und Mädchen unterscheiden, sind es die Mädchen, die bessere Leistungen erbringen (McElvany & Ohle, 2015, S. 48). Dieses Ergebnis wurde auch in der IGLU-Studie 2016 gefunden, wobei der Unterschied vor allem im literarischen Lesen und nur wenig beim informierenden Lesen zum Tragen kommt (Bos et al., 2017, S. 20). Es wurden verschiedene Theorien aufgestellt, die den Unterschied erklären sollen, wie beispielsweise evolutionär-biologische Gründe (McElvany, Kessels, Schwabe & Kasper, 2017, S. 177). Krinzinger und Günther (2013) geben einen Forschungsüberblick und fassen zusammen, dass es einen leichten Vorteil für Frauen beim Schriftspracherwerb gibt, wobei die praktische Relevanz angezweifelt werden kann (Krinzinger & Günther, 2013, S. 40). Studien mit Kindern zeigen insgesamt auf, dass die Entwicklung von Jungen generell, nicht nur auf die Sprachfertigkeit bezogen, langsamer zu sein scheint als die Entwicklung von Mädchen. Die Unterschiede in der Sprachfertigkeit sind allerdings im Alter von sechs Jahren nicht mehr vorhanden (Krinzinger & Günther, 2013, S. 41).

Krinzinger und Günther (2013, S. 42) führen an, dass Jungen generell etwas häufiger von Lese-Rechtschreibstörungen betroffen sind als Mädchen, die Größe des Verhältnisses allerdings oft überschätzt wird. In unselektierten Studien zeigen etwa doppelt so viele Jungen wie Mädchen Lese-Rechtschreibstörungen. Im Gegensatz dazu zeigt sich bei klinischen Stichproben ein Verhältnis von 3:1 oder gar 6:1 zu Ungunsten der Jungen. Lese-Rechtschreibstörungen werden daher bei Jungen häufiger erkannt, was wiederum dadurch bedingt zu sein scheint, dass Verhaltensprobleme bei Jungen meist externalisiert auftreten, bei Mädchen dagegen internalisiert. Kinder, die unruhig sind oder Verhaltensauffälligkeiten zeigen, fallen vermutlich eher auf (Krinzinger & Günther, 2013, S. 42). Quinn (2018) führte eine Metaanalyse mit 16 Studien über die Diagnose von Lese-Rechtschreibschwierigkeiten bei Jungen und Mädchen durch. Es zeigte sich, dass Jungen 1.83 mal häufiger mit Leseschwierigkeiten diagnostiziert werden als Mädchen, wobei Jungen im Verhältnis zu Mädchen umso häufiger

diagnostiziert werden, je gravierender die zu diagnostizierende Lesestörung definiert war (Quinn, 2018, S. 1054).

Die schwächeren Leseleistungen bei Jungen, die zum Ende der Pflichtschulzeit gefunden werden, können nach McElvany, Kessels, Schwabe und Kasper (2017, S. 180) vor allem auf Unterschiede in der Lesemotivation zurückgeführt werden, die zum Ende der Grundschulzeit bereits „substantiell“ sind. Der Unterschied, der in der IGLU-Studie 2016 in der Lesekompetenz gefunden wurde, verschwindet allerdings, wenn Leseselbstkonzept, Lesemotivation und Lesehäufigkeit, welche bei Mädchen ebenfalls höher ausgeprägt sind, kontrolliert werden (Bos et al., 2017, S. 20). Quinn (2018, S. 1056) führt die Unterschiede in der Lesefähigkeit in Bezug auf die Lesemotivation auf „stereotype threat“ (Bedrohung durch Stereotype) zurück. Mädchen sind motivierter als Jungen, sodass Lehrkräfte Mädchen möglicherweise stärker im Lesen unterstützen. Auf diese Weise erfahren Jungen, dass Lesen eine weniger erstrebenswerte Aktivität ist und dass sie mit negativeren Konsequenzen rechnen müssen, wenn sie nicht gut lesen können. Dies führt in der Folge zu schlechteren Leseleistungen. Retelsdorf, Schwartz und Asbrock (2015) gingen dieser Frage in ihrer Langzeitstudie ebenfalls nach. 1358 Schülerinnen und Schüler mit ihren 54 Lehrkräften nahmen Anfang der fünften Klasse und 18 Monate später zum zweiten Schulhalbjahr der sechsten Klasse an der Studie teil. Es zeigten sich keine signifikanten Unterschiede zwischen Jungen und Mädchen in Bezug auf das Leseselbstkonzept und die Lesefertigkeit in der fünften sowie in der sechsten Klasse (Retelsdorf et al., 2015, S. 189). Ebenfalls wurden Geschlechtsstereotype der Lehrkräfte erhoben. Es zeigte sich, dass Lehrkräfte, die über solche verfügen, Mädchen bevorzugen. Außerdem zeigten Jungen mit Lehrkräften, die Geschlechtsstereotype hatten, zum zweiten Messzeitpunkt in der sechsten Klasse ein niedrigeres Leseselbstkonzept als Jungen mit Lehrkräften ohne solche Stereotype. Für Mädchen galt dies nicht. Dieses Ergebnis zeigte sich auch dann noch, wenn die Leseleistung statistisch kontrolliert wurde (Retelsdorf et al., 2015, S. 191). Lesemotivation scheint demnach mit der Lesekompetenz zusammenzuhängen. Logan, Medford und Hughes (2011) führten eine Studie durch, um den Effekt von intrinsischer Lesemotivation auf das Leseverständnis zu untersuchen. Es zeigte sich, dass die intrinsische Lesemotivation nicht für alle Leserinnen und Leser das Leseverständnis vorhersagen kann (Logan et al., 2011, S. 126). Der größte Prädiktor für das Leseverständnis bei guten Leserinnen und Lesern waren stattdessen die verbalen Fähigkeiten. Für schwache Leserinnen und Leser waren dagegen die intrinsische Motivation und die Dekodierfähigkeit die größten Prädiktoren. Logan et al. heben hervor, dass andere Studien bereits zeigen konnten, dass die Lesemotivation mit der Lesekompetenz zusammenhängt und dass schwache Leserinnen und Leser über eine niedrigere Lesemotivation verfügen, dass aber zusätzlich deutlich wird, dass der Zusammenhang für schwache Leserinnen und Leser stärker ist als für gute. Es wird vermutet, dass motivierte Leserinnen und Leser mehr Zeit mit dem Lesen von Texten verbringen und sich auch bei schwierigen Texten mehr anstrengen (Logan et al., 2011, S. 127).

McGeown, Duncan, Griffiths und Stothard (2015) führten ihre Studie mit Jugendlichen durch. Im Fokus standen die Lesemotivation, das Leseverhalten sowie deren Auswirkungen auf die Leseleistung. Die Leseleistung wurde differenziert über das Wortlesen, das Leseverständnis, die Fähigkeit zum Zusammenfassen und die Lesegeschwindigkeit erfasst. Die Lesemotivation wie auch das Lesen fiktionaler Literatur waren Prädiktoren für das Leseverständnis und die Fähigkeit zum Zusammenfassen, auch nach statistischer Kontrolle von Wortlesen und Lesegeschwindigkeit. Ebenso waren sie Prädiktoren für die Lesegeschwindigkeit selbst (McGeown et al., 2015, S. 563). Mädchen im Speziellen sowie gute Leserinnen und Leser allgemein verfügten über eine höhere Lesemotivation und lasen häufiger fiktionale Bücher als Jungen bzw. schwache Leserinnen und Leser (McGeown et al., 2015, S. 564). Eine höhere Lesemotivation sowie vermehrtes Lesen fiktionaler Bücher sollten dementsprechend Ziele sein, um die Leseleistung zu erhöhen (McGeown et al., 2015, S. 566).

In der IGLU-Studie 2016 ergab sich insgesamt eine hohe Lesemotivation für die Viertklässlerinnen und -klässler in Deutschland. Weiterhin hängt die Lesemotivation positiv mit der Leseleistung zusammen. Allerdings zeigt sich über den Verlauf seit 2001, dass die Lesemotivation bei leseschwächeren Kindern abnimmt (Goy, Valtin & Hußmann, 2017, S. 150). Analog zu McGeown et al. (2015) besteht auch in der IGLU-Studie ein positiver Zusammenhang der Lesemotivation mit der Lesehäufigkeit. Gleichzeitig nimmt die Lesemenge in der Freizeit vor allem bei leseschwachen Kindern seit 2011 ab (Goy et al., 2017, S. 156).

Die IGLU-Studie 2016 kommt weiterhin zu dem Ergebnis, dass drei Viertel der Schülerinnen und Schüler über ein hohes Leseselbstkonzept verfügen und nur fünf Prozent über ein niedriges (Goy et al., 2017, S. 149). 87 Prozent der Kinder, die hohe Leseleistungen zeigen, haben auch ein hohes Leseselbstkonzept, während 44 Prozent der Kinder mit niedrigen Leseleistungen ein hohes Leseselbstkonzept haben. Im Vergleich zu den Ergebnissen aus der IGLU-Studie von 2011 fällt die Selbsteinschätzung der Kinder realistischer aus (Goy et al., 2017, S. 150). Katzir, Kim und Dotan (2018) konnten in ihrer Studie über Grundschul Kinder in der zweiten Klasse zwischen der Angst vor dem Lesen und dem Leseselbstkonzept einen negativen Zusammenhang finden (Katzir et al., 2018, S. 8). Auch hier verfügten Mädchen über eine bessere Leseleistung sowie eine höhere Lesegenauigkeit von Wörtern als Jungen. Gleichzeitig zeigte sich aber für Mädchen eine höhere Angst vor dem Lesen und ein niedrigeres Leseselbstkonzept (Katzir et al., 2018, S. 9). Es wird vermutet, dass Mädchen sich allgemein im akademischen Bereich weniger zutrauen oder dass sie kritischer mit Rückmeldungen umgehen. Über die Gründe kann aber nur spekuliert werden (Katzir et al., 2018, S. 10).

In der IGLU-Studie 2006 ergaben sich niedrigere Werte im Selbstkonzept bei schwachen Leserinnen und Lesern, sodass hier von einem Risikofaktor - allerdings im Sinne einer sekundären Symptomatik - gesprochen wird (Valtin et al., 2010, S. 48). Ein angenehmes Klassen- und Schulklima sorgt für ein höheres Wohlbefinden und damit auch für eine geringere psychische Belastung. Diese Studie zeigte, dass ein höherer Prozentsatz an leseschwachen

Schülerinnen und Schülern negative soziale Erfahrungen mit Mitschülerinnen und Mitschülern gemacht hat, als dies die Vergleichsgruppe der starken Leserinnen und Leser angab (Valtin et al., 2010, S. 57). Die Richtung des Zusammenhangs lässt sich aber nicht erfahren.

## 8.2 Soziale und kulturelle Herkunft

Erste Erfahrungen mit Büchern und Literatur machen Kinder in ihrer Familie. Aktivitäten wie Vorgelesen bekommen oder das gemeinsame Ansehen von Bilderbüchern prägen Kinder schon früh und zeigen sich als Prädiktoren für die spätere Lesekompetenz von Kindern (McElvany et al., 2009, S. 130).

Es wird zwischen primären und sekundären Herkunftseffekten unterschieden (Hußmann, Stubbe & Kasper, 2017, S. 196). Primäre Herkunftseffekte sind die Faktoren, die bereits zum Zeitpunkt der Einschulung vorhanden sind. Sie sorgen für „ungleiche Startbedingungen“. Hierzu zählen zum Beispiel kulturelle Aspekte wie Theaterbesuche oder gemeinsames Musizieren. Diese Effekte halten über den Verlauf der Schulzeit an und sorgen für Unterschiede z. B. in den Bildungsverlaufsentscheidungen, was zu den sekundären Herkunftseffekten gezählt wird (Hußmann, Stubbe, et al., 2017, S. 196).

In der IGLU-Studie 2016 wurden als Indikatoren für die soziale Herkunft der Bücherbestand und weitere Besitztümer im Elternhaus, der höchste Bildungsabschluss sowie der Berufsstatus der Eltern und das Einkommen herangezogen (Hußmann, Stubbe, et al., 2017, S. 199–200). Für internationale Vergleiche wurden nur die Indikatoren Bücherbestand, Bildungsabschluss und Berufsstatus verwendet (Hußmann, Stubbe, et al., 2017, S. 201). Für die vierte Klasse beträgt der Leistungsunterschied in Deutschland zwischen Kindern aus höheren und niedrigeren sozialen Schichten ein Lernjahr (Hußmann, Stubbe, et al., 2017, S. 197). Kinder, die in einem Haushalt mit mehr als 100 Büchern aufwachsen, zeigen im Schnitt demnach einen Lernfortschritt in der Lesekompetenz von einem Jahr gegenüber Kindern, die in einem Haushalt mit weniger als 100 Büchern aufwachsen. Es gibt kein Teilnehmerland, in dem diese Differenz noch höher ausfällt (Hußmann, Stubbe, et al., 2017, S. 201). Das gleiche gilt auch in Bezug auf den Berufsstatus. Für den Bildungsabschluss zeigt sich ein ähnliches Bild. Hier gibt es allerdings vier Teilnehmerländer (Bulgarien, Israel, Ungarn und Singapur), in denen die Differenz noch größer ausfällt (Hußmann, Stubbe, et al., 2017, S. 204). Die Fragebögen für Deutschland wurden um weitere Items erweitert, um ein detaillierteres Bild über den Zusammenhang der sozialen Herkunft mit der Lesekompetenz zu erhalten (Hußmann, Stubbe, et al., 2017, S. 208). Es zeigt sich eine konstante Entwicklung seit der IGLU-Studie 2001 (Hußmann, Stubbe, et al., 2017, S. 209). Während einige Länder die Unterschiede verringern konnten, gehört Deutschland zu den wenigen Nationen, in denen sich die Unterschiede seit 2001 sogar vergrößert haben (Hußmann, Stubbe, et al., 2017, S. 214).

Niklas, Möllers und Schneider (2013) untersuchten in einer Längsschnittstudie die familiäre Lernumwelt im Vorschulalter als möglichen Einflussfaktor auf die Leseleistungen von Kindern zum Ende der ersten Klasse. Als Teil der Lernumwelt erfassten sie zum Beispiel die Anzahl

an Büchern im Haushalt, das Vorhandensein eines Zeitungsabonnements, das elterliche Leseverhalten oder Leseaktivitäten mit dem Kind (Niklas et al., 2013, S. 99). Die familiäre Lernumwelt klärte nur eine „moderate“ Menge der Varianz der Leseleistungen auf (Niklas et al., 2013, S. 106). Ein niedriger sozioökonomischer Status sowie der Migrationshintergrund der Eltern wirkten sich jeweils negativ auf die familiäre Lernumwelt aus. Familien mit Migrationshintergrund hatten außerdem auch häufiger einen niedrigeren sozioökonomischen Status. Die phonologische Bewusstheit und der Wortschatz wirkten als Mediatoren der familiären Lernumwelt und der Leseleistungen zum Ende der ersten Klasse (Niklas et al., 2013, S. 106–107).

Retelsdorf und Möller (2008, S. 234) konnten in zwei Studien einen Zusammenhang zwischen dem familiären Hintergrund, erfasst über den sozioökonomischen Status und das Bildungsniveau, gemeinsamen sprachlichen Aktivitäten sowie der Lesemenge der Eltern aufzeigen. Diese Faktoren hängen wiederum mit der Lesemotivation und dem Leseselbstkonzept der Kinder zusammen, welche sich positiv auf die Lesekompetenz der Kinder auswirken. Es finden sich bei zehn Prozent der Grundschul Kinder mit Migrationshintergrund schulische Leistungen im förderbedürftigen Bereich (McElvany & Ohle, 2015, S. 48). In den IGLU-Studien wurde der Migrationshintergrund so definiert, dass mindestens ein Elternteil im Ausland geboren wurde (Wendt & Schwippert, 2017, S. 221). Kinder mit Migrationshintergrund verfügten über eine niedrigere Lesekompetenz (Wendt & Schwippert, 2017, S. 219). Im Vergleich zu Kindern ohne Migrationshintergrund lag dieser Rückstand in der IGLU-Studie 2016 für Kinder mit einem im Ausland geborenen Elternteil bei einem halben und für Kinder mit zwei im Ausland geborenen Elternteilen bei einem Schuljahr (Wendt & Schwippert, 2017, S. 226). Ein möglicher Grund hierfür ist ein niedrigerer sozialer Status dieser Familien. Allerdings zeigt sich auch bei Kontrolle des sozialen Status in den Ergebnissen der IGLU-Studie 2016 noch ein signifikanter Unterschied in der Leseleistung der Kinder mit und ohne Migrationshintergrund (Wendt & Schwippert, 2017, S. 228). Weiterhin gibt es auch Unterschiede hinsichtlich der Lesehäufigkeit, der Lesemotivation und des Lesekonzepts. Diese Faktoren fallen bei Kindern mit Migrationshintergrund etwas niedriger aus (Wendt & Schwippert, 2017, S. 230).

Die Studie von Leyendecker, Citlak, Schröpfer und Schölmerich (2014) beschäftigte sich mit Familien, die aus der Türkei, aus Russland sowie aus Polen zugewandert waren. Das Augenmerk lag in dieser Studie darauf, welche Einstellungen und Erwartungen die Eltern an die Bildungsinstitutionen hatten und welchen außercurricularen Aktivitäten die Kinder in der Vorschule nachgingen (Leyendecker et al., 2014, S. 73). Insgesamt gab es bei allen Familien unabhängig vom Migrationshintergrund einen hohen Zusammenhang zwischen dem Bildungsniveau der Mutter und den Aktivitäten der Kinder. So sahen Kinder beispielsweise weniger fern und die Eltern schätzten Vorlesen als wichtiger ein, je höher der Bildungsabschluss war (Leyendecker et al., 2014, S. 85). Weitere Unterschiede zeigten sich darin, dass Kinder aus Familien mit Migrationshintergrund häufiger fernsahen als Kinder ohne Migrations-

hintergrund. Ebenso gingen sie erst später in den Kindergarten, die Familien nahmen seltener an Eltern-Kind-Kursen teil und die Eltern fanden das Vorlesen weniger relevant. Auf der anderen Seite zeigte sich aber, dass Mütter mit türkischem oder polnischem Hintergrund dem dyadischen Spiel mit ihren Kindern eine höhere Bedeutung zumaßen als andere Mütter (Leyendecker et al., 2014, S. 85). Es zeigten sich auch Unterschiede hinsichtlich der Erwartungen an die Bildungsinstitutionen. So war die Erwartung, dass die Schule dafür verantwortlich ist, den Kindern Disziplin beizubringen, höher bei Eltern mit Migrationshintergrund und bei Eltern mit niedriger Schulbildung (Leyendecker et al., 2014, S. 86).

Migrationshintergrund kann demnach nicht pauschal als negativer Einfluss betrachtet werden. Ein weiterer Faktor, der hier eine wichtige Rolle spielt, ist die Familiensprache (Wendt & Schwippert, 2017, S. 220). Kinder, die zuhause so gut wie immer oder immer Deutsch sprechen, haben eine signifikant höhere Lesekompetenz als Kinder, die nur manchmal oder auch nie Deutsch zuhause sprechen, und sind diesen in etwa ein Schuljahr voraus (Wendt & Schwippert, 2017, S. 224). In der Studie von Tiedemann und Billmann-Mahecha (2007) konnte neben der Familiensprache auch die Freizeitsprache, also die Sprache, welche die Kinder und Jugendlichen außerhalb der Schule im Freundeskreis benutzen, einen großen Teil der Varianz der Leseverständnisleistung aufklären (Tiedemann & Billmann-Mahecha, 2007, S. 46).

In einer Studie von Treutlein, Roos und Schöler (2012) zeigte sich das gleiche Bild: Je besser Kinder die deutsche Sprache beherrschen, umso weniger Probleme haben sie beim Erwerb von Lesen und Schreiben (Treutlein et al., 2012, S. 130). Zusätzlich wurden Faktoren wie die Klassengröße und der Anteil an Kindern mit unzureichenden Deutschkenntnissen in der Klassengemeinschaft als mögliche Moderatorvariablen betrachtet. Beide Faktoren für sich genommen beeinflussten nicht das Leistungsniveau der Klasse. Es zeigte sich aber, dass der Anteil an Kindern mit unzureichenden Deutschkenntnissen einen negativen Einfluss auf die Leseleistungen eben dieser Kinder hat. In Klassen mit einem hohen Anteil haben niedrige Deutschkenntnisse demnach einen stärkeren negativen Einfluss auf Hörverständnis, Rekodieren und Dekodieren von Kindern mit niedrigen Deutschkenntnissen (Treutlein et al., 2012, S. 130). Es scheint, dass viele positive Sprachvorbilder, also Kinder mit guten Deutschkenntnissen, einen positiven Einfluss auf die Leseleistungen von Kindern mit unzureichenden Deutschkenntnissen haben.

McElvany, Ohle, El-Khechen, Hardy und Cinar (2017) untersuchten in ihrer Studie den Effekt des Einbeziehens der nicht-deutschen Familiensprache auf das Erlernen neuer Wörter aus Lesetexten. Hierfür wurden türkischsprachige Kinder vier Gruppen zugeteilt (McElvany, Ohle, et al., 2017, S. 18). Lesetexte wurden für Kinder in der ersten Gruppe nur auf Deutsch angeboten, während die Kinder in der zweiten Gruppe die Texte zunächst auf Türkisch und dann auf Deutsch lasen. In der dritten Gruppe wurden die Texte zunächst auf Deutsch präsentiert. Die zu lernenden Zielwörter wurden für diese Gruppe zusätzlich auf Türkisch angeboten. Der Text wurde im Anschluss noch einmal auf Deutsch gelesen. Die vierte Gruppe war die

Kontrollgruppe, in der die deutschen Zielwörter durch Synonyme ersetzt worden waren (McElvany, Ohle, et al., 2017, S. 17). Die zusätzliche Darbietung türkischer Wörter bzw. des türkischen Textes im Vergleich zur Kontrollgruppe oder zu der Gruppe mit den ausschließlich deutschen Texten gab keine Vorteile für die türkischsprachigen Kinder. Zweimaliges Lesen der Texte auf Deutsch war stellenweise effektiver, während die zusätzliche Darbietung türkischer Wörter sogar eher einen negativen Effekt hatte. Sie vermuten, dass cognitive load eine Rolle spielt und durch die Darbietung beider Sprachen weniger Ressourcen zum Lernen der neuen Wörter zur Verfügung standen (McElvany, Ohle, et al., 2017, S. 22).

### 8.3 Ursachen durch am Leseprozess beteiligte Teilfertigkeiten

Beim Leseerwerb kann es dazu kommen, dass einzelne Teilfertigkeiten nicht ausreichend ausgebildet werden. Dies kann sich wiederum negativ auf den weiteren Leseerwerb auswirken und so zu Schwierigkeiten beim Lesen führen, was sich zum Beispiel in einer erhöhten Zahl an Lesefehlern oder einer höheren benötigten Lesezeit zeigt (Fischer & Gasteiger-Klicpera, 2013, S. 67). Bei diesen Faktoren handelt es sich also nicht um grundsätzliche Ursachen für Leseschwierigkeiten.

Teilweise gilt die Buchstabenkenntnis als wichtigster Prädiktor für die spätere Lesefähigkeit (Foulin, 2005, S. 130). Sie ist eine „notwendige, aber keine hinreichende Bedingung für das erfolgreiche Dekodieren“ und somit eine wichtige Voraussetzung für den Leseerwerb (Juska-Bacher, Beckert, Gyger, Jakob & Schneider, 2015, S. 30). Die Kenntnis aller Buchstaben ist außerdem Grundlage der visuellen Worterkennung (Foulin, 2005, S. 129). Mangelnde Buchstabenkenntnis wird von Fischer und Gasteiger-Klicpera (2013, S. 67) als typische Schwierigkeit von schwachen Leserinnen und Lesern bezeichnet. Ursache einer schlechten Buchstabenkenntnis kann dabei fehlerhaftes Wissen der Graphem-Phonem-Korrespondenz oder auch dadurch bedingt sein, dass graphische Merkmale von Buchstaben nicht abgespeichert wurden (Fischer & Gasteiger-Klicpera, 2013, S. 67). Ebenso kann es vorkommen, dass schwache Leserinnen und Leser über die Stufe des logographischen Lesens nicht hinauskommen. So lesen sie zum Beispiel nur die Anfangsbuchstaben und orientieren sich an markanten Bestandteilen eines Wortes. Die Schwierigkeiten sind auf gering entwickelte phonologische Fertigkeiten zurückzuführen und es kommt nicht zum Dekodieren, sodass nur einige wenige bekannte Wörter erkannt werden können (Fischer & Gasteiger-Klicpera, 2013, S. 68). Studien zur Buchstabenkenntnis überprüfen häufig die Kenntnis von Großbuchstaben bei Kindern im Vorschulalter und die Lesefähigkeit nach Eintritt in die Grundschule (Foulin, 2005, S. 130). Juska-Bacher et al. (2015, S. 37) kamen in ihrer Studie zu dem Ergebnis, dass 20 Prozent der getesteten Kinder aus der deutschsprachigen Schweiz zu Beginn der Grundschule alle der 28 überprüften Groß- und Kleinbuchstaben benennen konnten. Die Buchstabenkenntnis der übrigen 80 Prozent der Kinder war gleichmäßig verteilt, wobei zwei Kinder keinen einzigen Buchstaben kannten. Sie erhielten erwartungsgemäß eine hohe positive Korrelation der Buchstabenkenntnis mit der Lesekompetenz. Das erfolgreiche Wortlesen begann

ab einer Kenntnis von 17 der hier getesteten 28 Buchstaben (Juska-Bacher et al., 2015, S. 38).

Klicpera et al. (2017, S. 152–153) berichten von Beobachtungen aus ihren Längsschnittuntersuchungen, wonach schwache Leserinnen und Leser bereits drei Monate nach Schuleintritt einen Rückstand dahingehend zeigten, dass sie die bis dahin eingeführten Buchstaben (ca. acht) nicht alle benennen konnten, was bei den guten Leserinnen und Lesern nicht vorkam. Entsprechend lasen sie auch weniger Wörter korrekt als die guten Leserinnen und Leser.

Weiterhin rücken Richter und Müller in ihrem Artikel die Relevanz hierarchieniedriger Prozesse in den Fokus (Richter & Müller, in Druck, S. 18). Zu diesen zählt unter anderem die Leseflüssigkeit, welche aus vier Komponenten besteht: Um flüssig lesen zu können, müssen Kinder über eine ausreichend hohe Lesegeschwindigkeit sowie Dekodiergenauigkeit verfügen. Diese Dekodierprozesse müssen außerdem automatisiert sein (Hochstadt, Krafft & Olsen, 2015, S. 121). Zudem können Kinder beim Vorlesen häufig sinngemäß betonen, was bedeutet, dass nicht nur einzelne Wörter sondern Wortgruppen und ihre Bedeutung innerhalb eines Satzes erfasst werden (Bremerich-Vos et al., 2017, S. 284; Hochstadt et al., 2015, S. 121). Bekannte Wörter können dabei über den direkten Weg erlesen werden und sind somit Bestandteil des Sichtwortschatzes. Diese Wörter müssen nicht mehr indirekt über die phonologische Rekodierung erlesen werden (Bremerich-Vos et al., 2017, S. 283–284). Defizitäre Worterkennungsprozesse und damit eine niedrige Leseflüssigkeit sind häufig das zugrundeliegende Problem für eine nicht ausreichende Lesekompetenz. Kinder, welche zum Ende des ersten Schuljahres zu den schwachen Leserinnen und Lesern gehören, verfügen auch häufig über eine niedrige Leseflüssigkeit (Richter & Müller, in Druck, S. 18). Ebenso ist der erwähnte Sichtwortschatz von großer Bedeutung. Er besteht aus Wörtern, die direkt erkannt werden und die somit keine zusätzlichen kognitiven Ressourcen im Arbeitsgedächtnis mehr benötigen. Die hierdurch freien Ressourcen können unter anderem für das Leseverständnis genutzt werden (Richter & Müller, in Druck, S. 13). Eine hohe Dekodiergenauigkeit allein reicht folglich nicht aus, um die benötigte Leseflüssigkeit zu erzielen. Wenn Wörter zwar dekodiert werden können, aber keine Automatisierung erfolgt, muss ein Kind die Wörter weiterhin über die phonologische Rekodierung erlesen, was ein langsamer Prozess ist, der viel Aufmerksamkeit erfordert (Bremerich-Vos et al., 2017, S. 284).

Vor allem für das Leseverständnis, welches die hierarchiehohe Fertigkeit der Lesekompetenz darstellt, ist der Umfang des Wortschatzes relevant (Klicpera et al., 2017, S. 67). Ein geringerer Wortschatz kann das Lesen negativ beeinflussen. Vor allem im unteren Leistungsbereich zeigen sich Ungleichheiten in Bezug auf den Wortschatz bei fünfjährigen Kindern mit Migrationshintergrund und bei Kindern mit niedrigem sozioökonomischen Status (Linberg & Wenz, 2017, S. 92). Es muss allerdings zwischen dem passiven und dem aktiven Wortschatz unterschieden werden. Der aktive Wortschatz ist vor allem beim Sprechen und Schreiben

relevant, wenn Kinder selbständig Sätze und Texte produzieren. Der passive Wortschatz spielt dagegen beim Zuhören in Gesprächen sowie beim Verständnis von gelesenen Sätzen und Texten eine Rolle (McElvany, El-Khechen, Schwabe & Kessels, 2016, S. 46). Für englischsprachige Schülerinnen und Schüler der sechsten bis achten Klassen zeigte sich ein signifikanter Zusammenhang zwischen dem Wortschatz und dem Wortlesen, wobei dieser Zusammenhang bei den unauffälligen Leserinnen und Lesern größer ausfiel als bei denen, die Schwierigkeiten im Lesen hatten. Weiterhin war der Wortschatz ein signifikanter Prädiktor für das Leseverständnis, wobei auch hier der Zusammenhang bei den unauffälligen Leserinnen und Lesern größer war (Oslund, Clemens, Simmons & Simmons, 2018, S. 370–371). Für Jungen werden häufiger Schwierigkeiten im Lesen berichtet als für Mädchen (s. Kapitel 8.1.6). Gleichzeitig zeigen sich allerdings oft keine quantitativen Unterschiede im Wortschatz (McElvany et al., 2016, S. 47). Aufgrund der hohen Relevanz des Wortschatzes sind McElvany et al. (2016) der Frage nachgegangen, ob zwischen den beiden Geschlechtern qualitative Unterschiede im Wortschatz vorhanden sind, was sich eventuell negativ auf die Leseleistungen von Jungen auswirkt (McElvany et al., 2016, S. 48). Sie untersuchten in ihrer Studie Kinder der dritten Klasse hinsichtlich des Wortschatzes zu weiblich konnotierten (zum Beispiel aus dem Bereich Mode, Puppen oder Emotionen), männlich konnotierten (zum Beispiel aus dem Bereich Sport, Computer oder Geschichte) und neutral konnotierten Wörtern (McElvany et al., 2016, S. 48–49). Der Umfang des Wortschatzes unterschied sich nicht zwischen Jungen und Mädchen (McElvany et al., 2016, S. 51). In tiefergehenden Analysen zeigte sich, dass Kinder die jeweils mit dem eigenen Geschlecht konnotierten Wörter besser kannten als die mit dem anderen Geschlecht konnotierten. Es wird vermutet, dass Kinder entsprechend häufiger mit den dem eigenen Geschlecht zugeschriebenen Begriffen in Kontakt kommen. Eine Unter- oder Überschätzung des Wortschatzes könnte dann geschehen, wenn ein Test vornehmlich männlich oder weiblich konnotierte Wörter enthält (McElvany et al., 2016, S. 53). In der Praxis spielt dieser Unterschied auch bei der Erweiterung des Wortschatzes eine Rolle, da entsprechend aus den Bereichen, die eher dem anderen Geschlecht zugerechnet werden, Themen und somit auch Begriffe ausgewählt werden sollten (McElvany et al., 2016, S. 54).

#### 8.4 Zwischenfazit

Die in diesem Kapitel aufgeführten Faktoren zeigen auf, dass der Leseerwerb ein sehr komplexer Vorgang ist mit vielen möglichen Ursachen für Leseschwierigkeiten. Gleichzeitig sind die genauen Ursachen noch unbekannt. Sicher ist, dass es nicht eine einzelne Ursache gibt und dass die Ursachen sich auch je nach Erscheinungsbild (isolierte Lesestörung vs. kombinierte Lese-Rechtschreibstörung, Komorbiditäten etc.) unterscheiden können. Weiterhin ist oft unklar, ob ein Faktor als Ursache oder als Folge einzuschätzen ist.

Sicher ist, dass der Lesestörung eine genetische Komponente zugrunde liegt, auch wenn die hierfür verantwortlichen Gene noch nicht sicher ausgemacht werden konnten. Die

Wirkrichtung bei beeinträchtigten exekutiven Funktionen ist dagegen noch unklar. Häufig kommt es zu komorbiden Aufmerksamkeitsstörungen, wobei auch hier die Wirkrichtung nicht erforscht ist. Spracherwerbsstörungen gelten dagegen als großer Risikofaktor für die Entwicklung von Lesestörungen und Leseschwächen.

Auch wenn Jungen häufiger Leseschwierigkeiten zeigen, scheinen die Ursachen hierfür zumindest teilweise in anderen Faktoren verortet zu sein, wie in einer niedrigeren Lesemotivation und einem niedrigeren Leseselbstkonzept. Auch die soziale Herkunft an sich ist keine Ursache für Leseschwierigkeiten, aber ein Risikofaktor, vermittelt über eine oft weniger anregende Lernumwelt der Kinder. Der Migrationshintergrund als solcher wirkt ebenfalls einmal vermittelt über den sozialen Status und außerdem über die Sprachkenntnisse der Kinder. Unzureichende Deutschkenntnisse sind ein Risikofaktor für den Leseerwerb. Auch bei den Teilprozessen im Leseerwerb kann es zu Problemen kommen, die dann den weiteren Leseerwerb gefährden können. Besonders hervorzuheben sind an dieser Stelle die Buchstabenkenntnis, die Leseflüssigkeit sowie der Wortschatz.

## 9 Förderdiagnostische Aspekte beim Leseerwerb in der Grundschule

Eine Lehrkraft sollte möglichst früh bemerken, wenn ein Kind Schwierigkeiten beim Leseerwerb hat, um entsprechend zu reagieren, bevor sich diese Lernschwierigkeiten verfestigen können. Dabei sind Grundschullehrkräfte neben formeller auf semiformelle Diagnostik angewiesen, da diese auch während der üblichen Lehrtätigkeiten durchgeführt werden kann. Mit der Diagnosestellung ist der Prozess allerdings nicht abgeschlossen, da die Lehrkraft im Anschluss entscheiden muss, auf welche Art eine Förderung erfolgen sollte. Diese einzelnen Schritte werden in diesem Kapitel unter Berücksichtigung des Forschungsstands näher beleuchtet.

### 9.1 Diagnostik von Störungen beim Leseerwerb

Wenn man die definitorische Breite des Begriffs „Lese-Rechtschreibstörung“ sowie deren mögliche Ursachen betrachtet, wird schnell deutlich, dass dieser Begriff allein die konkrete Situation einer Schülerin oder eines Schülers nur schlecht darstellen kann. Daher ist es unerlässlich, Diagnostik durchzuführen und hierbei alle Ebenen des Leseprozesses zu berücksichtigen. Formelle Diagnostik ermöglicht dabei auch den Vergleich mit dem Lernstand anderer Schülerinnen und Schüler zu einem gegebenen Zeitpunkt, beispielsweise zum Ende der ersten Klasse. Daneben greifen Lehrkräfte im Schulalltag oft auch auf semiformelle Diagnostik, wie Beobachtungen während des Unterrichts, zurück (Schmidt, 2015, S. 103). In diesem Abschnitt sollen nun Indikatoren und Möglichkeiten zur Diagnostik erläutert werden. Die Begriffe informelle und semiformelle Diagnostik werden synonym verwendet, da analog zu Haschers Definition hier diejenigen diagnostischen Tätigkeiten gemeint sind, die nicht Bestandteil der formellen Diagnostik sind, aber auch „nicht nur zu impliziten Urteilen“ führen (Hascher, 2008, S. 75).

#### 9.1.1 Indikatoren für Schwierigkeiten beim Leseerwerb

Zu diagnostizierende Konstrukte, wie beispielsweise mathematische Kompetenz oder Lesekompetenz, sind nicht direkt beobacht- und erfassbar, da Konstrukte nur über mit diesen in Beziehung stehende und beobachtbare Merkmale erfasst werden können. Lehrkräfte müssen daher wissen, welche beobachtbaren Merkmale Indikatoren für Schwierigkeiten darstellen, um diese ausmachen zu können (Möller, Machts & Retelsdorf, 2016, S. 14).

In ihrer Interviewstudie arbeitete Schmidt (2015, S. 99–100) heraus, welche Merkmale Hauptschullehrkräfte beim Erfassen der Lesekompetenz ihrer Schülerinnen und Schüler berücksichtigen. Die Lehrkräfte fokussierten hauptsächlich auf die Leseflüssigkeit in Form der Betonung und des „flüssigen Vorlesens“ sowie auf das Leseverständnis im Sinne der Wiedergabe eines Textes in eigenen Worten. Teilweise gaben sie auch resignierende Einschätzungen und merkten an, dass man in der Sekundarstufe „nicht mehr kompensatorisch wirken könne“ (Schmidt, 2015, S. 101). Als weiteres Ergebnis hält Schmidt (2015, S. 102) die

Tendenz der Lehrkräfte fest, ein einmal gefälltes Urteil über die Lesefähigkeit einer Schülerin oder eines Schülers als „statisch“ anzusehen. Sie befragte die Lehrkräfte allerdings nur zu einem einzigen Zeitpunkt, sodass Langzeiteffekte über den Verlauf der Hauptschulzeit nicht erfasst wurden (Schmidt, 2015, S. 98). Mit dieser Studie zeigt sich abermals die Notwendigkeit, Lesekompetenz bereits in der Grundschule und somit vor Eintritt in die Sekundarstufe ausreichend zu fördern.

Bailey und Drummond (2006, S. 152) fassen zusammen, dass für Lehrkräfte generell häufig Faktoren wie Geschlecht, Sprache, ethnische Herkunft sowie fehlende akademische Fortschritte, Fehlverhalten und augenscheinliches Aufgeben auf Seiten der Schülerinnen und Schüler ausschlaggebend sind, um festzustellen, ob Interventionen und Förderung notwendig sind. In ihrer Studie konzentrierten sich die Autorinnen auf den Bereich des Leseerwerbs (Bailey & Drummond, 2006, S. 165). Insgesamt nahmen 17 Lehrkräfte aus 5 südkalifornischen Schulen an der Studie teil. Bei den Kindern handelte es sich um Vorschulkinder sowie um Erstklässlerinnen und -klässler. Die meisten Lehrkräfte nannten zwei bis drei Gründe, woran sie mögliche entstehende Leseschwächen festmachen. Die Aussagen lassen sich auf vier Faktoren reduzieren: Lesen, Schreiben, mündliche Sprache und Verhalten/Soziales, wobei dem Faktor „Lesen“ die meisten Aussagen zugeschrieben werden konnten (Bailey & Drummond, 2006, S. 162). Einige Lehrkräfte führten sehr spezifische Gründe an (z. B. „die Schülerin hat ein Sprachproblem, wodurch sie Schwierigkeiten hat, Laute am Anfang und am Ende eines Wortes auseinanderzuhalten, wie beispielsweise ‚b‘ und ‚p‘ bei ‚because‘“), während andere eher vage blieben. Dabei gab es kaum Unterschiede zwischen den einzelnen Lehrkräften in Abhängigkeit ihrer Hintergrundcharakteristika wie der Lehrerfahrung (Bailey & Drummond, 2006, S. 164–165). Ebenso zeigten sich auch kaum Unterschiede in Bezug auf die Charakteristiken der Schülerinnen und Schüler, wie beispielsweise das Geschlecht (Bailey & Drummond, 2006, S. 170–171). Nur wenige Lehrkräfte nannten formale Tests als Grundlage ihrer Vermutungen (Bailey & Drummond, 2006, S. 173).

Lehrkräfte sollten bei ihrer Diagnose einzelne Teilfähigkeiten im Leseprozess als Grundlage ihrer Einschätzungen nutzen (Fischer & Gasteiger-Klicpera, 2013, S. 67; Schmidt, 2015, S. 94). Dies ist wichtig, da beispielsweise die basale Lesefertigkeit zuerst ausreichend erworben werden muss, bevor sich das Leseverständnis entwickeln kann und sich die einzelnen Teilfertigkeiten auch gegenseitig beeinflussen können. Die Buchstabenkenntnis ist eine der grundlegenden Fähigkeiten für den Leseerwerb: Zum erfolgreichen Wortlesen müssen mindestens 17 Buchstaben bekannt sein (Juska-Bacher et al., 2015, S. 38). Neben mangelnder Buchstabenkenntnis nennen Fischer und Gasteiger-Klicpera (2013, S. 67–71) noch weitere Merkmale, die auf Schwierigkeiten im Leseerwerb hinweisen können: ein Verharren im logographischen Lesen, Probleme beim phonologischen Rekodieren, Schwierigkeiten beim Synthetisieren wie auch im lexikalischen Lesen sowie eine niedrige Lesegeschwindigkeit.

### 9.1.2 Semiformelle Diagnostik

Fischer und Gasteiger-Klicpera (2013, S. 70) gehen davon aus, dass aufgrund unterschiedlicher Herangehensweisen der Lehrkräfte im Anfangsunterricht häufig semiformelle Verfahren genutzt werden, um Teilfähigkeiten wie die Buchstabenkenntnis zu erfassen und um qualitative Aussagen über den Lernstand der Kinder machen zu können. Da die Kenntnisse der Kinder im Erstunterricht auch davon abhängen, wie schnell die Lehrkraft welche Buchstaben einführt, bietet sich die semiformelle Diagnostik besonders an. Hierüber kann die Lehrkraft dann den Lernstand über die Klasse hinweg erfassen und einzelne Kinder mit ihren Klassenkameradinnen und -kameraden vergleichen (Fischer & Gasteiger-Klicpera, 2013, S. 70–71). Anstelle von formeller Diagnostik spielt daher bei Grundschullehrkräften die semiformelle Diagnostik eine wichtige Rolle (Nuerk, Zürcher & Gawrilow, 2015, S. 235). Auch die IGLU-Studie 2016 kommt zu dem Schluss, dass bevorzugt semiformelle Diagnostik in Form von Klassenarbeiten und Beobachtungen als Grundlage genutzt wird und nur selten formelle objektive Verfahren (Bremerich-Vos et al., 2017, S. 283). Dies liegt nicht nur daran, dass Grundschullehrkräfte häufig angeben, keine Zeit für formelle Diagnostik zu haben. Vielmehr findet semiformelle Diagnostik täglich mehr oder weniger „nebenbei“ statt. Hausaufgaben werden kontrolliert, Schülerinnen und Schüler werden gebeten vorzulesen, die Lehrkraft stellt Fragen zu einem gelesenen Text und vieles andere mehr. Auch können Lehrkräfte selbst Tests erstellen, welche an die Inhalte des eigenen Unterrichts angepasst sind. Um den Lernverlauf zu überwachen, schlagen Hesse und Latzko (2017, S. 266) außerdem Lerntagebücher und Portfolios vor, die während des Unterrichts oder auch zuhause von den Kindern ausgefüllt werden können und der Lehrkraft so kontinuierlich Informationen über Fortschritte der Kinder geben können.

Der Nachteil an diesen semiformellen Verfahren ist, dass sie die diagnostischen Gütekriterien nicht erfüllen (Hesse & Latzko, 2017, S. 261). Semiformelle Diagnostik bezeichnet Schrader (2013, S. 161) dennoch als unverzichtbar in der Handlungssteuerung des Unterrichts, während formelle Diagnostik vor allem bei „Entscheidungen mit schwerwiegenden und weitreichenden Konsequenzen“ relevant ist. Semiformelle Diagnostik ist weiterhin aufgrund der Nähe zum eigenen Unterricht geeignet, frühzeitig Schwierigkeiten im Lernen zu erkennen (Hesse & Latzko, 2017, S. 261). Gleichzeitig muss sich die Lehrkraft darüber bewusst sein, dass diese auch sehr fehleranfällig ist und an geeigneten Stellen durch formelle Diagnostik ergänzt werden muss (Schrader, 2013, S. 155).

Um die Lesefähigkeit zu prüfen, kann eine Lehrkraft ein Kind auffordern, einen Text oder auch eine Wortliste vorzulesen, um so festzustellen, ob und welche Schwierigkeiten sich bei dieser Aufgabe beobachten lassen. Durch Vorlesen kann beispielsweise festgestellt werden, wie schnell ein Kind liest, ob es zu Verlesungen kommt und ob diese vom Kind bemerkt und korrigiert werden. Gleichzeitig kann eine Lehrkraft hierdurch auch ausmachen, ob das Kind bereits synthetisierend lesen kann, also ob Laute miteinander verbunden werden und ob Silben zum Erlesen von Wörtern genutzt werden können (Fischer & Gasteiger-Klicpera, 2013,

S. 69). Auch die Buchstabenkenntnis lässt sich über semiformelle Diagnostik erfassen. Ungefähr zehn Wochen nach Schulbeginn oder wenn mindestens acht Buchstaben eingeführt wurden, kann die Lehrkraft die bekannten Buchstaben aufschreiben und vom Kind nennen lassen oder umgekehrt die Buchstaben nennen und vom Kind aufschreiben lassen. Ebenso können mit den bekannten Buchstaben Wörter gebildet werden, welche das Kind vorlesen soll (Klicpera et al., 2017, S. 223).

Mit standardisierten Tests kann eine Lehrkraft erkennen, ob ein Kind das Gelesene auch versteht, sie kann aber kaum Rückschlüsse auf die Lesestrategien des Kindes ziehen. Dieses Problem kann dadurch gelöst werden, dass die Lehrkraft das Kind beobachtet und gezielt Fragen stellt, wie es auf die Lösung von Fragen zum Text gekommen ist (Klicpera et al., 2017, S. 229). Auch hier wird deutlich, dass semiformelle und formelle Diagnostik sich ergänzen können, sodass die Lehrkraft zu einem umfangreichen Gesamtbild über den Leistungsstand kommen kann.

### 9.1.3 Formelle Diagnostik

In der IGLU-Studie 2006 zeigte sich, dass Lehrkräfte gut ein Drittel der leseschwachen Schülerinnen und Schüler nicht korrekt erkennen und diese teilweise sogar als gute Leserinnen und Leser bezeichnen (Valtin et al., 2010, S. 76). Daher spielt die formelle Diagnostik eine ungemein wichtige Rolle, da keine Einschätzungen durch die Lehrkraft vorgenommen werden müssen, sondern Normen den Vergleich mit dem Leistungsstand anderer Kinder der gleichen Klassenstufe ermöglichen und so der Lehrkraft Rückmeldung darüber geben, ob Schwierigkeiten vorliegen (Schmidt, 2015, S. 94).

Es gibt zahlreiche diagnostische Tests, die eingesetzt werden können, um die Lesekompetenz von Schülerinnen und Schülern festzustellen. Berichtet werden hier nur drei in der Grundschule durchführbare Verfahren, die sich auf die Diagnostik von Lesestörungen beziehen und welche in der S3-Leitlinie zur Diagnostik und Behandlung von Kindern und Jugendlichen mit Lese-Rechtschreibstörung empfohlen werden (Schulte-Körne, Galuschka, & Deutsche Gesellschaft für Kinder und Jugendpsychiatrie, Psychosomatik und Psychotherapie e.V. (DGKJP), 2015, S. 29): Der Ein-Minuten-Leseflüssigkeitstest des SLRT-II, die Würzburger Leise Leseprobe sowie der Leseverständnistest für Erst- bis Sechstklässler ELFE 1-6. Die in die Leitlinie aufgenommenen Verfahren erfüllen alle in der gekürzten Fassung der DIN 33430 (Anforderungen an berufsbezogene Eignungsdiagnostik) enthaltenen Anforderungen (Schulte-Körne et al., 2015, S. 32).

#### 9.1.3.1 Ein-Minuten-Leseflüssigkeitstest des SLRT-II

Der Lese- und Rechtschreibtest SLRT-II beinhaltet zwei Verfahren: den Rechtschreibtest und den Ein-Minuten-Leseflüssigkeitstest (Moll & Landerl, 2010). Dieser Test besteht wiederum aus zwei Subtests: Eine Wortliste und eine Pseudowortliste, die vom Kind jeweils möglichst schnell, aber korrekt, mit einer Zeitvorgabe von jeweils einer Minute vorgelesen werden

sollen. Dabei enthalten beide Listen je 156 Wörter in acht Spalten mit aufsteigendem Schwierigkeitsgrad (Moll & Landerl, 2010, S. 25–26). Die Wortliste beginnt mit einfachen Wörtern (Haus, Fisch etc.) und ist somit auch für Leseanfänger bzw. schwache Leserinnen und Leser geeignet. Die Komplexität der Wörter steigt zunehmend an, die Wortfrequenz nimmt ab (Zündschlüssel, Krankenzimmer, Suppenteller etc.), während die Schriftgröße ebenfalls verringert wird (Moll & Landerl, 2010, S. 25). Die Pseudowortliste ist analog aufgebaut. Hier finden sich Wörter wie „tuful“, „sika“ oder „katu“. Die Buchstabenfolgen ergeben keine sinnvollen Wörter, entsprechen aber ansonsten denen von deutschen Schriftwörtern (Moll & Landerl, 2010, S. 26). Diese Wörter müssen über die Buchstaben und Silben, also auf dem indirekten Weg, erlesen werden. Es wird gezählt, wie viele Wörter bzw. Pseudowörter das Kind in der einen Minute jeweils korrekt vorliest. Normen liegen ab der zweiten Hälfte der ersten Klasse bis zum Altersbereich für junge Erwachsene vor (Moll & Landerl, 2010, S. 15). Da die Kinder laut vorlesen und die Testleitung auf Lesefehler achten muss, lässt sich dieser Test nur im Einzelsetting durchführen. Es liegt eine Parallelförmigkeit vor, sodass der Test ein zweites Mal durchgeführt werden kann, ohne dass Wiederholungseffekte das Ergebnis verfälschen (Moll & Landerl, 2010, S. 9).

#### 9.1.3.2 Würzburger Leise Leseprobe (WLLP-R)

Mit der Würzburger Leise Leseprobe kann die Lesegeschwindigkeit, eine Komponente der Leseflüssigkeit, erhoben werden. Es liegen Normen ab Ende der ersten Klasse bis zum Ende der vierten Klasse vor (Schneider, Blanke, Faust & Küspert, 2011, S. 25). Der Test kann als Gruppen- oder Einzeltest durchgeführt werden. Im Manual wird darauf hingewiesen, dass er sich als Screening für Schwierigkeiten versteht, zur Bestimmung von Leseschwächen aber ein Einzelverfahren hinzugezogen werden soll (Schneider et al., 2011, S. 8). Es wird pro Aufgabe ein Wort mit vier Bildern dargeboten. Dazu soll immer das zum Wort passende Bild angekreuzt werden. Auch hier werden die Kinder angewiesen, so schnell und genau wie möglich zu arbeiten (Schneider et al., 2011, S. 10). Es gibt eine Parallelförmigkeit, die es erlaubt, den Test ohne Wiederholungseffekte ein weiteres Mal durchzuführen (Schneider et al., 2011, S. 13).

#### 9.1.3.3 Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6)

Der Leseverständnistest für Erst- bis Sechstklässler ELFE 1-6 erfasst in der Papierversion das Leseverständnis auf Wort-, Satz- und Textebene. Nur in der PC-Version liegt ein weiterer Subtest für die Lesegeschwindigkeit vor. Der Test kann ab Ende der ersten Klasse bis zur sechsten Klasse eingesetzt werden, wobei er für die fünfte und sechste Klasse nur noch als Screening verwendet werden sollte (Lenhard & Schneider, 2006, S. 12). Er kann in der Papierversion auch als Gruppentest durchgeführt werden.

Um das Wortverständnis zu erfassen, werden den Kindern 72 Bilder präsentiert. Neben jedem Bild finden sich vier Wörter: drei Distraktoren und das Zielwort, welches unterstrichen

werden soll (Lenhard & Schneider, 2006, S. 32). Die Kinder haben in diesem Subtest drei Minuten Zeit, um die Wörter zu bearbeiten (Lenhard & Schneider, 2006, S. 19). Das Satzverständnis wird mit einzelnen Sätzen überprüft. Jeder Satz enthält eine Lücke mit fünf vorgegebenen Wörtern, aus denen das in den Satz passende ausgewählt werden muss (Lenhard & Schneider, 2006, S. 33). Insgesamt können bis zu 28 Sätze bearbeitet werden, wofür die Kinder drei Minuten Zeit bekommen (Lenhard & Schneider, 2006, S. 19). Im dritten Subtest, dem Test zum Textverständnis, werden 13 kurze Texte mit jeweils ein bis drei zugehörigen Aufgaben präsentiert. Die Kinder sollen die Texte mit den Aufgaben nacheinander lesen und bearbeiten. Die Aufgaben bestehen dabei jeweils aus einer Frage oder einem Teilsatz mit vier Antwortmöglichkeiten, wobei immer eine Antwort ausgewählt werden soll, die zu der Frage bzw. dem Satz passt. Insgesamt werden 20 Aufgaben vorgegeben (Lenhard & Schneider, 2006, S. 33). Die Kinder haben sieben Minuten Zeit, die Texte mit ihren Aufgaben zu bearbeiten (Lenhard & Schneider, 2006, S. 19).

## 9.2 Rahmenbedingungen zur Leseförderung

Bei der Förderung von Kindern mit Lese-Rechtschreibschwierigkeiten kann nicht mit kurzfristigen Erfolgen gerechnet werden, sodass entsprechende Förderung immer langfristig, über einen Zeitraum von mindestens 20 Wochen, angelegt sein muss (Balke-Melcher, Schuchardt, Wolpers & Mähler, 2016, S. 25; Ise, Engel & Schulte-Körne, 2012, S. 130; von Suchodoletz, 2007, S. 351). Auch in der Metaanalyse von Galuschka, Ise, Krick und Schulte-Körne (2014) zeigte sich, dass Interventionen mit einer Dauer von mehr als 12 Wochen eine höhere Effektstärke hatten als solche mit einer Dauer von maximal 12 Wochen (Galuschka et al., 2014, S. 6). Neben der Intensivität der Förderung und dem Zeitpunkt, wann ein Kind zum ersten Mal Leseförderung erhält, spielt auch das Setting, in welchem die Leseförderung stattfindet, eine wichtige Rolle für die Wirksamkeit der Maßnahme (Schwab & Gasteiger-Klicpera, 2014, S. 65).

Eine maßgebliche Aufgabe der Schule ist es, vorbeugend zu agieren, um das Auftreten von Lernstörungen möglichst abzuwenden. Dies wird in den einzelnen Bundesländern in entsprechenden Gesetzen festgehalten, wie beispielsweise § 2 VOSB (Verordnung über Unterricht, Erziehung und sonderpädagogische Förderung von Schülerinnen und Schülern mit Beeinträchtigungen oder Behinderungen im Landesrecht Hessen). Dieser führt vorbeugende Maßnahmen als Aufgabe der allgemeinen Schulen an. Im Unterricht sollen demnach differenzierte Arbeitsmaterialien sowie weitere Fördermaßnahmen eingesetzt werden, welche in Einzelsettings oder in Kleingruppen stattfinden können. Weiterhin sollen Eltern, aber auch Schülerinnen und Schüler, beraten und informiert werden. Eine Zusammenarbeit mit den Beratungs- und Förderzentren (BFZ) mit seinen Förderschullehrkräften auf der einen und mit außerschulischen sowie vorschulischen Fördereinrichtungen auf der anderen Seite ist ebenfalls vorgegeben. Viele Lehrende in der Sekundarstufe fassen den Deutschunterricht allerdings nicht

mehr als Ort zur Förderung grundlegender Lesefähigkeiten auf und sehen die Förderung in diesem Alter nur im außerschulischen Bereich (Kamzela, 2015, S. 66).

Wenn die Lehrkraft Schwierigkeiten bemerkt, stehen ihr mehrere Handlungsmöglichkeiten zur Verfügung. Oft hängen diese von den an der Schule vorhandenen Ressourcen, wie beispielsweise räumlichen oder finanziellen Möglichkeiten, ab (Buda, 2017, S. 183). So besteht oft das Angebot von zusätzlichen Förderstunden in der Schule. Individuelle Förderung kann im Unterricht sowohl durch die Deutsch-Lehrkraft selbst wie auch durch eine zusätzliche Lehrkraft erfolgen. Im eigenen Unterricht kann eine Lehrkraft besser auf die heterogenen Bedingungen in der Klasse eingehen, indem sie weniger Aufgaben (quantitative Differenzierung) und/oder leichtere Aufgaben (qualitative Differenzierung) für Schülerinnen und Schüler mit Schwierigkeiten bereitstellt. Laut der IGLU-Studie 2016 kommt es in knapp 46 Prozent der Grundschulen zur qualitativen Differenzierung in Form von unterschiedlichen Schwierigkeitsgraden der Aufgaben. Zusätzliche Unterstützung über quantitative Differenzierung bieten 86 Prozent an (Bremerich-Vos et al., 2017, S. 304). Eckerth (2015, S. 179) fand Hinweise darauf, dass eine Doppelbesetzung in Grundschulklassen zu einem stärkeren Einsatz von Differenzierung führt. Eine Doppelbesetzung im Leseunterricht ist allerdings keineswegs die Norm. Knapp drei Viertel der Lehrkräfte geben an, im Leseunterricht in der Grundschule auf keine zweite Lehrperson zurückgreifen zu können (Tarelli, Lankes, Drossel & Gegenfurtner, 2012, S. 142). Dieser Wert hat sich in der IGLU-Studie 2016 sogar noch weiter erhöht. Hier gaben 84 Prozent an, keine weitere Fachkraft zur Verfügung zu haben und bei zwei Dritteln gab es auch sonst keine Hilfslehrkraft im Leseunterricht (Valtin, 2017, S. 321).

Eine weitere Möglichkeit, auf Leseschwierigkeiten im eigenen Unterricht zu reagieren, ist die Anpassung desselben oder der Einsatz von Förderprogrammen, die für die Verwendung in der gesamten Klasse konzipiert sind. Schwab und Gasteiger-Klicpera (2014, S. 66) kritisieren, dass es im deutschsprachigen Raum kaum evidenzbasierte Förderprogramme gibt, die im regulären Schulunterricht durchgeführt werden können, während es im englischsprachigen Raum solche Programme gibt. Beispielhaft nennen sie die Programme „Reciprocal Teaching“ von Brown und Palincsar (1984) und Peer-Assisted Learning Strategies (Fuchs, Fuchs, Mathes & Simmons, 1997). Zur Förderung des Leseverständnisses gibt es für den deutschsprachigen Raum beispielsweise das Programm „Wir werden Textdetektive“, welches allerdings erst ab der fünften Klasse eingesetzt werden kann (Buick, Gold, Mokhlesgerami, Rühl & Souvignier, 2006).

Der oben bereits benannte Förderunterricht hat die Aufgabe, Schülerinnen und Schüler mit Schwierigkeiten in Deutsch – oder auch in Mathematik – gezielt zu fördern. Der IGLU-Studie 2016 zufolge gibt es Förderangebote für Deutsch an knapp 82 Prozent der Schulen. Spezielle Förderkurse bzw. spezielle Lesekurse außerhalb des Unterrichts für Kinder mit entsprechenden Schwierigkeiten werden an 77 Prozent der Schulen angeboten. In Kleingruppen erfolgende Leseförderung durch ehrenamtlich tätige Eltern oder Großeltern findet an 59 Prozent der Schulen statt. Einen Zusatzunterricht in Deutsch gibt es an knapp 54 Prozent der Schulen

(Bremerich-Vos et al., 2017, S. 303–304). Während in den Ergebnissen der IGLU-Studie 2016 Lehrkräfte bei etwa 17 Prozent ihrer Schülerinnen und Schüler Förderbedarf im Lesen sehen, erhalten allerdings nur ca. 11 Prozent der Grundschülerinnen und -schüler diesen schulischen Förderunterricht. Insgesamt nehmen sogar zwei Drittel der laut dem Ergebnis des IGLU-Lesetests förderbedürftigen Kinder an keiner schulischen Förderung teil (Bos et al., 2017, S. 25–26).

Die Wirksamkeit von Förderkursen ist bisher kaum untersucht und wenn, dann konnte diese nicht nachgewiesen werden (von Suchodoletz, 2007, S. 351, 2010, S. 335). Vor allem für Kinder mit gravierenden Schwierigkeiten im Lesen und/oder im Schreiben und nicht nur kurzzeitigen Rückständen, sind diese daher vermutlich nicht ausreichend. Von Suchodoletz (2007, S. 351) empfiehlt daher zusätzlich außerschulische Förderung.

Nachhilfeinstitute oder private Nachhilfestunden sind laut der IGLU-Studie 2011 eine häufig von Eltern gewählte Methode zur Förderung ihrer Kinder. Diese Form der Förderung kommt bei Kindern mit Leseschwierigkeiten sogar häufiger vor als innerschulische Förderung (Valtin et al., 2010, S. 77). Generell ist bei dieser Art an Förderung von Vorteil, dass sie auf das Kind zugeschnitten sein kann. Allerdings ist fraglich, ob und wie gut Rücksprache mit der Deutsch-Lehrkraft erfolgt. Zudem kann Nachhilfe im privaten Bereich nicht auf ihre Qualität hin kontrolliert werden (Valtin et al., 2010, S. 85).

Von Suchodoletz (2010, S. 335) plädiert aufgrund der fehlenden Wirksamkeitsnachweise schulischer Förderkurse dafür, bei Vorliegen einer Lese-Rechtschreibstörung möglichst früh entsprechende außerschulische Lerntherapie zu suchen. Auch Buda (2017, S. 191) hält fest, dass regulärer Förderunterricht kein Ersatz für eine individuelle Lerntherapie bzw. Einzelförderung sein kann. Außerschulische Förderung erscheint notwendig, weil innerhalb der Schule auch im Förderunterricht oft nicht genügend Zeit für alternative Übungsmethoden vorhanden ist (Balke-Melcher et al., 2016, S. 18). Hier ist vor allem die Beratung der Eltern durch die Lehrkraft notwendig, um diese auf die Möglichkeit einer außerschulischen Lerntherapie hinzuweisen.

Lipka und von Orloff (2014, S. 73) merken kritisch an, dass es für den Begriff der (integrativen) Lerntherapie keine „rechtliche Verankerung“ gibt und auch keine Zulassungsvoraussetzungen, um den Beruf auszuüben. An dieser Stelle soll daher kurz das Berufsbild in Anlehnung an den Fachverband für integrative Lerntherapie e. V. (FiL) erläutert werden. Integrative Lerntherapie unterscheidet sich von anderen Maßnahmen wie Nachhilfe dadurch, dass diese von entsprechend qualifizierten Fachkräften im Einzelsetting durchgeführt wird und sowohl psychotherapeutische wie auch fachdidaktische Komponenten enthält (Bender et al., 2017, S. 65). Es findet zu Beginn und begleitend quantitative wie qualitative Diagnostik Anwendung (Lipka & von Orloff, 2014, S. 73). Darauf basierend werden individuelle Lernpläne erstellt, dokumentiert und fortgeführt. Ziele und Erwartungen an die integrative Lerntherapie werden sowohl vom Kind, von den Eltern und idealerweise auch von den jeweiligen Lehrkräften einbezogen. Die Lerntherapie erfolgt über eine Integration von Methoden aus der Pädagogik,

der Psychologie, der Medizin sowie weiteren Fachbereichen wie der Linguistik und den entsprechenden Fachdidaktiken (Bender et al., 2017, S. 65–66). Neben einer Verbesserung der schulischen Leistungen stehen auch eine Stärkung der Psyche des Kindes sowie die soziale Integration im Fokus. Der Fachverband sieht weiterhin ein abgeschlossenes pädagogisch oder psychologisch ausgerichtetes Hochschulstudium oder eine entsprechende Berufsausbildung mit mehrjähriger Berufspraxis zusätzlich zu einer Weiterbildung in der Lerntherapie als Zugangsvoraussetzung für den Beruf (Lipka & von Orloff, 2014, S. 73–74).

Bender et al. (2017, S. 68) führen an, dass eine integrative Lerntherapie für die Kinder relevant ist, die auch nach schulischen Fördermaßnahmen weiterhin Schwierigkeiten im Lernen aufweisen und die den Anschluss an ihre Mitschülerinnen und Mitschüler verloren haben oder wo dies droht. In der Studie von Schulz, Dertmann und Jagla (2003, S. 239) wurde nach deren Aussagen zum ersten Mal die Wirksamkeit der integrativen Lerntherapie überprüft. Es nahmen insgesamt 74 Kinder mit Lese-Rechtschreibschwierigkeiten an der Studie teil, wobei alle eine solche erhielten (Schulz et al., 2003, S. 234). Im Vergleich zu unauffälligen Kindern verfügten diese in der kognitiven Komponente über ein niedrigeres Selbstwertgefühl. Ebenso berichteten die Eltern häufiger über internalisierende Verhaltensstörungen (Schulz et al., 2003, S. 240). Sowohl in der genannten kognitiven Komponente des Selbstwertgefühls wie auch in den von den Eltern berichteten Verhaltensstörungen zeigten sich nach einem halben bis einem Jahr signifikante Verbesserungen. Durchschnittlich blieben die Kinder zwei Jahre in der Lerntherapie. Knapp die Hälfte der Kinder verbesserte sich in der Rechtschreibleistung um ca. eine Schulnote im Vergleich zur jeweiligen Klassennorm. Da in dieser Studie keine Kontrollgruppe vorhanden war, kann allerdings nur von Hinweisen auf eine Wirksamkeit gesprochen werden (Schulz et al., 2003, S. 241). Nach insgesamt zwei bis zweieinhalb Jahren wurde eine weitere Erhebung durchgeführt, um den Verlauf nach regulärer Beendigung der integrativen Lerntherapie zu überprüfen. An dieser Erhebung nahmen noch 51 Kinder teil (Bövers & Schulz, 2005, S. 193). Es zeigte sich, dass die kognitive Komponente des Selbstwertgefühls nur in der ersten Zeit nach Anlauf der Lerntherapie steigt und die Hälfte der Kinder nach Abschluss der Lerntherapie immer noch ein sehr niedriges Selbstwertgefühl hat. Die positiven Veränderungen in den von den Eltern berichteten Verhaltensstörungen hielten dagegen über den Verlauf der gesamten Lerntherapie an (Bövers & Schulz, 2005, S. 197). Mehr als 80 Prozent der Kinder verbesserten sich zudem deutlich in der Rechtschreibleistung (Bövers & Schulz, 2005, S. 199). Über die Leseleistung wird in beiden Studien nicht berichtet. Tornow (2012) berichtet über eine quasi-experimentelle Querschnittsuntersuchung zur Wirksamkeit der integrativen Lerntherapie, welche in Berlin stattgefunden hat. Es wurden neue Fälle mit Kindern verglichen, welche die Lerntherapie beendeten (Tornow, 2012, S. 5). In die Auswertung sind Daten von 194 neuen und 108 abgeschlossenen Lerntherapien eingeflossen. Es ergab sich eine Abbruchquote von neun Prozent, welche der Autor als niedrig einschätzt, sowie eine Verweildauer von 24 Monaten im Median bei gleichzeitig sehr hoher

Wirksamkeit in den Bereichen der schulischen Leistungen sowie der psychischen Stabilisierung durch integrative Lerntherapie (Tornow, 2012, S. 24).

Ricken (2014, S. 182) und Hülsmann (2018, S. 138) führen an, dass eine engere Zusammenarbeit von Schulen mit Lerntherapeutinnen und -therapeuten, zum Beispiel in Form von innerschulischer Lerntherapie, angestrebt werden sollte. Als Vorzug nennen sie, dass die Lerntherapeutinnen und -therapeuten, wenn sie innerhalb der Schule tätig sind, bei Auffälligkeiten sofort tätig werden können und Schülerinnen und Schüler so schneller unterstützen können. Weitere Vorteile innerschulischer Lerntherapie bestehen ferner darin, dass finanziell schwächere Kinder nicht benachteiligt werden und dass zwischen Lehrkräften und Lerntherapeutinnen und -therapeuten auf diese Weise eine viel engere Zusammenarbeit stattfinden kann (Balke-Melcher et al., 2016, S. 19).

Balke-Melcher et al. (2016, S. 19) überprüften in ihrer Studie, ob sich die Leistungen von Grundschulkindern über den Verlauf der dritten bis zur vierten Grundschulklasse verbessern, wenn diese über die beiden Schuljahre mit innerschulisch stattfindenden Lerntherapien gefördert wurden, ob Kinder mit isolierten oder mit kombinierten Störungen gleichermaßen von dieser Form der Lerntherapie profitieren und ob es zwischen Lernschwächen und Lernstörungen einen Unterschied gibt in der Wirksamkeit der innerschulischen Lerntherapie. Neben der Förderung der Kinder lag ein weiteres Ziel der Intervention darin, die Lehrkräfte durch die Unterstützung in den innerschulischen Lerntherapien zu entlasten. Durchgeführt wurde die Untersuchung mit 105 Kindern an neun Schulen, die über zwei Jahre jeweils zwei Schulstunden pro Woche gefördert wurden. Es wurde geprüft, dass weder die Kinder, die an der innerschulischen Lerntherapie teilnahmen, noch die Kinder in der Kontrollgruppe außerschulische Fördermaßnahmen erhielten. Nach zwei Jahren hatten in beiden Gruppen gleich viele Kinder ihre Leseschwierigkeiten überwunden. Weder bei isoliert leseschwachen noch bei kombiniert schwachen Kindern zeigte sich eine stärkere Verbesserung in der Leseleistung im Vergleich zur Kontrollgruppe (Balke-Melcher et al., 2016, S. 23). Für die Rechtschreibschwierigkeiten ergab sich dagegen ein signifikanter Unterschied zugunsten der Kinder mit der innerschulischen Lerntherapie. Mehr als doppelt so viele Kinder hatten ihre Rechtschreibschwierigkeiten überwunden (Balke-Melcher et al., 2016, S. 20–21).

Nach den zwei Jahren hatten von den geförderten Kindern noch mehr als die Hälfte der leseschwachen und mehr als zwei Drittel der rechtschreibschwachen Kinder unterdurchschnittliche Leistungen (Balke-Melcher et al., 2016, S. 25). Folglich hatten sich mehr leseschwache als rechtschreibschwache Kinder verbessert. Es wird daher vermutet, dass das Lesen schulintern auch auf anderen Wegen gefördert werden kann. Zusammenfassend zeigt diese Studie, dass schulinterne Lerntherapie förderlich und sinnvoll sein kann, aber dennoch keine Einzeltherapie ersetzt (Balke-Melcher et al., 2016, S. 26).

Auch das Elternhaus kann in die Förderung einbezogen werden. Nach Wandeler, Niggli, Trautwein und Villiger Hugo (2013, S. 2) ist die Schule auf die Unterstützung im Elternhaus sogar angewiesen. Villiger, Niggli, Wandeler und Kutzelmann (2012, S. 90) halten

zusammenfassend fest, dass kooperativ zwischen Schule und Elternhaus stattfindende Förderung effektiver ist als rein schulische Fördermaßnahmen. Wandeler et al. (2013) untersuchten in ihrer Längsschnittstudie in der deutschsprachigen Schweiz, ob Hausaufgabenbetreuung durch Eltern eine effektive Begleitmaßnahme in der Leseförderung sein kann. Sie führen an, dass das Verhalten von Eltern auf der einen Seite lernförderlichen Einfluss haben kann, aber auch beeinträchtigende Effekte auftreten können, weswegen sie der Frage nachgingen, ob ein Elterntraining positive Effekte auf die Lesekompetenz von Kindern der vierten Klasse haben kann (Wandeler et al., 2013, S. 2). Das Elterntraining fand an zwei Zeitpunkten über jeweils drei Stunden statt. Beim zweiten Termin waren die Kinder ebenfalls anwesend und es wurden gemeinsam Lesestrategien eingeübt (Wandeler et al., 2013, S. 4). Insgesamt führte das Elterntraining zu erhöhter emotionaler Unterstützung, aber nicht zur Abnahme von Einmischung, im Sinne einer unaufgeforderten Unterstützung durch die Eltern. Die von den Kindern wahrgenommene Einmischung der Eltern stand mit einer niedrigeren Lesekompetenz zum zweiten Messzeitpunkt in Zusammenhang. Die Leseleistung wurde durch die elterliche Unterstützung allerdings nicht gefördert (Wandeler et al., 2013, S. 10–11). Das Elterntraining hatte allerdings nur an zwei Abenden stattgefunden und fiel somit sehr kurz aus. Die positiven Effekte auf die emotionale Unterstützung zeigen dagegen, dass auch ein solch kurzes Elterntraining bereits wirkungsvoll sein kann. Wandeler et al. (2013, S. 12) schlagen daher Auffrischungsstunden und Vertiefungskurse für die Eltern vor. Weiterhin weisen diese Ergebnisse darauf hin, dass durch solche Maßnahmen das Elternhaus eine wichtige Rolle einnehmen kann.

Multhauf, Buschmann und Soellner (2016) führten eine gruppenbasierte kognitive Verhaltenstherapie für Mütter durch, deren Kinder sich in der dritten Klasse befanden und welche Lese- und/oder Rechtschreibschwierigkeiten aufwiesen. Inhalte waren Psychoedukation über die Lese- und Rechtschreibschwierigkeiten der Kinder, kognitive Umstrukturierung in Bezug auf Schuldgefühle und überhöhte Erwartungen gegenüber den schulischen Leistungen der Kinder, Empathieförderung für die Schwierigkeiten der Kinder sowie Förderung der Selbstwirksamkeit der Mütter (Multhauf et al., 2016, S. 1209–1210). Die Intervention lief über drei Monate und zeigte eine verzögerte Wirkung, da sich im Posttest keine Unterschiede zur Wartekontrollgruppe ergaben (Multhauf et al., 2016, S. 1218). Im Follow-Up Test weitere drei Monate später wiesen die Mütter allerdings ein geringeres Stressniveau und erhöhte Erziehungskompetenzen auf (Multhauf et al., 2016, S. 1217).

Kopp et al. (2014) boten Feriencamps zur Leseförderung von Kindern mit Deutsch als Zweitsprache an. Es fand jeden Tag zwei bis drei Stunden Leseförderung statt. Für Kinder ab der dritten Klasse wurden sowohl basale Lesefähigkeiten wie auch Lesestrategien trainiert (Kopp et al., 2014, S. 147–148). Zum Ende des Feriencamps wie auch fünf Monate später zeigte sich eine signifikant bessere Leseleistung (Kopp et al., 2014, S. 149). Dies deutet darauf hin, dass auch unter diesen Rahmenbedingungen eine Leseförderung wirksam sein kann.

### 9.3 Inhaltliche Aspekte der Leseförderung

Neben einer Entlastung in der Schule, zum Beispiel über Differenzierung in Form von weniger oder leichteren Aufgaben, muss gleichzeitig auch ein Kompetenzaufbau erfolgen, der am Lernstand des Kindes ansetzt (Ricken, 2014, S. 181). Eine generelle Empfehlung zur Verbesserung der Leseleistung ist die Anpassung der Lesematerialien, indem die Schrift und die Zeilenabstände vergrößert werden (Schulte-Körne et al., 2015, S. 51).

Die oben bereits genannte S3-Leitlinie zur Diagnostik und Behandlung von Kindern und Jugendlichen mit Lese-Rechtschreibstörung empfiehlt, die Symptome in den Fokus zu rücken. Das bedeutet, dass diejenigen Methoden verwendet werden sollen, die „an der verminderten Lese- und / oder Rechtschreibleistung ansetzen“ bzw. an Fähigkeiten, die direkt mit dem Lese- und Rechtschreiberwerb in Beziehung stehen (Schulte-Körne et al., 2015, S. 34–35).

Um die Wirksamkeit verschiedener Fördermaßnahmen zu überprüfen, wurde von Galuschka et al. (2014) eine Metaanalyse durchgeführt. Diese berücksichtigte insgesamt 22 randomisierte kontrollierte Studien zu Fördermaßnahmen bei Leseschwierigkeiten (Galuschka et al., 2014, S. 1–2). Eine weitere Metaanalyse, auf deren Ergebnisse sich die S3-Leitlinie ebenfalls stützt, berücksichtigt 28 Studien (Ise et al., 2012, S. 123). Im Gegensatz zur Metaanalyse von Galuschka et al. (2014) sind bei der Metaanalyse von Ise et al. (2012, S. 123) nur deutschsprachige Studien einbezogen worden, gleichzeitig wurden auch nicht-randomisierte Studien berücksichtigt. Die Empfehlungen der S3-Leitlinie berücksichtigen die eingeschränkte Übertragbarkeit von Studien aus nicht-deutschsprachigen Ländern und gewichten bei voneinander abweichenden Ergebnissen Studien aus deutschsprachigen Ländern entsprechend stärker (Schulte-Körne et al., 2015, S. 34).

Für kausale Therapien, die an Grundfunktionen ansetzen, welche an der Entstehung von Lese-Rechtschreibschwierigkeiten beteiligt sein können, konnte keine Wirksamkeit nachgewiesen werden. Dazu zählen auditive und visuelle Wahrnehmungs- sowie Verarbeitungstrainings, medikamentöse und alternativmedizinische Verfahren sowie Irlen-Linsen (Galuschka et al., 2014, S. 10; Ise et al., 2012, S. 130; Schulte-Körne et al., 2015, S. 35). Diese werden hier entsprechend nicht weiter berücksichtigt.

Im Folgenden wird zunächst auf Fördermaßnahmen der basalen Lesefertigkeiten eingegangen, die grundlegend erworben werden müssen, bevor sinnentnehmendes Lesen möglich wird. Im Anschluss erfolgt ein Blick auf die Förderung von sinnentnehmendem Lesen, wobei der Schwerpunkt dieser Förderung in der Sekundarstufe I zu finden ist.

#### 9.3.1 Förderung der basalen Lesefertigkeiten

In der S3-Leitlinie wird zunächst die Förderung auf Phonem- und Silbenbasis als wirkungsvolle Fördermaßnahme für basale Lesefertigkeiten empfohlen, vor allem, wenn die Kinder und Jugendlichen in diesen Bereichen Schwierigkeiten haben. Hierzu zählen neben der systematischen Instruktion der Graphem-Phonem-Korrespondenz wie auch der Phonem-Graphem-Korrespondenz die Segmentierung einzelner Wörter in Phoneme und Morpheme

sowie in Silben, Onset oder Reim. Ebenso zählt auch das Verbinden einzelner Phoneme zu einem Wort dazu (Schulte-Körne et al., 2015, S. 36). Die Förderung durch die Ganzwortmethode, bei der ganze Wörter wiederholt gelesen werden sollen, wird dagegen nicht empfohlen bzw. sollte, wenn sie eingesetzt wird, nicht als alleinige Maßnahme durchgeführt werden, da eine Wirksamkeit nicht nachgewiesen werden konnte (Schulte-Körne et al., 2015, S. 38).

Phonologietrainings, welche die phonologischen Fertigkeiten fördern sollen, sollten nicht als alleinige Fördermaßnahme angewendet werden, da sich für den deutschsprachigen Raum bisher keine Wirksamkeit nachweisen ließ (Schulte-Körne et al., 2015, S. 41). Ise et al. (2012, S. 132), vermuten, dass dies daran liegt, dass deutschsprachige Kinder sich die benötigten phonologischen Kenntnisse im Verlauf der Grundschule aneignen, sodass eine Förderung dieser keine weiteren Effekte bringt.

Nach Galuschka et al. (2014, S. 9) stellen die systematische Instruktion von Buchstaben-Laut-Korrespondenzen sowie von Dekodierstrategien die effektivsten Fördermethoden dar. Allerdings kann diese Metaanalyse keine Aussage darüber treffen, ob dies auch spezifisch für den deutschsprachigen Raum gilt, da die Studien in anderen Ländern stattgefunden haben.

Ein isoliertes Leseflüssigkeitstraining ist dagegen keine effektive Methode, um Lesen zu fördern. Ein Phonics-Training, welches ein Leseflüssigkeitstraining mit der Förderung phonologischer Bewusstheit kombiniert, gilt dagegen als nachweislich effektiv (Galuschka et al., 2014, S. 9). Es muss allerdings ebenfalls wieder berücksichtigt werden, dass die zugrundeliegenden Studien aus anderen Ländern stammen und die Ergebnisse somit nicht unbedingt auf den deutschsprachigen Raum übertragen werden können. Müller, Križan, Hecht, Richter und Ennemoser (2013) führten eine Studie an hessischen Grundschulen durch, in der drei Arten der Leseförderung und deren Effekte auf die Leseflüssigkeit untersucht wurden: ein Phonics-Training, ein Leseflüssigkeitstraining sowie ein Lesestrategietraining (Müller et al., 2013, S. 131). Als Phonics-Training kam eine Lautiermethode zum Einsatz, sodass nicht nur die phonologische Bewusstheit, sondern auch explizit die Graphem-Phonem-Korrespondenz bei geschriebenen Wörtern gefördert wurde (Müller et al., 2013, S. 132). Die zweite untersuchte Methode war die des wiederholten Lesens, wobei Textabschnitte zunächst vorgelesen und dann gemeinsam chorisch gelesen wurden, bevor der Abschnitt in Zweiergruppen erneut wiederholt gemeinsam und schließlich vom trainierten Kind allein laut gelesen wurde (Müller et al., 2013, S. 134). Im durchgeführten Lesestrategietraining wurden drei Lesestrategien vermittelt, welche eigentlich hierarchiehöhere Leseprozesse einüben sollten, wodurch angenommen wurde, dass nur indirekt eine Förderung der Leseflüssigkeit erfolgen sollte (Müller et al., 2013, S. 133). Es wurden die jeweils fünf schwächsten Leserinnen und Leser aus den Jahrgangsstufen zwei und vier ausgewählt, welche jeweils 10 Wochen lang eines der zuvor genannten Trainings oder – als Kontrollgruppe – kein Training erhielten. Die Stichprobe bestand aus 58 Kindern in der zweiten und 51 Kindern in der vierten Klasse. Die Zuteilung zu den drei Trainings bzw. zur Kontrollgruppe erfolgte klassenweise randomisiert. Die jeweils fünf besten

Kinder in den Klassen wurden zudem als Tutorinnen und Tutoren eingesetzt. Auf diese Weise konnten die Trainings als peer-gestützte Maßnahmen umgesetzt werden (Müller et al., 2013, S. 134). Es wurde postuliert, dass Kinder in der zweiten Klasse durch die Förderung einen größeren Zuwachs in der Leseflüssigkeit zeigen sollten als Kinder der vierten Klasse, da hierarchieniedrige Prozesse wie die Leseflüssigkeit in der vierten Klasse, auch bei schwachen Kindern, bereits einigermaßen routiniert sein sollten (Müller et al., 2013, S. 133). Es wurde weiterhin vermutet, dass das Lesestrategietraining in der zweiten Klasse zu keiner Erhöhung der Leseflüssigkeit führt, da das Arbeitsgedächtnis der Kinder mit dem Erlesen der Wörter bereits ausgelastet ist und das zusätzliche Strategietraining eher überlastet, als dass es die Leseflüssigkeit trainieren könnte. In der vierten Klasse sollte dagegen die Worterkennung schon ausreichend routiniert sein, sodass die kognitiven Ressourcen für die Lesestrategien zur Verfügung stehen sollten (Müller et al., 2013, S. 133–134). In der abschließenden Testung nach Abschluss des zehnwöchigen Trainings wurde deutlich, dass sich die Leseflüssigkeit in beiden Klassenstufen jeweils gleich schnell entwickelt hatte. Dies entsprach nicht der Annahme der Autorinnen und Autoren, dass die Entwicklung in der vierten Klasse schneller verlaufen würde als in der zweiten. Zusätzlich zeigte sich entsprechend ihrer Erwartungen in der vierten Klasse ein stärkerer Anstieg der Leseflüssigkeit im Vergleich zur Kontrollgruppe für alle drei Trainingsmethoden. In der zweiten Klasse führten dagegen lediglich das Leseflüssigkeitstraining sowie das Lesestrategietraining zu einer Beschleunigung der Leseflüssigkeitsentwicklung (Müller et al., 2013, S. 140). Für die zweite Klasse waren für das Phonics-Training eine Beschleunigung und für das Lesestrategietraining keine Beschleunigung erwartet worden. In der Studie konnte dagegen gezeigt werden, dass hierarchiehöhere Prozesse einen positiven Einfluss auf hierarchieniedrigere Prozesse wie die Leseflüssigkeit haben können und dass es einen Zusammenhang zwischen diesen Prozessen gibt (Müller et al., 2013, S. 141).

Eine weitere Studie kam zu dem Ergebnis, dass die Fähigkeit zur Worterkennung einen moderierenden Einfluss auf die Wirksamkeit von Förderprogrammen zur Steigerung des Leseverständnisses hatte (Müller et al., 2015, S. 59). Es ergab sich insgesamt kein Effekt der Förderung auf die Lesekompetenz der Kinder, wohl aber ein Interaktionseffekt der Worterkennung in dem Sinne, dass schwache Leserinnen und Leser dann von der Förderung profitierten und sich im Leseverständnis verbessern konnten, wenn sie auch über eine gute Worterkennung verfügten (Müller et al., 2015, S. 59).

Eine weitere Fördermaßnahme stellen Lautleseverfahren dar. Das Ziel von Lautleseverfahren ist die Verbesserung der Leseflüssigkeit (Fadler & Böhme, 2018, S. 4). Diese Verfahren sind allerdings nicht mit dem üblichen Reihum-Lesen im Unterricht gleichzusetzen, für das es keine Wirksamkeitsnachweise gibt (Stahns & Rieser, 2018, S. 134). Bei den Lautleseverfahren können zwei Grundformen unterschieden werden: das wiederholte Lautlesen und das begleitende Lautlesen. In der ersten Form wird ein Text wiederholt gelesen. In der zweiten Form lesen je ein schwaches und ein starkes Kind einen Text gemeinsam halblaut vor (Fadler

& Böhme, 2018, S. 4). Laut den Autorinnen werden die Lesegenauigkeit und Lesegeschwindigkeit, Teilaspekte der Leseflüssigkeit, über Lautleseverfahren gefördert, wobei dem kooperativen Lesen ein besonderes Augenmerk gilt (Fadler & Böhme, 2018, S. 4).

Über die Verbreitung von Lautleseverfahren im deutschen Unterricht ist allerdings nicht viel bekannt. Die Autorinnen vermuten, dass die Lautleseverfahren in der Grundschule wohl nur in geringem Maße eingesetzt werden und sehen als Ursache einen niedrigen Bekanntheitsgrad dieser Methode (Fadler & Böhme, 2018, S. 5). Laut der IGLU-Studie 2016 kommen Lautlese-Tandems generell bei etwas über der Hälfte der Kinder in der vierten Klasse vor. Nur bei 20 Prozent der befragten Lehrkräfte wird dieses Verfahren allerdings mindestens ein- bis zweimal die Woche angewendet (Bremerich-Vos et al., 2017, S. 287). Die Wahl für dieses Vorgehen ist dabei unabhängig vom Anteil schwacher Leserinnen und Leser in der Klasse (Bremerich-Vos et al., 2017, S. 288).

Ein reines Viellesen, bei dem Kinder für sich große Mengen an Texten oder Büchern lesen sollen, führt zu keinen Verbesserungen der Leseleistungen bei schwächeren Schülerinnen und Schülern und fördert nur starke Leserinnen und Leser in ihren Leseleistungen (Hochstadt et al., 2015, S. 122; Rosebrock & Nix, 2015, S. 65–66). Vielleseverfahren können aber eingesetzt werden, um Kinder an das Lesen von Büchern heranzuführen und die Lesemotivation zu steigern (Rosebrock & Nix, 2015, S. 66–67). Wichtig ist allerdings, dass das Lesen „Selbstzweck“ ist und zum Beispiel in freien Lesezeiten in der Schule stattfinden kann. Das Gelesene sollte zudem nicht im Unterricht besprochen werden (Stahns & Rieser, 2018, S. 134).

### 9.3.2 Förderung des Leseverständnisses

Selbst bei ausreichender Leseflüssigkeit kann es vorkommen, dass Kinder Schwierigkeiten haben, sinnentnehmend zu lesen. Daher sollte dem Leseverständnis ebenfalls Beachtung geschenkt werden. Eine ausschließliche Verwendung von Textverständnisstrategien sollte allerdings nicht zur Förderung von Kindern und Jugendlichen mit Lese-Rechtschreibschwierigkeiten angewendet werden. Als Grund führt die S3-Leitlinie an, dass Studien bisher keine signifikanten Effekte nachweisen konnten bzw. dass einige Studien nicht ausschließlich Kinder mit Lese-Rechtschreibschwierigkeiten beinhalteten, sodass über die Wirksamkeit für diese bestimmte Zielgruppe keine Aussage gemacht werden kann (Schulte-Körne et al., 2015, S. 39–40).

Stahns und Rieser (2018, S. 134) empfehlen eine explizite Vermittlung von kognitiven und metakognitiven Lesestrategien. Zu diesen zählen Strategien wie das Zusammenfassen wichtiger Informationen in einem Text, das Aktivieren von Vorwissen zu einem Text bzw. das Verknüpfen mit dem Vorwissen oder das Fragenstellen zum Text. Hierzu analysierten Stahns und Rieser (2018, S. 136) Daten der IGLU-Studie 2011 im Hinblick auf die Unterrichtsgestaltung von Lehrkräften und auf die Frage, ob diese mit der Leseleistung der Kinder zusammenhängt. Die Leseleistung in den Klassen war umso niedriger, je häufiger Lesestrategien instruiert wurden. Eine mögliche Erklärung für dieses konterintuitive Ergebnis sehen Stahns und

Rieser in der Ausgangssituation: Eventuell wenden Lehrkräfte häufiger Lesestrategien an, je niedriger die ursprünglichen Leseleistungen in der Klasse sind, sodass die niedrige Leseleistung nicht das Ergebnis der Anwendung der Lesestrategien ist, sondern der Grund, warum diese häufiger verwendet werden. Ebenfalls möglich ist, dass die Schwierigkeiten im Lesen auf hierarchieniedrigerer Ebene zu finden sind und die Schülerinnen und Schüler mit der Vermittlung von Lesestrategien überfordert werden, da sie zunächst eine ausreichende Leseflüssigkeit erwerben müssen (Stahns & Rieser, 2018, S. 142).

#### 9.4 Schlussfolgernd: Anforderungen an die Lehrkraft

Lehrkräfte müssen in ihrer Klasse mit einer sehr heterogenen Lerngruppe umgehen. Dabei ist es wichtig, die Lernfortschritte aller Kinder im Auge zu behalten, um zu bemerken, wenn einzelne Kinder Schwierigkeiten im Leseerwerb zeigen. Hierfür bietet sich vor allem im Anfangsunterricht semiformelle Diagnostik an, da diese sehr gut auf den eigenen Unterricht angepasst werden kann. Dennoch sollte eine Lehrkraft zusätzlich formelle Diagnostik heranziehen, um den Lernstand von Kindern mit einer Norm außerhalb der eigenen Klasse vergleichen zu können. Die Lehrkraft muss entsprechend über ausreichende diagnostische Expertise verfügen, um den Test korrekt durchführen, auswerten und interpretieren zu können. Bei Bedarf sollte sie sich daher entsprechende Unterstützung, zum Beispiel von Förderschullehrkräften, holen.

Bei Anzeichen von drohenden Lernschwierigkeiten sollte eine Lehrkraft möglichst schnell mit ersten unterstützenden Maßnahmen vorbeugend tätig werden. Hierzu sollte sie den eigenen Unterricht an die Bedürfnisse der Klasse anpassen oder zusätzlich für einzelne Schülerinnen und Schüler differenzierende Aufgaben und Materialien hinzuziehen. Die Rahmenbedingungen sind dabei auch von den Möglichkeiten der Schule abhängig. So zeigt sich, dass Differenzierung vor allem durch das Vorhandensein einer Doppelbesetzung im Unterricht gefördert wird. Spezifische Lese-Förderprogramme für den Einsatz in der ganzen Klasse gibt es dagegen im deutschsprachigen Raum vor allem für die Grundschule kaum. Zur Förderung der basalen Lesefertigkeit haben sich Lautleseverfahren als wirksam erwiesen, welche in der ganzen Klasse eingesetzt werden können. Ebenso kann bei gravierenderen Schwierigkeiten eine außerschulische Lerntherapie sinnvoll sein. Über diesen Schritt muss die Lehrkraft die Eltern entsprechend beraten können. Die Lehrkraft sollte den Austausch mit den Lehrkräften im Förderunterricht bzw. den Förderkräften in der Lerntherapie suchen und umgekehrt.

Im eigenen Unterricht sollte auf Fördermaßnahmen verzichtet werden, die nachweislich nicht wirksam sind. Hierzu zählen vor allem die oben genannten kausalen Therapien. Vielleseverfahren können dagegen eingesetzt werden, um die Lesemotivation zu steigern. Diese sind aber nicht geeignet, um schwache Leserinnen und Leser in ihrer Lesefertigkeit zu fördern. Neben Lautleseverfahren sollte auch eine Förderung auf Phonem- und Silbenbasis erfolgen. Die basalen Lesefertigkeiten müssen erworben und ausreichend automatisiert sein, damit die höheren Fertigkeiten der Lesekompetenz wie sinnentnehmendes Lesen überhaupt erst

erworben werden können. Erst bei ausreichenden basalen Lesefertigkeiten sollten dann Lesestrategien explizit vermittelt werden.

## 10 Ziel der Studie und Fragestellungen

In dieser Studie stehen die Leseentwicklung von Grundschulkindern im Verlauf der ersten Grundschuljahre und die förderdiagnostische Kompetenz von Grundschullehrkräften in Bezug auf diese Entwicklung im Fokus. Das Ziel der Studie, die methodische Umsetzung sowie die Fragestellungen, die beantwortet werden sollen, werden im Folgenden erläutert.

### 10.1 Ziel der Studie

Wie aus den vorangegangenen Kapiteln ersichtlich geworden ist, ist die Lesekompetenz eine Schlüsselkomponente, um Schülerinnen und Schülern ein selbstständiges und erfolgreiches Lernen und somit auch eine aktive Teilnahme an der Gesellschaft zu ermöglichen. Die Lehrkraft mit ihrem (förder-)diagnostischen Blick ist besonders gefragt, da einige Kinder bereits wenige Wochen nach Grundschuleintritt Schwierigkeiten beim Leseerwerb zeigen und diese sich, wenn sie nicht rechtzeitig erkannt und mit entsprechenden Fördermaßnahmen aufgefangen werden, manifestieren und zu einer Leseschwäche oder einer -störung entwickeln können. Eine Lehrkraft muss entsprechend über ausreichend umfangreiche diagnostische Kompetenzen verfügen, um diese Schwierigkeiten zu bemerken, die Ursachen zu ermitteln und mit entsprechender Förderung reagieren zu können. Studien haben allerdings gezeigt, dass leseschwache Schülerinnen und Schüler ihren Lehrkräften teilweise nicht auffallen (Artelt et al., 2001, S. 119; Valtin et al., 2010, S. 76).

Sehr häufig werden die diagnostischen Fähigkeiten von Lehrkräften in Form der Urteilsakkuratheit geprüft, indem die Übereinstimmungen zwischen gezeigter Leistung und vorhergesagter Leistung errechnet werden. Es wird hier meist nach Einschätzungen zum Beispiel über eine Vergabe von Noten oder ähnlichen Ratings gefragt und selten danach, was den Lehrkräften auffällt. Meistens wird daher nicht untersucht, wieso die Lehrkräfte die Schülerinnen und Schüler für förderbedürftig halten (Schmidt, 2015, S. 95). Dies könnte aber weiterhelfen, um Aufschluss darüber zu bekommen, wie Lehrkräfte zu ihren Urteilen gelangen. Es fehlt allerdings im deutschsprachigen Raum an längsschnittlichen Untersuchungen im Grundschulbereich, die die förderdiagnostische Kompetenz von Lehrkräften beim Leseerwerb in den Blick nehmen. Die vorliegende Untersuchung konzentriert sich daher nicht nur auf die Akkuratheit von Lehrkräfteurteilen, sondern auch darauf, anhand welcher Faktoren sie Schwierigkeiten beim Leseerwerb ausmachen, ob – und falls ja, welche – Fördermaßnahmen sie in der Folge ergreifen und ob sich all dies positiv auf die Leseentwicklung der Grundschul-kinder auswirkt. Die Einschätzungen der Lehrkräfte sollen hierbei mit quantitativen Daten über die schulischen Leistungen der Kinder verglichen werden. Hierzu wird ein Blick auf die Entwicklung der Lesekompetenz im Verlauf der ersten Grundschuljahre unter Berücksichtigung der förderdiagnostischen Kompetenz der Lehrkräfte sowie wichtiger Einflussfaktoren, welche in den vorherigen Kapiteln erläutert wurden, geworfen werden.

Der Zeitraum des ersten Schuljahres, aber auch der weitere Verlauf im zweiten Schuljahr sowie die Leseleistung der Kinder zum Ende der Grundschulzeit sind von besonderem

Interesse für die vorliegende Studie. Häufig sind anfänglich auftretende Schwierigkeiten andauernd und ein Großteil der Kinder, die auch im zweiten Schuljahr noch Schwierigkeiten haben, kann diese bis zum Ende der Schulzeit nur dann überwinden, wenn gezielte intensive Hilfe und Unterstützung gegeben werden (Fischer & Gasteiger-Klicpera, 2013, S. 78; Klicpera et al., 2013, S. 137). Wenn die untersuchten Schülerinnen und Schüler also im zweiten Schuljahr weiterhin Schwierigkeiten aufweisen, muss davon ausgegangen werden, dass diese sich hier bereits manifestieren.

Des Weiteren hat sich auch gezeigt, dass die Berufserfahrung nicht im Zusammenhang mit diagnostischen Kompetenzen zu stehen scheint (Christophel et al., 2017, S. 277; Hesse & Latzko, 2017, S. 27; Ohle & McElvany, 2015, S. 6). Das Ziel dieser Untersuchung ist es folglich, aufzuzeigen, wie es mit der förderdiagnostischen Kompetenz von Grundschullehrkräften im Unterrichtsfach Deutsch bestellt ist und wie sich diese auf die Leseentwicklung der Grundschulkindern auswirkt, um daraus ableiten zu können, ob und in welchen Bereichen die förderdiagnostische Kompetenz von Grundschullehrkräften in Bezug auf den Leseerwerb zukünftig stärker geschult werden müsste.

## 10.2 Methodisches Vorgehen

Die vorliegende Untersuchung ist Teil einer international angelegten Studie, die zeitgleich in Finnland, in Südtirol und in Deutschland stattgefunden hat (Hofmann et al., 2015; Koch et al., 2012). In dieser Studie sind die Lese- und Rechtschreibleistungen der Kinder über den Verlauf der Grundschulzeit erhoben worden. In der hier vorliegenden Untersuchung werden allerdings nur die Daten der Kinder in Deutschland mit dem oben genannten Fokus auf den Leseerwerb und die förderdiagnostische Kompetenz der Grundschullehrkräfte im Fach Deutsch betrachtet werden.

Tabelle 1.

### *Übersicht über die Zeitpunkte der Datenerhebungen*

<b>Schuljahr</b>	<b>Messzeitpunkt</b>	<b>Zeitpunkt im Schuljahr</b>	<b>Zeitraum</b>
1. Schuljahr	T1	Beginn des 1. Schuljahres	September 2010
	T2	Mitte des 1. Schuljahres	Februar 2011
	T3	Ende des 1. Schuljahres	Mai 2011
2. Schuljahr	T4	Beginn des 2. Schuljahres	September 2011
	T5	Mitte des 2. Schuljahres	Januar 2012
	T6	Ende des 2. Schuljahres	Mai 2012
4. Schuljahr	T7	Ende des 4. Schuljahres	Mai 2014

Die Datenerhebungen haben jeweils zu Beginn, zur Mitte und zum Ende des ersten sowie des zweiten Schuljahres stattgefunden (siehe Tabelle 1). Zudem ist eine Nacherhebung zum Ende der vierten Klasse erfolgt, um die Entwicklung der Schülerinnen und Schüler über die gesamte Grundschulzeit und ihren Lernstand vor dem Übergang in die Sekundarstufe I

erfassen zu können. Zu jedem dieser insgesamt sieben Messzeitpunkte sind mehrere Testverfahren durchgeführt worden.

Da die Untersuchung Teil einer international angelegten Studie ist, sind möglichst ähnliche Messinstrumente verwendet worden. Um die Erhebungen mit einer großen Anzahl an Schülerinnen und Schülern durchführen zu können, sind zudem, wann immer möglich, ökonomische Verfahren eingesetzt worden – so wurden Fragebögen und Gruppentestungen gegenüber einzelnen Interviews bevorzugt.

Im deutschsprachigen Raum ist die Erhebung der Leseleistung der Kinder mit dem ELFE 1-6 (Lenhard & Schneider, 2006) erfolgt. Dieser bietet sich an, da er zum einen ökonomisch als Gruppentest durchzuführen ist. Weiterhin erfasst er das Leseverständnis auf drei Ebenen, auf der Wortebene, der Satzebene und der Textebene. Des Weiteren sind Vorläuferfertigkeiten und mögliche Einflussfaktoren auf den Leseerwerb erhoben worden. Alle verwendeten Testverfahren werden in Kapitel 11 detailliert vorgestellt und beschrieben. Um leseschwache Schülerinnen und Schüler zu identifizieren, wird analog zu zuvor genannten Studien der zweite geplante Messzeitpunkt T2 betrachtet. Dieser hat zum Ende des ersten Schulhalbjahres, also einige Wochen nach Schulbeginn, stattgefunden. Für diesen recht frühen Messzeitpunkt liegt allerdings keine Normierung im ELFE 1-6 vor, da die früheste Normierung erst für das Ende des ersten Schuljahres existiert. Da aber der Lesetest zu den folgenden Messzeitpunkten wiederholt worden ist, muss hier anschließend von einem Lerneffekt ausgegangen werden. Folglich würde sich die Verwendung der Normierungen zur Auffindung der leseschwachen Schülerinnen und Schüler nicht eignen. Aus diesem Grund wird der Mittelwert der Gesamtstichprobe betrachtet: Die Schülerinnen und Schüler, die mindestens eine Standardabweichung unterhalb dieses Mittelwerts liegen, was einem T-Wert  $< 40$  entspricht, gelten in dieser Untersuchung als gefährdet. Diese Einteilung geschieht analog zu anderen Studien (Balke-Melcher et al., 2016, S. 20; Buschmann & Multhauf, 2016, S. 46). Stang und Urhahne (2016a, S. 215) bezeichnen eine Kontaktdauer von drei Monaten als eventuell zu kurz, sodass Lehrkräfte unter Umständen noch nicht genügend Möglichkeiten gehabt haben, alle Schülerinnen und Schüler ausreichend kennenlernen und ihre Fähigkeiten entsprechend einschätzen zu können. Oerke et al. (2016) konnten keine geringere Urteilsakkuratheit bei einem halben Jahr Kontaktdauer mit den Schülerinnen und Schülern im Vergleich zu eineinhalb Jahren erkennen. Daher soll in dieser Studie neben dem Messzeitpunkt T2, welcher einige Wochen nach Schulbeginn liegt und wo sich die ersten Schwierigkeiten im Leseerwerb bereits zeigen, auch der dritte Messzeitpunkt T3 zum Ende des ersten Schuljahres hinzugezogen werden.

Die förderdiagnostische Kompetenz der Lehrkräfte soll zu allen sieben Messzeitpunkten überprüft werden, um Aussagen über den Verlauf machen zu können. Hier sind keine gesonderten Testverfahren zum Einsatz gekommen. Die Lehrkräfte haben stattdessen in einer kurzen Befragung angegeben, welche Schülerinnen und Schüler ihnen aufgefallen sind. Zu den daraufhin genannten Schülerinnen und Schülern haben die Lehrkräfte anschließend

angegeben, was ihnen aufgefallen ist, wobei hier inhaltlich nicht weiter spezifiziert worden ist, um die Lehrkräfte nicht zu beeinflussen, und somit auch positive Auffälligkeiten oder zum Beispiel auch schwache Leistungen in Mathematik genannt werden konnten. Zuletzt sollten die Lehrkräfte angeben, was sie entsprechend unternehmen, um die genannten Schülerinnen und Schüler zu unterstützen.

### 10.3 Fragestellungen und Hypothesen

Aus dem zuvor herausgearbeiteten Stand der Forschung und der definierten Zielsetzung dieser Untersuchung leiten sich die folgenden Fragestellungen und Hypothesen ab.

#### 10.3.1 Fragestellung 1: Erkennen von Leseschwierigkeiten

Mit der ersten Fragestellung soll untersucht werden, ob Lehrkräften Schwierigkeiten im Leseerwerb auffallen und wie viele Kinder mit Leseschwierigkeiten ihnen korrekt bzw. wie viele Kinder mit Leseschwierigkeiten ihnen nicht auffallen. Weiterhin ist hier von Interesse, woran die Lehrkräfte schwache Leserinnen und Leser ausmachen.

**Fragestellung 1: Fällt den Grundschullehrkräften, die das Unterrichtsfach Deutsch unterrichten, auf, welche Schülerinnen und Schüler Leseschwierigkeiten haben und woran machen sie dies fest?**

Wie gezeigt werden konnte, unterschätzen Lehrkräfte den Anteil leseschwacher Schülerinnen und Schüler zum Teil sehr deutlich. In der IGLU-Studie 2006 lag der Anteil der leseschwachen Grundschulkinder, der von den Lehrkräften nicht als leseschwach identifiziert wurde, bei einem Drittel (Valtin et al., 2010, S. 76). Analog zu den Ergebnissen in der IGLU-Studie 2006, die mit Grundschulkindern und deren Lehrkräften durchgeführt wurde, kann daher davon ausgegangen werden, dass auch in der hier geplanten Studie ca. ein Drittel der leseschwachen Grundschulkinder den Lehrkräften im Unterrichtsfach Deutsch nicht auffällig wird.

In der Studie von Bailey und Drummond (2006) nannten die Lehrkräfte im Schnitt zwei bis vier Gründe, woran sie Leseschwierigkeiten bei ihren Schülerinnen und Schülern feststellten. Einige Lehrkräfte machten sehr detaillierte, andere dagegen nur vage Aussagen (Bailey & Drummond, 2006, S. 164). Diese genannten Gründe gehörten überwiegend zu den Bereichen Lesen, Schreiben, mündliche Sprache und Verhalten/Soziales. Analog zu den Studienergebnissen wird davon ausgegangen, dass die Lehrkräfte insgesamt auch hier sowohl vage wie auch detaillierte und spezifische Aussagen treffen werden.

Als leseschwach gelten in dieser Studie die Kinder, deren Leseleistungen im ELFE 1-6 zu den Messzeitpunkten T2 bzw. T3 mindestens eine Standardabweichung unter dem jeweiligen Mittelwert der Gesamtstichprobe liegen. Wenn eine Lehrkraft ein solches Kind als im Lesen auffällig benennt (entweder durch allgemeine Aussagen über Auffälligkeiten im Bereich

Lesen oder durch spezifische Aussagen über Auffälligkeiten in Teilfertigkeiten des Lesens), dann hat sie dieses Kind korrekt als leseschwach erkannt.

Ebenfalls sollte zu beiden Messzeitpunkten der Frage nachgegangen werden, ob die schwächsten Schülerinnen und Schüler einer Klasse genannt werden, unabhängig von ihrer absoluten Leseleistung. Dies würde für das Vorliegen des Referenzgruppeneffekts sprechen, wonach schwächere Schülerinnen und Schüler, welche z. B. in einer sehr leistungsstarken Klasse sind, schlechter bewertet werden als Schülerinnen und Schüler mit der gleichen Leistung in einer weniger starken Klasse (Trautwein & Baeriswyl, 2007, S. 122).

Auch zwei mögliche Einflussfaktoren – das Geschlecht der Kinder und der sozioökonomische Status – sollen in Bezug auf die diagnostische Kompetenz der Grundschullehrkräfte näher betrachtet werden. Studien konnten zeigen, dass Jungen im Verlauf der Schulzeit über eine niedrigere Lesekompetenz verfügen (McElvany & Ohle, 2015, S. 48). Zum Einfluss des Geschlechts auf die Urteilsakkuratheit von Lehrkräften gibt es dagegen gemischte Befunde: Karing, Matthäi et al. (2011) konnten keinen Einfluss finden. Andere Studien dagegen konnten diesen finden, aber immer in der Richtung, dass Mädchen bei gleicher schulischer Leistung besser bewertet wurden (Hinnant et al., 2009; Kaiser et al., 2015; Kuhl & Hannover, 2012). Krinzinger und Günther (2013) führen an, dass Kinder, die externalisierende Verhaltensauffälligkeiten aufweisen, der Lehrkraft auch eher insgesamt und in Bezug auf schulische Leistungen negativ auffallen. Da diese externalisierenden Störungen bei Jungen häufiger auftreten als bei Mädchen, wird dies als möglicher Grund aufgeführt, dass Störungen im Schriftspracherwerb bei Jungen eher auffallen als bei Mädchen (Krinzinger & Günther, 2013, S. 42). Auf die gleiche Art kann sich der sozioökonomische Status negativ auf die Urteilsakkuratheit auswirken. Maaz et al. (2008) sowie Ready und Wright (2011) konnten dies für die Beurteilung der Lesekompetenz bzw. des Schriftspracherwerbs aufzeigen. Karing, Matthäi et al. (2011) und Kaiser et al. (2015) konnten allerdings keine Belege finden, wobei letztere sich auf die Beurteilung der Mathematikleistung konzentriert hatten.

Interessant ist daher, ob leseschwache Jungen bzw. ob leseschwache Kinder mit einem niedrigen sozioökonomischen Status eher auffallen als leseschwache Mädchen bzw. leseschwache Kinder mit einem höheren sozioökonomischen Status.

Um diese Fragestellung beantworten zu können, wird zunächst analysiert werden, welche Aussagen von den Lehrkräften gemacht werden. Hierbei wird unterschieden werden, ob die Lehrkräfte lesespezifische Faktoren nennen, also Aussagen über Teilfertigkeiten des Lesens treffen. Des Weiteren wird angenommen, dass Lehrkräfte auch allgemeine Ursachen als Auffälligkeiten nennen, die nicht lesespezifisch sind, welche aber auch den Leseerwerb negativ beeinflussen können, wie beispielsweise Verhaltensauffälligkeiten. Im Anschluss werden die genannten Hypothesen zur Beantwortung der Fragestellung überprüft.

### **Hypothese 1.1**

*Ein Drittel der leseschwachen Schülerinnen und Schüler (1 SD unter M im ELFE 1-6) fällt den Lehrkräften zum Messzeitpunkt T2 nicht auf - das heißt, dass die Lehrkräfte keine Aussagen über Auffälligkeiten im Lesen allgemein oder über Teilfertigkeiten im Spezifischen nennen.*

### **Hypothese 1.2**

*Ein Drittel der leseschwachen Schülerinnen und Schüler (1 SD unter M im ELFE 1-6) fällt den Lehrkräften zum Messzeitpunkt T3 nicht auf - das heißt, dass die Lehrkräfte keine Aussagen über Auffälligkeiten im Lesen allgemein oder über Teilfertigkeiten im Spezifischen nennen.*

### **Hypothese 1.3**

*Sollten sich zu den Messzeitpunkten T2 oder T3 einzelne Schulklassen als lesestark herausstellen (der Mittelwert der Leseleistung der Klasse liegt signifikant über dem Mittelwert der Gesamtstichprobe), wird überprüft, ob es zum Referenzgruppeneffekt kommt und entsprechend im Lesen schwächere Kinder als leseschwach genannt werden, welche verglichen mit der Gesamtstichprobe nicht leseschwach sind.*

### **Hypothese 1.4**

*Lehrkräfte bezeichnen Mädchen, die leseschwach sind, seltener als leseauffällig, als dies für leseschwache Jungen geschieht.*

### **Hypothese 1.5**

*Lehrkräfte bezeichnen Kinder mit einem höheren sozioökonomischen Status, die leseschwach sind, seltener als leseauffällig, als dies für leseschwache Kinder mit einem niedrigeren sozioökonomischen Status geschieht.*

#### 10.3.2 Fragestellung 2: Einfluss der Diagnose auf die Leseentwicklung

Im nächsten Schritt soll überprüft werden, ob das Auffallen von Schwierigkeiten – entweder allgemein, für den Bereich Deutsch oder spezifisch für den Leseerwerb – einen Einfluss auf die Leseentwicklung der Grundschul Kinder hat.

#### **Fragestellung 2: Wird die Leseentwicklung von Schülerinnen und Schülern davon beeinflusst, ob den Lehrkräften Schwierigkeiten bei diesen auffallen?**

Ausgehend vom Forschungsstand lässt sich festhalten, dass Grundschullehrkräfte sowohl über diagnostische als auch über didaktische Kompetenzen verfügen. Es ist folglich anzunehmen, dass Schülerinnen und Schüler, die ihnen aufgrund von Schwierigkeiten auffallen, indirekt oder direkt Zuwendung und Unterstützung erhalten, welche zu einer Verbesserung in den schulischen Leistungen führen. Die gesteigerte Aufmerksamkeit der Lehrkraft auf die

Schülerin oder den Schüler könnte für sich genommen bereits zu einer Verbesserung der schulischen Leistung führen (Diaper, 1990, S. 266). Da in Bezug auf Schwierigkeiten beim Leseerwerb von der Notwendigkeit intensiver Förderung gesprochen wird, sollten Lehrkräfte hier aber nicht untätig bleiben (Fischer & Gasteiger-Klicpera, 2013, S. 63). Die IGLU-Studien von 2006 und 2016 zeigten allerdings auf, dass nur etwa ein Drittel der als förderbedürftig erkannten Kinder eine schulische Förderung erhält (Bremerich-Vos et al., 2017, S. 281; Valtin et al., 2010, S. 76). Es wird aber ebenfalls davon ausgegangen, dass Lehrkräfte über ausreichend didaktische Kompetenzen verfügen und Aufgaben entsprechend auf den Leistungsstand der Kinder sowie auf deren Bedürfnisse anpassen können (Anders et al., 2010, S. 178; 180; Kaiser & Möller, 2017, S. 71–72), sodass eine gesteigerte Aufmerksamkeit der Lehrkraft schon zu Anpassungen und damit einhergehend zu einer Leistungssteigerung führen könnte. Daher ist es interessant, diesen Aspekt hier genauer zu betrachten.

### **Hypothese 2.1**

*Die Kinder, die zu Messzeitpunkt T2 von der Lehrkraft als auffällig im Bereich Deutsch bezeichnet werden, werden sich – unabhängig von anderen Einflussfaktoren – positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die den Lehrkräften zu Beginn der Schulzeit nicht auffällig geworden sind.*

### **Hypothese 2.2**

*Die Kinder, die zu Messzeitpunkt T3 von der Lehrkraft als auffällig im Bereich Deutsch bezeichnet werden, werden sich – unabhängig von anderen Einflussfaktoren – positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die den Lehrkräften zu Beginn der Schulzeit nicht auffällig geworden sind.*

### **Hypothese 2.3**

*Die Kinder, die zu Messzeitpunkt T2 von der Lehrkraft als leseauffällig bezeichnet werden, werden sich – unabhängig von anderen Einflussfaktoren – positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die den Lehrkräften zu Beginn der Schulzeit nicht auffällig geworden sind.*

### **Hypothese 2.4**

*Die Kinder, die zu Messzeitpunkt T3 von der Lehrkraft als leseauffällig bezeichnet werden, werden sich – unabhängig von anderen Einflussfaktoren – positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die der Lehrkraft zu T3 nicht auffällig geworden sind.*

### **Hypothese 2.5**

*Die Kinder, die zu Messzeitpunkt T2 der Lehrkraft allgemein auffällig werden, werden sich in der Folge positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die der Lehrkraft zu T2 nicht aufgefallen sind.*

### **Hypothese 2.6**

*Die Kinder, die zu Messzeitpunkt T3 der Lehrkraft allgemein auffällig werden, werden sich in der Folge positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die der Lehrkraft zu T3 nicht aufgefallen sind.*

#### 10.3.3 Fragestellung 3: Eingeleitete (Förder-)Maßnahmen bei Auffälligkeiten

Mit dieser dritten Fragestellung soll überprüft werden, inwiefern Lehrkräfte mit Förder- und Unterstützungsangeboten auf die Leseschwierigkeiten der Kinder reagieren.

### **Fragestellung 3: Werden Schülerinnen und Schüler, deren Leseschwierigkeiten erkannt wurden, mit Unterstützungsangeboten versorgt und sind diese spezifisch für die Leseschwierigkeiten?**

Nach Schmidt zeichnet sich die förderdiagnostische Kompetenz einer Lehrkraft dadurch aus, dass diese „im Idealfall prozessorientiert die Fähigkeiten ihrer Lernenden diagnostizieren und daraus Konsequenzen für die Auswahl von Lehrangeboten ziehen“ kann (Schmidt, 2015, S. 92). Nur etwa jedes dritte förderbedürftige leseschwache Kind erhält allerdings eine schulische Förderung (Bremerich-Vos et al., 2017, S. 281; Valtin et al., 2010, S. 76). Es soll betrachtet werden, ob dieses Ergebnis hier reproduziert wird. Es soll aber ebenso betrachtet werden, welcher Art die durchgeführten Fördermaßnahmen sind, um darauf aufbauend nicht nur eine quantitative Einschätzung über den Anteil an leseschwachen Kindern, die Unterstützung erhalten, treffen zu können, sondern auch eine qualitative Einschätzung: Welche Arten von Maßnahmen werden eingeleitet? Werden Angebote außerschulisch oder innerschulisch durchgeführt? Die Inhalte der Förderung werden daher auch eingehend betrachtet werden. So ist es denkbar, dass ein Teil der Kinder zwar keine schulische, wohl aber eine außerschulische Förderung erhält. Ebenso könnte es sein, dass die Lehrkraft über Differenzierung der Aufgaben ihren Unterricht entsprechend anpasst, aber keine explizite schulische Förderung erfolgt. Des Weiteren stellt auch das Elternhaus eine Ressource dar und ein Einbeziehen der Eltern in die Förderung kann somit ebenfalls sinnvoll sein (Villiger et al., 2012, S. 90).

Um diese Fragestellung beantworten zu können, werden die Antworten der Lehrkräfte in einem ersten Schritt im Hinblick darauf, welche Arten von Maßnahmen genannt werden, kategorisiert werden. Dies können zum einen innerschulische spezifische und unspezifische Förderangebote sein, aber auch außerschulische, wie zum Beispiel Nachhilfe oder Lerntherapie. Im Anschluss werden die beiden Hypothesen zur Beantwortung der Fragestellung überprüft.

### **Hypothese 3.1**

*Zwei Drittel der als leseauffällig genannten Kinder (zu T2 und T3) bleiben in der Folge zu den Messzeitpunkten T2 bis T7 ohne schulinterne Leseförderung.*

### **Hypothese 3.2**

*Analog zu Hypothese 3.1 bleiben auch zwei Drittel der allgemein als auffällig genannten Kinder (zu T2 und T3) in der Folge zu den Messzeitpunkten T2 bis T7 ohne schulinterne Förderung.*

#### 10.3.4 Fragestellung 4: Erfolg der durchgeführten (Förder-)Maßnahmen

Während die dritte Fragestellung herausarbeiten soll, ob leseschwache Schülerinnen und Schüler mit Unterstützungsangeboten versorgt werden, geht es hier nun um die Frage nach dem Erfolg der durchgeführten Maßnahmen.

#### **Fragestellung 4: Führen die eingeleiteten (Förder-)Maßnahmen bei den Schülerinnen und Schülern zu einer Steigerung der Lesekompetenz?**

Aufgrund der diagnostischen und didaktischen Fähigkeiten der Lehrkräfte kann angenommen werden, dass die Kinder, die korrekt als leseschwach identifiziert werden und daraufhin Förderung erhalten, später über eine höhere Lesekompetenz verfügen als diejenigen, die zwar als leseschwach erkannt werden, aber keine Förderung erhalten und als diejenigen, die nicht als leseschwach erkannt werden. Daher soll in dieser Fragestellung darauf geblickt werden, ob eingeleitete Unterstützungsmaßnahmen Auswirkungen auf die Lesekompetenz der Grundschul Kinder haben. Nach Fischer und Gasteiger-Klicpera (2013, S. 78) sowie Klicpera et al. (2013, S. 137) ist eine intensive Förderung notwendig, um Schülerinnen und Schüler mit Leseschwierigkeiten zu unterstützen, weswegen auch differenziert auf die Fördermaßnahmen und deren Auswirkungen geblickt werden soll. Da nicht mit kurzfristigen Ergebnissen von Fördermaßnahmen gerechnet werden kann (Balke-Melcher et al., 2016, S. 25), wird die Leseleistung zum Ende der vierten Klasse, also zum siebten Messzeitpunkt, als relevanter Zeitpunkt für die Entscheidung über die Wirksamkeit der Fördermaßnahmen hinzugezogen.

### **Hypothese 4.1**

*Wenn ein Kind zu Beginn der Schulzeit (T2 bzw. T3) leseschwach ist, die Lehrkraft dies bemerkt und lesespezifische Fördermaßnahmen eingeleitet werden, dann erzielt das Kind zu T7 eine höhere Leseleistung als ein leseschwaches Kind, bei dem dies nicht geschehen ist.*

**Hypothese 4.2**

*Je früher die Lehrkraft von einer Leseförderung für eine Schülerin oder einen Schüler berichtet, der bzw. die zu Beginn der Schulzeit (T2 bzw. T3) leseschwach ist, desto besser wird die Leseleistung zu T7 sein.*

**Hypothese 4.3**

*Leseschwache Schülerinnen und Schüler, die spezifische Förderprogramme im Lesen als Unterstützung erhalten, verfügen zum Ende der vierten Klasse (T7) über eine bessere Leseleistung als Schülerinnen und Schüler, die keine solche Förderung erhalten haben.*

## 11 Methodik

Die durchgeführte Untersuchung ist Teil einer international angelegten Studie, welche parallel über vier Jahre hinweg in Deutschland ( $N = 281$  Grundschul Kinder aus 13 Klassen), Finnland ( $N = 322$  Grundschul Kinder aus 21 Klassen) und Südtirol ( $N = 162$  deutschsprachige Grundschul Kinder aus 10 Klassen) stattgefunden hat (Hofmann et al., 2015; Koch et al., 2012). Hierbei wurden sowohl die Leseentwicklung wie auch die Entwicklung der Rechtschreibleistung im Verlauf der ersten beiden Grundschuljahre und in einer Follow-Up Untersuchung im vierten Schuljahr betrachtet. Die Datenerhebungen fanden zwischen September 2010 (zu Beginn der ersten Klasse) und Mai 2014 (gegen Ende der vierten Klasse) an insgesamt sieben Messzeitpunkten statt.

Gegenstand der vorliegenden Untersuchung sind die Daten aus Deutschland, weshalb nur auf die Erhebungen in den deutschen Grundschulen sowie auf die deutschsprachigen Testverfahren eingegangen wird. Der Bereich der Rechtschreibentwicklung wird nur insofern betrachtet, wie er relevant zur Beantwortung der vorliegenden Fragestellungen erscheint.

### 11.1 Stichprobenakquise

Zur Rekrutierung der deutschen Stichprobe wurde eine Anfrage an das Staatliche Schulamt für den Landkreis Gießen und den Vogelsbergkreis gestellt, um städtische sowie ländliche Grundschulen zu finden, die an der Untersuchung teilnehmen möchten. Sieben Grundschulen mit insgesamt 13 Schulklassen aus der mittelhessischen Region konnten auf diese Weise für die Teilnahme an der Studie gewonnen werden. Die Schulen wurden persönlich kontaktiert und erhielten zu Beginn ein Informationsschreiben über die geplante Studie. Nach Zusage der Schule wurden Einverständniserklärungen und Informationsschreiben an die Eltern verschickt. Für jede Klasse wurde zudem die Grundschullehrkraft, die das Fach Deutsch unterrichtet, befragt. Insgesamt haben 13 Deutschlehrkräfte und 281 Grundschul Kinder an den Testungen teilgenommen.

### 11.2 Ablauf der Studie

Die Durchführung fand im Zeitraum von September 2010 bis Mai 2014 an sieben Messzeitpunkten statt (s. Tabelle 2).

Tabelle 2  
*Datenerhebungen in den Grundschulen*

	1. Klasse			2. Klasse			4. Klasse
<b>Messzeitpunkt</b>	T1	T2	T3	T4	T5	T6	T7
<b>Zeitraum</b>	Sep 2010	Feb 2011	Mai 2011	Sep 2011	Jan 2012	Mai 2012	Mai 2014

Mitarbeiterinnen und Mitarbeiter des Instituts für Förderpädagogik und Inklusive Bildung sowie geschulte Studentinnen des Förderschullehramts führten die Erhebungen jeweils als

Gruppentestungen im Klassenzimmer oder als Einzeltestungen in ruhigen Nebenräumen in der jeweiligen Schule durch. Die Lehrkräfte waren bei den Erhebungen im Klassenraum anwesend. Im Anschluss fand jeweils die Befragung der Lehrkraft zu den Grundschulkindern statt (Details hierzu finden sich in Kapitel 11.5).

Zu Messzeitpunkt T7 konnte die Erhebung an einer der Schulen, die mit zwei Klassen an der Studie teilgenommen hat, nicht mehr durchgeführt werden. Des Weiteren haben im Verlauf der Grundschulzeit einzelne Schülerinnen und Schüler die Schule gewechselt, eine Klasse wiederholt oder auch übersprungen, sodass für den siebten Messzeitpunkt insgesamt noch Daten von 171 Grundschulkindern vorliegen, was einem Anteil von 61 Prozent entspricht.

### 11.3 Soziodemographische Daten

Die Geburtsdaten der Kinder wurden erhoben, um das genaue Alter zu den Messzeitpunkten berechnen zu können. Zudem wurde das Geschlecht jedes Kindes erfasst. Die Kinder gaben außerdem an, ob sie zuhause (vorwiegend) deutsch, deutsch und eine andere Sprache oder vorwiegend eine andere Sprache sprechen. Analog zu den PISA-Studien (Baumert & Schümer, 2001, S. 331) wurde der sozioökonomische Status indirekt über die Angabe der Berufstätigkeit der Eltern festgestellt, welche in einem Elternbrief erfragt wurde.

### 11.4 Testverfahren bei den Grundschulkindern

Zu jedem Messzeitpunkt wurden mehrere Testverfahren durchgeführt. Tabelle 3 zeigt kompakt, welche Verfahren zu den jeweiligen Messzeitpunkten zur Anwendung gekommen sind. Nach Möglichkeit sind für die Testungen standardisierte Testverfahren ausgewählt worden. Teilweise war es notwendig, selbsterstellte Testverfahren zu verwenden, um ein annähernd identisches Vorgehen in Deutschland, Finnland und Südtirol zu gewährleisten. Die phonologische Bewusstheit ist daher durchgehend mit selbst erstellten Testverfahren überprüft worden. Für die Bereiche Rechtschreib- und Leseleistung sowie Selbstkonzept, Klassenklima und soziale Integration kamen standardisierte sowie selbst erstellte Verfahren zum Einsatz. Zur Überprüfung der kognitiven Fähigkeiten kamen nur standardisierte Verfahren zur Anwendung.

Tabelle 3

*Testverfahren bei den Grundschulkindern zu den 7 Messzeitpunkten*

	1. Klasse			2. Klasse			4. Klasse
	Sep 2010	Feb 2011	Mai 2011	Sep 2011	Jan 2012	Mai 2012	Mai 2014
<b>Phonologische Bewusstheit</b>							
Silbensynthese*	•						
<b>Silbensegmentation (Bilder)*</b>	•	•					
<b>Silbensegmentation (Wörter)*</b>				•	•	•	•
Phonem Identifikation des Anlautes*	•						
Phonem Identifikation des Endlautes*		•					
Phonemsynthese I*	•	•					
Phonemsynthese II*			•	•			
<b>Satzsegmentation*</b>				•	•		•
<b>Rechtschreib- und Leseleistung</b>							
<b>Buchstabenkenntnis*</b>		•					
<b>Laut-Buchstaben-Zuordnung*</b>	•		•				
<b>Schreiben nach Bildern*</b>	•	•	•	•	•		
<b>Rechtschreibtest (DERET 1-2+)</b>			•			•	
<b>Rechtschreibtest (DERET 3-4+)</b>							•
<b>Wortlesetest (ELFE 1-6)</b>		•	•	•		•	•
<b>Satzlesetest (ELFE 1-6)</b>				•		•	•
<b>Textlesetest (ELFE 1-6)</b>						•	•
<b>Selbstkonzept, Klassenklima und soziale Integration</b>							
<b>Schulisches Selbstkonzept und Lernmotivation**</b>		•					
<b>Klassenklima (FEES 1-2)</b>			•			•	
<b>Fähigkeitsselbstkonzept (FEES 1-2)</b>			•			•	
<b>Soziale Integration (FEES 1-2)</b>			•			•	
<b>Klassenklima (FEES 3-4)</b>							•
<b>Fähigkeitsselbstkonzept (FEES 3-4)</b>							•
<b>Soziale Integration (FEES 3-4)</b>							•
<b>Kognitive Fähigkeiten</b>							
<b>Logisches Denken (RPM)</b>					•		
<b>Benennungsgeschwindigkeit</b>					•		
<b>Arbeitsgedächtnis</b>					•		

*Anmerkung:* Die mit \* versehenen Tests entstammen der deutschsprachigen Testbatterie von Koch und Georg (2010). \*\*Fragebogen von Stiensmeier-Pelster und Schöne (o. J.). Fett markiert sind die Skalen, die in die vorliegende Untersuchung eingegangen sind.

Zudem sollte darauf geachtet werden, die Kinder zu den einzelnen Messzeitpunkten nicht übermäßig zu belasten oder zu überfordern, sodass nicht alle Verfahren zu jedem Messzeitpunkt angewendet wurden.

#### 11.4.1 Vorgaben zu den Itemanalysen für die selbst erstellten Verfahren

Zur Prüfung der Güte der selbst erstellten deutschsprachigen Verfahren wird jeweils die interne Konsistenz berechnet. McNeish (2018, S. 422–423) argumentiert, dass Cronbachs Alpha nicht mehr verwendet werden sollte, da dieser Koeffizient auf Annahmen beruht, die zu starr sind und die auf die zu untersuchenden Daten häufig nicht zutreffen. Dies führt auch in

der Interpretation des Koeffizienten zu Schwierigkeiten. Er kommt zu dem Schluss, dass sämtliche alternativen Methoden besser geeignet seien als Cronbachs Alpha (McNeish, 2018, S. 423). Dunn, Baguley und Brunnsden (2014, S. 405) nennen Omega  $\omega$  als Alternative, da dieser Wert auch bei Verletzung der Annahmen einen guten Index für die interne Konsistenz darstellt. McNeish (2018, S. 417) bezeichnet den Koeffizienten auch als „Omega Total“ ( $\omega_{Total}$ ), um ihn von anderen ähnlichen Koeffizienten abzugrenzen, die ebenfalls als Omega bezeichnet werden. Die Angabe des Koeffizienten soll dabei stets unter Hinzunahme des Konfidenzintervalls geschehen (Dunn et al., 2014, S. 409). Es wird für die verwendeten selbst erstellten Testverfahren dieser Studie entsprechend Omega  $\omega_{Total}$  berechnet.

Über die benötigte Höhe des Koeffizienten zur internen Konsistenz gibt es keine übereinstimmenden Angaben, weshalb sich an Janssen und Laatz (2017, S. 611) sowie Kuckartz, Rädiker, Ebert und Schehl (2013, S. 247) orientiert wird. Erstere empfehlen Werte von mindestens .70 oder .80. Letztere bezeichnen diesen Wertebereich als brauchbar, Werte ab .80 als gut und Werte ab .90 als sehr gut. Der Wert von .70 wird demnach als untere Grenze festgelegt. Cortina (1993, S. 102) hält fest, dass 14 Items bereits zu einem Wert von über .70 führen und die Werte für solche Skalen entsprechend vorsichtig interpretiert werden müssen. Streiner (2003, S. 103) legt die Grenze bei 20 Items und mehr. Da der Wert folglich auch immer abhängig ist von der Anzahl der Items, die in die Berechnung einfließen, kann er nur als Orientierung dienen. Für die jeweilige Itemanalyse werden neben der internen Konsistenz daher ebenso die Trennschärfe sowie die Itemschwierigkeit berichtet. Es gilt, dass die Trennschärfe zwischen .40 und .70 liegen soll, um als gut bezeichnet zu werden (Kelava & Moosbrugger, 2012, S. 86). Kuckartz et al. (2013, S. 247) bezeichnen Trennschärfen ab .50 als hoch. Ferner schlagen sie einen unteren Grenzwert von .25 vor. Entsprechend dieser Angaben sollen die verwendeten Items eine Trennschärfe von mindestens .25 aufweisen. Werte ab .25 bis .39 werden als brauchbar, ab .40 als gut und ab .50 als hoch bezeichnet.

Die Itemschwierigkeit soll zwischen 20 und 80 liegen, was bedeutet, dass zwischen 20 und 80 Prozent der Probanden das jeweilige Item lösen können. Berechnet wird die Itemschwierigkeit über die Anzahl der Probanden, die das Item korrekt lösen konnten, geteilt durch die Gesamtzahl an Probanden  $\times 100$ . Je höher die Itemschwierigkeit ist, umso *leichter* ist folglich das Item zu lösen. Bei der Itemauswahl sollte allerdings auch der Einsatzzweck berücksichtigt werden, sodass Extremwerte eingeschlossen werden können (Kelava & Moosbrugger, 2012, S. 87). Da die vorliegende Untersuchung leseschwache Kinder im Fokus hat, sind auch solche Items von Interesse, die in extremeren Merkmalsbereichen differenzieren, weswegen die Itemschwierigkeit auch zwischen 5 und 95 liegen kann, sofern „gleichzeitig ausreichend hohe Trennschärfen“ vorliegen (Kelava & Moosbrugger, 2012, S. 87).

Die Itemanalysen erfolgen mit der Software R in der Version 3.6.1 (R Core Team, 2019) und RStudio in der Version 1.2.1335 (RStudio Team, 2019). Hierüber stehen aktuelle statistische Verfahren wie die Berechnung von  $\omega_{Total}$  mit dem Paket *MBESS* (Kelley, 2019) zur Verfügung.

## 11.4.2 Vor- und Teilleistungen des Schriftspracherwerbs

Die phonologische Bewusstheit sowie weitere Vor- und Teilleistungen des Schriftspracherwerbs wurden über mehrere Subskalen erhoben. Zur Erfassung dieser Fertigkeiten wurden selbst erstellte Verfahren verwendet, die den durchgeführten Verfahren in Finnland weitestgehend entsprechen (Koch & Georg, 2010). In dieser Untersuchung finden die Silbensegmentation und die Satzsegmentation als Teil der phonologischen Bewusstheit Anwendung. Neben den standardisierten Verfahren zur Erfassung der Rechtschreib- und Lesefähigkeiten wurden noch die Buchstabenkenntnis, Laut-Buchstaben-Zuordnung sowie das Schreiben nach Bildern über selbst erstellte Verfahren erhoben.

Tabelle 4

*Interne Konsistenz  $\omega_{Total}$  für die Skalen der selbst erstellten Tests*

	T1	T2	T3	T4	T5	T6	T7
<b>Phonologische Bewusstheit</b>							
Silbensegmentation (Bilder)	.83 [.79, .86]	.89 [.87, .92]					
Silbensegmentation (Wörter)*				.76 [.62, .91]	.84 [.79, .89]	.86 [.83, .89]	.89 [.86, .92]
Satzsegmentation				.88 [.86, .90]	.88 [.86, .90]		.87 [.79, .95]
<b>Rechtschreib- und Leseleistung</b>							
Buchstabenkenntnis		.72 [.63, .82]					
Laut-Buchstaben-Zuordnung*	.92 [.91, .93]		.77 [.73, .82]				
Schreiben nach Bildern	.89 [.87, .92]	.82 [.79, .85]	.73 [.68, .79]	.77 [.71, .83]	.71 [.65, .78]		

*Anmerkungen.* Das Konfidenzintervall (95 %) steht in Klammern. Teilweise variierte die Itemanzahl zu den Messzeitpunkten. Informationen hierüber finden sich in den entsprechenden Abschnitten zu den Skalen. \* Skala besteht aus mehr als 20 Items.

Tabelle 4 zeigt die interne Konsistenz der Skalen zu den einzelnen Messzeitpunkten. Die verwendeten Skalen werden nachfolgend vorgestellt. Die Angaben zu den jeweiligen Trennschärfen und Itemschwierigkeiten der Skalen sind in Anhang A zu finden.

## 11.4.2.1 Silbensegmentation mit Bildern

Die Fähigkeit Silben zu segmentieren wurde überprüft, indem die Kinder gesprochene Wörter durch Klatschen in Silben zerlegen sollten (z. B. „La-ma“). Dies geschah in der ersten Klasse zu den Messzeitpunkten T1 und T2 mithilfe von Bildern. Jedes so dargestellte Wort bestand aus einer bis vier Silben. Unterhalb des Bildes befanden sich jeweils vier stilisierte Hände. Die Kinder sollten für jede Silbe im Wort eine Hand ausmalen.

Geübt wurde die Aufgabe zu Beginn an drei Beispielen. Die Testleiterin las das jeweils nächste Wort laut vor. Die Kinder durften jedes für sich leise klatschen, um die Anzahl der Silben festzustellen. Zum ersten Messzeitpunkt wurden den Kindern acht Bilder gezeigt. Zu

Messzeitpunkt T2 wurden zwei weitere Bilder hinzugefügt, sodass insgesamt zehn Aufgaben zu bearbeiten waren. Die Bearbeitungszeit lag bei 5 bis 7 Minuten.

Mit  $\omega_{Total} = .83$  und  $.89$  sind die internen Konsistenzen zu beiden Messzeitpunkten als gut zu bezeichnen (s. Tabelle 4). Die Trennschärfen für alle Items liegen zwischen  $.39$  und  $.71$  und sind damit brauchbar bis hoch. Gleichzeitig liegen die Itemschwierigkeiten zwischen 49 und 80 (s. Tabelle 121 in Anhang A). Insgesamt entsprechen somit alle Items den Anforderungen.

#### 11.4.2.2 Silbensegmentation mit Wörtern

Ab der zweiten Klasse wurde die Fähigkeit Wörter in Silben zu zerlegen nicht mehr mit Bildern überprüft, sondern mit geschriebenen Wörtern, in die durch senkrechte Striche die Silbentrennung markiert werden sollte (z. B. „Bleistift“ – „Blei|stift“). Es gab auf dem Aufgabenblatt neun Beispiele, bei denen die Striche bereits eingefügt waren. Geübt wurde anschließend mit 17 Wörtern. Zur Bearbeitung der insgesamt 44 Test-Wörter wurde eine Zeit von 1 Minute und 15 Sekunden vorgegeben. Item 9 („Stift“) wurde allerdings aus der Analyse entfernt, da nicht differenziert werden kann zwischen Kindern, die die Antwort nicht wussten und deswegen nichts angestrichen haben, und Kindern, die das Item korrekt bearbeitet haben, da zur korrekten Bearbeitung ein Nichtmarkieren erfolgen musste.

Je nach Anzahl der Silben schwankt die maximal erreichbare Punktzahl pro Wort zwischen 1 und 5. Bevor für diese Skala die interne Konsistenz berechnet wurde, wurden die Items daher z-transformiert, um eine Vergleichbarkeit der Items untereinander zu garantieren.

Ab Item 28 liegen keine brauchbaren Trennschärfen mehr vor (s. Tabelle 122 in Anhang A). Für Item 27 ist nur die Trennschärfe zu T7 im brauchbaren Bereich. Nur 22 Kinder haben dieses Item bearbeitet, wobei nur 5 dieser Kinder das Item auch korrekt beantwortet haben. Selbst zu Messzeitpunkt T7 scheint dieses Item daher sehr schwer zu sein. Zu den ersten beiden Messzeitpunkten wurde das Item von keinem Kind beantwortet. Zu T6 ist die Trennschärfe mit  $.17$  nicht akzeptabel.

Da für diesen Test ein Zeitlimit vorgegeben war, muss davon ausgegangen werden, dass die letzten Items aufgrund der Zeitbeschränkung von vielen Kindern nicht mehr bearbeitet werden konnten und somit für die besonders relevanten Messzeitpunkte in den ersten beiden Grundschuljahren keine oder kaum Informationen liefern. Die Items 27 bis 44 werden deshalb aus der Skala entfernt.

Die internen Konsistenzen für die endgültige Skala sind mit  $\omega_{Total} = .76$  bis  $.89$  als brauchbar bis gut zu bewerten (s. Tabelle 4). In der Neuberechnung erreicht Item 26 nun nur noch zu einem Messzeitpunkt, T7, eine brauchbare Trennschärfe (s. Tabelle 123 in Anhang A). Da allerdings nach Eliminierung dieses Items die interne Konsistenz nicht ansteigt und die Trennschärfe zu T7 ausreichend hoch ist, wird das Item beibehalten. Die Itemschwierigkeit gibt an, wie viel Prozent der Kinder die volle Punktzahl erreichten. Da die Kinder die Items nicht immer in der vorgegebenen Reihenfolge bearbeiteten und teilweise längere Wörter übersprangen,

erfolgte die Berechnung der Itemschwierigkeiten nicht unter Berücksichtigung der zeitlichen Beschränkung. Die Angaben müssen entsprechend eingeschränkt betrachtet werden.

#### 11.4.2.3 Satzsegmentation

Um die Fähigkeit zum Erkennen von Wortgrenzen zu überprüfen, wurden den Kindern bei diesem Test zehn Aufgaben vorgelegt. Es wurden zu Beginn drei Beispiele vorgegeben und im Anschluss drei Übungsaufgaben zusammen durchgeführt. Jede Aufgabe bestand aus einem Satz, welcher ohne Lücken als ein Wort geschrieben war. Nur der erste Buchstabe des Satzes war großgeschrieben. Aufgabe der Kinder war, die Lücken zwischen den Wörtern jeweils mit einem senkrechten Strich zu markieren (z. B. „Esregnetnicht“ – „Es|regnet|nicht“). Für die Bearbeitung der zehn Aufgaben hatten die Kinder 1 Minute und 25 Sekunden Zeit. Wie bei der Silbensegmentation waren auch für die Satzsegmentation abhängig von der Wortanzahl der Sätze unterschiedliche maximal erreichbare Punkte pro Aufgabe möglich. Daher wurden diese Items ebenfalls zunächst z-transformiert, um eine Vergleichbarkeit untereinander zu ermöglichen. Auch hier war es den Kindern möglich, Aufgaben zu überspringen, sodass die Itemschwierigkeit für alle teilnehmenden Kinder berechnet wurde und die Zeitbegrenzung nicht berücksichtigt wird.

Die internen Konsistenzen für die Satzsegmentation liegen zu den drei Messzeitpunkten T4, T5 und T7 bei  $\omega_{Total} = .87$  bis  $.88$  (s. Tabelle 4). Die Trennschärfen können bis auf wenige Ausnahmen als gut bis hoch bezeichnet werden. Kein Wert ist unterhalb der Grenze von  $.25$  (s. Tabelle 124 in Anhang A). Die Itemschwierigkeiten beziehen sich auf die Häufigkeit, alle Grenzen im vorgegebenen Satz korrekt erkannt zu haben. Einige Itemschwierigkeiten liegen bei unter 5 bzw. über 95. Vor allem zu T7 sind einige Items sehr leicht. Dagegen sind die Items 8 bis 9 zu Messzeitpunkt T4 mit einer Itemschwierigkeit von 5 und darunter sehr schwer, was sich aber zu T7 nicht mehr zeigt. Da die Trennschärfen durchgehend oberhalb des Grenzwerts liegen und die interne Konsistenz gut ist, werden alle Items beibehalten.

#### 11.4.2.4 Buchstabenkenntnis

Die Buchstabenkenntnis wurde zum zweiten Messzeitpunkt erfasst. Die Kinder sollten zu zwei vorgeschprochenen Buchstaben die entsprechenden verschriftlichten Buchstaben in einer Zeile markieren („Könnt ihr in dieser Zeile das versteckte U und das versteckte T finden?“). Zu jeder Aufgabe wurde eine Reihe mit sechs Buchstaben bestehend aus den zwei genannten sowie vier Distraktoren vorgelegt. Es wurden alle 26 Grundbuchstaben in Großschreibung abgefragt (ohne Umlaute und ß). Bei jeder einzelnen Aufgabe wurden zwei Buchstaben erfragt, sodass sich insgesamt 13 Aufgaben ergeben. Die Bearbeitungszeit lag bei 5 Minuten. Die interne Konsistenz für die Skala mit 13 Items beträgt  $\omega_{Total} = .72$  und ist somit brauchbar (s. Tabelle 4). Vier Items haben eine niedrige Trennschärfe von unter  $.25$  (s. Tabelle 125 in Anhang A). Gleichzeitig liegt die Itemschwierigkeit bei Items 1 bis 4 oberhalb der Grenze von 95. Diese Items wurden entsprechend von sehr vielen Kindern korrekt beantwortet. Dennoch

werden für diese Skala alle Items im Itempool behalten, da die interne Konsistenz oberhalb der Grenze von .70 liegt und es aus inhaltlichen Gründen sinnvoll erscheint, alle Buchstaben zu erfassen.

### 11.4.2.5 Laut-Buchstaben-Zuordnung

Dieser Test wurde zum ersten und zum dritten Messzeitpunkt durchgeführt. Alle 26 Grundbuchstaben (ohne Umlaute und ß) sollten von den Kindern aufgeschrieben werden. In einer vorher festgelegten zufälligen Reihenfolge wurden die Buchstaben laut aufgesagt, woraufhin die Schülerinnen und Schüler den entsprechenden Buchstaben in ein leeres Feld eintragen sollten. Die Buchstaben o und l wurden als Übungsitems genutzt, in der Folge im Test aber noch einmal abgefragt.

Die interne Konsistenz liegt zu T1 mit  $\omega_{Total} = .92$  im sehr guten und zu T3 mit  $\omega_{Total} = .77$  im brauchbaren Bereich (siehe Tabelle 4). Item 9 (Buchstabe „o“) ist das einzige Item zu T1, welches mit .20 keine ausreichende Trennschärfe erreicht (s. Tabelle 126 in Anhang A). Alle übrigen Items liegen im brauchbaren bis sehr hohen Bereich. Zu T3 haben dagegen 11 der 26 Items keine ausreichende Trennschärfe von über .25. Ein weiteres Item (Buchstabe „e“) wurde zu T3 von allen teilnehmenden Kindern korrekt beantwortet. Insgesamt sind diese Items für die Kinder zu diesem Zeitpunkt sehr leicht, was sich auch an Werten von über 80 in der Itemschwierigkeit zeigt. Lediglich für die Buchstaben q, j, c, x und y zeigen sich gute Itemschwierigkeiten zwischen 20 und 80. Es muss im Folgenden berücksichtigt werden, dass zu T3 aufgrund der geringen Schwierigkeit der Items mit Deckeneffekten zu rechnen ist. Auch wenn für den Buchstaben „o“ ebenfalls nur eine niedrige Trennschärfe (.20 zu T1 und .25 zu T3) vorliegt, wird das Item im Itempool behalten, da auf diese Weise sämtliche Buchstaben mit der Skala erfasst werden.

### 11.4.3 Schriftspracherwerb

Zur Erfassung der Rechtschreib- und Leseleistung kamen das selbst erstellte Verfahren „Schreiben nach Bildern“ sowie standardisierte normierte Tests zum Einsatz. Die Rechtschreibleistung wurde mit den DERET-Verfahren („Deutscher Rechtschreibtest für das erste und zweite Schuljahr“ sowie zu Messzeitpunkt T7 „Deutscher Rechtschreibtest für das dritte und vierte Schuljahr“; Stock & Schneider, 2008a, 2008b) erhoben. Die Leseleistung wurde mit dem Test ELFE 1-6 (Lenhard & Schneider, 2006) erfasst. Die Verfahren werden im Folgenden näher erläutert.

#### 11.4.3.1 Schreiben nach Bildern

Angelehnt an Dehn und Hüttis-Graff (2010) wurde eine Wortliste erstellt, welche die frühe Rechtschreibkompetenz erfassen soll. Es wurden Bilder vorgegeben, zu denen die passenden Wörter geschrieben werden sollten. Das jeweilige Wort zum Bild wurde zudem vorgelesen. Dieser Test wurde zu den ersten fünf Messzeitpunkten vorgelegt. Im Verlauf des ersten

Grundschuljahres (Messzeitpunkte T1, T2 und T3) waren es sechs Wörter, welche die Kinder schreiben sollten. Die ersten vier entsprechen den von den Autorinnen vorgeschlagenen Wörtern: Sofa, Mund, Limonade und Turm. Zwei weitere kurze Wörter wurden diesen hinzugefügt: Kamm und Feder. Zu den im zweiten Schuljahr folgenden Messzeitpunkten T4 und T5 wurden fünf weitere schwieriger zu schreibende Wörter hinzugefügt: Leiter, Strom, Biene, Hahn und Regenschirm. Für jedes Item kann ein Kind keinen, einen (für eine phonematisch korrekte Schreibweise) oder zwei Punkte (für eine orthographisch korrekte Schreibweise) erhalten.

Die interne Konsistenz liegt zu T1 bei  $\omega_{Total} = .89$  und zu T2 bei  $\omega_{Total} = .82$  und ist somit als gut anzusehen (s. Tabelle 4). Die Trennschärfen liegen zu T1 durchgehend im hohen Bereich (s. Tabelle 127 in Anhang A). Zum zweiten Messzeitpunkt liegen die Trennschärfen niedriger, aber immer noch im brauchbaren bis hohen Bereich. Die ermittelten Itemschwierigkeiten beziehen sich auf den Anteil der Kinder, die diese Silben orthographisch korrekt schreiben konnten. Zu T1 gibt es sieben Items mit einer Itemschwierigkeit von unter 20, was bedeutet, dass diese Items nur von wenigen Kindern korrekt bearbeitet wurden. Da gleichzeitig die Trennschärfen aber oberhalb des Grenzwertes von .25 liegen, sind diese Items gut geeignet, Unterschiede zwischen den Kindern festzustellen, sodass alle Items in der Skala beibehalten werden. Zu T2 liegen die Werte der Itemschwierigkeiten höher. Somit konnten mehr Kinder die Items lösen. Es gibt allerdings kein Item mit einem Wert von unter 5 oder über 95. Drei Items haben eine Itemschwierigkeit von über 80. Diese Items wurden von vielen Kindern korrekt bearbeitet. Da auch hier die Trennschärfen oberhalb des unteren Grenzwerts von .25 liegen, kann die Skala auch zu diesem Messzeitpunkt eingesetzt werden.

Zu T3 liegen die Trennschärfen für die ersten beiden Items mit .22 und .11 unterhalb des brauchbaren Bereichs. Diese beiden Items stellen die zwei Silben des getesteten Wortes „Sofa“ dar. Die sehr hohen Werte in den Itemschwierigkeiten von 98 und 87 zeigen, dass diese beiden Silben zum Ende der ersten Klasse bereits sehr leicht sind und von den meisten Kindern korrekt bearbeitet werden. Auch die Items 5 und 10 liegen mit einer Itemschwierigkeit von 90 bzw. 91 in einem sehr hohen Bereich, während die Trennschärfe jeweils brauchbar ist. Insgesamt ist die Skala zu T3 trotz eines einfachen Wortes mit einer internen Konsistenz von  $\omega_{Total} = .73$  als brauchbar anzusehen (s. Tabelle 4) und kann somit auch zu diesem Zeitpunkt noch eingesetzt werden.

Zu T4 liegt die interne Konsistenz bei  $\omega_{Total} = .77$  und zu T5 bei  $\omega_{Total} = .71$  (s. Tabelle 4). Somit sind auch diese Werte als brauchbar anzusehen. Allerdings zeigen sich für die ersten beiden Silben „so“ und „fa“ zu T4 und T5 sehr niedrige Trennschärfen zwischen .00 und .18. Gleichzeitig liegen die Itemschwierigkeiten (bezogen auf die orthographisch korrekte Schreibweise) der zweiten Silbe bei 83 und 84 und somit im sehr guten Bereich. Zu beiden Messzeitpunkten existieren allerdings weitere einzelne Items mit nicht akzeptablen Trennschärfen von unter .25. Diese sind allerdings jeweils Bestandteile von Wörtern, die über mindestens eine Silbe verfügen, deren Trennschärfe im brauchbaren bis guten Bereich liegt, wie zum

Beispiel bei dem Wort „Biene“. Die erste Silbe erreicht eine Trennschärfe von .45 zu T4 und .30 zu T5. Die zweite Silbe erreicht dagegen eine Trennschärfe von .21 zu T4 und .09 zu T5. Die Itemschwierigkeiten zeigen, dass einzelne Silben vergleichsweise leicht sind und von vielen Kindern korrekt geschrieben werden, wie beispielsweise die zweite Silbe „ne“ des Wortes Biene mit einer Itemschwierigkeit von 97 zu beiden Messzeitpunkten.

Da die internen Konsistenzen zu beiden Messzeitpunkten brauchbar sind und bei allen Wörtern mindestens eine Silbe über eine brauchbare Trennschärfe verfügt (mit Ausnahme des ersten Wortes „Sofa“), werden alle Items als Teil der Skala beibehalten.

### 11.4.3.2 Rechtschreibleistung

Die Rechtschreibleistung der Kinder wurde mit dem standardisierten Verfahren DERET im Klassensetting erhoben. Der DERET 1-2+ (Stock & Schneider, 2008a) wurde zu T3 und T6 verwendet. Zum letzten Messzeitpunkt befanden sich die Kinder in der vierten Klassenstufe, sodass der DERET 3-4+ (Stock & Schneider, 2008b) zum Einsatz kam.

Der DERET 1-2+ kann unter anderem zum Ende des ersten Schuljahres (T3) sowie zum Ende des zweiten Schuljahres (T6) durchgeführt werden. Der Test besteht jeweils aus einem Lückentext und aus einem kurzen Fließtext, der den Kindern diktiert wird. In dieser Studie wurde nur der Fließtext diktiert. Zu Messzeitpunkt T3 wurde Form A für die erste Klasse verwendet. Der diktierte Text besteht aus 29 Wörtern. Die interne Konsistenz liegt laut Manual bei  $\alpha = .90$  (Stock & Schneider, 2008a, S. 39). In der zweiten Klasse zu Messzeitpunkt T6 wurde Form A des Tests für die zweiten Klassenstufen verwendet. Der Fließtext besteht aus 52 Wörtern. Die interne Konsistenz liegt laut Manual bei  $\alpha = .92$  (Stock & Schneider, 2008a, S. 39). In der vierten Klassenstufe zu Messzeitpunkt T7 wurde Form B des DERET 3-4+ verwendet. Für diese Klassenstufe besteht der Fließtext aus 92 Wörtern. Das Manual gibt die interne Konsistenz mit  $\alpha = .92$  an (Stock & Schneider, 2008b, S. 41).

### 11.4.3.3 Lesekompetenz

Die Leseleistung wurde mit ELFE 1-6 (Lenhard & Schneider, 2006) erhoben. Mithilfe dieses Testverfahrens kann das Leseverständnis auf Wort-, Satz- und Textebene in den Schulklassen 1 bis 6 erhoben werden. Details zum Aufbau des Tests sind in Kapitel 9.1.3.3 beschrieben. Wie Tabelle 3 zu entnehmen ist, wurde das Leseverständnis auf Wortebene an fünf Messzeitpunkten (T2 bis T4, T6 und T7) erhoben. Das Leseverständnis auf Satzebene wurde an drei (T4, T6, T7) und das Leseverständnis auf Textebene an zwei Messzeitpunkten (T6 und T7) erhoben.

Das Manual gibt die interne Konsistenz für den Wortverständnistest mit  $\alpha = .97$  an. Für den Satzverständnistest liegt eine interne Konsistenz von  $\alpha = .93$  vor und für den Textverständnistest von  $\alpha = .92$  (Lenhard & Schneider, 2006, S. 9).

#### 11.4.4 Schulisches Selbstkonzept und Lernmotivation

Das schulische Selbstkonzept und die Lernmotivation wurden zu Messzeitpunkt T2 mit einem Test erhoben, der eine Adaptation des SELLMO (Spinath, Stiensmeier-Pelster, Schöne & Dickhäuser, 2012) und des SESSKO (Schöne, Dickhäuser, Spinath & Stiensmeier-Pelster, 2012) für die erste Klasse darstellt (Stiensmeier-Pelster & Schöne, o. J.). Es wurde auf einen unveröffentlichten Fragebogen in Entwicklung zurückgegriffen, da es nichts Vergleichbares für Kinder in der ersten Klasse gab und SELLMO sowie SESSKO erst ab der dritten Klasse einsetzbar sind.

Der verwendete Fragebogen besteht aus insgesamt 22 Items mit einer fünfstufigen Antwortskala. Analog zum SESSKO sollen die Kinder ihre schulischen Leistungen und Lernfortschritte einschätzen. Es gibt dabei Aussagen zu vier Vergleichsskalen: sozial (im Vergleich mit Mitschülerinnen und -schülern), kriterial (verglichen mit den Anforderungen in der Schule), individuell (im Vergleich mit früheren Zeitpunkten) und absolut (hier wird keine Bezugsnorm verwendet). Erfragt werden ebenso die Lern- und Leistungsziele der Kinder. Diese Fragen sind an den SELLMO angelehnt. Als dritten Bereich finden sich Fragen zur intrinsischen Lernmotivation in Bezug auf Lesen, Schreiben und Rechnen.

Die einzelnen Antwortmöglichkeiten werden auf dem Fragebogen durch je fünf Sterne repräsentiert, wobei eine höhere Zustimmung durch einen von links nach rechts immer größer werdenden Stern dargestellt wird. So lautet die erste Frage zum Beispiel „Wie gut bist du in der Schule?“. Es liegt eine fünfstufige Antwortskala von *gar nicht gut* bis *sehr gut* vor. Die Antwort *gar nicht gut* wird auf der linken Seite der Antwortskala mit einem sehr kleinen Stern verbildlicht, die Antwort *sehr gut* dagegen auf der rechten Seite mit einem sehr großen Stern. Alle Fragen sowie alle möglichen Antworten wurden den Kindern der Reihe nach vorgelesen, sodass sie nichts selbst erlesen mussten und folglich Lesefehler aufgrund schwacher Lesekenntnisse ausgeschlossen werden können. Die Bearbeitungsdauer betrug ca. 10 Minuten.

Tabelle 5

*Interne Konsistenzen der Skalen des Fragebogens zu schulischem Selbstkonzept und Lernmotivation für die erste Klasse*

Skala	Items	Cronbachs $\alpha^*$	$\omega_{Total}^{**}$
Subskala SESSKO	01, 02, 03, 10, 11, 12, 13	.75	.73 [.65, .81]
Selbstkonzept Mathematik	04, 05, 06	.72	.82 [.75, .88]
Selbstkonzept Deutsch	07, 08, 09	.71	.68 [.58, .79]

*Anmerkungen.* \*Interne Konsistenz Cronbachs  $\alpha$  aus den unveröffentlichten Daten der Vortestung von Stiensmeier-Pelster und Schöne (o. J.). \*\* $\omega_{Total}$  wurde mit den Daten dieser Untersuchung berechnet.

Eine Erhebung von Stiensmeier-Pelster und Schöne (o. J.) mit ca. 800 Schülerinnen und Schülern ergab die in Tabelle 5 aufgeführten internen Konsistenzen  $\alpha$ . Da die internen Konsistenzen nur für die Subskalen *Gesamtskala SESSKO*, *Selbstkonzept Mathe* sowie

*Selbstkonzept Deutsch* ausreichend hoch ausgefallen sind, gehen auch nur diese Subskalen in weitere Berechnungen ein.

Mit Ausnahme von Item 2 der Subskala SESSKO erreichen alle Items gute bis hohe Trennschärfen von .41 bis .64 (s. Tabelle 6). Die interne Konsistenz liegt sowohl in der Vortestung wie auch in der vorliegenden Stichprobe im brauchbaren Bereich. Die Itemschwierigkeiten sind mit 82 bis 90 hoch, übersteigen aber nicht den kritischen Wert von 95.

Für das Selbstkonzept Mathematik lag in der Vortestung eine brauchbare und für die vorliegende Stichprobe eine gute interne Konsistenz vor. Die berechneten Trennschärfen sind mit .64 bis .70 hoch (s. Tabelle 6). Die Itemschwierigkeiten der drei Items sind mit 84 bis 88 hoch, aber ebenfalls unterhalb des kritischen Bereichs von 95.

Tabelle 6  
*Trennschärfe und Itemschwierigkeit für die Subskalen*

		Trennschärfe	Itemschwierigkeit
Subskala SESSKO	Item 1	.43	86
	Item 2	.24	82
	Item 3	.41	90
	Item 10	.56	88
	Item 11	.54	88
	Item 12	.64	90
	Item 13	.41	88
Selbstkonzept Mathematik	Item 4	.70	88
	Item 5	.66	84
	Item 6	.64	88
Selbstkonzept Deutsch	Item 7	.57	88
	Item 8	.39	84
	Item 9	.54	86

Die Subskala zum Selbstkonzept Deutsch erreicht in der vorliegenden Untersuchung mit  $\omega_{Total} = .68$  nicht die untere Grenze von .70. Unter Berücksichtigung des Konfidenzintervalls liegt die interne Konsistenz mit einer Wahrscheinlichkeit von 95 Prozent zwischen .58 und .79 (s. Tabelle 5). In der unveröffentlichten Originaluntersuchung lag die interne Konsistenz dagegen bei  $\alpha = .71$ , weswegen noch die Trennschärfen und die Itemschwierigkeiten für die vorliegende Stichprobe betrachtet werden: Die Trennschärfen für die drei Items dieser Subskala liegen mit .39 bis .57 im brauchbaren bis hohen Bereich.

Die Itemschwierigkeiten liegen zwischen 84 und 88 und sind somit recht hoch, aber unterhalb des kritischen Bereichs. Diese Skala wird aufgrund des Ergebnisses der Vortestung sowie der brauchbaren bis hohen Trennschärfen verwendet.

#### 11.4.5 Klassenklima und Fähigkeitsselbstkonzept

Das Klassenklima und das Fähigkeitsselbstkonzept wurden mit dem Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen (FEES 1-2 von Rauer & Schuck, 2003a) bzw. zu Messzeitpunkt T7 mit der Version für die dritten und vierten Klassen (FEES 3-4 von Rauer & Schuck, 2003b) erhoben. Dieser

Test umfasst zwei Teilfragebögen, welche auch separat verwendet werden können. In dieser Studie wurde nur der erste der beiden Teilfragebögen, der SIKS, eingesetzt. Dieser beinhaltet drei Subskalen: das Fähigkeitsselbstkonzept (SK), die erlebte soziale Integration in der Klasse (SI) sowie das wahrgenommene Klassenklima (KK).

Um Vergleiche zwischen den Klassenstufen zu gewährleisten, sind die Aussagen im FEESS 1-2 mit den Aussagen im FEESS 3-4 identisch. Im FEESS 1-2 werden alle zu beantwortenden Aussagen vorgelesen. Als Antwortmöglichkeiten gibt es *stimmt* und *stimmt nicht*. Für die dritten und vierten Klassen stehen jeweils vier Antwortmöglichkeiten zur Verfügung (*stimmt gar nicht*, *stimmt kaum*, *stimmt ziemlich* und *stimmt genau*). Die Bearbeitungsdauer für den Teilfragebogen SIKS beträgt jeweils ca. 30 Minuten. Es liegen für alle vier Klassenstufen in den jeweils entsprechenden Testversionen sowohl Individualnormen wie auch Normen für die ganze Klasse vor.

Das Manual des FEESS 1-2 gibt die interne Konsistenz für die Individualdaten der Eichstichproben in der ersten Klasse für die Subskala SI mit  $\alpha = .69$  an, für SK mit  $\alpha = .78$  und für KK mit  $\alpha = .63$ . Für die zweite Klasse beträgt die interne Konsistenz für die Subskala SI  $\alpha = .72$ , für SK  $\alpha = .76$  und für KK  $\alpha = .63$  (Rauer & Schuck, 2003a, S. 56). Die interne Konsistenz für den FEESS 3-4 beträgt laut Manual in der vierten Klasse für SK  $\alpha = .85$ , für SI  $\alpha = .84$  und für KK  $\alpha = .77$  (Rauer & Schuck, 2003b, S. 51).

Sollten Kinder einzelne Items auslassen oder nicht eindeutig angekreuzt haben, so ist für den FEESS 1-2 per Münzwurf zu entscheiden, welcher Wert, 1 oder 0, für diesen fehlenden Wert eingefügt wird. Dies gilt aber nur, solange nur ein Item pro Skala ersetzt werden muss. Bei Fehlen mehrerer Items kann die jeweilige Skala für dieses Kind nicht ausgewertet werden (Rauer & Schuck, 2003a, S. 26). Das gleiche gilt für den FEESS 3-4, wobei per Münzwurf zwischen den beiden mittleren Werten, 1 und 2, entschieden wird (Rauer & Schuck, 2003b, S. 21).

#### 11.4.6 Erfassung der kognitiven Fähigkeiten

Mitte des zweiten Schuljahres (T5) wurden Tests zum logischen Denken sowie zur Bestimmung der Benennungsgeschwindigkeit und zur Kapazität des Kurzzeit- und des Arbeitsgedächtnisses durchgeführt (s. Tabelle 3). Die Tests werden in den folgenden Unterkapiteln vorgestellt.

##### 11.4.6.1 Logisches Denken

Das logische Denken der Kinder wurde mit den Coloured Progressive Matrices (CPM) erhoben (Bulheller & Häcker, 2002). Es handelt sich hierbei um einen standardisierten und sprachfreien Intelligenztest, der das logische Schlussfolgern erfasst und sowohl in Einzel- als auch in Gruppensettings durchgeführt werden kann. Normen liegen für den Altersbereich von 3;9 bis 11;8 Jahren vor. Die interne Konsistenz wird als zufriedenstellend bis gut bezeichnet, wobei Werte von .71 bis .90 angeführt werden (Bulheller & Häcker, 2002, S. 23; S. 28). Ein

höherer Wert in den CPM drückt nicht aus, dass ein Kind Aufgaben schneller löst als ein Kind mit einem niedrigeren Wert. Es ist vielmehr ein Zeichen dafür, dass das Kind mit einem höheren Wert eine qualitativ andere Art des Denkens erworben hat (Bulheller & Häcker, 2002, S. 47).

Es liegen insgesamt 36 zu bearbeitende Items vor. Für jedes Item wird ein buntes Muster vorgegeben, in dem sich eine Lücke befindet. Aus sechs Auswahlmöglichkeiten soll das Kind diejenige auswählen, die das Muster richtig ergänzt. Durchgeführt wurde der Test in der Testheftversion im Klassenverband und erfolgte entsprechend der Angaben des Manuals ohne zeitliche Begrenzung, um den Kindern die Möglichkeit zu geben, alle Items zu bearbeiten.

### 11.4.6.2 Benennungsgeschwindigkeit

Die Benennungsgeschwindigkeit wurde in Einzelsitzungen mithilfe zweier Aufgaben aus dem Rapid Automated Naming Test von Wolf und Denckla (2005) erhoben. Dem Kind ist zunächst eine Karte mit Bildern vorgelegt worden. Die oberste Reihe dient als Beispiel und zeigt fünf Gegenstände (Auto, Haus, Fisch, Stift und Ball), welche vom Kind benannt werden sollen. Hierüber wird sichergestellt, dass das Kind die abgebildeten Gegenstände entsprechend benennen kann. Anschließend werden fünf Zeilen bearbeitet. In jeder Zeile sind nebeneinander zehn Gegenstände abgebildet, wobei jeder der fünf oben genannten zweimal in zufälliger Reihenfolge vorkommt. Die Aufgabe des Kindes ist es, die gezeigten Bilder in der Reihenfolge, in der sie dargestellt sind, so schnell wie möglich zu benennen. Der Subtest mit Buchstaben ist vom Aufbau und Ablauf her identisch. Statt der fünf Bilder sind es allerdings die Buchstaben O, A, S, T und P, die genannt werden sollen. Gemessen wird die Zeit, welche das Kind für alle 5 Reihen jeweils benötigt. Zusätzlich werden die Anzahl der Fehler, die korrigiert werden und die Anzahl derer, die nicht korrigiert werden, notiert. Für jede Subskala werden die korrekt genannten Buchstaben bzw. Bilder pro Sekunde berechnet. Der Gesamtscore wird anschließend aus den beiden Subskalen und ermittelt.

Die Test-Retest-Reliabilität wird von den Autorinnen mit .84 bis .92 für alle Altersgruppen angegeben (Wolf & Denckla, 2005, S. 29).

### 11.4.6.3 Arbeitsgedächtnis

In Einzelsitzungen wurde der Untertest Zahlen nachsprechen aus dem HAWIK-IV (Petermann & Petermann, 2010) durchgeführt. Dieser ist einer von drei Untertests, die zur Erfassung des Arbeitsgedächtnisses verwendet werden. Er besteht aus zwei Teilen, dem Zahlen nachsprechen (vorwärts) und dem Zahlen nachsprechen (rückwärts). Die interne Konsistenz der beiden Skalen liegt zwischen .71 und .83 und ist somit als zufriedenstellend bis hoch zu bezeichnen (Petermann & Petermann, 2010, S. 115).

Mit dem Untertest Zahlen nachsprechen (vorwärts) wird das auditive Kurzzeitgedächtnis erfasst. Das Kind soll Zahlen, die die Testleiterin vorgesprochen hat, in der gleichen Reihenfolge wiederholen. Mit dem Untertest Zahlen nachsprechen (rückwärts) wird die Fähigkeit zur

Manipulation verbaler Informationen in einem kurzen Zeitraum ermittelt, wobei das Zahlen nachsprechen (vorwärts) als Vorläuferfertigkeit gesehen wird (Petermann & Daseking, 2009, S. 229). In diesem Test soll das Kind die Zahlen, die die Testleiterin vorgegeben hat, in umgekehrter Reihenfolge wiederholen.

### 11.5 Befragung der Lehrkräfte

Während der Erhebungen waren die Grundschullehrkräfte im Klassenraum anwesend. Über die Zielsetzung und die Fragestellungen der Untersuchung waren sie nicht informiert worden, wussten aber, dass die Schriftsprachentwicklung im Fokus steht. Im Anschluss an die jeweilige Erhebung, insgesamt folglich sieben Mal, wurden die Lehrkräfte jeweils von einer Untersuchungsleiterin zu ihren Schülerinnen und Schülern befragt. Hierüber soll Einblick in die förderdiagnostische Kompetenz der Lehrkräfte erlangt werden. Da die Lehrkräfte in ihren Antworten möglichst wenig beeinflusst werden sollten, wurde während des gesamten Erhebungszeitraums kein Fokus auf die diagnostischen Fertigkeiten der Lehrkräfte gelegt, weswegen von weiteren Fragen oder Nachfragen abgesehen wurde.

Die Befragung wurde mit dem ersten Satz eingeleitet „Wir haben jetzt die Tests durchgeführt“, worauf sich direkt die Frage anschloss: „Jetzt sagen Sie bitte: Welches Kind fällt aus Ihrer Sicht auf?“ Wenn ein Kind genannt wurde, schlossen sich die beiden Fragen an: „Was fällt Ihnen auf?“ sowie „Was wird getan?“. Die Antworten wurden stichpunktartig mitgeschrieben. Die Klassenliste lag aus und die Lehrkräfte konnten sich zu jeder Schülerin und zu jedem Schüler äußern.

Da es sich um Antworten auf offene Fragen handelt, müssen die Antworten zunächst Kategorien zugewiesen werden, bevor sie für weitere Analysen verwendet werden können. Die Antwortkategorien wurden vor Beginn der Auswertungen festgelegt. Die Kodierung der Antworten erfolgte durch zwei nicht an den Erhebungen beteiligte Beurteilerinnen. Die Antwortkategorien beziehen sich jeweils auf aus der Theorie hergeleitete relevante Faktoren. Da sich die Art der Kodierung für die beiden offenen Fragen unterscheidet, wird das Vorgehen im Folgenden getrennt für beide Fragen beschrieben. Abschließend wird die Beurteilerübereinstimmung für die Antwortkategorien dargestellt.

#### 11.5.1 Kodierung zu Frage 1: „Was fällt auf?“

Tabelle 7 listet die Inhaltskategorien auf, die den beiden Beurteilerinnen zur Kodierung der Antworten vorlagen. Diese wurden in Oberkategorien noch weiter zusammengefasst. Entsprechend des Schwerpunkts der vorliegenden Untersuchung wurden zunächst Inhalte ausgewählt, die sich in der Oberkategorie *Spezifisch auf den Schriftspracherwerb bezogen* subsumieren lassen. Hierzu gehören Aussagen, die sich direkt auf den Leseprozess beziehen, aber auch Aussagen über Teilfaktoren oder Voraussetzungen, wie die Buchstabenkenntnis. In der Oberkategorie *Allgemeiner auf den Schriftspracherwerb bezogen* wurden Inhalte erfasst, die keine spezifischen Aspekte benennen, sondern sich allgemein auf Lesen oder

Schreiben beziehen. Da sich Lesen und Schreiben nicht unabhängig voneinander entwickeln, ist es sinnvoll, beide Kompetenzen zu berücksichtigen.

Tabelle 7

*Einteilung der Faktoren zu Frage 1: „Was fällt auf?“*

Oberkategorie	Inhalte
Spezifisch auf den Schrift-spracherwerb bezogen	Phonologische Bewusstheit; Lesegeschwindigkeit; Buchstabenkenntnis; Laut-Buchstaben-Zuordnung; Lesefertigkeit; Leseverständnis
Allgemeiner auf den Schrift-spracherwerb bezogen	Lese-Rechtschreibstörung; Lesen allgemein; Schreiben allgemein; Deutsch allgemein
Sprache	Sprachauffälligkeiten oder niedrige Deutschkenntnisse
Intelligenz / FSP Lernen	Negative Aussagen zur Intelligenz des Kindes; Förderschwerpunkt Lernen (besteht oder wird vermutet)
Motorik	Motorische Auffälligkeiten
Verhalten und Motivation	Konzentrationsfähigkeit / ADHS; Verhaltensauffälligkeiten; Arbeitsverhalten; soziale Interaktion; Motivation
Sonstige Auffälligkeiten	Emotionale Probleme; Wahrnehmungsstörungen; negative Aussagen über das Elternhaus; sonstige Auffälligkeiten
Gut im Schreiben	Positive Aussagen zu Schreibfähigkeiten
Gut im Lesen	Positive Aussagen zu Lesefähigkeiten
Sonstige positive Nennungen	Positive Aussagen über: Elternhaus; Intelligenz des Kindes; sonstige positive Auffälligkeiten

*Anmerkungen.* Die Spalte Inhalte zeigt die einzelnen Kategorien, welche den Beurteilerinnen vorlagen.

In weiteren Oberkategorien wurden Inhalte berücksichtigt, die einen negativen Einfluss auf den Leseerwerb haben können. Zum einen sind das sprachliche Auffälligkeiten des Kindes, aber auch kognitive Einflussfaktoren wie die Intelligenz der Kinder oder motorische Einschränkungen. Zur Oberkategorie *Verhalten und Motivation* zählen Inhalte zur Konzentrationsfähigkeit, zur sozialen Interaktion, zur Motivation und zum Arbeitsverhalten.

Unter der Oberkategorie *Sonstige Auffälligkeiten* wurden verschiedene Inhalte zusammengefasst, die sich nicht in die übrigen Oberkategorien einordnen lassen und sich nicht auf die für die Fragestellung relevanten Inhalte beziehen, wie zum Beispiel negative Aussagen über das Elternhaus, Aussagen über Schwächen in anderen Fächern als Deutsch und weitere Auffälligkeiten, die nicht explizit als Inhaltskategorie berücksichtigt wurden.

Lehrkräfte fokussieren sich nicht nur auf Schwächen, sondern auch auf Stärken und Ressourcen der Kinder. Daher wurden für positive Aussagen drei weitere Kategorien gebildet. Analog zum Fokus der Studie gibt es jeweils eine Oberkategorie für Antworten, die sich auf positive Leistungen des Kindes im Lesen bzw. im Schreiben beziehen. Eine dritte Oberkategorie beinhaltet sonstige positive Nennungen, darunter auch Intelligenz und Aussagen über das Elternhaus.

### 11.5.2 Kodierung zu Frage 2: „Was wird getan?“

Um die Antworten auf die Frage auszuwerten, was bei Auffälligkeiten getan wird, wurde sich an § 2 VOSB (Verordnung über Unterricht, Erziehung und sonderpädagogische Förderung

von Schülerinnen und Schülern mit Beeinträchtigungen oder Behinderungen im Landesrecht Hessen) orientiert. Dieser Paragraf gibt an, dass von der Schule vorbeugende Maßnahmen getroffen werden sollen. In Absatz 1 werden differenzierende und individualisierende Arbeitsformen, Förderung in Einzel- oder Kleingruppen und Beratung von Eltern sowie Zusammenarbeit mit Beratungs- und Förderzentren (BFZ) und außerschulischen Einrichtungen genannt. Diese werden in Absatz 3 desselben Paragrafen weiter spezifiziert.

Tabelle 8  
Faktoren zu Frage 2: „Was wird getan?“

Oberkategorie	Inhalte
Organisatorische innerschulische Aspekte	Förderung / Unterstützung im Klassenzimmer Lehrkraft selbst fördert BFZ-Kraft fördert sonstige Person oder unklar, wer fördert Förderunterricht Einzelförderung Gruppenförderung Sonstige innerschulische Förderung
Organisatorische Aspekte der Zusammenarbeit	Elterngespräche Außerschulische Hilfe BFZ ist eingeschaltet
Inhaltliche Aspekte	Differenzierung Lob bzw. positive Verstärkung Motivationsförderung Förderplan Spezifische Fördermethoden Sonstige Inhalte
Sonstiges	Klassenwiederholung Anspruch auf sonderpädagogische Förderung festgestellt Sonstige professionelle Förderung

*Anmerkungen.* Spalte Inhalte listet die einzelnen Kategorien auf, welche den Beurteilerinnen vorlagen

Es wurden entsprechende Kategorien für die Antworten der Lehrkräfte gebildet. Eine vollständige Auflistung findet sich in Tabelle 8. In den ersten beiden Oberkategorien wurden Inhalte zur organisatorischen Umsetzung berücksichtigt. Die erste hiervon fokussiert auf innerschulische Aspekte, während die zweite sich auf die oben genannte Zusammenarbeit mit außerschulischen Institutionen, mit dem BFZ sowie mit den Eltern bezieht.

Inhaltliche Aspekte wurden in einer dritten Oberkategorie erfasst. Hier finden sich Differenzierung und weitere Maßnahmen, die eine Lehrkraft durchführen kann, wie positive Verstärkung, das Erstellen eines Förderplans oder die Anwendung spezifischer Fördermethoden. Auch für diese Frage wurde eine Oberkategorie für sonstige Antworten gebildet, die aus Aussagen über eine Klassenwiederholung, Feststellen des Anspruchs auf sonderpädagogische Förderung durch das BFZ und sonstige professionelle Förderung, welche nicht zu den oben genannten inhaltlichen Aspekten gehört, besteht.

### 11.5.3 Ablauf der Kodierung

Die Kodierung wurde nach Abschluss der Datenerhebung und der Dateneingaben Mitte bis Ende 2017 durch zwei Beurteilerinnen vorgenommen, die an der Datenerhebung nicht beteiligt gewesen waren. Es handelte sich um studentische Hilfskräfte, die das Lehramt an Förderschulen studierten. Zum Zeitpunkt der Beurteilungen hatten beide bereits zwei Schulpraktika an einer Förderschule und an einer regulären Grundschule im Rahmen eines Praxissemesters abgeschlossen. Über die Fragestellungen der Studie waren die beiden Beurteilerinnen nicht informiert. Es war lediglich bekannt, dass es sich um eine Studie zum Schriftspracherwerb mit sieben Messzeitpunkten handelt und Lehrkräfte die genannten Fragen beantwortet haben.

Zunächst wurde den Beurteilerinnen der Leitfaden zur Kodierung der ersten Frage gegeben. Sie wurden anschließend in persönlichen Gesprächen im Vorgehen geschult. Dabei konnten ebenfalls Fragen von Seiten der Beurteilerinnen gestellt werden, wenn Angaben im Leitfaden unklar waren. Da sich viele Inhalte zur ersten Frage ähnlich sind, wurden die Beurteilerinnen angehalten, die Aussagen der Lehrkräfte in höchstens drei Variablen zu kodieren. Zugleich sollten die Beurteilerinnen Rückmeldung geben, falls sie mehr als drei Variablen benötigen sollten. Dies sei jedoch nicht notwendig gewesen. Zu jeder Aussage wurden dementsprechend bis zu drei Ziffern vergeben.

Auch für die zweite Frage erhielten die beiden Beurteilerinnen einen Leitfaden zur Kodierung. Es schloss sich ebenfalls die Schulung in persönlichen Gesprächen an. Im Unterschied zur ersten Frage sollten bei dieser Frage für jede Aussage der Lehrkräfte alle in Tabelle 8 genannten Inhalte geprüft und entsprechend kodiert werden. Die Beurteilerinnen sollten hierzu zunächst entscheiden, ob der Inhalt in der Aussage genannt wurde und, wenn ja, ob dies a) nicht zur Förderung von Lesen, b) vermutlich (auch) zur Förderung von Lesen oder c) ganz sicher (auch) zur Förderung von Lesen geschieht. Diese Unterscheidung wurde gewählt, da Schülerinnen und Schüler in vielen Bereichen Schwierigkeiten haben können und dementsprechend auch in verschiedenen Bereichen Unterstützung erhalten. Nicht immer wird dies aber explizit für Lesen genannt. Die Unterscheidung zwischen b) und c) ist daher notwendig, da häufig von Schwierigkeiten im Schriftspracherwerb allgemein oder „im Fach Deutsch“ gesprochen wird und hier sowohl Lesen wie auch Schreiben gefördert werden können. Eine Aussage „erhält Förderung im Fach Deutsch“ würde entsprechend kodiert werden unter „b) vermutlich (auch) zur Förderung von Lesen“, da Lesen nicht explizit genannt wird und vielleicht auch nicht im Fokus der Förderung steht, man aber davon ausgehen kann, dass eine solche Förderung alle Bereiche des Schriftspracherwerbs und somit auch Lesen beinhaltet. Gleichzeitig würde man bei dieser Aussage unter *Organisatorische inhaltliche Aspekte* kodieren, dass unklar ist, wer diese Förderung durchführt, da keine Information über die Form der Förderung gegeben wird.

Einige Variablen wurden dichotom erhoben, da diese Bereiche nicht explizit für Lesen betrachtet werden können oder sollten, wie zum Beispiel die Klassenwiederholung, die nicht

allein wegen Schwierigkeiten im Lesen veranlasst wird. Ebenso wird kein Anspruch auf sonderpädagogische Förderung gegeben, wenn ein Kind nur im Lesen Schwierigkeiten hat. Da den beiden Beurteilerinnen die Einteilung der Antworten schwerer fiel als bei der ersten Frage, wurden die Antwortkategorien noch einmal besprochen. So war es ihnen zum Beispiel schwergefallen, zwischen zusätzlicher Förderung und Differenzierung zu unterscheiden. Es wurde sich dann gemeinsam darauf festgelegt, dass Differenzierung die Antworten betrifft, in denen über mehr Zeit oder leichtere Aufgaben bzw. weniger Lernmaterial (und schwerere oder mehr Aufgaben bei leistungsstarken Kindern) gesprochen wurde. Nach gemeinsamer Klärung der Fragen gingen die Beurteilerinnen ihre jeweiligen Kodierungen noch einmal durch.

#### 11.5.4 Beurteilerübereinstimmungen

Bei kategorialen Daten und zwei Beurteilenden kann zur Berechnung der Beurteilerübereinstimmung Cohens Kappa verwendet werden (Gisev, Bell & Chen, 2013, S. 333). Zusätzlich wird empfohlen, die prozentuale Übereinstimmung anzugeben, welche allerdings nicht die Wahrscheinlichkeit überprüft, dass eine Übereinstimmung nur durch Zufall entstanden ist (Gisev et al., 2013, S. 333). Wirtz und Kutschmann (2007, S. 2) halten entsprechend fest, dass die prozentuale Übereinstimmung allein kein ausreichendes Maß ist, da eine Übereinstimmung von 80 Prozent vorliegen kann, weil die Beurteilenden in 80 Prozent der Fälle das Nichtvorliegen eines Merkmals festgestellt haben, aber keine Übereinstimmungen bezüglich des Vorliegens eines Merkmals hatten.

Cohens Kappa kann sowohl für dichotome wie auch für mehrstufige Antwortkategorien verwendet werden (Wirtz & Kutschmann, 2007, S. 5). Die Formel zur Berechnung dieses Koeffizienten lautet wie folgt :  $\kappa = \frac{\%Ü_o - \%Ü_e}{100 - \%Ü_e}$  (Wirtz & Kutschmann, 2007, S. 3). Der Wert  $\%Ü_o$  steht für den prozentualen Anteil an übereinstimmenden Urteilen und  $\%Ü_e$  für den prozentualen Anteil an durch Zufall zu erwartenden Übereinstimmungen. Cohens Kappa kann Werte zwischen -1 und +1 annehmen, wobei +1 eine perfekte Übereinstimmung darstellt. Ein Wert über 0 bedeutet, dass die Übereinstimmung höher ist, als dies durch reinen Zufall geschehen würde. Gisev et al. (2013, S. 333) führen an, dass Werte unter .40 sehr häufig als schlechte Übereinstimmung, zwischen .40 und .75 als angemessene bis gute Übereinstimmung und Werte ab .75 als exzellente Übereinstimmung angegeben werden. Dieser Einteilung schließen sich auch Wirtz und Kutschmann (2007, S. 3) an, wobei diese noch einmal differenzieren zwischen Werten von .40 bis .59 als mäßige und Werten von .60 bis .75 als gute Übereinstimmung.

Ein Problem bei Cohens Kappa ist allerdings dessen Abhängigkeit von der Prävalenz der Antworten (Gisev et al., 2013, S. 334). So wird kritisiert, dass der Wert für Kappa umso niedriger ausfällt, je geringer die Grundrate des Merkmals ist (Wirtz & Caspar, 2002, S. 59; Wirtz & Kutschmann, 2007, S. 5). Daher kann auch für ein schwer zu erfassendes Merkmal bereits ab einem Wert von 0.5 eine zufriedenstellende Übereinstimmung vorherrschen (Wirtz &

Caspar, 2002, S. 59). Wirtz und Caspar (2002, S. 59) halten schließlich fest, dass alle Werte zu Cohens Kappa nur allgemein und als Faustregel gelten.

Es ist davon auszugehen, dass einige Antwortkategorien, wie beispielsweise die Klassenwiederholung oder das Hinzuziehen des BFZs, erwartungsgemäß nur auf wenige Schülerinnen und Schüler zutreffen und die Häufigkeit dieser Merkmale entsprechend niedrig sein wird. Wirtz und Kutschmann (2007, S. 5) schlagen zumindest für dichotome Antwortkategorien zusätzlich Yules  $Y$  vor, welches den Einfluss der Grundrate des Merkmals korrigiert:

$$\text{Yules } Y = \frac{\sqrt{OR}-1}{\sqrt{OR}+1}$$

Cohens Kappa wird daher in dieser Untersuchung nicht als alleiniges Maß herangezogen. Zusätzlich werden die prozentuale Übereinstimmung sowie Yules  $Y$  für dichotome Antwortkategorien berichtet. Für mehrstufige Antwortformate liegt kein entsprechender Koeffizient vor, sodass für diese Fälle nur auf Cohens Kappa zurückgegriffen werden kann (Wirtz & Kutschmann, 2007, S. 7). Es werden, ebenfalls in Anlehnung an Wirtz und Kutschmann (2007), Werte ab 0.4 als mäßig, ab 0.5 als zufriedenstellend, ab 0.6 als gut und Werte ab 0.75 als sehr gut angesehen.

Bei Betrachtung der absoluten Häufigkeiten der Beurteilerinnen zur Frage „Was fällt auf?“ (s. Tabelle 128 in Anhang B) zeigt sich, dass einige Inhalte von beiden Beurteilerinnen gar nicht oder nur sehr selten gewählt wurden. Beurteilerin S hat die Kategorie Lesefertigkeit kein einziges Mal und Beurteilerin K nur für eine Aussage gewählt. Andere Kategorien, beispielsweise Sprachauffälligkeiten oder Konzentrationsfähigkeit / ADHS wurden häufiger verwendet. Insgesamt betrachtet zeigt sich, dass mehrere Inhaltskategorien nicht häufig genannt wurden. Es muss daher davon ausgegangen werden, dass Cohens Kappa für diese Aussagen schon aufgrund der geringen Raten niedriger ausfallen wird.

Für die zweite Frage ergibt sich insgesamt ein etwas ausgeglicheneres Bild (s. Tabelle 129 in Anhang B). Auch hier gibt es allerdings Inhaltskategorien, die nur sehr selten gewählt wurden. So wurde Motivationsförderung zu mehreren Messzeitpunkten von Beurteilerin S kein einziges Mal kodiert. Andere Inhaltskategorien wurden deutlich häufiger gewählt, wie beispielsweise Förderung im Klassenzimmer und Förderung durch die Lehrkraft selbst.

#### 11.5.4.1 Beurteilerübereinstimmung zu Frage 1: „Was fällt auf?“

Da es in einigen Inhaltskategorien nur zu wenig Nennungen kam, wurde die Beurteilerübereinstimmung für die Oberkategorien (*Spezifisch auf den Schriftspracherwerb bezogen, Allgemeiner auf den Schriftspracherwerb bezogen, Verhalten und Motivation, Sonstige Auffälligkeiten* und *Sonstige positive Nennungen*) ebenfalls betrachtet. Einige Kategorien bestehen nur aus je einer Inhaltskategorie und werden entsprechend nicht weiter zusammengefasst: *Sprache, Intelligenz / Förderschwerpunkt Lernen, Motorik* sowie *Gut im Schreiben* und *Gut im Lesen*.

Tabelle 9

*Beurteilerübereinstimmungen für die Oberkategorien und einzelnen Inhaltskategorien*

		T1 n = 70	T2 n = 87	T3 n = 67	T4 n = 76	T5 n = 65	T6 n = 96	T7 n = 23
Spezifisch auf den Schriftspracherwerb be- zogen	κ	1.00**	.85**	.93**	.85**	.79**	.71**	.65**
	PÜ	100	97	99	99	99	97	96
	Y	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Allgemeiner auf den Schriftspracherwerb be- zogen	κ		.85**	.83**	.86**	.88**	.74**	1.00**
	PÜ		97	94	95	95	91	100
	Y		1.00	.86	1.00	1.00	.81	1.00
Verhalten und Motivation	κ	.85**	.94**	.81**	.73**	.89**	.89**	.62**
	PÜ	93	98	91	88	95	95	91
	Y	1.00	1.00	.83	.83	1.00	1.00	1.00
Sonstige Auffälligkeiten	κ	.77**	.83**	.71**	.81**	.78**	.72**	.78**
	PÜ	90	93	87	91	89	88	91
	Y	.80	.85	.81	.82	.80	.74	1.00
Sprache	κ	1.00**	.93**	.94**	.75**	.72**	.80**	.83**
	PÜ	100	99	99	96	94	94	96
	Y	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Intelligenz / FSP Lernen	κ	.74**	.85**	1.00**	.85**	1.00**	.93**	1.00**
	PÜ	97	99	100	99	100	99	100
	Y	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Motorik	κ	1.00**		.85**	1.00**	1.00**	.80**	
	PÜ	100		99	100	100	99	
	Y	1.00		1.00	1.00	1.00	1.00	
Gut im Schreiben	κ		.79**					
	PÜ		99					
	Y		1.00					
Gut im Lesen	κ		.94**	.66**			1.00**	
	PÜ		99	99			100	
	Y		1.00	1.00			1.00	
Sonstige positive Nen- nungen	κ	.94**	.85**	.79**	.82**	.93**	.96**	.88**
	PÜ	99	94	97	97	99	99	96
	Y	1.00	.87	1.00	.90	1.00	1.00	1.00

*Anmerkungen.* n = Zahl der Kinder zu jedem Messzeitpunkt, zu denen Aussagen getätigt wurden. FSP = Förderschwerpunkt. Leere Zellen bedeuten, keine Beurteilerin hat diese Kategorie gewählt. κ = Cohens Kappa; PÜ = Prozentuale Übereinstimmung; Y = Yules Y. \*p < .05, \*\*p < .01

Die Beurteilerübereinstimmungen für diese Bereiche werden in Tabelle 9 dargestellt. Für die fünf Oberkategorien sowie die Kategorien, die nur aus einem inhaltlichen Punkt bestehen, liegen mit Werten von κ = .65 bis 1.00 durchgehend gute bis sehr gute Übereinstimmungen vor. Y liegt mit .74 bis 1.00 ebenfalls im guten bis sehr guten Bereich. Diese Kategorien gehen entsprechend alle in weitere Auswertungen ein.

Die Lesefertigkeit wurde nur zu einem Messzeitpunkt als Inhaltskategorie gewählt und auch nur von Beurteilerin K für eine Aussage (s. Tabelle 128 in Anhang B). Daher wird diese Inhaltskategorie nicht weiter berücksichtigt. Die Inhaltskategorie Wahrnehmungsstörungen erreicht nur zu zwei von sechs Messzeitpunkten sehr gute Übereinstimmungen (κ = .79 und .74 sowie jeweils Y = 1.00). Zu vier Messzeitpunkten liegt mit κ = 0.00 keine ausreichende Beurteilerübereinstimmung vor. Y kann nicht bestimmt werden, da Beurteilerin K diese Kategorie zu diesen Zeitpunkten nicht ausgewählt hat. Daher wird diese Inhaltskategorie ebenfalls entfernt. Die Beurteilerübereinstimmungen für die übrigen Inhaltskategorien, die in den oben

genannten Oberkategorien zusammengefasst werden können, finden sich in Tabelle 130 in Anhang B.

Es haben 14 Inhaltskategorien zu allen Messzeitpunkten eine Beurteilerübereinstimmung von  $\kappa = .40$  oder höher (s. Tabelle 130 in Anhang B). Zwei Inhaltskategorien (Deutsch allgemein zu T6 sowie soziale Interaktion zu T2) zeigen zu je einem Messzeitpunkt ein  $\kappa$  von unter  $.40$ , während  $Y$  über dem Kriterium von mindestens  $.40$  liegt. Diese beiden Kategorien sind über alle sieben Messzeitpunkte mit insgesamt mäßiger bis sehr guter Beurteilerübereinstimmung zu bewerten und können somit für weitere Analysen verwendet werden.

Tabelle 10

*Endgültige Auswahl der Kategorien zu Frage 1: „Was fällt auf?“*

Oberkategorie	Inhalte
Spezifisch auf den Schriftspracherwerb bezogen	Phonologische Bewusstheit (nicht T4, T6 und T7); Lesegeschwindigkeit (nicht T4 und T6); Buchstabenkenntnis (nicht T4 und T5); Laut-Buchstaben-Zuordnung (nicht T2); Leseverständnis
Allgemeiner auf den Schriftspracherwerb bezogen	Lese-Rechtschreibstörung; Lesen allgemein; Schreiben allgemein (nicht T3); Deutsch allgemein
Sprache	Sprachauffälligkeiten oder niedrige Deutschkenntnisse
Intelligenz / FSP Lernen	Negative Aussagen zur Intelligenz des Kindes; Förderschwerpunkt Lernen (besteht oder wird vermutet)
Motorik	Motorische Auffälligkeiten
Verhalten und Motivation	Konzentrationsfähigkeit / ADHS; Verhaltensauffälligkeiten; Arbeitsverhalten (nicht T7); soziale Interaktion; Motivation (nicht T5)
Sonstige Auffälligkeiten	Emotionale Probleme; negative Aussagen über das Elternhaus (nicht T1); sonstige Auffälligkeiten
Gut im Schreiben	Positive Aussagen zu Schreibfähigkeiten
Gut im Lesen	Positive Aussagen zu Lesefähigkeiten
Sonstige positive Nennungen	Positive Aussagen über: Elternhaus (nicht T6); Intelligenz des Kindes (nicht T4); sonstige positive Auffälligkeiten

*Anmerkungen.* Die Spalte Inhalte zeigt die einzelnen Kategorien, die in die Analysen eingehen. Ausnahmen aufgrund zu niedriger Beurteilerübereinstimmungen stehen jeweils in Klammern. Die Oberkategorien gehen durchgehend in die Analysen ein.

Die Inhaltskategorie Arbeitsverhalten zeigt zu zwei Messzeitpunkten  $\kappa < .40$ . Zu T3 hat  $Y$  dagegen einen Wert von 1.00. Da  $Y$  einen sehr guten Wert aufweist, wird diese Variable in weitere Berechnungen eingehen. Zu T7 weisen weder  $\kappa$  noch  $Y$  einen ausreichenden Wert auf, sodass die Inhaltskategorie zu T7 nicht einzeln berücksichtigt wird.

Die übrigen zwölf Kategorien erreichen zu einzelnen Messzeitpunkten sowohl mit  $\kappa$  wie auch mit  $Y$  keine ausreichenden Werte. Insgesamt betrachtet liegen jedoch für mindestens die Hälfte der Messzeitpunkte, zu denen diese Inhaltskategorien durch die Beurteilerinnen ausgewählt wurden, mäßige bis sehr gute Übereinstimmungen in  $\kappa$  und  $Y$  vor. Die Inhaltskategorien werden daher beibehalten mit Ausnahme der Messzeitpunkte, zu denen sowohl  $\kappa$  als auch  $Y$  unter  $.40$  liegen (in Tabelle 130 in Anhang B grau hinterlegt).

Zur Übersicht befindet sich in Tabelle 10 die endgültige Auswahl an verwendeten Kategorien. Die Oberkategorien sowie die Inhaltskategorien, die keine weiteren Inhalte subsumieren, können durchgehend in die Analysen einbezogen werden.

#### 11.5.4.2 Lehrkrafturteil leseauffällig

Um die Fragestellungen der Untersuchung beantworten zu können, wurde aus den einzelnen für den Leseerwerb relevanten und mit mindestens mäßiger Beurteilerübereinstimmung vorliegenden Inhaltskategorien die Oberkategorie *Lehrkrafturteil leseauffällig* erstellt. Kinder sind dann leseauffällig geworden, wenn Auffälligkeiten, die sich auf die Fähigkeit des Lesens beziehen, genannt wurden. Dies beinhaltet die Inhaltskategorien *Lesen allgemein*, *Leseverständnis* sowie *Lesegeschwindigkeit/-flüssigkeit*. Da die Antworten der Lehrkräfte ohne Nachfragen teilweise recht allgemein gehalten waren, werden ebenfalls Aussagen berücksichtigt, die sich allgemeiner auf den Schriftspracherwerb beziehen: *Lese-Rechtschreibstörung*, *Deutsch allgemein*, *Laut-Buchstaben-Zuordnung* sowie *Buchstabenkenntnis*. Kinder, auf die hiervon nichts zutrifft, werden als nicht leseauffällig geworden eingestuft. Dies trifft auch auf solche Kinder zu, zu denen nur Aussagen explizit zur Rechtschreibung getroffen wurden. Die Angaben und auch die hieraus resultierenden Beurteilerübereinstimmungen finden sich in Tabelle 11.

Tabelle 11

*Zusammensetzung und Beurteilerübereinstimmungen für das Lehrerurteil „leseauffällig“*

	T1 n = 70	T2 n = 87	T3 n = 67	T4 n = 76	T5 n = 65	T6 n = 96	T7 n = 23
Lehrkrafturteil <i>leseauffällig</i>	κ	.77**	.83**	.82**	.78**	.72**	.88**
	PÜ	93	93	94	94	91	96
	Y	.83	.87	1.00	1.00	.85	1.00

*Anmerkungen.* n = Zahl der Kinder zu jedem Messzeitpunkt, zu denen Aussagen getätigt wurden. Leere Zellen bedeuten, keine Beurteilerin hat diese Kategorie gewählt. κ = Cohens Kappa; PÜ = Prozentuale Übereinstimmung; Y = Yules Y. \*p < .05, \*\*p < .01

Weder Beurteilerin S noch Beurteilerin K verwenden zu T1 eine der Kategorien für das Lehrkrafturteil *leseauffällig*. Es liegen mit Werten über κ = .70 durchgehend sehr gute Beurteilerübereinstimmungen vor. Y ist ebenfalls durchgehend im sehr guten Bereich.

#### 11.5.4.3 Beurteilerübereinstimmung zu Frage 2: „Was wird getan?“

Tabelle 12 zeigt die Beurteilerübereinstimmungen für die Antworten zur zweiten Frage, die mindestens mäßige Beurteilerübereinstimmungen erreicht haben und somit in die endgültige Analyse eingehen. Eine Gesamtaufzählung inklusive der Inhaltskategorien, die aufgrund niedriger Beurteilerübereinstimmung aus weiteren Analysen entfernt wurden, findet sich in Tabelle 131 in Anhang B.

Für fast alle Variablen liegen mehrstufige Antwortformate vor, sodass für diese nur auf Cohens Kappa und die prozentuale Übereinstimmung zurückgegriffen werden kann. Für die dichotomen Kategorien wird ebenfalls Yules Y berichtet.

Tabelle 12

*Beurteilerübereinstimmung für Kategorien mit mindestens mäßiger Übereinstimmung*

		T1 n = 21	T2 n = 45	T3 n = 39	T4 n = 45	T5 n = 57	T6 n = 68	T7 n = 21
<b>Organisatorische innerschulische Aspekte</b>								
Förderung / Unterstützung in der Klasse	K	.83**	.62**	.53**	.59**	.61**	.53**	.57**
	PÜ	90	78	72	76	83	78	71
Lehrkraft fördert selbst	K	1.00**	.85**	.71**	.55**	.51**	.50**	.78**
	PÜ	100	91	85	73	84	79	90
Förderunterricht	K		.76**	.71**	1.00**	.83**	.92**	.73**
	PÜ		93	82	100	91	96	86
<b>Organisatorische Aspekte der Zusammenarbeit</b>								
Elterngespräche <sup>+</sup>	K		.73**	.77**	.90**	.79**	.81**	
	PÜ		93	95	98	98	94	
	Y		1.00	.84	1.00	1.00	.87	
Außerschulische Hilfe	K	.76**	.73**	.77**	.79**	.55**	.57**	.79**
	PÜ	86	91	92	98	91	84	90
BFZ ist eingeschaltet <sup>+</sup>	K	1.00**	.81**	.84**	.79**	.83**	.78**	.83**
	PÜ	100	96	97	98	91	90	95
	Y	1.00	1.00	1.00	1.00	.83	.85	1.00
<b>Inhaltliche Aspekte</b>								
Differenzierung	K	1.00**	1.00**	1.00**	.66**	.68**	.74**	.78**
	PÜ	100	100	100	98	95	97	86
Lob, positive Verstärkung	K		.65**		.65**	.74**		
	PÜ		96		96	98		
Förderplan	K		.82**		1.00**	1.00**	1.00**	.73**
	PÜ		98		100	100	100	86
Spezifische Fördermethode genannt	K	1.00**	.91**		1.00**	.50**	.50**	.64**
	PÜ	100	98		100	98	99	91
<b>Sonstiges</b>								
Klassenwiederholung <sup>+</sup>	K		1.00**	.88**	1.00**	1.00**	1.00**	
	PÜ		100	98	100	100	100	
	Y		1.00	1.00	1.00	1.00	1.00	
ASF ist festgestellt <sup>+</sup>	K		.88**	1.00**	.66**	.73**	1.00**	1.00**
	PÜ		98	100	98	97	100	100
	Y		1.00	1.00	1.00	1.00	1.00	1.00
Sonstige professionelle Förderung	K	.68**	.74**		1.00**	.55**	.40**	.52**
	PÜ	81	93		100	91	79	81

*Anmerkungen.* n = Zahl der Kinder zu jedem Messzeitpunkt, zu denen Aussagen getätigt wurden. Leere Zellen bedeuten, keine Beurteilerin hat diese Kategorie gewählt. κ = Cohens Kappa; PÜ = Prozentuale Übereinstimmung; Y = Yules Y. \*p < .05, \*\*p < .01. <sup>+</sup>dichotome Variable. Grau hinterlegt ist der Zeitpunkt, zu dem die Kategorie aufgrund unzureichender Übereinstimmung nicht in die Analysen eingeht.

Im Gegensatz zu genannten Auffälligkeiten finden sich mehrere Inhalte, die zu fast keinem oder keinem Messzeitpunkt eine ausreichende Beurteilerübereinstimmung von mindestens κ = .40 aufweisen (in Tabelle 131 in Anhang B grau hervorgehoben). Dies ist vor allem im Bereich der organisatorischen innerschulischen Aspekte der Fall. Ob eine BFZ-Kraft oder eine sonstige Person fördert, ob es sich um Einzel- oder auch um Gruppenförderung handelt und auch ob die Antwort in die Restkategorie für sonstige innerschulische Förderung gehört, kann zu fast keinem Zeitpunkt mit akzeptabler Beurteilerübereinstimmung festgestellt werden

(s. Tabelle 131 in Anhang B). Diese Kategorien gehen entsprechend in keine weiteren Berechnungen ein. Ob dagegen Förderung im Klassenunterricht stattfindet, die Lehrkraft selbst in der Förderung tätig wird und ob das Kind am Förderunterricht teilnimmt, wird durchgehend mit zufriedenstellenden bis sehr guten Beurteilerübereinstimmungen entschieden. Die organisatorischen Aspekte der Zusammenarbeit mit den Kategorien Elterngespräche, außerschulische Hilfe sowie dem Hinzuziehen des BFZs erreichen ebenfalls zufriedenstellende bis sehr gute Übereinstimmungen.

Im Bereich der inhaltlichen Aspekte können die Kategorien Differenzierung, Lob / positive Verstärkung sowie die Nennung eines Förderplans mit guter bis sehr guter Beurteilerübereinstimmung bestimmt werden. Ob eine spezifische Fördermethode genannt wird, kann mit zufriedenstellender bis sehr guter Beurteilerübereinstimmung festgestellt werden.

Für die Kategorie Motivationsförderung sowie für die Restkategorie sonstige Inhalte liegen nur zu jeweils einem Messzeitpunkt eine gute bzw. zufriedenstellende Beurteilerübereinstimmung vor. Entsprechend gehen diese beiden Kategorien in keine weiteren Berechnungen ein. Ob die Antwort der Lehrkraft eine Aussage über eine Klassenwiederholung des Kindes enthält, kann durchgehend mit sehr guter Beurteilerübereinstimmung bestimmt werden. Für den Anspruch auf sonderpädagogische Förderung liegen gute bis sehr gute Beurteilerübereinstimmungen vor. Für die Restkategorie sonstige professionelle Förderung liegt zu T3 eine nicht akzeptable Beurteilerübereinstimmung vor (s. Tabelle 131 in Anhang B). Die Kategorie geht für diesen Messzeitpunkt entsprechend nicht in die Analyse ein. Für die übrigen Messzeitpunkte liegen mäßige bis sehr gute Beurteilerübereinstimmungen vor.

#### 11.5.4.4 Fazit zur Beurteilerübereinstimmung

Um die Antworten der Lehrkräfte den vorgegebenen Kategorien zuzuordnen, müssen subjektive Interpretationen vorgenommen werden. Verschiedene Beurteiler sollten unabhängig voneinander zu den gleichen Zuordnungen kommen. Wie die Beurteilerübereinstimmungen zeigen, ist dies in vielen Fällen gut bis sehr gut, in anderen Fällen dagegen nicht gelungen. In einem Gespräch, welches mit jeder Beurteilerin einzeln durchgeführt wurde, gaben beide an, dass einige Aussagen mehrdeutig und schwer zu interpretieren gewesen seien. Zudem erwies sich das Kategoriensystem zu „Was wird getan?“ stellenweise als sehr komplex, da die Restkategorien mit sonstigen Antworten nur dann gewählt werden sollten, wenn keine der spezifischeren Kategorien besser zutrafen. Beurteilerin S gab explizit an, dass die Einschätzung, wer die Förderung durchführt und wo diese stattgefunden hat, sehr schwierig gewesen sei. Dies spiegelt sich in der inakzeptablen Beurteilerübereinstimmung der Kategorien zur Einzel- oder Gruppenförderung wider. Es lässt sich scheinbar anhand der getätigten Aussagen nicht zuverlässig bestimmen, ob im Einzel- oder Gruppensetting gefördert wird. Der Aspekt, wer die Förderung durchführt, ist differenzierter zu betrachten. Es kann mit zufriedenstellender bis sehr guter Beurteilerübereinstimmung bestimmt werden, ob die Grundschullehrkraft selbst involviert ist. Ob eine BFZ-Kraft oder ob eine sonstige Person fördert, erreicht

dagegen jeweils keine ausreichend hohe Beurteilerübereinstimmung. Vor allem für die Frage, ob eine sonstige Person fördert bzw. ob es unklar ist, wer fördert, liegen die Beurteilerübereinstimmungen zu sechs der sieben Messzeitpunkte nicht signifikant oberhalb des Bereichs, in der eine zufällige Übereinstimmung zu finden ist. Die Kategorien, die nicht mindestens in der Hälfte der Messzeitpunkte ausreichende Beurteilerübereinstimmungen zeigen, gehen entsprechend nicht in die Analysen ein.

Nach Berechnung der Beurteilerübereinstimmungen wurde zufällig gewählt, wessen Beurteilungen für die Auswertung herangezogen werden. Das Los fiel auf Beurteilerin S.

### 11.6 Allgemeine Aspekte zu den Auswertungsmethoden

Die Auswertung erfolgte mittels IBM SPSS Statistics 25.0 (IBM Corp, 2017) sowie dem Statistikprogramm R in der Version 3.6.1 (R Core Team, 2019) und RStudio in der Version 1.2.1335 (RStudio Team, 2019). Die jeweils für R verwendeten Statistikpakete werden an den entsprechenden Stellen gesondert angegeben.

Alle inferenzstatistischen Überprüfungen der Signifikanz erfolgten auf dem zuvor festgelegten Alpha-Niveau von .05, was den Fehler 1. Art auf 5 Prozent festlegt. Sofern mehrere abhängige Variablen zur Beantwortung der Hypothese betrachtet wurden, wurde zur Interpretation der Ergebnisse von multiplen statistischen Tests die Benjamini-Hochberg-Korrektur (Benjamini & Hochberg, 1995) angewendet, welche die False Discovery Rate bei multiplen Hypothesentests kontrolliert (Benjamini & Hochberg, 1995, S. 298; Victor, Elsässer, Hommel & Blettner, 2010, S. 53).

Da zur Auswertung Vergleiche mit Teilstichproben durchgeführt wurden und aufgrund des Längsschnittdesigns auch Ausfälle zu verzeichnen waren, mussten stellenweise Gruppenvergleiche mit einer recht geringen Stichprobengröße durchgeführt werden. Bei extrem kleinen Stichproben ( $n \leq 5$ ) liegt die Wahrscheinlichkeit, einen Effekt in der Stichprobe statistisch nachzuweisen, wenn dieser auch wirklich existiert, bei fünf Prozent (de Winter, 2013, S. 8). De Winter (2013, S. 8) hält dennoch fest, dass  $t$ -Tests selbst mit solch extrem kleinen Stichproben durchgeführt werden können, wenn eine große Effektstärke erwartet wird. Vom Mann-Whitney-U-Test oder dem Welch-Test als Alternative rät er dagegen ab.

Für eine bessere Einschätzung der Ergebnisse wird zusätzlich zu den statistischen Kennwerten der Tests auch die Effektstärke mit ihrem 95 %-Konfidenzintervall angegeben. Die Berechnung der Effektstärken wird mit R und dem Paket *effsize* (Torchiano, 2016) durchgeführt. Die Interpretation der Effektstärken orientiert sich dabei an Cohen (1992). Für  $t$ - und  $F$ -Tests bei unabhängigen Stichproben bedeutet dies bei  $|d| < 0.2$  einen vernachlässigbaren Effekt, bei  $|d| < 0.5$  einen kleinen Effekt, bei  $|d| < 0.8$  einen mittleren Effekt und bei allen Werten darüber einen großen Effekt (Cohen, 1992, S. 157; Torchiano, 2019, S. 6).

## 11.6.1 Bestimmung der Leseschwäche

Da der Lesetest wiederholt vorgelegt wurde, ist mit einem Lerneffekt bei den teilnehmenden Kindern zu rechnen, sodass die Normierung aus dem ELFE 1-6-Manual nicht verwendet werden kann. Die Ergebnisse werden deshalb anhand der vorliegenden Stichprobe normiert. Analog zum Vorgehen der Autoren des Tests wird auch für diese Stichprobe zunächst betrachtet, ob die erzielten Ergebnisse der Untertests der Normalverteilung folgen. Im Anschluss erfolgt eine Flächentransformation nach Blom:  $\frac{\text{Rang} - \frac{3}{8}}{\text{Anzahl gültiger Werte} + \frac{1}{4}}$  (Lenhard & Schneider, 2006, S. 52).

Als Signifikanztest wird der Kolmogorov-Smirnov-Test verwendet, da dieser für Stichproben  $> 50$  empfohlen wird (Brosius, 2018, S. 480). Brosius (2018, S. 480) weist allerdings auch darauf hin, dass der Test auf *perfekte* Normalverteilung testet und daher auch dann zu einem signifikanten Ergebnis kommt, wenn die Normalverteilung nur leicht verletzt ist.

Tabelle 13 zeigt die Ergebnisse und die Schiefe der jeweiligen Verteilungen. Mit einer Ausnahme sind die Untertests zu allen Messzeitpunkten nicht normalverteilt. Tendenziell zeigt sich, dass die Kinder zu den früheren Messzeitpunkten häufiger niedrigere Werte im Test erreichen und zum letzten Messzeitpunkt Ende der vierten Klasse eher höhere Werte erreichen. Aufgrund dieser Ergebnisse wird die Flächentransformation nach Blom durchgeführt.

Tabelle 13

*Kolmogorov-Smirnov-Test auf Normalverteilung und Schiefe für die einzelnen Subtests*

	Kolmogorov-Smirnov			Schiefe
	Statistik	df	p	
T2 Wortlesen	0.152	248	.00**	2.731
T3 Wortlesen	0.075	252	.00**	1.123
T4 Wortlesen	0.091	240	.00**	0.997
T6 Wortlesen	0.084	238	.00**	0.627
T7 Wortlesen	0.051	170	.20	-0.224
T4 Satzlesen	0.107	240	.00**	0.520
T6 Satzlesen	0.070	238	.01*	0.109
T7 Satzlesen	0.169	170	.00**	-1.153
T6 Textlesen	0.115	238	.00**	0.444
T7 Textlesen	0.192	169	.00**	-1.349

Anmerkungen. \* $p < .05$ , \*\* $p < .01$

Um eine Vergleichbarkeit zwischen den Messzeitpunkten und Untertests zu vereinfachen, wird die Transformation für alle Subtests durchgeführt. Anschließend werden die z-Werte in die anschaulichere T-Wert-Skala überführt (Schmidt-Atzert & Amelang, 2012, S. 165–166). Während die z-Wert-Skala einen Mittelwert von 0 und eine Standardabweichung von 1 aufweist, hat die T-Wert-Skala einen Mittelwert von 50 und eine Standardabweichung von 10.

## 11.6.2 Geschachtelte Datenstruktur

Die Datenerhebungen fanden im Klassenverband statt. Aus diesem Grund muss die geschachtelte Struktur berücksichtigt werden, wobei die Schülerinnen und Schüler der Level-1-Einheit und die Schulklassen der Level-2-Einheit, den Clustern, zuzuordnen sind (Geiser, 2011, S. 199). Mittlerweile ist das Mittel der Wahl für die Auswertung die Mehrebenenanalyse. Diese setzt allerdings sehr große Stichproben voraus, beispielsweise je 30 Kinder in je 30 Schulklassen (Walter & Rack, 2007, S. 283). Aufgrund der zugrundeliegenden Fragestellung und dem Fokus auf die Kinder, die in einer Klasse leseschwach sind bzw. leseauffällig werden, sind solche Zahlen nicht zu erreichen gewesen, sodass diese Art der Datenanalyse nicht in Frage kommt.

Ob die geschachtelte Struktur stets berücksichtigt werden muss, wird teilweise in Frage gestellt (Schoppek, 2015, S. 201). Um darüber entscheiden zu können, soll die Intraklassenkorrelation (ICC) berechnet werden. Diese „ist definiert als das Verhältnis der Varianz zwischen den Clustern zur Gesamtvarianz“ (Geiser, 2011, S. 204) und berechnet sich wie folgt:

$$\rho_{IC} = \frac{\sigma^2_B}{(\sigma^2_W + \sigma^2_B)}$$

Die Intraklassenkorrelation gibt demnach den Anteil an der Gesamtvarianz in der abhängigen Variablen an, beispielsweise der Leseleistung zum Messzeitpunkt T7, der durch die Kontextzugehörigkeit, in dieser Studie folglich der Klassenzugehörigkeit, erklärt wird. Wenn dieser Anteil nicht mehr als „trivial“ angesehen werden kann, muss die Klasse als Cluster berücksichtigt werden. Lee (2000, S. 128) setzt diesen Wert bei  $ICC \geq .10$  an, wenn folglich 10 Prozent der Gesamtvarianz durch die Kontextzugehörigkeit erklärt werden. Schoppek (2015, S. 201) führt zudem den Designeffekt  $e_d$  an, der sich aus der ICC und der durchschnittlichen Gruppengröße berechnet:  $e_d = 1 + (Gruppengröße - 1) * ICC$ . Wenn der Designeffekt geringer als 2.0 ausfällt, kann nach ihm die geschachtelte Struktur in der Analyse ignoriert werden.

Tabelle 14

*Höhe der Intraklassenkorrelationen und des Designeffekts für die Leseleistung zu T6 und T7*

Abhängige Variable	Durchschnittliche Gruppengröße	ICC	$e_d$
Leseleistung zu T6	18.31	.170	3.94
Leseleistung zu T7	15.36	.166	3.38

*Anmerkungen.* ICC = Intraklassenkorrelation.  $e_d$  = design effect.

Tabelle 14 zeigt die Höhe der Intraklassenkorrelationen und der Designeffekte. Die durchschnittliche Gruppengröße konnte nur mit den an der Studie teilnehmenden Kindern jeder Klasse berechnet werden, sodass die Gruppengröße in Wirklichkeit noch höher ausfällt. Der Designeffekt wird dementsprechend eher unterschätzt. Für beide Messzeitpunkte liegen die Werte oberhalb der Grenzen, sodass die Klassenstruktur in der Auswertung entsprechend berücksichtigt werden muss.

### 11.6.3 Umgang mit fehlenden Daten

Bei Langzeitstudien, insbesondere bei Felduntersuchungen wie der vorliegenden, kommt es designbedingt zu Ausfällen und somit zu fehlenden Daten. Teilweise fehlten Kinder zu bestimmten Zeitpunkten etwa durch Umzug, Krankheit, Wiederholen oder Überspringen einer Klasse etc. Dieser Wegfall wird auch als statistische Mortalität bezeichnet (Rost, 2013, S. 150). Aus organisatorischen Gründen und da die Zeitspanne der Erhebung nicht unbegrenzt ausgeweitet werden konnte, war es nicht möglich, die Daten für erkrankte Kinder nachzuerheben. Fehlende Daten dürfen allerdings nicht einfach ignoriert werden, da diese zu verzerrten Parameterschätzungen führen können (Lüdtke, Robitzsch, Trautwein & Köller, 2007, S. 104). Zunächst wird kurz auf die Art an fehlenden Daten eingegangen, bevor die Vorgehensweise für die vorliegende Untersuchung beschrieben wird.

Gemeinhin ist eine Unterteilung in die folgenden drei Arten an fehlenden Daten üblich, welche auf Rubin (1976) zurückzuführen ist:

1. MCAR – Missing completely at Random
2. MAR – Missing at Random
3. MNAR – Missing not at Random

Wenn MCAR vorliegt, dann stehen die fehlenden Daten in keinerlei Zusammenhang mit den beobachteten Daten (Sauer, 2019, S. 118). Die Probanden, bei denen fehlende Daten auftreten, stellen entsprechend eine Zufallsstichprobe dar (Lüdtke et al., 2007, S. 104). Wenn dagegen MAR vorliegt, kann jeder systematische Unterschied zwischen den fehlenden Daten und den beobachteten Daten durch Unterschiede in den beobachteten Daten erklärt werden. Sauer (2019, S. 118) nennt als Beispiel, wenn Männer seltener ihr Gewicht angeben. In diesem Fall hängt die Wahrscheinlichkeit, dass in der Variable „Gewicht“ Werte fehlen, mit der Variable „Geschlecht“ zusammen. Die Ausprägung des Wertes ist allerdings zufallsbedingt, da die Wahrscheinlichkeit des fehlenden Wertes nicht mit seiner Ausprägung zusammenhängt. Dagegen liegt MNAR vor, wenn die Wahrscheinlichkeit, dass Daten in einer Variable fehlen, durch den Wert in dieser Variablen bedingt wird. Dies liegt zum Beispiel dann vor, wenn die Wahrscheinlichkeit steigt, dass das Gewicht nicht angegeben wird, je mehr eine Person wiegt (Sauer, 2019, S. 118). In dieser Untersuchung ist das beispielsweise der Fall, wenn Kinder mit niedrigeren schulischen Leistungen zu den späteren Messzeitpunkten fehlen, weil sie die Klasse wiederholen.

Die Schwierigkeit bei der Bestimmung der Art an fehlenden Daten besteht darin, dass man häufig den Grund nicht kennt. Ein Test, der auf MAR oder MNAR prüft, würde die Kenntnis über die fehlenden Werte voraussetzen (Lüdtke et al., 2007, S. 105). Auf das Vorliegen von MCAR kann man dagegen testen (Enders, 2010, S. 17). Dies geschieht beispielsweise über Little's MCAR-Test, der die Annahme prüft, ob MCAR gegenüber MAR haltbar ist (Little, 1988). Er kann aber nicht ausschließen, dass die fehlenden Daten von der Ausprägung in eben dieser Variablen abhängen, also ob MNAR vorliegt. Enders (2010, S. 17) findet diesen Test für einen Datensatz nicht sinnvoll. Zum einen wird es in diesem Fall seiner Meinung

nach recht sicher einzelne Variablen geben, die systematisch fehlende Werte aufweisen, so dass MCAR zurückgewiesen werden muss. Zum anderen ändert aber auch das Vorliegen von MCAR nichts an der vorherrschenden Empfehlung, fehlende Daten über multiple Imputation zu ersetzen, da dieses Vorgehen sowohl für MAR wie auch für MCAR geeignet ist (Enders, 2010, S. 13). Bevor auf die multiple Imputation eingegangen wird, sollen noch kurz andere in der Vergangenheit verwendete oder vorgeschlagene Verfahren sowie deren Vor- und Nachteile dargestellt werden.

Bis in die 2000er Jahre waren vor allem der fallweise und der paarweise Ausschluss geläufig (Enders, 2017, S. 4). Beim fallweisen Ausschluss werden die Fälle, die fehlende Daten aufweisen, komplett aus der Analyse entfernt. Es kann allerdings zu stark verzerrten Parameterschätzungen kommen, sobald die fehlenden Daten nicht nur MCAR sind. Zudem kann sich die Stichprobengröße stark reduzieren (Lüdtke et al., 2007, S. 107). Beim paarweisen Ausschluss werden nur die Personen aus der Analyse herausgenommen, bei denen Variablen mit fehlenden Werten in die Auswertung einbezogen werden würden, was laut Rost bei  $N > 200$  unproblematisch sein sollte (Rost, 2013, S. 195).

Eine weitere Methode ist die einfache Imputation von fehlenden Daten. Bei dieser wird der fehlende Wert geschätzt und die Schätzung in der Folge für Analysen verwendet. Die Schätzung selbst kann z. B. darüber erfolgen, dass der Mittelwert der jeweiligen Skala eingefügt wird. Dies führt allerdings zu einer Reduzierung in der Varianz der Daten, weswegen von der Verwendung solcher Methoden abgeraten wird (Enders, 2010, S. 42; Lüdtke et al., 2007, S. 108).

Dagegen werden sowohl Maximum Likelihood wie auch multiple Imputation als Verfahren bei fehlenden Daten empfohlen. Dabei ist der multiplen Imputation dann Vorzug zu geben, wenn die Daten sowohl kategoriale wie auch kontinuierliche Variablen enthalten, was bei pädagogisch-psychologischen Studien häufig der Fall ist (Enders, 2017, S. 4–5; Lüdtke et al., 2007, S. 106). Da dies auch für die vorliegende Untersuchung der Fall ist, wird sich im Folgenden nur auf die multiple Imputation bezogen. Das Verfahren geht auf Rubin (1987) zurück. Es werden mehrere Datensätze erzeugt, von denen jeder unterschiedliche Schätzungen für die fehlenden Daten enthält (Enders, 2017, S. 5). Die Daten werden über Regressionsgleichungen geschätzt, wobei die vollständigen Variablen als Prädiktoren und unvollständige als Outcome-Variablen verwendet werden. Dabei kommen iterative Algorithmen zum Einsatz, so dass die Parameter des Regressionsmodells angepasst werden. Die neuen Schätzungen werden dann für den nächsten Datensatz verwendet, bis die zuvor gewählte Anzahl an  $m$  Imputationen erstellt wurde.

Der chained equations-Ansatz führt eine Reihe an univariaten Regressionsanalysen statt einer einzigen multivariaten Regressionsanalyse durch, um Imputationen zu erzeugen (Enders, 2017, S. 8). Dieser Ansatz ist vor allem bei einer Mischung aus kategorialen und kontinuierlichen Variablen geeignet (Enders, 2017, S. 9). Die verwendete Regression ist dabei abhängig vom Skalenniveau (Urban, Mayerl & Wahl, 2016, S. 50–51). Für metrische Variablen wird

ein lineares Regressionsmodell angewendet, für kategoriale und ordinale dagegen logistische Regressionen.

Oft findet man in der Literatur den Vorschlag von  $m = 3$  bis 5 Imputationen (Graham, Olchowski & Gilreath, 2007, S. 206; Lüdtke et al., 2007, S. 113). Dies wird jedoch auch immer wieder kritisiert, weil die Menge an Imputationen und die Höhe des Powerverlusts aufgrund einer zu geringen Zahl an Imputationen von dem Anteil fehlender Daten abhängig ist. Mitra und Reiter (2016, S. 197) empfehlen  $m = 20$  Imputationen, da ihrer Meinung nach eine weitere Erhöhung keinen nennenswerten Vorteil mehr bringt. Graham, Olchowski und Gilreath (2007, S. 210) führen dagegen an, dass die Rate an Fällen, in denen fehlende Daten vorliegen, stets berücksichtigt werden muss. Wenn bei 30 Prozent aller Fälle fehlende Daten auftreten, dann ist der Verlust an Power zwischen  $m = 100$  und  $m = 20$  gering (von .79 zu .78). Von  $m = 20$  zu  $m = 5$  verliert man dagegen deutlich mehr Power (von .78 zu .73). Das gleiche Bild ergibt sich auch, wenn bei 50 Prozent der Fälle Daten fehlen. Hier sinkt die Power zunächst von .78 auf .77 (für  $m = 100$  und  $m = 20$ ) und dann von .77 ( $m = 20$ ) zu .68 ( $m = 5$ ). Bei 10 Prozent fehlenden Daten gibt es keinen Abfall an Power zwischen  $m = 20$  und  $m = 5$ . Es ist also davon abhängig, bei wie vielen Fällen fehlende Daten vorkommen. Bei einer höheren Rate als .50 sollte man mindestens  $m = 40$  Imputationen vornehmen. Bei einer Rate von 30 Prozent liegt der Powerverlust bei unter einem Prozent, wenn 20 Imputationen durchgeführt werden (Graham et al., 2007, S. 212).

Nicht imputieren sollte man dagegen, wenn MNAR vorliegt, da statistische Methoden in diesem Fall nicht zufriedenstellend sind (Rost, 2013, S. 198). Lüdtke et al. (2007, S. 116) geben an, dass es für diesen Fall zwar statistische Vorgehensweisen gibt, dass hierbei allerdings der Ausfallprozess explizit modelliert werden muss und dies nicht ohne weiteres möglich ist. Auch Allison (2000, S. 302) rät davon ab, bei Vorliegen von MNAR Daten zu imputieren, da das Erstellen und Schätzen von Modellen „komplex, nicht testbar“ ist und spezialisierte Software benötigt wird. Göthlich (2009, S. 129) kommt zu dem Schluss, dass das Problem des Datenausfalls durch MNAR „von keinem Verfahren geheilt werden“ kann.

Nach der Datenimputation können Standardverfahren zur Auswertung verwendet werden (Peugh & Enders, 2004, S. 550). Eine einfache Mittelwertbildung einer Variablen über die  $m$  Imputationen ist dabei nicht möglich. Vielmehr geschieht dies nach Rubins Regeln (Rubin, 1987, S. 76). Es wird mit den  $m$  Datensätzen der jeweilige statistische Test durchgeführt. Die daraus resultierenden  $m$  Parameterschätzungen und Standardfehler werden „gepooled“, also entsprechend gemittelt. Aus diesem Ergebnis kann die Signifikanz oder das Fehlen dieser berechnet werden (Peugh & Enders, 2004, S. 535).

Da die vorliegende Untersuchung sowohl kategoriale wie auch kontinuierliche Variablen mit fehlenden Werten enthält, wird das Verfahren der multiplen Imputation durch chained equations (MICE) durchgeführt. An den entsprechenden Stellen in der Auswertung wird der Anteil fehlender Daten zunächst berichtet, bevor daran die Anzahl an Imputationen festgelegt wird.

Es ist weiterhin damit zu rechnen, dass vor allem leistungsschwache bzw. sehr leistungsstarke Schülerinnen und Schüler zu späteren Messzeitpunkten fehlen, weil diese die Klasse wiederholen bzw. übersprungen haben. Diese Art an fehlenden Daten lässt sich in einer wie der vorliegenden Felduntersuchung nur schwer vermeiden. Da die schulischen Leistungen zum Ende der Grundschulzeit daher mit hoher Wahrscheinlichkeit auch MNAR sind, werden diese nicht imputiert.

Zu den frühen Messzeitpunkten muss berücksichtigt werden, dass das Wortlesen als Indikator für die Korrektheit der Einschätzung der Lehrkraft verwendet wird. Da nicht erhoben wurde, warum und wie lange die Kinder fehlen, wird die Leistung im Wortlesen auch zu früheren Messzeitpunkten nicht imputiert. Es kann nicht eingeschätzt werden, warum eine Lehrkraft zu einem fehlenden Kind keine Aussage über dessen schulische Leistungen tätigt – beispielweise, ob die Lehrkraft bewusst keine Einschätzung vornehmen möchte oder kann, weil das Kind bereits seit längerem fehlt, oder ob sie das Kind vielleicht auch nicht nennt, weil es an der Testung zu dem entsprechenden Messzeitpunkt nicht teilgenommen hat. Eine Imputation der Leseleistungen könnte zu gravierenden Verzerrungen in Bezug auf das Lehrkräfteurteil führen. Aus diesem Grund werden nur die Daten von den Kindern verwendet, die zum betrachteten Messzeitpunkt in der Klasse waren.

Auf den fallweisen Ausschluss wird zugunsten des paarweisen Ausschlusses verzichtet. Krombholz (2019, S. 25) schlägt bei Langzeitstudien ein „Aufbrechen“ des Datensatzes nach Erhebungszeitpunkten vor, um so Daten mit weniger fehlenden Werten zu erhalten. Da der Fokus nicht nur auf den schulischen Leistungen der Kinder liegt, sondern begleitend auf den Aussagen der Lehrkräfte, erscheint dies als das sinnvollste Vorgehen. Die jeweilige Stichprobengröße wird entsprechend genannt werden.

#### 11.6.4 Statistisches Matching zur Bestimmung der Kontrollgruppen

Da es sich bei dieser Untersuchung um eine Beobachtungsstudie im Feld handelt, ist es nicht möglich, Kinder den zu untersuchenden Merkmalen (z. B. Leseauffälligkeit durch die Lehrkraft) randomisiert zuzuweisen. Aufgrund dieser fehlenden randomisierten Zuteilung ist bei einem einfachen Vergleich der nicht auffälligen mit auffälligen Kindern nicht auszuschließen, dass eine unterschiedliche Leseleistung zu T6 oder T7 auf andere Faktoren als die Auffälligkeit zurückzuführen ist, da Einflussfaktoren nicht kontrolliert werden können. Dies wird als Selection Bias bzw. Stichprobenverzerrung bezeichnet (Müller, 2012, S. 4). Ebenso widerspricht dieses Vorgehen der „Conditional Independence Assumption“, nach der die Unterschiede in den Outcome-Variablen zwischen der Treatment- und der Kontrollgruppe unabhängig sein müssen von Selektionsprozessen. Wenn diese Annahme verletzt ist, kann ein Unterschied in der Outcome-Variable nicht auf das Treatment zurückgeführt werden (Müller, 2012, S. 5).

Um dennoch Gruppenvergleiche durchführen zu können, müssen Einflussfaktoren kontrolliert werden. Ziel ist es dabei, eine möglichst optimale Parallelisierung der Gruppen zu erreichen.

In Beobachtungsstudien wird daher häufig auf statistisches Matching zurückgegriffen, um geeignete Kontrollgruppen zu bilden (Bacher, 2002; Mansournia, Jewell & Greenland, 2018; Stuart, 2010). Eine Randomisierung kann hierdurch nicht ersetzt werden, da beim statistischen Matching nur bekannte Einflussfaktoren berücksichtigt werden können. Für nicht-randomisierte Studien ist dieses Vorgehen dennoch eine sehr gute Alternative. Es muss bei der Interpretation der Ergebnisse allerdings berücksichtigt werden, dass unbekannte Störfaktoren nicht adjustiert werden können (Kuss, Blettner & Börgermann, 2016, S. 602).

Um eine passende Kontrollgruppe zu ermitteln, gibt es mehrere Verfahren, die angewendet werden können. Häufig wird das statistische Matching mit dem Propensity Score-Verfahren vorgeschlagen (Bacher, 2002, S. 47), welches auf Rosenbaum und Rubin (1983) zurückgeht. Mit dem Begriff Propensity Score bezeichnen die Autoren die bedingte Wahrscheinlichkeit, bei Vorhandensein einer Reihe von Einflussfaktoren einem Treatment zugeordnet zu werden (Rosenbaum & Rubin, 1983, S. 41).

Nach Bacher (2002, S. 47) müssen zunächst geeignete Variablen ausgewählt werden, mit denen der Propensity Score ermittelt wird. Diese beeinflussen sowohl die Treatmentzugehörigkeit (z. B., ob das Kind leseauffällig geworden ist oder nicht) als auch die interessierende Outcome-Variable (z. B. die Lesefähigkeit zu den späteren Messzeitpunkten). Sie dürfen allerdings nicht vom Treatment abhängig sein (Caliendo & Kopeinig, 2008, S. 38). Es sollten weiterhin nicht zu wenige Variablen gewählt werden (Bacher, 2002, S. 47). Die perfekte Wahl an Einflussvariablen gibt es hierbei nicht, da Personen mit bestimmten Ausprägungen dann entsprechend stets entweder auffällig oder nicht auffällig sind. Daher raten Caliendo und Kopeinig (2008, S. 38) dazu, das Modell nicht übermäßig zu spezifizieren, vor allem bei kleineren Stichproben.

Sogenannte „Covariates of convenience“ sind allerdings alleine nicht ausreichend (Thoemmes, 2012, S. 4). Damit bezeichnet Thoemmes Kovariaten wie Geschlecht, Alter oder den sozioökonomischen Status. Diese sind zweifelsohne relevant und sollten mit in das Modell aufgenommen werden. Zusätzlich müssen aber noch spezifischere Variablen in die Berechnung des Propensity Scores einfließen. Für die vorliegende Untersuchung sind das beispielsweise die im Theorieteil herausgearbeitete, für das Lesen relevante Buchstabenkenntnis sowie die Ausgangsleseleistung. Neben diesen Einflussfaktoren soll auch die Klassenzugehörigkeit als Kovariable aufgenommen werden, da über diese Methode die geschachtelte Datenstruktur berücksichtigt werden kann (Thoemmes & West, 2011). Nur bei sehr großen Stichproben ist es sinnvoll, innerhalb der Cluster, also der Klassen, zu matchen, da bei kleineren Stichproben die Gefahr besteht, keine passenden Personen für die Vergleichsgruppe innerhalb des Clusters zu finden und es somit zu großem Datenverlust kommen kann (Thoemmes & West, 2011, S. 525). Bei Untersuchungen, in denen der Cluster kein zentraler Bestandteil des Studiendesigns ist und nicht explizit in den Fokus rückt, ist das Matching zwischen den Clustern eine Alternative, die ebenfalls zu guten Ergebnissen führt (Thoemmes & West, 2011, S. 541).

Nachdem die Einflussfaktoren bestimmt sind, wird mit diesen eine logistische Regression auf die interessierende Variable, zum Beispiel „Lehrkrafturteil leseauffällig“, als abhängige Variable gebildet (Thoemmes, 2012, S. 4). Nominal- und ordinalskalierte Variablen werden hierzu in Dummyvariablen überführt. Metrische Kovariablen müssen nicht angepasst werden. Aus den ermittelten Parametern wird abschließend für jede Person der Propensity Score ermittelt (Müller, 2012, S. 11–12). Um die Kontrollgruppe zu bestimmen, müssen nun Personen gefunden werden, deren Propensity Scores denen der Treatmentgruppe möglichst ähnlich sind. Hierfür stehen verschiedene Verfahren zur Verfügung, wobei die Methode des „nearest neighbour“-Matchings am geläufigsten ist. Weitere Auswahlmöglichkeiten ergeben sich dadurch, dass die Personen mit oder ohne Replacement, mit one-to-one-Matching oder mit one-to-many-Matching bestimmt werden können (Müller, 2012, S. 12). One-to-one bedeutet, dass jeder Person in der Treatmentgruppe genau eine Person in der Kontrollgruppe zugeordnet wird. Beim Matching mit one-to-many wird eine Person in der Treatmentgruppe mehreren Personen in der Kontrollgruppe zugeordnet. Hierüber kann das Matching optimiert werden, wobei das Maximum bei 1:5 liegt. Empfohlen wird es vor allem dann, wenn die Stichprobengrößen der beiden Gruppen stark voneinander abweichen (Thoemmes, 2012, S. 10). Caliendo und Kopeinig (2008, S. 42) empfehlen ebenfalls das one-to-many-Matching, gegebenenfalls auch mit Replacement. Dies bedeutet, dass eine Person in der Kontrollgruppe, nachdem sie bereits einer Person in der Treatmentgruppe zugeordnet wurde, auch noch einer weiteren Person zugeordnet werden kann. Sollte man sich gegen Replacement entscheiden, dann muss darauf geachtet werden, dass das Matching in zufälliger Reihenfolge erfolgt. Insgesamt gibt es keine Methode, die den anderen überlegen wäre. So wird die Anwendung von Replacement vor allem dann vorgeschlagen, wenn es weniger Probanden in der Kontrollgruppe gibt, die den Probanden in der Treatmentgruppe zugeordnet werden können. Das Vorgehen sollte je nach Situation bzw. Verteilung auf Treatment- und Kontrollgruppe angepasst werden (Caliendo & Kopeinig, 2008, S. 45). Stuart (2010, S. 15) rät dazu, mehrere Methoden parallel zu verwenden und sich für das beste Ergebnis zu entscheiden. Dieses wird u. a. über die standardisierten Mittelwertdifferenzen über alle Kovariaten hinweg bestimmt, welche möglichst gering ausfallen sollen. Ebenso kann man auch besonders relevante Einflussfaktoren ins Auge fassen und die Methode wählen, die für diese Variablen die kleinsten standardisierten Mittelwertdifferenzen erzielt oder die Methode, welche zu den wenigsten großen standardisierten Mittelwertdifferenzen ( $> .25$ ) führt. Die genannten standardisierten Mittelwertdifferenzen beziehen sich bereits auf einen Teil der Adäquatheitsprüfungen, welche zum Ende dieses Abschnitts noch einmal explizit erläutert werden.

Wie in Kapitel 11.6.3 festgehalten, liegen teilweise multipel imputierte Datensätze vor, da einige der ausgewählten Einflussfaktoren fehlende Daten aufweisen (beispielsweise der sozioökonomische Status). Auch für diesen speziellen Fall gibt es bestimmte Verfahren, mit denen das Propensity Score-Matching adäquat mit diesen multipel imputierten Datensätzen durchgeführt werden kann. Mitra und Reiter (2016) gehen auf zwei Vorgehensweisen ein: der

Within- und der Across-Ansatz. Beim Within-Ansatz wird das Propensity Score-Matching für jede Imputation einzeln durchgeführt. Im Anschluss werden die jeweiligen statistischen Auswertungsmethoden  $m$  mal berechnet, was zu  $m$  Treatmenteffekten führt (Mitra & Reiter, 2016, S. 190). Als Treatmenteffekte werden die Auswirkungen des Treatments bezeichnet (Müller, 2012, S. 1), also beispielsweise die spätere Leseleistung der leseauffällig gewordenen Kinder im Vergleich zur Gruppe der nicht leseauffällig gewordenen Kinder. Diese  $m$  Treatmenteffekte werden abschließend zu einem Treatmenteffekt gemittelt. Der Across-Ansatz sieht dagegen vor, dass für jede  $m$  Imputation der Propensity Score berechnet wird. Dieser Propensity Score wird im Anschluss für alle  $m$  Imputationen gemittelt. Mit dem gemittelten Propensity Score wird dann das Matching durchgeführt. Mitra und Reiter (2016, S. 203) empfehlen aufgrund der besseren Ergebnisse in ihrer Simulationsstudie, das Matching mit der Across-Methode durchzuführen.

Auch wenn das Propensity Score-Verfahren sehr häufig empfohlen und durchgeführt wird, ist auch dieses Vorgehen nicht unproblematisch. King und Nielsen (2019, S. 16) sprechen sogar von dem „PSM-Paradox“, wonach die Daten nach dem statistischen Matching schlechter sein können als zuvor, indem das Ungleichgewicht zwischen den zugrundeliegenden Verteilungen der Kovariaten erhöht wird. Dieser Effekt verstärkt sich, wenn viele Kovariaten zu berücksichtigen sind. Nach der Prozedur müssen entsprechend Adäquatheitsprüfungen stattfinden. Die Frage lautet, ob ein Balancing bezüglich der Kovariaten erreicht wurde. Betrachtet werden statistische Größen vor und nach dem Matching. Zum einen werden die standardisierten Mittelwertdifferenzen betrachtet, welche möglichst niedrig und nahe 0 sein sollten. Außerdem werden statistische Tests durchgeführt, die die Gruppen vor und nach dem Matching hinsichtlich der Kovariaten vergleichen, um festzustellen, ob es signifikante Unterschiede zwischen den Gruppen gibt (Thoemmes, 2012, S. 5).

Mit SPSS lassen sich zwar multiple Imputationen durchführen, allerdings enthält die Software kein natives Matching-Verfahren (IBM Corp, 2016). Für R gibt es dagegen Pakete, welche die dargestellte und empfohlene Vorgehensweise erlauben. Verwendet werden hierfür in dieser Untersuchung die Pakete *mice* (van Buuren & Groothuis-Oudshoorn, 2011), *MatchIT* (Ho, Imai, King & Stuart, 2011) und *cobalt* (Greifer, 2019). Zunächst werden mit dem Paket *mice* für den jeweiligen Datensatz  $m$  Imputationen durchgeführt. Anschließend wird mit dem Paket *MatchIT* der Propensity Score für alle  $m$  Datensätze berechnet. Dieser wird dann für alle  $m$  Datensätze gemittelt, sodass das Matching im letzten Schritt mit diesem ermittelten Propensity Score durchgeführt werden kann. Mit dem Paket *cobalt* lassen sich die standardisierten Mittelwertdifferenzen vor und nach dem Matching kompakt darstellen. Detaillierte Informationen zur jeweiligen Berechnung finden sich in den Kapiteln zu den entsprechenden Hypothesenauswertungen, denen das statistische Matching zur Bildung der Kontrollgruppe zugrunde liegt. Entsprechend der Vorschläge aus der Literatur wird für jede Berechnung jeweils die Matching-Methode verwendet, die zum besten Ergebnis führt.

## 12 Ergebnisse

Bevor die Ergebnisse der in Kapitel 10 aufgeführten Fragestellungen und deren Hypothesen vorgestellt werden, wird zunächst die Stichprobe deskriptiv beschrieben. Zusätzlich wird bereits ein besonderes Augenmerk auf die Gruppe der leseschwachen Kinder geworfen.

### 12.1 Lehrkräfte

Alle 13 Grundschullehrkräfte waren weiblich und unterrichteten das Fach Deutsch in den an der Studie teilnehmenden Klassen. Das Alter der Lehrkräfte wurde zum sechsten Messzeitpunkt, im Mai 2012, erhoben. Vier Lehrkräfte gaben ihr Alter nicht an. Die übrigen Lehrkräfte waren zwischen 28 und 60 Jahre alt ( $M = 46.56$ ,  $SD = 9.02$ ).

Zu jedem Messzeitpunkt beantworteten die Lehrkräfte Fragen zu ihren Schülerinnen und Schülern. Da eine Schule im Mai 2014 nicht mehr an der Erhebung teilnahm, liegen für die Lehrkräfte der Klassen M und N nur Antworten für die Messzeitpunkte 1 bis 6 vor.

### 12.2 Schülerinnen und Schüler

Es haben 7 Grundschulen mit insgesamt 13 Klassen und 281 Kindern an der Studie teilgenommen. Die Gruppe der Kinder setzt sich aus 122 Mädchen (43 Prozent) und 159 Jungen (57 Prozent) zusammen. Für 206 Kinder liegen zusätzlich die Altersangaben vor: Die Kinder waren zum ersten Messzeitpunkt zwischen 5 und 7 Jahre alt ( $M = 6.20$ ,  $SD = 0.47$ ).

Insgesamt haben 194 Schülerinnen und Schüler bis zu Messzeitpunkt 6 durchgehend an der Studie teilgenommen, was 69 Prozent der Gesamtstichprobe entspricht. Über den gesamten Verlauf von Messzeitpunkt 1 bis 7 liegen von 128 Kindern vollständige Datensätze vor. Für die hier zugrundeliegenden Fragestellungen sind insbesondere die Messzeitpunkte ab T2 von Interesse. Für T2 bis T6 liegt die Stichprobengröße bei 201 Kindern (72 Prozent) und für T2 bis T7 liegt sie bei 131 Kindern (47 Prozent).

#### 12.2.1 Familiensprache

Da die Familiensprache relevant für die Leseentwicklung der Kinder in der Schule sein kann (McElvany et al., 2009; Tiedemann & Billmann-Mahecha, 2007), wurde die in der Familie hauptsächlich gesprochene Sprache erfasst. Zu 253 Kindern liegen Informationen darüber vor, bei 28 Kindern (10 Prozent) fehlt die Angabe (s. Abbildung 1). Etwa zwei Drittel der Kinder ( $n = 190$ ) sprechen Deutsch als Muttersprache und diese zudem als alleinige Sprache zuhause. Jeweils 11 Prozent der Kinder sprechen sowohl Deutsch als auch eine andere Sprache ( $n = 32$ ) oder kein Deutsch ( $n = 31$ ).

## 12 Ergebnisse

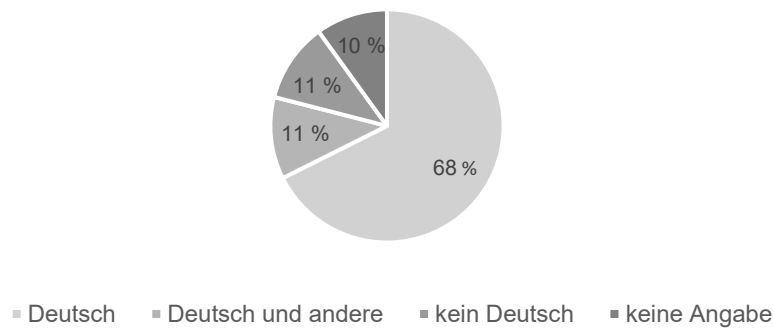


Abbildung 1. Häufigkeiten der Antworten auf die Frage, welche Sprache(n) zuhause gesprochen werden

### 12.2.2 Sozioökonomischer Hintergrund

Von insgesamt 384 Elternteilen (195 Mütter und 189 Väter) wurde der Elternbrief beantwortet, sodass für diese entsprechende Informationen über deren Berufstätigkeit vorliegen. Über 86 Mütter und 92 Väter fehlt diese Information. Von insgesamt 187 Haushalten (67 Prozent) liegen Informationen für beide Elternteile vor. Da der Familienstand der Eltern nicht erhoben wurde, kann keine Aussage darüber getroffen werden, ob die Haushalte, bei denen die Angabe des Vaters oder der Mutter fehlt, von alleinerziehenden Elternteilen beantwortet wurden oder ob die Angabe aus anderen Gründen fehlt.

Da sich die vorliegende Untersuchung auf den deutschen Raum konzentriert und die Anzahl an Angaben aufgrund der Rücklaufquote recht niedrig ist, wurde sich dafür entschieden, auf eine Unterteilung mit drei Kategorien zurückzugreifen. Die erste Gruppe beinhaltet akademische Berufe und Berufe mit hoher Verantwortung wie beispielsweise im Bereich der Unternehmensleitung. Die zweite Gruppe beinhaltet Elternteile mit Ausbildungsberufen. Die dritte Gruppe besteht aus unqualifizierten Arbeitskräften sowie erwerbslosen Elternteilen (inkl. Hausfrauen und -männer).

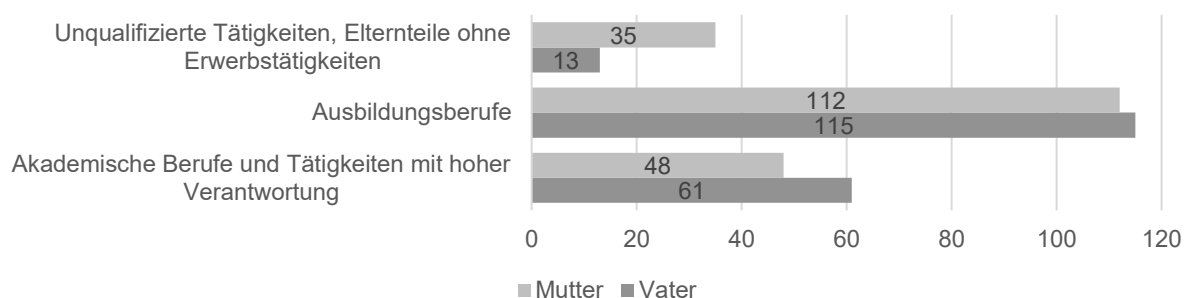


Abbildung 2. Berufstätigkeit der Elternteile

Abbildung 2 zeigt die Häufigkeit der einzelnen Kategorien für die Väter und die Mütter. Die Mehrheit mit 59 Prozent aller Angaben ( $n = 227$  Elternteile) geht einem Ausbildungsberuf nach. Etwas über ein Viertel der Eltern ( $n = 109$ ) geht einem akademischen Beruf oder einer

Tätigkeit mit hoher Verantwortung nach. Die übrigen knapp 13 Prozent ( $n = 48$ ) üben unqualifizierte Tätigkeiten aus oder sind nicht erwerbstätig.

Analog zu Trautwein und Baeriswyl (2007, S. 124) wird der jeweils höhere Wert einer der beiden Elternteile als Indikator für den sozioökonomischen Status im Haushalt betrachtet. Daraus ergibt sich die folgende Verteilung: Über insgesamt 187 Kinder liegen Informationen für beide Elternteile vor. Davon leben drei Prozent der Kinder ( $n = 5$ ) in Haushalten mit einem niedrigen sozioökonomischen Status, da hier kein Elternteil mindestens einem Ausbildungsberuf nachgeht. Etwas über die Hälfte der Kinder ( $n = 100$ ) kommt aus Haushalten mit einem mittleren sozioökonomischen Status, in denen mindestens ein Elternteil einem Ausbildungsberuf nachgeht, und fast die Hälfte der Kinder ( $n = 82$ ) lebt in Haushalten mit einem hohen sozioökonomischen Status, in denen mindestens ein Elternteil einem Beruf mit hoher Verantwortung oder einem akademischen Beruf nachgeht. Über neun Kinder liegen nur die Informationen für einen Elternteil vor. Diese teilen sich auf in zwei Kinder, zu denen nur Informationen über den Vater und sieben Kinder, zu denen nur Informationen der Mutter vorliegen. Ein Vater geht einem Ausbildungsberuf nach, der zweite Vater sowie eine der sieben Mütter gehen einem akademischen bzw. einem Beruf mit hoher Verantwortung nach. Die übrigen sechs Mütter gehen einem Ausbildungsberuf nach. Für 85 Kinder (30 Prozent) liegen keine entsprechenden Informationen vor.

### 12.2.3 Leseschwache Kinder zu den einzelnen Messzeitpunkten

Die Prävalenz der Leseschwäche ist, wie in Kapitel 6 dargestellt, abhängig davon, wie die Leseschwäche definiert und erfasst wird. In dieser Studie werden jeweils die Kinder als lese-schwach identifiziert, deren Leseleistung im Subtest Wortlesen des ELFE 1-6 mindestens eine Standardabweichung unterhalb des Durchschnitts der Gesamtstichprobe der hier unter-suchten Klassen liegt. Zur Darstellung werden T-Werte verwendet, deren Mittelwert bei  $T = 50$  ( $SD = 10$ ) liegt. Somit sind die Kinder lese-schwach, deren Leseleistung  $T < 40$  ist.

Tabelle 15

*Absolute Häufigkeiten der lese-schwachen, durchschnittlichen und lese-starken Kinder*

Messzeitpunkt	$n$	Leseschwach ( $T < 40$ )	Durchschnittlich ( $40 \leq T \leq 60$ )	Lesestark ( $T > 60$ )
T2	248	36 (15 %)	173 (70 %)	39 (16 %)
T3	252	43 (17 %)	168 (67 %)	41 (16 %)
T4	240	37 (15 %)	167 (70 %)	36 (15 %)
T6	238	38 (16 %)	166 (70 %)	34 (14 %)
T7	170	27 (16 %)	116 (68 %)	27 (16 %)

*Anmerkungen.* T = T-Wert im Wortlesen des ELFE 1-6, normiert an der untersuchten Stichprobe

Tabelle 15 zeigt die absoluten Häufigkeiten der als lese-schwach, lese-stark oder durchschnitt-lich im Lesen ausgemachten Schülerinnen und Schüler, wobei die lese-schwachen bzw. le-se-starken Kinder jeweils eine Standardabweichung unter bzw. über dem Mittelwert der Stich-probe liegen. Zu den Messzeitpunkten T1 und T5 wurde kein Lesetest durchgeführt.

## 12 Ergebnisse

Das Vorgehen der Normierung an der vorliegenden Stichprobe ist aufgrund der wiederholten Testungen notwendig. Es ist allerdings möglich, dass die untersuchten Kinder insgesamt im Vergleich mit den Normgruppen aus dem Testmanual überdurchschnittlich hohe Leistungen erzielen. Dementsprechend wären die Schülerinnen und Schüler dieser Stichprobe, die einen T-Wert unter 40 erzielen, vielleicht nicht leseschwach im Vergleich zur Normgruppe. Dies soll zunächst kurz geprüft werden.

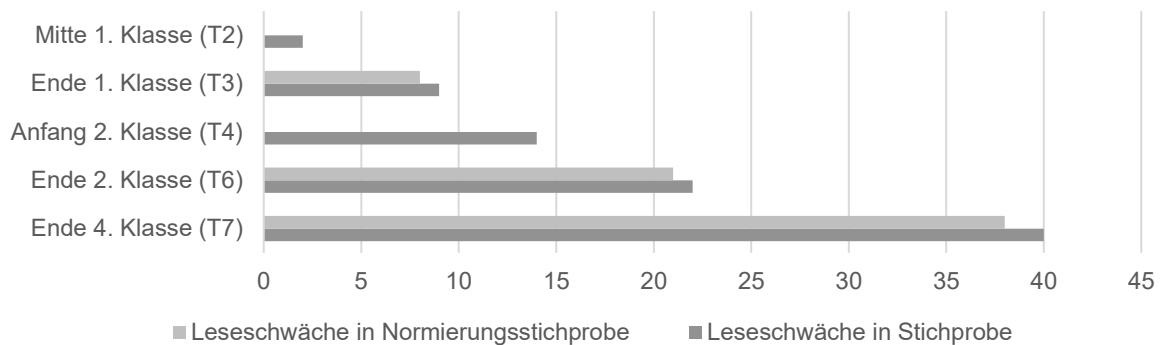


Abbildung 3. Dargestellt ist die Anzahl an maximal korrekt erlesenen Wörtern, für die eine Leseschwäche (im Sinne von  $T < 40$ ) festgestellt wird.

Abbildung 3 zeigt die Rohsummen der maximal erlesenen Wörter für die Kinder, die nach der jeweiligen Normierung einen T-Wert unter 40 erreicht haben. Für T2 und T4 liegen im Manual des Tests keine entsprechenden Normen vor. Zu den übrigen Messzeitpunkten zeigt sich, dass die in der vorliegenden Stichprobe als leseschwach eingestuft Kinder ein bis zwei Wörter mehr als die Normierungsstichprobe erlesen (je ein Wort mehr zu T3 und T6 und zwei Wörter mehr zu T7). Gleichzeitig fällt dieser Unterschied sehr gering aus, da diese im Vergleich zur Normierungsstichprobe mit 1 bzw. 2 von 72 Wörtern einen Vorsprung von nur einem bis drei Prozent gegenüber der Normierungsstichprobe aufweisen. Da die Kinder den gleichen Test wiederholt vorgelegt bekommen haben, erscheint diese Verschiebung plausibel. In T-Werten ausgedrückt erreichen sie bei Anlegen der Normen aus dem Manual Werte zwischen 40.2 und 40.7 und liegen somit auch im Vergleich mit dieser Referenzgruppe nur knapp über der Grenze von 40.

Das durchschnittliche Alter der zu T3 bis T7 jeweils als leseschwach identifizierten Kinder sowie die Anzahl an leseschwachen Jungen und Mädchen können Tabelle 16 entnommen werden. Zwischen 14 und 19 Prozent der teilnehmenden Jungen und zwischen 11 und 19 Prozent der teilnehmenden Mädchen sind jeweils als leseschwach identifiziert worden. Zu Beginn der Schulzeit bis zum Anfang der zweiten Klasse sind anteilmäßig etwas mehr Mädchen leseschwach. Ab Ende der zweiten Klasse sind es anteilig mehr leseschwache Jungen als Mädchen.

## 12 Ergebnisse

Tabelle 16

### *Soziodemographische Daten der leseschwachen Kinder*

	T2 (n = 36)	T3 (n = 43)	T4 (n = 37)	T6 (n = 38)	T7 (n = 27)
Alter zu T1: M (SD)	6.68 (0.51)	6.70 (0.50)	6.72 (0.51)	6.64 (0.54)	6.81 (0.58)
Geschlecht					
Männlich	19 (13 %)	22 (15 %)	19 (14 %)	27 (19 %)	17 (18 %)
Weiblich	17 (16 %)	21 (19 %)	18 (18 %)	11 (11 %)	10 (14 %)
Sozioökonomischer Status					
Unqualifiziert / nicht erwerbstätig	1 (17 %)	-	-	-	-
Ausbildungsberuf	15 (17 %)	15 (15 %)	18 (18 %)	14 (15 %)	18 (23 %)
Akademisch / Hohe Verantwortung	5 (6 %)	10 (13 %)	7 (9 %)	10 (12 %)	3 (6 %)
Familiensprache					
Nur Deutsch	18 (11 %)	31 (18 %)	23 (14 %)	25 (16 %)	15 (14 %)
Deutsch und andere	9 (31 %)	5 (19 %)	8 (28 %)	7 (23 %)	1 (4 %)
Nur andere	4 (14 %)	3 (10 %)	4 (15 %)	3 (11 %)	6 (33 %)

*Anmerkungen.* Die Prozentangaben beziehen sich auf die jeweilige Gesamtstichprobe der in der jeweiligen Zeile angeführten Gruppe.

In der Gesamtstichprobe gibt es lediglich fünf Kinder, die in einem Haushalt leben, in dem beide Elternteile unqualifizierten Tätigkeiten nachgehen oder nicht erwerbstätig sind. Davon ist ein Kind zu T2 leseschwach. Zu den folgenden Messzeitpunkten ist keines davon als leseschwach ausgemacht worden. Jeweils 15 bis 23 Prozent der Kinder, die in Haushalten leben, in denen mindestens ein Elternteil einem Ausbildungsberuf nachgeht, sind leseschwach. Für Haushalte mit mindestens einem Elternteil mit einem akademischen oder einem Beruf mit hoher Verantwortung sind 6 bis 13 Prozent leseschwach.

Tabelle 17

### *Häufigkeit der Leseschwäche je Klasse*

	T2		T3		T4		T6		T7	
	n	%	n	%	n	%	n	%	n	%
Klasse A	0	0	2	13	3	19	1	7	2	10
Klasse B	4	20	6	29	2	12	4	22	2	12
Klasse C	4	19	3	15	2	9	2	9	5	24
Klasse D	3	16	5	28	4	22	4	24	5	31
Klasse E	6	30	4	21	5	28	8	44	3	20
Klasse F	6	32	8	42	6	33	6	33	3	21
Klasse G	0	0	1	5	1	5	0	0	0	0
Klasse H	2	14	3	20	2	15	2	15	3	27
Klasse I	1	7	2	12	1	7	3	20	1	13
Klasse K	4	16	1	4	3	12	1	4	-*	-*
Klasse L	3	13	6	23	5	20	3	12	-*	-*
Klasse M	3	16	2	11	3	18	4	22	2	13
Klasse N	0	0	0	0	0	0	0	0	1	6
Gesamt	36	15	43	17	37	15	38	16	28	16

*Anmerkungen.* Die Prozentangaben beziehen sich jeweils auf die Anzahl der teilnehmenden Kinder in der Klasse. \*Klassen K und L haben zu T7 nicht mehr an der Untersuchung teilgenommen.

Von den Kindern, die zuhause nur Deutsch sprechen, sind jeweils 11 bis 18 Prozent leseschwach. Von den Kindern, die zuhause sowohl Deutsch wie auch mindestens eine andere

## 12 Ergebnisse

Sprache sprechen, sind 4 bis 31 Prozent leseschwach. Ähnlich sieht es für Kinder aus, die zuhause kein Deutsch sprechen. Hiervon sind zu T2 bis T6 jeweils 10 bis 14 Prozent leseschwach. Zu T7 ist aus dieser Gruppe jedes dritte Kind, welches zu diesem Zeitpunkt noch an der Studie teilgenommen hat, leseschwach.

In allen Klassen gibt es zu mindestens einem Messzeitpunkt leseschwache Kinder (s. Tabelle 17). Die Häufigkeit unterscheidet sich allerdings zwischen den Klassen. Die niedrigste Rate hat Klasse N, in der nur ein Kind zu T7 als leseschwach identifiziert wird. Klasse G hat eine ähnlich niedrige Prävalenz: Zu T3 und T4 gibt es jeweils ein Kind mit Leseschwäche. Dagegen steht Klasse E mit der höchsten Prävalenz: Hier haben 20 bis 44 Prozent der Kinder eine Leseschwäche.

Mit Ausnahme von T4 bestehen etwa bei der Hälfte der leseschwachen Kinder kombinierte Schwierigkeiten im Lesen und im Schreiben mit T-Werten unter 40 sowohl im Wortlesen wie auch im zeitgleich erhobenen Rechtschreibtest. Zu T4 gilt dies nur für etwa ein Drittel der leseschwachen Kinder (s. Tabelle 18).

Tabelle 18

### *Kombinierte Schwierigkeiten im Lesen und der Rechtschreibung bei leseschwachen Kindern*

	T2 (n = 36)		T3 (n = 43)		T4 (n = 37)		T6 (n = 38)		T7 (n = 27)	
	n	%	n	%	n	%	n	%	n	%
Kombinierte LRS	18	50	23	54	12	32	19	51	14	52

*Anmerkungen.* Verwendete Rechtschreibtests: Schreiben nach Bildern zu T2 und T4, DERET 1-2+ zu T3 und T6 sowie DERET 3-4+ zu T7

Die Leistungen der leseschwachen Kinder im Rechtschreiben sowie in der Laut-Buchstaben-Zuordnung und der Buchstabenkenntnis sind Tabelle 19 zu entnehmen. Um diese einordnen zu können, sind alle Ergebnisse in T-Werten dargestellt.

Tabelle 19

### *Schulische Leistungen bei zu T2 bis T7 jeweils leseschwachen Kindern in T-Werten*

	Laut-Buchstaben-Zuordnung <sup>+</sup>			Buchstabenkenntnis			Rechtschreibung		
	M (SD)	Min	Max	M (SD)	Min	Max	M (SD)	Min	Max
T2	43.18 (7.21)	32.44	59.92	42.43 (8.78)	26.14	57.02	39.44 (6.93)	26.14	55.20
T3	40.47 (8.40)	21.90	59.98	-	-	-	40.01 (9.27)	21.90	63.68
T4	-	-	-	-	-	-	43.65 (10.36)	22.04	67.06
T6	-	-	-	-	-	-	39.94 (7.28)	22.11	54.09
T7	-	-	-	-	-	-	40.39 (7.58)	23.23	53.82

*Anmerkungen.* <sup>+</sup>Zu T2 Daten aus der Erhebung von T1, Rechtschreibleistung T2 und T4 erhoben mit Schreiben nach Bildern, zu T3 und T6 erhoben mit DERET 1-2+ und zu T7 erhoben mit DERET 3-4+ (T7).

Es zeigen sich insgesamt knapp unterdurchschnittliche ( $T < 40$ ) bzw. grenzwertig durchschnittliche Rechtschreibleistungen für die jeweils als leseschwach identifizierten Kinder. Die Leistungen in der Laut-Buchstaben-Zuordnung und der Buchstabenkenntnis sind insgesamt durchschnittlich, wobei einige Werte im unteren durchschnittlichen Bereich ( $T < 45$ ) zu verorten sind. Die Minimalwerte liegen durchgehend im unterdurchschnittlichen ( $T < 40$ ) bis weit

unterdurchschnittlichen Bereich ( $T < 30$ ). Die Maximalwerte liegen fast immer im oberen durchschnittlichen Bereich mit T-Werten über 50. Für fünf Skalen liegen die Maximalwerte auch im überdurchschnittlichen Bereich ( $T > 60$ ). Die Leistungen der leseschwachen Kinder werden entsprechend als heterogen eingeschätzt. In Rohwerten ausgedrückt kannten die leseschwachen Kinder Mitte der ersten Klasse 9 bis 25 von insgesamt 26 Buchstaben, was ebenfalls die Heterogenität dieser Gruppe verdeutlicht.

Tabelle 20

*Silben- und Satzsegmentation leseschwacher Kinder in T-Werten*

	Silbensegmentation			Satzsegmentation <sup>+</sup>		
	<i>M (SD)</i>	Min	Max	<i>M (SD)</i>	Min	Max
T2	47.31 (8.48)	33.01	60.85	-	-	-
T4	44.50 (8.78)	30.82	64.51	42.29 (7.89)	31.73	56.80
T6	40.51 (6.38)	24.04	54.57	-	-	-
T7	43.91 (8.53)	30.78	64.02	48.50 (6.79)	26.58	50.89

*Anmerkungen.* <sup>+</sup>Satzsegmentation wurde erst ab T4 erhoben. Zu T6 fand keine Testung in Satzsegmentation statt.

Ein erfolgreiches Bearbeiten der Items zur Erfassung der Silben- und Satzsegmentation ab T3 setzt voraus, dass Kinder die einzelnen Wörter bzw. die Sätze erlesen können. Zu T1 und T2 wurden die Wörter im Test als Bilder dargestellt und von der Testleiterin vorgelesen, so dass die Kinder diese nicht selbst lesen mussten. Tabelle 20 gibt die Mittelwerte für die Skalen zur Silben- und Satzsegmentation für die jeweils als leseschwach identifizierten Kinder an. Alle Werte befinden sich im durchschnittlichen Bereich, wobei diese durchgehend unterhalb von  $T = 50$  liegen. Zu T6 erreichen die Kinder in der Silbensegmentation nur eine grenzwertig noch durchschnittliche Leistung.

Tabelle 21

*Selbstkonzept und logisches Denken bei leseschwachen Kindern in T-Werten*

	Schulisches SK Deutsch (T2)			SK der Schulfähigkeit			Logisches Denken (T5)		
	<i>M (SD)</i>	Min	Max	<i>M (SD)</i>	Min	Max	<i>M (SD)</i>	Min	Max
T2	47.25 (9.66)	27.09	57.04	-	-	-	48.16 (10.14)	28.21	66.16
T3	43.87 (9.28)	27.09	57.04	44.56 (9.21)	27.09	57.04	48.68 (9.75)	28.21	71.79
T4	48.16 (9.66)	27.09	57.04	-	-	-	47.44 (9.55)	31.31	71.79
T6	50.17 (8.19)	32.98	57.04	48.65 (10.16)	28.00	63.00	49.10 (9.95)	28.21	71.79
T7	49.66 (8.53)	32.98	57.04	45.52 (8.00)	33.00	59.00	48.68 (9.82)	28.21	71.79

*Anmerkungen.* SK = Selbstkonzept. Schulisches Selbstkonzept Deutsch wurde zu T2 erhoben. Logisches Denken wurde zu T5 erhoben. Selbstkonzept der Schulfähigkeit wurde zu T3 und T6 jeweils erhoben mit FEES 1-2 und zu T7 mit FEES 3-4.

Betrachtet man die erreichten Minimal- und Maximalwerte, wird auch an dieser Stelle deutlich, dass die Kinder sehr heterogene Leistungen erzielt haben. In der Silbensegmentation liegt der Maximalwert bei drei von vier Messzeitpunkten im überdurchschnittlichen Bereich mit  $T > 60$ . Gleichzeitig liegt der Minimalwert in der Silbensegmentation zu einem Zeitpunkt und in der Satzsegmentation zu zwei Zeitpunkten im weit unterdurchschnittlichen Bereich von

$T < 30$ . Die Werte im logischen Denken liegen ebenfalls durchgehend im durchschnittlichen Bereich. Auch das schulische Selbstkonzept im Fach Deutsch und das Fähigkeitsselbstkonzept des FEES sind insgesamt durchschnittlich (s. Tabelle 21). Die Minimal- und Maximalwerte deuten wieder auf eine große Heterogenität hin. Die Maximalwerte im logischen Denken liegen zu T3 bis T7 sogar bei  $T > 70$ .

#### 12.2.4 Zu Beginn der Grundschulzeit leseschwache Kinder

Zur Prüfung einiger Hypothesen werden Kinder betrachtet, die zu *Beginn der Schulzeit*, also zu T2 und/oder T3, leseschwach waren. Diese Gruppe besteht aus insgesamt 62 Kindern. Jeweils die Hälfte sind Jungen bzw. Mädchen. Im Schnitt sind die Kinder zu T1 6.67 Jahre alt ( $SD = 0.48$ ). Ein Kind kommt aus einem Haushalt, in dem beide Eltern unqualifizierten oder keinen Tätigkeiten nachgehen. Knapp ein Viertel der Kinder kommt aus Haushalten, in denen beide Elternteile einem akademischen oder einem Beruf mit hoher Verantwortung nachgehen ( $n = 14$ ). Die übrigen, knapp 40 Prozent ( $n = 24$ ), leben in einem Haushalt, in dem mindestens ein Elternteil einem Ausbildungsberuf nachgeht. Für 23 Kinder liegt keine Information über den sozioökonomischen Status vor. Für sieben Kinder gibt es keine Information zur Familiensprache. Etwas über 60 Prozent ( $n = 38$ ) sprechen zuhause nur Deutsch. Zehn Prozent ( $n = 6$ ) sprechen dort nur eine andere Sprache. Die übrigen Kinder ( $n = 11$ ) sprechen zuhause Deutsch und eine andere Sprache.

Tabelle 22

*Absolute Häufigkeit der zu T2 oder T3 leseschwachen Kinder je Klasse*

Klasse	A	B	C	D	E	F	G	H	I	K	L	M	N
<i>n</i>	2	7	6	7	8	9	1	3	2	5	8	4	-
%	10	30	26	33	36	41	5	18	11	19	30	19	-

*Anmerkungen.* Die Prozentangabe bezieht sich auf die Anzahl der Kinder der jeweiligen Klasse.

Mit Ausnahme von Klasse N befindet sich in allen Klassen zu T2 oder T3 mindestens ein leseschwaches Kind (s. Tabelle 22). Es zeigen sich große Schwankungen zwischen den einzelnen Klassen mit Prävalenzen zwischen null und 41 Prozent.

Zu T6 liegen noch Daten von 49 Kindern von den ursprünglichen 62 vor. Knapp die Hälfte dieser Kinder ist auch zu diesem Messzeitpunkt noch leseschwach (s. Tabelle 23). Dies gilt sowohl für die basalen Lesefertigkeiten, dem Wort- und dem Satzlesen, wie auch für die höhere Ebene des Leseverständnisses, welche mit dem Subtest Textlesen des ELFE 1-6 erhoben wurde. Für T7 liegen noch von 31 Kindern Daten vor. Von diesen ist ein Drittel weiterhin als leseschwach einzustufen. Knapp die Hälfte ist auf den höheren Leseebenen – dem Satzlesen und dem Textverständnis – sowie in der Gesamtleistung leseschwach (s. Tabelle 23).

## 12 Ergebnisse

Tabelle 23

*Leseleistungen zu T6 und T7 von Kindern, die zu T2 oder T3 leseschwach waren*

	Basales Lesen		Sinnentnehmendes Lesen			Gesamtleseleistung		
	Wortlesen		Satzlesen		Textlesen		ELFE 1-6	
	T6	T7	T6	T7	T6	T7	T6	T7
<i>n</i>	49	31	49	31	49	30	49	30
<i>M</i>	40.51	43.95	40.22	40.72	39.99	41.76	39.43	41.10
<i>SD</i>	7.44	8.01	6.68	8.04	7.37	8.16	6.76	6.98
Min	22.09	23.19	22.09	23.19	22.09	23.21	22.09	23.21
Max	58.35	57.69	50.05	56.01	54.11	57.55	51.00	53.63
<i>n</i> < 40	23	10	23	14	24	13	23	14
% <i>n</i> < 40	47 %	32 %	47 %	45 %	49 %	43 %	47 %	47 %

*Anmerkungen.* *n* < 40 gibt die Zahl der Kinder an, die im jeweiligen Test einen T-Wert unter 40 erzielt haben. Die Prozentangabe berechnet sich jeweils aus den Kindern, für die zu T6 bzw. zu T7 noch Daten vorliegen.

Die Rechtschreibleistung der Kinder liegt zu T2 und T3 insgesamt im unteren durchschnittlichen Bereich (s. Tabelle 24). Von diesen 62 Kindern weisen zur Mitte der ersten Klasse 24 und Ende der ersten Klasse 28 Kinder unterdurchschnittliche Rechtschreibleistungen auf. Dies entspricht 39 bzw. 46 Prozent.

Tabelle 24

*Rechtschreibleistungen der zu T2 oder T3 leseschwachen Kindern*

	<i>n</i> < 40	%	<i>M</i> ( <i>SD</i> )	Min	Max
T2	24	39	41.61 (7.59)	26.14	59.66
T3	28	46	40.86 (8.24)	21.90	63.68

*Anmerkungen.* *n* < 40 bezieht sich auf die absolute Häufigkeit an Kindern mit einem T-Wert unter 40 in der Rechtschreibleistung. Die Prozentangabe bezieht sich auf die 62 Kinder, die zu T2 oder T3 leseschwach waren.

Die Leistungen der Kinder in der Laut-Buchstaben-Zuordnung, der Buchstabenkenntnis und der Silbensegmentation zu T1 bis T3 sind Tabelle 25 zu entnehmen. Alle Werte liegen im durchschnittlichen Bereich.

Tabelle 25

*Schulische Leistungen bei zu T2 oder T3 leseschwachen Kindern in T-Werten*

	Laut-Buchstaben-Zuordnung			Buchstabenkenntnis			Silbensegmentation		
	<i>M</i> ( <i>SD</i> )	Min	Max	<i>M</i> ( <i>SD</i> )	Min	Max	<i>M</i> ( <i>SD</i> )	Min	Max
T1	43.70 (8.17)	25.07	59.92	-	-	-	49.58 (8.18)	30.54	61.21
T2	-	-	-	43.16 (8.83)	21.94	57.02	48.14 (9.01)	33.01	60.85
T3	41.22 (7.87)	21.90	59.98	-	-	-	-	-	-

*Anmerkungen.* Die einzelnen Skalen wurden nicht zu jedem Messzeitpunkt erhoben.

Auch für diese Gruppe werden die Leistungen in der Silben- sowie der Satzsegmentation zu den späteren Messzeitpunkten T5 bis T7 betrachtet (s. Tabelle 26). Alle Werte liegen im durchschnittlichen Bereich. Der Maximalwert bei der Silbensegmentation liegt zu T5 im überdurchschnittlichen Bereich mit *T* = 65.66. Zu den übrigen Messzeitpunkten liegen die

## 12 Ergebnisse

Maximalwerte der Silben- und der Satzsegmentation im durchschnittlichen Bereich mit Werten unter 60.

Tabelle 26

### *Silben- und Satzsegmentation leseschwacher Kinder in T-Werten*

	Silbensegmentation			Satzsegmentation		
	<i>M (SD)</i>	Min	Max	<i>M (SD)</i>	Min	Max
T5	43.34 (8.31)	25.17	65.66	45.71 (8.94)	28.18	53.06
T6	43.43 (6.99)	27.77	55.41	-	-	-
T7	41.33 (7.55)	23.21	56.59	46.21 (8.76)	26.58	50.89

*Anmerkungen.* Zu T6 fand keine Testung in Satzsegmentation statt.

Auch die Werte im logischen Denken und dem schulischen Selbstkonzept (s. Tabelle 27) liegen im durchschnittlichen Bereich. Die Minimal- und Maximalwerte deuten auch bei diesen Werten auf eine große Heterogenität hin. Im logischen Denken liegt der Maximalwert bei T > 70.

Tabelle 27

### *Selbstkonzept und logisches Denken bei leseschwachen Kindern in T-Werten*

Schulisches SK Deutsch (T2)			SK der Schulfähigkeit (T3)			Logisches Denken (T5)		
<i>M (SD)</i>	Min	Max	<i>M (SD)</i>	Min	Max	<i>M (SD)</i>	Min	Max
45.41 (9.74)	27.09	57.04	44.80 (9.11)	29.00	63.00	48.22 (9.54)	28.21	71.79

*Anmerkungen.* SK = Selbstkonzept. Selbstkonzept der Schulfähigkeit aus der Erhebung des FEES 1-2 zu T3.

Für die Benennungsgeschwindigkeit zu T5 liegen die Daten von  $n = 51$  Kindern vor (s. Tabelle 28). Insgesamt liegen die Werte sowohl für das Benennen von Bildern wie auch von Buchstaben im durchschnittlichen Bereich.

Tabelle 28

### *Benennungsgeschwindigkeit bei leseschwachen Kindern in T-Werten*

Benennungsgeschwindigkeit Bilder			Benennungsgeschwindigkeit Buchstaben			Benennungsgeschwindigkeit Gesamtscore		
<i>M (SD)</i>	Min	Max	<i>M (SD)</i>	Min	Max	<i>M (SD)</i>	Min	Max
44.63 (10.21)	22.00	71.00	43.10 (8.91)	22.00	63.00	42.86 (8.82)	22.00	61.00

*Anmerkungen.* Die Benennungsgeschwindigkeit wurde zu T5 erhoben.

#### 12.2.5 Drop-Outs von Schülerinnen und Schülern

Zwei Schulklassen nahmen an der Nacherhebung Ende der vierten Klasse (T7) nicht mehr teil. Weitere Kinder haben aus anderen Gründen nicht mehr teilgenommen oder sind bei einzelnen Erhebungen nicht anwesend gewesen, beispielsweise wegen Umzug mit Schulwechsel, Klassenwiederholung oder Krankenhausaufenthalt. Die Gründe sind meistens unbekannt, da die Lehrkräfte sich nicht zu jedem abwesenden Kind entsprechend geäußert haben. Kinder, die eine Klasse wiederholen, verfügen sehr wahrscheinlich über niedrigere schulische Leistungen. Hierüber kommt es vermutlich auch zu Drop-Out von leseschwachen Kindern.

Tabelle 29

*Mittelwertunterschiede im Wortlesen für Kinder, die zu T7 anwesend bzw. abwesend waren*

	Abwesend zu T7		Anwesend zu T7		T (df)	p	d [95 %-KI]
	n	M (SD)	n	M (SD)			
T2	96	49.80 (10.66)	152	50.18 (9.34)	0.301 (246)	.76	-0.04 [-0.30, 0.22]
T3	98	49.39 (11.37)	154	50.40 (8.89)	0.750 <sup>+</sup> (170.74)	.45	-0.10 [-0.36, 0.15]
T4	84	49.38 (10.65)	156	50.34 (9.52)	0.710 (238)	.48	-0.10 [-0.36, 0.17]
T6	81	47.95 (9.06)	157	51.06 (10.22)	2.309 (236)	<b>.02*</b>	-0.32 [-0.59, -0.04]

Anmerkungen. <sup>+</sup>F (df), da der Welch-Test aufgrund nicht homogener Varianzen durchgeführt wurde. \*p < .05

Tabelle 29 zeigt die Leistungen im Wortlesen für die Kinder, die zu T7 an der Nacherhebung teilgenommen haben und die Kinder, die nicht mehr teilgenommen haben. Demnach gibt es nur für T6 einen signifikanten Unterschied. Kinder, die zu T7 nicht mehr an der Erhebung teilgenommen haben, haben zum Ende der zweiten Klasse (T6) signifikant niedrigere Leistungen im Wortlesen gezeigt. Allerdings liegen beide Gruppen mit ihren Leistungen im durchschnittlichen Bereich mit T-Werten von 47.95 und 51.06. Mit  $d = -0.32$  liegt ein mittlerer Effekt vor.

### 12.3 Fragestellung 1: Erkennen von Leseschwierigkeiten

Mit der ersten Fragestellung soll betrachtet werden, ob Schülerinnen und Schüler mit Leseschwierigkeiten von den Lehrkräften auch als solche erkannt und benannt werden und ob bestimmte Faktoren wie die Klassenleistung, das Geschlecht und der sozioökonomische Status diese Einschätzung beeinflussen. Bevor auf die einzelnen Hypothesen eingegangen wird, werden zunächst die Aussagen der Lehrkräfte über Auffälligkeiten genauer betrachtet.

#### 12.3.1 Genannte Auffälligkeiten der Lehrkräfte

In Kapitel 11.5 wurde das Vorgehen zur Kategorisierung der durch die Lehrkräfte gegebenen Antworten beschrieben. An dieser Stelle folgt nun die deskriptive Beschreibung der genannten Auffälligkeiten. Die Lehrkräfte konnten sowohl positive wie auch negative Auffälligkeiten nennen. Es wird zunächst allgemein auf die genannten Auffälligkeiten eingegangen, bevor positive Aussagen und die Leseauffälligkeit im Einzelnen in den Fokus rücken.

##### 12.3.1.1 Auffälligkeiten allgemein

Abbildung 4 kann entnommen werden, wie viele Kinder als auffällig bezeichnet wurden, wobei an dieser Stelle sowohl negative wie auch positive Auffälligkeiten berücksichtigt werden. Ebenso beinhalten diese Werte Aussagen über das Elternhaus der jeweiligen Kinder. In die Darstellung eingegangen sind alle Kinder, die zum jeweiligen Messzeitpunkt Teil der Klasse waren oder zwischenzeitlich die Klasse verlassen haben.

## 12 Ergebnisse

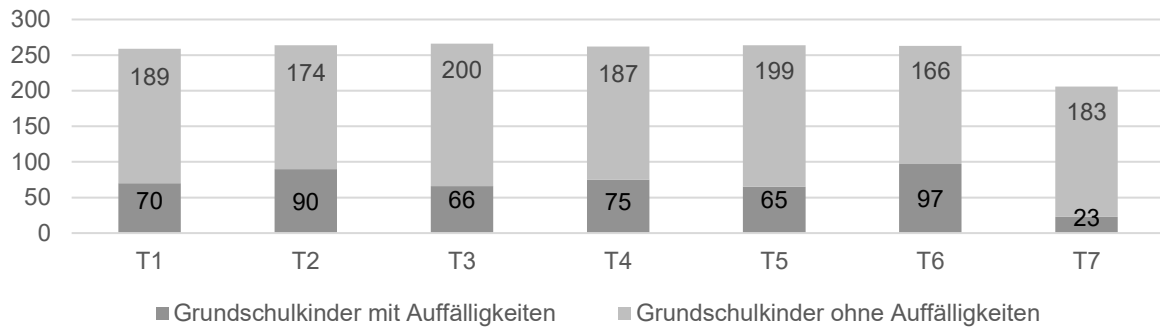


Abbildung 4. Anzahl der Grundschul Kinder, zu denen die Lehrkräfte zu den sieben Messzeitpunkten generell Auffälligkeiten genannt haben.

Zu Messzeitpunkt T7 war die Häufigkeit der Nennungen mit 23 Kindern (11 Prozent) am niedrigsten und zu Messzeitpunkt T6 mit 97 Kindern (37 Prozent) am höchsten. Lehrkräfte machten sowohl oberflächlichere Aussagen wie „allgemein schwach“ und „positiv“, aber auch spezifischere Aussagen wie „Buchstaben merken fällt ihm/ihr schwer“, „kein Buchstabenverständnis“, „Kontaktschwierigkeiten in der Klasse“ (alle Aussagen von Lehrkräften zu T2).

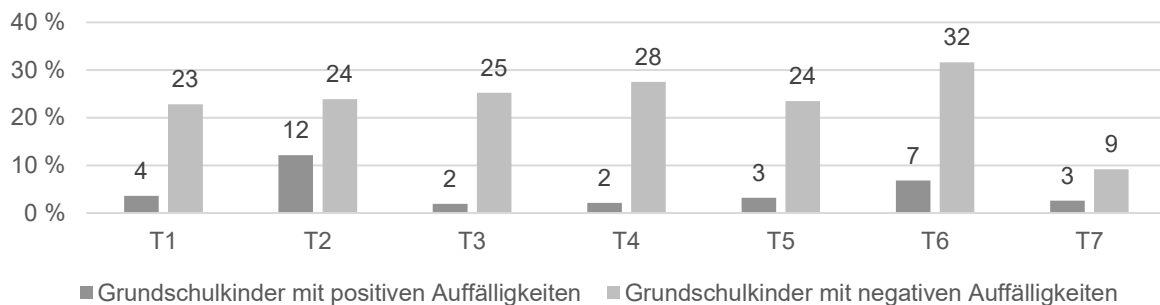


Abbildung 5. Relative Häufigkeiten der positiven und negativen Auffälligkeiten zu den sieben Messzeitpunkten.

Abbildung 5 zeigt die genannten Auffälligkeiten, unterteilt in negative und positive. In dieser Unterteilung sind die Aussagen zum Elternhaus nicht berücksichtigt worden, sodass sich die Nennungen immer auf die Kinder selbst beziehen. Eine Lehrkraft konnte zu jedem Kind sowohl positive wie auch negative Aussagen tätigen. Zu Messzeitpunkt T2 wurden mit einem Anteil von 12 Prozent die meisten positiven Aussagen gemacht. Negative Aussagen wurden durchgehend häufiger gemacht als positive. Hier war die niedrigste Häufigkeit mit 9 Prozent zu Messzeitpunkt T7 und die höchste mit 32 Prozent zu T6.

Tabelle 30 zeigt die Häufigkeit von positiven und negativen Auffälligkeiten, aufgeteilt auf die Klassen und die Messzeitpunkte. Da die Klassengrößen variieren, erfolgen die Angaben in Prozent. Es wurden jeweils zu 0 bis 88 Prozent der Schülerinnen und Schüler der Klasse Auffälligkeiten genannt. Zwei Lehrkräfte (Klassen E und N) haben zu zwei Messzeitpunkten keine Auffälligkeiten genannt. Zwei weitere Lehrkräfte (Klassen C und D) haben nur zu einem Messzeitpunkt, jeweils T7, keine Auffälligkeiten genannt.

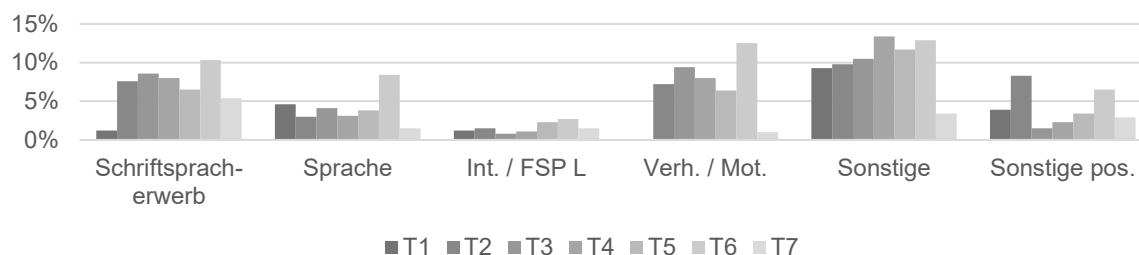
Tabelle 30

*Relative Häufigkeiten von Auffälligkeiten nach Klasse und Messzeitpunkt*

Klasse	T1	T2	T3	T4	T5	T6	T7
A	39	32	21	32	35	22	11
B	33	35	29	16	16	11	37
C	24	23	14	14	13	9	0
D	48	43	52	25	30	35	0
E	0	50	26	44	0	78	17
F	40	35	25	33	20	19	11
G	15	21	30	35	25	30	10
H	44	63	40	27	38	63	20
I	27	33	29	38	38	88	13
K	32	15	27	27	27	27	-
L	20	26	22	30	33	59	-
M	11	48	10	25	25	35	10
N	22	33	0	32	21	21	0

*Anmerkungen.* T1 bis T7 bezeichnen die Messzeitpunkte. Alle Angaben in Prozent. Klassen K und L haben zu Messzeitpunkt T7 nicht mehr an der Erhebung teilgenommen.

Entsprechend der Vorannahme nannten die Lehrkräfte Auffälligkeiten, welche sich nicht ausschließlich auf den Schriftspracherwerb bezogen. Über das Elternhaus wurde insgesamt wenig gesagt. Nur zu Messzeitpunkt T2 und auch nur für ein Kind wurden positive Aussagen über das Elternhaus gemacht: Die Lehrkraft bezeichnete die Eltern als sehr interessiert. Negative Aussagen zum Elternhaus wurden zu T2 (5 Nennungen), T3 (6 Nennungen), T4 (3 Nennungen), T5 (5 Nennungen) sowie T6 (2 Nennungen) getätigt. Dies betrifft jeweils ein bis zwei Prozent der Kinder.



*Abbildung 6.* Relative Häufigkeiten der Oberkategorien zu den sieben Messzeitpunkten. Angaben in Prozent. Int. = Intelligenz; FSP L = Förderschwerpunkt Lernen; Verh. = Verhalten; Mot. = Motivation; Sonstige pos. = sonstige positive Nennungen

Abbildung 6 zeigt die prozentualen Häufigkeiten für die Oberkategorien zu den sieben Messzeitpunkten. Nicht abgebildet sind die Oberkategorien „Motorik“, „Gut im Schreiben“ sowie „Gut im Lesen“. Diese Kategorien wurden jeweils nur selten bis gar nicht verwendet, weswegen auf eine Darstellung in der Tabelle verzichtet wurde. Negative Auffälligkeiten zum Schriftspracherwerb wurden zu einem bis zehn Prozent der Kinder gemacht. In sprachlichen Fertigkeiten sind zwischen zwei und fünf Prozent der Kinder negativ aufgefallen. Zwischen einem und drei Prozent der Kinder sind bezüglich ihrer kognitiven Fertigkeiten aufgefallen. Zu T1 wurden keine Aussagen über Motivation und Verhalten der Kinder getätigt. Zu den übrigen

Messzeitpunkten variierte der Anteil zwischen einem und 13 Prozent der Kinder, zu denen Auffälligkeiten in diesen Bereichen genannt wurden. Insgesamt wurden zu vielen Kindern Aussagen getätigt, die nur der Kategorie für sonstige Auffälligkeiten zuzuordnen waren (z. B. „allgemein schwach“ oder „tagesformabhängig“). Die Häufigkeit schwankt hier zwischen drei und 13 Prozent der Kinder. Sonstige positive Auffälligkeiten wurden zu zwei bis acht Prozent der Kinder getätigt.

### 12.3.1.2 Positive Auffälligkeiten

Tabelle 31 zeigt die relativen Häufigkeiten der Kinder, zu denen die Lehrkräfte positive Aussagen getätigt haben. Zu T2 wurden in 11 von 13 Klassen positive Auffälligkeiten genannt, zu T7 dagegen in nur drei Klassen. Zu allen anderen Messzeitpunkten waren es jeweils vier bis sechs Klassen. In fünf Klassen wurden nur zu einem oder zwei Messzeitpunkten positive Auffälligkeiten genannt. Lehrkraft G hat sich zu fünf Messzeitpunkten über einzelne Kinder positiv geäußert. Lehrkraft I hat zu Messzeitpunkt T6 zur Hälfte ihrer Klasse positive Auffälligkeiten benannt.

Tabelle 31

#### *Relative Häufigkeiten von positiv auffälligen Kindern in den Klassen*

	A	B	C	D	E	F	G	H	I	K	L	M	N
T1	10	26	-	-	-	5	-	-	-	-	-	-	5
T2	5	25	-	-	20	5	5	38	13	8	7	19	22
T3	-	5	5	5	-	5	-	7	-	-	-	-	-
T4	5	-	-	-	-	9	5	-	-	-	4	5	-
T5	-	-	-	-	-	-	10	6	17	7	-	-	5
T6	6	-	-	5	-	-	5	-	50	-	22	-	5
T7	-	17	-	-	-	-	5	6	-	*	*	-	-

*Anmerkungen.* Die Buchstaben A bis N bezeichnen die Lehrkräfte der 13 Klassen. Alle Angaben in Prozent.

\*Die Klassen haben zu T7 nicht mehr an der Studie teilgenommen.

### 12.3.1.3 Leseauffälligkeit

Die erste Erhebung, T1, fand kurz nach Einschulung der Kinder statt. Zu diesem Messzeitpunkt wurde noch kein Kind als leseauffällig eingeschätzt.

Tabelle 32

#### *Absolute und relative Häufigkeiten der als leseauffällig bzw. auffällig im Bereich Deutsch bezeichneten Kinder*

	T1 (n = 259)	T2 (n = 264)	T3 (n = 266)	T4 (n = 262)	T5 (n = 264)	T6 (n = 263)	T7 (n = 206)
Lesen	-	15 (6 %)	19 (7 %)	16 (6 %)	9 (3 %)	17 (7 %)	5 (2 %)
Bereich Deutsch	14 (5 %)	25 (10 %)	28 (11 %)	25 (10 %)	23 (9 %)	39 (15 %)	11 (5 %)

*Anmerkungen.* Die vordere Zahl gibt jeweils die absolute Häufigkeit an. Prozentuale Häufigkeit in Klammern. Zu Messzeitpunkt T1 war kein Kind leseauffällig.

Tabelle 32 zeigt die absoluten und relativen Häufigkeiten der Leseauffälligkeit zu den weiteren Messzeitpunkten. Die relative Häufigkeit schwankt zwischen zwei und sieben Prozent der Stichprobe.

Neben der Leseauffälligkeit wird ebenso betrachtet, wie viele Kinder nach Aussagen der Lehrkräfte allgemein im Bereich Deutsch auffällig geworden sind: In diesem Bereich auffällig geworden sind die Kinder, die von den Lehrkräften als auffällig im Lesen, in der Sprache, im Schreiben allgemein oder in der phonologischen Bewusstheit bezeichnet wurden. Diese Häufigkeiten finden sich ebenfalls in Tabelle 32. Die Werte liegen entsprechend höher: Es werden zu den einzelnen Messzeitpunkten zwischen 5 und 15 Prozent der Kinder als auffällig im Bereich Deutsch bezeichnet. Für diesen Bereich gibt es auch schon zu T1 Auffälligkeiten.

### 12.3.2 Hypothese 1.1

*Ein Drittel der leseschwachen Schülerinnen und Schüler (1 SD unter M im ELFE 1-6) fällt den Lehrkräften zum Messzeitpunkt T2 nicht auf - das heißt, dass die Lehrkräfte keine Aussagen über Auffälligkeiten im Lesen allgemein oder über Teilfertigkeiten im Spezifischen machen.*

Zum zweiten Messzeitpunkt liegen für 248 von 264 teilnehmenden Kindern vollständige Wortlesetests des ELFE 1-6 vor. Die Kinder haben zu diesem Zeitpunkt 0 bis 69 der insgesamt 72 Aufgaben korrekt beantwortet, wobei der Median bei 7.5 Aufgaben liegt. Ein T-Wert von unter 40 ergibt sich für die Rohwerte von 0 bis 2 korrekt gelösten Aufgaben. Dies trifft auf 36 Kinder zu. Zu einem knappen Viertel der Kinder ( $n = 63$ ) wurden negative Auffälligkeiten genannt. Als leseauffällig wurden davon 15 Kinder benannt, wobei ein Kind zu T2 nicht an den Testungen teilgenommen hat. Die übrigen 14 Kinder haben einen durchschnittlichen T-Wert von 45.37 ( $SD = 8.48$ ) im Wortlesetest erreicht. Die im Lehrkräfteurteil nicht im Lesen auffällig gewordenen Kinder ( $n = 234$ ) erreichen im Mittel einen T-Wert von 50.31 ( $SD = 9.87$ ). Um zu überprüfen, ob sich die Gruppen der leseauffälligen und der nicht im Lesen auffälligen Kinder in ihrer Leseleistung signifikant voneinander unterscheiden, wird der  $t$ -Test für unabhängige Stichproben durchgeführt. Der Levene-Test ergibt mit  $p = .31$  ( $F = 1.025$ ) keinen signifikanten Unterschied der Varianzen beider Gruppen, sodass trotz ungleich großer Stichproben der  $t$ -Test durchgeführt werden kann. Es ergibt sich mit  $p = .07$  ( $t = -1.833$ ,  $df = 246$ ,  $g = 0.50$  (95 %-KI[-0.04,1.04])) kein signifikanter Unterschied. Die leseauffälligen und nicht leseauffälligen Kinder unterscheiden sich demnach nicht signifikant in ihrer Leseleistung.

Tabelle 33 zeigt die Häufigkeiten der einzelnen Antwortkategorien für die 14 leseauffällig gewordenen Kinder. Zwei Kinder, die leseschwach waren, sind auch im Lesen auffällig geworden. Zu einem dieser beiden Kinder wurde etwas über die Buchstabenkenntnis gesagt. Zu dem zweiten Kind hat sich die Lehrkraft allgemein zur Leistung im Fach Deutsch geäußert. Zwölf Kinder waren nicht leseschwach, wurden aber als leseauffällig bezeichnet. Hierzu wurden zu sechs Kindern differenzierte Aussagen zum Schriftspracherwerb getätigt. Diese setzen sich aus Aussagen über die phonologische Bewusstheit, die Lesegeschwindigkeit/-

flüssigkeit sowie die Buchstabenkenntnis zusammen. Zu acht Kindern wurden undifferenzierte Aussagen getätigt, welche sich auf fünf Aussagen zu Lesen allgemein und drei allgemein auf das Fach Deutsch aufteilen.

Tabelle 33

*Häufigkeiten der Aussagen zu den leseauffälligen Kindern nach Kategorie für T2*

	Leseschwach (n = 2)	Nicht leseschwach (n = 12)
<b>Schriftsprache differenziert</b>	<b>1</b>	<b>6</b>
Phonologische Bewusstheit	-	1
Lesegeschwindigkeit/-flüssigkeit	-	1
Buchstabenkenntnis	1	4
<b>Schriftsprache undifferenziert</b>	<b>1</b>	<b>8</b>
Lesen allgemein	-	5
Deutsch allgemein	1	3

*Anmerkungen.* Die Oberkategorie (Schriftsprache differenziert bzw. Schriftsprache undifferenziert) setzt sich jeweils aus den darunter stehenden Unterkategorien zusammen. Pro Kind ist mehr als eine Nennung möglich.

Mit dem Binomialtest kann geprüft werden, ob die vorgefundenen Ausprägungen einer dichotomen Variable mit einer theoretisch angenommenen Verteilung übereinstimmen (Brosius, 2018, S. 997). Dieser Test wird mit der Software R durchgeführt, da R eine zweiseitige Testung erlaubt und das 95 %-Konfidenzintervall angibt. Um die Größe des Effekts zu bestimmen, wird im Falle der Binomialverteilung auf das Odds Ratio oder auf das Risk Ratio zurückgegriffen (Ellis, 2010, S. 6–7). Häufig wird das Risk Ratio bevorzugt, da es einfacher zu interpretieren ist und das Odds Ratio bei kleinen Differenzen oft schon hohe Werte annehmen kann (Ellis, 2010, S. 7).

Das Risk Ratio wird berechnet, indem das Auftreten der Leseauffälligkeit in der Gruppe der leseschwachen Kinder durch das Auftreten der Leseauffälligkeit in der Gruppe der nicht leseschwachen Kinder geteilt wird (Viera, 2008, S. 731):

$$RR = \frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}}$$

Zusätzlich zum Risk Ratio wird auch das jeweilige 95 %-Konfidenzintervall angegeben, welches mithilfe eines Rechners (MedCalc Software Ltd, o. J.) ermittelt wurde, der das Konfidenzintervall nach der Formel von Altman (1991) berechnet.

Das Odds Ratio berechnet sich, indem das Odds, dass ein leseschwaches Kind als leseauffällig bezeichnet wird, durch das Odds, dass ein nicht leseschwaches Kind als leseauffällig bezeichnet wird, geteilt wird (Viera, 2008, S. 731):

$$OR = \frac{a/b}{c/d}$$

Auch für das Odds Ratio wird jeweils das 95 %-Konfidenzintervall angegeben, welches ebenfalls mithilfe eines Rechners ermittelt wurde (Select Statistical Services Ltd, o. J.). Tabelle 34

stellt die Häufigkeitsverteilung der Leseschwäche und der Leseauffälligkeit dar. Von den 36 Kindern, die im Wortlesetest als leseschwach ausgemacht wurden, sind 2 Kinder (6 Prozent) den Lehrkräften leseauffällig geworden. Dagegen sind 34 leseschwache Kinder den Lehrkräften nicht leseauffällig geworden. Es wurden 12 Kinder als leseauffällig bezeichnet, die im Wortlesetest nicht als leseschwach identifiziert worden sind. Diese 12 Kinder haben im Wortlesetest T-Werte zwischen 41 und 59 erreicht. Die übrigen 200 Kinder, die nicht leseschwach sind, wurden von den Lehrkräften auch nicht als leseauffällig bezeichnet. Insgesamt zeigt sich somit, dass den Lehrkräften 34 von 36 leseschwachen Kinder (94 Prozent) zum zweiten Messzeitpunkt nicht leseauffällig werden.

Tabelle 34

*Häufigkeitsverteilung der Leseschwäche und Leseauffälligkeit zu Messzeitpunkt T2*

	Leseschwach	Nicht leseschwach	Summe
Leseauffällig	2 <sup>a</sup>	12 <sup>c</sup>	14
Nicht leseauffällig	34 <sup>b</sup>	200 <sup>d</sup>	234
Summe	36	212	248

*Anmerkungen.*  $n = 247$  Kinder. Die hochgestellten Buchstaben a bis d bezeichnen die Parameter zur Berechnung des Odds Ratio und des Risk Ratio.

Mit  $p < .001$  (Binomialtest, zweiseitige Testung) ergibt sich mit 94 Prozent nicht erkannter leseschwacher Kinder (95 %-KI[81 %, 99 %]) eine statistisch signifikante Abweichung zum erwarteten Wert von 33 Prozent nicht erkannter leseschwacher Kinder. Die Hypothese, dass der Anteil nicht erkannter leseschwacher Kinder in dieser Stichprobe dem Anteil von 33 Prozent entspricht, muss daher zurückgewiesen werden. Die Lehrkräfte bemerken zum zweiten Messzeitpunkt signifikant mehr leseschwache Kinder nicht. Das Odds Ratio beträgt 0.98 (95 %-KI[0.21, 4.58]). Das Risk Ratio beträgt ebenfalls 0.98 (95 %-KI[0.24, 4.44]). Die Wahrscheinlichkeit, als leseauffällig bezeichnet zu werden, wenn man leseschwach ist, ist zu diesem Zeitpunkt demnach etwa genauso hoch wie die Wahrscheinlichkeit, als leseauffällig bezeichnet zu werden, wenn man nicht leseschwach ist.

Tabelle 35

*Häufigkeit negativer Auffälligkeiten bei leseschwachen Kindern zu T2*

	Nicht leseauffällig	leseauffällig	Summe
Allgemein negativ auffällig	12	2	14
Nicht negativ auffällig	22	0	22
Summe	34	2	36

*Anmerkungen.*  $n = 36$  leseschwache Kinder

In einem weiteren Schritt wird der Faktor „Auffallen“ ausgeweitet und es wird betrachtet, ob leseschwache Kinder den Lehrkräften allgemein negativ auffällig werden. Dies beinhaltet alle negativen Aussagen, auch solche im Bereich des Lesens. Die meisten der allgemein negativen Nennungen ( $n = 7$ ) für die leseschwachen Kinder fallen in die Kategorie Sonstige Auffälligkeiten, die sich den anderen Kategorien nicht eindeutig zuordnen ließen (zum Beispiel die

Aussagen „schwach“ oder „teilweise schwach“). Jeweils drei Mal wurden Kinder in der Sprache sowie im Verhalten und der Motivation als auffällig bezeichnet. Tabelle 35 zeigt die entsprechende Häufigkeitsverteilung.

Es sind 14 leseschwache Kinder allgemein negativ auffällig geworden, zwei davon spezifisch im Lesen. Dagegen sind 22 der leseschwachen Kinder den Lehrkräften auch nicht allgemein auffällig geworden. Dies entspricht einem Anteil von 61 Prozent der leseschwachen Kinder (95 %-KI[43 %, 77 %]). Dies ist mit  $p = .001$  (Binomialtest, zweiseitige Testung) ein statistisch signifikanter Unterschied zu dem vermuteten Drittel an leseschwachen Kindern, die nicht auffallen.

Insgesamt muss die Hypothese, dass ein Drittel der leseschwachen Kinder den Lehrkräften zu T2 nicht auffällt, zurückgewiesen werden, da der Anteil der leseschwachen Kinder, die den Lehrkräften nicht negativ auffällig geworden sind, in dieser Stichprobe signifikant über dem angenommenen Anteil von 33 Prozent liegt.

### 12.3.3 Hypothese 1.2

*Ein Drittel der eigentlich leseschwachen Schülerinnen und Schüler (1 SD unter M im ELFE 1-6) fällt den Lehrkräften zum Messzeitpunkt T3 nicht auf - das heißt, dass die Lehrkräfte keine Aussagen über Auffälligkeiten im Lesen allgemein oder über Teilfertigkeiten im Spezifischen machen.*

Die Auswertung der Hypothese 1.2 erfolgt analog zur Auswertung der Hypothese 1.1. Zum Messzeitpunkt T3 liegen für 252 Kinder vollständige Wortlesetests vor. Insgesamt beantworteten die Kinder 1 bis 59 Aufgaben im Wortlesetest korrekt. Der Median liegt bei 17 Aufgaben. Rohwerte von 1 bis 9 korrekt gelösten Aufgaben entsprechen einem T-Wert von unter 40. Kinder, die höchstens 9 Aufgaben im Wortlesetest korrekt beantwortet haben, gelten demnach als leseschwach. Dies trifft zu Messzeitpunkt T3 auf 43 Kinder zu.

Insgesamt konnten die Lehrkräfte zu 266 Kindern, die sich bis zu diesem Zeitpunkt in der Klasse befanden bzw. zuvor in der Klasse befunden hatten, Aussagen machen. Die Anzahl an Kindern unterscheidet sich von der Anzahl der Kinder, die an der Testung teilgenommen haben, da die Lehrkräfte sich auch zu Kindern, die aus Gründen wie Rückstufung, Krankenhausaufenthalte usw. nicht in der Klasse waren, äußern konnten. Zu 66 Kindern wurden Aussagen über positive oder negative Auffälligkeiten gemacht, was einem Anteil von 25 Prozent entspricht.

Es wurden 19 Kinder als leseauffällig bezeichnet, wobei eines nicht an der Testung zu T3 teilgenommen hat. Im Schnitt haben die übrigen 18 Kinder einen T-Wert von 38.03 ( $SD = 5.20$ ) erreicht. Die nicht leseauffälligen Kinder ( $n = 234$ ) haben im Durchschnitt einen T-Wert von 50.93 ( $SD = 9.59$ ) erreicht. Der Levene-Test ergibt mit  $p = .02$  ( $F = 5.758$ ) einen signifikanten Unterschied der Varianzen beider Gruppen, sodass der Welch-Test statt des  $t$ -

Tests Anwendung findet. Die Leseleistungen der beiden Gruppen unterscheiden sich signifikant voneinander ( $p < .001$ ,  $t = -9.365$ ,  $df = 26.92$ ,  $g = 1.37$  (95 %-KI[0.88,1.87])).

Tabelle 36 sind die Häufigkeiten der Aussagen zu den schriftsprachlichen Antwortkategorien für die leseauffälligen Kinder zu entnehmen. Hierbei wird unterschieden zwischen leseschwachen und nicht leseschwachen Kindern.

Tabelle 36

*Häufigkeiten der Aussagen zu den leseauffälligen Kindern nach Kategorie für T3*

	Leseschwach (n = 11)	Nicht leseschwach (n = 7)
<b>Schriftsprache differenziert</b>	<b>4</b>	<b>3</b>
Phonologische Bewusstheit	2	1
Lesegeschwindigkeit/-flüssigkeit	-	1
Buchstabenkenntnis	2	-
Leseverständnis	2	1
<b>Schriftsprache undifferenziert</b>	<b>8</b>	<b>5</b>
Lesen allgemein	7	4
Deutsch allgemein	1	2

*Anmerkungen.* Die Oberkategorie (Schriftsprache differenziert bzw. Schriftsprache undifferenziert) setzt sich jeweils aus den darunter stehenden Unterkategorien zusammen. Pro Kind ist mehr als eine Nennung möglich.

Zu vier der 11 leseschwachen Kinder, die im Lesen auffällig geworden sind, wurden spezifische Aussagen über den Schriftspracherwerb getätigt. Diese teilen sich auf in zwei Aussagen zur phonologischen Bewusstheit, zwei Aussagen zur Buchstabenkenntnis und zwei Aussagen zum Leseerwerb. Zu acht Kindern wurden undifferenzierte Aussagen über den Schriftspracherwerb gemacht. Sieben Mal waren dies allgemeine Aussagen zum Lesen und einmal eine Aussage zu Deutsch allgemein. Zu drei nicht leseschwachen Kindern wurden differenziertere Aussagen über den Schriftspracherwerb getätigt. Diese setzten sich aus jeweils einer Aussage über die phonologische Bewusstheit, die Lesegeschwindigkeit/-flüssigkeit sowie das Leseverständnis zusammen. Zu fünf nicht leseschwachen Kindern wurden allgemeine negative Aussagen über den Schriftspracherwerb gemacht. Vier Nennungen finden sich für Lesen allgemein und zwei weitere Nennungen für Deutsch allgemein.

Tabelle 37

*Häufigkeitsverteilung der Leseschwäche und Leseauffälligkeit zu Messzeitpunkt T3*

	Leseschwach	Nicht leseschwach	Summe
Leseauffällig	11 <sup>a</sup>	7 <sup>c</sup>	18
Nicht leseauffällig	32 <sup>b</sup>	202 <sup>d</sup>	234
Summe	43	209	252

*Anmerkungen.* n = 247 Kinder. Die hochgestellten Buchstaben a bis d bezeichnen die Parameter zur Berechnung des Odds Ratio und des Risk Ratio.

Tabelle 37 zeigt die Häufigkeit der Leseschwäche und der Leseauffälligkeiten zu T3. Sieben Kinder, die nicht leseschwach sind, wurden als leseauffällig bezeichnet. Elf leseschwache Kinder wurden korrekt als leseauffällig genannt. Zu 32 leseschwachen Kindern wurden keine

Auffälligkeiten genannt. Daraus ergibt sich, dass 74 Prozent (95 %-KI[59 %, 86 %]) und somit die Mehrheit der leseschwachen Kinder den Lehrkräften zu diesem Messzeitpunkt nicht auffallen. Dieser Anteil unterscheidet sich signifikant vom erwarteten Anteil von 33 Prozent an nicht erkannten leseschwachen Kindern ( $p < .001$ , Binomialtest, zweiseitige Testung).

Insgesamt muss auch diese Hypothese, dass der Anteil an nicht erkannten leseschwachen Kindern einem Drittel entspricht, zurückgewiesen werden. Es werden signifikant mehr lese-schwache Kinder nicht korrekt erkannt. Das Odds Ratio beträgt 9.92 (95 %-KI[3.58,27.46]) und das Risk Ratio beträgt 7.65 (95 %-KI[3.13, 18.58]). Die Wahrscheinlichkeit, als lese-auffällig bezeichnet zu werden, wenn man leseschwach ist, ist fast acht Mal so hoch wie die Wahrscheinlichkeit, als leseauffällig bezeichnet zu werden, wenn man nicht leseschwach ist.

Tabelle 38

*Häufigkeit negativer Auffälligkeiten bei leseschwachen Kindern zu T3*

	Nicht leseauffällig	leseauffällig	Summe
Allgemein negativ auffällig	14	11	25
Nicht negativ auffällig	18	0	18
Summe	32	11	43

Anmerkungen.  $n = 43$

Auch bei dieser Hypothese wird in einem weiteren Schritt betrachtet, ob die leseschwachen Kinder den Lehrkräften allgemein negativ auffällig werden. Von den 43 leseschwachen Kindern trifft dies auf 25 Kinder zu, wobei 11 davon spezifisch im Lesen auffallen (s. Tabelle 38). Zu 18 leseschwachen Kindern haben sich die Lehrkräfte nicht geäußert, was einem Anteil von 42 Prozent entspricht (95 %-KI[27 %, 58 %]). Dieser Anteil unterscheidet sich nicht signifikant von der Nullhypothese, welche einen Anteil von 33 Prozent annimmt ( $p = .26$ , Binomialtest, zweiseitige Testung). Dementsprechend wird hier die Nullhypothese beibehalten: Von den leseschwachen Kindern fällt den Lehrkräften ein Drittel gar nicht auf. Die übrigen fallen ihnen negativ auf, auch wenn sich diese Auffälligkeiten nicht auf das Lesen beziehen. Wie zu T2 finden sich auch hier die häufigsten negativen Aussagen zu sonstigen Auffälligkeiten ( $n = 10$ ), die sich keiner anderen spezifischeren Kategorie zuordnen ließen (zum Beispiel „allgemein schwach“). Zu sieben Kindern wurden Aussagen über Verhalten und Motivation gemacht. Zu je ein bis drei Kindern wurden Aussagen über die Sprache, über kognitive Fähigkeiten und Motorik gemacht.

Die Hypothese, dass ein Drittel der leseschwachen Kinder den Lehrkräften zu T3 nicht im Lesen auffällt, muss zurückgewiesen werden, da der Anteil in dieser Stichprobe signifikant über 33 Prozent liegt. Wenn man sich allgemein negative Auffälligkeiten bei leseschwachen Kindern betrachtet, ergibt sich dagegen keine signifikante Abweichung von den angenommenen 33 Prozent.

## 12.3.4 Hypothese 1.3

Sollten sich zu den Messzeitpunkten T2 oder T3 einzelne Schulklassen als lesestark herausstellen (der Mittelwert der Leseleistung der Klasse liegt signifikant über dem Mittelwert der Gesamtstichprobe), wird überprüft, ob es zum Referenzgruppeneffekt kommt und entsprechend im Lesen schwächere Kinder als leseschwach genannt werden, welche verglichen mit der Gesamtstichprobe nicht leseschwach sind.

Um diese Hypothese auszuwerten, muss zunächst geprüft werden, ob es lesestarke Klassen gibt. Diese sollen mit in der Leseleistung durchschnittlichen Klassen verglichen werden. Leseschwache Klassen gehen nicht in die Analyse ein. Das Kriterium liegt in Form einer dichotomen Variable vor: Ist die Klasse insgesamt lesestark oder durchschnittlich in ihrer Leseleistung? Die zweite Variable ist ebenfalls dichotom: Hat die Lehrkraft Kinder mit mindestens durchschnittlichen Leseleistungen als leseauffällig bezeichnet oder nicht? Es liegt eine 2x2 Kreuztabelle vor, sodass der Chi-Quadrat-Test verwendet wird.

Tabelle 39

Wortleseleistung und Unterschiede zum erwarteten Mittelwert ( $T = 50$ ) in den Klassen

Klasse	T2					T3				
	<i>n</i>	<i>M</i> ( <i>SD</i> )	<i>t</i>	<i>df</i>	<i>p</i>	<i>n</i>	<i>M</i> ( <i>SD</i> )	<i>t</i>	<i>df</i>	<i>p</i>
A	16	58.08 (8.00)	4.040	15	.00**+	16	54.32 (10.37)	1.665	15	.12
B	20	46.56 (9.52)	-1.617	19	.12	21	45.51 (11.20)	-1.837	20	.08
C	21	45.53 (7.49)	-2.738	20	.01*~	20	46.26 (5.64)	-2.964	19	.01*~
D	19	46.04 (7.34)	-2.353	18	.03*~	18	46.41 (9.71)	-1.566	17	.14
E	20	45.00 (8.95)	-2.499	19	.02*~	19	46.72 (7.51)	-1.904	18	.07
F	19	45.37 (9.64)	-2.095	18	.05	19	43.79 (9.94)	-2.725	18	.01*~
G	19	56.90 (7.85)	3.827	18	.00**+	19	55.95 (9.22)	2.814	18	.01*+
H	14	46.18 (8.00)	-1.785	13	.10	15	48.35 (10.45)	-.613	14	.55
I	15	50.44 (7.83)	0.218	14	.83	17	53.31 (8.14)	1.674	16	.11
K	25	51.43 (8.67)	0.822	24	.42	25	54.63 (7.06)	3.280	24	.00**+
L	23	52.89 (10.87)	1.278	22	.22	26	49.70 (10.93)	-.140	25	.89
M	19	48.38 (10.91)	-0.646	18	.53	19	46.19 (6.26)	-2.652	18	.02*~
N	18	58.32 (8.05)	4.385	17	.00**+	18	59.25 (8.16)	4.806	17	.00**+

Anmerkungen. Die Buchstaben bezeichnen die 13 teilnehmenden Klassen. Die Mittelwerte beziehen sich auf die durchschnittlichen T-Werte im Wortlesetest. \* $p < .05$ ; \*\* $p < .01$ . +signifikant lesestarke Klasse (dunkelgrau hervorgehoben). ~signifikant leseschwache Klasse (hellgrau hervorgehoben)

Tabelle 39 zeigt die durchschnittlich erreichten T-Werte der Leseleistung für die Messzeitpunkte T2 und T3. Zum zweiten Messzeitpunkt liegen die Mittelwerte der T-Werte zwischen 45.00 ( $SD = 8.95$ ) und 58.32 ( $SD = 8.05$ ). Zu Messzeitpunkt T3 liegen die durchschnittlich erreichten T-Werte zwischen 43.79 ( $SD = 9.94$ ) und 59.25 ( $SD = 8.16$ ). Im nächsten Schritt wird für jede Klasse betrachtet, ob diese sich in ihrer durchschnittlichen Leseleistung signifikant vom Mittelwert ( $T = 50$ ) unterscheidet.

Um dies zu prüfen, wird für jede Klasse der Einstichproben- $t$ -Test durchgeführt. Der erwartete Mittelwert liegt bei  $T = 50$ . Tabelle 39 zeigt die Ergebnisse der Einstichproben- $t$ -Tests für alle Klassen. Es ergeben sich für T2 drei leseschwache Klassen (C, D und E). Klasse C ist als

## 12 Ergebnisse

einzig auch zu T3 noch lese schwach. Zusätzlich sind zum dritten Messzeitpunkt auch die Klassen F und M in ihrer mittleren Leseleistung unterdurchschnittlich.

Die Klassen A, G und N sind zu T2 überdurchschnittlich in ihrer Leseleistung. Zu T3 sind die Klassen G, K und N lese stark. Die lese starken Klassen mit ihrer Leseleistung sind Abbildung 7 zu entnehmen.

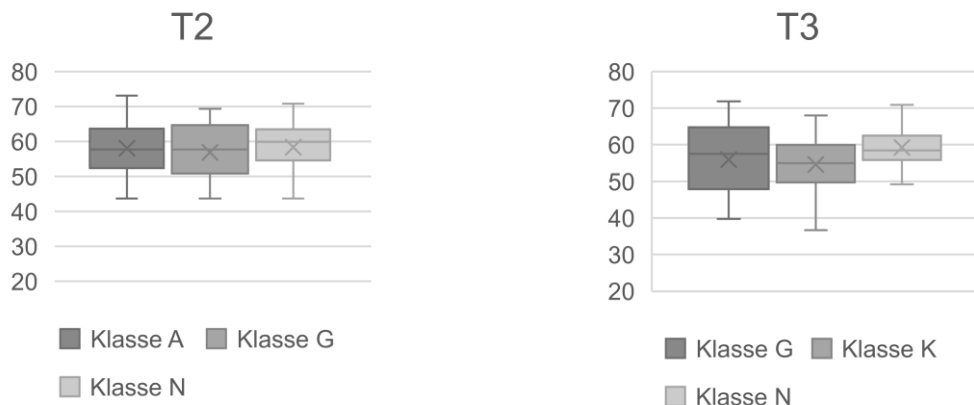


Abbildung 7. Boxplots zur Leistung im Wortlesetest für die Klassen mit signifikant höherem Mittelwert als  $T = 50$ . Das X markiert den Mittelwert. Die horizontale Linie stellt den Median dar.

Um Hypothese 1.3 zu prüfen, werden nur die Kinder betrachtet, die eine mindestens durchschnittliche Leseleistung aufweisen. Tabelle 40 zeigt die Häufigkeit, mit der die Lehrkräfte diese Schülerinnen und Schüler als im Lesen auffällig bezeichnet haben.

Tabelle 40

Häufigkeit von lese auffälligen, aber nicht lese schwachen Kindern sowie deren durchschnittliche Leseleistung getrennt nach Schulklassen (A bis N)

Klasse	T2		T3	
	<i>n</i>	<i>M</i>	<i>n</i>	<i>M</i>
A	1	43.70	0	-
B	1	49.24	0	-
C	0	-	1	42.69
D	1	46.17	3	42.69
E	2	46.59	0	-
G	0	-	1	42.69
H	2	44.94	0	-
I	1	54.77	0	-
K	0	-	0	-
M	3	46.19	2	43.55
N	1	59.24	0	-

Anmerkungen. Lesestärke Klassen sind dunkelgrau hervorgehoben, lese schwache Klassen hellgrau. *M* bezeichnet die durchschnittliche Leistung lese auffälligen Kindern im Wortlesetest in T-Werten. Klassen, in denen keine Fehleinschätzung vorkam, sind nicht dargestellt.

In keiner der lese starken Klassen wurde mehr als eine Fehleinschätzung in Richtung dieses Effekts vorgenommen. Zu T2 wurden insgesamt 2 von 53 Kinder (vier Prozent) in lese starken

Klassen fälschlicherweise als lese auffällig bezeichnet, während in den durchschnittlich lesestarken Klassen 7 von 135 Kinder (fünf Prozent) zu Unrecht als lese auffällig bezeichnet wurden. Da die Zelhäufigkeit in einer von vier Zellen eine erwartete Häufigkeit unter fünf aufweist, ist die Voraussetzung für den  $X^2$ -Test nicht erfüllt. Stattdessen wird Fishers exakter Test verwendet (Brosius, 2018, S. 502–503). Es gibt keinen statistisch signifikanten Unterschied zur theoretisch erwarteten Verteilung ( $p = 1.00$ , Fishers exakter Test bei zweiseitiger Testung). Es liegt ein Odds Ratio von 0.72 (95 %-KI[0.14, 3.57]) und ein Risk Ratio von 0.73 (95 %-KI[0.16, 3.39]) vor.

Zu T3 wurde in den lesestarken Klassen ein Kind von 62 (zwei Prozent) bei durchschnittlicher Leseleistung als lese auffällig bezeichnet. In den Klassen mit durchschnittlicher Leseleistung wurden 3 Kinder von 132 (zwei Prozent) mit durchschnittlicher Leseleistung als lese auffällig bezeichnet. Für diese Daten liegt ein Odds Ratio von 0.72 (95 %-KI[0.07, 7.08]) und ein Risk Ratio von 0.73 (95 %-KI[0.08, 6.84]) vor.

Auch hier ist die Voraussetzung für den  $X^2$ -Test nicht erfüllt, da zwei von vier Zellen eine erwartete Häufigkeit von unter fünf haben, sodass Fishers exakter Test verwendet wird. Es liegt kein statistisch signifikanter Unterschied zur theoretisch erwarteten Verteilung vor ( $p = 1.00$ , Fishers exakter Test bei zweiseitiger Testung).

Insgesamt zeigt sich, dass es in Bezug auf fehlerhafte Einschätzungen von Leseauffälligkeiten, wenn durchschnittliche Leseleistungen vorliegen, keine Unterschiede zwischen den lesestarken und den im Lesen durchschnittlichen Klassen gibt und die Lehrkräfte sich in ihrer Genauigkeit nicht unterscheiden. Die Hypothese, dass sich der Referenzgruppeneffekt zeigt und in lesestarken Klassen häufiger nicht leseschwache Kinder als lese auffällig bezeichnet werden, muss daher zurückgewiesen werden.

#### 12.3.5 Hypothese 1.4

*Lehrkräfte bezeichnen Mädchen, die leseschwach sind, seltener als lese auffällig, als dies für leseschwache Jungen geschieht.*

Diese Hypothese wird ebenfalls getrennt für die einzelnen Messzeitpunkte ausgewertet. Kapitel 8.1.6 hat aufgezeigt, dass sich Jungen und Mädchen im Verlauf der Schulzeit unterschiedlich entwickeln hinsichtlich der Leseleistung. Daher wird Messzeitpunkt T4 zusätzlich hinzugezogen. Zu T2 liegen Daten über die Wortleseleistung für 106 Mädchen und 142 Jungen vor, zu T3 für 109 Mädchen und 143 Jungen und zu T4 für 102 Mädchen und 138 Jungen. Jungen haben zu T2 eine durchschnittliche Leseleistung von  $T = 50.70$  ( $SD = 9.58$ ). Bei Mädchen liegt die durchschnittliche Leseleistung bei  $T = 49.14$  ( $SD = 10.18$ ). Dieser Unterschied ist nicht signifikant ( $t(246) = -1.24$ ,  $p = .22$ ,  $d = 0.16$  (95 %-KI[-0.09, 0.41])). Zu T3 liegt die durchschnittliche Leseleistung für Jungen bei  $50.70$  ( $SD = 9.77$ ) und für Mädchen bei  $49.09$  ( $SD = 10.08$ ). Dieser Unterschied ist ebenfalls nicht signifikant ( $t(250) = -1.28$ ,  $p = .20$ ,  $d = 0.16$  (95 %-KI[-0.09, 0.41])). Zu T4 liegt die durchschnittliche Leseleistung für Jungen bei

## 12 Ergebnisse

50.85 ( $SD = 10.19$ ) und für Mädchen bei 48.86 ( $SD = 9.47$ ). Auch dieser Unterschied ist nicht signifikant ( $t(238) = -1.54, p = .13, d = 0.20$  (95 %-KI[-0.06, 0.46])). Insgesamt unterscheiden sich die Leseleistungen von Jungen und Mädchen Mitte der ersten Klasse bis Anfang der zweiten Klasse nicht voneinander.

Tabelle 41

*Absolute Häufigkeiten von Leseschwäche und Leseauffälligkeit für Jungen und Mädchen*

		Jungen		Mädchen		Gesamt	$p$
		$n$	%	$n$	%		
T2	Leseschwach	19		17		36	1.00
	Leseauffällig und leseschwach	1	5	1	6	2	
T3	Leseschwach	22		21		43	.11
	Leseauffällig und leseschwach	8	36	3	14	11	
T4	Leseschwach	19		18		37	.01*
	Leseauffällig und leseschwach	9	47	1	6	10	

Anmerkungen. \* $p < .05$ , \*\* $p < .01$ .

Tabelle 41 ist zu entnehmen, wie viele leseschwache Jungen und Mädchen jeweils auch als leseauffällig bezeichnet wurden. Ob leseschwache Mädchen seltener im Lesen auffällig werden, wird anhand dieser Zahlen statistisch überprüft, wobei die jeweilige Anzahl an tatsächlich leseschwachen Jungen und Mädchen dieser Stichprobe die Vergleichsreferenz darstellen, welche im  $X^2$ -Test zur Berechnung der erwarteten relativen Häufigkeit verwendet wird (Brosius, 2018, S. 995).

Zu T2 wurde einer von 19 leseschwachen Jungen und eins von 17 leseschwachen Mädchen leseauffällig. Die erwarteten Zellhäufigkeiten liegen bei zwei Zellen unter 5, sodass Fishers exakter Test Anwendung findet. Es gibt keinen statistisch signifikanten Unterschied ( $p = 1.00$ , Fishers exakter Test, zweiseitige Testung). Das Odds Ratio beträgt 1.23 (95 %-KI [0.06, 19.50]) und das Risk Ratio beträgt 1.12 (95 %-KI[0.08,16.52]).

Zu T3 liegen die Zahlen höher. Knapp ein Drittel ( $n = 8$ ) der leseschwachen Jungen und 14 Prozent ( $n = 3$ ) der leseschwachen Mädchen wurden leseauffällig. Auch dieser Unterschied ist nicht statistisch signifikant ( $X^2 = 2.513, df = 1, p = .11$ , zweiseitige Testung). Das Odds Ratio beträgt 0.29 (95 %-KI [0.07, 1.31]) und das Risk Ratio beträgt 0.39 (95 %-KI[0.12, 1.28]). Zu T4 zeigt sich für Jungen die höchste Erkennungsrate mit 47 Prozent ( $n = 9$ ). Von den Mädchen wurde dagegen eins von insgesamt 18 leseschwachen Kindern auch leseauffällig, was sechs Prozent entspricht. Die erwartete Häufigkeit liegt bei einer von zwei Zellen unter fünf. Es wird entsprechend Fishers exakter Test verwendet. Der Unterschied ist statistisch signifikant ( $p = .01$ , Fishers exakter Test, zweiseitige Testung). Das Odds Ratio beträgt 0.07 (95 %-KI [0.01, 0.60]) und das Risk Ratio beträgt 0.12 (95 %-KI[0.02, 0.84]). Leseschwache Mädchen werden signifikant seltener als leseauffällig erkannt als leseschwache Jungen. Die Hypothese, dass leseschwache Mädchen seltener leseauffällig werden als leseschwache Jungen, muss differenziert betrachtet werden. In der ersten Klasse werden leseschwache Mädchen nicht signifikant seltener leseauffällig. Während die Erkennungsrate bei den

Mädchen von T2 bis T4 zwischen 6 und 14 Prozent liegt, steigt die Erkennungsrate bei den leseschwachen Jungen kontinuierlich an. In der ersten Klasse (zu T2 und T3) ist diese noch vergleichbar mit den Mädchen. Zu Beginn der zweiten Klasse weicht sie signifikant von denen der Mädchen ab. Zu diesem Zeitpunkt wird fast die Hälfte der leseschwachen Jungen lese-auffällig, während es bei den Mädchen nur sechs Prozent sind.

### 12.3.6 Hypothese 1.5

*Lehrkräfte bezeichnen Kinder mit einem höheren sozioökonomischen Status, die leseschwach sind, seltener als leseauffällig, als dies für leseschwache Kinder mit einem niedrigeren sozioökonomischen Status geschieht.*

Da nicht von allen Eltern Angaben zur Berufstätigkeit gemacht wurden, liegen für den sozio-ökonomischen Status fehlende Werte vor, die über das MICE-Verfahren imputiert werden. Insgesamt fehlen bei 84 Fällen die Angaben zur Berufstätigkeit. Dies entspricht 30 Prozent des Datensatzes. Es werden 40 Imputationen vorgenommen, da dies sicherstellt, dass der Powerverlust möglichst gering ist (Graham et al., 2007, S. 210; Mitra & Reiter, 2016, S. 197). Die zu betrachtenden Informationen liegen in einer 2x3 Kreuztabelle vor. Die Leseauffälligkeit ist eine dichotome Variable (leseauffällig, nicht leseauffällig). Der sozioökonomische Status hat drei Ausprägungen: *Erwerbslosigkeit bzw. unqualifizierte Tätigkeiten, Ausbildungsberufe sowie akademische Berufe bzw. Tätigkeiten mit hoher Verantwortung*. Mit den  $m$  Datensätzen wird jeweils der Fisher-Freeman-Halton-Test durchgeführt, der für Kreuztabellen mit mehr als vier Zellen ausgelegt ist (Freeman & Halton, 1951). Da es für das Poolen dieser Ergebnisse kein statistisches Paket gibt, wird auf die Median P Rule (MPR) zurückgegriffen (Eekhout, van de Wiel & Heymans, 2017, S. 3). Die MPR sieht vor, dass aus den  $m$   $p$ -Werten der Median gebildet wird, um dann eine Entscheidung für oder gegen die Beibehaltung der Nullhypothese zu fällen. Die MPR zeichnet sich dadurch aus, dass sie nur die  $p$ -Werte und keine weiteren Parameterschätzungen benötigt (Eekhout et al., 2017, S. 3).

Üblicherweise soll bei multipler Imputation die Outcome-Variable im Imputationsmodell enthalten sein. Für die Imputation kategorialer Daten, mit denen im Anschluss Signifikanztests über die MPR durchgeführt werden sollen, wird explizit empfohlen, die Outcome-Variable nicht in das Imputationsmodell aufzunehmen, da dies zu über-optimistischen Testergebnissen führt (Heymans & Eekhout, 2019, Kap. 3.1.4). Entsprechend dieser Vorgaben wird die multiple Imputation ohne die Outcome-Variable, in diesem Fall also die Leseauffälligkeit, durchgeführt. Die durchgeführten Leistungstests fließen in das Imputationsmodell ein. Ebenso werden die soziodemographischen Daten, die Schul- und die Klassenzugehörigkeit, Alter und Geschlecht der Kinder sowie die Tests zum Selbstkonzept und der FEES aufgenommen.

Tabelle 42

*Sozioökonomischer Status bei leseschwachen Kindern*

	T2 (n = 36)	T3 (n = 43)	T4 (n = 37)
Nicht erwerbstätig / unqualifizierte Tätigkeiten	1	0	0
davon leseauffällig	0	0	0
Ausbildungsberuf	15	15	18
davon leseauffällig	0	5	4
Akademiker / Beruf mit hoher Verantwortung	5	10	7
davon leseauffällig	1	2	3
Sozioökonomischer Status <i>nicht</i> angegeben	15	18	12
davon leseauffällig	1	4	3

Über 15 leseschwache Kinder liegen zu T2 keine Informationen zum sozioökonomischen Status vor (s. Tabelle 42). Der sozioökonomische Status für diese Kinder wird entsprechend imputiert. Nur zwei der leseschwachen Kinder wurden den Lehrkräften leseauffällig. Bei  $p = .19$  ( $m = 40$  Fisher-Freeman-Halton-Tests, Anwendung der MPR) liegt kein statistisch signifikanter Unterschied vor. Die Lehrkräfte haben entsprechend leseschwache Kinder mit einem höheren sozioökonomischen Status nicht seltener oder häufiger als leseauffällig bezeichnet. Zu T3 liegen für 25 von 43 leseschwachen Kindern Informationen über den sozioökonomischen Status vor (s. Tabelle 42). Für 18 Kinder, hiervon vier leseauffällige, fehlt diese Information und wird entsprechend imputiert. Mit  $p = .69$  ( $m = 40$  Fisher-Freeman-Halton-Tests, Anwendung der MPR) liegt kein statistisch signifikanter Unterschied vor. Für T4 liegen für 25 leseschwache Kinder Informationen über den sozioökonomischen Status vor. Für 12 Kinder fehlen diese Informationen. Mit  $p = .27$  ( $m = 40$  Fisher-Freeman-Halton-Tests, Anwendung der MPR) liegt kein statistisch signifikanter Unterschied vor.

Die Hypothese, dass leseschwache Kinder mit einem höheren sozioökonomischen Status seltener leseauffällig werden, muss zurückgewiesen werden.

#### 12.4 Fragestellung 2: Einfluss der Diagnose auf die Leseentwicklung

Dieser Abschnitt fokussiert auf den Entwicklungsverlauf von nach Lehrkräfteurteil auffällig gewordenen Kindern, unabhängig davon, ob auch Förderung einsetzt. Es werden hierfür jeweils auffällig gewordene Kinder mit Kindern verglichen, die nicht auffällig geworden sind. Ausgang der Betrachtung sind die Messzeitpunkte T2 (Mitte des ersten Schuljahres) und T3 (Ende des ersten Schuljahres). Endpunkt zur Prüfung, ob sich die Leseentwicklung unterscheidet, sind T6 (Ende des zweiten Schuljahres) und T7 (Ende des vierten Schuljahres).

Die aufgestellten Hypothesen differenzieren zwischen Auffälligkeiten in Bezug auf die Lesefähigkeit, den Bereich Deutsch und auf allgemein negative Auffälligkeiten. Der Bereich Deutsch beinhaltet neben Auffälligkeiten im Lesen auch solche im Schreiben sowie in sprachlichen Fähigkeiten. Die Leseentwicklung wird sowohl über den Lese-Lernzuwachs im Wortlesen erfasst wie auch durch die Leistungen in den Subtests Satzlesen und Textlesen, welche erst zu späteren Messzeitpunkten erhoben wurden.

### 12.4.1 Bestimmen des Lese-Lernzuwachses

Die Leseentwicklung wird sowohl auf der basalen wie auch auf der höheren Ebene des sinnentnehmenden Lesens betrachtet. Auf der basalen Ebene wird die Leseleistung im Wortlesen zu den späteren Messzeitpunkten erfasst. Die höhere Ebene wird durch die Leistungen in den Untertests Satzlesen und Textlesen berücksichtigt. Um zu betrachten, wie sich Kinder im Lesen entwickeln, wird der Zuwachs an korrekt erlesenen Wörtern im ELFE 1-6 berechnet. Hierzu wird die zum späteren Zeitpunkt erreichte Leseleistung in Relation zur Ausgangsleseleistung gesetzt. Der Zuwachs an gelesenen Wörtern stellt damit den Lese-Lernzuwachs dar. Das Ergebnis wird über die Blom-Transformation in z-Werte transformiert und anschließend in T-Werte übertragen. Auf diese Weise liegen vergleichbare Ergebnisse auch für die unterschiedlichen Messzeitpunkte vor.

Der Vorteil einer Beobachtungsstudie ist, dass man die Heterogenität im Feld, also in der natürlichen Umgebung der Klasse, berücksichtigen kann. Dadurch bedingt ergeben sich in den Klassen auch Ausreißerwerte in Form von sehr hohen oder sehr niedrigen Leistungen bei einzelnen Kindern. Eine verbindliche Vorgabe, wie mit Ausreißerwerten umzugehen ist, gibt es nicht (Goerke, 2009, S. 24). Generell werden in dieser Untersuchung keine Kinder aus Analysen ausgeschlossen, auch wenn sie Ausreißerwerte zeigen, da sie gleichzeitig die Heterogenität der Klassen widerspiegeln. Ein Ausschluss ist gleichbedeutend mit dem Verlust wertvoller Informationen für die Datenanalyse. Dieses Vorgehen stößt allerdings an seine Grenzen, wenn Deckeneffekte die statistische Auswertung negativ beeinflussen und so Ergebnisse verzerren können. Zur Auswertung der folgenden Hypothesen wird der Lernzuwachs der Kinder im Wortlesetest betrachtet. Die Kinder können in diesem Test in einer vorgegebenen Zeit maximal 72 Wörter korrekt erlesen. Der Lernzuwachs berechnet sich für jedes Kind aus der Summe der korrekt erlesenen Wörter zum betrachteten Endzeitpunkt (T6 oder T7) abzüglich der Anzahl der ursprünglich korrekt erlesenen Wörter (zu T2 oder T3). Kinder, die zu Beginn bereits eine weit überdurchschnittliche Leistung erzielen und viele Wörter korrekt erlesen haben, können deshalb nur einen geringen Lernzuwachs aufweisen. Um solche Deckeneffekte zu vermeiden, werden daher zunächst die positiven Ausreißerwerte betrachtet. Auf dieser Grundlage wird entschieden, ob einzelne Kinder aufgrund sehr hoher Ausgangsleistungen im Wortlesetest aus der weiteren Analyse herausgenommen werden. Hierfür werden die Ergebnisse in den Wortlesetests zu T2 und T3 auf Ausreißerwerte untersucht.

Insgesamt erzielen acht Kinder eine weit überdurchschnittliche Leseleistung zu T2 (s. Abbildung 8). Für vier Kinder liegt sogar ein extremer Ausreißerwert von mehr als 3 Standardabweichungen über dem Mittelwert vor, für die übrigen vier Kinder ein leichter Ausreißerwert mit einer Leistung von mehr als 1.5 Standardabweichungen über dem Mittelwert. Zu T3 ergibt sich nur für ein Kind ein extremer Ausreißerwert von mindestens 3 Standardabweichungen über dem Mittelwert. Fünf Kinder zeigen einen leichten Ausreißerwert, der mindestens 1.5 Standardabweichungen über dem Mittelwert liegt.

## 12 Ergebnisse

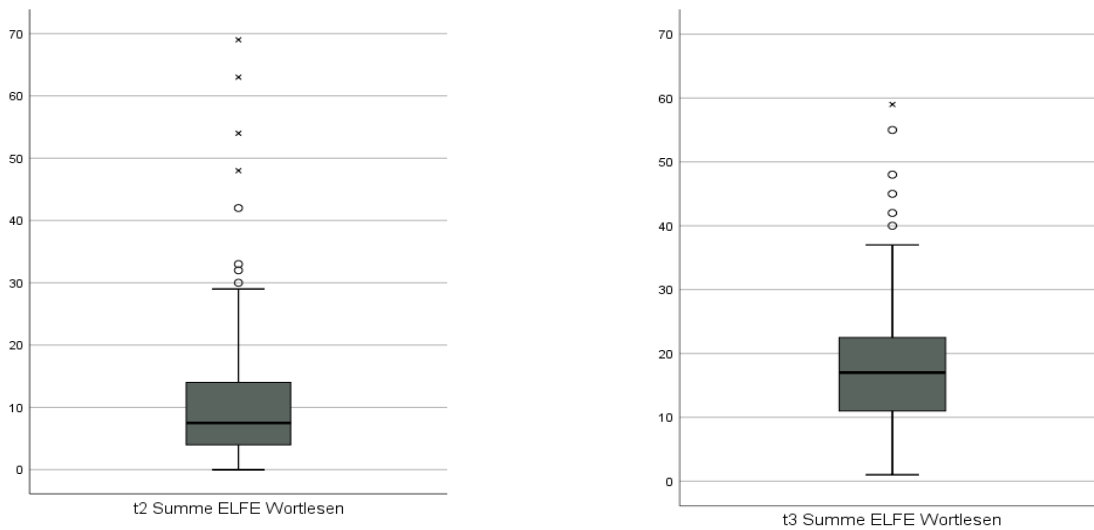


Abbildung 8. Ausreißerwerte im Wortlesetest von ELFE 1-6 (in T-Werten) zu T2 und T3. \* = extreme Ausreißerwerte von  $3 \text{ SD} \geq M$ ; o = Ausreißerwerte  $1.5 \text{ SD} \geq M$

Die Kinder mit extrem positiven Ausreißerwerten haben zu T2 zwischen 42 und 69 Wörter korrekt beantwortet, was jeweils mehr als die Hälfte der erreichbaren Punkte darstellt. Das Kind, welches 69 Wörter richtig erlesen hat, könnte somit einen maximalen Zugewinn von 3 Wörtern erreichen. Um eine möglichst große Vielfalt in den Informationen beizubehalten, werden nur die Kinder, welche zu T2 oder zu T3 einen extrem positiven Ausreißerwert aufweisen, aus der Analyse für diese Fragestellung ausgeschlossen. Das Kriterium von  $\geq 3 \text{ SD}$  über dem Mittelwert entspricht auch gängigem Vorgehen (Goerke, 2009, S. 23). Dies trifft auf insgesamt fünf Kinder zu.

### 12.4.2 Statistisches Matching über Propensity Scores

Es müssen jeweils Kontrollgruppen gefunden werden, die sich aus unauffälligen Schülerinnen und Schülern zusammensetzen, welche sich in weiteren bedeutsamen Faktoren aber nicht von den auffälligen Kindern unterscheiden. Wie in Kapitel 11.6.4 erläutert, geschieht dies für jeden Gruppenvergleich mithilfe des Matchings über das Propensity Score-Verfahren. Bevor das Verfahren durchgeführt werden kann, müssen zunächst die fehlenden Werte behandelt werden. Es werden nur die Daten der Kinder berücksichtigt, die zum Ausgangs- sowie zum Endzeitpunkt am Lesetest teilgenommen haben. Zwischen 41 und 54 Prozent der Daten enthalten mindestens einen fehlenden Wert (s. Tabelle 43). Der relativ hohe Anteil ergibt sich daraus, dass für zwei Klassen die Geburtsdaten fehlen. Ebenso ist der sozioökonomische Status in einigen Fällen unbekannt, da die Eltern hierzu keine Angabe gemacht haben. Einige Kinder fehlten zu einzelnen Terminen, z. B. aufgrund einer längeren Erkrankung, sodass auch bei diesen Skalen vereinzelt fehlende Werte vorhanden sind. Bei einer Rate von etwa 50 Prozent entsprechen  $m = 40$  Imputationen einem Powerverlust von unter einem Prozent (Graham et al., 2007, S. 212). Da eine Erhöhung der Imputationen auch bei

einer geringeren Rate an fehlenden Werten keinen Nachteil mit sich bringt, wird für alle vorhandenen Datensätze einheitlich  $m = 40$  festgelegt. Im Anschluss an die multiple Imputation kann das Matching durchgeführt werden. Es werden insgesamt vier Datensätze imputiert, die jeweils zwei relevante Messzeitpunkte (T2-T6, T2-T7, T3-T6, T3-T7) im Fokus haben.

Tabelle 43

*Anteil fehlender Werte nach betrachteten Messzeitpunkten*

Relevante Messzeitpunkte	T2-T6	T2-T7	T3-T6	T3-T7
Anteil fehlender Werte	51 %	41 %	54 %	51 %

Faktoren, die sich sowohl auf die Auffälligkeit wie auch auf die zu betrachtende Variable – hier entsprechend die Leseleistung – auswirken können, sollen in das Matching eingeschlossen werden. Alle vier Datensätze berücksichtigen als relevante Faktoren das Geschlecht, das Alter, das logische Denken, den sozioökonomischen Hintergrund, die Familiensprache sowie das Selbstkonzept im Fach Deutsch. Die Klassenzugehörigkeit wird ebenfalls als Kovariable aufgenommen. Als relevante Einflussfaktoren werden die Silbensegmentation und die Buchstabenkenntnis zu T2 berücksichtigt. Für Berechnungen mit T2 als Ausgangspunkt werden weiterhin die Laut-Buchstaben-Zuordnung zu T1 sowie das Schreiben nach Bildern zu T1 und das Wortlesen zu T2 in das Modell aufgenommen. Für T3 werden stattdessen die Laut-Buchstaben-Zuordnung zu T3, die Rechtschreibleistung im DERET 1-2 zu T3 sowie die Leseleistung im Wortlesen zu T2 und zu T3 berücksichtigt.

Der Propensity Score für das Matching wird durch eine logistische Regression auf die interessierende Variable berechnet. Diese Variable unterscheidet sich je nach Hypothese: Hypothesen 2.1 und 2.2 betrachten Auffälligkeiten im Bereich Deutsch. Hypothesen 2.3 und 2.4 fokussieren auf die Leseauffälligkeit und die letzten beiden Hypothesen 2.5 und 2.6 betrachten allgemein negative Auffälligkeiten. Entsprechend wird das Matching für die oben genannten vier Datensätze jeweils mit einer dieser drei Variablen als Outcome-Variable berechnet, um passende Kontrollgruppen zu finden. Es ergeben sich somit insgesamt 12 Datensätze.

Um die Qualität des Matchings beurteilen zu können, werden die standardisierten Mittelwertdifferenzen berichtet, die möglichst nahe null sein sollen. Außerdem werden die Ergebnisse der statistischen Mittelwertvergleiche und der Verteilungstests vor und nach dem Matching dargestellt. Nach dem Matching sollen keine statistisch signifikanten Unterschiede zwischen der Treatment- und der Kontrollgruppe vorliegen. Für metrische Variablen werden hierfür zweiseitige  $t$ -Tests verwendet. Um zu prüfen, ob kategoriale Verteilungen in der Kontroll- und der Treatmentgruppe signifikant voneinander abweichen, werden  $X^2$ -Tests durchgeführt. Die Ergebnisse für die imputierten Datensätze werden mit dem Paket `miceadds` (Robitzsch, Grund & Henke, 2019) zu einem Gesamtergebnis gepoolt. In der Verteilung der kategorialen Variablen existieren teilweise leere Zellen, wenn beispielsweise weder in der Treatment- noch in der Kontrollgruppe Kinder einer bestimmten Klasse gewählt wurden. Für diese Fälle kann  $X^2$  nicht berechnet werden. Die Variablen Klasse und Geschlecht enthalten allerdings keine

## 12 Ergebnisse

fehlenden Daten, sodass die Verteilung vor und nach dem Matching in allen 40 Datensätzen identisch ist. Bei Vorhandensein leerer Zellen kann daher jeweils ein einzelner Fisher-Freeman-Halton-Test angewendet werden.

Für die imputierten Daten kann dieser Test mit anschließendem Pooling über die Median-*P*-Regel nicht angewendet werden, da die Outcome-Variable Teil des Imputationsmodells war. Dies trifft auf die Variablen sozioökonomischer Status und Familiensprache zu. Für diese Fälle gibt es auch keine kompakt darstellbare gepoolte Verteilung. Die Berechnung erfolgt an diesen Stellen mit dem  $X^2$ -Test und anschließendem Pooling, wobei nur die Datensätze ohne leere Zellen in die Berechnung eingehen.

### 12.4.2.1 Kontrollgruppen für T2-T6

Für die Berechnung T2-T6 mit Leseauffälligkeit als interessierende Variable ergibt sich das beste Matching für die „nearest neighbour“-Methode mit Ratio 3:1 ohne Replacement mit zufälliger Reihenfolge.

Tabelle 44

*Mittelwerte vor und nach dem Matching für T2-T6 Leseauffälligkeit*

Variable		Prä-Test				Post-Test			
		<i>M</i>	<i>SD</i>	<i>p</i>	<i>t</i> ( <i>df</i> )	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t</i> ( <i>df</i> )
Wortlesen T2	KG	50.40	8.95	.14	1.469 (209.03)	47.80	8.24	.64	0.474 (44.12)
	TG	46.53	7.32			46.53	7.32		
Alter	KG	6.66	1.43	.30	1.032 (103.38)	6.49	0.38	.98	0.021 (44.12)
	TG	6.49	0.28			6.49	0.28		
Logisches Denken	KG	50.40	10.09	.61	0.511 (163.82)	51.76	9.86	.95	-0.068 (37.44)
	TG	51.99	9.36			51.99	9.36		
Silbensegmentation T2	KG	50.10	8.91	.78	0.276 (209.03)	49.67	8.90	.92	0.104 (44.12)
	TG	49.37	7.92			49.37	7.92		
Buchstabenkenntnis T2	KG	50.40	9.23	.84	0.199 (209.03)	48.99	9.25	.78	0.287 (44.12)
	TG	49.86	8.64			49.86	8.64		
Laut-Buchstaben-Zuordnung T1	KG	50.76	12.02	.48	0.705 (203.50)	48.90	11.33	.94	0.075 (38.63)
	TG	48.66	8.47			48.66	8.47		
Schreiben nach Bildern T1	KG	51.08	10.28	.27	1.110 (204.79)	48.89	9.77	.79	0.268 (39.86)
	TG	48.14	6.93			48.14	6.93		
Selbstkonzept Deutsch T2	KG	48.91	12.90	.55	-0.597 (198.92)	47.42	10.12	.32	-0.998 (43.36)
	TG	50.67	9.44			50.67	9.44		

*Anmerkungen.* Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe ( $n = 12$ ), KG = Kontrollgruppe ( $n_{Prä} = 201$ ,  $n_{Post} = 36$ ), \* $p < .05$ , \*\* $p < .01$ .

Die Treatmentgruppe besteht aus 12 Kindern, die zu T2 leseauffällig wurden. Die Kontrollgruppe mit Kindern, die nicht leseauffällig wurden, besteht ursprünglich aus 201 Kindern. Nach dem Matching befinden sich 36 Kinder in der Kontrollgruppe. Die größte standardisierte Mittelwertdifferenz der einzelnen Variablen liegt bei -0.173 für das Wortlesen zu T2. Alle anderen Werte liegen näher an null. Über alle Variablen verändert sich die standardisierte Mittelwertdifferenz durch das Matching von 1.340 auf 0.677.

## 12 Ergebnisse

Weder vor noch nach dem Matching gibt es signifikante Unterschiede zwischen den jeweiligen Gruppen (s. Tabelle 44). Insgesamt nähern sich für alle Variablen außer der Buchstabenkenntnis und dem Selbstkonzept in Deutsch die Mittelwerte der Kontrollgruppe an die Treatmentgruppe an.

Tabelle 45

### Unterschiede in den kategorialen Daten für T2-T6 Leseauffälligkeit

Variable	Prä-Test			Post-Test		
	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>
Klasse	.01*+	-	-	1.00+	-	-
Familiensprache	.45	0.804	2	.64	0.444	2
Sozioökonomischer Status	.40	0.910	2	.66++	0.415	2
Geschlecht	.83	0.044	1	.85	0.035	1

*Anmerkungen.* \**p* < .05, +Fisher-Freeman-Halton-Test, ++Berechnung mit 24 von 40 Datensätzen. Jeweils leere Zellen für „unqualifizierte Tätigkeiten / nicht erwerbstätig“ der beiden Gruppen Prä- und Post-Test.

Die Ergebnisse für die kategorialen Variablen werden in Tabelle 45 berichtet. Vor dem Matching unterscheidet sich die Verteilung über die Klassen hinweg. Nach dem Matching besteht dieser Unterschied nicht mehr.

Tabelle 46

### Mittelwerte vor und nach dem Matching für T2-T6 Bereich Deutsch

Variable		Prä-Test				Post-Test			
		<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>
Wortlesen T2	KG	50.73	8.88	.01*	2.788 (209.03)	44.89	7.83	.91	-0.112 (59.09)
	TG	45.12	7.50			45.12	7.50		
Alter	KG	6.67	1.38	.55	0.598 (77.94)	6.57	0.44	.85	-0.191 (59.09)
	TG	6.59	0.40			6.59	0.40		
Logisches Denken	KG	50.47	10.16	.79	0.266 (178.5)	48.83	10.11	.71	-0.375 (53.66)
	TG	49.84	9.39			49.84	9.39		
Silbensegmentation T2	KG	50.09	8.90	.88	0.148 (209.03)	49.08	7.80	.74	-0.327 (59.09)
	TG	49.79	8.49			49.79	8.49		
Buchstabenkenntnis T2	KG	50.78	9.02	.05	1.970 (209.03)	46.72	8.83	.98	0.028 (59.09)
	TG	46.65	10.02			46.65	10.02		
Laut-Buchstaben-Zuordnung T1	KG	51.07	11.56	.10	1.655 (182.61)	48.98	9.35	.45	0.755 (52.18)
	TG	47.18	9.68			47.18	9.68		
Schreiben nach Bildern T1	KG	51.37	10.18	.03*	2.227 (192.03)	46.77	7.63	.99	0.008 (51.69)
	TG	46.76	7.65			46.76	7.65		
Selbstkonzept Deutsch T2	KG	48.87	13.21	.91	-0.111 (189.11)	48.42	13.27	.81	-0.240 (48.07)
	TG	49.13	8.80			49.13	8.80		

*Anmerkungen.* Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe (*n* = 21), KG = Kontrollgruppe (*n*<sub>Prä</sub> = 192, *n*<sub>Post</sub> = 42), \**p* < .05, \*\**p* < .01.

Für die Berechnung von T2-T6 mit Auffälligkeiten im Bereich Deutsch ergibt sich das beste Matching für die „nearest neighbour“-Methode mit Ratio 2:1 ohne Replacement mit zufälliger Reihenfolge.

Die Treatmentgruppe besteht aus 21 Kindern, die zu T2 auffällig im Bereich Deutsch wurden. In der Kontrollgruppe befinden sich ursprünglich 192 Kinder. Nach dem Matching besteht

## 12 Ergebnisse

diese aus 42 Kindern. Mit 0.206 liegt die größte standardisierte Mittelwertdifferenz der einzelnen Variablen für die Buchstaben-Laut-Zuordnung zu T1 vor. Alle anderen Werte liegen entsprechend darunter. Über alle Variablen verringert sich die standardisierte Mittelwertdifferenz von 1.245 auf 0.472. Vor dem Matching ergeben sich für zwei Variablen signifikante Unterschiede: Sowohl in der Rechtschreibung wie auch im Wortlesen erzielt die Treatmentgruppe niedrigere Leistungen. Nach dem Matching gibt es keine signifikanten Unterschiede zwischen den Gruppen (s. Tabelle 46).

Tabelle 47

### Unterschiede in den kategorialen Daten für T2-T6 im Bereich Deutsch

Variable	Prä-Test			Post-Test		
	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>
Klasse	.03*	22.474	12	.94+	-	-
Familiensprache	.96	0.041	2	.88	0.123	2
Sozioökonomischer Status	.14	1.948	2	.65++	0.427	2
Geschlecht	.99	0.000	1	.93	0.008	1

*Anmerkungen.* \*  $p < .05$ , + = Fisher-Freeman-Halton-Test, ++ Berechnung mit 36 von 40 Datensätzen. Jeweils leere Zellen für „unqualifizierte Tätigkeiten / nicht erwerbstätig“ der beiden Gruppen Prä- und Post-Test.

Für die kategorialen Variablen ergibt sich im Prä-Test nur für die Verteilung der Klassen ein signifikanter Unterschied zwischen auffälligen und nicht auffälligen Kindern (s. Tabelle 47). Im Post-Test zeigen sich keine signifikanten Unterschiede.

Tabelle 48

### Mittelwerte vor und nach dem Matching für T2-T6 negative Auffälligkeiten

Variable		Prä-Test				Post-Test			
		<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>
Wortlesen T2	KG	51.08	8.60	.01*	2.796 (209.03)	46.84	8.41	.90	-0.126 (92.06)
	TG	47.07	9.27			47.07	9.27		
Alter	KG	6.65	1.31	.40	-0.850 (99.07)	6.65	0.56	.44	-0.772 (85.28)
	TG	6.73	0.62			6.73	0.62		
Logisches Denken	KG	50.62	10.03	.56	0.588 (196.41)	51.61	11.84	.39	0.873 (85.04)
	TG	49.66	10.30			49.66	10.30		
Silbensegmentation T2	KG	50.49	8.86	.19	1.313 (209.03)	49.91	8.95	.46	0.736 (92.06)
	TG	48.58	8.71			48.58	8.71		
Buchstabenkenntnis T2	KG	50.83	8.75	.17	1.3629 (209.03)	48.79	9.48	1.00	0.002 (92.06)
	TG	48.78	10.47			48.78	10.47		
Laut-Buchstaben-Zuordnung T1	KG	51.62	10.42	.01*	2.486 (176.23)	48.89	11.64	.50	0.681 (83.87)
	TG	47.45	12.22			47.45	12.22		
Schreiben nach Bildern T1	KG	51.56	9.57	.06	1.911 (181.49)	49.39	10.71	.70	0.386 (86.43)
	TG	48.68	9.76			48.68	9.76		
Selbstkonzept Deutsch T2	KG	48.98	12.62	.84	0.206 (183.69)	49.43	11.53	.71	0.373 (76.14)
	TG	48.63	10.74			48.63	10.74		

*Anmerkungen.* Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe ( $n = 48$ ), KG = Kontrollgruppe ( $n_{Prä} = 165$ ,  $n_{Post} = 48$ ), \* $p < .05$ , \*\* $p < .01$ .

Für den Datensatz T2-T6 mit allgemein negativen Auffälligkeiten ergibt sich das beste Matching für die „nearest-neighbour“-Methode mit einer 1:1-Ratio ohne Replacement mit zufälliger Reihenfolge.

Tabelle 49

*Unterschiede in den kategorialen Daten für T2-T6 negative Auffälligkeiten*

Variable	Prä-Test			Post-Test		
	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>
Klasse	.23	15.241	12	1.00+	-	-
Familiensprache	.50	0.691	2	.70	0.352	2
Sozioökonomischer Status	.48	0.742	2	.81	0.215	2
Geschlecht	.53	0.395	1	.83	0.047	1

Anmerkungen. +Fisher-Freeman-Halton-Test

In der Treatmentgruppe befinden sich 48 Kinder. Vor dem Matching besteht die Kontrollgruppe aus 165 Kindern, nach dem Matching ebenfalls aus 48. Die höchste standardisierte Mittelwertdifferenz liegt bei 0.219 für die Silbensegmentation zu T2. Alle anderen Werte liegen unter 0.2. Für das gesamte Matching sinkt die standardisierte Mittelwertdifferenz von 0.934 auf 0.257. Im Prä-Test ergeben sich zwei signifikante Unterschiede: Die Kinder der Kontrollgruppe erreichen eine signifikant bessere Leistung im Lesen und im Test zur Laut-Buchstaben-Zuordnung (s. Tabelle 48). Im Post-Test gibt es keine signifikanten Unterschiede zwischen der Treatment- und der Kontrollgruppe. Für die kategorialen Daten ergeben sich weder in den Prä-Tests noch in den Post-Tests signifikante Unterschiede (s. Tabelle 49).

## 12.4.2.2 Kontrollgruppen für T2-T7

Für die Daten von T2-T7 mit leseauffälligen Kindern in der Treatmentgruppe ergibt sich das beste Matching für das „nearest neighbour“-Vorgehen mit Ratio 2:1 ohne Replacement und mit zufälliger Reihenfolge. Die Treatmentgruppe besteht aus 9 Kindern. Vor dem Matching befinden sich in der Kontrollgruppe 130 Kinder. Nach dem Matching besteht die Kontrollgruppe aus 18 Kindern. Die größte standardisierte Mittelwertdifferenz nach dem Matching ergibt sich für die Variable logisches Denken mit einem Wert von 0.418. Insgesamt über alle Variablen gesehen verringert sich die standardisierte Mittelwertdifferenz von 1.551 auf 0.921. Weder im Prä-Test noch im Post-Test ergeben sich signifikante Unterschiede in den metrischen Variablen (s. Tabelle 50). Im logischen Denken und im Selbstkonzept Deutsch liegen die Werte nach dem Matching weiter auseinander als vor dem Matching. Die übrigen Werte der Kontrollgruppe liegen nach dem Matching näher an der Treatmentgruppe.

Tabelle 51 zeigt die Ergebnisse der Signifikanzprüfungen für die kategorialen Daten. Sowohl vor wie auch nach dem Matching zeigen sich keine signifikanten Unterschiede zwischen den leseauffälligen Kindern und den nicht leseauffälligen Kindern. Die Variable Familiensprache war im Post-Test für alle Imputationen identisch, sodass ein einfacher Fisher-Hamilton-Freeman-Test angewendet werden konnte.

## 12 Ergebnisse

Tabelle 50

*Mittelwerte vor und nach dem Matching für T2-T7 leseauffällig*

Variable		Prä-Test				Post-Test			
		M	SD	p	t (df)	M	SD	p	t (df)
Wortlesen T2	KG	50.05	8.97	.13	1.521 (144.04)	48.18	7.24	.36	0.935 (23.21)
	TG	45.41	7.29			45.41	7.29		
Alter	KG	6.65	0.43	.30	1.008 (143.92)	6.57	0.39	.62	0.509 (23.21)
	TG	6.50	0.27			6.50	0.27		
Logisches Denken	KG	51.13	11.06	.83	-0.215 (141.82)	48.62	9.60	.39	-0.877 (23.21)
	TG	51.86	7.76			51.86	7.76		
Silbensegmentation T2	KG	50.65	9.01	.50	0.683 (144.04)	47.69	6.85	.77	-0.292 (23.21)
	TG	48.55	7.74			48.55	7.74		
Buchstabenkenntnis T2	KG	50.27	9.20	.79	0.388 (144.04)	48.41	9.30	.85	-0.191 (23.21)
	TG	49.06	6.17			49.06	6.17		
Laut-Buchstaben-Zuordnung T1	KG	51.40	10.14	.44	0.783 (142.03)	49.39	9.59	.90	0.129 (21.82)
	TG	48.93	7.33			48.93	7.33		
Schreiben nach Bildern T1	KG	51.51	9.75	.22	1.235 (142.12)	47.78	8.21	.99	0.008 (21.79)
	TG	47.76	6.78			47.76	6.78		
Selbstkonzept Deutsch T2	KG	49.13	10.37	.17	-1.391 (141.32)	46.61	8.40	.06	-2.008 (23.21)
	TG	53.52	8.50			53.52	8.50		

*Anmerkungen.* Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe ( $n = 9$ ), KG = Kontrollgruppe ( $n_{Prä} = 130$ ,  $n_{Post} = 18$ ), \* $p < .05$ , \*\* $p < .01$ .

Für Kinder, die im Bereich Deutsch auffällig wurden, ergibt sich das beste Matching für das „nearest-neighbour“-Verfahren mit Replacement und Ratio 5:1. In der Treatmentgruppe befinden sich 16 Kinder. Vor dem Matching besteht die Kontrollgruppe aus 132 Kindern. Nach dem Matching verbleiben 34 Kinder in der Kontrollgruppe. Die größte standardisierte Mittelwertdifferenz nach dem Matching mit 0.136 ergibt sich für eine Dummyvariable des sozioökonomischen Status (Ausbildungsberuf: Ja / Nein). Alle übrigen Mittelwertdifferenzen liegen bei  $< 0.10$ . Die standardisierte Mittelwertdifferenz über alle Variablen sinkt von 1.172 auf 0.219.

Tabelle 51

*Unterschiede in den kategorialen Daten für T2-T7 leseauffällig*

Variable	Prä-Test			Post-Test		
	p	F / X <sup>2</sup>	df	p	F / X <sup>2</sup>	df
Klasse	.11+	-	-	1.00+	-	-
Familiensprache	.21	1.569	2	1.00+	-	-
Sozioökonomischer Status	.69	0.378	2	.69++	0.372	2
Geschlecht	.75	0.098	1	.88	0.024	1

*Anmerkungen.* +Fisher-Freeman-Halton-Test, ++Berechnung mit 5 von 40 imputierten Datensätzen. Jeweils leere Zellen für „unqualifizierte Tätigkeiten / nicht erwerbstätig“ der beiden Gruppen Prä- und Post-Test.

Vor dem Matching zeigt sich bei zwei Variablen ein signifikanter Unterschied zwischen der Kontrollgruppe und der Treatmentgruppe (s. Tabelle 52): Die Leistung im Wortlesen und im Schreiben nach Bildern zu T1 liegt in der Kontrollgruppe signifikant über der Leistung der Kinder der Treatmentgruppe. Im Post-Test gibt es keine signifikanten Unterschiede.

## 12 Ergebnisse

Tabelle 52

### Mittelwerte vor und nach dem Matching für T2-T7 Bereich Deutsch

Variable		Prä-Test				Post-Test			
		M	SD	p	t (df)	M	SD	p	t (df)
Wortlesen T2	KG	50.44	8.85	.01*	2.685 (144.04)	44.05	7.51	.94	-0.081 (46.12)
	TG	44.23	7.70			44.23	7.70		
Alter	KG	6.64	0.43	.77	0.291 (143.81)	6.63	0.50	.90	0.127 (45.42)
	TG	6.61	0.41			6.61	0.41		
Logisches Denken	KG	51.54	11.40	.28	1.092 (138.79)	48.61	10.30	.99	-0.015 (40.79)
	TG	48.65	8.13			48.65	8.13		
Silbensegmentation T2	KG	50.77	8.98	.34	0.956 (144.04)	48.72	8.14	.93	0.088 (46.12)
	TG	48.50	8.50			48.50	8.50		
Buchstabenkenntnis T2	KG	50.59	9.02	.13	1.527 (144.04)	46.45	9.77	.86	-0.177 (46.12)
	TG	46.95	8.73			46.95	8.73		
Laut-Buchstaben-Zuordnung T1	KG	51.88	10.34	.05	1.990 (139.24)	48.32	9.59	.63	0.490 (39.60)
	TG	47.06	7.89			47.06	7.89		
Schreiben nach Bildern T1	KG	51.83	9.73	.02*	2.418 (140.54)	46.75	7.73	.81	0.243 (40.29)
	TG	46.24	6.10			46.24	6.10		
Selbstkonzept Deutsch T2	KG	48.89	8.67	.70	0.389 (122.55)	48.74	9.06	.78	0.275 (41.09)
	TG	47.98	8.77			47.98	8.77		

Anmerkungen. Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe ( $n = 16$ ), KG = Kontrollgruppe ( $n_{Prä} = 132$ ,  $n_{Post} = 34$ ), \* $p < .05$ , \*\* $p < .01$ .

Für die kategorialen Variablen ergeben sich weder vor noch nach dem Matching signifikante Unterschiede (s. Tabelle 53).

Tabelle 53

### Unterschiede in den kategorialen Daten für T2-T7 auffällig im Bereich Deutsch

Variable	Prä-Test			Post-Test		
	p	F / X <sup>2</sup>	df	p	F / X <sup>2</sup>	df
Klasse	.27+	-	-	.98+	-	-
Familiensprache	.65	0.428	2	.56	0.577	2
Sozioökonomischer Status	.57	0.558	2	.77	0.265	2
Geschlecht	1.00	0.000	1	.79	0.069	1

Anmerkungen. +Fisher-Freeman-Halton-Test

Für Kinder, die allgemein negativ auffallen, ergibt sich das beste Matching für die „nearest-neighbour“-Methode mit Replacement mit Ratio 4:1. In der Treatmentgruppe befinden sich 29 Kinder. Vor dem Matching besteht die Kontrollgruppe aus 119 Kindern, nach dem Matching aus 51. Die größte standardisierte Mittelwertdifferenz liegt für die einzelnen Variablen bei 0.196 für das Selbstkonzept Deutsch. Vor dem Matching liegt die standardisierte Mittelwertdifferenz für alle Variablen bei 1.135. Durch das Matching sinkt diese auf 0.123.

## 12 Ergebnisse

Tabelle 54

Mittelwerte vor und nach dem Matching für T2-T7 negative Auffälligkeiten

Variable		Prä-Test				Post-Test			
		M	SD	p	t (df)	M	SD	p	t (df)
Wortlesen T2	KG	50.61	9.06	.02*	2.348 (144.04)	49.64	7.80	.07	1.840 (76.07)
	TG	46.33	7.55			46.33	7.55		
Alter	KG	6.63	0.41	.51	-0.664 (143.54)	6.66	0.40	.84	-0.206 (74.92)
	TG	6.68	0.49			6.68	0.49		
Logisches Denken	KG	51.56	11.15	.24	1.171 (134.32)	52.32	10.75	.19	1.338 (70.76)
	TG	49.11	10.15			49.11	10.15		
Silbensegmentation T2	KG	51.06	8.82	.14	1.500 (144.04)	48.48	9.07	.93	0.084 (76.07)
	TG	48.30	9.22			48.30	9.22		
Buchstabenkenntnis T2	KG	50.61	8.90	.27	1.119 (144.04)	50.11	9.49	.47	0.729 (76.07)
	TG	48.52	9.19			48.52	9.19		
Laut-Buchstaben-Zuordnung T1	KG	52.18	9.22	.01*	2.531 (131.88)	50.60	8.79	.13	1.529 (65.84)
	TG	47.34	10.08			47.34	10.08		
Schreiben nach Bildern T1	KG	52.12	9.58	.03*	2.192 (138.08)	49.38	8.42	.48	0.714 (71.43)
	TG	48.09	7.74			48.09	7.74		
Selbstkonzept Deutsch T2	KG	48.46	8.74	.31	-1.011 (134.99)	48.64	8.48	.41	-0.828 (72.27)
	TG	50.25	8.26			50.25	8.26		

Anmerkungen. Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe ( $n = 29$ ), KG = Kontrollgruppe ( $n_{Prä} = 119$ ,  $n_{Post} = 51$ ), \* $p < .05$ , \*\* $p < .01$ .

Tabelle 54 zeigt die Ergebnisse der Signifikanztests für die metrischen Variablen. Vor dem Matching gibt es bei drei Variablen signifikante Unterschiede zwischen der Treatment- und der Kontrollgruppe: Die Kinder der Kontrollgruppe erzielen signifikant höhere Leistungen im Wortlesen, in der Laut-Buchstaben-Zuordnung und im Schreiben nach Bildern. Nach dem Matching unterscheiden sich die Kontroll- und die Treatmentgruppe in keiner Variable signifikant voneinander.

Tabelle 55

Unterschiede in den kategorialen Daten für T2-T7 negative Auffälligkeiten

Variable	Prä-Test			Post-Test		
	p	F / X <sup>2</sup>	df	p	F / X <sup>2</sup>	df
Klasse	.26+	-	-	.99+	-	-
Familiensprache	.94	0.063	2	.37	0.982	2
Sozioökonomischer Status	.61	0.487	2	.69++	0.377	2
Geschlecht	1.00	0.000	1	1.00	0.000	1

Anmerkungen. +Fisher-Freeman-Halton-Test. ++Berechnung mit 33 von 40 imputierten Datensätzen. Jeweils leere Zellen für „unqualifizierte Tätigkeiten / nicht erwerbstätig“ der beiden Gruppen Prä- und Post-Test.

Für die kategorialen Variablen ergeben sich weder vor noch nach dem Matching signifikante Unterschiede in den Verteilungen zwischen der Kontroll- und der Treatmentgruppe (s. Tabelle 55).

## 12 Ergebnisse

### 12.4.2.3 Kontrollgruppen für T3-T6

Für die Daten von T3-T6 mit 15 leseauffälligen Kindern in der Treatmentgruppe ergibt sich das beste Matching für die „nearest neighbour“-Methode mit Ratio 1:1 ohne Replacement und mit zufälliger Reihenfolge. Vor dem Matching besteht die Kontrollgruppe aus 207 Kindern, danach aus 15 Kindern. Die höchste standardisierte Mittelwertdifferenz ergibt sich für die Variable des Rechtschreibtests DERET 1-2 mit -0.436. Für alle Variablen sinkt die standardisierte Mittelwertdifferenz von 1.649 auf 0.456. Bei fünf der neun kontinuierlichen Variablen zeigen sich vor dem Matching statistisch signifikante Unterschiede (s. Tabelle 56). In allen Fällen erreichen die Kinder der Treatmentgruppe signifikant niedrigere Leistungen. Nach dem Matching gibt es zwischen der Treatment- und der Kontrollgruppe keine signifikanten Unterschiede mehr.

Tabelle 56

*Mittelwerte vor und nach dem Matching für T3-T6 leseauffällig*

Variable		Prä-Test				Post-Test			
		M	SD	p	t (df)	M	SD	p	t (df)
Wortlesen T2	KG	50.60	14.12	.00**	3.450 (167.15)	41.42	8.56	.97	-0.044 (22.07)
	TG	41.54	7.46			41.54	7.46		
Wortlesen T3	KG	50.91	8.71	.00**	5.230 (218.03)	38.44	5.85	.75	-0.324 (26.19)
	TG	39.04	4.03			39.04	4.03		
Alter	KG	6.68	1.59	.22	-1.235 (92.11)	6.74	0.59	.49	-0.708 (24.04)
	TG	6.89	0.57			6.89	0.57		
Logisches Denken	KG	50.57	10.21	.16	1.394 (216.65)	46.58	9.65	.92	-0.100 (26.194)
	TG	46.93	9.03			46.93	9.03		
Silbensegmentation T2	KG	49.84	11.70	.91	-0.109 (182.40)	50.75	10.79	.86	0.173 (24.30)
	TG	50.12	9.73			50.12	9.73		
Buchstabenkenntnis T2	KG	50.85	12.81	.00**	3.040 (194.56)	43.66	11.02	.84	0.202 (24.40)
	TG	42.93	9.16			42.93	9.16		
Laut-Buchstaben-Zuordnung T3	KG	50.62	8.34	.00**	3.883 (218.03)	43.55	7.54	.60	0.531 (26.19)
	TG	41.84	9.95			41.84	9.95		
DERET T3	KG	51.13	8.98	.00**	5.622 (218.03)	40.25	6.45	.35	0.957 (26.19)
	TG	37.72	7.95			37.72	7.95		
Selbstkonzept Deutsch T2	KG	49.22	15.06	.36	0.922 (176.04)	48.89	9.97	.56	0.595 (24.55)
	TG	46.89	8.86			46.89	8.86		

*Anmerkungen.* Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe ( $n = 15$ ), KG = Kontrollgruppe ( $n_{\text{Prä}} = 207$ ,  $n_{\text{Post}} = 15$ ), \* $p < .05$ , \*\* $p < .01$ .

Für die kategorialen Variablen ergeben sich weder vor dem Matching noch nach dem Matching statistisch signifikante Unterschiede in den Verteilungen (s. Tabelle 57).

Für die Treatmentgruppe mit Kindern, die im Bereich Deutsch auffällig wurden, ergibt sich das beste Matching für das „nearest neighbour“-Vorgehen bei einer Ratio von 3:1 mit Replacement. Die Treatmentgruppe besteht aus 22 Kindern. Die Kontrollgruppe besteht vor dem Matching aus 200 Kindern. Nach dem Matching befinden sich noch 26 Kinder in der Kontrollgruppe.

## 12 Ergebnisse

Tabelle 57

### Unterschiede in den kategorialen Daten für T3-T6 Leseauffälligkeit

Variable	<i>p</i>	Prä-Test		<i>p</i>	Post-Test	
		<i>F</i> / <i>X</i> <sup>2</sup>	<i>df</i>		<i>F</i> / <i>X</i> <sup>2</sup>	<i>df</i>
Klasse	.22	15.393	12	1.00+	-	-
Familiensprache	.70	0.355	2	1.00	-0.025	2
Sozioökonomischer Status	.16	1.835	2	.69++	0.377	2
Geschlecht	.33	0.935	1	1.00	0.00	1

*Anmerkungen.* +Fisher-Freeman-Halton-Test. ++Berechnung mit 24 von 40 imputierten Datensätzen. Jeweils leere Zellen für „unqualifizierte Tätigkeiten / nicht erwerbstätig“ der beiden Gruppen Prä- und Post-Test.

Die höchste standardisierte Mittelwertdifferenz findet sich bei der Variable Wortlesen zu T3 mit einem Wert von 0.299. Alle übrigen Differenzen liegen unter 0.200. Für alle Variablen verringert sich die standardisierte Mittelwertdifferenz von 1.546 auf 0.011.

Tabelle 58

### Mittelwerte vor und nach dem Matching für T3-T6 Bereich Deutsch

Variable		Prä-Test				Post-Test			
		<i>M</i>	<i>SD</i>	<i>p</i>	<i>t</i> ( <i>df</i> )	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t</i> ( <i>df</i> )
Wortlesen T2	KG	50.72	13.72	.00**	3.271 (158.40)	44.53	7.67	.63	0.480 (39.47)
	TG	43.38	8.63			43.38	8.63		
Wortlesen T3	KG	50.83	8.77	.00**	3.685 (218.03)	44.40	9.58	.76	0.306 (44.12)
	TG	43.59	8.43			43.59	8.43		
Alter	KG	6.70	1.60	.31	-1.0164 (85.31)	6.72	0.86	.45	-0.765 (34.74)
	TG	6.85	0.66			6.85	0.66		
Logisches Denken	KG	50.77	10.34	.02*	2.3155 (215.07)	47.28	9.50	.57	0.578 (44.12)
	TG	45.68	9.61			45.68	9.61		
Silbensegmentation T2	KG	49.84	12.23	.97	0.033 (189.56)	51.96	8.86	.40	0.853 (42.11)
	TG	49.77	8.78			49.77	8.78		
Buchstabenkenntnis T2	KG	50.94	12.91	.00**	3.000 (185.59)	45.17	10.72	.76	0.301 (41.83)
	TG	44.29	9.04			44.29	9.04		
Laut-Buchstaben-Zuordnung T3	KG	50.55	8.32	.00**	2.752 (218.03)	45.60	8.66	.90	0.126 (44.12)
	TG	45.24	10.84			45.24	10.84		
DERET T3	KG	51.18	8.90	.00**	4.753 (218.03)	44.15	9.06	.35	0.938 (44.12)
	TG	41.49	10.64			41.49	10.64		
Selbstkonzept Deutsch T2	KG	49.31	13.49	.46	0.739 (132.66)	46.6	9.30	.70	-0.387 (40.57)
	TG	47.63	9.35			47.63	9.35		

*Anmerkungen.* Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe (*n* = 22), KG = Kontrollgruppe (*n*<sub>Prä</sub> = 200, *n*<sub>Post</sub> = 26), \**p* < .05, \*\**p* < .01.

Vor dem Matching ergeben sich für sechs der neun metrischen Variablen statistisch signifikante Unterschiede (s. Tabelle 58). Die Kinder der Treatmentgruppe erzielen jeweils niedrigere Leistungen. Nach dem Matching gibt es keine statistisch signifikanten Unterschiede mehr zwischen der Treatment- und der Kontrollgruppe.

Tabelle 59 zeigt die Ergebnisse der Signifikanzprüfungen für die kategorialen Daten. Vor dem Matching gibt es einen signifikanten Unterschied in der Verteilung des sozioökonomischen

## 12 Ergebnisse

Status in der Treatment- und der Kontrollgruppe. Nach dem Matching finden sich keine signifikanten Unterschiede mehr.

Tabelle 59

### Unterschiede in den kategorialen Daten für T3-T6 Bereich Deutsch

Variable	Prä-Test			Post-Test		
	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>
Klasse	.35	13.301	12	.99+	-	-
Familiensprache	.06	2.752	2	.70	0.350	2
Sozioökonomischer Status	.03*	3.381	2	.66	0.422	2
Geschlecht	.22	1.5292	1	1.00	0.000	1

Anmerkungen \**p* < .05. +Fisher-Freeman-Halton-Test.

Für die Treatmentgruppe mit allgemein negativ auffälligen Kindern ergibt sich das beste Matching für die „nearest-neighbour“-Methode mit Ratio 4:1 mit Replacement. Die höchste standardisierte Mittelwertdifferenz findet sich bei der Variable Selbstkonzept Deutsch mit einem Wert von 0.154. Insgesamt verringert sich die standardisierte Mittelwertdifferenz über alle Variablen von 1.384 auf 0.053. Die Treatmentgruppe besteht aus 53 Kindern. In der Kontrollgruppe befinden sich vor dem Matching 169 Kinder und nach dem Matching 66 Kinder.

Tabelle 60

### Mittelwerte vor und nach dem Matching für T3-T6 allgemein negative Auffälligkeiten

Variable	Prä-Test				Post-Test				
	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>	
Wortlesen T2	KG	51.16	13.19	.01*	2.849 (114.13)	48.51	10.90	.25	1.160 (78.51)
	TG	46.30	10.47			46.30	10.47		
Wortlesen T3	KG	51.45	8.24	.00**	4.132 (218.03)	48.14	8.00	.16	1.418 (115.05)
	TG	45.81	9.93			45.81	9.93		
Alter	KG	6.64	1.47	.31	-1.018 (51.56)	6.72	1.06	.77	-0.297 (40.99)
	TG	6.76	0.78			6.76	0.78		
Logisches Denken	KG	50.87	10.59	.12	1.563 (211.17)	48.59	10.60	.93	0.083 (110.77)
	TG	48.44	8.63			48.44	8.63		
Silbensegmentation T2	KG	50.28	11.00	.32	0.993 (166.89)	48.90	10.12	.93	0.0847 (94.72)
	TG	48.75	10.02			48.75	10.02		
Buchstabenkenntnis T2	KG	51.22	12.17	.00**	2.869 (160.18)	48.60	11.32	.29	1.056 (98.09)
	TG	46.57	10.91			46.57	10.91		
Laut-Buchstaben-Zuordnung T3	KG	51.28	8.26	.00**	3.930 (218.03)	48.23	9.10	.19	1.304 (115.05)
	TG	46.05	9.02			46.05	9.02		
DERET T3	KG	51.90	8.48	.00**	4.941 (218.03)	48.11	8.25	.06	1.874 (115.05)
	TG	44.86	10.69			44.86	10.69		
Selbstkonzept Deutsch T2	KG	49.31	13.65	.55	0.592 (104.31)	49.06	12.95	.71	0.372 (65.11)
	TG	48.31	11.87			48.31	11.87		

Anmerkungen. Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe (*n* = 53), KG = Kontrollgruppe (*n*<sub>Prä</sub> = 169, *n*<sub>Post</sub> = 66), \**p* < .05, \*\**p* < .01.

Es zeigen sich vor dem Matching bei fünf der neun metrischen Variablen statistisch signifikante Unterschiede (s. Tabelle 60).

## 12 Ergebnisse

Tabelle 61

*Unterschiede in den kategorialen Daten für T3-T6 allgemein negative Auffälligkeiten*

Variable	Prä-Test			Post-Test		
	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>
Klasse	.33	13.538	12	.70+	-	-
Familiensprache	.16	1.836	2	.70	0.362	2
Sozioökonomischer Status	.27	1.315	2	.73	0.314	2
Geschlecht	.00**	9.585	1	.11	2.514	1

Anmerkungen. \**p* < .05, \*\**p* < .01. +Fisher-Freeman-Halton-Test.

Die Kinder der Treatmentgruppe haben jeweils niedrigere Leistungen. Nach dem Matching ergeben sich zwischen der Kontrollgruppe und der Treatmentgruppe keine signifikanten Unterschiede mehr. Tabelle 61 ist zu entnehmen, dass sich nur die Verteilung der Geschlechter in den beiden Gruppen vor dem Matching signifikant unterscheidet. Nach dem Matching gibt es keine signifikanten Unterschiede zwischen der Treatment- und der Kontrollgruppe.

### 12.4.2.4 Kontrollgruppen für T3-T7

Für die Treatmentgruppe der leseauffälligen Kinder für T3-T7 ergibt sich das beste Matching für die „nearest neighbour“-Methode mit Ratio 1:1 ohne Replacement mit zufälliger Reihenfolge.

Tabelle 62

*Mittelwertunterschiede vor und nach dem Matching für T3-T7 Leseauffälligkeit*

Variable		Prä-Test				Post-Test			
		<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>
Wortlesen T2	KG	50.52	14.54	.02*	2.285 (103.12)	40.43	9.51	.67	-0.434 (11.79)
	TG	42.38	9.61			42.38	9.61		
Wortlesen T3	KG	50.54	8.12	.00**	3.798 (147.04)	40.93	7.44	.79	0.274 (14.32)
	TG	40.18	3.43			40.18	3.43		
Alter	KG	6.61	0.41	.01*	-2.577 (147.04)	6.53	0.45	.06	-2.066 (14.32)
	TG	6.97	0.47			6.97	0.47		
Logisches Denken	KG	51.53	10.27	.70	0.384 (145.86)	46.68	11.65	.47	-0.747 (14.32)
	TG	50.26	8.40			50.26	8.40		
Silbensegmentation T2	KG	50.09	13.37	.32	-0.989 (113.35)	51.19	8.20	.60	-0.542 (12.49)
	TG	53.53	10.01			53.53	10.01		
Buchstabenkenntnis T2	KG	50.49	13.28	.05	1.958 (101.69)	40.04	9.05	.05	-0.696 (12.31)
	TG	43.31	10.81			43.31	10.81		
Laut-Buchstaben-Zuordnung T3	KG	51.55	7.99	.00**	3.425 (147.04)	42.41	6.54	.92	0.106 (14.32)
	TG	41.98	10.33			41.98	10.33		
DERET T3	KG	52.38	9.08	.00**	5.185 (147.04)	39.99	4.99	.20	1.356 (14.32)
	TG	36.43	6.08			36.43	6.08		
Selbstkonzept Deutsch T2	KG	48.44	14.08	.33	0.970 (118.52)	50.49	7.93	.25	1.217 (12.69)
	TG	45.27	10.16			45.27	10.16		

Anmerkungen. Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe (*n* = 9), KG = Kontrollgruppe (*n*<sub>Prä</sub> = 143, *n*<sub>Post</sub> = 9), \**p* < .05, \*\**p* < .01.

Die Treatmentgruppe besteht aus 9 Kindern. In der Kontrollgruppe befinden sich vor dem Matching 143 Kinder und nach dem Matching ebenfalls 9 Kinder. Die höchste standardisierte

## 12 Ergebnisse

Mittelwertdifferenz zeigt sich mit einem Wert von 0.954 für die Variable Alter. Insgesamt kann die standardisierte Mittelwertdifferenz über alle Variablen mit dem Matching nur von 41.885 auf 41.549 gesenkt werden. Tabelle 62 können die Mittelwerte vor und nach dem Matching für die metrischen Variablen entnommen werden. Für fünf von neun Variablen ergeben sich vor dem Matching statistisch signifikante Unterschiede. Die Kinder der Treatmentgruppe sind signifikant älter als die Kinder der Kontrollgruppe. Ebenso bestehen signifikante Unterschiede in den beiden Wortlesetests, in der Laut-Buchstaben-Zuordnung und der Rechtschreibung. Die Kinder der Treatmentgruppe erzielen signifikant niedrigere Leistungen. Nach dem Matching gibt es keine signifikanten Unterschiede.

Tabelle 63

### Unterschiede in den kategorialen Daten für T3-T7 Leseauffälligkeit

Variable	Prä-Test			Post-Test		
	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>
Klasse	.57+	-	-	1.00+	-	-
Familiensprache	.84	0.174	2	.77	0.261	2
Sozioökonomischer Status	.48	0.728	2	.75++	0.294	2
Geschlecht	.32	0.987	1	1.00	0.000	1

Anmerkungen. \**p* < .05, \*\**p* < .01. + Fisher-Freeman-Halton-Test. ++Berechnung mit 25 von 40 imputierten Datensätzen. Jeweils leere Zellen für „unqualifizierte Tätigkeiten / nicht erwerbstätig“ der beiden Gruppen Prä- und Post-Test.

Für die kategorialen Variablen ergeben sich weder vor noch nach dem Matching statistisch signifikante Unterschiede (s. Tabelle 63).

Tabelle 64

### Mittelwertunterschiede vor und nach dem Matching für T3-T7 Bereich Deutsch

Variable		Prä-Test				Post-Test			
		<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>
Wortlesen T2	KG	50.18	14.61	.02*	2.367 (113.02)	44.69	12.91	.63	0.489 (23.18)
	TG	43.02	8.33			43.02	8.33		
Wortlesen T3	KG	50.47	8.18	.01*	2.812 (147.04)	44.31	7.58	.79	0.263 (32.16)
	TG	43.62	7.12			43.62	7.12		
Alter	KG	6.61	0.41	.06	-1.886 (147.04)	6.75	0.53	.59	-0.541 (32.16)
	TG	6.85	0.47			6.85	0.47		
Logisches Denken	KG	51.70	10.29	.30	1.049 (145.47)	47.02	10.04	.62	-0.502 (30.77)
	TG	48.66	8.11			48.66	8.11		
Silbensegmentation T2	KG	50.24	12.91	.39	-0.867 (121.89)	52.78	11.88	.99	-0.012 (26.26)
	TG	52.82	8.98			52.82	8.98		
Buchstabenkenntnis T2	KG	50.78	13.60	.02*	2.379 (125.14)	46.56	13.84	.47	0.730 (26.72)
	TG	43.66	9.21			43.66	9.21		
Laut-Buchstaben-Zuordnung T3	KG	51.52	7.95	.01*	2.725 (147.04)	47.88	6.82	.31	1.032 (32.16)
	TG	44.76	11.35			44.76	11.35		
DERET T3	KG	52.42	9.01	.00**	4.561 (147.04)	43.91	5.20	.13	1.564 (32.16)
	TG	39.92	10.18			39.92	10.18		
Selbstkonzept Deutsch T2	KG	48.47	13.50	.26	1.123 (91.34)	48.32	11.15	.35	0.950 (25.45)
	TG	45.02	11.39			45.02	11.39		

Anmerkungen. Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe (*n* = 12), KG = Kontrollgruppe (*n*<sub>Prä</sub> = 139, *n*<sub>Post</sub> = 24), \**p* < .05, \*\**p* < .01.

Für die Kinder, die im Bereich Deutsch auffällig geworden sind, ergibt sich das beste Matching für die „nearest neighbour“-Methode mit Ratio 2:1 ohne Replacement mit zufälliger Reihenfolge. Die Treatmentgruppe besteht aus 12 Kindern. In der Kontrollgruppe befinden sich vor dem Matching 139 Kinder und nach dem Matching 24 Kinder. Die höchste standardisierte Mittelwertdifferenz zeigt sich mit einem Wert von -0.391 in der Rechtschreibleistung zu T3. Für alle Variablen sinkt die standardisierte Mittelwertdifferenz von 2.074 auf 1.635.

Für fünf der neun metrischen Variablen ergibt sich ein statistisch signifikanter Unterschied vor dem Matching (s. Tabelle 64). Die Leistungen sind in der Treatmentgruppe signifikant niedriger als in der Kontrollgruppe. Nach dem Matching finden sich keine statistisch signifikanten Unterschiede mehr. Für die kategorialen Daten gibt es weder vor noch nach dem Matching signifikante Unterschiede (s. Tabelle 65).

Tabelle 65

*Unterschiede in den kategorialen Daten für T3-T7 Bereich Deutsch*

Variable	Prä-Test			Post-Test		
	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>
Klasse	.59+	-	-	1.00+	-	-
Familiensprache	.20	1.615	2	.66	0.422	2
Sozioökonomischer Status	.29	1.247	2	.66++	0.412	2
Geschlecht	.10	2.773	1	.68	0.167	1

*Anmerkungen.* \**p* < .05, \*\**p* < .01. +Fisher-Freeman-Halton-Test. ++Berechnung mit 34 von 40 imputierten Datensätzen. Jeweils leere Zellen für „unqualifizierte Tätigkeiten / nicht erwerbstätig“ der beiden Gruppen Prä- und Post-Test.

Für die Kinder, die allgemein negativ auffällig geworden sind, ergibt sich das beste Matching für die „nearest-neighbour“-Methode mit Ratio 8:1 mit Replacement. In der Treatmentgruppe befinden sich 30 Kinder. Vor dem Matching besteht die Kontrollgruppe aus 121 Kindern, nach dem Matching aus 54 Kindern. Die größte standardisierte Mittelwertdifferenz liegt bei -0.277 für die Variable Selbstkonzept Deutsch. Alle übrigen standardisierten Mittelwertdifferenzen liegen unter 0.180. Insgesamt sinkt die standardisierte Mittelwertdifferenz über alle Variablen von 1.404 auf 0.248.

Vor dem Matching ergeben sich bei drei Leistungstests signifikante Unterschiede zwischen der Treatment- und der Kontrollgruppe (s. Tabelle 66). Die Kinder der Treatmentgruppe erzielen im Wortlesen zu T3, in der Laut-Buchstaben-Zuordnung und in der Rechtschreibung signifikant niedrigere Ergebnisse. Nach dem Matching bestehen keine signifikanten Unterschiede mehr zwischen der Treatment- und der Kontrollgruppe.

## 12 Ergebnisse

Tabelle 66

*Mittelwerte vor und nach dem Matching für T3-T7 allgemein negative Auffälligkeiten*

Variable		Prä-Test				Post-Test			
		M	SD	p	t (df)	M	SD	p	t (df)
Wortlesen T2	KG	50.78	13.59	.06	1.898 (119.73)	48.53	11.59	.48	0.713 (70.98)
	TG	46.90	12.13			46.90	12.13		
Wortlesen T3	KG	51.05	7.94	.00**	3.468 (147.04)	48.08	7.79	.14	1.487 (80.07)
	TG	45.39	8.23			45.39	8.23		
Alter	KG	6.60	0.41	.12	-1.585 (147.04)	6.62	0.45	.23	-1.203 (80.07)
	TG	6.74	0.42			6.74	0.42		
Logisches Denken	KG	52.03	10.72	.11	1.628 (141.71)	49.64	11.26	.71	0.373 (74.64)
	TG	48.80	7.52			48.80	7.52		
Silbensegmentation T2	KG	50.60	11.88	.60	0.530 (127.47)	50.27	10.20	.74	0.329 (72.47)
	TG	49.54	11.78			49.54	11.78		
Buchstabenkenntnis T2	KG	50.82	12.32	.06	1.891 (121.24)	49.26	12.07	.35	0.950 (71.52)
	TG	46.90	12.06			46.90	12.06		
Laut-Buchstaben-Zuordnung T3	KG	52.18	7.83	.00**	3.637 (147.04)	49.65	8.90	.09	1.701 (80.07)
	TG	46.17	9.11			46.17	9.11		
DERET T3	KG	52.97	8.86	.00**	4.139 (147.04)	49.27	8.68	.06	1.906 (80.07)
	TG	45.20	10.52			45.20	10.52		
Selbstkonzept Deutsch T2	KG	48.37	12.10	.92	0.105 (99.34)	49.17	10.19	.64	0.465 (62.05)
	TG	48.15	12.25			48.15	12.25		

*Anmerkungen.* Alle Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. TG = Treatmentgruppe ( $n = 30$ ), KG = Kontrollgruppe ( $n_{Prä} = 121$ ,  $n_{Post} = 54$ ), \* $p < .05$ , \*\* $p < .01$ .

Für die kategorialen Variablen ergibt sich vor dem Matching nur für das Geschlecht ein signifikanter Unterschied (s. Tabelle 67). Nach dem Matching gibt es keine signifikanten Unterschiede mehr zwischen den beiden Gruppen.

Tabelle 67

*Unterschiede in den kategorialen Daten für T3-T7 allgemein negative Auffälligkeiten*

Variable	Prä-Test			Post-Test		
	p	F / X <sup>2</sup>	df	p	F / X <sup>2</sup>	df
Klasse	.17+	-	-	.92+	-	-
Familiensprache	.76	0.278	2	.83	0.191	2
Sozioökonomischer Status	.75	0.286	2	.70++	0.362	2
Geschlecht	.00**	9.798	1	.48	0.490	1

*Anmerkungen.* \* $p < .05$ , \*\* $p < .01$ . +Fisher-Freeman-Halton-Test. ++Berechnung mit 36 von 40 imputierten Datensätzen. Jeweils leere Zellen für „unqualifizierte Tätigkeiten / nicht erwerbstätig“ der beiden Gruppen Prä- und Post-Test.

### 12.4.3 Hypothese 2.1

*Die Kinder, die zu Messzeitpunkt T2 von der Lehrkraft als auffällig im Bereich Deutsch bezeichnet werden, werden sich – unabhängig von anderen Einflussfaktoren – positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die den Lehrkräften zu Beginn der Schulzeit nicht auffällig geworden sind.*

## 12 Ergebnisse

Für die Auswertung dieser Hypothese liegen zwei Datensätze vor. Der erste Datensatz besteht aus 21 Kindern in der Treatmentgruppe und 42 Kindern in der Kontrollgruppe, die sowohl zu T2 wie auch zu T6 an den Lesetests teilgenommen haben. Die Ausgangsleseleistung im Wortlesen liegt sowohl für die Treatmentgruppe ( $M = 44.89$ ,  $SD = 7.83$ ) wie auch für die Kontrollgruppe ( $M = 45.12$ ,  $SD = 7.50$ ) im durchschnittlichen Bereich.

Tabelle 68  
Leseleistungen zu T6

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	50.65 (10.65)	50.93 (9.62)	-0.106 (61)	.92	0.03 [-0.51, 0.56]
Wortlesetest	46.01 (9.76)	46.03 (10.14)	-0.010 (61)	.99	0.00 [-0.53, 0.54]
Satzlesetest	44.41 (10.14)	45.76 (9.47)	-0.520 (61)	.61	0.14 [-0.40, 0.67]
Textlesetest	43.74 (10.82)	45.49 (7.46)	-0.753 (61)	.46	0.20 [-0.33, 0.74]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

Tabelle 68 kann entnommen werden, dass es keine signifikanten Unterschiede zwischen der Treatment- und der Kontrollgruppe gibt. Auch in den ELFE-Subtests Satzlesen und Textlesen, welche zu T6 erhoben wurden, ergeben sich keine signifikanten Unterschiede.

Der Datensatz mit Kindern, die zu T2 und zu T7 an den Lesetests teilgenommen haben, besteht aus 16 Kindern in der Treatmentgruppe und 34 Kindern in der Kontrollgruppe. Die Leseleistung der beiden Gruppen liegt zu T2 mit  $M = 44.05$  ( $SD = 7.51$ ) in der Treatment- bzw.  $M = 44.23$  ( $SD = 7.70$ ) in der Kontrollgruppe im durchschnittlichen Bereich. Tabelle 69 ist zu entnehmen, dass auch für T7 keine signifikanten Unterschiede zwischen der Kontroll- und der Treatmentgruppe bestehen. Beide Gruppen verbessern sich in gleichem Maße im Wortlesen. Auch im Satz- und Textlesen unterscheiden sich die beiden Gruppen nicht signifikant voneinander.

Die Hypothese, dass Kinder, die Mitte der ersten Klasse auffällig im Bereich Deutsch werden, eine positivere Leseentwicklung zeigen als Kinder, die in diesem Bereich nicht auffällig werden, muss zurückgewiesen werden.

Tabelle 69  
Leseleistungen zu T7

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	51.52 (8.25)	50.50 (10.33)	0.347 (48)	.73	-0.11 [-0.72, 0.50]
Wortlesetest	47.15 (8.03)	46.12 (9.44)	0.374 (48)	.71	-0.11 [-0.72, 0.50]
Satzlesetest	41.60 (7.11)	45.99 (9.04)	-1.707 (48)	.09	0.52 [-0.10, 1.14]
Textlesetest	41.36 (9.40)	46.64 (9.05)	-1.903 (48)	.06	0.58 [-0.04, 1.20]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

### 12.4.4 Hypothese 2.2

Die Kinder, die zu Messzeitpunkt T3 von der Lehrkraft als auffällig im Bereich Deutsch bezeichnet werden, werden sich – unabhängig von anderen Einflussfaktoren – positiver im

## 12 Ergebnisse

*Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die den Lehrkräften zu Beginn der Schulzeit nicht auffällig geworden sind.*

Der erste Datensatz besteht aus 22 Kindern, die zu T3 auffällig im Bereich Deutsch wurden, und 26 Kindern in der Kontrollgruppe. Die Ausgangsleseleistung der Kinder ist mit 44.53 ( $SD = 7.67$ ) bzw. 43.38 ( $SD = 8.63$ ) jeweils durchschnittlich.

Die Leseleistungen der Treatment- und der Kontrollgruppe zu T6 unterscheiden sich nicht signifikant voneinander (s. Tabelle 70). In beiden Gruppen liegen diese sowohl im Wortlesen wie auch im Satz- und Textlesen im durchschnittlichen Bereich. Auch im Lese-Lernzuwachs unterscheiden sich beide Gruppen nicht signifikant voneinander und zeigen einen durchschnittlichen Anstieg.

Tabelle 70  
*Leseleistungen zu T6*

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	46.66 (7.74)	47.84 (9.58)	-0.464 (46)	.65	0.13 [-0.45, 0.72]
Wortlesetest	41.61 (6.55)	43.21 (10.25)	-0.629 (46)	.53	0.18 [-0.40, 0.77]
Satzlesetest	41.71 (9.04)	43.65 (10.67)	-0.673 (46)	.50	0.19 [-0.39, 0.78]
Textlesetest	41.23 (8.79)	44.20 (9.93)	-1.088 (46)	.28	0.32 [-0.27, 0.90]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p = < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

Der zweite Datensatz betrachtet die Leseleistungen zu T7. Dieser besteht aus 12 Kindern, die der Lehrkraft zu T3 im Bereich Deutsch auffällig geworden sind, sowie aus 24 Kindern in der Kontrollgruppe. Die Leseleistungen beider Gruppen zu T3 sind mit  $M = 44.69$  ( $SD = 12.91$ ) und  $M = 43.02$  ( $SD = 8.03$ ) durchschnittlich.

Tabelle 71  
*Leseleistungen zu T7*

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	47.52 (9.03)	49.15 (9.36)	-0.499 (34)	.62	0.18 [-0.54, 0.90]
Wortlesetest	42.99 (7.20)	45.11 (9.62)	-0.672 (34)	.51	0.24 [-0.48, 0.96]
Satzlesetest	41.10 (7.26)	44.11 (9.98)	-0.926 (34)	.36	0.33 [-0.40, 1.05]
Textlesetest	41.00 (7.87)	45.97 (7.81)	-1.794 (34)	.08	0.63 [-0.10, 1.37]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p = < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

Weder im Wortlesen noch im Satz- oder Textlesen unterscheiden sich die beiden Gruppen zu T7 (s. Tabelle 71). Auch im Lese-Lernzuwachs gibt es keinen signifikanten Unterschied. Beide Gruppen erreichen jeweils durchschnittliche Leistungen.

Die Hypothese, dass Kinder, die zum Ende der ersten Klasse im Bereich Deutsch auffällig werden, eine positivere Leseentwicklung zeigen als unauffällige Kinder, muss zurückgewiesen werden.

## 12 Ergebnisse

### 12.4.5 Hypothese 2.3

Die Kinder, die zu Messzeitpunkt T2 von der Lehrkraft als leseauffällig bezeichnet werden, werden sich – unabhängig von anderen Einflussfaktoren – positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die den Lehrkräften zu Beginn der Schulzeit nicht auffällig geworden sind.

Der erste Datensatz, der die Leseentwicklung bis zu T6 betrachtet, besteht aus den Daten von 12 Kindern, die zu T2 leseauffällig geworden sind, und 36 Kindern, die nicht leseauffällig geworden sind. Beide Gruppen haben zu T2 mit  $M = 47.80$  ( $SD = 8.24$ ) und  $M = 46.53$  ( $SD = 7.32$ ) durchschnittliche Leseleistungen im Wortlesen erreicht. Zu T6 unterscheiden sich die Leseleistungen und auch der Lese-Lernzuwachs der beiden Gruppen nicht (s. Tabelle 72).

Tabelle 72  
Leseleistungen zu T6

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	50.14 (12.21)	50.72 (10.29)	-0.163 (46)	.87	0.05 [-0.62, 0.73]
Wortlesetest	45.92 (11.70)	47.73 (10.28)	-0.510 (46)	.61	0.17 [-0.50, 0.84]
Satzlesetest	45.11 (12.49)	47.70 (8.50)	-0.810 (46)	.42	0.27 [-0.40, 0.94]
Textlesetest	46.78 (10.90)	48.49 (8.60)	-0.555 (46)	.58	-0.19 [-0.49, 0.86]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

In die Berechnung für die Leseentwicklung bis zum Ende der vierten Klasse gehen die Daten von 9 leseauffälligen Kindern und 18 unauffälligen Kindern ein. Zu T2 erreichen beide Gruppen im Schnitt durchschnittliche Leseleistungen mit  $M = 48.18$  ( $SD = 7.24$ ) bzw.  $M = 45.41$  ( $SD = 7.29$ ). Zu T7 unterscheiden sich die Leseleistungen sowie der Lese-Lernzuwachs der beiden Gruppen nicht signifikant voneinander (s. Tabelle 73). Beide Gruppen erzielen jeweils durchschnittliche Leistungen.

Tabelle 73  
Leseleistungen zu T7

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	51.55 (9.72)	49.86 (11.33)	0.383 (25)	.71	-0.16 [-1.00, 0.66]
Wortlesetest	47.41 (9.20)	47.26 (10.10)	0.036 (25)	.97	-0.01 [-0.86, 0.83]
Satzlesetest	42.09 (9.01)	47.02 (9.20)	-1.321 (25)	.20	0.54 [-0.31, 1.39]
Textlesetest	45.09 (8.90)	47.30 (7.90)	-0.657 (25)	.52	0.27 [-0.58, 1.11]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

Die Hypothese, dass Kinder, die der Lehrkraft zu T2 leseauffällig werden, eine positivere Leseentwicklung zeigen als unauffällige Kinder, muss zurückgewiesen werden.

## 12 Ergebnisse

### 12.4.6 Hypothese 2.4

Die Kinder, die zu Messzeitpunkt T3 von der Lehrkraft als leseauffällig bezeichnet werden, werden sich – unabhängig von anderen Einflussfaktoren – positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die den Lehrkräften zu T3 nicht auffällig geworden sind.

Die erste Berechnung betrachtet die Leseentwicklung bis T6. Die Treatmentgruppe besteht aus 15 Kindern, die der Lehrkraft zu T3 leseauffällig werden. Die Kontrollgruppe besteht ebenfalls aus 15 Kindern. Die Kinder erreichen zu T3 eine durchschnittliche Leseleistung von  $M = 41.42$  ( $SD = 8.56$ ) bzw.  $M = 41.54$  ( $SD = 7.46$ ).

Tabelle 74  
Leseleistungen zu T6

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	49.56 (5.83)	49.10 (9.53)	0.157 (23.17)	.88	-0.06 [-0.81, 0.69]
Wortlesetest	39.73 (5.38)	38.73 (8.27)	0.392 (28)	.70	-0.14 [-0.89, 0.61]
Satzlesetest	37.96 (7.27)	37.09 (6.56)	0.341 (28)	.74	-0.12 [-0.87, 0.62]
Textlesetest	37.29 (6.02)	39.26 (7.85)	-0.768 (28)	.45	0.28 [-0.47, 1.03]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

Die Leseleistung der beiden Gruppen unterscheidet sich zu T6 nicht. Auch der Lese-Lernzuwachs ist nicht signifikant unterschiedlich in den beiden Gruppen (s. Tabelle 74).

Tabelle 75  
Leseleistungen zu T7

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	50.23 (8.67)	48.90 (15.09)	0.228 (16)	.82	-0.11 [-1.11, 0.89]
Wortlesetest	42.84 (7.90)	42.52 (11.13)	0.070 (16)	.95	-0.03 [-1.03, 0.97]
Satzlesetest	39.99 (7.88)	40.06 (8.83)	-0.016 (16)	.99	0.00 [-0.99, 1.00]
Textlesetest	38.42 (6.37)	42.63 (7.53)	-1.282 (16)	.22	0.60 [-0.42, 1.63]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

Die Stichprobe für die Leseleistung zu T7 besteht aus 9 Kindern, die zu T3 leseauffällig geworden sind, und aus einer Kontrollgruppe, die ebenfalls aus 9 Kindern besteht. Auch diese beiden Gruppen erreichen zu T3 eine noch durchschnittliche Leseleistung von  $M = 40.43$  ( $SD = 9.51$ ) bzw.  $M = 42.38$  ( $SD = 9.61$ ). Zu T7 zeigen sich keine signifikanten Unterschiede in der Leseleistung der beiden Gruppen. Der Lese-Lernzuwachs unterscheidet sich in beiden Gruppen ebenfalls nicht signifikant voneinander (s. Tabelle 75).

## 12.4.7 Hypothese 2.5

Die Kinder, die zu Messzeitpunkt T2 der Lehrkraft allgemein auffällig werden, werden sich in der Folge positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die der Lehrkraft zu T2 nicht aufgefallen sind.

Der Datensatz zur Berechnung der Leseentwicklung bis zu T6 besteht aus 48 Kindern, die den Lehrkräften allgemein auffällig geworden sind und ebenfalls 48 Kindern in der Kontrollgruppe. Mit  $M = 46.84$  ( $SD = 8.41$ ) und  $M = 47.07$  ( $SD = 9.27$ ) erreichen sie zu T2 eine durchschnittliche Ausgangsleseleistung.

Tabelle 76 kann entnommen werden, dass es keine signifikanten Unterschiede in der Leseleistung zu T6 oder im Lese-Lernzuwachs gibt. Beide Gruppen erzielen jeweils durchschnittliche Ergebnisse.

Tabelle 76  
Leseleistungen zu T6

	$M_t$ (SD)	$M_o$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	49.48 (10.30)	50.43 (10.37)	-0.449 (94)	.65	0.09 [-0.31, 0.50]
Wortlesetest	46.76 (9.70)	46.88 (10.59)	-0.056 (94)	.96	0.01 [-0.39, 0.42]
Satzlesetest	46.01 (10.14)	47.37 (10.16)	-0.656 (94)	.51	0.13 [-0.27, 0.54]
Textlesetest	45.34 (10.26)	47.60 (9.88)	-1.100 (94)	.27	0.22 [-0.18, 0.63]

Anmerkungen.  $M_t$  = Mittelwert der Treatmentgruppe.  $M_o$  = Mittelwert der Kontrollgruppe. \* $p < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls

Für die Berechnung der Leseentwicklung bis zu T7 werden die Daten von 29 Kindern herangezogen, die den Lehrkräften zu T2 allgemein auffällig geworden sind, sowie die Daten von 51 Kindern als Kontrollgruppe. Die Leseleistung beider Gruppen war zu T2 mit  $M = 49.64$  ( $SD = 7.80$ ) und  $M = 46.33$  ( $SD = 7.55$ ) im durchschnittlichen Bereich.

Tabelle 77  
Leseleistungen zu T7

	$M_t$ (SD)	$M_o$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	50.32 (7.84)	51.16 (10.37)	-0.382 (78)	.70	0.09 [-0.37, 0.55]
Wortlesetest	47.25 (8.35)	49.92 (10.28)	-1.189 (78)	.24	0.28 [-0.19, 0.74]
Satzlesetest	43.95 (9.07)	50.83 (9.27)	-3.216 (78)	<b>.00**</b>	0.75 [0.27, 1.23]
Textlesetest	44.06 (9.09)	50.67 (10.11)	-2.911 (78)	<b>.01*</b>	0.68 [0.20, 1.15]

Anmerkungen.  $M_t$  = Mittelwert der Treatmentgruppe.  $M_o$  = Mittelwert der Kontrollgruppe. \* $p < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls. Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Der Lernzuwachs im Wortlesen unterscheidet sich nicht signifikant in den beiden Gruppen (s. Tabelle 77). Im Wortlesen zu T7 unterscheiden sich die Gruppen ebenfalls nicht signifikant voneinander. Im Satzlesen und im Textlesen zu T7 zeigen sich dagegen signifikante Unterschiede. Kinder, die den Lehrkräften zu T2 allgemein auffällig wurden, erzielen zu T7

signifikant niedrigere Leistungen. Für beide Tests liegen mit  $d = 0.75$  und  $d = 0.68$  hohe Effektstärken vor. Unter Berücksichtigung der 95 %-Konfidenzintervalle liegen die Effektstärken jeweils im mittleren bis großen Bereich.

Die Hypothese, dass sich Kinder, die der Lehrkraft zu T2 allgemein negativ auffallen, positiver im Lesen entwickeln als Kinder, die nicht auffallen, muss zurückgewiesen werden. Lediglich im Satzlesen und im Textlesen zeigt sich für die Überprüfung zu T7 ein signifikanter Unterschied, allerdings entgegen der vermuteten Richtung. Kinder, die der Lehrkraft allgemein auffällig geworden sind, zeigen niedrigere Leistungen im Satz- und Textlesen zu T7.

#### 12.4.8 Hypothese 2.6

*Die Kinder, die zu Messzeitpunkt T3 der Lehrkraft allgemein auffällig werden, werden sich in der Folge positiver im Lesen entwickeln (überprüft zu Messzeitpunkten T6 und T7) als die Kinder, die der Lehrkraft zu T3 nicht aufgefallen sind.*

Der Datensatz zum Vergleich der Leseentwicklung bis zu T6 besteht aus 53 Kindern in der Treatmentgruppe und 66 Kindern in der Kontrollgruppe. Die Leseleistung der Kinder zu T3 ist mit  $M = 48.51$  ( $SD = 10.90$ ) bzw.  $M = 46.30$  ( $SD = 10.47$ ) jeweils durchschnittlich.

Der Lese-Lernzuwachs unterscheidet sich in beiden Gruppen nicht signifikant voneinander (s. Tabelle 78). Auch im Wortlesen und im Textlesen zu T6 zeigen sich keine signifikanten Unterschiede. Im Satzlesen erzielen die Kinder der Kontrollgruppe dagegen signifikant bessere Leistungen als die Kinder der Treatmentgruppe.

Tabelle 78  
Leseleistungen zu T6

	$M_t$ (SD)	$M_o$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	48.61 (9.06)	50.53 (10.52)	-1.050 (117)	.30	0.19 [-0.17, 0.56]
Wortlesetest	45.06 (9.18)	48.26 (9.73)	-1.826 (117)	.07	0.34 [-0.03, 0.70]
Satzlesetest	43.47 (8.43)	48.27 (9.50)	-2.882 (117)	<b>.01*</b>	0.53 [0.16, 0.90]
Textlesetest	44.88 (9.22)	48.34 (9.87)	-1.954 (117)	.05	0.36 [-0.01, 0.73]

Anmerkungen.  $M_t$  = Mittelwert der Treatmentgruppe.  $M_o$  = Mittelwert der Kontrollgruppe. \* $p = < .05$ , \*\* $p < .01$ . Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls. Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Zur Prüfung der Leseentwicklung bis T7 werden die Daten von 30 Kindern in der Treatmentgruppe und von 54 Kindern in der Kontrollgruppe herangezogen. Die Kinder erreichen zu T3 jeweils durchschnittliche Leistungen im Wortlesen mit  $M = 48.53$  ( $SD = 11.59$ ) und  $M = 46.90$  ( $SD = 12.13$ ). Im Wortlesen zu T7 und im Lese-Lernzuwachs zeigen sich keine signifikanten Unterschiede (s. Tabelle 79). Die Kinder der Kontrollgruppe erreichen allerdings signifikant höhere Leistungen im Satzlesen und im Textlesen zu T7.

Tabelle 79  
Leseleistungen zu T7

	$M_1$ (SD)	$M_0$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Lese-Lernzuwachs	49.10 (10.33)	50.56 (9.90)	-0.634 (82)	.53	0.14 [-0.31, 0.60]
Wortlesetest	45.40 (8.20)	48.89 (9.63)	-1.676 (82)	.10	0.38 [-0.29, 0.62]
Satzlesetest	43.52 (8.49)	48.68 (9.81)	-2.419 (82)	<b>.02*</b>	0.55 [0.09, 1.01]
Textlesetest	43.21 (8.42)	49.54 (10.34)	-2.862 (82)	<b>.01*</b>	0.65 [0.19, 1.12]

Anmerkungen.  $M_1$  = Mittelwert der Treatmentgruppe.  $M_0$  = Mittelwert der Kontrollgruppe. \* $p$  = < .05, \*\* $p$  < .01. Effektstärke Cohens  $d$  mit Angabe des 95 %-Konfidenzintervalls. Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Für T6 zeigt sich im Satzlesen, entgegen der Hypothese, eine bessere Leistung für die Kontrollgruppe. Für T7 zeigt sich ebenfalls entgegen der Hypothese sowohl im Satz- wie auch im Textlesen eine bessere Leistung in der Kontrollgruppe. Daher muss die Hypothese zurückgewiesen werden: Die Kinder, die den Lehrkräften zu T3 allgemein auffällig geworden sind, entwickeln sich nicht positiver im Lesen als Kinder, die nicht auffällig geworden sind.

### 12.5 Fragestellung 3: Eingeleitete (Förder-)Maßnahmen bei Auffälligkeiten

Um diese Fragestellung beantworten zu können, werden zunächst die von den Lehrkräften genannten Maßnahmen betrachtet. Im Anschluss erfolgt die Überprüfung der Hypothesen.

#### 12.5.1 Genannte Maßnahmen

Je nach Messzeitpunkt wurden von den Lehrkräften zu 21 bis 67 Kindern Maßnahmen genannt (s. Tabelle 80). Zu allen Oberkategorien wurden zu jedem Messzeitpunkt Aussagen getätigt. Am häufigsten werden organisatorische innerschulische Aspekte angegeben. Förderung und Unterstützung finden demnach über alle Messzeitpunkte hinweg im Klassenzimmer für 24 bis 60 Prozent der Kinder statt, für die Maßnahmen genannt wurden. Die Lehrkraft führt diese oft auch selbst durch (zwischen 24 und 58 Prozent). Förderunterricht wird ab Mitte der ersten Klasse durchgehend genannt. Auch über außerschulische Hilfe und Gespräche mit Eltern berichten die Lehrkräfte. Das BFZ wird ebenfalls als Ansprechpartner oder Unterstützung genannt. Über inhaltliche Aspekte wird dagegen seltener berichtet. Differenzierung findet für 2 bis 52 Prozent der Kinder statt, die Unterstützung erhalten.

Lob und positive Verstärkung werden selten explizit als Maßnahme genannt. Es werden auch nur sehr wenig spezifische Fördermethoden aufgezählt.

## 12 Ergebnisse

Tabelle 80

### *Absolute Häufigkeiten genannter Maßnahmen über alle Kinder*

Messzeitpunkt	T1 <i>n</i> = 21	T2 <i>n</i> = 44	T3 <i>n</i> = 39	T4 <i>n</i> = 43	T5 <i>n</i> = 54	T6 <i>n</i> = 67	T7 <i>n</i> = 21
<b>Organisatorische innerschulische Aspekte</b>							
Förderung / Unterstützung in der Klasse	8	22	21	26	13	20	11
Lehrkraft fördert selbst	7	18	20	25	13	16	6
Förderunterricht	-	10	21	26	19	34	8
<b>Organisatorische Aspekte der Zusammenarbeit</b>							
Elterngespräche	-	5	5	5	3	12	-
Außerschulische Hilfe	11	7	7	2	6	14	6
BFZ ist eingeschaltet	1	5	3	3	28	24	3
<b>Inhaltliche Aspekte</b>							
Differenzierung	7	10	5	1	5	3	11
Lob, positive Verstärkung	-	2	-	4	2	-	-
Förderplan	-	3	-	1	1	2	8
Spezifische Fördermethode genannt	4	7	-	1	1	1	4
<b>Sonstiges</b>							
Klassenwiederholung	-	1	5	1	2	2	-
ASF ist festgestellt	-	4	1	1	5	1	2
Sonstige professionelle Förderung	11	5	*	2	6	10	4

*Anmerkungen.* Mehrfachnennungen möglich. *n* = Anzahl an Kindern, zu denen Lehrkräfte etwas gesagt haben. ASF = Anspruch auf sonderpädagogische Förderung. \* = Kategorie wird aufgrund zu niedriger Beurteilerübereinstimmung nicht berücksichtigt.

Förderpläne werden am häufigsten zum Ende der vierten Klasse für 38 Prozent der Kinder angeführt, für die Maßnahmen benannt werden.

#### 12.5.1.1 Lesespezifische Maßnahmen

Tabelle 81 gibt eine Übersicht, welche zwischen Maßnahmen differenziert, die sich explizit auf das Lesen beziehen und solchen Maßnahmen, die Lesen nicht explizit benennen, aber Leseförderung wahrscheinlich beinhalten, wie beispielsweise die Teilnahme am Förderunterricht Deutsch oder Sprachförderung. Ebenso wurden Maßnahmen genannt, die sich nicht auf Leseförderung bzw. auf Förderung im Bereich Deutsch beziehen.

Tabelle 81

### *Absolute Häufigkeiten genannter Maßnahmen zu den sieben Messzeitpunkten*

Messzeitpunkt	T1	T2	T3	T4	T5	T6	T7
Nicht im Lesen	10	15	9	16	31	21	2
Auch im Lesen	6	22	30	24	20	44	12
Explizit im Lesen	5	7	0	3	3	2	7

Lesespezifische Fördermaßnahmen erhalten zu den sieben Messzeitpunkten zwischen null (T3) und sieben (T2 und T7) Kinder. Dies entspricht einem Anteil von null bis drei Prozent aller Schülerinnen und Schüler.

## 12 Ergebnisse

Die Angaben sind an dieser Stelle unabhängig von bestehenden Leseschwächen oder Leseauffälligkeiten dargestellt. Explizite Leseförderung wurde seltener genannt als Förderung, die sich allgemein auf den Bereich Deutsch bezieht und Lesen auch beinhalten kann. Andere Bereiche, die sich nicht auf das Lesen beziehen, werden ebenfalls zu allen sieben Messzeitpunkten genannt und mit Ausnahme von T7 auch häufiger als explizite Leseförderung.

### 12.5.1.2 Genannte Maßnahmen für allgemein auffällige Kinder

Tabelle 82 zeigt die Häufigkeiten genannter Maßnahmen für Kinder, die den Lehrkräften allgemein negativ auffällig geworden sind. Zu jeder Kategorie gibt es zu mindestens einem Messzeitpunkt Nennungen. Am häufigsten kamen Aussagen in Bezug auf organisatorische innerschulische Aspekte vor. Allgemein negative Auffälligkeiten beziehen sich nicht immer auf schulische Leistungen. Entsprechend werden für diese Kinder auch häufiger Maßnahmen genannt, die sonstiger professioneller Förderung zugerechnet werden. Zu jedem Messzeitpunkt trifft dies auf zwei bis elf Kinder zu, was 3 bis 21 Prozent der negativ auffällig gewordenen Kinder entspricht.

Tabelle 82

#### *Absolute Häufigkeiten genannter Maßnahmen für allgemein negativ auffällige Kinder*

Messzeitpunkt	T1 <i>n</i> = 59	T2 <i>n</i> = 63	T3 <i>n</i> = 67	T4 <i>n</i> = 72	T5 <i>n</i> = 62	T6 <i>n</i> = 83	T7 <i>n</i> = 19
Wird etwas getan?	17 (29 %)	23 (37 %)	35 (52 %)	42 (58 %)	38 (61 %)	63 (76 %)	16 (84 %)
<b>Organisatorische innerschulische Aspekte</b>							
Förderung / Unterstützung in der Klasse	4	8	17	25	9	20	8
Lehrkraft fördert selbst	3	6	16	24	9	16	3
Förderunterricht	-	6	17	25	16	34	7
<b>Organisatorische Aspekte der Zusammenarbeit</b>							
Elterngespräche	-	1	5	5	2	12	-
Außerschulische Hilfe	11	6	7	2	4	14	6
BFZ ist eingeschaltet	1	5	3	3	19	24	2
<b>Inhaltliche Aspekte</b>							
Differenzierung	3	4	5	1	4	3	8
Lob, positive Verstärkung	-	2	-	4	1	-	-
Förderplan	-	2	-	1	1	2	7
Spezifische Fördermethode genannt	-	2	-	1	1	1	4
<b>Sonstiges</b>							
Klassenwiederholung	-	1	5	1	2	1	-
ASF festgestellt	-	3	1	1	2	-	2
Sonstige professionelle Förderung	11	4	*	2	4	10	4

*Anmerkungen.* Mehrfachnennungen möglich. \*Kategorie wird aufgrund zu niedriger Beurteilerübereinstimmung nicht berücksichtigt. Die Prozentangabe bezieht sich auf allgemein auffällige Kinder, für die Maßnahmen genannt werden.

### 12.5.1.3 Genannte Maßnahmen für leseauffällige Kinder

Maßnahmen, die für leseauffällige Kinder genannt wurden, finden sich in Tabelle 83, wobei T1 nicht abgebildet ist, da zu diesem Zeitpunkt kein Kind leseauffällig wurde. Es wurden keine

Aussagen über einen Anspruch auf sonderpädagogische Förderung oder über den Einsatz von Lob und positiver Verstärkung getätigt. Zu drei Zeitpunkten (T5 bis T7) wird jeweils einmal von einer spezifischen Fördermethode berichtet.

Zur Mitte des ersten Schuljahres (T2) werden 15 Kinder leseauffällig. Für drei dieser Kinder werden Maßnahmen genannt, was 20 Prozent der leseauffälligen Kinder entspricht. Keine dieser Maßnahmen (Elterngespräche, außerschulische Hilfe und Klassenwiederholung) findet im Unterricht im Klassenzimmer statt. Ab Ende der ersten Klasse werden dagegen zu zwei Dritteln bis zu allen leseauffälligen Kindern Maßnahmen genannt, auch solche, die im Klassenzimmer stattfinden.

Tabelle 83

*Absolute Häufigkeiten genannter Maßnahmen für leseauffällige Kinder je Messzeitpunkt*

Messzeitpunkt	T2 <i>n</i> = 15	T3 <i>n</i> = 19	T4 <i>n</i> = 16	T5 <i>n</i> = 9	T6 <i>n</i> = 17	T7 <i>n</i> = 5
Wird etwas getan?	3 (20 %)	12 (63 %)	14 (88 %)	7 (78 %)	13 (76 %)	5 (100 %)
<b>Organisatorische innerschulische Aspekte</b>						
Förderung / Unterstützung in der Klasse	-	9	9	3	8	3
Lehrkraft fördert selbst	-	9	9	3	7	2
Förderunterricht	-	8	12	5	9	4
<b>Organisatorische Aspekte der Zusammenarbeit</b>						
Elterngespräche	1	1	-	-	2	-
Außerschulische Hilfe	1	2	1	1	2	2
BFZ ist eingeschaltet	-	-	-	1	7	-
<b>Inhaltliche Aspekte</b>						
Differenzierung	-	2	-	1	-	4
Förderplan	-	-	-	-	2	3
Spezifische Fördermethode genannt	-	-	-	1	1	1
<b>Sonstiges</b>						
Klassenwiederholung	1	2	-	-	-	-
Sonstige professionelle Förderung	-	*	1	1	-	2

*Anmerkungen.* Mehrfachnennungen möglich. \*Kategorie wird aufgrund zu niedriger Beurteilerübereinstimmung nicht berücksichtigt. Die Prozentangabe bezieht sich auf leseauffällige Kinder, für die Maßnahmen genannt werden.

Der Fokus der vorliegenden Untersuchung liegt auf Maßnahmen, die von den Lehrkräften eingeleitet werden. Daher wird noch einmal spezifisch auf die organisatorischen innerschulischen Aspekte geblickt. Abbildung 9 zeigt die prozentualen Häufigkeiten für diese Maßnahmen für die Messzeitpunkte T3 bis T7.

## 12 Ergebnisse

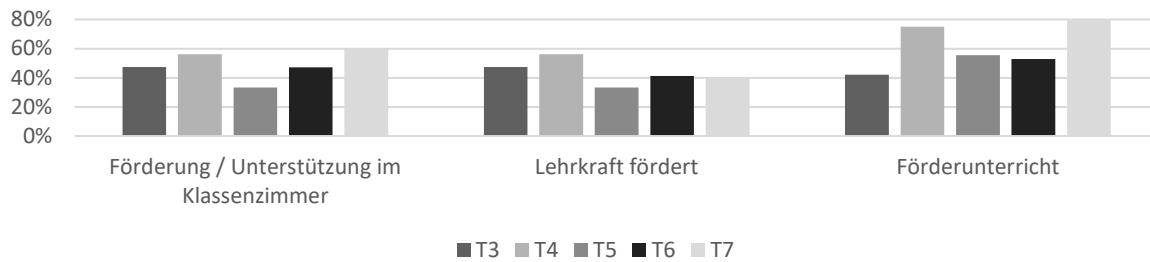


Abbildung 9. Prozentuale Häufigkeiten der genannten organisatorischen innerschulischen Maßnahmen bei leseauffälligen Kindern

Die Lehrkräfte fördern jeweils zwischen 33 und 56 Prozent der leseauffälligen Kinder selbst. Zwischen 33 und 60 Prozent dieser Kinder erhalten Förderung bzw. Unterstützung im Klassenzimmer und zwischen 42 und 80 Prozent der Kinder erhalten Förderunterricht.

### 12.5.1.4 Genannte Maßnahmen für leseschwache Kinder

Tabelle 84 gibt die Maßnahmen an, die für leseschwache Kinder zu T2 bis T7 genannt wurden. T1 und T5 sind nicht abgebildet, da zu diesen Zeitpunkten kein Lesetest durchgeführt wurde und die leseschwachen Kinder dort nicht ermittelt wurden. Insgesamt werden jeweils 26 bis 53 Prozent der leseschwachen Kinder mit Maßnahmen versorgt. Nur zwei Mal werden spezifische Fördermethoden genannt.

Tabelle 84

*Absolute Häufigkeiten genannter Maßnahmen für leseschwache Kinder*

Messzeitpunkt	T2 <i>n</i> = 36	T3 <i>n</i> = 43	T4 <i>n</i> = 37	T6 <i>n</i> = 38	T7 <i>n</i> = 27
Wird etwas getan?	10 (28 %)	20 (47 %)	16 (43 %)	20 (53 %)	7 (26 %)
<b>Organisatorische innerschulische Aspekte</b>					
Förderung / Unterstützung in der Klasse	4	15	9	7	5
Lehrkraft fördert selbst	3	14	9	3	3
Förderunterricht	4	11	12	11	3
<b>Organisatorische Aspekte der Zusammenarbeit</b>					
Elterngespräche	2	3	-	7	-
Außerschulische Hilfe	1	4	1	7	2
BFZ ist eingeschaltet	1	1	2	12	1
<b>Inhaltliche Aspekte</b>					
Differenzierung	1	4	1	2	7
Förderplan	-	-	-	2	2
Spezifische Fördermethode genannt	-	-	-	1	1
<b>Sonstiges</b>					
Klassenwiederholung	1	-	-	-	-
ASF festgestellt	1	-	1	-	2
Sonstige professionelle Förderung	1	*	1	4	1

*Anmerkungen.* Mehrfachnennungen möglich. ASF = Anspruch auf sonderpädagogische Förderung. \*Kategorie wird aufgrund zu niedriger Beurteilerübereinstimmung nicht berücksichtigt.

## 12 Ergebnisse

Die meisten Nennungen entfallen auf organisatorische innerschulische Aspekte. Zu T6 wurde für 12 der 38 leseschwachen Kinder, also knapp einem Drittel, das BFZ eingeschaltet. Zu anderen Messzeitpunkten geschah dies seltener. Abbildung 10 zeigt die prozentualen Häufigkeiten für genannte organisatorische innerschulische Maßnahmen.

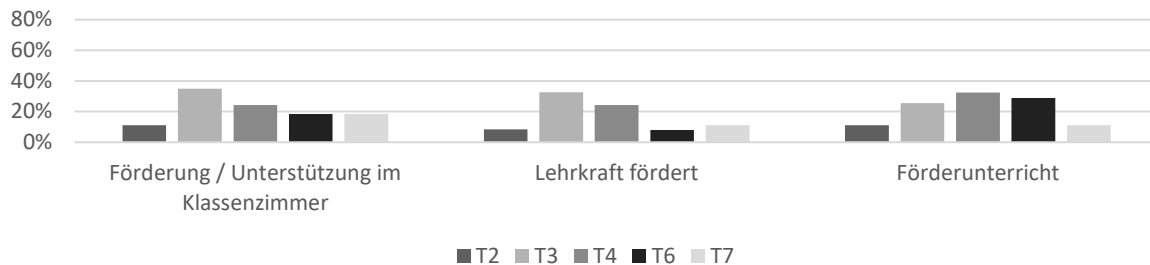


Abbildung 10. Prozentuale Häufigkeiten der genannten organisatorischen innerschulischen Maßnahmen bei leseschwachen Kindern

Die Zahlen liegen zwischen 8 und 35 Prozent und somit niedriger als die genannten Maßnahmen bei leseauffälligen Kindern (s. Abbildung 9).

### 12.5.1.5 Schulinterne Fördermaßnahmen

Außerschulische Hilfe wurde in einer einzelnen Kategorie erfasst. Schulinterne Förderung ergibt sich dagegen aus verschiedenen Kategorien. Als schulinterne Förderung zusammengefasst werden organisatorische innerschulische Aspekte sowie inhaltliche Aspekte, welche sich eindeutig auf innerschulische Maßnahmen beziehen. Es wird hierbei differenziert zwischen schulinterner Förderung allgemein, schulinterner Förderung im Bereich Deutsch sowie schulinterner Förderung im Lesen. Schulinterne Förderung allgemein beinhaltet alle genannten Maßnahmen, auch wenn sie sich nicht auf das Lesen oder auf den Bereich Deutsch beziehen. Schulinterne Förderung im Bereich Deutsch beinhaltet dagegen nur Maßnahmen, die sich auf Förderung im Lesen, in der Rechtschreibung oder im sprachlichen Bereich beziehen. Zur schulinternen Leseförderung werden die Kategorien nur dann gezählt, wenn die Beurteilerin sich dafür entschieden hat, dass diese Förderung eindeutig im Bereich Lesen durchgeführt wird.

Tabelle 85

*Absolute Häufigkeiten schulinterner Fördermaßnahmen für leseauffällige Kinder*

	T2 <i>n</i> = 15	T3 <i>n</i> = 19	T4 <i>n</i> = 16	T5 <i>n</i> = 9	T6 <i>n</i> = 17	T7 <i>n</i> = 5
Leseförderung	0	0	3	1	0	4
Deutschförderung	0	10	11	6	11	5
Förderung allgemein	0	10	12	6	11	5

*Anmerkungen.* Förderung allgemein beinhaltet Deutschförderung und Leseförderung. Deutschförderung beinhaltet Leseförderung.

Zu Messzeitpunkt T2 findet keine schulinterne Förderung für leseauffällig gewordene Kinder statt (s. Tabelle 85). Insgesamt zeigt sich, dass schulinterne Förderung für leseauffällige Kinder mit Ausnahme einer Maßnahme zu T4 immer mindestens im Bereich Deutsch angelegt ist und nicht nur allgemeine Förderung genannt wird. Seltener wird spezifisch eine schulinterne Leseförderung genannt. Zu T7 erhalten allerdings vier der fünf zu diesem Messzeitpunkt leseauffälligen Kinder explizit eine schulinterne Leseförderung.

Tabelle 86

*Absolute Häufigkeiten schulinterner Fördermaßnahmen für leseschwache Kinder*

	T2 <i>n</i> = 36	T3 <i>n</i> = 43	T4 <i>n</i> = 37	T6 <i>n</i> = 38	T7 <i>n</i> = 27
Leseförderung	0	0	2	1	4
Deutschförderung	3	15	10	13	7
Förderung allgemein	4	16	13	13	7

*Anmerkungen.* Förderung allgemein beinhaltet Deutschförderung und Leseförderung. Deutschförderung beinhaltet Leseförderung. Zu T1 und T5 wurde kein Lesetest durchgeführt.

Ein sehr ähnliches Bild ergibt sich auch für die leseschwachen Kinder (s. Tabelle 86). Von den 36 zu T2 leseschwachen Kindern erhalten vier eine allgemeine schulinterne Förderung, wobei diese bei drei Kindern im Bereich Deutsch stattfindet. Für kein leseschwaches Kind wird eine explizite schulinterne Leseförderung genannt. Zu den übrigen Messzeitpunkten liegen fast alle genannten schulinternen Förderungen mindestens im Bereich Deutsch. Zu T7 erhalten vier von insgesamt sieben Kindern, die schulinterne Förderung erhalten, diese im Bereich Lesen.

## 12.5.2 Hypothese 3.1

*Zwei Drittel der als leseauffällig genannten Kinder (zu T2 und T3) bleiben in der Folge zu den Messzeitpunkten T2 bis T7 ohne schulinterne Leseförderung.*

Es sind insgesamt 30 Kinder zu T2 und/oder zu T3 als leseauffällig bezeichnet worden, davon 15 zu T2 und 19 zu T3. Tabelle 87 zeigt die Häufigkeiten genannter schulinterner Fördermaßnahmen zu den Messzeitpunkten T2 bis T7. Da es sich bei Nennung von schulinterner Förderung immer auch um mindestens eine Maßnahme im Bereich Deutsch handelte, wird auf die Darstellung allgemeiner Förderung verzichtet. Von den 15 Kindern, die zu T2 leseauffällig wurden, erhalten in der Folge pro Messzeitpunkt nur jeweils 2 bis 3 eine schulinterne Deutschförderung. Über alle Messzeitpunkte hinweg erhalten 6 der 15 Kinder diese Förderung. Eine explizite Leseförderung wurde nur für ein Kind genannt.

Von den 19 Kindern, die zu T3 leseauffällig sind, erhalten in der Folge jeweils 4 bis 10 eine schulinterne Deutschförderung. Über alle Messzeitpunkte sind es 17 von 19 Kindern mit mindestens schulinterner Deutschförderung, wobei sich diese für 4 Kinder explizit auf den Bereich Lesen bezieht.

Tabelle 87

*Absolute Häufigkeiten schulinterner Fördermaßnahmen für leseauffällige Kinder*

Messzeitpunkt		T2	T3	T4	T5	T6	T7	T2-T7
T2 (n = 15)	Leseförderung	0	0	0	0	0	1	1
	Deutschförderung	0	3	2	0	2	3	6
T3 (n = 19)	Leseförderung	2	0	0	1	0	2	4
	Deutschförderung	6	10	5	4	5	3	17
T2 mit T3 (n = 30)	Leseförderung	2	0	0	1	0	3	5
	Deutschförderung	6	11	6	4	7	6	20

*Anmerkungen.* Deutschförderung beinhaltet auch die Nennungen der Leseförderung.

Zusammengefasst haben von den 30 Kindern, die zu T2 und/oder T3 leseauffällig geworden sind, 5 Kinder zu mindestens einem der folgenden Messzeitpunkte eine schulinterne Leseförderung erhalten.

Tabelle 88

*Anteil leseauffälliger Kinder ohne schulinterne Förderung*

	n	Ohne Förderung	%	95 %-KI	p
Leseförderung	30	25	83	65 %, 94 %	.05
Deutschförderung	30	10	33	17 %, 53 %	<b>.00**</b>

*Anmerkungen.* n bezieht sich auf die Zahl leseschwacher Kinder zu T2 oder T3. % ist der prozentuale Anteil der Kinder, die in der Folge ohne Förderung bleiben. \*  $p < .05$ , \*\* $p < .01$ . Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Insgesamt wurde zu 25 von 30 zu T2 und/oder T3 leseauffälligen Kindern keine schulinterne Leseförderung genannt, was einem Anteil von 83 Prozent (s. Tabelle 88) entspricht. Diese 83 Prozent werden im Binomialtest mit der erwarteten Häufigkeit von 67 Prozent verglichen (zweiseitige Testung). Es ergibt sich mit  $p = .05$  kein signifikanter Unterschied zu dem theoretisch erwarteten Anteil. Zwei Drittel der zu T2 oder T3 als leseauffällig bezeichneten Kinder bleiben zu T2 bis T7 ohne explizit genannte schulinterne Leseförderung.

Die schulinterne Deutschförderung wird für die leseauffälligen Kinder ebenfalls berücksichtigt, da die Lehrkräfte eventuell Leseförderung nicht explizit benannt haben, obwohl diese Bestandteil der Deutschförderung ist. Von den 30 leseschwachen Kindern haben 20 eine schulinterne Deutschförderung erhalten. Somit sind 10 Kinder ohne schulinterne Deutschförderung geblieben, was einem Anteil von 33 Prozent entspricht (s. Tabelle 88). Der Binomialtest ergibt mit  $p < .001$  (zweiseitige Testung) eine signifikante Abweichung von den erwarteten 67 Prozent. Somit haben signifikant weniger als zwei Drittel der leseauffällig gewordenen Kinder keine schulinterne Deutschförderung erhalten.

Betrachtet man die genannte schulinterne Leseförderung, so wird die Hypothese beibehalten: Zwei Drittel der zu T2 oder T3 leseauffälligen Kinder bleibt in der Folge ohne schulinterne Leseförderung. Es bleiben allerdings signifikant weniger als zwei Drittel dieser Kinder ohne schulinterne Deutschförderung.

## 12.5.3 Hypothese 3.2

Analog zu Hypothese 3.1 bleiben auch zwei Drittel der allgemein als auffällig genannten Kinder (zu T2 und T3) in der Folge zu den Messzeitpunkten T2 bis T7 ohne schulinterne Förderung.

Insgesamt sind 99 Kinder zu T2 oder zu T3 allgemein auffällig geworden (s. Tabelle 89). Davon haben 54 Kinder in der Folge eine allgemeine schulinterne Förderung erhalten. Für 45 dieser Kinder wurde keine solche Förderung von der Lehrkraft genannt, was einem Anteil von 45 Prozent entspricht. Dieser Anteil wird analog zu Hypothese 3.1 ebenfalls mit dem Binomialtest mit der erwarteten Häufigkeit von 67 Prozent verglichen.

Tabelle 89

Häufigkeit allgemeiner schulinterner Fördermaßnahmen für zu T2 oder T3 allgemein auffällige Kinder

		T2	T3	T4	T5	T6	T7	T2-T7
Förderung allgemein	T2 (n = 63)	11	12	11	9	13	8	33
	T3 (n = 67)	15	23	18	14	19	10	46
	T2 mit T3 (n = 99)	18	25	20	15	23	12	54

Es ergibt sich für diesen Anteil mit  $p < .001$  (zweiseitiger Binomialtest) ein signifikanter Unterschied zum erwarteten Anteil von 67 Prozent (s. Tabelle 90).

Tabelle 90

Anteil allgemein auffälliger Kinder ohne schulinterne Förderung

	n	Ohne Förderung	%	95 %-KI	p
Schulinterne Förderung	99	45	45	35 %, 56 %	.00*

Anmerkungen. n bezieht sich auf die Zahl leseschwacher Kinder zu T2 oder T3. % ist der prozentuale Anteil der Kinder, die in der Folge ohne Förderung bleiben. \*  $p < .05$ , \*\* $p < .01$

Die Nullhypothese wird daher zurückgewiesen. Es haben mit 45 Prozent signifikant weniger als zwei Drittel der Kinder, die zu T2 oder zu T3 allgemein negativ aufgefallen sind, in der Folge keine schulinterne Förderung erhalten.

## 12.6 Fragestellung 4: Erfolg der durchgeführten (Förder-)Maßnahmen

Während die dritte Fragestellung sich darauf fokussierte, inwiefern Schülerinnen und Schüler, die den Lehrkräften auffallen, mit Unterstützungsangeboten versorgt werden, soll nun geprüft werden, ob die genannten Maßnahmen zu einer positiveren Entwicklung im Lesen führen.

## 12.6.1 Hypothese 4.1

Wenn ein Kind zu Beginn der Schulzeit (T2 bzw. T3) leseschwach ist, die Lehrkraft dies bemerkt und lesespezifische Fördermaßnahmen eingeleitet werden, dann erzielt das Kind zu T7 eine höhere Leseleistung als ein leseschwaches Kind, bei dem dies nicht geschehen ist.

## 12 Ergebnisse

Zur Auswertung der Hypothese werden Nennungen zur schulinternen Förderung in Deutsch (ohne explizite Nennung von Lese- oder Rechtschreibförderung) sowie zur schulinternen Leseförderung berücksichtigt. Die genannten Maßnahmen können dabei innerhalb der Klasse stattfinden wie Differenzierung und zusätzliche Hilfen bei der Aufgabenstellung oder innerhalb der Schule, zum Beispiel in Form von Förderunterricht. Abbildung 11 zeigt die Häufigkeit genannter schulinterner Fördermaßnahmen für leseschwache Kinder zu T2 (links in der Abbildung) und für alle Messzeitpunkte ab T2 (rechts in der Abbildung). Zu T2 sind 2 von 36 leseschwachen Kindern leseauffällig. Von diesen beiden erhält zum gleichen Zeitpunkt keines eine schulinterne Förderung. Im weiteren Verlauf bis zu T7 erhält eines der beiden Kinder eine schulinterne Fördermaßnahme im Bereich Deutsch. Das zweite Kind erhält durchgehend keine schulinterne Förderung (s. Abbildung 11 rechts).

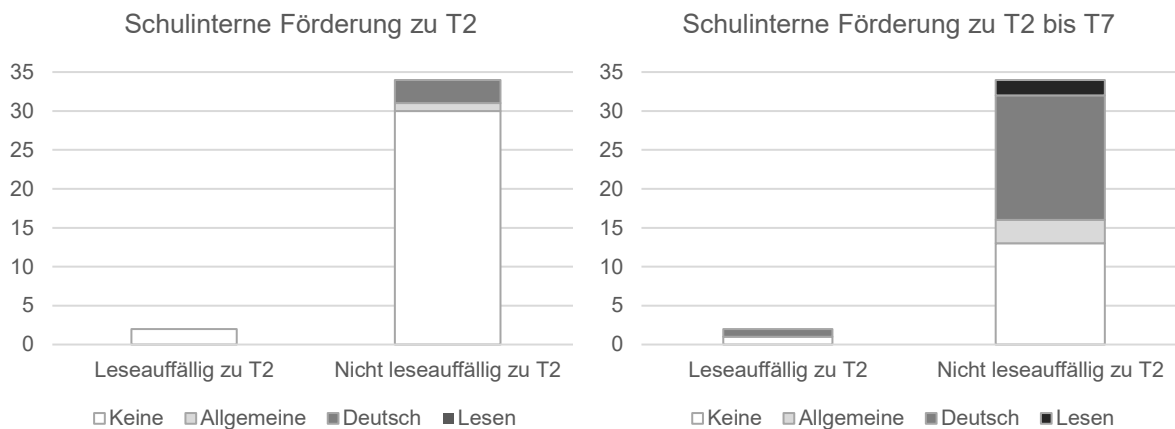


Abbildung 11. Genannte schulinterne Fördermaßnahmen zu T2 (links) und zu T2 bis T7 (rechts) für Kinder, die zu T2 leseschwach sind

Von den unauffälligen leseschwachen Kindern erhält die Mehrheit zu T2 keine schulinterne Förderung. Es wird für keins der Kinder eine schulinterne Lesefördermaßnahme genannt. Im weiteren Verlauf bis zu T7 erhält die Mehrheit ( $n = 16$ ) eine schulinterne Förderung in Deutsch. Für zwei Kinder wird im Verlauf eine schulinterne Leseförderung genannt. Dreizehn Kinder bleiben dagegen bis zum Ende der vierten Klasse ohne schulinterne Förderung.

Zu T3 sind 11 von 43 leseschwachen Kindern leseauffällig. Für kein Kind wird zu diesem Messzeitpunkt explizit eine schulinterne Leseförderung genannt (s. Abbildung 12 links). Von den leseauffälligen Kindern erhält die Mehrheit ( $n = 9$ ) allerdings eine schulinterne Förderung in Deutsch.

## 12 Ergebnisse

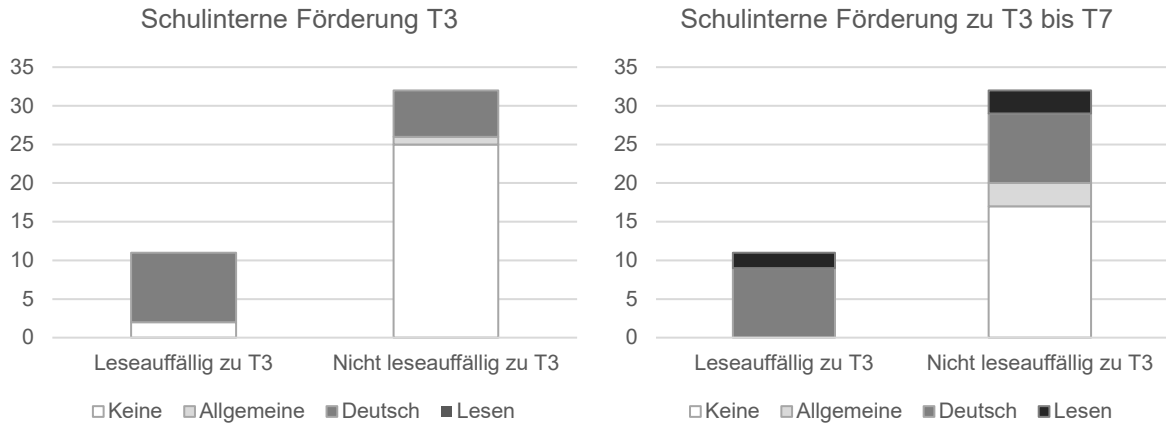


Abbildung 12. Genannte schulinterne Fördermaßnahmen zu T3 (links) und zu T3 bis T7 (rechts) für Kinder, die zu T3 leseschwach sind

Von den Kindern, die nicht auffallen, erhalten zu T3 sechs Kinder Förderung in Deutsch und ein Kind allgemeine Förderung. Die Mehrheit der unauffälligen Kinder erhält keine Förderung. Alle zu T3 leseschwachen Kinder, die den Lehrkräften leseauffällig geworden sind, erhalten in der Folge zu mindestens einem Messzeitpunkt schulinterne Förderung in Deutsch oder Leseförderung (s. Abbildung 12 rechts).

Auch von den leseschwachen Kindern, die zu T3 unauffällig gewesen sind, erhalten 38 Prozent zu mindestens einem der Messzeitpunkte Förderung im Lesen ( $n = 3$ ) oder in Deutsch ( $n = 9$ ). Drei Kinder erhalten nur allgemeine schulinterne Förderung. Die übrigen unauffälligen Kinder ( $n = 17$ ) erhalten durchgehend keine schulinterne Förderung.

Insgesamt sind 62 Kinder zu Beginn der Schulzeit, also Mitte bzw. Ende der ersten Klasse, schwach im Lesen. Hiervon sind zu T2 fünf und zu T3 14 Kinder leseauffällig geworden. Schulinterne Förderung in Deutsch haben zu jedem Messzeitpunkt zwischen 3 und 14 dieser Kinder erhalten (s. Tabelle 91). Schulinterne Leseförderung haben allerdings nur ein bis vier Kinder erhalten. Aufgrund dieser sehr geringen Anzahl an Nennungen von spezifischen Fördermaßnahmen im Bereich Lesen kann die Hypothese nicht wie ursprünglich geplant ausgewertet werden.

Tabelle 91

### Häufigkeit von schulinternen Fördermaßnahmen für (zu T2 oder T3) leseschwache Kinder

Messzeitpunkt	T2	T3	T4	T5	T6	T7
Keine Förderung	54	44	43	45	40	34
Allgemeine schulinterne Förderung	2	1	4	4	1	-
Schulinterne Förderung in Deutsch	5	17	10	7	14	3
Schulinterne Leseförderung	1	-	-	1	1	4
Keine Information*	-	-	5	5	6	21

Anmerkungen. \*Hierunter fallen Kinder, die die Klasse zwischenzeitlich verlassen haben, sowie zu T7 zusätzlich die Kinder aus den zwei Klassen, die nicht an der Nacherhebung teilgenommen haben.

Oftmals berichteten die Lehrkräfte nicht spezifisch über Lese- oder Rechtschreibförderung, sondern allgemeiner über Deutschförderung. So haben zu jedem Messzeitpunkt 3 bis 17 dieser Kinder eine schulinterne Förderung in Deutsch erhalten, welche sowohl Unterstützung im Lesen oder im Rechtschreiben oder in beiden Bereichen beinhalten kann. Daher wird zur Beantwortung der Hypothese auf die Nennung spezifischer Fördermaßnahmen in Deutsch wie auch spezifisch im Lesen zurückgegriffen. Leseschwache Kinder, für die schulinterne Förderung in Deutsch oder schulinterne Leseförderung – im weiteren Verlauf zusammengefasst als *schulinterne Deutschförderung* bezeichnet – genannt wurden, werden mit ebenfalls leseschwachen Kindern verglichen, für die keine solche Förderung genannt wurde.

#### 12.6.1.1 Erfolg schulinterner Deutschförderung

Von den 62 leseschwachen Kindern erhalten zu T2 insgesamt sechs Kinder eine schulinterne Fördermaßnahme in Deutsch bzw. im Lesen (s. Tabelle 91). Zwei Kinder erhalten eine andere schulinterne Förderung. Zu T3 sind es 17 Kinder, die eine schulinterne Förderung in Deutsch erhalten. Kein Kind erhält explizit schulinterne Leseförderung. Ein Kind erhält eine schulinterne Förderung, die nicht im Bereich Deutsch stattfindet. Der Messzeitpunkt T4 wird ebenfalls berücksichtigt, da Kinder, die zu T3 aufgefallen sind, eventuell erst in der Folge Förderung erhalten haben: Zu T4 erhalten zehn Kinder eine schulinterne Förderung in Deutsch. Für kein Kind wird explizit eine schulinterne Leseförderung genannt. Weitere vier Kinder haben eine andere schulinterne Förderung erhalten.

Insgesamt betrachtet erhalten von den zu Schulbeginn 62 leseschwachen Kindern 26 eine schulinterne Deutschförderung zu T2, T3 und/oder T4. Dies entspricht einem Anteil von 44 Prozent. Über drei Kinder liegen keine vollständigen Informationen vor.

Für 31 Kinder, die zu T2 oder T3 leseschwach waren, liegen auch zu T7 noch vollständige Daten vor. Von diesen haben 14 Kinder zu T2, T3 und/oder T4 eine schulinterne Deutschförderung erhalten und 17 Kinder keine. Die Ausgangsleseleistungen dieser beiden Gruppen unterscheiden sich weder zu T2 noch zu T3 signifikant voneinander (s. Tabelle 92).

Tabelle 92

#### *Ausgangsleseleistungen der Kinder mit und ohne schulinterne Deutschförderung*

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Wortlesen T2	14	39.71 (4.74)	17	38.02 (7.46)	0.766 (27.43)	.45	-0.26 [-1.01, 0.48]
Wortlesen T3	14	39.34 (3.70)	16	39.84 (6.17)	-0.265 (28)	.79	0.10 [-0.65, 0.85]

*Anmerkungen.*  $M_1$  = Kinder mit schulinterner Deutschförderung.  $M_2$  = Kinder ohne schulinterne Deutschförderung

Im nächsten Schritt wird überprüft, ob die Kinder mit schulinterner Deutschförderung im Vergleich zu Kindern ohne eine solche Förderung ihre Lesekompetenz steigern konnten. In ihrer Wortleseleistung unterscheiden sich die beiden Gruppen nicht signifikant voneinander zu T7 (s. Tabelle 93). Beide erreichen in diesem Subtest jeweils durchschnittliche Werte.

Tabelle 93

*Leseleistungen zu T7 für leseschwache Kinder mit und ohne Deutschförderung*

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Wortlesen	14	41.18 (8.37)	17	46.23 (7.16)	-1.811 (29)	.08	0.65 [-0.10, 1.41]
Satzlesen	14	35.92 (6.41)	17	44.66 (7.15)	-3.546 (29)	<b>.00**</b>	1.28 [0.47, 2.09]
Textlesen	14	38.85 (6.74)	16	44.30 (8.63)	-1.907 (28)	.07	0.70 [-0.07, 1.47]

Anmerkungen.  $M_1$  = Kinder mit schulinterner Deutschförderung.  $M_2$  = Kinder ohne schulinterne Deutschförderung. \* $p < .05$ , \*\* $p < .01$ . Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Betrachtet man komplexere Leseleistungen, ergibt sich ein signifikanter Unterschied zwischen den beiden Gruppen für das Satzlesen, nicht aber für das Textlesen (s. Tabelle 93). Der Mittelwertunterschied für das Satzlesen entspricht mit  $d = 1.28$  einem großen Effekt, allerdings entgegen der erwarteten Richtung. Kinder, die zu T2 oder zu T3 leseschwach waren und in der Folge schulinterne Deutschförderung erhalten haben, verfügen zum Ende der vierten Klasse über eine signifikant niedrigere Leistung im sinnentnehmenden Lesen von Sätzen als ebenfalls leseschwache Kinder, die keine solche Förderung erhalten haben.

Da zur Berechnung allerdings keine Kontrollgruppe gebildet wurde, werden zusätzlich mögliche Einflussfaktoren für beide Gruppen betrachtet. Die leseschwachen Kinder mit schulinterner Deutschförderung sind im Schnitt ein knappes halbes Jahr älter als die Kinder ohne diese Fördermaßnahmen (s. Tabelle 94). Ebenso unterscheiden sich die Gruppen in der Rechtschreibleistung und im Selbstkonzept für das Fach Deutsch. Während die Kinder mit schulinterner Deutschförderung allerdings signifikant schlechtere Leistungen im Schreiben erzielt haben, verfügen sie über ein höheres Selbstkonzept im Fach Deutsch als die Kinder ohne schulinterne Deutschförderung.

Tabelle 94

*Mittelwerte für mögliche Einflussfaktoren*

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Alter	14	6.89 (0.52)	17	6.48 (0.35)	2.554 (21.71)	.02*	-0.96 [-1.74, -0.18]
Logisches Denken	14	48.07 (7.98)	16	50.79 (11.36)	-0.748 (28)	.46	0.27 [-0.48, 1.03]
Silbensegmentation T2	14	50.42 (11.03)	17	51.26 (7.31)	-0.245 (21.79)	.81	0.09 [-0.65, 0.83]
Buchstabenkenntnis T2	14	42.00 (8.25)	17	45.24 (9.82)	-0.981 (29)	.34	0.35 [-0.39, 1.10]
LBZ T1	13	44.18 (9.39)	16	48.39 (6.93)	-1.389 (27)	.18	0.52 [-0.26, 1.30]
Rechtschreibung T1	13	43.26 (6.36)	16	48.23 (5.79)	-2.202 (27)	.04*	0.82 [0.02, 1.62]
SK Deutsch T2	14	48.96 (9.31)	17	41.47 (6.90)	2.570 (29)	.02*	-0.93 [-1.70, -0.15]

Anmerkungen.  $M_1$  = Kinder mit schulinterner Deutschförderung.  $M_2$  = Kinder ohne schulinterne Deutschförderung. LBZ = Laut-Buchstaben-Zuordnung. SK = Selbstkonzept. Rechtschreibung zu T1 erhoben mit „Schreiben nach Bildern“. \* $p < .05$ , \*\* $p < .01$ . Nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode (Benjamini & Hochberg, 1995) sind keine Werte mehr signifikant.

Hinsichtlich der Klassenverteilung unterscheiden sich die Kinder nicht signifikant voneinander. Leseschwache Kinder einzelner Klassen erhalten weder häufiger noch seltener schulinterne Deutschförderung als leseschwache Kinder anderer Klassen ( $p = .24$ , Fishers exakter Test bei zweiseitiger Testung). Auch in der Verteilung der Familiensprache und des

sozioökonomischen Status gibt es keine signifikanten Unterschiede ( $p = .22$  für Familiensprache und  $p = .65$  für sozioökonomischen Status, Fishers exakter Test bei zweiseitiger Testung).

In Bezug auf das Geschlecht ergibt sich dagegen ein signifikanter Unterschied. Von den 31 untersuchten Kindern sind 18 Jungen und 13 Mädchen. Nur knapp ein Viertel der leseschwachen Mädchen ( $n = 3$ ), aber 61 Prozent der leseschwachen Jungen ( $n = 11$ ) erhalten Deutschförderung. Leseschwache Mädchen erhalten entsprechend signifikant seltener schulinterne Deutschförderung als leseschwache Jungen ( $p = .04$ ,  $X^2 = 4.409$ ,  $df = 1$ ).

#### 12.6.1.2 Erfolg von Förderunterricht Deutsch

Eine Möglichkeit, einem leseschwachen Kind gezielt und regelmäßig Unterstützung zukommen zu lassen, ist der Förderunterricht an den Schulen. Dies sind zusätzliche Stunden, in denen die Kinder meist in kleineren Gruppen gefördert werden. Zum ersten Messzeitpunkt wird noch nicht über Förderunterricht berichtet. Nur zu T5 und T7 erhalten 3 der 31 leseschwachen Kinder lesespezifischen Förderunterricht. Bei beiden Messzeitpunkten handelt es sich um dieselben Kinder. Zu allen anderen Messzeitpunkten wird kein lesespezifischer Förderunterricht genannt (s. Tabelle 95). Diese drei sowie elf weitere Kinder haben über den Verlauf der ersten bis zur vierten Klasse Förderunterricht in Deutsch erhalten. 17 Kinder haben zu keinem Messzeitpunkt Förderunterricht in Deutsch erhalten.

Tabelle 95

#### Häufigkeit von Förderunterricht für zu T2 oder T3 leseschwache Kinder

Messzeitpunkt	T2	T3	T4	T5	T6	T7
Kein Förderunterricht	29	24	26	28	25	27
Allgemein Deutsch	2	7	5	2	6	1
Lesespezifisch	0	0	0	3	0	3

Anmerkungen. Allgemein Deutsch beinhaltet nicht lesespezifischen Förderunterricht.

Zur Prüfung der Hypothese, dass eine Teilnahme am Förderunterricht Deutsch zu Schulbeginn die Leseentwicklung positiv beeinflusst, werden analog zu Kapitel 12.6.1.1 die Messzeitpunkte Mitte der ersten bis Anfang der zweiten Klasse (T2 bis T4) herangezogen. Zwölf Kinder haben zu mindestens einem dieser Zeitpunkte Förderunterricht Deutsch erhalten. Die übrigen 19 Kinder haben keinen Förderunterricht in Deutsch erhalten. Im Vergleich zur Gruppe mit schulinterner Deutschförderung befinden sich zwei Kinder mehr in der Gruppe, die keine Förderung erhalten haben.

Tabelle 96

#### Ausgangsleseleistungen der Kinder mit und ohne Förderunterricht Deutsch

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Wortlesen T2	12	39.27 (4.98)	19	38.47 (7.18)	0.336 (29)	.74	-0.12 [-0.88, 0.63]
Wortlesen T3	12	39.82 (3.53)	18	39.47 (6.00)	0.183 (28)	.86	-0.07 [-0.83, 0.70]

Anmerkungen.  $M_1$  = Kinder mit Förderunterricht Deutsch.  $M_2$  = Kinder ohne Förderunterricht Deutsch

## 12 Ergebnisse

Es gibt keine signifikanten Unterschiede in den Ausgangsleseleistungen zu T2 und zu T3 zwischen den Kindern, die Förderunterricht Deutsch erhalten haben und den Kindern, die keinen erhalten haben (s. Tabelle 96). Beide Gruppen haben insgesamt sowohl zu T2 wie auch zu T3 unterdurchschnittliche Leseleistungen.

Tabelle 97

Leseleistungen zu T7 für Kinder mit und ohne Förderunterricht Deutsch

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$d$ [95 %-KI]
Wortlesen	12	40.49 (7.39)	19	46.13 (7.79)	-1.998 (29)	.06	0.74 [-0.04, 1.51]
Satzlesen	12	36.34 (6.41)	19	43.48 (7.86)	-2.637 (29)	<b>.01*</b>	0.97 [0.18, 1.77]
Textlesen	12	38.51 (7.17)	18	43.92 (8.23)	-1.853 (28)	.07	0.69 [-0.09, 1.48]

*Anmerkungen.*  $M_1$  = Kinder mit Förderunterricht Deutsch.  $M_2$  = Kinder ohne Förderunterricht Deutsch, \* $p < .05$ , \*\* $p < .01$ . Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Im basalen Wortlesen zu T7 unterscheiden sich die beiden Gruppen nicht signifikant voneinander (s. Tabelle 97). Auf der höheren Ebene des Leseverständnisses zeigt sich nur im Satzlesen ein signifikanter Unterschied. Hier liegt mit  $d = 0.97$  ein großer Effekt vor, allerdings entgegen der Hypothese. Kinder, die Förderunterricht Deutsch erhalten haben, zeigen eine signifikant niedrigere Leistung im Satzlesetest. Im Textlesen unterscheiden sich beide Gruppen nicht signifikant voneinander.

Tabelle 98

Mittelwertunterschiede für mögliche Einflussfaktoren

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t/F$ (df)	$p$	$d$ [95 %-KI]
Alter	12	6.95 (0.53)	19	6.48 (0.34)	2.721 (16.74)	.02*	-1.11 [-1.91, -0.30]
Logisches Denken	12	48.01 (8.61)	18	50.52 (10.73)	-0.675 (28)	.51	0.25 [-0.51, 1.02]
Silbensegmentation T2	12	48.68 (10.99)	19	52.27 (7.52)	-0.995 (17.54)	.33	0.40 [-0.36, 1.16]
Buchstabenkenntnis T2	12	42.22 (8.89)	19	44.75 (9.40)	-0.745 (29)	.46	0.27 [-0.48, 1.03]
Laut-Buchstaben-Zuordnung T1	11	44.99 (10.05)	18	47.43 (7.09)	-0.768 (27)	.45	0.29 [-0.50, 1.08]
Rechtschreibung T1	11	43.58 (6.91)	18	47.48 (5.86)	-1.627 (27)	.12	0.62 [-0.18, 1.43]
Selbstkonzept Deutsch T2	12	50.60 (8.94)	19	41.22 (6.64)	3.349 (29)	.00**	-1.23 [-2.05, -0.42]

*Anmerkungen.*  $M_1$  = Kinder mit Förderunterricht Deutsch.  $M_2$  = Kinder ohne Förderunterricht Deutsch, Rechtschreibung erhoben mit „Schreiben nach Bildern“. \* $p < .05$ , \*\* $p < .01$

Auch an dieser Stelle sollen die möglichen Einflussfaktoren betrachtet werden. Tabelle 98 gibt die Mittelwertunterschiede für die Kinder mit und ohne Förderunterricht an. Es zeigen sich zwei signifikante Unterschiede: Kinder mit Förderunterricht sind im Schnitt ein knappes halbes Jahr älter. Zu T2 verfügen sie außerdem über ein signifikant höheres Selbstkonzept

in Deutsch. Im Gegensatz zur Gruppe mit schulinterner Deutschförderung ergibt sich kein signifikanter Unterschied in den Rechtschreibleistungen.

Hinsichtlich der Klassenverteilung unterscheiden sich die Kinder mit und ohne Förderunterricht Deutsch nicht signifikant voneinander ( $p = .19$ , Fishers exakter Test, zweiseitige Testung). Die Verteilungen der Familiensprache und des sozioökonomischen Status unterscheiden sich ebenfalls nicht signifikant ( $p = .06$  für Familiensprache und  $p = 1.00$  für sozioökonomischen Status, Fishers exakter Test bei zweiseitiger Testung).

In Bezug auf das Geschlecht ergibt sich ein signifikanter Unterschied. Von den 31 untersuchten Kindern sind 18 Jungen und 13 Mädchen. Es erhalten nur 15 Prozent der Mädchen (2 von 13), aber 56 Prozent der Jungen (10 von 18) Förderunterricht. Leseschwache Mädchen erhalten signifikant seltener Förderunterricht in Deutsch als leseschwache Jungen ( $p = .02$ ,  $X^2 = 5.134$ ,  $df = 1$ ).

### 12.6.1.3 Schulinterne Deutschförderung nach statistischem Matching

Aufgrund der gefundenen Gruppenunterschiede hinsichtlich möglicher Einflussfaktoren wird über das statistische Matching mit dem Propensity Score-Verfahren nach einer Kontrollgruppe gesucht, welche hinsichtlich der Ausgangsleseleistungen und möglicher Einflussfaktoren parallelisiert wird.

Analog zu Fragestellung 2 werden Geschlecht, Alter, logisches Denken, der sozioökonomische Hintergrund, die Familiensprache, das Selbstkonzept im Fach Deutsch sowie die Klassenzugehörigkeit berücksichtigt. Für die schulischen Leistungen werden Silbensegmentation und Buchstabenkenntnis zu T2 einbezogen sowie die Laut-Buchstaben-Zuordnung und Schreiben nach Bildern zu T1. Die Leistungen im Wortlesen sowie die Verteilung von leseschwachen und nicht leseschwachen Kindern zu T2 und T3 werden ebenfalls in das Modell aufgenommen. Auf Seiten der Lehreraussagen wird die Leseauffälligkeit zu T2 und T3 berücksichtigt. Für das Matching wird zur Bestimmung des Propensity Scores eine logistische Regression auf die interessierende Variable für die Treatmentgruppe, also die schulinterne Deutschförderung, gerechnet.

Die ursprüngliche Kontrollgruppe besteht aus 113 Kindern, die zu T2 bis T4 keine schulinterne Deutschförderung erhalten haben. Die Treatmentgruppe besteht aus 29 Kindern, davon 19 Jungen und 10 Mädchen, die zu T2, T3 oder T4 schulinterne Deutschförderung erhalten haben. Im Datensatz enthalten 33 Prozent der Fälle fehlende Werte. Die Imputation der fehlenden Daten erfolgt analog zu Fragestellung 2 mit  $m = 40$  Imputationen. Das beste Matching ergab sich für die „nearest neighbour“-Methode mit Ratio 2:1 mit Replacement. Die daraus resultierende Kontrollgruppe besteht aus 27 Kindern, davon 20 Jungen und 7 Mädchen. Die größte standardisierte Mittelwertdifferenz liegt bei 0.264 für die Variable Alter. Die standardisierte Mittelwertdifferenz über alle Variablen sinkt von 1.345 auf 0.072.

## 12 Ergebnisse

Tabelle 99

*Mittelwerte vor und nach dem Matching für Kinder mit bzw. ohne schulinterne Deutschförderung*

Variable	Prä-Test				Post-Test			
	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>t (df)</i>
Wortlesen T2								
KG	50.75	8.57	.00**	2.999 (138.04)	45.32	7.59	.97	-0.035 (52.11)
TG	45.40	8.55			45.40	8.55		
Wortlesen T3								
KG	50.84	8.05	.01*	2.718 (138.04)	47.31	7.13	.60	0.522 (52.11)
TG	46.16	9.11			46.16	9.11		
Alter								
KG	6.60	0.41	.07	-1.826 (138.04)	6.64	0.46	.35	-0.938 (52.11)
TG	6.75	0.42			6.75	0.42		
Logisches Denken								
KG	52.49	9.51	.01*	2.780 (138.04)	47.87	7.34	.72	0.354 (52.11)
TG	47.14	8.13			47.14	8.13		
Silbensegmentation T2								
KG	50.99	8.50	.25	1.165 (138.04)	50.00	8.85	.65	0.455 (52.11)
TG	48.82	10.42			48.82	10.42		
Buchstabenkenntnis T2								
KG	51.56	9.00	.00**	3.537 (138.04)	45.15	8.20	.98	0.019 (52.11)
TG	45.11	7.70			45.11	7.70		
Laut-Buchstaben-Zuordnung T1								
KG	52.19	8.06	.00**	3.193 (137.94)	49.55	9.12	.25	1.169 (52.04)
TG	46.59	9.75			46.59	9.75		
Schreiben nach Bildern T1								
KG	51.98	8.30	.00**	3.339 (138.04)	48.43	6.89	.28	1.089 (52.11)
TG	46.26	7.96			46.26	7.96		
Selbstkonzept Deutsch								
KG	48.20	8.17	.23	-1.216 (138.04)	46.72	8.57	.12	-1.582 (52.11)
TG	50.27	8.25			50.27	8.25		

*Anmerkungen.* TG = Treatmentgruppe, KG = Kontrollgruppe. Werte basieren auf den gepoolten Ergebnissen der 40 Imputationen. \* $p < .05$ , \*\* $p < .01$

Tabelle 99 zeigt die Mittelwertunterschiede vor und nach dem Matching. Vor dem Matching gibt es bei sechs der neun Variablen signifikante Unterschiede. Die Ergebnisse der entsprechenden Leistungstests fallen in der Treatmentgruppe dabei signifikant niedriger aus. Auch das logische Denken liegt signifikant unter dem Wert der übrigen Kinder. Nach dem Matching gibt es zwischen der Treatment- und der Kontrollgruppe keine signifikanten Unterschiede. Die Ergebnisse für die kategorialen Variablen werden in Tabelle 100 berichtet. Es zeigt sich vor dem Matching für die beiden Variablen Leseauffälligkeit zu T3 und Leseschwäche zu T2 oder T3 ein signifikanter Unterschied zwischen der Treatmentgruppe und den übrigen Kindern. Nach dem Matching gibt es keine signifikanten Unterschiede zwischen der Treatment- und der Kontrollgruppe. Beide Gruppen unterscheiden sich entsprechend nicht im Hinblick auf die Häufigkeit von Leseschwäche und Leseauffälligkeit.

Tabelle 100

Verteilung der kategorialen Variablen für Kinder mit bzw. ohne schulinterne Deutschförderung

Variable	Prä-Test			Post-Test		
	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>	<i>p</i>	<i>F / X<sup>2</sup></i>	<i>df</i>
Klasse	.14	14.685	10	.95+	-	-
Familiensprache	.06	2.844	2	.41	0.900	2
Sozioökonomischer Status	.31	1.162	2	.88	0.132	2
Geschlecht	.36	0.823	1	.69	0.164	1
Leseauffällig T2	.77	0.083	1	1.00	0.000	1
Leseauffällig T3	.00**	15.865	1	.07	3.245	1
Leseschwach zu T2 oder T3	.00**	14.137	1	.25	1.331	1

Anmerkungen. \**p* < .05, \*\**p* < .01, +Fisher-Freeman-Halton-Test

Mit der gefundenen Kontrollgruppe wird geprüft, ob Kinder, die zu T2, T3 oder T4 eine schulinterne Deutschförderung erhalten haben, zu T7 über eine bessere Leseleistung verfügen als die Kinder, die eine solche Förderung nicht erhalten haben. Die Ausgangsleseleistungen der beiden Gruppen unterscheiden sich nicht (s. Tabelle 99). Sowohl zu T2 wie auch zu T3 liegen die Leseleistungen der Treatment- und der Kontrollgruppe im unteren durchschnittlichen Bereich mit Werten zwischen 45 und 47. Dies trifft auch auf die Leistungen im Wortlesen zum Ende der vierten Klasse zu (s. Tabelle 101).

Tabelle 101

Leseleistungen zu T7 für die Treatment- und die Kontrollgruppe

	<i>n<sub>1</sub></i>	<i>M<sub>1</sub> (SD)</i>	<i>n<sub>2</sub></i>	<i>M<sub>2</sub> (SD)</i>	<i>t (df)</i>	<i>p</i>	<i>d [95 %-KI]</i>
Wortlesen	29	46.69 (9.72)	27	47.06 (7.51)	-0.162 (54)	.87	0.04 [-0.49, 0.58]
Satzlesen	29	44.63 (10.96)	27	46.89 (6.65)	-0.941 (46.68)	.35	0.25 [-0.29, 0.79]
Textlesen	29	42.79 (7.59)	26	49.39 (8.97)	-2.955 (53)	<b>.00**</b>	0.80 [0.24, 1.36]

Anmerkungen. *M<sub>1</sub>* = Kinder mit Deutschförderung. *M<sub>2</sub>* = Kinder ohne Deutschförderung, \**p* < .05, \*\**p* < .01. Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Die Kinder der Treatmentgruppe und der Kontrollgruppe unterscheiden sich nicht signifikant voneinander. Tabelle 101 sind ebenfalls die Mittelwerte für das sinnentnehmende Satz- und Textlesen zu T7 zu entnehmen. Das sinnentnehmende Lesen auf Satzebene unterscheidet sich in den beiden Gruppen nicht signifikant voneinander. Im Textlesen zeigt sich dagegen ein signifikanter Unterschied. Mit *d* = 0.80 liegt ein großer Effekt entgegen der erwarteten Richtung vor. Die Kontrollgruppe erzielt signifikant bessere Leistungen als die Treatmentgruppe.

#### 12.6.1.4 Erfolg der durchgeführten Maßnahmen

Die Hypothese konnte nicht wie ursprünglich geplant ausgewertet werden, da zu wenig lese-spezifische Maßnahmen genannt wurden. Geprüft wurden stattdessen die schulinterne Deutschförderung und Förderunterricht Deutsch für Kinder mit Leseschwäche zu T2 oder T3. Zusätzlich wurde über statistisches Matching ein Vergleich zwischen Kindern mit und ohne

schulinterne Deutschförderung durchgeführt, wobei beide Vergleichsgruppen sich in ihrer Zusammensetzung hinsichtlich der Leseschwäche und Leseauffälligkeit nicht unterscheiden, auch wenn beide eine im unteren Durchschnitt befindliche Leseleistung zeigen.

Auch für die Subgruppe der Kinder mit Förderunterricht Deutsch wurde versucht, über statistisches Matching eine Kontrollgruppe zu finden. Dies gelang nicht, da selbst nach Eliminierung einiger Einflussfaktoren noch signifikante Unterschiede zwischen der Kontroll- und der Treatmentgruppe bestehen blieben. Daher wird auf die Berechnung an dieser Stelle verzichtet.

In keinem überprüften Bereich verfügten die Kinder mit entsprechender Förderung zum Ende der vierten Klasse über bessere Lesefertigkeiten als die jeweilige Kontrollgruppe. Es zeigten sich stellenweise dagegen Effekte entgegen der erwarteten Richtung.

#### 12.6.2 Hypothese 4.2

*Je früher die Lehrkraft von einer Leseförderung für eine Schülerin oder einen Schüler berichtet, der bzw. die zu Beginn der Schulzeit (T2 bzw. T3) leseschwach ist, desto besser wird die Leseleistung zu T7 sein.*

Um diese Hypothese auszuwerten, wird eine Rangreihenfolge gebildet, zu welchem Messzeitpunkt die Lehrkraft zum ersten Mal von einer schulinternen Lese- bzw. Deutschfördermaßnahme für eine Schülerin oder einen Schüler berichtet. Tabelle 102 kann entnommen werden, zu welchem Zeitpunkt diese erste Nennung geschehen ist. Die Häufigkeiten für innerschulische Deutschförderung beinhalten auch jeweils die Leseförderung, sodass diese Zahlen durchgehend höher sind.

Tabelle 102

#### *Beginn der innerschulischen Lese- und Deutschförderung*

Erste Nennung der Förderung	<i>Ohne</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>	<i>T7</i>
<i>Alle Kinder</i>								
Leseförderung	264	5	4	0	2	2	1	3
Deutschförderung	204	7	18	19	13	3	12	5
<i>Zu T2 oder T3 leseschwache Kinder</i>								
Leseförderung	58	0	1	0	0	1	1	1
Deutschförderung	31	2	5	13	5	2	4	0

Anmerkungen. *Ohne* bezieht sich auf die Anzahl an Kindern ohne innerschulische Förderung. Die Angaben zur Deutschförderung beinhalten auch die Leseförderung.

Berücksichtigt man alle Kinder unabhängig von ihrer Leseleistung, wird bereits zu Beginn der ersten Klasse für fünf Kinder eine Leseförderung genannt. Dies stellt gleichzeitig die höchste Zahl an ersten Nennungen für innerschulische Leseförderung dar. Von den 264 Kindern, für die keine innerschulische Leseförderung genannt wird, liegen für 175 Kinder Daten zu allen Messzeitpunkten vor. Für 89 Kinder liegen dagegen keine durchgehenden Informationen vor,

## 12 Ergebnisse

da sie entweder zu T7 nicht mehr an der Erhebung teilgenommen haben oder vorher die Klasse verlassen haben. Bis zum Zeitpunkt des Ausscheidens wurde für diese Kinder allerdings keine entsprechende Förderung genannt.

Die häufigsten Erstnennungen für innerschulische Deutschförderung kommen zu T3 mit 19 Nennungen und zu T2 mit 18 Nennungen vor. Insgesamt wurde für 204 Kinder keine innerschulische Deutschförderung genannt, wobei für 134 dieser Kinder durchgehend Daten vorliegen und 70 Kinder vor dem Ende der Studie ausschieden, ohne dass bis dahin eine innerschulische Deutschförderung für sie genannt wurde.

Von den Kindern, die zu T2 oder zu T3 leseschwach waren, erhält jeweils ein Kind zum ersten Mal schulinterne Leseförderung zu T2, T5, T6 und T7. Von den 58 Kindern, zu denen keine schulinterne Leseförderung genannt wird, erhalten 35 durchgehend keine solche Förderung. Zu 23 Kindern liegen hierzu keine durchgehenden Informationen vor. Bis zum Zeitpunkt ihres Ausscheidens aus der Studie wurde allerdings keine schulinterne Leseförderung genannt.

Schulinterne Deutschförderung erhalten zwei leseschwache Kinder bereits zu T1. Die meisten Kinder ( $n = 13$ ) erhalten zum ersten Mal eine schulinterne Deutschförderung zu T3, also Ende der ersten Klasse. Im Verlauf der zweiten Klasse erhalten 11 leseschwache Kinder zum ersten Mal Deutschförderung. Zum Ende der vierten Klasse wird für kein weiteres Kind schulinterne Deutschförderung zum ersten Mal genannt. Für 31 der leseschwachen Kinder wird keine schulinterne Deutschförderung genannt. Davon liegen für 13 Kinder keine durchgehenden Informationen vor. Bis zu ihrem jeweiligen Ausscheiden aus der Untersuchung wurde keine entsprechende Förderung genannt. Für die übrigen 18 Kinder wurde bis zu T7 keine schulinterne Deutschförderung genannt.

Tabelle 103

*Durchschnittliche Leseleistung leseschwacher Kinder zu T7 nach Zeitpunkt der ersten Nennung der Förderung*

Erste Nennung der Förderung	<i>ohne</i>	T2	T3	T4	T5	T6	T7	$r_s$	$p$
<b>Schulinterne Leseförderung</b>									
<i>n</i>	27	-	-	-	1	-	1		
<i>M</i>	44.54	-	-	-	35.75	-	45.26	-	-
<i>SD</i>	5.18	-	-	-	-	-	-		
<b>Schulinterne Deutschförderung</b>									
<i>n</i>	14	3	7	3	-	2	-		
<i>M</i>	46.67	45.69	40.44	37.74	-	48.39	-	-.08	.79
<i>SD</i>	7.15	0.74	8.86	12.93	-	5.85	-		
	<i>ohne</i>	Klasse 1			Klasse 2				
<i>n</i>	14	10			5		-		
<i>M</i>	46.67	42.02			42.00		-	.07	.82
<i>SD</i>	7.15	7.68			11.23		-		

*Anmerkungen.* Hier sind nur die Kinder dargestellt, für die zu T7 noch Daten vorliegen. *M* = mittlere Leseleistung im Wortlesen zu T7.

## 12 Ergebnisse

Von den Kindern, die zu T2 oder T3 leseschwach sind, erhalten insgesamt nur vier eine schulinterne Leseförderung. Hiervon liegen für zwei Kinder Daten zu T7 vor. Das Kind, für welches zu T5 zum ersten Mal eine Leseförderung genannt wurde, erzielt im Wortlesen zu T7 einen T-Wert von 35.75 (s. Tabelle 103). Das zweite Kind, welches zum ersten Mal zu T7 eine Leseförderung genannt bekommen hat, erreicht einen T-Wert von 45.26. Aufgrund der extrem kleinen Stichprobe kann hierfür keine Korrelation berechnet werden.

Eine schulinterne Deutschförderung erhalten 31 Kinder, wobei die erste Nennung am häufigsten zu T3 geschieht. Für 15 dieser Kinder liegen die Ergebnisse des Lesetests zu T7 vor (s. Tabelle 103). Aufgrund der recht geringen Stichprobengröße werden die Messzeitpunkte zusätzlich noch weiter zusammengefasst in erste Klasse (T1 bis T3) und zweite Klasse (T4 bis T6). Für die vierte Klasse gibt es keine Werte. Zur Auswertung der Hypothese wurden Rangkorrelationen nach Spearman gerechnet, wobei die Kinder ohne Förderung nicht in die Berechnung eingingen. Weder bei Berücksichtigung der einzelnen Messzeitpunkte noch bei Berücksichtigung der Klassen ergeben sich nennenswerte Korrelationen.

Tabelle 104

*Beginn der Förderung und durchschnittliche Leseleistung zu T7 für alle Kinder*

Messzeitpunkt	ohne	T1	T2	T3	T4	T5	T6	T7	$r_s$	$p$
<b>Schulinterne Leseförderung</b>										
<i>n</i>	145	4	2	0	0	1	0	3		
<i>M</i>	50.39	50.39	52.44	-	-	35.75	-	48.55	-.20	.58
<i>SD</i>	9.81	6.94	10.15	-	-	-	-	11.04		
		Klasse 1			Klasse 2			Klasse 4		
<i>n</i>	145	6			1			3		
<i>M</i>	50.39	51.07			35.75			48.55	-.22	.54
<i>SD</i>	9.81	7.12			-			11.04		
Messzeitpunkt	ohne	T1	T2	T3	T4	T5	T6	T7	$r_s$	$p$
<b>Schulinterne Deutschförderung</b>										
<i>n</i>	110	4	11	8	9	0	8	5		
<i>M</i>	51.77	50.39	47.37	41.20	49.40	-	46.98	45.50	-.09	.55
<i>SD</i>	8.96	6.94	10.93	8.48	12.54	-	11.62	11.85		
		Klasse 1			Klasse 2			Klasse 4		
<i>n</i>	110	23			17			5		
<i>M</i>	51.77	45.75			48.29			45.50	.03	.85
<i>SD</i>	8.96	9.83			11.80			11.85		

*Anmerkungen.* Hier sind nur die Kinder dargestellt, für die zu T7 noch Daten vorliegen.

Betrachtet man alle Kinder, die schulinterne Lese- oder Deutschförderung erhalten haben, unabhängig vom Vorliegen einer Leseschwäche zu T2 oder T3, ergeben sich die in Tabelle 104 aufgeführten Zahlen. Es wurde wieder die Rangkorrelation nach Spearman zur Prüfung der Hypothese herangezogen. Die Kinder, die keine schulinterne Förderung erhalten haben, gingen nicht in die Berechnungen ein. Die Leseleistung im Wortlesen zu T7 zeigt für diese Stichprobe eine schwache negative, aber nicht signifikante Korrelation zum Zeitpunkt der ersten Nennung der Leseförderung ( $n = 10$ ). Das gleiche Ergebnis zeigt sich, wenn die Berechnung nicht nach Messzeitpunkt, sondern nach Klassenstufe durchgeführt wird. Für

Beginn in Klasse 2 liegt allerdings nur ein einzelner Wert vor, der unter den Werten für Beginn in Klasse 1 und Beginn in Klasse 4 liegt.

Für den Zeitpunkt der ersten Nennung einer schulinternen Deutschförderung nach Messzeitpunkt ( $n = 45$ ) ergibt sich eine sehr geringe, ebenfalls nicht signifikante Korrelation. Auch hier liegt das gleiche Ergebnis vor, wenn die Klassenstufen statt der Messzeitpunkte betrachtet werden.

Die Hypothese, dass der Zeitpunkt der ersten Nennung einer innerschulischen Förderung mit der Leseleistung zum Ende der vierten Klasse in Zusammenhang steht, muss zurückgewiesen werden. Eine frühere Nennung von Förderung steht nicht im Zusammenhang mit einer besseren Leseleistung.

### 12.6.3 Hypothese 4.3

*Leseschwache Schülerinnen und Schüler, die spezifische Förderprogramme im Lesen als Unterstützung erhalten, verfügen zum Ende der vierten Klasse (T7) über eine bessere Leseleistung als Schülerinnen und Schüler, die keine solche Förderung erhalten haben.*

Insgesamt sind 75 Kinder in der ersten bis zu Beginn der zweiten Klasse (T4) als leseschwach identifiziert worden. Die genannten Fördermaßnahmen beziehen sich alle entweder auf den Bereich Deutsch oder spezifisch auf Lesen. Pro Messzeitpunkt wurden aber nur sehr selten spezifische Fördermaßnahmen für diese Kinder genannt (s. Tabelle 105). Bei Betrachtung aller Kinder ( $n = 281$ ) wurden etwas mehr spezifische Fördermaßnahmen genannt, mit Ausnahme von T3. Zu diesem Messzeitpunkt nannte keine Lehrkraft eine spezifische Fördermaßnahme. Lesespezifische Fördermaßnahmen wurden am häufigsten zu T1 und T2 genannt.

Tabelle 105

#### *Häufigkeit spezifischer Fördermaßnahmen*

Messzeitpunkt		T1	T2	T3	T4	T5	T6	T7
Leseschwach zu T2, T3 oder T4 ( $n = 75$ )	Im Bereich Deutsch	0	0	0	0	0	1	1
	Lesespezifisch	0	1	0	0	1	0	1
	Anderer Bereich	0	1	0	0	0	0	2
Alle Kinder ( $n = 281$ )	Im Bereich Deutsch	0	0	0	1	0	1	1
	Lesespezifisch	4	6	0	0	1	0	1

Die einzelnen Nennungen der lesespezifischen Maßnahmen sowie die jeweilige Leseleistung des entsprechenden Kindes finden sich in Tabelle 106.

Die vier Nennungen zum ersten Messzeitpunkt stammen von derselben Lehrkraft. Sie wendete bei vier Kindern „differenzierte Lese-Mal-Blätter“ sowie „Zusatzmaterial“ an. Da zu diesem Zeitpunkt kein Lesetest durchgeführt wurde, kann über die Ausgangsleseleistung keine Aussage getroffen werden. Die Kinder, die zu T2 mit spezifischen Fördermaßnahmen versorgt wurden, haben durchgehend durchschnittliche Leseleistungen im Wortlesen erzielt. Nur eines der mit einer lesespezifischen Fördermethode unterstützten Kinder ist nahe an der

Grenze zu einer unterdurchschnittlichen Leseleistung mit  $T = 40.99$ . Zu T5 wurde kein Lesetest durchgeführt, sodass für diesen Zeitpunkt stattdessen die Leseleistung des Kindes zum vorherigen Messzeitpunkt T4 betrachtet wird, welche unterdurchschnittlich ist. Das Kind, welches zu T7 eine lesespezifische Förderung erhalten hat, erzielt ebenfalls eine unterdurchschnittliche Leseleistung.

Tabelle 106

*Vollständige Aussagen der Lehrkräfte für die Kinder, für die lesespezifische Fördermaßnahmen genannt wurden*

	Vollständige Antwort der Lehrkraft	Leseleistung
T1	differenzierte Lese-Mal-Blätter, Zusatzmaterial	-*
T1	differenzierte Lese-Mal-Blätter, Zusatzmaterial	-*
T1	differenzierte Lese-Mal-Blätter, Zusatzmaterial	-*
T1	differenzierte Lese-Mal-Blätter, Zusatzmaterial	-*
T2	Förderplan erstellt; tägliches Lesen und Silbentraining	57.72
T2	Förderung Deutsch (Hexe Susi)	40.99
T2	Lernspiele am PC, Lese-Mal-Blätter	50.81
T2	Lernspiele am PC, Lese-Mal-Blätter	57.72
T2	Lernspiele am PC, Lese-Mal-Blätter	54.77
T2	Lernspiele am PC, Lese-Mal-Blätter	49.24
T5	Förderstunde, Silbeneinteilen	33.79**
T7	Förderstunde: Arbeit an individueller Fördermappe aus Grundschuldiagnose von Schrödel/LRS Förderung mit Silbenstrategie	34.41

*Anmerkungen.* Leseleistung im zeitgleich erhobenen Wortlesetest des ELFE 1-6. \*Leseleistung nicht erhoben. \*\*Wortlesen zu T4.

Von den insgesamt zwölf Nennungen lesespezifischer Fördermaßnahmen wurden nur drei mit Kindern durchgeführt, die als leseschwach bezeichnet werden können, da sie zu Beginn der Schulzeit (T2 bis T4) mindestens einmal eine unterdurchschnittliche Leseleistung zeigten. Aus diesem Grund kann diese Hypothese nicht inferenzstatistisch ausgewertet werden.

## 12.7 Zusätzliche Analysen

Auf Basis der erhaltenen Ergebnisse und aufgrund theoretischer Überlegungen wurden weitere Fragestellungen abgeleitet, deren Analysen an dieser Stelle dargestellt werden. Da diese Analysen allerdings rein explorativer Natur sind (Hollenbeck & Wright, 2017, S. 16), sollen sie lediglich als Anregung zu weiterer Forschung dienen.

### 12.7.1 Urteilsakkuratheit der Lehrkräfte

In der Literatur wird die Kontaktdauer als möglicher Einflussfaktor auf die Urteilsakkuratheit der Lehrkräfte diskutiert (Oerke et al., 2016; Ready & Wright, 2011). Deshalb wird der Beginn der zweiten Klasse (T4) in einer zusätzlichen Berechnung speziell betrachtet. Zudem können Lehrkräfte mit semiformaler Diagnostik unter Umständen nur schwerwiegendere Leseschwächen feststellen, während die Akkuratheit für weniger gravierende Leseschwächen niedrig ausfällt. Daher wird in einer weiteren Analyse die Ausprägung der Leseschwäche

berücksichtigt. Die Rechtschreibleistung geht als möglicher Einflussfaktor ebenfalls in eine zusätzliche Analyse ein, da es denkbar ist, dass die Lehrkräfte Rechtschreibleistungen akkurater beurteilen können und Kinder, die sowohl im Rechtschreiben wie auch im Lesen schwach sind, eher auffallen als Kinder, die lediglich im Lesen Schwierigkeiten haben.

#### 12.7.1.1 Kontaktdauer

Zu Messzeitpunkt T4 kennen die Lehrkräfte ihre Schülerinnen und Schüler schon seit einem Jahr, sodass basierend auf dem bisherigen Forschungsstand eine ausreichende Kontaktdauer vorausgesetzt werden kann. 240 Kinder haben zu T4 den Wortlesetest bearbeitet. 37 Kinder haben eine Leseschwäche mit einem T-Wert < 40 im Wortlesen. Insgesamt werden 16 Kinder als leseauffällig benannt. Zu 15 dieser Kinder liegen auch die Ergebnisse des Wortlesetests vor. Sie haben im Mittel einen T-Wert von 39.91 ( $SD = 8.37$ ) im Wortlesetest erreicht. Die nicht leseauffällig gewordenen Kinder ( $n = 225$ ) haben im Mittel einen T-Wert von 50.68 ( $SD = 9.66$ ) erreicht. Der Levene-Test ergibt mit  $p = .47$  ( $F = 0.521$ ) keinen signifikanten Unterschied der Varianzen beider Gruppen, sodass der  $t$ -Test angewendet wird. Die Leseleistungen der beiden Gruppen unterscheiden sich signifikant voneinander ( $t = -4.209$ ,  $df = 238$ ,  $p < .001$ ,  $g = 1.12$  (95 %-KI[0.59,1.65])).

Tabelle 107

#### *Häufigkeitsverteilung der Leseschwäche und Leseauffälligkeit zu Messzeitpunkt T4*

	Leseschwach	Nicht leseschwach	Summe
Leseauffällig	10 <sup>a</sup>	5 <sup>c</sup>	15
Nicht leseauffällig	27 <sup>b</sup>	198 <sup>d</sup>	225
Summe	37	203	240

*Anmerkungen.*  $n = 240$  Kinder. Die hochgestellten Buchstaben a bis d bezeichnen die Parameter zur Berechnung des Odds Ratio und des Risk Ratio.

Tabelle 107 zeigt die Trefferquoten der Lehrkräfte. Zehn leseschwache Kinder wurden korrekt als leseauffällig genannt. Fünf nicht leseschwache Kinder wurden ebenfalls als leseauffällig bezeichnet. Diese Kinder haben T-Werte zwischen 41 und 57 erreicht. Die 27 Kinder, die nicht aufgefallen sind, aber als leseschwach identifiziert wurden, haben T-Werte zwischen 24 und 39 erreicht. Das Odds Ratio beträgt 14.67 (95 %-KI[4.66, 46.15]). Das Risk Ratio liegt bei 10.97 (95 %-KI[3.98, 30.27]). Die Wahrscheinlichkeit, als leseauffällig bezeichnet zu werden, wenn man leseschwach ist, ist demnach etwa elf Mal so hoch wie die Wahrscheinlichkeit, als leseauffällig bezeichnet zu werden, wenn man nicht leseschwach ist.

Insgesamt sind 73 Prozent der leseschwachen Kinder (95 %-KI[56 %, 86 %]) den Lehrkräften nicht aufgefallen. Der Binomialtest ergibt mit  $p < .001$  (zweiseitige Testung) einen signifikanten Unterschied zur Erwartung von 33 Prozent an nicht erkannten leseschwachen Kindern. Somit muss auch zu diesem Zeitpunkt die Nullhypothese zurückgewiesen werden, dass sich der in der vorliegenden Studie gefundene Anteil nicht von den in anderen Studien gefundenen

33 Prozent unterscheidet. Auch zu Beginn der zweiten Klasse erkennen die Lehrkräfte dieser Stichprobe signifikant mehr leseschwache Kinder als erwartet nicht.

Tabelle 108

*Häufigkeit von allgemeinen negativen bzw. lesespezifischen Auffälligkeiten bei leseschwachen Kindern zu T4*

	Nicht leseauffällig	leseauffällig	Summe
Allgemein negativ auffällig	15	10	25
Nicht negativ auffällig	12	0	12
Summe	27	10	37

Anmerkungen.  $n = 37$  leseschwache Kinder

Es wird ebenfalls geprüft, wie viele der 37 leseschwachen Kinder allgemein negativ auffällig werden. Tabelle 108 zeigt die entsprechenden Häufigkeiten. Zusätzlich zu den 10 leseauffällig gewordenen Kindern sind den Lehrkräften insgesamt 15 weitere Kinder allgemein negativ auffällig geworden. Entsprechend sind 12 Kinder den Lehrkräften gar nicht negativ auffällig geworden, was einem Anteil von 32 Prozent (95 %-KI [18 %, 50 %]) entspricht. Der Binomialtest ergibt mit  $p = 1.00$  bei zweiseitiger Testung keinen signifikanten Unterschied zur Erwartung von 33 Prozent an nicht erkannten leseschwachen Kindern. Zu T4 ergibt sich somit das gleiche Bild wie zu T3: Während der Anteil an nicht leseauffällig gewordenen leseschwachen Kindern mit etwa drei Viertel der Kinder signifikant höher liegt als 33 Prozent, fällt ein Drittel der leseschwachen Kinder auch allgemein negativ nicht auf.

#### 12.7.1.2 Ausprägung der Leseschwäche

Lehrkräfte diagnostizieren meist semiformal während ihres Unterrichts. Es ist denkbar, dass sie die Schülerinnen und Schüler, die deutlich unterdurchschnittliche Leseleistungen erzielen, korrekt identifizieren, aber nicht diejenigen, die knapp unter dem Durchschnitt liegen.

Tabelle 109

*Leseleistung nach Leseauffälligkeit bei leseschwachen Kindern*

	Aufgefallen <i>M (SD)</i>	Nicht aufgefallen <i>M (SD)</i>	<i>t (df)</i>	<i>p</i>	<i>d/g [95%-KI]</i>
T2	Leseauffälligkeit $n = 2$	30.21 (3.67) $n = 34$	-1.773 (34)	.09	1.26 [-0.21, 2.74] <sup>+</sup>
	Allgemeine Auffälligkeiten $n = 14$	34.63 (4.13) $n = 22$	0.030 (34)	.98	-0.01 [-0.71, 0.68]
T3	Leseauffälligkeit $n = 11$	34.91 (4.26) $n = 32$	-0.316 (41)	.75	0.11 [-0.58, 0.80] <sup>+</sup>
	Allgemeine Auffälligkeiten $n = 25$	34.57 (4.82) $n = 18$	-1.252 (41)	.22	0.39 [-0.24, 1.02]
T4	Leseauffälligkeit $n = 10$	34.98 (3.66) $n = 27$	0.222 (35)	.83	-0.08 [-0.82, 0.66] <sup>+</sup>
	Allgemeine Auffälligkeiten $n = 25$	34.39 (4.35) $n = 12$	-0.699 (35)	.49	0.24 [-0.46, 0.94] <sup>+</sup>

Anmerkungen. <sup>+</sup>Hedges  $g$

Daher soll nun geprüft werden, ob sich die Leseleistung der Kinder, die nicht auffällig sind, von der Leseleistung der Kinder, die von den Lehrkräften als leseauffällig bzw. allgemein schriftsprachlich auffällig bezeichnet werden, unterscheidet.

Zu keinem der drei Messzeitpunkte unterscheiden sich leseschwache auffällige und nicht auffällige Kinder voneinander in ihrer Leseleistung (s. Tabelle 109). Lediglich für die Leseauffälligkeit zu T2 liegt eine Tendenz vor, was sich in der Höhe der Effektstärke von  $g = 1.26$  widerspiegelt. Es wurden allerdings nur zwei der leseschwachen Kinder leseauffällig, sodass sich aus diesem Ergebnis keine allgemeingültige Aussage ableiten lassen kann. Dies wird auch aus der großen Spanne des Konfidenzintervalls deutlich. Insgesamt zeigt sich, dass sich in der hier vorliegenden Stichprobe die durchschnittliche Leseleistung der leseschwachen und im Lesen auffälligen Kinder nicht von der durchschnittlichen Leseleistung der unauffälligen leseschwachen Kinder unterscheidet.

### 12.7.1.3 Rechtschreibleistung

Leseschwierigkeiten können unabhängig von Rechtschreibschwierigkeiten auftreten. Dennoch ist es denkbar, dass die Rechtschreibleistung die Leseauffälligkeit beeinflusst – in dem Sinne, dass Leseschwierigkeiten bei rechtschreibstarken Kindern nicht auffallen, wie auch der umgekehrte Fall, dass rechtschreibschwachen Kindern auch Leseschwierigkeiten zugerechnet werden könnten. Daher soll geprüft werden, ob sich leseauffällige leseschwache Kinder in der Rechtschreibleistung von leseschwachen Kindern unterscheiden, die nicht leseauffällig werden. Tabelle 110 zeigt die Ergebnisse.

Leseschwache Kinder, die den Lehrkräften als solche aufgefallen sind, haben zu Messzeitpunkten T2 bis T4 unterdurchschnittliche Rechtschreibleistungen mit T-Werten von 31.63 bis 38.05. Leseschwache Kinder, die den Lehrkräften nicht aufgefallen sind, haben knapp unterdurchschnittliche bis durchschnittliche Rechtschreibleistungen mit T-Werten zwischen 39.90 und 45.72.

Tabelle 110

#### *Rechtschreibleistung von leseauffälligen und unauffälligen leseschwachen Kindern*

	Aufgefallen		Nicht aufgefallen		<i>t</i> ( <i>df</i> )	<i>p</i>	<i>g</i> [95%-KI]
	<i>M</i> ( <i>SD</i> )	<i>n</i>	<i>M</i> ( <i>SD</i> )	<i>n</i>			
T2	31.63 (7.76)	2	39.90 (6.72)	34	-1.684 (34)	.10	1.20 [-0.27, 2.67]
T3	34.33 (6.37)	11	41.96 (9.38)	32	-2.500 (41)	.02*	0.86 [0.14, 1.57]
T4	38.05 (5.74)	10	45.72 (10.98)	27	-2.094 (35)	.04*	0.76 [0.00, 1.51]

*Anmerkungen.* T2, T4: Schreiben nach Bildern, T3: DERET.

Der Unterschied zu T2 ist nicht signifikant. Allerdings sind zu diesem Messzeitpunkt nur zwei Kinder leseauffällig geworden. Die Ergebnisse für T3 und T4 sind signifikant. Diese Ergebnisse deuten darauf hin, dass leseauffällige leseschwache Kinder auch über eine niedrigere Rechtschreibleistung verfügen.

## 12 Ergebnisse

### 12.7.2 Mädchen, Jungen und Lesen

In der ersten Klasse wurden gleich viele leseschwache Jungen und Mädchen jeweils lese-auffällig. Zu Beginn der zweiten Klasse zeigte sich dagegen ein signifikanter Unterschied: Es wurden weniger leseschwache Mädchen leseauffällig, als dies für leseschwache Jungen der Fall war. Ebenso haben weniger der leseschwachen Mädchen eine schulinterne Deutschförder-ung oder Förderunterricht in Deutsch erhalten als leseschwache Jungen. Tabelle 111 ist zu entnehmen, dass sich die Leseleistungen der zu T2 oder T3 als leseschwach identifizier-ten Jungen und Mädchen zu keinem der Messzeitpunkte signifikant voneinander unterschei-den.

Tabelle 111

*Leseleistungen im Wortlesen von zu T2 oder T3 leseschwachen Mädchen und Jungen*

	Mädchen		Jungen		<i>t</i> ( <i>df</i> )	<i>p</i>	<i>d</i> [95%-KI]
	<i>M</i> ( <i>SD</i> )	<i>n</i>	<i>M</i> ( <i>SD</i> )	<i>n</i>			
T2	38.69 (6.92)	31	39.52 (6.57)	29	-0.471 (58)	.64	0.12 [-0.40, 0.64]
T3	38.20 (6.64)	30	38.48 (6.03)	31	-0.176 (59)	.86	0.04 [-0.47, 0.56]
T4	39.95 (6.66)	26	40.82 (7.12)	24	-0.448 (48)	.66	0.13 [-0.44, 0.70]
T6	41.15 (7.98)	23	39.94 (7.04)	26	0.565 (47)	.58	-0.16 [-0.74, 0.42]
T7	45.84 (8.09)	13	42.58 (7.90)	18	1.121 (29)	.27	-0.41 [-1.16, 0.34]

Die in Abschnitt 12.6.1.1 durchgeführten Berechnungen werden nun getrennt für Mädchen und Jungen wiederholt, da sich in der anschließenden Analyse möglicher Einflussfaktoren gezeigt hatte, dass anteilig mehr Jungen als Mädchen schulinterne Deutschförderung erhal-ten haben. Weder bei den zu T2 oder T3 leseschwachen Jungen noch bei den entsprechen-den leseschwachen Mädchen unterscheiden sich die Leseausgangsleistungen der Gruppen mit bzw. ohne schulinterne Deutschförderung signifikant voneinander (s. Tabelle 112).

Tabelle 112

*Ausgangsleseleistungen von leseschwachen Jungen und Mädchen mit bzw. ohne schulin-terne Deutschförderung*

	<i>n</i> <sub>1</sub>	<i>M</i> <sub>1</sub> ( <i>SD</i> )	<i>n</i> <sub>2</sub>	<i>M</i> <sub>2</sub> ( <i>SD</i> )	<i>t</i> ( <i>df</i> )	<i>p</i>	<i>d/g</i> + [95%-KI]
<i>Jungen</i>							
Wortlesen T2	11	39.80 (5.15)	7	37.31 (7.85)	0.809 (16)	.43	-0.39 [-1.43, 0.64]
Wortlesen T3	11	38.69 (3.89)	7	39.82 (7.11)	-0.437 (16)	.67	-0.21 [-0.82, 1.24]
<i>Mädchen</i>							
Wortlesen T2	3	39.48 (3.66)	10	38.51 (7.56)	0.302 (7.63)	.77	-0.13+ [-1.48, 1.22]
Wortlesen T3	3	41.72 (1.69)	9	39.86 (5.78)	0.534 (10)	.61	-0.33+ [-1.71, 0.92]

Anmerkungen. *M*<sub>1</sub> = Kinder mit schulinterner Deutschförderung. *M*<sub>2</sub> = Kinder ohne schulinterne Deutschför-derung. +Hedges *g*

In keinem der betrachteten Lesetests oder dem Lese-Lernzuwachs gibt es signifikante Un-terschiede bei Mädchen mit bzw. ohne schulinterne Deutschförderung (s. Tabelle 113). Mäd-chen mit bzw. ohne eine solche Förderung erzielen zu T7 durchschnittliche Leseleistungen im Wortlesen und im Textlesen. Im Satzlesen erreichen die Mädchen mit schulinterner

## 12 Ergebnisse

Deutschförderung unterdurchschnittliche Leistungen. Es sind allerdings zu T7 nur noch drei leseschwache Mädchen, die eine solche Förderung erhalten hatten, sodass eine extrem kleine Treatmentgruppe vorliegt.

Tabelle 113

### *Leseleistung zu T7 getrennt nach Geschlecht und Deutschförderung*

	<i>n</i> <sub>1</sub>	<i>M</i> <sub>1</sub> (SD)	<i>n</i> <sub>2</sub>	<i>M</i> <sub>2</sub> (SD)	<i>t</i> (df)	<i>p</i>	<i>d/g</i> <sup>+</sup> [95%-KI]
<i>Jungen</i>							
Wortlesen T7	11	39.09 (7.68)	7	48.06 (4.64)	-2.766 (16)	<b>.01*</b>	1.34 [0.21, 2.47] <sup>+</sup>
LZW seit T2	11	43.95 (8.80)	7	55.22 (7.23)	-2.827 (16)	<b>.01*</b>	1.37 [0.23, 2.50] <sup>+</sup>
LZW seit T3	11	45.81 (8.88)	7	56.23 (9.47)	-2.367 (16)	<b>.03*</b>	1.14 [0.04, 2.46] <sup>+</sup>
Satzlesen T7	11	35.71 (6.32)	7	45.46 (6.01)	-3.250 (16)	<b>.01*</b>	1.57 [0.41, 2.74]
Textlesen T7	11	38.23 (6.84)	7	46.19 (7.02)	-2.384 (16)	<b>.03*</b>	1.15 [0.05, 2.26]
<i>Mädchen</i>							
Wortlesen T7	3	48.81 (7.04)	10	44.94 (8.50)	0.711 (11)	.49	-0.47 <sup>+</sup> [-1.93, 0.99]
LZW seit T2	3	57.32 (12.06)	9	51.98 (8.26)	0.875 (10)	.40	-0.45 <sup>+</sup> [-1.91, 1.01]
LZW seit T3	3	56.64 (9.08)	10	51.74 (11.23)	0.685 (11)	.51	0.58 <sup>+</sup> [-2.09, 0.93]
Satzlesen T7	3	36.71 (8.15)	10	44.11 (8.12)	-1.383 (11)	.19	0.85 <sup>+</sup> [-0.54, 2.24]
Textlesen T7	3	41.13 (7.18)	9	42.83 (9.85)	-0.272 (10)	.79	0.17 <sup>+</sup> [-1.21, 1.54]

Anmerkungen. LZW = Lese-Lernzuwachs. *M*<sub>1</sub> = Kinder mit schulinterner Deutschförderung. *M*<sub>2</sub> = Kinder ohne schulinterne Deutschförderung. \**p* < .05. <sup>+</sup> Hedges *g*. Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Zu T7 liegen noch für elf Jungen mit und sieben Jungen ohne schulinterne Deutschförderung Daten vor. In allen Subtests des ELFE 1-6 sowie im Lese-Lernzuwachs gibt es signifikante Unterschiede zwischen diesen beiden Gruppen. Die jeweiligen Effektstärken liegen zwischen 1.14 und 1.57 und sind somit als groß zu bewerten, gehen allerdings nicht in die erwartete Richtung: Leseschwache Jungen, die eine Deutschförderung erhalten haben, erzielen signifikant schlechtere Ergebnisse zu T7. Die Leseleistung dieser Gruppe liegt auch zu T7 noch im unterdurchschnittlichen Bereich, während die zu T2 oder T3 leseschwachen Jungen ohne schulinterne Deutschförderung in allen Subtests durchschnittliche Leistungen erzielen mit *T*-Werten über 45. Weitere mögliche Einflussfaktoren wie Alter oder Klassenzugehörigkeit wurden an dieser Stelle allerdings nicht berücksichtigt.

### 12.7.3 Häufigkeit von außerschulischen Hilfen

Auch wenn der Schule der Bildungsauftrag zukommt, greifen Eltern ebenso auf außerschulische Hilfen zurück, um ihre Kinder zusätzlich zu unterstützen. Dies kann beispielsweise durch Nachhilfe oder auch durch Lerntherapie geschehen. Es wird hierfür wieder die Gruppe der zu T2 oder T3 leseschwachen Kinder (*n* = 62) betrachtet. Die Antworten der Lehrkräfte wurden während der Kodierung darauf geprüft, ob die Fördermaßnahme ganz sicher zur Förderung des Lesens geschieht, vermutlich (auch) zur Förderung des Lesens oder nicht zur Förderung des Lesens (s. Kapitel 11.5.3). Tabelle 114 zeigt die entsprechenden Häufigkeiten für außerschulische Hilfen.

Für keins der leseschwachen Kinder wurde spezifisch im Lesen eine außerschulische Hilfe genannt. Im Bereich Deutsch haben jeweils null bis fünf Kinder außerschulische Hilfe erhalten. Auch in Bereichen, die sich nicht auf das Lesen beziehen, gab es außerschulische Hilfe für einige dieser Kinder.

Tabelle 114

*Häufigkeit von genannter außerschulischer Hilfe für zu T2 oder T3 leseschwache Kinder*

	T1	T2	T3	T4	T5	T6	T7
Nicht im Lesen	4	1	2	1	0	1	2
Eventuell im Lesen	0	3	2	0	2	5	1
Explizit im Lesen	0	0	0	0	0	0	0
Gesamt	4	4	4	1	2	6	3

Der Begriff „Nachhilfe“ fällt nur von einer Lehrkraft einmal zu T2. „Lerntherapie“ wird nicht genannt. Ein Kind erhält allerdings zu T7 „Legasthenietherapie“. Insgesamt erhalten somit pro Messzeitpunkt null bis zehn Prozent der leseschwachen Kinder außerschulische Hilfen.

#### 12.7.4 Die Gruppe der ehemals leseschwachen Kinder

Die Ergebnisse von Fragestellung 4 zeigen, dass leseschwache Kinder mit schulinterner Deutschförderung keine positivere Leseentwicklung aufweisen als Kinder, die keine solche Förderung erhalten haben. Tabelle 115 ist zu entnehmen, wie viele der zu T2 oder T3 leseschwachen Kinder zu T6 und T7 jeweils noch leseschwach sind. Zu T6 liegen Daten von 49 Kindern vor. Hiervon haben 26 Kinder durchschnittliche Wortleseleistungen. Zu T7 sind es 21 von 31 Kindern (68 Prozent) mit einer durchschnittlichen Leistung im Wortlesen. Kein Kind hat überdurchschnittliche Leseleistungen von  $T > 60$  erreicht.

Tabelle 115

*Absolute Häufigkeiten für weit unterdurchschnittliche, unterdurchschnittliche und durchschnittliche Leseleistungen von zu T2 oder T3 leseschwachen Kindern*

	<i>n</i>	Weit unterdurchschnittlich ( $T < 30$ )	Unterdurchschnittlich ( $T < 40$ )	Durchschnittlich ( $40 \leq T \leq 60$ )
T6	49	3 (6 %)	20 (41 %)	26 (53 %)
T7	31	1 (3 %)	9 (29 %)	21 (68 %)

Es soll ein kurzer Blick auf die Frage gelegt werden, inwieweit sich die Gruppe der ehemals leseschwachen Kinder, die zu T7 durchschnittliche Leseleistungen zeigt, von den ehemals leseschwachen Kindern unterscheidet, die weiterhin eine Leseschwäche aufweisen, wobei noch einmal darauf hingewiesen wird, dass die Berechnungen in diesem Abschnitt nicht Bestandteil der ursprünglichen Fragestellung und lediglich explorativer Natur sind.

Zu T7 sind 10 Kinder, davon jeweils fünf Jungen und Mädchen, weiterhin leseschwach. Insgesamt sind 21 Kinder nicht mehr leseschwach (s. Tabelle 116). Hiervon sind acht Mädchen und 13 Jungen. Der Unterschied in Bezug auf die Geschlechterverteilung ist nicht signifikant

( $p = .70$ , Fishers exakter Test, zweiseitige Testung). Hinsichtlich der Klassenverteilung zeigt sich ebenfalls kein signifikanter Unterschied ( $p = .45$ , Fishers exakter Test, zweiseitige Testung).

Tabelle 116

*Leseleistungen der weiterhin bzw. nicht mehr leseschwachen Kinder*

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$g$ [95 %-KI]
Wortlesen T6	9	34.79 (7.58)	19	42.42 (7.12)	2.598 (26)	<b>.02*</b>	-1.02 [-1.87, -0.17]
Wortlesen T7	10	34.70 (4.72)	21	48.35 (4.82)	7.423 (29)	<b>.00**</b>	-2.78 [-3.82, -1.74]
LZW T2	10	39.45 (6.21)	21	55.37 (7.34)	5.910 (29)	<b>.00**</b>	-2.21 [-3.16, -1.26]
LZW T3	10	42.33 (7.39)	20	55.70 (7.59)	4.584 (28)	<b>.00**</b>	-1.73 [-2.62, -0.84]

Anmerkungen. LZW = Lese-Lernzuwachs.  $M_1$  = Weiterhin leseschwache Kinder.  $M_2$  = Nicht mehr leseschwache Kinder. \* $p < .05$ , \*\*  $p < .01$ . Fett gedruckte Werte sind nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode signifikant (Benjamini & Hochberg, 1995).

Es gibt in der betrachteten Gruppe kein Kind, dessen Eltern beide einer unqualifizierten Tätigkeiten nachgehen oder erwerbslos sind. Für die übrigen beiden Ausprägungen im sozioökonomischen Status gibt es keinen signifikanten Unterschied ( $p = .19$ , Fishers exakter Test, zweiseitige Testung). Auch hinsichtlich der Familiensprache unterscheiden sich die beiden Gruppen nicht ( $p = .09$ , Fishers exakter Test, zweiseitige Testung).

Tabelle 117

*Ausgangsleseleistungen der weiterhin bzw. der nicht mehr leseschwachen Kinder*

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$g$ [95 %-KI]
Wortlesen T2	10	39.52 (5.45)	21	38.43 (6.82)	-0.442 (29)	.66	0.16 [-0.60, 0.93]
Wortlesen T3	10	36.78 (3.51)	20	41.02 (5.23)	2.310 (28)	.03*	-0.87 [-1.67, -0.07]

Anmerkungen.  $M_1$  = Weiterhin leseschwache Kinder.  $M_2$  = Nicht mehr leseschwache Kinder. \* $p < .05$ , \*\*  $p < .01$ . Nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode (Benjamini & Hochberg, 1995) sind keine Werte mehr signifikant.

Tabelle 117 ist zu entnehmen, dass sich die nicht mehr leseschwachen Kinder in ihrer Leseleistung zu T2 nicht signifikant von den weiterhin leseschwachen Kindern unterscheiden haben. Beide Gruppen zeigten Mitte der ersten Klasse insgesamt unterdurchschnittliche Wortleseleistungen. Zu T3 unterschieden sich beide Gruppen dagegen signifikant. Die Kinder, die zu T7 nicht mehr leseschwach sind, verfügten zum Ende der ersten Klasse bereits über signifikant höhere Leseleistungen im Wortlesen, welche insgesamt betrachtet im unteren durchschnittlichen Bereich liegen. Dieser Unterschied ist allerdings nach der Benjamini-Hochberg-Korrektur nicht mehr signifikant.

Die Gruppe der weiterhin leseschwachen Kinder ist im Schnitt ein knappes halbes Jahr älter als die Kinder, die nicht mehr leseschwach sind (s. Tabelle 118). Gleichzeitig schneiden sie im Test zur Silbensegmentation besser ab. Beide Werte sind nach der Korrektur des Alpha-Fehler-Niveaus bei multiplen Hypothesentests allerdings nicht signifikant. In den übrigen schulischen Leistungen, im logischen Denken sowie im Selbstkonzept im Fach Deutsch unterscheiden sich die beiden Gruppen nicht signifikant voneinander.

## 12 Ergebnisse

Tabelle 118

*Mittelwertunterschiede zwischen weiterhin bzw. nicht mehr leseschwachen Kinder*

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$g$ [95 %-KI]
Alter	10	6.92 (0.52)	21	6.54 (0.41)	-2.240 (29)	.03*	0.84 [0.04, 1.63]
Logisches Denken	10	49.33 (12.12)	20	49.61 (8.87)	0.072 (28)	.94	-0.03 [-0.80, 0.74]
Silbensegmentation T2	10	56.55 (5.89)	21	48.18 (9.10)	-2.645 (29)	.01*	0.99 [0.18, 1.80]
Buchstabenkenntnis T2	10	42.43 (9.33)	21	44.42 (9.21)	0.560 (29)	.58	-0.21 [-0.98, 0.56]
LBZ T1	9	48.86 (6.17)	20	45.45 (8.97)	-1.034 (27)	.31	0.40 [-0.40, 1.21]
Rechtschreibung T1	9	46.94 (8.68)	20	45.58 (5.38)	-0.432 (10.88)	.67	0.20 [-0.60, 1.00]
SK Deutsch T2	10	45.97 (9.10)	21	44.32 (8.81)	-0.481 (29)	.63	0.18 [-0.59, 0.95]

Anmerkungen. LBZ = Laut-Buchstaben-Zuordnung. SK = Selbstkonzept.  $M_1$  = Weiterhin leseschwache Kinder.  $M_2$  = Nicht mehr leseschwache Kinder. \* $p < .05$ , \*\* $p < .01$ . Nach Korrektur des Alpha-Fehler-Niveaus mit der Benjamini-Hochberg-Methode (Benjamini & Hochberg, 1995) sind keine Werte mehr signifikant.

Es wird zusätzlich betrachtet, ob sich die Kinder, die nicht mehr leseschwach sind, in den von den Lehrkräften genannten Maßnahmen von den Kindern unterscheiden, die weiterhin leseschwach sind. Tabelle 119 sind die entsprechenden Verteilungen zu entnehmen.

Tabelle 119

*Genannte Maßnahmen für weiterhin bzw. nicht mehr leseschwache Kinder*

Maßnahme	Leseschwäche zu T7				Keine Leseschwäche zu T7				$p$
	Keine	Allgemein	Deutsch	Lesen	Keine	Allgemein	Deutsch	Lesen	
T1	7	2	0	0	19	1	0	0	.22
T2	10	0	0	0	13	1	7	0	.07
T3	6	0	4	0	14	1	6	0	.79
T4	7	1	2	0	16	1	4	0	1.00
T5	7	0	2	1	15	3	3	0	.38
T6	6	1	3	0	12	2	7	0	1.00
T7	8	0	0	2	16	0	3	4	.54

Anmerkungen. \* $p < .05$ , \*\* $p < .01$ . Fishers exakter Test, zweiseitige Testung

Es zeigt sich für keinen der sieben Messzeitpunkte ein signifikanter Unterschied (Fishers exakter Test, zweiseitige Testung). Lediglich zu T2 werden tendenziell für mehr Kinder, die zu T7 nicht mehr leseschwach sind, Maßnahmen genannt. Dieser Unterschied ist mit  $p = .07$  allerdings nicht signifikant.

Tabelle 120

*Benennungsgeschwindigkeit bei leseschwachen und nicht mehr leseschwachen Kindern*

	$n_1$	$M_1$ (SD)	$n_2$	$M_2$ (SD)	$t$ (df)	$p$	$g$ [95 %-KI]
RAN Bilder	10	42.58 (9.35)	19	44.53 (9.55)	0.525 (27)	.60	-0.20 [-0.98, 0.58]
RAN Buchstaben	10	41.06 (7.89)	19	45.80 (9.02)	1.400 (27)	.17	-0.53 [-1.32, 0.26]
RAN Gesamtscore	10	40.43 (9.10)	19	44.41 (7.89)	1.225 (27)	.23	-0.47 [-1.25, 0.32]

Anmerkungen.  $M_1$  = Weiterhin leseschwache Kinder.  $M_2$  = Nicht mehr leseschwache Kinder. \* $p < .05$ , \*\* $p < .01$ . RAN = Rapid Automated Naming.

Die Benennungsgeschwindigkeit wird als Prädiktor für spätere Leseleistungen diskutiert. Da diese in der vorliegenden Untersuchung allerdings erst zu T5 erhoben wurde, kann sie nicht

## 12 Ergebnisse

als ursächliche Variable betrachtet werden. Dennoch ist an dieser Stelle interessant, ob sie im Zusammenhang steht mit den Lesefähigkeiten der Kinder.

Tabelle 120 zeigt die Mittelwerte für die beiden Gruppen. Weder in den beiden Subtests noch im Gesamtscore unterscheiden sich die beiden Gruppen signifikant voneinander. Insgesamt zeigt sich für die Gruppe der weiterhin leseschwachen Kinder eine knapp durchschnittliche Gesamtleistung mit einem T-Wert von 40 und für die Gruppe der nicht mehr leseschwachen Kinder ein T-Wert von 44. Beide Gruppen liegen somit im unteren durchschnittlichen Bereich. Insgesamt lässt sich festhalten, dass sich die Gruppen der nicht mehr leseschwachen und der weiterhin leseschwachen Kinder in den hier untersuchten Variablen nach der Benjamini-Hochberg-Korrektur in keiner der untersuchten Variablen signifikant voneinander unterscheiden.

## 13 Diskussion

Nachdem im vorhergehenden Kapitel die erlangten Ergebnisse ausführlich dargestellt wurden, sollen diese nun abschließend kritisch betrachtet und in Bezug zum bisherigen Forschungsstand eingeordnet werden. Im Anschluss werden die Limitationen der vorliegenden Untersuchung dargestellt, bevor das Kapitel mit praktischen Implikationen abschließt.

### 13.1 Erkennen von Leseschwierigkeiten

Entsprechend der Ergebnisse von Bailey und Drummond (2006) wurden auch hier von den Lehrkräften neben Angaben zum Lesen und Schreiben Aussagen über andere Bereiche wie beispielsweise das Verhalten der Kinder, Motivation oder Sprachkenntnisse getätigt. Einige Lehrkrafturteile blieben oberflächlich mit allgemeinen Bewertungen, andere Aussagen gingen ins Detail. Es zeigte sich, dass Grundschullehrkräfte in den ersten beiden Grundschuljahren zu etwa jedem vierten Kind positive wie negative Auffälligkeiten nennen. Die negativen Auffälligkeiten überwiegen hierbei durchgehend die positiven, wobei es möglich ist, dass der Aufbau der Studie den Fokus der Grundschullehrkräfte verstärkt auf negative Auffälligkeiten gelenkt hat. Auch wenn die Fragen allgemein und offen gestellt waren („Wer fällt auf?“, „Was fällt Ihnen auf?“, gefolgt von „Was wird getan?“), könnte die Kombination einer Studie zur Prüfung der Lese- und Rechtschreibfertigkeiten mit diesen Fragen die Lehrkräfte dazu verleitet haben, sich vor allem auf die Defizite der Kinder zu konzentrieren. Allerdings wurden prozentual gesehen zu mehreren Messzeitpunkten ebenso viele oder mehr negative Aussagen über das Verhalten oder die Motivation der Kinder gemacht als Aussagen über den Schriftspracherwerb.

Das Ergebnis der IGLU-Studie 2006 mit einer Rate von 33 Prozent leseschwacher Kinder, die von den Grundschullehrkräften nicht erkannt wurden, galt als Orientierung für diese Untersuchung (Valtin et al., 2010, S. 76). In der vorliegenden Stichprobe sind deutlich mehr Kinder in der Wahrnehmung der Lehrkräfte nicht leseauffällig geworden, als erwartet worden war. Mitte des ersten Schuljahres, also wenige Monate nach Schuleintritt, fiel so gut wie kein leseschwaches Kind den Lehrkräften auf. Die Erkennungsrate stieg zum Ende der ersten Klasse an. Aber auch hier blieben etwa drei von vier leseschwachen Kindern unauffällig. Dies gilt allerdings nur für die Leseauffälligkeit im Speziellen. Werden sämtliche negative Auffälligkeiten berücksichtigt, so fallen den Lehrkräften 58 Prozent der leseschwachen Kinder auf – allerdings eben nicht spezifisch durch Leseschwierigkeiten.

Die steigende Erkennungsrate der Leseauffälligkeit im Speziellen könnte darauf hindeuten, dass die Lehrkräfte mit längerer Kontaktdauer die Kinder besser in ihren Leistungen einschätzen können. In vorhergegangenen Untersuchungen war die Kontaktdauer bereits als möglicher Einflussfaktor diskutiert worden, wobei ein halbes Jahr als ausreichend festgestellt wurde (Oerke et al., 2016, S. 43; Stang & Urhahne, 2016b, S. 215). Oerke et al. (2016) sowie Ready und Wright (2011) hatten bei Betrachtung der Längsschnittdaten eine steigende Urteilsakkuratheit gefunden, was in der hier vorliegenden Studie ebenfalls der Fall ist. In einer

zusätzlichen Analyse ergab sich allerdings bei Betrachtung der Erkennungsrate von Mitte des ersten Schuljahres bis zum Anfang des zweiten Schuljahres kein weiterer Anstieg dieser. Die Erkennungsrate für die in dieser Untersuchung teilnehmenden Lehrkräfte blieb auch nach einem Jahr Kontaktdauer unter der Erkennungsrate der IGLU-Studie. In der IGLU-Studie wurden die leseschwachen Kinder durch Vergabe einer Schulnote identifiziert. Die deutlich niedrigere Erkennungsrate in dieser Studie könnte auch hier aufgrund des Designs entstanden sein. Lehrkräfte mussten sich nicht zu jedem Kind äußern. Die Schulnote sollte dagegen für jedes Kind vergeben werden. Leseschwache Kinder, die höchstens die Schulnote „ausreichend“ erhalten hatten, galten als erkannt. Für die vorliegende Untersuchung ist eine Abweichung in dem Sinne denkbar, dass Lehrkräfte die leseschwachen Kinder nicht genannt haben, welche in ihren Augen die Schulnote „ausreichend“ erhalten würden. Eventuell werden nur die Kinder als auffällig benannt, die in den Augen der Lehrkräfte gravierendere Schwierigkeiten im Lesen haben, während Kinder, deren Leseschwäche weniger stark ausgeprägt ist, nicht genannt werden. Auch dieser Gedankengang wurde in einer zusätzlichen Analyse geprüft. Für die vorliegende Stichprobe lässt sich ein solcher Effekt für Ende des ersten Schuljahres und Anfang des zweiten Schuljahres ausschließen. Nur für Mitte des ersten Schuljahres ergab sich eine nicht-signifikante Tendenz mit einer Effektstärke von  $g = 1.26$ . Die beiden leseschwachen Kinder, die als auffällig bezeichnet wurden, hatten im Schnitt einen T-Wert von 30.21 erreicht. Die übrigen leseschwachen Kinder lagen im Durchschnitt 4.5 T-Werte darüber.

Neben der Kontaktdauer wurden weitere mögliche Einflussfaktoren betrachtet. In bisherigen Studien konnte oftmals ein Referenzgruppeneffekt gefunden werden, wonach in lesestarken Klassen häufiger Kinder als schwach bezeichnet werden als Kinder mit vergleichbaren Leistungen in durchschnittlichen Klassen (Maaz et al., 2008; Südkamp & Möller, 2009; Trautwein & Baeriswyl, 2007). Dieser Effekt trat in dieser Studie allerdings nicht auf.

Auch der sozioökonomische Status von Kindern zeigte sich in bisherigen Studien oft als Einflussfaktor auf die Urteilsakkuratheit der Lehrkräfte (Kaiser et al., 2015; Maaz et al., 2008; Ready & Wright, 2011), seltener konnte dagegen kein Einfluss festgestellt werden (Karing, Matthäi, et al., 2011). Karing, Matthäi und Artelt (2011, S. 165) sowie Maaz et al. (2008, S. 530) betrachteten den sozioökonomischen Status über das Berufsprestige, wobei der jeweils höchste Wert beider Elternteile berücksichtigt wurde. Ready und Wright (2011, S. 342) betrachteten mehrere Faktoren: Einkommen, Bildungsstand der Eltern sowie das Berufsprestige. Kaiser et al. (2015, S. 290) gingen anders vor und nahmen das kulturelle Kapital in Form des Bücherbesitzes als Merkmal für den familiären Hintergrund. In der vorliegenden Untersuchung wurde der sozioökonomische Status über die Angabe der Berufe der Eltern erfasst, wobei auch hier die jeweils höchste Einschätzung verwendet wurde. Entgegen den Erwartungen zeigte sich kein Zusammenhang mit der Leseauffälligkeit. Leseschwache Kinder mit einem höheren sozioökonomischen Status wurden entsprechend nicht seltener leseauffällig als Kinder mit einem niedrigeren Status. Allerdings wurde der sozioökonomische Status nur

in drei Stufen erfasst. Das Vorgehen bei Karing, Matthäi und Artelt (2011) sowie Maaz et al. (2008) war hingegen feingliedriger mit 16 bis 90 Abstufungen (Ganzeboom, De Graaf & Treiman, 1992).

Das Geschlecht als Einflussfaktor auf die Urteilsakkuratheit wurde bereits häufiger betrachtet, da Jungen im Verlauf der Grundschulzeit oft eine geringere Lesekompetenz ausbilden als Mädchen (McElvany & Ohle, 2015, S. 48). In dieser Studie zeigte sich ein Zusammenhang der Leseauffälligkeit mit dem Geschlecht, allerdings in der ersten Klasse nur als Tendenz. Erst zu Beginn der zweiten Klasse wurden signifikant weniger leseschwache Mädchen erkannt als leseschwache Jungen. Bisher liegen in der Forschung gemischte Ergebnisse vor, wobei bei Vorliegen eines Effekts stets Mädchen besser bewertet werden als Jungen bei gleichen Leistungen (Hinnant et al., 2009; Kaiser et al., 2015; Kuhl & Hannover, 2012). Die vorliegende Untersuchung konnte aufzeigen, dass der Effekt erst im Verlauf der ersten beiden Grundschuljahre aufzutreten scheint. Die Leseleistungen von Jungen und Mädchen unterscheiden sich zu diesen Zeitpunkten dagegen nicht. Unklar bleibt an dieser Stelle die Frage, welche Konsequenzen sich daraus im weiteren Verlauf für nicht auffällige, aber leseschwache Mädchen ergeben.

### 13.2 Durch Auffälligkeiten beeinflusste Leseentwicklung

Auch wenn nicht alle Kinder, die mit Leseschwierigkeiten auffallen, eine entsprechende Förderung erhalten, so war dennoch aufgrund bisheriger Forschung vermutet worden, dass sich Kinder durch die gesteigerte Aufmerksamkeit der Lehrkräfte in der Folge stärker verbessern würden als unauffällige Kinder (Anders et al., 2010; Diaper, 1990; Kaiser & Möller, 2017). Es wurden zur Überprüfung dieser Annahme dabei drei Einteilungen für die genannten Auffälligkeiten in den Blick genommen: im Bereich Deutsch, spezifischer nur im Lesen sowie allgemein negative Auffälligkeiten, welche auch Auffälligkeiten im Bereich Deutsch beinhalten. Auffälligkeiten im Bereich Deutsch beziehen sich sowohl auf Aussagen zur Rechtschreib- und Leseleistung der Kinder als auch auf hierfür relevante Vorläuferfertigkeiten wie die phonologische Bewusstheit und auf Schwierigkeiten mit der Sprache. Die in diesem Bereich auffällig gewordenen Kinder zeigen allerdings im Vergleich zu einer Kontrollgruppe keine Unterschiede in ihrer Leseentwicklung. Sowohl zum Ende der zweiten wie auch zum Ende der vierten Klasse unterscheiden sich die Gruppen nicht im basalen Lesen, im Leseverständnis oder im Lese-Lernzuwachs. Das gleiche Bild ergibt sich für Kinder, die den Lehrkräften spezifisch nur im Lesen auffällig geworden sind. Auch für diese gibt es keine Unterschiede zum Ende der zweiten oder zum Ende der vierten Klasse zu der jeweiligen Kontrollgruppe. Etwas differenzierter müssen dagegen die Kinder betrachtet werden, die den Lehrkräften allgemein auffällig wurden. Hierzu zählen sowohl die bereits genannten Auffälligkeiten im Bereich Deutsch, aber auch alle anderen negativen Auffälligkeiten wie beispielsweise im Verhalten, in der Motivation oder auch Schwierigkeiten in anderen Fächern. Verglichen mit der Kontrollgruppe zeigen sich für diese Kinder ebenfalls keine Unterschiede im Wortlesen oder im Lese-

Lernzuwachs zum Ende der zweiten und zum Ende der vierten Klasse. Die Kontrollgruppe erreicht allerdings sowohl zum Ende der zweiten wie zum Ende der vierten Klasse bessere Leistungen im Satzlesen. Dieses zählt ebenfalls zum basalen Lesen, ist aber ein komplexerer Vorgang als das Lesen einzelner Wörter. Das Leseverständnis der beiden Gruppen unterscheidet sich zum Ende der zweiten Klasse nicht voneinander. Zum Ende der vierten Klasse erreicht auch hier die Kontrollgruppe bessere Leistungen. Insgesamt betrachtet gibt es in dieser Stichprobe daher keine Hinweise darauf, dass sich Kinder, die ihren Lehrkräften auffällig werden, in der Folge positiver im Lesen entwickeln. Stattdessen zeigte sich vereinzelt im Satzlesen und im Textverständnis eine negativere Entwicklung, aber nur für die Kinder, die allgemein – also nicht nur spezifisch im Lesen oder im Bereich Deutsch – auffällig geworden sind.

Da es sich um eine Beobachtungsstudie handelt, konnten keine randomisierten Gruppen gebildet werden. Für die Gruppenvergleiche wurden deshalb über statistisches Matching Kontrollgruppen erstellt. Berücksichtigt wurden hierzu mehrere Einflussfaktoren wie das Alter, das Geschlecht, der sozioökonomische Status, die Familiensprache, die Klassenzugehörigkeit sowie das logische Denken und schulische Ausgangsleistungen im Bereich des Schriftspracherwerbs. Denkbar ist, dass weitere Hintergrundmerkmale, die nicht erfasst wurden, die Unterschiede bzw. das Fehlen dieser, erklären können. So war nicht vermutet worden, dass auffällig gewordene Kinder sogar eine negativere Leseleistung zeigen als unauffällige Kinder, die über die gleichen Ausgangsbedingungen verfügen. Es ist denkbar, dass andere Faktoren dafür verantwortlich sind, dass Kinder auffällig werden, und diese sich auch auf deren Leseentwicklung auswirken.

Das Matching fand über die Leseleistung einzelner Wörter statt. Es wurde in der ersten Klasse noch nicht mit den beiden Subtests zum Satz- und Textlesen gearbeitet, sodass hierüber keine Informationen vorlagen. Auffällige Kinder könnten in diesen höheren Leseleistungen bereits früher schwächere Leistungen gezeigt haben. Möglich ist, dass gerade die Kinder, die allgemein auffällig geworden sind, Schwierigkeiten in der Aufmerksamkeit und Konzentration haben. Dies würde erklären, warum sich bei diesen Kindern das Lesen auf höheren Ebenen schlechter entwickelte als bei nicht auffälligen Kindern. Da diese Faktoren allerdings nicht zu Beginn der Studie erhoben wurden, konnte ein möglicher Einfluss nicht in die Bildung der Kontrollgruppe einbezogen werden.

Eine weitere Einschränkung ist die teilweise sehr geringe Stichprobengröße. Auch wenn  $t$ -Tests mit sehr kleinen Stichproben ( $n \leq 5$ ) durchgeführt werden können (de Winter, 2013, S. 8), fällt die Power des Tests niedrig aus. Entsprechend können nur große Effekte als signifikant erkannt werden. Der kleinste Stichprobenumfang in der vorliegenden Studie ergab sich für den Vergleich von Ende der ersten Klasse leseauffällig gewordenen Kindern ( $n = 9$ ) mit ihren Leseleistungen zum Ende der vierten Klasse im Vergleich zu einer Gruppe von Kindern, die nicht leseauffällig geworden sind (ebenfalls  $n = 9$ ). Der für diesen Gruppenvergleich größte Effekt mit  $d = 0.60$  erreicht eine Power von 0.22 (siehe Anhang C). Dies spiegelt sich

auch in der Größe der Konfidenzintervalle der Effektstärken wider. Für diesen Gruppenvergleich liegt die Effektstärke mit einer Wahrscheinlichkeit von 95 Prozent zwischen -0.42 und 1.63. Eine A priori-Berechnung gibt den benötigten Stichprobenumfang bei einer statistischen Power von 0.80 zur Ermittlung eines großen Effekts mit  $n = 52$  an (siehe Anhang C). Sieben der insgesamt zwölf Gruppenvergleiche wurden mit Gruppen durchgeführt, die aus insgesamt weniger als 52 Kindern bestehen (je  $n = 18$  bis  $n = 50$ ). Dies ergab sich daraus, dass den Lehrkräften unerwartet viele Kinder im Lesen oder im Bereich Deutsch trotz niedriger Leseleistungen nicht auffällig geworden waren und die Treatmentgruppen entsprechend klein ausgefallen sind. Die übrigen fünf Gruppenvergleiche übersteigen die Grenze. Dennoch ergibt sich für alle zwölf Gruppenvergleiche ein sehr konsistentes Bild: Auffällige Kinder zeigen auch bei Betrachtung der absoluten Werte in keinem Vergleich eine positivere Leseentwicklung oder Leseleistung zum Ende der zweiten und zum Ende der vierten Klasse. Da, wo sich – auch bei Fehlen von signifikanten Unterschieden – mindestens mittlere Effektstärken von  $d = 0.50$  ergeben haben, liegen die Leistungen der Kinder, die auffällig geworden sind, stets unter den Leistungen der unauffälligen Kinder. Entsprechend gibt es in dieser Untersuchung keinen Hinweis darauf, dass auffällig gewordene Kinder eine positivere Leseentwicklung zeigen. Kinder, die allgemein auffällig geworden sind, zeigen eher eine gegensätzliche Tendenz, indem sie zum Ende der zweiten bzw. zum Ende der vierten Klasse geringere Leistungen in kognitiv anspruchsvolleren Aufgaben zum sinnentnehmenden Lesen zeigen. Entsprechend der Schlussfolgerung von Hesse und Latzko (2017, S. 35) muss auch hier resümiert werden, dass die Urteilsakkuratheit für sich nicht hinreichend ist und konkrete Fördermaßnahmen ebenfalls berücksichtigt werden müssen.

### 13.3 Eingeleitete (Förder-)Maßnahmen bei Auffälligkeiten

Wenn ein Kind auffällig wird, sollte eine Lehrkraft entsprechend reagieren und Unterstützungsmaßnahmen anbieten. Bisherige Untersuchungen wie die IGLU-Studie 2016 zeigten, dass leseauffälligen Kindern nur wenig Unterstützung angeboten wird und so zwei Drittel dieser Kinder ohne schulinterne Leseförderung bleiben (Bos et al., 2017, S. 25–26). Analog zu den Ergebnissen der IGLU-Studie 2016 blieben auch hier etwa zwei Drittel der Kinder, die in der ersten Klasse leseauffällig wurden, ohne eine solche Förderung. In dieser Untersuchung wurden neben der schulinternen Leseförderung weitere Formen der schulinternen Förderung berücksichtigt, so vor allem die allgemeine Deutschförderung, welche Lese- wie auch Rechtschreibförderung beinhaltet. Betrachtet man diese schulinternen Maßnahmen im Bereich Deutsch, sinkt der Anteil auf ein Drittel an leseauffälligen Kindern, welche keine solche Förderung erhalten haben.

Ebenfalls wurde betrachtet, wie viele der allgemein auffällig gewordenen Kinder eine schulinterne Förderung erhalten. Das Ergebnis ähnelt dem der leseauffälligen Kinder: Nur knapp die Hälfte der Kinder, und damit immerhin weniger als die angenommenen zwei Drittel, bleibt ohne schulinterne Förderung. Auch wenn dies auf den ersten Blick erfreulich ist, muss auch

die Kehrseite betrachtet werden: Knapp die Hälfte an allgemein auffälligen Kindern und etwa jedes dritte leseauffällige Kind erhält diesbezüglich in der Folge keine oder kaum Förderung in der Schule.

Der Fokus dieser Untersuchung lag auf der förderdiagnostischen Kompetenz der Lehrkräfte, sodass außerschulische Maßnahmen nicht berücksichtigt wurden. Sicherlich könnten solche Maßnahmen aber zu einer verbesserten Leseentwicklung von schwächeren Kindern führen. Ebenso ist es denkbar, dass eine Lehrkraft weniger Unterstützung anbietet, wenn sie zum Beispiel über stattfindende Lerntherapie außerhalb der Schule informiert ist. Außerschulische Hilfen für leseschwache Kinder wurden allerdings nur wenige genannt. Daher kann an dieser Stelle über die Gründe, warum Kinder auf der einen Seite leseauffällig werden, auf der anderen Seite aber keine innerschulische Unterstützung erhalten, nur spekuliert werden. Auch sind die Rahmenbedingungen oft ausschlaggebend. So scheinen finanzielle und zeitliche Ressourcen eine große Rolle zu spielen (Buda, 2017, S. 183; Eckerth, 2015, S. 179). Hierüber wurden allerdings keine Daten erhoben und die Lehrkräfte auch nicht dahingehend befragt.

### 13.4 Erfolg der durchgeführten (Förder-)Maßnahmen

Bisherige Forschung konnte zeigen, dass eine langfristige Förderung über mindestens 12, besser über mindestens 20 Wochen erfolgen muss, um leseschwache Kinder ausreichend fördern zu können (Balke-Melcher et al., 2016, S. 25; Galuschka et al., 2014, S. 6; Ise et al., 2012, S. 130; von Suchodoletz, 2007, S. 351). Dies kann in der vorliegenden Untersuchung nicht geprüft werden, da über die Dauer der Maßnahmen keine Angaben gemacht wurden und diese auch nicht explizit erfragt wurden. Über den Beginn liegen dagegen Informationen vor in Form des ersten Zeitpunkts, zu dem Lehrkräfte eine Förderung nennen. Studien heben hervor, dass es wichtig ist, möglichst früh mit Leseförderung zu beginnen (Fischer & Gasteiger-Klicpera, 2013, S. 78; Klicpera et al., 2013, S. 137). In dieser Untersuchung gab es allerdings keinen Zusammenhang zwischen der Leseleistung zum Ende der vierten Klasse und dem Zeitpunkt der ersten Nennung einer Leseförderung für Kinder, die in der ersten Klasse leseschwach waren. Einschränkend festzuhalten ist an dieser Stelle aber, dass entsprechende Angaben von den Lehrkräften nur für wenige Kinder gemacht wurden.

Es sollte ebenso geprüft werden, ob spezifische Leseförderprogramme zu einer besseren Leseentwicklung und zu besseren Leseleistungen zum Ende der zweiten bzw. zum Ende der vierten Klasse führen. Diese Hypothese konnte nicht ausgewertet werden, da nur für sehr wenige Kinder spezifische Fördermaßnahmen genannt wurden. An dieser Stelle zeigt sich deutlich ein Nachteil der Vorgehensweise, keine Nachfragen zu stellen, da so unklar bleibt, ob die Lehrkräfte von sich aus keine exakten Angaben gemacht haben oder ob spezifische Fördermaßnahmen tatsächlich sehr selten eingesetzt werden. Letzteres entspricht zumindest dem Erkenntnis von Fadler und Böhme (2018, S. 5) oder auch Bremerich-Vos et al. (2017, S.

287), welche zum Beispiel für Lautleseverfahren zur Leseförderung nur eine sehr geringe Bekanntheit und Verbreitung angeben.

Recht häufig wurde dagegen „Deutschförderung“ genannt, ohne dabei spezifisch auf Rechtschreibung oder Lesen einzugehen. Entsprechend wurden statt der spezifischen Fördermaßnahmen die genannten Maßnahmen, die sich allgemein auf Deutsch beziehen, betrachtet. Die Teilnahme an schulinterner Deutschförderung scheint nicht zu einer besseren Leseentwicklung zu führen. Leseschwache Kinder ohne Deutschförderung zeigen zum Ende der vierten Klasse sogar bessere Leistungen als leseschwache Kinder, die Deutschförderung erhalten haben. Lediglich im Satzlesen ist dieser Unterschied allerdings signifikant mit einer großen Effektstärke von  $d = 1.28$ . Aufgrund der Stichprobengröße und der geringen Power der Tests kann für das Wort- und Textlesen nur von einer Tendenz gesprochen werden mit mittleren Effektstärken von  $d = 0.65$  sowie  $d = 0.70$ . Es zeigt sich folglich insgesamt im Lesen eine Tendenz, aber entgegen der erwarteten Richtung.

Dieses Ergebnis muss allerdings kritisch betrachtet werden, da sich die Kinder mit und ohne Deutschförderung in einigen Faktoren unterscheiden: Auch wenn beide Gruppen in der ersten Klasse die gleiche Leseleistung zeigen, gab es bei näherer Betrachtung Unterschiede in der Rechtschreibleistung, im Selbstkonzept im Fach Deutsch, im Alter und in der Geschlechterverteilung. Entsprechend wurde auch für diesen Vergleich über statistisches Matching eine Kontrollgruppe erstellt. Die Treatmentgruppe besteht aus Kindern, die Deutschförderung erhalten haben, unabhängig davon, ob sie auch leseschwach waren. Kinder, die keine Deutschförderung erhalten haben und sich in den entsprechenden Einflussvariablen nicht von der Treatmentgruppe unterscheiden, bilden die Kontrollgruppe. Es zeigt sich für diese beiden Gruppen ein ähnliches Bild: Im basalen Lesen auf Wort- und Satzebene und im Lese-Lernzuwachs gibt es keine Unterschiede. Allerdings ist die Kontrollgruppe zum Ende der vierten Klasse auch hier signifikant besser im Textverständnis.

Da für das Matching, um eine ausreichend große Stichprobe zu haben, Kinder nur aufgrund der erhaltenen schulinternen Deutschförderung der Treatmentgruppe zugewiesen wurden, sind sowohl in der Treatment- wie auch in der Kontrollgruppe Kinder, die über durchschnittliche Leseleistungen verfügen. Entsprechend ist es denkbar, dass dieser Umstand dafür verantwortlich ist, dass keine Lesesteigerung gefunden werden konnte. Gleichwohl zeigte sich auch im Vergleich der Gruppe der leseschwachen Kinder kein signifikanter Unterschied nach Erhalt schulinterner Deutschförderung.

Insgesamt muss daher festgehalten werden, dass sich keine Hinweise ergeben haben, dass schulinterne Deutschförderung einen positiven Einfluss auf die Lesefähigkeiten der Kinder hat. Über den Inhalt, den Umfang oder die Dauer der jeweiligen Maßnahmen ist allerdings nichts bekannt.

Die zusätzliche explorative Analyse der erhobenen Daten zeigt tendenziell, dass es Unterschiede im Erfolg von Fördermaßnahmen zwischen Jungen und Mädchen geben könnte. Leseschwache Jungen mit Förderung waren im weiteren Verlauf signifikant schlechter im Lesen

als Jungen ohne Förderung. Für leseschwache Mädchen mit und ohne Deutschförderung ergab sich kein signifikanter Unterschied, aber ein mittlerer Effekt in Richtung eines besseren Lese-Lernzuwachses für die drei Mädchen, die Deutschförderung erhalten hatten. Dies sollte bei künftiger Forschung genauer betrachtet werden, da ein  $n = 3$  sicherlich nicht aussagekräftig oder gar repräsentativ ist. Durch die Untersuchung hindurch ziehen sich aber leichte Unterschiede zwischen Jungen und Mädchen, wie man auch aufgrund bisheriger Forschung annehmen kann (s. Kapitel 8.1.6).

### 13.5 Limitationen und Anregungen zu weiterer Forschung

Die vorliegende Studie hatte zum Ziel, neben der Urteilsakkuratheit auch den Aspekt der folgenden Fördermaßnahmen und die Effekte auf die Leseentwicklung zu prüfen. Dies geschah, indem neben standardisierten Tests zur Prüfung der schulischen Leistungen der Kinder auch ihre Deutsch-Lehrkräfte hinsichtlich Auffälligkeiten und Fördermaßnahmen befragt wurden. Es wurde entsprechend nur der Status Quo an einigen mittelhessischen Grundschulen betrachtet. Eine experimentelle Veränderung fand dabei nicht statt. Da die Studie über den Verlauf der vier Grundschuljahre angelegt war, wurde versucht, die Grundschullehrkräfte möglichst nicht zu beeinflussen, indem auf sämtliche Nachfragen verzichtet wurde. Dies führte allerdings auch zu entsprechend sehr allgemein gehaltenen und schwierig zu kategorisierenden Antworten wie beispielsweise: „Das Kind ist schwach“. Einige Lehrkräfte erwähnten in ihren Antworten weitere Personen, die im Klassenzimmer anwesend waren. Es wurde nicht nachgefragt, wie häufig diese Personen im Unterricht anwesend sind und inwiefern die Personen sich um alle oder nur um spezifische Kinder in der Klasse kümmerten. Sinnvoll wäre daher eine ähnliche Studie, in der gezielt detailliertere Fragen über Umfang und Inhalte der Unterstützung gestellt werden, die im Vorhinein schriftlich formuliert werden sollten, so dass auch hier keine weitere Beeinflussung über individuelle Nachfragen stattfinden kann. Denkbar wäre ebenfalls eine Ergänzung der Frage, was getan wird, um den Hinweis, dass auch konkrete Beispiele genannt werden sollten. Das Format des Interviews erscheint sehr gut geeignet, da dieses im Vergleich zu einer schriftlichen Befragung für die Lehrkräfte einfacher und weniger zeitaufwändig ist. Ein Nachteil kann darin gesehen werden, dass die Lehrkraft spontan antwortet und schlichtweg vergessen kann, Auffälligkeiten oder durchgeführte Maßnahmen für ein Kind zu nennen. Bei einem schriftlichen Fragebogen, der zeitlich aufwändiger ist, kann eine Lehrkraft ihre Antworten dagegen im Nachhinein noch einmal durchlesen und gegebenenfalls noch Informationen hinzufügen oder korrigieren.

Eine weitere Einschränkung dieser Langzeituntersuchung ist dadurch bedingt, dass es im Verlauf zu Drop-Outs kam. Nicht immer äußerten sich die Lehrkräfte zum Verbleib der Kinder und auch hier wurde nicht nachgefragt. Daher kann nur für einige Kinder, zu denen die Lehrkräfte sich im Nachhinein noch einmal geäußert haben, etwas über den Grund des Wegfalls berichtet werden. Dies geschah zum einen aufgrund von Klassenwiederholungen, aber auch durch lange Krankenhausaufenthalte oder den Wegzug der Familie. Es muss darum

zweifelsohne berücksichtigt werden, dass die vorliegenden Daten einen entsprechenden Bias enthalten, da die Kinder, welche die Klasse wiederholten oder übersprangen, nicht mehr Teil der Stichprobe sind. Dies ist vor allem für Vergleiche zwischen der ersten mit dem Ende der vierten Klasse der Fall. In weiteren Studien könnten mehrere Klassenstufen betrachtet werden, um dieses Problem zumindest für die Kinder, die eine Klasse wiederholen oder überspringen, zu umgehen. Kinder, welche die Schule wechseln oder längere Zeit krank sind, werden jedoch auch in diesem Fall aus der Stichprobe herausfallen.

Auf Seiten der Kinder kann es zu mangelnder Motivation oder Fehlern im Ausfüllen der Testunterlagen gekommen sein, da die Testungen oft in Gruppen stattgefunden haben und die Kinder sich über die Dauer einer Schulstunde konzentrieren mussten.

Eine Vergrößerung der Stichprobe wäre wünschenswert gewesen, ist aber nicht ohne weiteres möglich. Die Erhebungen in den Klassen mussten zeitnah erfolgen, um einen Vergleich untereinander gewährleisten zu können, sodass aufgrund von personellen Ressourcen nur eine begrenzte Anzahl teilnehmen konnte. Da die Studie zeitgleich auch in Finnland und Südtirol stattfand, musste auch auf diese Erhebungszeiträume Rücksicht genommen werden. Der Wegfall zweier Klassen zum Ende der vierten Klasse fällt zusätzlich negativ ins Gewicht. Neben der Vergrößerung der Stichprobe ist für eine weitere Untersuchung auch denkbar, gezielt leseschwache Kinder auszumachen und diese zu begleiten. Allerdings würden dadurch sowohl die Lehrkräfte wie auch die teilnehmenden Kinder entsprechend beeinflusst werden, da auf diesen Kindern gezielt der Fokus liegen würde. Eine Vergleichsgruppe ließe sich bei diesem Studiendesign auch nicht bilden.

In dieser Studie wurden leseschwache Kinder unabhängig von ihrer Rechtschreibleistung betrachtet. In einer zusätzlichen Analyse wurde die Rechtschreibleistung explorativ ebenfalls einbezogen. Nicht jedes leseschwache Kind dieser Studie hat auch Schwierigkeiten in der Rechtschreibung. Mitte bis Ende der ersten Klasse lagen bei etwa jedem zweiten leseschwachen Kind auch unterdurchschnittliche Rechtschreibleistungen vor. Zu Beginn der zweiten Klasse war dies bei jedem dritten leseschwachen Kind der Fall. Mitte des ersten Schuljahres ergab sich mit einer großen Effektstärke von  $g = 1.20$  eine Tendenz, dass die Kinder, die den Lehrkräften als leseschwach auffällig geworden sind, auch über eine niedrigere Rechtschreibleistung verfügen. Ende des ersten und Anfang des zweiten Schuljahres war dieser Unterschied jeweils signifikant. Dies könnte ein Hinweis darauf sein, dass die Leseschwäche an sich den Lehrkräften nicht auffällt, sondern erst in Kombination mit anderen Auffälligkeiten, in diesem Fall der Rechtschreibleistung. Es sollte in Zukunft geprüft werden, ob Lehrkräfte in der Diagnostik der Leseschwäche stärker geschult werden sollten, vor allem in Bezug auf eine isolierte Lesestörung. Bei Ausweitung der Stichprobengröße wäre es sinnvoll, zwischen Kindern, die eine isolierte Leseschwäche aufweisen, und Kindern mit kombinierten Leserechtschreibschwierigkeiten zu unterscheiden.

Zudem zeigte sich, dass ehemals leseschwache Kinder bereits zum Ende der ersten Klasse tendenziell bessere Leistungen zeigten als Kinder, die zum Ende der vierten Klasse weiterhin

leseschwach waren. Mitte der ersten Klasse unterschieden sich die beiden Gruppen noch nicht voneinander. Warum dies so ist, kann nicht geprüft werden. Es kann aber als Hinweis gesehen werden, dass Kinder früh gefördert werden sollten, da sich ansonsten die Leseschwierigkeiten verfestigen können. Eine früh einsetzende und gezielte Förderung ist nötig, um den Kindern ein Aufholen zu ermöglichen. Dies entspricht auch dem bisherigen Stand der Forschung (Fischer & Gasteiger-Klicpera, 2013; Klicpera et al., 2013).

Lohnenswert wäre ebenso ein Blick auf den an den Schulen stattfindenden Förderunterricht Deutsch: Wie häufig werden dort jeweils explizit die Rechtschreibung und das Lesen gefördert? Nach welchen Methoden findet die Förderung statt? Wie wirkungsvoll sind die Methoden? Die vorliegende Stichprobe zeigte im Vergleich zu Kindern ohne Fördermaßnahmen teilweise sogar eine negativere Leseentwicklung. Ob dies allerdings auf Inhalt, Umfang und Art der Fördermaßnahmen zurückzuführen ist, lässt sich nicht sagen, da die Gruppen sich auch in anderen Punkten unterscheiden. Eine experimentelle Studie zum Förderunterricht könnte lohnenswert sein, um weitere Einsichten zu erhalten.

### 13.6 Praktische Implikationen

Die Ergebnisse zeigen, dass Lehrkräfte vor allem zu Beginn der Schulzeit Schwierigkeiten haben, leseschwache Kinder zu identifizieren. In dieser Studie blieben im Verlauf der ersten Klasse mehr als Dreiviertel der leseschwachen Kinder mit ihren Leseschwierigkeiten unerkannt. Es wurden verschiedene mögliche Einflussfaktoren betrachtet, von denen sich lediglich das Geschlecht ab der zweiten Klasse als signifikant erwies. Eventuell überlagert hier das Klischee, dass Mädchen im Lesen und Schreiben besser sind als Jungen, die Urteilsfähigkeit der Lehrkräfte. Dies kann zu negativen Konsequenzen führen, wenn die Leseschwäche aufgrund des Geschlechts nicht bemerkt wird. Bei der Diagnostik sollten Lehrkräfte daher stets ein besonderes Augenmerk auf mögliche Leseschwächen bei Mädchen haben, um diese nicht zu übersehen. Ebenso zeigte sich, dass die Kinder, deren Leseschwächen erkannt wurden, häufig auch über niedrige Rechtschreibleistungen verfügten. Kinder mit isolierten Leseschwierigkeiten waren den Lehrkräfte seltener auffällig geworden.

Da die Erkennungsrate insgesamt sehr niedrig ausfiel, sollten regelmäßig Screeningverfahren zur Überprüfung der Lesefähigkeit durchgeführt werden. Hierbei ist es wichtig, dass alle Kinder einer Klasse teilnehmen, um den Einfluss einzelner Faktoren, wie beispielsweise das Geschlecht des Kindes, auf die Urteilsakkuratheit zu minimieren. Der präventiv ausgelegte Response-to-Intervention-Ansatz, der in den USA mittlerweile auch gesetzlich verankert ist und in Deutschland im Zuge des Rügener Inklusionsmodells als Pilotprojekt umgesetzt wurde, sieht vor, dass solche Screenings mehrere Male im Verlauf des jeweiligen Schuljahres stattfinden, um Kinder mit Schwierigkeiten frühzeitig zu identifizieren und eingreifen zu können, bevor sich Lernschwierigkeiten verfestigen oder Rückstände vergrößern können (Ennemoser & Besca, 2021, S. 44; Voß et al., 2014, S. 115–117).

Ein Screeningverfahren hat den Vorteil, dass dieses meist in kurzer Zeit und mit der gesamten Klasse durchführbar ist. Zudem sollte solch ein Verfahren über entsprechende Vergleichsnormen verfügen, sodass leseschwache Kinder gut erkannt werden können. Im Anschluss sollten bei auffällig gewordenen Kindern weitere Testverfahren oder eine qualitative Analyse der Lesefähigkeiten angewendet werden, um genauer festzustellen, wo die Schwierigkeiten des Kindes liegen und welche Fördermaßnahmen notwendig sind. Grundschullehr- amtsstudierende sollten schon in ihrer Studienzeit mit entsprechenden Testverfahren vertraut gemacht werden.

Aus der Forschung ist bekannt, dass Förderung möglichst früh einsetzen soll. Hieraus ergibt sich die Notwendigkeit, Screeningverfahren schon in der ersten Klasse durchzuführen. In den Grundschulen sollten verpflichtend Zeiträume zum Durchführen von Screeningverfahren zur Überprüfung der Lesefähigkeiten eingeführt werden, sodass sichergestellt werden kann, dass Lehrkräfte über zeitliche und auch materielle Ressourcen verfügen, um diese durchzuführen zu können. Mit dem Erkennen von Leseschwierigkeiten allein ist den betroffenen Kindern noch nicht geholfen, wie auch die Ergebnisse dieser Studie zeigen. Viele Kinder, deren Leseschwierigkeiten die Lehrkräfte benannt hatten, blieben in der Folge ohne gezielte Förderung. Die Gründe hierfür wurden in dieser Untersuchung nicht eruiert. Lehrkräfte sollten daher generell für die Relevanz von frühzeitig einsetzender und gezielter Leseförderung sensibilisiert werden. Es sollte darüber hinaus verbindlich festgehalten werden, wann Kinder als förderbedürftig eingeschätzt werden und welche Schritte erfolgen sollen. Das Rügener Inklusionsmodell sieht beispielsweise vor, die Kinder, die zu den leistungsschwächsten 25 Prozent gehören, zunächst in einer Kleingruppe zu fördern. Wenn Lernfortschritte ausbleiben oder die Kinder zu den schwächsten zehn Prozent gehören, soll auf der nächsten Förderebene eine intensive Einzelförderung stattfinden (Voß et al., 2014, S. 120).

Das Erkennen der Leseschwierigkeiten an sich führte in dieser Studie zu keinen besseren Leseleistungen im weiteren Verlauf der Grundschulzeit. Allerdings zeigte sich auch keine positivere Leseentwicklung für Kinder, die an einer Deutschförderung teilgenommen hatten. Aufgrund der fehlenden positiven Ergebnisse ist davon auszugehen, dass die Lesefähigkeiten der Kinder in dieser Stichprobe nicht oder zumindest nicht ausreichend gefördert wurden. Basierend auf den Aussagen der Lehrkräfte ist anzunehmen, dass Kinder nur sehr wenig spezifische Fördermethoden durchlaufen haben. Das kann ein Hinweis darauf sein, dass die Lehrkräfte diese nur selten einsetzen oder dass ihnen entsprechende Methoden erst gar nicht bekannt sind. Ebenso berichtete keine Lehrkraft von klassenweiten Fördermaßnahmen. An dieser Stelle sollte erhoben werden, inwiefern Grundschullehrkräfte im Fach Deutsch im Zuge ihrer universitären Ausbildung entsprechende Förder- und Unterstützungsmaßnahmen kennengelernt haben, bzw. ob sich diese im Unterricht gut einsetzen lassen. Zusätzlich benötigen auch bereits ausgebildete Grundschullehrkräfte die Möglichkeit, sich über Förderprogramme und -materialien fortbilden zu können. Dabei sollte berücksichtigt werden, dass eine Grundschullehrkraft in ihrer Klasse nur über begrenzte zeitliche Ressourcen verfügt. Eine Art

der Unterstützung liegt im klassenweiten Einsatz von Lautlesemethoden vor, welche entsprechend mit allen Kindern durchgeführt wird. Diese können Teil des regulären Leseunterrichts sein.

Bei gravierenderen Leseschwierigkeiten benötigen Kinder zudem zusätzliche Unterstützung, welche nicht im Regelunterricht von der Grundschullehrkraft allein umgesetzt werden kann. Für diese Kinder sind zusätzliche Förderstunden in Kleingruppen oder auch, wenn trotz Förderung in der Kleingruppe weiterhin Schwierigkeiten bestehen, im Einzelsetting notwendig (Ennemoser & Besca, 2021, S. 44). Auch die außerschulisch durchgeführte Lerntherapie sieht regulär Einzelförderung vor. Diese ist jedoch von den Sorgeberechtigten zu bezahlen und nicht Bestandteil des Unterrichts. Bestrebungen, Lerntherapie auch in die Schulen zu bringen und damit für alle unabhängig vom Einkommen zugänglich zu machen, gibt es allerdings (Balke-Melcher et al., 2016; Ricken, 2014).

Der Förderunterricht für den Bereich Deutsch sollte daraufhin untersucht werden, inwiefern die Lesefähigkeit dort spezifisch gefördert wird. Die vorliegende Untersuchung kann keine Aussage hierüber treffen, da über die Inhalte des Förderunterrichts keine Informationen erhoben wurden. Es ist denkbar, dass die Lehrkräfte sich verstärkt auf die Rechtschreibleistungen fokussiert haben und die Leseförderung nur eine untergeordnete Rolle spielte. Hilfreich wäre ein Blick darauf, welche Möglichkeiten bestehen, an die Inhalte des Deutschförderkurses im Regelunterricht anzuknüpfen oder passend zu ergänzen.

Der RTI-Ansatz fordert den Einsatz evidenzbasierter Methoden in der Förderung, da diese eine entsprechende Qualität gewährleisten und hierüber eher eine positive Leistungsentwicklung zu erwarten ist (Balke-Melcher et al., 2016, S. 18; Voß et al., 2014, S. 117).

Zusammenfassend ist festzuhalten, dass in der Aus- und Fortbildung von Grundschullehrkräften der Diagnostik von isolierten wie auch kombinierten Leseschwächen ebenso wie Unterstützungsmöglichkeiten, welche Lehrkräfte innerhalb der gesamten Klasse durchführen können, eine stärkere Bedeutung zukommen müssen. Auch wenn es für den Bereich der Grundschule nur sehr wenig solcher Förderprogramme gibt, sollte die Bekanntheit dieser entsprechend gesteigert werden. Lautleseverfahren zur Förderung des basalen Lesens, welches in den ersten Grundschuljahren im Fokus steht, wurden beispielsweise von keiner Lehrkraft als Unterstützung genannt, auch wenn sie Kinder korrekt als leseschwach identifiziert hatten. Diese Verfahren lassen sich als Teil des Leseunterrichts im klassenweiten Setting durchführen, sodass die Implementierung keine große Herausforderung darstellen sollte. Dies zeigt sich auch darin, dass laut der IGLU-Studie 2016 Lautlese-Tandems bereits bei über der Hälfte der Kinder in der vierten Klasse eingesetzt werden, allerdings nur ein- bis zweimal die Woche (Bremerich-Vos et al., 2017, S. 287). Eine frühere und regelmäßige Durchführung in den Klassen könnte zu einer höheren Effektivität der Methode führen. Bei Kindern mit bereits eingetretenen Schwierigkeiten im Leseerwerb sollte darüber hinaus gezielt Förderung in Kleingruppen oder in Einzelsettings mit evidenzbasierten Methoden zum Einsatz kommen.

### 13.7 Fazit und Ausblick

Die Studie konnte einen Einblick in die förderdiagnostische Kompetenz von Deutschlehrkräften an hessischen Grundschulen geben. Es zeigte sich, dass viele leseschwache Kinder den Lehrkräften nicht als leseschwach auffällig wurden. Insgesamt blieben aber auch viele Kinder, die den Lehrkräften aufgefallen waren, ohne Förderung. Ebenso wurden kaum spezifische Fördermethoden im Lesen genannt. Im Hinblick auf das deutsche Schulsystem stellt sich die Frage, inwiefern Grundschullehrkräfte neben der entsprechenden fachlichen Kompetenz und Erfahrung über die notwendige Zeit und entsprechende finanzielle Ressourcen verfügen, um Leseschwächen gezielt erkennen und auch im Regelunterricht bereits fördern zu können.

Häufig wurde eine allgemeine Deutschförderung als Unterstützungsmaßnahme angegeben. Kinder mit Deutschförderung zeigten allerdings entweder keine Unterschiede in ihrer Lesefähigkeit zum Ende der Grundschulzeit oder sogar niedrigere Leseleistungen als Kontrollgruppen ohne Deutschförderung. Die Gründe hierfür konnten in dieser Studie aufgrund nicht erhobener Daten nicht herausgearbeitet werden. Wichtig erscheint es an dieser Stelle daher, die in vielen Grundschulen stattfindenden Deutschförderkurse genauer in den Blick der Forschung zu rücken: Wie findet die Auswahl der Kinder statt, die an den Förderkursen teilnehmen? Was findet in den Förderkursen statt? Wie häufig und wie lange nehmen Kinder an der Förderung teil? Und welche Fortschritte erzielen Kinder, die an der Förderung teilnehmen? Um die Förderdiagnostik in den Grundschulen zu verbessern, sollten Grundschullehrkräfte sowie die Förderlehrkräfte der regionalen Beratungs- und Förderzentren außerdem gezielt nach ihren Bedürfnissen bezüglich Fortbildung für Lesediagnostik und -förderung befragt werden, um festzustellen, wo sie Schwierigkeiten sehen und welche Hilfsmittel wie beispielsweise Screening-Verfahren, quantitative oder auch qualitative Testverfahren sie benötigen, um leseschwache Kinder sicherer zu erkennen und ihnen im eigenen Unterricht oder in zusätzlichen Förderstunden effektive Unterstützung zukommen lassen zu können. Darauf aufbauend sollten entsprechende Fortbildungen gestaltet werden.

Neben dem Bereich der Fortbildung sollte auch der erste Abschnitt der Ausbildung, das Studium, in den Fokus rücken: Welche Kenntnisse werden in diesem Zeitraum bereits vermittelt und wo besteht Optimierungsbedarf? Gleichzeitig sollte an dieser Stelle neben einem Fokus auf Lesediagnostik bereits die Bekanntheit von vorhandenen evidenzbasierten Fördermaßnahmen gesteigert werden.

## Literaturverzeichnis

- Abs, H. J. (2006). Zur Bildung diagnostischer Kompetenz in der zweiten Phase der Lehrerbildung. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern (Zeitschrift für Pädagogik, Beiheft; 51)* (S. 217–234). Weinheim und Basel: Beltz.
- Allison, P. D. (2000). Multiple Imputation for Missing Data: A Cautionary Tale. *Sociological Methods & Research*, 28(3), 301–309. <https://doi.org/10.1177/0049124100028003003>
- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman & Hall.
- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57(3), 175–193. <https://doi.org/10.2378/peu2010.art13d>
- Artelt, C. & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(34), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>
- Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J. et al. (2007). *Förderung von Lesekompetenz - Expertise*. Verfügbar unter: [http://www.bmbf.de/pub/bildungsreform\\_band\\_siebzehn.pdf](http://www.bmbf.de/pub/bildungsreform_band_siebzehn.pdf)
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–140). Opladen: Leske + Budrich.
- Bacher, J. (2002). Statistisches Matching: Anwendungsmöglichkeiten, Verfahren und ihre praktische Umsetzung in SPSS. *Zentralarchiv für Empirische Sozialforschung*, 51, 38–66.
- Bailey, A. L. & Drummond, K. V. (2006). Who Is at Risk and Why?: Teachers' Reasons for Concern and Their Understanding and Assessment of Early Literacy. *Educational Assessment*, 11(3 & 4), 149–178.
- Balke-Melcher, C., Schuchardt, K., Wolpers, J.-G. & Mähler, C. (2016). LeFIS-Lernförderung in Schulen – Evaluation eines Modellprojekts zur schulinternen Lerntherapie für Kinder mit Lese- & Rechtschreibschwierigkeiten. *Lernen und Lernstörungen*, 5(1), 17–31. <https://doi.org/10.1024/2235-0977/a000122>
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. et al. (Hrsg.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. & Maaz, K. (2010). Bildungsungleichheit und Bildungsarmut – Der Beitrag von Large-Scale-Assessments. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer: neue Ungleichheiten* (1. Auflage., S. 159–179). Wiesbaden: VS, Verlag für Sozialwissenschaften.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Opladen: Leske + Budrich.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000:*

- Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Beck, M., Jäpel, F. & Becker, R. (2010). Determinanten des Bildungserfolgs von Migranten. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer: neue Ungleichheiten* (1. Auflage., S. 313–337). Wiesbaden: VS, Verlag für Sozialwissenschaften.
- Becker, D. & Birkelbach, K. (2013). Lehrer als Gatekeeper? Eine theoriegeleitete Annäherung an Determinanten und Folgen prognostischer Lehrerurteile. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte: Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen*. (S. 207–237). Dordrecht: Springer. Zugriff am 28.3.2018. Verfügbar unter: <http://public.eblib.com/choice/publicfullrecord.aspx?p=1466438>
- Behrmann, L. & Souvignier, E. (2013). The Relation Between Teachers' Diagnostic Sensitivity, their Instructional Activities, and their Students' Achievement Gains in Reading. *Zeitschrift für Pädagogische Psychologie*, 27(4), 283–293. <https://doi.org/10.1024/1010-0652/a000112>
- Bender, F., Brandelik, K., Jeske, K., Lipka, M., Löffler, C., Mannhaupt, G. et al. (2017). Die integrative Lerntherapie: Therapieform zur Behandlung von Lernstörungen. *Lernen und Lernstörungen*, 6(2), 65–73. <https://doi.org/10.1024/2235-0977/a000167>
- Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Bilz, L., Steger, J. & Fischer, S. M. (2016). Die Genauigkeit des Lehrerurteils bei der Identifikation von an Mobbing beteiligten Schülerinnen und Schülern. *Psychologie in Erziehung und Unterricht*, 63(PrePrint Online), 1–15. <https://doi.org/10.2378/peu2016.art08d>
- Boardman, A. G., Klingner, J. K., Buckley, P., Annamma, S. & Lasser, C. J. (2015). The efficacy of Collaborative Strategic Reading in middle school science and social studies classes. *Reading and Writing*, 28(9), 1257–1283. <https://doi.org/10.1007/s11145-015-9570-3>
- Bos, W., Buddeberg, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.). (2012). *IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster / New York / München / Berlin: Waxmann.
- Bos, W., Valtin, R., Hußmann, A., Wendt, H. & Goy, M. (2017). IGLU 2016: Wichtige Ergebnisse im Überblick. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 13–28). Münster / New York: Waxmann.
- Bövers, S. & Schulz, W. (2005). Integrative Lerntherapie bei Kindern mit Lese- und Rechtschreibstörungen: Ergebnisse einer katamnestic Studie. *Kindheit und Entwicklung*, 14(3), 191–200. <https://doi.org/10.1026/0942-5403.14.3.191>
- Brandenburg, J., Kleszczewski, J., Fischbach, A., Büttner, G., Grube, D., Mähler, C. et al. (2013). Arbeitsgedächtnisfunktionen von Kindern mit Minderleistungen in der Schriftsprache. *Lernen und Lernstörungen*, 2(3), 147–159. <https://doi.org/10.1024/2235-0977/a000037>
- Brem, S. & Maurer, U. (2015). Lesen als neurobiologischer Prozess. In U. Rautenberg & U. Schneider (Hrsg.), *Lesen: ein interdisziplinäres Handbuch* (S. 117–140). Berlin; Boston: De Gruyter.
- Bremerich-Vos, A., Stahns, R., Hußmann, A. & Schurig, M. (2017). Kapitel X. Förderung von Leseflüssigkeit und Leseverstehen im Leseunterricht. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016*:

- Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 279–296). Münster / New York: Waxmann.
- Brosius, F. (2018). *SPSS: umfassendes Handbuch zu Statistik und Datenanalyse* (8. Auflage). Frechen: mitp.
- Brühwiler, C. (2017). Diagnostische und didaktische Kompetenz als Kern adaptiver Lehrkompetenz (Pädagogische Psychologie und Entwicklungspsychologie). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: theoretische und methodische Weiterentwicklungen* (S. 123–134). Münster New York: Waxmann.
- Buda, C. (2017). Projekt „Modulares Förderkonzept bei Rechtschreibproblemen“: Ein Praxisbericht. *Lernen und Lernstörungen*, 6(4), 183–192. <https://doi.org/10.1024/2235-0977/a000187>
- Buick, S., Gold, A., Mokhlesgerami, J., Rühl, K. & Souvignier, E. (2006). *Wir werden Textdetektive. Lehrermanual* (2. Auflage). Göttingen: Vandenhoeck & Ruprecht.
- Bulheller, S. & Häcker, H. (2002). *Coloured progressive matrices: mit der Parallelform des Tests und der Puzzle-Form; Deutsche Bearbeitung und Normierung nach J. C. Raven*. Frankfurt: Swets Test Services.
- Buschmann, A. & Multhaus, B. (2016). LRS-Elterngruppenprogramm: Teilnehmerzufriedenheit und subjektive Effektivität. *Lernen und Lernstörungen*, 5(1), 44–57. <https://doi.org/10.1024/2235-0977/a000089>
- van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3). <https://doi.org/10.18637/jss.v045.i03>
- Caliendo, M. & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31–72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Carrion-Castillo, A., Franke, B. & Fisher, S. E. (2013). Molecular Genetics of Dyslexia: An Overview. *Dyslexia*, 19(4), 214–240. <https://doi.org/10.1002/dys.1464>
- Carrion-Castillo, A., Maassen, B., Franke, B., Heister, A., Naber, M., van der Leij, A. et al. (2017). Association analysis of dyslexia candidate genes in a Dutch longitudinal sample. *European Journal of Human Genetics*, 25(4), 452–460. <https://doi.org/10.1038/ejhg.2016.194>
- Catts, H. W., Fey, M. E., Tomblin, J. B. & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research*, 45, 1142–1157.
- Christophel, E., Baadte, C., Heyne, N. & Schnotz, W. (2017). Diagnostische und didaktische Kompetenz von Lehrkräften zur Förderung der Text-/ Bild-Integrationsfähigkeit bei Schülerinnen und Schülern der Sekundarstufe I (DIKOL). In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals: interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 263–281). Wiesbaden: Springer VS.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155–159.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78(2), 141–146.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Hrsg.), *Strategies of information processing* (S. 151–216). London: Academic Press.
- Coltheart, M. (2005a). Modeling Reading: The Dual-Route Approach. In M.J. Snowling & C. Hulme (Hrsg.), *The science of reading: a handbook* (S. 6–23). Malden, MA: Blackwell Pub. Zugriff am 12.7.2016. Verfügbar unter: [http://www.blackwellreference.com/subscriber/uid=3/book?show=all&id=g9781405114882\\_9781405114882](http://www.blackwellreference.com/subscriber/uid=3/book?show=all&id=g9781405114882_9781405114882)

- Coltheart, M. (2005b). Modeling Reading: The Dual-Route Approach. In M.J. Snowling & C. Hulme (Hrsg.), *The science of reading: a handbook* (S. 6–23). Malden, MA: Blackwell Pub. Zugriff am 12.7.2016. Verfügbar unter: [http://www.blackwellreference.com/subscriber/uid=3/book?show=all&id=g9781405114882\\_9781405114882](http://www.blackwellreference.com/subscriber/uid=3/book?show=all&id=g9781405114882_9781405114882)
- Coltheart, M. (2006). Dual route and connectionist models of reading: an overview. *London Review of Education*, 4(1), 5–17. <https://doi.org/10.1080/13603110600574322>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 87(1), 98–104.
- Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A. & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities*, 39(6), 507–514.
- Daucourt, M. C., Schatschneider, C., Connor, C. M., Al Otaiba, S. & Hart, S. A. (2018). Inhibition, Updating Working Memory, and Shifting Predict Reading Disability Symptoms in a Hybrid Model: Project KIDS. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00238>
- DeFries, J. C. & Alarcón, M. (1996). Genetics of specific reading disability. *Mental Retardation and Developmental Disabilities Research Reviews*, 2(1), 39–47. [https://doi.org/10.1002/\(SICI\)1098-2779\(1996\)2:1<39::AID-MRDD7>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1098-2779(1996)2:1<39::AID-MRDD7>3.0.CO;2-S)
- Dehn, M. & Hüttis-Graff, P. (2010). *Beobachtung, Diagnose, Lernhilfen: Schulanfangsbeobachtung, Lernbeobachtung Schreiben und Lesen, für Schulanfang und Jahrgang 1 (Zeit für die Schrift)* (2. Auflage). Berlin: Cornelsen-Scriptor.
- Denckla, M. B. & Rudel, R. G. (1976). Rapid „automatized“ naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14, 471–479.
- Deutsches Institut für Medizinische Dokumentation und Information (Hrsg.). (2018). *ICD-10. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme. 10. Revision - WHO-Ausgabe* (Band 1-Systematisches Verzeichnis). Zugriff am 26.6.2018. Verfügbar unter: <https://www.dimdi.de/dynamic/de/klassi/downloadcenter/icd-10-gm/version2018/>
- Diamond, A. (2014). Biologische und soziale Einflüsse auf kognitive Kontrollprozesse, die vom präfrontalen Kortex abhängen (Programmbereich Psychologie). In S. Kubesch (Hrsg.), A. Hildenbrand (Übers.), *Exekutive Funktionen und Selbstregulation: neurowissenschaftliche Grundlagen und Transfer in die pädagogische Praxis* (1. Auflage, S. 19–47). Bern: Huber.
- Diaper, G. (1990). The Hawthorne Effect: a fresh examination. *Educational Studies*, 16(3), 261–267. <https://doi.org/10.1080/0305569900160305>
- Dummer-Smoch, L. (2007). Schwierigkeiten beim Lesenlernen aufgrund einer spezifischen Störung (Legasthenie) oder Ausdruck einer allgemeinen Lernbeeinträchtigung (Handbuch Sonderpädagogik). In H. Schöler & A. Welling (Hrsg.), *Sonderpädagogik der Sprache* (S. 370–395). Göttingen: Hogrefe.
- Dunn, T. J., Baguley, T. & Brunsdon, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Eckerth, M. (2015). Die Differenzierungspraxis von Lehrkräften im schriftsprachlichen Anfangsunterricht der Grundschule. In K. Liebers, B. Landwehr, A. Marquardt & K. Schlotter (Hrsg.), *Lernprozessbegleitung und adaptives Lernen in der Grundschule Forschungsbezogene Beiträge* (S. 175–180). Wiesbaden: Springer VS.
- Eekhout, I., van de Wiel, M. A. & Heymans, M. W. (2017). Methods for significance testing of categorical covariates in logistic regression models after multiple imputation: power

- and applicability analysis. *BMC Medical Research Methodology*, 17(1), 129.  
<https://doi.org/10.1186/s12874-017-0404-7>
- Ellis, P. D. (2010). *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*. Cambridge; New York: Cambridge University Press.
- Enders, C. K. (2010). *Applied missing data analysis* (Methodology in the social sciences). New York: Guilford Press.
- Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour Research and Therapy*, 98, 4–18.  
<https://doi.org/10.1016/j.brat.2016.11.008>
- Endlich, D., Dummert, F., Schneider, W. & Schwenck, C. (2014). Verhaltensprobleme bei Kindern mit umschriebener und kombinierter schulischer Minderleistung. *Kindheit und Entwicklung*, 23(1), 61–69. <https://doi.org/10.1026/0942-5403/a000128>
- Ennemoser, M. & Besca, M. (2021). Präventive Mehrebenenförderung. *Pädagogik*, (4), 44–47.
- Esser, G., Wyschkon, A. & Schmidt, M. H. (2002). Was wird aus Achtjährigen mit einer Lese- und Rechtschreibstörung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 31(4), 235–242. <https://doi.org/10.1026/0084-5345.31.4.235>
- Fadler, S. & Böhme, K. (2018). Basale Lesefertigkeiten erfolgreich fördern. Die Nutzung von Lautlese-Tandems an Grundschulen. *Potsdamer Zentrum für empirische Inklusionsforschung (ZEIF)*, 3, 1–9.
- Falkai, P. & American Psychiatric Association. (2015). *Diagnostisches und statistisches Manual psychischer Störungen DSM-5*. Göttingen: Hogrefe.
- Feinberg, A. B. & Shapiro, E. S. (2009). Teacher Accuracy: An Examination of Teacher-Based Judgments of Students' Reading With Differing Achievement Levels. *The Journal of Educational Research*, 102(6), 453–462.  
<https://doi.org/10.3200/JOER.102.6.453-462>
- Fischbach, A., Schuchardt, K., Brandenburg, J., Kleszczewski, J., Balke-Melcher, C., Schmidt, C. et al. (2013). Prävalenz von Lernschwächen und Lernstörungen: Zur Bedeutung der Diagnosekriterien. *Lernen und Lernstörungen*, 2(2), 65–76.  
<https://doi.org/10.1024/2235-0977/a000035>
- Fischbach, A., Schuchardt, K., Mähler, C. & Hasselhorn, M. (2010). Zeigen Kinder mit schulischen Minderleistungen sozio-emotionale Auffälligkeiten? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42(4), 201–210.  
<https://doi.org/10.1026/0049-8637/a000025>
- Fischer, U. & Gasteiger-Klicpera, B. (2013). Prävention von Leseschwierigkeiten. Diagnose und Förderung im Anfangsunterricht. *Didaktik Deutsch*, 18(35), 62–81.
- Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, 18(2), 129–155. <https://doi.org/10.1007/s11145-004-5892-2>
- Freeman, G. H. & Halton, J. H. (1951). Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika*, 38(1–2), 141–149.  
<https://doi.org/10.1093/biomet/38.1-2.141>
- Fricke, S., Szczerbinski, M., Fox-Boyer, A. & Stackhouse, J. (2015). Preschool Predictors of Early Literacy Acquisition in German-Speaking Children. *Reading Research Quarterly*, 51(1), 29–53.
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of dyslexia*, 36(1), 67–81.

- Fuchs, D., Fuchs, L. S., Mathes, P. G. & Simmons, D. C. (1997). Peer-Assisted Learning Strategies: Making Classrooms More Responsive to Diversity. *American Educational Research Journal*, 34(1), 174–206. <https://doi.org/10.3102/00028312034001174>
- Gailberger, S. & Willenberg, H. (2008). Leseverstehen Deutsch (Beltz Pädagogik). In E. Klieme, W. Eichler, A. Helmke, R.H. Lehmann, G. Nold, H.-G. Rolff et al. (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (S. 60–71). Weinheim: Beltz.
- Galuschka, K., Ise, E., Krick, K. & Schulte-Körne, G. (2014). Correction: Effectiveness of Treatment Approaches for Children and Adolescents with Reading Disabilities: A Meta-Analysis of Randomized Controlled Trials. *PLoS ONE*, 9(2), e105843. <https://doi.org/10.1371/journal.pone.0105843>
- Ganzeboom, H. B. G., De Graaf, P. M. & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56. [https://doi.org/10.1016/0049-089X\(92\)90017-B](https://doi.org/10.1016/0049-089X(92)90017-B)
- Garbe, C., Holle, K. & Jesch, T. (2009). *Texte lesen: Lesekompetenz - Textverstehen - Lesedidaktik - Lesesozialisation*. Paderborn: Schöningh.
- Geiser, C. (2011). *Datenanalyse mit Mplus: eine anwendungsorientierte Einführung* (2., durchgesehene Auflage.). Wiesbaden: VS Verlag.
- Georgiou, G. K., Parrila, R., Cui, Y. & Papadopoulos, T. C. (2013). Why is rapid automatized naming related to reading? *Journal of Experimental Child Psychology*, 115(1), 218–225. <https://doi.org/10.1016/j.jecp.2012.10.015>
- Gisev, N., Bell, J. S. & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, 9(3), 330–338. <https://doi.org/10.1016/j.sapharm.2012.04.004>
- Goerke, B. (2009). Ausreißerwerte. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Ergänzungen zur Methodik der empirischen Forschung* (S. 23–43). Kiel. Verfügbar unter: <https://d-nb.info/1122110839/34#page=40>
- Goldston, D. B., Walsh, A., Mayfield Arnold, E., Reboussin, B., Sergent Daniel, S., Erkanli, A. et al. (2007). Reading Problems, Psychiatric Disorders, and Functional Impairment from Mid- to Late Adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(1), 25–32. <https://doi.org/10.1097/01.chi.0000242241.77302.f4>
- Gorecki, B. & Landerl, K. (2015). Phonologische Bewusstheit: Ist die phonologische Bewusstheit ein Prädiktor für die Leseleistung? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(3), 139–146. <https://doi.org/10.1026/0049-8637/a000135>
- Göthlich, S. E. (2009). Zum Umgang mit fehlenden Daten in großzahligen empirischen Erhebungen. In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (3., überarb. und erw. Aufl., S. 119–135). Wiesbaden: Gabler.
- Gough, P. B. & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10.
- Goy, M., Valtin, R. & Hußmann, A. (2017). Kapitel IV. Leseselbstkonzept, Lesemotivation, Leseverhalten und Lesekompetenz. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 143–176). Münster New York: Waxmann.

- Graham, J. W., Olchowski, A. E. & Gilreath, T. D. (2007). How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prevention Science*, 8(3), 206–213. <https://doi.org/10.1007/s11121-007-0070-9>
- Greifer, N. (2019). *cobalt: Covariate Balance Tables and Plots*. Verfügbar unter: <https://CRAN.R-project.org/package=cobalt>
- Groeben, N. (2009). Zur aktuellen Struktur des Konstrukts „Lesekompetenz“. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (S. 11–21). Weinheim: Juventa.
- Grünke, M. & Grosche, M. (2014). Lernbehinderung. In G.W. Lauth, M. Grünke & J.C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen: Förderung, Training und Therapie in der Praxis* (S. 76–89). Göttingen; Bern: Hogrefe, Verl. für Psychologie.
- Günther, K. B. (1986). Ein Stufenmodell der Entwicklung kindlicher Lese- und Schreibstrategien (Libelle: Wissenschaft Lesen und Schreiben). In H. Brügelmann (Hrsg.), *ABC und Schriftsprache: Rätsel für Kinder, Lehrer und Forscher* (S. 32–54). Konstanz: Faude.
- Günther, T., Peters, K., Scharke, W., Horbach, J., Kraatz, C., Herpertz-Dahlmann, B. et al. (2016). Aufmerksamkeit und Leseleistungen bei Kindern mit ADHS, Lese- und Rechtschreibstörung und der komorbiden Störung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 44(5), 351–363. <https://doi.org/10.1024/1422-4917/a000451>
- Hascher, T. (2008). Diagnostische Kompetenzen im Lehrberuf. In C. Kraler & M. Schratz (Hrsg.), *Wissen erwerben, Kompetenzen entwickeln: Modelle zur kompetenzorientierten Lehrerbildung* (S. 71–86). Münster: Waxmann.
- Hasselhorn, M. & Marx, H. (2000). Phonologisches Arbeitsgedächtnis und Leseleistungen. Ein Vergleich zwischen dysphasisch-sprachentwicklungsgestörten und sprachunauffälligen Kindern. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Lese-Rechtschreibschwierigkeiten. Jahrbuch der pädagogisch-psychologischen Diagnostik* (S. 135–148). S.l.: Hogrefe Verlag. Zugriff am 20.7.2018. Verfügbar unter: [http://e-books.ciando.com/shop/book/short/index.cfm?fuseaction=short&bok\\_id=653](http://e-books.ciando.com/shop/book/short/index.cfm?fuseaction=short&bok_id=653)
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (Schule weiterentwickeln, Unterricht verbessern Orientierungsband) (6. Auflage). Seelze-Velber: Klett/Kallmeyer.
- Hesse, I. & Latzko, B. (2009). *Diagnostik für Lehrkräfte* (UTB Pädagogik). Opladen: Budrich.
- Hesse, I. & Latzko, B. (2017). *Diagnostik für Lehrkräfte* (3. Auflage). Opladen & Toronto: Verlag Barbara Budrich.
- Heymans, M. W. & Eekhout, I. (2019). *Applied missing data analysis with SPSS and (R)Studio*. Amsterdam. Verfügbar unter: [https://bookdown.org/mwheyman/Book\\_ML/](https://bookdown.org/mwheyman/Book_ML/)
- Hinnant, J. B., O'Brien, M. & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101(3), 662–670.
- Ho, D. E., Imai, K., King, G. & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 42(8).
- Hochstadt, C., Krafft, A. & Olsen, R. (2015). *Deutschdidaktik: Konzeptionen für die Praxis* (UTB) (2., überarbeitete und erweiterte Auflage.). Tübingen: A. Francke Verlag.
- Hofmann, C., Koch, A., Holopainen, L., Kofler, D., Baur, S., Bauer, K. et al. (2015). Schriftspracherwerb in Finnland, Südtirol und Deutschland in den ersten vier Schuljahren - Eine Vergleichsstudie von 2010 bis 2014. Unveröffentlichtes Manuskript, Justus-Liebig-Universität Gießen.

- Hoge, R. D. (1983). Psychometric properties of teacher-judgment measures of pupil attitudes, classroom behaviors, and achievement levels. *Journal of Special Education*, 17(4), 401–429.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, 59(3), 297–313.
- Hollenbeck, J. R. & Wright, P. M. (2017). Harking, Sharking, and Tharking: Making the Case for Post Hoc Analysis of Scientific Data. *Journal of Management*, 43(1), 5–18. <https://doi.org/10.1177/0149206316679487>
- Huck, L. & Schröder, A. (2016). Psychosoziale Belastungen und Lernschwierigkeiten: Befunde zum Zusammenhang von Lese-Rechtschreib-Schwäche, Rechenschwäche und psychosozialen Belastungen in einer Inanspruchnahme-Stichprobe aus einer lerntherapeutischen Einrichtung. *Lernen und Lernstörungen*, 5(3), 157–164. <https://doi.org/10.1024/2235-0977/a000143>
- Hülsmann, M. (2018). Kommentare zu Bender et al. (2017). Die integrative Lerntherapie: Therapieform zur Behandlung von Lernstörungen. *Lernen und Lernstörungen*, 6 (2), 65–73: Wirkungsfelder der Integrativen Lerntherapie. *Lernen und Lernstörungen*, 7(3), 138–138. <https://doi.org/10.1024/2235-0977/a000220>
- Hurrelmann, B. (2013). Ein erweitertes Konzept von Lesekompetenz und Konsequenzen für die Leseförderung. In G. Auernheimer (Hrsg.), *Schieflagen im Bildungssystem* (S. 161–176). Wiesbaden: Springer Fachmedien Wiesbaden.
- Hußmann, A., Stubbe, T. C. & Kasper, D. (2017). Kapitel VI. Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195–218). Münster New York: Waxmann.
- Hußmann, A., Wendt, H., Kasper, D., Bos, W. & Goy, M. (2017). Ziele, Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU 2016). In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 29–77). Münster New York: Waxmann.
- IBM Corp. (2016, September 7). IBM Propensity Score Matching in SPSS Statistics. *Propensity Score Matching in SPSS Statistics*. Zugriff am 8.8.2019. Verfügbar unter: <http://www.ibm.com/support>, <https://www-304.ibm.com/support/docview.wss?uid=swg21639710>
- IBM Corp. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.
- Ise, E., Engel, R. R. & Schulte-Körne, G. (2012). Was hilft bei der Lese-Rechtschreibstörung?: Ergebnisse einer Metaanalyse zur Wirksamkeit deutschsprachiger Förderansätze. *Kindheit und Entwicklung*, 21(2), 122–136. <https://doi.org/10.1026/0942-5403/a000077>
- Janssen, J. & Laatz, W. (2017). *Statistische Datenanalyse mit SPSS*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-53477-9>
- Jürgens, E. & Lissmann, U. (2015). *Pädagogische Diagnostik: Grundlagen und Methoden der Leistungsbeurteilung in der Schule* (Reihe „BildungsWissen Lehramt“). Weinheim Basel: Beltz. Verfügbar unter: [https://content-select-com.ezproxy.uni-giessen.de/media/moz\\_viewer/552557cf-6624-4dec-912d-4cc3b0dd2d03/language:de](https://content-select-com.ezproxy.uni-giessen.de/media/moz_viewer/552557cf-6624-4dec-912d-4cc3b0dd2d03/language:de)
- Juska-Bacher, B., Beckert, C., Gyger, M., Jakob, S. & Schneider, H. (2015). Leserelevante Kompetenzen an der Schwelle vom Kindergarten zur Schule. *Zeitschrift für Literaturwissenschaft und Linguistik*, 45(2), 28–42. <https://doi.org/10.1007/BF03379722>

- Kaiser, J. & Möller, J. (2017). Diagnostische Kompetenz von Lehramtsstudierenden. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals: interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 55–74). Wiesbaden: Springer VS.
- Kaiser, J., Möller, J., Helm, F. & Kunter, M. (2015). Das Schülerinventar: Welche Schülermerkmale die Leistungsurteile von Lehrkräften beeinflussen. *Zeitschrift für Erziehungswissenschaft*, 18(2), 279–302. <https://doi.org/10.1007/s11618-015-0619-5>
- Kamzela, K. (2015). Die Perspektive von Lehrenden auf basale Lesefähigkeiten zu Beginn der Sekundarstufe I (Research). In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik* (S. 49–70). Wiesbaden: Springer VS.
- Karing, C., Matthäi, J. & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25(3), 159–172. <https://doi.org/10.1024/1010-0652/a000041>
- Karing, C., Pfof, M. & Artelt, C. (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? *Journal for Educational Research Online*, 3(2), 119–147.
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern* (Empirische Erziehungswissenschaft). Münster: Waxmann.
- Katzir, T., Kim, Y.-S. G. & Dotan, S. (2018). Reading Self-Concept and Reading Anxiety in Second Grade Children: The Roles of Word Reading, Emergent Literacy Skills, Working Memory and Gender. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01180>
- Kelava, A. & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen (Springer-Lehrbuch). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage., S. 75–102). Berlin; Heidelberg: Springer.
- Kelley, K. (2019). *MBESS: The MBESS R Package*. Verfügbar unter: <https://CRAN.R-project.org/package=MBESS>
- King, G. & Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*.
- Klicpera, C., Schabmann, A. & Gasteiger-Klicpera, B. (2013). *Legasthenie - LRS* (UTB Pädagogik, Psychologie) (4., aktualisierte Auflage.). München: Reinhardt.
- Klicpera, C., Schabmann, A., Gasteiger-Klicpera, B. & Schmidt, B. (2017). *Legasthenie - LRS Modelle, Diagnose, Therapie und Förderung*. Stuttgart: UTB GmbH.
- Klieme, E. (2008). Systemmonitoring für den Sprachunterricht (Beltz Pädagogik). In E. Klieme, W. Eichler, A. Helmke, R.H. Lehmann, G. Nold, H.-G. Rolff et al. (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (S. 1–10). Weinheim: Beltz.
- Klingner, J. K., Urbach, J., Golos, D., Brownell, M. & Menon, S. (2010). Teaching reading in the 21st century: A glimpse at how special education teachers promote reading comprehension. *Learning Disability Quarterly*, 33(2), 59–74.
- Klug, J. (2011). *Modeling and training a new concept of teachers' diagnostic competence*. Dissertation. Darmstadt: Technische Universität Darmstadt.
- KMK. (2005). Bildungsstandards im Fach Deutsch für den Primarbereich (Beschluss vom 15.10.2004). München: Wolters.

- KMK. (2014a). Standards für die Lehrerbildung: Bildungswissenschaften (Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 12.06.2014). München: Wolters.
- KMK. (2014b, Juni 13). Inklusion: KMK aktualisiert Standards für die Lehrerbildung. Zugriff am 7.6.2018. Verfügbar unter: <https://www.kmk.org/aktuelles/artikelansicht/inklusion-kmk-aktualisiert-standards-fuer-die-lehrerbildung.html>
- Koch, A. & Georg, K. (2010). Testbatterie zur Erfassung von Vor- und Teilleistungen des Schriftspracherwerbs. Unveröffentlichte Testmaterialien, Justus-Liebig-Universität Gießen.
- Koch, A., Holopainen, L., Hofmann, C., Bauer, K., Mäkihonko, M. & Kofler, D. (2012). Cross-linguistic study on literacy development at grade 1 and 2 in Finland, Germany and Italy. Unveröffentlichtes Manuskript, Justus-Liebig-Universität Gießen.
- Kohn, J., Wyschkon, A., Ballaschk, K., Ihle, W. & Esser, G. (2013). Verlauf von Umschriebenen Entwicklungsstörungen: Eine 30-Monats-Follow-up-Studie. *Lernen und Lernstörungen*, 2(2), 77–89. <https://doi.org/10.1024/2235-0977/a000032>
- Kohn, J., Wyschkon, A. & Esser, G. (2013). Psychische Auffälligkeiten bei Umschriebenen Entwicklungsstörungen: Gibt es Unterschiede zwischen Lese-Rechtschreib- und Rechenstörungen? *Lernen und Lernstörungen*, 2(1), 7–20. <https://doi.org/10.1024/2235-0977/a000027>
- Kopp, B., Martschinke, S., Munser-Kiefer, M., Haider, M., Kirschhock, E.-M., Ranger, G. et al. (Hrsg.). (2014). *Individuelle Förderung und Lernen in der Gemeinschaft* (Jahrbuch Grundschulforschung) (Band 17). Wiesbaden: Springer VS.
- Krinzinger, H. & Günther, T. (2013). Lesen, Schreiben, Rechnen – gibt es Unterschiede zwischen den Geschlechtern? *Lernen und Lernstörungen*, 2(1), 35–49. <https://doi.org/10.1024/2235-0977/a000029>
- Krombholz, H. (2019). *Zum Einfluss fehlender Werte („Missing Data“) auf Ergebnisse einer Längsschnittstudie im Vorschulalter – Wie zufällig sind „fehlende Beobachtungen“?* report. PsychArchives. <https://doi.org/http://dx.doi.org/10.23668/psycharchives.2432>
- Kuckartz, U., Rädiker, S., Ebert, T. & Schehl, J. (2013). *Statistik*. Wiesbaden: VS Verlag für Sozialwissenschaften. Zugriff am 8.11.2016. Verfügbar unter: <http://link.springer.com/10.1007/978-3-531-19890-3>
- Kuhl, P. & Hannover, B. (2012). Differenzielle Benotungen von Mädchen und Jungen: Der Einfluss der von der Lehrkraft eingeschätzten Kompetenz zum selbstgesteuerten Lernen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(3), 153–162. <https://doi.org/10.1026/0049-8637/a000066>
- Kuss, O., Blettner, M. & Börgermann, J. (2016). Propensity Score - eine alternative Methode zur Analyse von Therapieeffekten. *Deutsches Ärzteblatt Online*. <https://doi.org/10.3238/arztebl.2016.0597>
- Lee, V. E. (2000). Using Hierarchical Linear Modeling to Study Social Contexts: The Case of School Effects. *Educational Psychologist*, 35(2), 125–141.
- Leinhardt, G. (1983). Novice and Expert Knowledge of Individual Student's Achievement. *Educational Psychologist*, 18(3), 165–179.
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.
- Leuders, T., Leuders, J. & Philipp, K. (2014). Fachbezogene diagnostische Kompetenzen - Forschungsstand und Forschungsdesiderata. In J. Roth & J. Ames (Hrsg.), *Beiträge zum Mathematikunterricht 2014*. Zugriff am 1.9.2016. Verfügbar unter: <https://doi.org/10.17877/DE290R-5262>

- Leyendecker, B., Citlak, B., Schräpler, J.-P. & Schölmerich, A. (2014). Diversität elterlicher Einstellungen und vorschulischer Lernerfahrungen: ein Vergleich deutscher und zugewanderter Familien aus der Türkei, Russland und Polen. *Zeitschrift für Familienforschung*, 26(1), 70–93.
- Linberg, T. & Wenz, S. E. (2017). Ausmaß und Verteilung sozioökonomischer und migrationspezifischer Ungleichheiten im Sprachstand fünfjähriger Kindergartenkinder. *Journal for educational research online*, 9(1), 77–98.
- Lindberg, S. (2016). Ursachen der Lese-Rechtschreibstörung (UTB). In A. Mayer (Hrsg.), *Lese-Rechtschreibstörungen (LRS)* (S. 53–66). München Basel: Ernst Reinhardt Verlag.
- Lipka, M. & von Orloff, M. (2014). Ein Beruf stellt sich vor: Lerntherapeutin/Lerntherapeut. *Lernen und Lernstörungen*, 3(1), 72–75. <https://doi.org/10.1024/2235-0977/a000055>
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202.
- Logan, S., Medford, E. & Hughes, N. (2011). The importance of intrinsic motivation for high and low ability readers' reading comprehension performance. *Learning and Individual Differences*, 21(1), 124–128. <https://doi.org/10.1016/j.lindif.2010.09.011>
- Lorenz, C. & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23(34), 211–222. <https://doi.org/10.1024/1010-0652.23.34.211>
- Lüdtkke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- Maaz, K., Neumann, M., Trautwein, U., Wendt, W., Lehmann, R. & Baumert, J. (2008). Der Übergang von der Grundschule in die weiterführende Schule. Die Rolle von Schüler- und Klassenmerkmalen beim Einschätzen der individuellen Lernkompetenz durch die Lehrkräfte. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30(3), 519–548.
- Machts, N., Kaiser, J., Schmidt, F. T. C. & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, 19, 85–103. <https://doi.org/10.1016/j.edurev.2016.06.003>
- Mahling, M. (2015). Bildungspolitische und gesellschaftliche Rahmenbedingungen: Lesen und Schule (De Gruyter Reference). In U. Rautenberg & U. Schneider (Hrsg.), *Lesen: ein interdisziplinäres Handbuch* (S. 547–566). Berlin; Boston: De Gruyter.
- Maier, U. (2015). *Leistungsdiagnostik in Schule und Unterricht: Schülerleistungen messen, bewerten und fördern* (UTB Schulpädagogik). Bad Heilbrunn: Klinkhardt.
- Mansournia, M. A., Jewell, N. P. & Greenland, S. (2018). Case-control matching: effects, misconceptions, and recommendations. *European Journal of Epidemiology*, 33(1), 5–14. <https://doi.org/10.1007/s10654-017-0325-0>
- Marx, H. (2007). Theorien und Determinanten des Erwerbs der Schriftsprache (Handbuch Sonderpädagogik). In H. Schöler & A. Welling (Hrsg.), *Sonderpädagogik der Sprache* (Band 1, S. 92–147). Göttingen: Hogrefe.
- Mayer, A. (2016). *Lese-Rechtschreibstörungen (LRS)* (UTB). München Basel: Ernst Reinhardt Verlag.
- McElvany, N., Becker, M. & Lüdtkke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 41(3), 121–131. <https://doi.org/10.1026/0049-8637.41.3.121>

- McElvany, N., El-Khechen, W., Schwabe, F. & Kessels, U. (2016). Qualitative Wortschatzunterschiede zwischen Mädchen und Jungen im Grundschulalter. *Zeitschrift für Pädagogische Psychologie*, 30(1), 45–55. <https://doi.org/10.1024/1010-0652/a000167>
- McElvany, N., Kessels, U., Schwabe, F. & Kasper, D. (2017). Kapitel V. Geschlecht und Lesekompetenz. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 177–194). Münster New York: Waxmann.
- McElvany, N. & Ohle, A. (2015). Lese- und Sprachförderung - Theoretische und empirische Implikationen für die Unterrichtsentwicklung (Pädagogik). In H.-G. Rolff (Hrsg.), *Handbuch Unterrichtsentwicklung: eine Veröffentlichung der Deutschen Akademie für Pädagogische Führungskräfte (DAPF)* (S. 44–59). Weinheim: Beltz.
- McElvany, N., Ohle, A., El-Khechen, W., Hardy, I. & Cinar, M. (2017). Förderung sprachlicher Kompetenzen – Das Potenzial der Familiensprache für den Wortschatzerwerb aus Texten. *Zeitschrift für Pädagogische Psychologie*, 31(1), 13–25. <https://doi.org/10.1024/1010-0652/a000189>
- McGeown, S. P., Duncan, L. G., Griffiths, Y. M. & Stothard, S. E. (2015). Exploring the relationship between adolescent's reading skills, reading motivation and reading habits. *Reading and Writing*, 28(4), 545–569. <https://doi.org/10.1007/s11145-014-9537-9>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- MedCalc Software Ltd. (o. J.). Free statistical calculators - Relative risk calculator. Verfügbar unter: [https://www.medcalc.org/calc/relative\\_risk.php](https://www.medcalc.org/calc/relative_risk.php)
- Mienert, M. & Pitcher, S. (2011). *Pädagogische Psychologie*. Wiesbaden: VS Verlag für Sozialwissenschaften. Zugriff am 27.1.2015. Verfügbar unter: <http://link.springer.com/book/10.1007/978-3-531-92095-5>
- Mitra, R. & Reiter, J. P. (2016). A comparison of two methods of estimating propensity scores after multiple imputation. *Statistical Methods in Medical Research*, 25(1), 188–204. <https://doi.org/10.1177/0962280212445945>
- Moll, K., Kunze, S., Neuhoff, N., Bruder, J. & Schulte-Körne, G. (2014). Specific Learning Disorder: Prevalence and Gender Differences. *PLoS ONE*, 9(7), e103537. <https://doi.org/10.1371/journal.pone.0103537>
- Moll, K. & Landerl, K. (2010). *Lese- und Rechtschreibtest (SLRT-II). Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)*. Bern: Verlag Hans Huber.
- Moll, K., Wallner, R. & Landerl, K. (2012). Kognitive Korrelate der Lese-, Leserechtschreib- und der Rechtschreibstörung. *Lernen und Lernstörungen*, 1(1), 7–19. <https://doi.org/10.1024/2235-0977/a000002>
- Möller, J., Machts, N. & Retelsdorf, J. (2016). Diagnostische Kompetenz von Lehrkräften. Merkmale, Instrumente, Urteilsverzerrungen. *Schulmanagement*, 4, 14–17.
- Morgan, W. P. (1896). A Case of Congenital Word Blindness. *BMJ*, 2(1871), 1378–1378. <https://doi.org/10.1136/bmj.2.1871.1378>
- Mugnaini, D., Lassi, S., La Malfa, G. & Albertini, G. (2009). Internalizing correlates of dyslexia. *World Journal of Pediatrics*, 5(4), 255–264. <https://doi.org/10.1007/s12519-009-0049-7>
- Müller, B. (2014, Dezember 19). *Förderung des Leseverständnisses durch Lesestrategien: Eine Interventionsstudie in der Grundschule*. Dissertation. Kassel: Universität Kassel. Zugriff am 25.3.2015. Verfügbar unter: <https://kobra.bibliothek.uni-kassel.de/bitstream/urn:nbn:de:hebis:34-2015012347267/5/DissertationBettinaMueller.pdf>

- Müller, B., Križan, A., Hecht, T., Richter, T. & Ennemoser, M. (2013). Leseflüssigkeit im Grundschulalter: Entwicklungsverlauf und Effekte systematischer Leseförderung. *Lernen und Lernstörungen*, 2(3), 131–146. <https://doi.org/10.1024/2235-0977/a000039>
- Müller, B., Richter, T., Križan, A., Hecht, T. & Ennemoser, M. (2015). Word recognition skills moderate the effectiveness of reading strategy training in Grade 2. *Learning and Individual Differences*, 40, 55–62. <https://doi.org/10.1016/j.lindif.2015.05.001>
- Müller, C. E. (2012). *Quasiexperimentelle Wirkungsevaluation mit Propensity Score Matching: ein Leitfaden für die Umsetzung mit Stata*. CEval-Arbeitspapier, 19. (S. 1–40). Saarbrücken: Universität des Saarlandes, Fak. 05 Empirische Humanwissenschaften, CEval - Centrum für Evaluation. Verfügbar unter: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-428947>
- Multhauf, B., Buschmann, A. & Soellner, R. (2016). Effectiveness of a group-based program for parents of children with dyslexia. *Reading and Writing*, 29(6), 1203–1223. <https://doi.org/10.1007/s11145-016-9632-1>
- Munser-Kiefer, M. & Kirschhock, E.-M. (2009). Entwicklung und Evaluation eines Lesekompetenztrainings - Eine Vorstudie zur Frage nach der geeigneten Sozialform. In C. Röhrner, C. Henrichwark & M. Hopf (Hrsg.), *Europäisierung der Bildung. Konsequenzen und Herausforderungen für die Grundschulpädagogik* (S. 166–171). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH, Wiesbaden.
- Nagler, T., Lindberg, S. & Hasselhorn, M. (2018). Leseentwicklung im Grundschulalter. Kognitive Grundlagen und Risikofaktoren. *Lernen und Lernstörungen*, 7(1), 33–34. <https://doi.org/10.1024/2235-0977/a000185>
- Nation, K. (2005). Children's Reading Comprehension Difficulties. In M.J. Snowling & C. Hulme (Hrsg.), *The science of reading: a handbook* (S. 248–265). Malden, MA: Blackwell Pub. Zugriff am 12.7.2016. Verfügbar unter: [http://www.blackwellreference.com/subscriber/uid=3/book?show=all&id=g9781405114882\\_9781405114882](http://www.blackwellreference.com/subscriber/uid=3/book?show=all&id=g9781405114882_9781405114882)
- Niklas, F., Möllers, K. & Schneider, W. (2013). Die frühe familiäre Lernumwelt als Mediator zwischen strukturellen Herkunftsmerkmalen und der basalen Lesefähigkeit am Ende der ersten Klasse. *Psychologie in Erziehung und Unterricht*, 60(2), 94–111. <https://doi.org/10.2378/peu2013.art08d>
- Nuerk, H.-C., Zürcher, A. & Gawrilow, C. (2015). Schulpsychologische Diagnostik: Eine Einführung. *Lernen und Lernstörungen*, 4(4), 233–243. <https://doi.org/10.1024/2235-0977/a000113>
- Oerke, B., McElvany, N., Ohle, A., Ullrich, M. & Horz, H. (2016). Verbessert sich die diagnostische Urteilsgenauigkeit von Lehrkräften bei längerem Kontakt mit der Klasse? *Psychologie in Erziehung und Unterricht*, 63(1), 34–47. <https://doi.org/10.2378/peu2016.art04d>
- Ohle, A. & McElvany, N. (2015). Teachers' diagnostic competences and their practical relevance. *Journal for educational research online*, 7(2), 5–10.
- Organisation for economic co-operation and development. (2000). *Literacy in the information age. Final report of the international adult literacy survey*. OECD Publishing.
- Oslund, E. L., Clemens, N. H., Simmons, D. C. & Simmons, L. E. (2018). The direct and indirect effects of word reading and vocabulary on adolescents' reading comprehension: Comparing struggling and adequate comprehenders. *Reading and Writing*, 31(2), 355–379. <https://doi.org/10.1007/s11145-017-9788-3>
- Paetsch, J., Radmann, S., Felbrich, A., Lehmann, R. & Stanat, P. (2016). Sprachkompetenz als Prädiktor mathematischer Kompetenzentwicklung von Kindern deutscher und nicht-

- deutscher Familiensprache. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48(1), 27–41. <https://doi.org/10.1026/0049-8637/a000142>
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension - fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117–275.
- Petermann, F. (2016). Sprachentwicklungsstörungen. *Kindheit und Entwicklung*, 25(3), 131–134. <https://doi.org/10.1026/0942-5403/a000196>
- Petermann, F. & Daseking, M. (Hrsg.). (2009). *Fallbuch HAWIK-IV*. Göttingen: Hogrefe.
- Petermann, F. & Petermann, U. (Hrsg.). (2010). *Hamburg-Wechsler-Intelligenztest für Kinder - IV: HAWIK-IV; Manual* (3., erg. Aufl.). Bern: Huber [u.a.].
- Peugh, J. L. & Enders, C. K. (2004). Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement. *Review of Educational Research*, 74(4), 525–556. <https://doi.org/10.3102/00346543074004525>
- Pfost, M. (2015). Children's Phonological Awareness as a Predictor of Reading and Spelling: A Systematic Review of Longitudinal Research in German-Speaking Countries. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(3), 123–138. <https://doi.org/10.1026/0049-8637/a000141>
- Philipp, M. (2012). Einige theoretische und begriffliche Grundlagen. In M. Philipp & A. Schilcher (Hrsg.), *Selbstreguliertes Lesen: Ein Überblick über wirksame Leseförderansätze* (S. 38–58). Seelze: Friedrich.
- Praetorius, A.-K. & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften (Pädagogische Psychologie und Entwicklungspsychologie). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: theoretische und methodische Weiterentwicklungen* (S. 13–18). Münster New York: Waxmann.
- Quinn, J. M. (2018). Differential identification of females and males with reading difficulties: A meta-analysis. *Reading and Writing*, 31(5), 1039–1061. <https://doi.org/10.1007/s11145-018-9827-8>
- R Core Team. (2019). *R: A language and environment for statistical computing*. Wien, Österreich. Verfügbar unter: <https://www.R-project.org/>
- Rauer, W. & Schuck, K.-D. (2003a). *Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen*. Göttingen: Beltz.
- Rauer, W. & Schuck, K.-D. (2003b). *Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen*. Göttingen: Beltz.
- Ready, D. D. & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: the role of child background and classroom context. *American Educational Research Journal*, 48(2), 335–360.
- Retelsdorf, J. & Möller, J. (2008). Familiäre Bedingungen und individuelle Prädiktoren der Lesekompetenz von Schülerinnen und Schülern. *Psychologie in Erziehung und Unterricht*, 55, 227–237.
- Retelsdorf, J., Schwartz, K. & Asbrock, F. (2015). "Michael can't read!" Teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology*, 107(1), 186–194. <https://doi.org/10.1037/a0037107>
- Richlan, F., Kronbichler, M. & Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Human Brain Mapping*, 30(10), 3299–3308. <https://doi.org/10.1002/hbm.20752>
- Richter, T. & Müller, B. (in Druck). Entwicklung hierarchieniedriger Leseprozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 1–23). Weinheim: Juventa. Zugriff am 18.6.2018. Verfügbar unter:

- [http://www.i4.psychologie.uni-wuerzburg.de/fileadmin/06020400/Richter\\_Muel-ler\\_2016\\_Entwicklung\\_hierarchieniedriger\\_Lese-faehigkeiten.pdf](http://www.i4.psychologie.uni-wuerzburg.de/fileadmin/06020400/Richter_Muel-ler_2016_Entwicklung_hierarchieniedriger_Lese-faehigkeiten.pdf)
- Ricken, G. (2014). Lerntherapie geht in die Schule: Überlegungen zu einer Ressourcenverknüpfung. *Lernen und Lernstörungen*, 3(3), 179–184. <https://doi.org/10.1024/2235-0977/a000074>
- Robitzsch, A., Grund, S. & Henke, T. (2019). *miceadds: Some additional multiple imputation functions, especially for mice. R package version 3.4-17*. Verfügbar unter: <https://CRAN.R-project.org/package=miceadds>
- Ronniger, P., Melzer, J., Petermann, F. & Rißling, J.-K. (2016). Klassifikation von Sprachentwicklungsstörungen. *Kindheit und Entwicklung*, 25(3), 135–144. <https://doi.org/10.1026/0942-5403/a000197>
- Roos, S., Lohbeck, A., Petermann, F., Petermann, U., Schultheiß, J., Nitkowski, D. et al. (2016). Fremd- und Selbsturteile von Lehrern und Schülern im Rahmen psychologischer Diagnostik. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 64(3), 199–207. <https://doi.org/10.1024/1661-4747/a000279>
- Rosebrock, C. & Nix, D. (2015). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung* (7., überarb. und erw. Aufl.). Baltmannsweiler: Schneider-Verl. Hohengehren.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41–55.
- Rost, D. H. (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien: eine Einführung* (UTB Erziehungswissenschaften, Psychologie) (3., vollständig überarbeitete und erweiterte Auflage.). Bad Heilbrunn: Verlag Julius Klinkhardt.
- RStudio Team. (2019). *RStudio: Integrated Development for R. RStudio*. Boston, MA. Verfügbar unter: <https://www.rstudio.com/>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* (Wiley series in probability and mathematical statistics). New York: Wiley.
- Sauer, S. (2019). *Moderne Datenanalyse mit R: Daten einlesen, aufbereiten, visualisieren, modellieren und kommunizieren* (FOM-Edition). Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-21587-3>
- Scerri, T. S. & Schulte-Körne, G. (2010). Genetics of developmental dyslexia. *European Child & Adolescent Psychiatry*, 19(3), 179–197. <https://doi.org/10.1007/s00787-009-0081-0>
- Schabmann, A., Schmidt, B. M., Klicpera, C., Gasteiger-Klicpera, B. & Klingebiel, K. (2009). Does systematic reading instruction impede prediction of reading a shallow orthography? *Psychology Science Quarterly*, 51(3), 315–338.
- Schilcher, A. (2012). Zur Integration der Leseforschung in den deutschdidaktischen Diskurs. In M. Philipp & A. Schilcher (Hrsg.), *Selbstreguliertes Lesen: Ein Überblick über wirksame Leseförderansätze* (S. 19–37). Seelze: Friedrich.
- Schmidt, F. (2015). Den diagnostischen Blick schärfen - Vorstellungen und Orientierungen von Deutschlehrerinnen und Deutschlehrern zur Diagnose von Lesekompetenz. In C. Bräuer & D. Wieser (Hrsg.), *Lehrende im Blick: empirische Lehrerforschung in der Deutschdidaktik* (S. 89–109). Wiesbaden: Springer VS.
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik* (Springer-Lehrbuch) (5., vollständig überarbeitete und erweiterte Auflage.). Berlin Heidelberg: Springer.
- Schneider, W., Blanke, I., Faust, V. & Küspert, P. (2011). *Würzburger Leise Leseprobe - Revision (WLLP-R)*. Göttingen: Hogrefe Verlag.

- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2012). *Skalen zur Erfassung des schulischen Selbstkonzepts: SESSKO (2., überarbeitete und neu normierte Auflage.)*. Göttingen: Hogrefe.
- Schoppek, W. (2015). Mehrebenenanalyse oder Varianzanalyse?: Ein simulationsbasierter Vergleich von Verfahren zur Auswertung pädagogisch-psychologischer Experimente. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(4), 199–209. <https://doi.org/10.1026/0049-8637/a000136>
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerbildung*, 31(2), 154–165.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27–52.
- Schrader, F.-W. & Helmke, A. (2014). Alltägliche Leistungsbeurteilungen durch Lehrer (Beltz Pädagogik). In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (3. aktualisierte Auflage., S. 45–58). Weinheim: Beltz.
- Schrader, F.-W., Helmke, A., Hosenfeld, I., Halt, A. C. & Hochweber, J. (2006). Komponenten der Diagnosegenauigkeit von Lehrkräften. In F. Eder, A. Gastager & F. Hofmann (Hrsg.), *Qualität durch Standards? Beiträge zum Schwerpunktthema der 67. Tagung der AEPF* (S. 265–278). Münster: Waxmann.
- Schründer-Lenzen, A. (2013). *Schriftspracherwerb* (Lehrbuch) (4., völlig überarbeitete Auflage.). Wiesbaden: Springer VS.
- Schuchardt, K., Brandenburg, J., Fischbach, A. & Mähler, C. (2017). Kognitive Profile bei lese-rechtschreibschwachen Kindern mit und ohne Aufmerksamkeitsprobleme. *Lernen und Lernstörungen*, 6(4), 169–181. <https://doi.org/10.1024/2235-0977/a000188>
- Schuchardt, K., Fischbach, A., Balke-Melcher, C. & Mähler, C. (2015). Die Komorbidität von Lernschwierigkeiten mit ADHS-Symptomen im Grundschulalter. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 43(3), 185–193. <https://doi.org/10.1024/1422-4917/a000352>
- Schuchardt, K. & Mähler, C. (2016). Exekutive Funktionen bei Kindern mit Lernstörungen. *Praxis der Kinderpsychologie und Kinderpsychiatrie - Ergebnisse aus Psychotherapie, Beratung und Psychiatrie*, 65(6), 389–405.
- Schulte-Körne, G. (2010). Diagnostik und Therapie der Lese-Rechtschreib-Störung. *Deutsches Ärzteblatt*, 107(41), 718–727. <https://doi.org/10.3238/arztebl.2010.0718>
- Schulte-Körne, G. (2014). Spezifische Lernstörungen: Vom DSM-IV zum DSM-5. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 42(5), 369–374. <https://doi.org/10.1024/1422-4917/a000312>
- Schulte-Körne, G., Deimel, W., Jungermann, M. & Remschmidt, H. (2003). Nachuntersuchung einer Stichprobe von lese-rechtschreibgestörten Kindern im Erwachsenenalter. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 31(4), 267–276. <https://doi.org/10.1024/1422-4917.31.4.267>
- Schulte-Körne, G., Galuschka, K., & Deutsche Gesellschaft für Kinder und Jugendpsychiatrie, Psychosomatik und Psychotherapie e.V. (DGKJP) (Hrsg.). (2015). *Diagnostik und Behandlung von Kindern und Jugendlichen mit Lese- und / oder Rechtschreibstörung. S3-Leitlinie AWMF-Register-Nr. 028-044*. Zugriff am 31.7.2018. Verfügbar unter: <https://www.awmf.org/leitlinien/detail/II/028-044.html>
- Schulte-Körne, G. & Remschmidt, H. (2003). Legasthenie – Symptomatik, Diagnostik, Ursachen, Verlauf und Behandlung. *Deutsches Ärzteblatt*, 100(7), A 396-A 406.

- Schulz, W., Dertmann, J. & Jagla, A. (2003). Kinder mit Lese- und Rechtschreibstörungen: Selbstwertgefühl und Integrative Lerntherapie. *Kindheit und Entwicklung*, 12(4), 231–242. <https://doi.org/10.1026//0942-5403.12.4.231>
- Schwab, S. & Gasteiger-Klicpera, B. (2014). Förderung der Lesekompetenzen bei Kindern der zweiten Schulstufe – Evaluierung eines differenzierten Sprach- und Leseförderprogramms im Rahmen des Grundschulunterrichts. *Zeitschrift für Bildungsforschung*, 4(1), 63–79. <https://doi.org/10.1007/s35834-013-0082-4>
- Seidenberg, M. S. (2005). Connectionist models of word reading. *Current directions in psychological science*, 14(5), 238–242.
- Select Statistical Services Ltd. (o. J.). Odds ratio – Confidence Interval. Verfügbar unter: <https://select-statistics.co.uk/calculators/confidence-interval-calculator-odds-ratio/>
- Seyda, M. (2018). Können Sportlehrkräfte die Perspektive ihrer Schülerinnen und Schüler einnehmen? Eine Untersuchung über die Akkuratheit von Beurteilungen physischer Fähigkeitsselbstwahrnehmungen. *Unterrichtswissenschaft*, 1–17. <https://doi.org/10.1007/s42010-017-0006-4>
- Skrowonek, H. & Marx, H. (1989). Die Bielefelder Längsschnittstudie zur Früherkennung von Risiken der Lese-Rechtschreibschwäche: Theoretischer Hintergrund und erste Befunde. *Heilpädagogische Forschung*, 15(1), 38–49.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 85–95.
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (2012). *Skalen zur Erfassung der Lern- und Leistungsmotivation: SELLMO* (2., überarbeitete und neu normierte Auflage.). Göttingen: Hogrefe.
- Spinner, K. H. (2010). Lesekompetenz ausbilden, Lesestandards erfüllen. In G. Schulz (Hrsg.), *Lesen lernen in der Grundschule: Lesekompetenz und Leseverstehen; Förderung und Bücherwelten* (1. Auflage, S. 48–61). Berlin: Cornelsen Scriptor.
- Spörer, N., Brunstein, J. C. & Arbeiter, K. (2007). Förderung des Leseverständnisses in Lerntandems und in Kleingruppen: Ergebnisse einer Trainingsstudie zu Methoden des reziproken Lehrens. *Psychologie in Erziehung und Unterricht*, 54, 298–313.
- Stahns, R. & Rieser, S. (2018). Qualität des Leseunterrichts in vierten Klassen in der Grundschule unter den Bedingungen von Mehrsprachigkeit. Ergebnisse der Lehrkräftebefragung von IGLU 2011. *Zeitschrift für Grundschulforschung*, 11(1), 131–145. <https://doi.org/10.1007/s42278-018-0007-3>
- Stang, J. & Urhahne, D. (2016a). Wie gut schätzen Lehrkräfte Leistung, Konzentration, Arbeits- und Sozialverhalten ihrer Schülerinnen und Schüler ein? Ein Beitrag zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 63(3), 204. <https://doi.org/10.2378/peu2016.art18d>
- Stang, J. & Urhahne, D. (2016b). Stabilität, Bezugsnormorientierung und Auswirkungen der Urteilsgenauigkeit. *Zeitschrift für Pädagogische Psychologie*, 30(4), 251–262. <https://doi.org/10.1024/1010-0652/a000190>
- Steinbrink, C. & Lachmann, T. (2014). *Lese-Rechtschreibstörung: Grundlagen, Diagnostik, Intervention*. Berlin: Springer VS.
- Stevens, C., Harn, B., Chard, D. J., Currin, J., Parisi, D. & Neville, H. (2013). Examining the Role of Attention and Instruction in At-Risk Kindergarteners: Electrophysiological Measures of Selective Auditory Attention Before and After an Early Literacy Intervention. *Journal of Learning Disabilities*, 46(1), 73–86. <https://doi.org/10.1177/0022219411417877>

- Stiensmeier-Pelster, J. & Schöne, C. (o. J.). *Fragebogen zu schulischem Selbstkonzept und Lernmotivation für die erste Klasse*. Unveröffentlichte Rohdaten. Gießen: Justus-Liebig-Universität.
- Stock, C. & Schneider, W. (2008a). *DERET 1-2+. Deutscher Rechtschreibtest für das erste und zweite Schuljahr*. Göttingen: Hogrefe.
- Stock, C. & Schneider, W. (2008b). *DERET 3-4+. Deutscher Rechtschreibtest für das dritte und vierte Schuljahr*. Göttingen: Hogrefe.
- Strehlow, U. & Haffner, J. (2007). Störungen/Schwierigkeiten des Lesenlernens (Dyslexie - Legasthenie) und des Rechtschreibens aus kinderpsychiatrischer Perspektive (Handbuch Sonderpädagogik). In H. Schöler & A. Welling (Hrsg.), *Sonderpädagogik der Sprache* (S. 396–418). Göttingen: Hogrefe.
- Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, 80(1), 99–103.  
[https://doi.org/10.1207/S15327752JPA8001\\_18](https://doi.org/10.1207/S15327752JPA8001_18)
- Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25(1), 1–21. <https://doi.org/10.1214/09-STS313>
- von Suchodoletz, W. (2007). Welche Behandlung ist bei der Legasthenie wirksam? *Monatsschrift Kinderheilkunde*, 155(4), 351–356. <https://doi.org/10.1007/s00112-007-1478-9>
- von Suchodoletz, W. (2010). Konzepte in der LRS-Therapie. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 38(5), 329–339. <https://doi.org/10.1024/1422-4917/a000057>
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>
- Südkamp, A., Kaiser, J. & Möller, J. (2017). Ein heuristisches Modell der Akkuratheit diagnostischer Urteile von Lehrkräften (Pädagogische Psychologie und Entwicklungspsychologie). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: theoretische und methodische Weiterentwicklungen* (S. 33–38). Münster New York: Waxmann.
- Südkamp, A. & Möller, J. (2009). Referenzgruppeneffekte im Simulierten Klassenraum: Direkte und indirekte Einschätzungen von Schülerleistungen. *Zeitschrift für Pädagogische Psychologie*, 23(34), 161–174. <https://doi.org/10.1024/1010-0652.23.34.161>
- Tarelli, I., Lankes, E.-M., Drossel, K. & Gegenfurtner, A. (2012). Lehr- und Lernbedingungen an Grundschulen im internationalen Vergleich. In W. Bos, I. Buddeberg, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 137–173). Münster / New York / München / Berlin: Waxmann.
- Tarelli, I., Valtin, R., Bos, W., Bremerich-Vos, A. & Schwippert, K. (2012). IGLU 2011: Wichtige Ergebnisse im Überblick. In W. Bos, I. Buddeberg, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 11–25). Münster / New York / München / Berlin: Waxmann.
- Therrien, W. J. (2004). Fluency and Comprehension Gains as a Result of Repeated Reading: A Meta-Analysis. *Remedial and Special Education*, 25(4), 252–261.  
<https://doi.org/10.1177/07419325040250040801>
- Thoemmes, F. (2012). Propensity score matching in SPSS. *arXiv:1201.6385 [stat]*.

- Thoemmes, F. J. & West, S. G. (2011). The Use of Propensity Scores for Nonrandomized Designs With Clustered Data. *Multivariate Behavioral Research*, 46(3), 514–543. <https://doi.org/10.1080/00273171.2011.569395>
- Tiedemann, J. & Billmann-Mahecha, E. (2007). Leseverständnis, Familiensprache und Freizeitsprache: Ergebnisse aus der Hannoverschen Grundschulstudie. *Zeitschrift für Pädagogische Psychologie*, 21(1), 41–49. <https://doi.org/10.1024/1010-0652.21.1.41>
- Tischler, T., Daseking, M. & Petermann, F. (2017). Kognitive Fähigkeiten und Lesekompetenz: Zur Bedeutung der Intelligenz für die Güte der Lesefertigkeit. *Kindheit und Entwicklung*, 26(1), 48–57. <https://doi.org/10.1026/0942-5403/a000215>
- Torchiano, M. (2016). *Effsize - A Package For Efficient Effect Size Computation*. <https://doi.org/10.5281/ZENODO.1480624>
- Torchiano, M. (2019, Juli 18). Package „effsize“ - Efficient Effect Size Computation. Zugriff am 13.11.2019. Verfügbar unter: <https://cran.r-project.org/web/packages/effsize/effsize.pdf>
- Tornow, H. (2012). *Abschlussbericht für das Projekt: Untersuchung zur Bestimmung steuerungsrelevanter Wirkungsfaktoren im Hilfeplanprozess im Rahmen des gesamtstädtischen Fachcontrollings. Hilfen zur Erziehung in Berlin*. (S. 1–31). Verfügbar unter: [https://www.berlin.de/sen/jugend/familie-und-kinder/hilfe-zur-erziehung/fachinfo/fachcontrolling-hilfen-zur-erziehung/mdb-sen-jugend-jugendhilfeleistungen-hilfen\\_zur\\_erziehung-abschlussbericht\\_wirkungsfaktoren.pdf](https://www.berlin.de/sen/jugend/familie-und-kinder/hilfe-zur-erziehung/fachinfo/fachcontrolling-hilfen-zur-erziehung/mdb-sen-jugend-jugendhilfeleistungen-hilfen_zur_erziehung-abschlussbericht_wirkungsfaktoren.pdf)
- Trautwein, U. & Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind: Referenzgruppeneffekte bei Übertrittsentscheidungen. *Zeitschrift für Pädagogische Psychologie*, 21(2), 119–133. <https://doi.org/10.1024/1010-0652.21.2.119>
- Treutlein, A., Roos, J. & Schöler, H. (2012). Lese- und Rechtschreibleistungen und ihre Abhängigkeit von individuellen Deutschkenntnissen, sprachlicher Zusammensetzung der Klasse und Klassengröße. *Psychologie in Erziehung und Unterricht*, 59(2), 122–131. <https://doi.org/10.2378/peu2012.art07d>
- Urban, D., Mayerl, J. & Wahl, A. (2016). Regressionsanalyse bei fehlenden Variablenwerten (missing values): Imputation oder Nicht-Imputation?: eine Anleitung für die Regressionsspraxis mit SPSS. <https://doi.org/10.18419/opus-9112>
- Urhahne, D., Zhou, J., Stobbe, M., Chao, S.-H., Zhu, M. & Shi, J. (2010). Motivationale und affektive Merkmale unterschätzter Schüler: Ein Beitrag zur diagnostischen Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 24(3–4), 275–288. <https://doi.org/10.1024/1010-0652/a000021>
- Valtin, R. (2017). Kapitel XII. Einordnung der IGLU-2016-Befunde in das europäische Rahmenkonzept für gute Leseförderung. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 315–328). Münster / New York: Waxmann.
- Valtin, R., Hornberg, S., Buddeberg, M., Voss, A., Kowoll, M. E. & Potthoff, B. (2010). Schülerinnen und Schüler mit Leseproblemen - eine ökosystemische Betrachtungsweise. In W. Bos, K. Schwippert & K.-H. Arnold (Hrsg.), *IGLU 2006 - die Grundschule auf dem Prüfstand: vertiefende Analysen zu Rahmenbedingungen schulischen Lernens* (S. 43–90). Münster: Waxmann.
- Vander Stappen, C. & Reybroeck, M. V. (2018). Phonological Awareness and Rapid Automated Naming Are Independent Phonological Competencies With Specific Impacts on Word Reading and Spelling: An Intervention Study. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00320>

- Victor, A., Elsässer, A., Hommel, G. & Blettner, M. (2010). Wie bewertet man die p-Wert-Flut? *Deutsches Ärzteblatt*, 107(4), 50–56. <https://doi.org/10.3238/arztebl.2010.0050>
- Viera, A. J. (2008). Odds Ratios and Risk Ratios: What's the Difference and Why Does It Matter? *Southern Medical Journal*, 101(7), 730–734. <https://doi.org/10.1097/SMJ.0b013e31817a7ee4>
- Villiger, C., Niggli, A., Wandeler, C. & Kutzelnann, S. (2012). Does family make a difference? Mid-term effects of a school/home-based intervention program to enhance reading motivation. *Learning and Instruction*, 22(2), 79–91. <https://doi.org/10.1016/j.learninstruc.2011.07.001>
- Voß, S., Blumenthal, Y., Sikora, S., Mahlau, K., Diehl, K. & Hartke, B. (2014). Rügener Inklusionsmodell (RIM) – Effekte eines Beschulungsansatzes nach dem Response to Intervention-Ansatz auf die Rechen- und Leseleistungen von Grundschulkindern. *Empirische Sonderpädagogik*, 2, 114–132.
- Walter, S. G. & Rack, O. (2007). Eine anwendungsbezogene Einführung in die Hierarchische Lineare Modellierung (HLM). In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (2., überarb. und erw. Aufl., S. 277–292). Wiesbaden: Gabler.
- Wandeler, C., Niggli, A., Trautwein, U. & Villiger Hugo, C. (2013). Hausaufgabenbetreuung als Begleitmaßnahme schulischer Leseförderung: Präventive Effekte eines Elternkurses. *Psychologie in Erziehung und Unterricht*, 60(Preprint 2013), 1–14. <https://doi.org/10.2378/peu2013.art13d>
- Wechsler, D. (1958). *The measurement and appraisal for adult intelligence*. Baltimore: Williams & Wilkins.
- Weinert, F. E. (2000). Lehren und Lernen für die Zukunft - Ansprüche an das Lernen in der Schule. *Pädagogische Nachrichten Rheinland-Pfalz*, 2, 1–16.
- Weinert, F. E. & Schrader, F.-W. (1986). Diagnose des Lehrers als Diagnostiker. In H. Petillon, J. W. L. Wagner & B. Wolf (Hrsg.), *Schülergerechte Diagnose: theoretische und empirische Beiträge zur pädagogischen Diagnostik: Festschrift zum 60. Geburtstag von Karlheinz Ingenkamp* (S. 11–29). Weinheim: Beltz.
- Wendt, H. & Schwippert, K. (2017). Kapitel VII. Lesekompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 219–234). Münster / New York: Waxmann.
- Wilhelm, O. & Kunina-Habenicht, O. (2015). Pädagogisch-psychologische Diagnostik (Springer-Lehrbuch). In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 305–328). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. Zugriff am 27.1.2015. Verfügbar unter: <http://link.springer.com/10.1007/978-3-642-41291-2>
- de Winter, J. C. F. (2013). Using the Student's t-test with extremely small sample sizes. *Practical assessment, research and evaluation*, 18(10), 1–12.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe, Verl. für Psychologie.
- Wirtz, M. & Kutschmann, M. (2007). Analyse der Beurteilerübereinstimmung für kategoriale Daten mittels Cohens Kappa und alternativer Maße. *Die Rehabilitation*, 46(6), 370–377. <https://doi.org/10.1055/s-2007-976535>

## Literaturverzeichnis

- Wolf, K. M., Schroeders, U. & Kriegbaum, K. (2016). Metaanalyse zur Wirksamkeit einer Förderung der phonologischen Bewusstheit in der deutschen Sprache. *Zeitschrift für Pädagogische Psychologie*, 30(1), 9–33. <https://doi.org/10.1024/1010-0652/a000165>
- Wolf, M. & Denckla, M. B. (2005). *RAN/RAS: Rapid Automatized Naming and Rapid Alternating Stimulus Tests*. Austin, Tx: Pro-Ed.
- Wyschkon, A., Schulz, F., Gallit, F. S., Poltz, N., Kohn, J., Moraske, S. et al. (2018). 5-Jahres-Verlauf der LRS: Stabilität, Geschlechtseffekte, Schriftsprachniveau und Schulerfolg. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 46(2), 107–122. <https://doi.org/10.1024/1422-4917/a000535>
- Yamada, Y., Stevens, C., Dow, M., Harn, B. A., Chard, D. J. & Neville, H. J. (2011). Emergence of the neural network for reading in five-year-old beginning readers of different levels of pre-literacy abilities: An fMRI study. *NeuroImage*, 57(3), 704–713. <https://doi.org/10.1016/j.neuroimage.2010.10.057>

## Anhang A: Itemanalysen der Testbatterie von Koch und Georg (2010)

### I. Silbensegmentation

Tabelle 121

*Trennschärfe und Itemschwierigkeit für Silbensegmentation in Bildern*

	Trennschärfe		Itemschwierigkeit	
	T1	T2	T1	T2
Item 1 Löwe	.63	.71	.77	.80
Item 2 Melone	.39	.59	.67	.71
Item 3 Igel	.64	.69	.78	.76
Item 4 Baum	.54	.63	.49	.58
Item 5 Kuh	.52	.60	.66	.71
Item 6 Elefant	.59	.62	.75	.67
Item 7 Ameise	.56	.72	.75	.79
Item 8 Schokolade	.53	.51	.75	.71
Item 9 Krokodil	.*	.65	.*	.62
Item 10 Schmetterling	.*	.66	.*	.65

*Anmerkungen.* \* Items 9 und 10 wurden der Skala erst zu T2 hinzugefügt.

Tabelle 122

*Trennschärfe für Silbensegmentation mit Wörtern für 43 Items*

	T4	T5	T6	T7
Item 1 Kinder	.27	.12	.24	.24
Item 2 Flasche	.19	.26	.26	.17
Item 3 ausmalen	.20	.19	.29	.24
Item 4 Regen	.30	.25	.20	.06
Item 5 Gabel	.44	.29	.22	.27
Item 6 Musik	.44	.30	.30	.13
Item 7 Blumentopf	.31	.31	.26	.18
Item 8 Esel	.29	.23	.36	.24
Item 10 schlafen	.40	.45	.49	.46
Item 11 Sonnenschirm	.48	.56	.50	.41
Item 12 schreiben	.49	.54	.55	.28
Item 13 Augen	.57	.54	.60	.15
Item 14 Autobahn	.49	.58	.65	.48
Item 15 glücklich	.38	.51	.52	.42
Item 16 Wasserhahn	.46	.62	.65	.52
Item 17 denken	.41	.56	.64	.53
Item 18 Telefon	.38	.52	.60	.67
Item 19 Bilderbuch	.48	.54	.62	.62
Item 20 Ameise	.26	.43	.59	.65
Item 21 Kaugummi	.18	.48	.50	.64
Item 22 Pinsel	.41	.31	.45	.64
Item 23 Telefonbuch	.31	.39	.43	.61
Item 24 Marmelade	.17	.38	.23	.61
Item 25 Feuerwehr	.29	.31	.12	.57
Item 26 Briefmarke	.20	.28	.15	.50
Item 27 Geburtstagskuchen	-	-	.17	.48
Item 28 Nudelsuppe	-	-	.00	.28
Item 29 laufen	.22	.13	.04	.26
Item 30 Erdbeere	-	.08	.07	.20
Item 31 Wochenende	-	.11	.08	.12
Item 32 Wolke	-	.11	.09	.19
Item 33 spielen	.22	.09	.11	.10
Item 34 schwimmen	-	.06	.09	.07
Item 35 lernen	-	-	-	.00
Item 36 Briefträger	-	-	-	.08
Item 37 Federball	.15	.05	.05	.08
Item 38 Osterhase	.18	.07	-	.08

## Anhang A: Itemanalysen der Testbatterie von Koch und Georg (2010)

	T4	T5	T6	T7
Item 39 helfen	-	-	-	.05
Item 40 Regenwurm	-	-	-	.08
Item 41 langsam	.23	.09	-	.08
Item 42 Blumenwiese	-	-	-	.09
Item 43 vorlesen	-	-	-	.08
Item 44 Einkaufswagen	-	-	-	.08

Anmerkungen. Zellen ohne Werte bedeuten, dass kein Kind das Item korrekt bearbeiten konnte.

Tabelle 123

Trennschärfe und Itemschwierigkeit für Silbensegmentation mit Wörtern der finalen 25 Items

	Trennschärfe				Itemschwierigkeit			
	T4	T5	T6	T7	T4	T5	T6	T7
Item 1 Kinder	.26	.12	.24	.26	73	78	84	92
Item 2 Flasche	.21	.29	.28	.27	69	78	82	88
Item 3 ausmalen	.24	.28	.33	.46	29	37	55	79
Item 4 Regen	.34	.33	.25	.32	64	75	83	94
Item 5 Gabel	.44	.31	.26	.25	71	82	87	97
Item 6 Musik	.48	.34	.32	.31	68	79	88	96
Item 7 Blumentopf	.38	.39	.32	.41	37	52	68	89
Item 8 Esel	.32	.26	.38	.36	21	30	39	76
Item 10 schlafen	.41	.46	.50	.45	32	54	63	89
Item 11 Sonnenschirm	.51	.57	.51	.53	6	18	39	76
Item 12 schreiben	.52	.59	.57	.48	18	40	55	89
Item 13 Augen	.61	.58	.62	.35	22	45	57	92
Item 14 Autobahn	.48	.57	.66	.49	9	19	32	77
Item 15 glücklich	.36	.50	.51	.49	11	25	34	79
Item 16 Wasserhahn	.48	.66	.67	.64	1	10	13	67
Item 17 denken	.43	.58	.64	.62	10	19	28	71
Item 18 Telefon	.32	.47	.59	.63	2	9	14	63
Item 19 Bilderbuch	.44	.51	.63	.69	1	6	12	61
Item 20 Ameise	.28	.45	.58	.67	1	2	8	42
Item 21 Kaugummi	.19	.48	.49	.68	1	2	4	40
Item 22 Pinsel	.34	.24	.40	.52	7	12	10	43
Item 23 Telefonbuch	.26	.33	.39	.54	1	2	2	31
Item 24 Marmelade	.18	.37	.18	.57	0	0	1	18
Item 25 Feuerwehr	.30	.30	.05	.53	3	1	0	21
Item 26 Briefmarke	.08	.16	.04	.25	1	2	1	14

## II. Satzsegmentation

Tabelle 124

Trennschärfe und Itemschwierigkeit für die Satzsegmentation

	Trennschärfe			Itemschwierigkeit		
	T4	T5	T7	T4	T5	T7
Satz 1	.55	.53	.67	74	92	98
Satz 2	.63	.58	.68	55	78	94
Satz 3	.62	.60	.55	65	78	95
Satz 4	.72	.61	.64	61	85	96
Satz 5	.74	.68	.66	48	73	95
Satz 6	.68	.70	.60	19	28	80
Satz 7	.63	.72	.70	18	40	92
Satz 8	.56	.69	.62	5	17	75
Satz 9	.42	.63	.65	5	15	80
Satz 10	.36	.43	.58	3	4	69

### III. Buchstabenkenntnis

Tabelle 125

*Trennschärfe und Itemschwierigkeit für die Buchstabenkenntnis*

	Trennschärfe		Itemschwierigkeit	
	T1	T3	T1	T3
Item 1 N und E	.14	-	96	-
Item 2 S und I	.19	.11	96	100
Item 3 A und R	.27	.16	98	100
Item 4 U und T	.49	.11	96	98
Item 5 D und H	.56	.06	75	98
Item 6 C und L	.40	.06	77	98
Item 7 M und G	.22	.18	58	91
Item 8 B und O	.37	.23	91	98
Item 9 F und W	.49	.25	88	99
Item 10 Z und K	.48	.30	84	98
Item 11 V und P	.38	.18	86	95
Item 12 J und Q	.37	.29	41	94
Item 13 X und Y	.08	.20	79	95

### IV. Laut-Buchstaben-Zuordnung

Tabelle 126

*Trennschärfe und Itemschwierigkeit für die Laut-Buchstaben-Zuordnung*

	Trennschärfe		Itemschwierigkeit	
	T1	T3	T1	T3
Item 1 e	.35	-	90	-
Item 2 l	.39	.11	86	100
Item 3 a	.30	.16	95	100
Item 4 s	.46	.11	69	98
Item 5 n	.62	.06	52	98
Item 6 r	.56	.06	57	98
Item 7 m	.46	.18	58	91
Item 8 i	.54	.23	79	98
Item 9 o	.20	.25	96	99
Item 10 t	.63	.30	53	98
Item 11 f	.59	.18	45	95
Item 12 k	.55	.36	38	96
Item 13 u	.54	.14	65	98
Item 14 h	.62	.41	30	96
Item 15 d	.61	.37	27	96
Item 16 g	.58	.40	18	92
Item 17 p	.58	.15	40	97
Item 18 b	.61	.29	40	94
Item 19 w	.61	.20	24	95
Item 20 v	.60	.33	41	91
Item 21 q	.42	.44	20	62
Item 22 z	.58	.48	40	88
Item 23 j	.42	.47	17	73
Item 24 c	.57	.46	30	65
Item 25 x	.52	.45	42	76
Item 26 y	.48	.42	35	70

V. Schreiben nach Bildern

Tabelle 127

*Trennschärfe und Itemschwierigkeit für Schreiben nach Bildern*

	Trennschärfe					Itemschwierigkeit				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
Item 1 so	.59	.37	.22	.00	.16	47	90	98	100	99
Item 2 fa	.57	.33	.11	.04	.18	32	85	87	83	84
Item 3 mund	.55	.51	.46	.51	.43	2	22	46	62	76
Item 4 li	.59	.52	.29	.36	.33	15	59	77	76	85
Item 5 mo	.60	.40	.27	.15	.17	33	77	90	90	94
Item 6 na	.65	.50	.34	.26	.35	16	62	79	92	92
Item 7 de	.70	.67	.53	.46	.31	6	42	71	83	91
Item 8 turm	.55	.50	.52	.35	.28	5	37	70	85	89
Item 9 kamm	.77	.45	.37	.17	.29	0	7	17	26	37
Item 10 fe	.76	.39	.36	.03	.04	24	84	91	98	98
Item 11 der	.74	.56	.45	.38	.34	3	31	53	72	80
Item 12 leit				.48	.34				90	94
Item 13 ter				.44	.39				81	80
Item 14 strom				.42	.35				67	79
Item 15 bie				.43	.33				41	47
Item 16 ne				.22	.09				95	95
Item 17 hahn				.26	.26				15	23
Item 18 re				.11	.01				99	100
Item 19 gen				.47	.23				84	93
Item 20 schirm				.42	.33				53	71

*Anmerkungen.* Items 12 bis 20 wurden dem Test erst ab T4 hinzugefügt.

## Anhang B: Beurteilerübereinstimmungen

### I. Absolute Häufigkeiten

Tabelle 128

Absolute Häufigkeiten für die Antwortkategorien zu Frage 1: „Was fällt auf?“

	T1 n = 70	T2 n = 87	T3 n = 67	T4 n = 76	T5 n = 65	T6 n = 96	T7 n = 23
<b>Spezifisch auf den Schriftspracherwerb bezogen</b>							
Phonologische Bewusstheit	3   3	3   5	3   3	1   0	1   3	1   0	1   0
Lesegeschwindigkeit/-flüssigkeit		1   2	2   3	0   1		1   2	
Buchstabenkenntnis		6   4	2   2	1   0	1   0		
Laut-Buchstaben-Zuordnung		0   2				2   1	
Lesefertigkeit		0   1					
Leseverständnis			3   4	2   2		5   3	1   1
<b>Allgemeiner auf den Schriftspracherwerb bezogen</b>							
Lese-Rechtschreibstörung				1   2	3   3	6   8	4   5
Lesen allgemein		5   7	11   11	3   4	3   3	5   4	1   1
Schreiben allgemein		1   1	2   2	2   3	6   6	9   9	5   5
Deutsch allgemein		4   5	3   5	11   13	3   8	2   9	
Sprache	12   12	8   7	11   10	8   5	10   6	22   16	3   4
Intelligenz / FSP Lernen	3   5	4   3	2   2	3   4	6   6	7   8	3   3
Motorik	7   7		3   4	2   2	2   2	2   3	
<b>Verhalten und Motivation</b>							
Konzentrationsfähigkeit / ADHS	10   9	7   4	13   13	9   10	8   9	22   21	1   1
Verhaltensauffälligkeiten	11   15	7   8	11   12	6   12	6   8	4   9	
Arbeitsverhalten	4   9	4   7	1   5	6   11	4   4	2   7	0   2
soziale Interaktion	2   2	1   7	3   2	1   3	2   2	6   8	1   1
Motivation	1   3			3   3	0   1		
<b>Sonstige Auffälligkeiten</b>							
Emotionale Probleme	4   5	1   1	4   3	2   2	4   4	5   5	1   1
Wahrnehmungsstörungen	1   0	3   0	1   0		3   2	5   3	1   0
Neg. über das Elternhaus	1   0	5   4	6   7	3   3	5   4	2   3	
sonstige Auffälligkeiten	19   18	22   22	24   16	33   31	22   20	24   22	5   4
Gut im Schreiben		3   2					
Gut im Lesen		10   9	2   1			1   1	
<b>Sonstige positive Nennungen</b>							
Pos. über das Elternhaus		1   1				0   1	
Intelligenz des Kindes	3   2	2   3	1   1	0   2			1   1
Sonstige positive Auffälligkeiten	7   7		3   5	6   4	9   8	17   16	5   4

Anmerkungen. Die erste Zahl in jeder Zelle gibt die Anzahl der Häufigkeiten für Beurteilerin S an, die zweite die Häufigkeiten für Beurteilerin K. Leere Zellen bedeuten, keine Beurteilerin hat diese Kategorie gewählt. n = Zahl der Kinder zu jedem Messzeitpunkt, zu denen Aussagen getätigt wurden

Tabelle 129

Absolute Häufigkeiten für die Antwortkategorien zu Frage 2: „Was wird getan?“

	T1 n = 21	T2 n = 45	T3 n = 39	T4 n = 45	T5 n = 57	T6 n = 68	T7 n = 21
<b>Organisatorische innerschulische Aspekte</b>							
Förderung im Klassenzimmer	8   8	22   17	21   29	26   22	13   19	20   27	11   14
Lehrkraft fördert selbst	7   7	18   18	20   17	25   20	13   8	16   21	6   6
BFZ-Kraft fördert	0   3	3   10	1   15	2   15	5   9	8   24	4   5
Sonstige Person fördert	1   3	1   4	3   14	0   15	0   17	4   20	0   9
Förderunterricht		10   11	21   21	26   26	19   19	34   34	8   7
Einzelförderung	0   1	0   2	2   4	1   3	0   7	2   8	0   7
Gruppenförderung	0   1	0   4	0   11	0   21	1   12	4   25	2   8
Sonstige innerschulische Förderung	0   2	6   5	1   9	1   9	9   13	7   12	4   6
<b>Organisatorische Aspekte der Zusammenarbeit</b>							
Elterngespräche		5   8	5   5	5   6	3   2	12   14	
Außerschulische Hilfe	11   11	7   10	7   8	2   3	6   6	14   16	6   6
BFZ ist eingeschaltet	1   1	5   7	3   4	3   2	28   29	24   29	3   4

## Anhang B: Beurteilerübereinstimmungen

	T1 n = 21	T2 n = 45	T3 n = 39	T4 n = 45	T5 n = 57	T6 n = 68	T7 n = 21
<b>Inhaltliche Aspekte</b>							
Differenzierung	7   7	10   10	5   5	1   2	5   5	3   5	11   12
Lob / positive Verstärkung		2   4		4   2	2   2		
Motivationsförderung	0   4	0   7		0   2	1   2		
Förderplan		3   3		1   1	1   1	2   2	8   8
Spezifische Fördermethoden	4   4	7   6		1   1	1   1	1   1	4   2
Sonstige Inhalte	1   7	6   19	6   27	10   20	13   32	27   39	6   6
<b>Sonstiges</b>							
Klassenwiederholung		1   1	5   4	1   1	2   2	2   2	
Anspruch auf sonderpäd. Förderung		4   5	1   1	1   2	5   3	1   1	2   2
Sonstige professionelle Förderung	11   10	5   8	4   9	2   2	6   6	10   16	4   6

*Anmerkungen.* Die erste Zahl in jeder Zelle gibt die Anzahl der Häufigkeiten für Beurteilerin S an, die zweite die Häufigkeiten für Beurteilerin K. Leere Zellen bedeuten, keine Beurteilerin hat diese Kategorie gewählt. n = Zahl der Kinder zu jedem Messzeitpunkt, zu denen Aussagen getätigt wurden

## II. Beurteilerübereinstimmungen für alle Inhaltskategorien

Tabelle 130

*Beurteilerübereinstimmungen für die Inhaltskategorien zu Frage 1: „Was fällt auf?“*

		T1 n = 70	T2 n = 87	T3 n = 67	T4 n = 76	T5 n = 65	T6 n = 96	T7 n = 23
Phonologische Bewusstheit	K	1.00**	.48**	.65**	.00	.49**	.00	.00
	PÜ	100	95	97	99	97	99	96
	Y	1.00	.76	.84	-	1.00	-	-
Lesegeschwindigkeit / -flüssigkeit	K		.66**	.79**	.00		-.01	
	PÜ		99	99	99		97	
	Y		1.00	1.00	-		-1.00	
Buchstabenkenntnis	K		.58**	.49**	.00	.00		
	PÜ		95	97	99	99		
	Y		.80	.78	-	-		
Laut-Buchstaben-Zuordnung	K		.00				.66**	
	PÜ		98				99	
	Y		-				1.00	
Leseverständnis	K			.85**	1.00**		.74**	1.00**
	PÜ			99	100		98	100
	Y			1.00	1.00		1.00	1.00
Lese-Rechtschreibstörung	K				.66**	1.00**	.85**	.86**
	PÜ				99	100	98	96
	Y				1.00	1.00	1.00	1.00
Lesen allgemein	K		.82**	1.00**	.85**	1.00**	.88**	1.00**
	PÜ		98	100	99	100	99	100
	Y		1.00	1.00	1.00	1.00	1.00	1.00
Schreiben allgemein	K		1.00**	-.031	.79**	.82**	.88**	1.00**
	PÜ		100	94	99	97	98	100
	Y		1.00	-1.00	1.00	.89	.93	1.00
Deutsch allgemein	K		.88**	.47**	.90**	.51**	.15*	
	PÜ		99	94	97	92	91	
	Y		1.00	.73	1.00	1.00	.53	
Konzentrationsfähigkeit / ADHS	K	.82**	.71**	1.00**	.94**	.93**	.97**	1.00**
	PÜ	96	97	100	99	99	99	100
	Y	.88	1.00	1.00	1.00	1.00	1.00	1.00
Verhaltensauffälligkeiten	K	.62**	.93**	.84**	.50**	.68**	.43**	
	PÜ	89	99	96	90	94	93	
	Y	.73	1.00	.89	.74	.81	.74	
Arbeitsverhalten	K	.58**	.71**	.32**	.54**	.47**	.43**	.00
	PÜ	93	97	94	91	94	95	91
	Y	1.00	1.00	1.00	.76	.69	1.00	-
Soziale Interaktion	K	1.00**	.24**	.79**	.49**	1.00**	.85**	1.00**
	PÜ	100	93	99	97	100	98	100
	Y	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Motivation	K	.49**			1.00**	.00		

## Anhang B: Beurteilerübereinstimmungen

		T1 n = 70	T2 n = 87	T3 n = 67	T4 n = 76	T5 n = 65	T6 n = 96	T7 n = 23
	PÜ	97			100	99		
	Y	1.00			1.00	-		
	K	.88**	1.00**	.85**	1.00**	1.00**	1.00**	1.00**
Emotionale Probleme	PÜ	99	100	99	100	100	100	100
	Y	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	K	.00	.88**	.92**	1.00**	.88**	.80**	
Negative Aussagen über das Elternhaus	PÜ	99	99	99	100	99	99	
	Y	-	1.00	1.00	1.00	1.00	1.00	
	K	.67**	.76**	.65**	.78**	.72**	.60**	.59**
Sonstige Auffälligkeiten	PÜ	87	91	85	89	88	86	87
	Y	.70	.79	.79	.79	.74	.65	.67
	K		1.00**				.00	
Positive Aussagen über das Elternhaus	PÜ		100				99	
	Y		1.00				-	
	K	.79**	.79**	1.00**	.00			1.00**
Intelligenz des Kindes	PÜ	99	99	100	97			100
	Y	1.00	1.00	1.00	-			1.00
	K	.84**	.86**	.74**	.57**	.93**	.96**	.86**
Sonstige positive Auffälligkeiten	PÜ	97	95	97	95	99	99	96
	Y	.90	.90	1.00	.79	1.00	1.00	1.00

Anmerkungen. n = Zahl der Kinder zu jedem Messzeitpunkt, zu denen Aussagen getätigt wurden. Leere Zellen bedeuten, keine Beurteilerin hat diese Kategorie gewählt. κ = Cohens Kappa; PÜ = Prozentuale Übereinstimmung; Y = Yules Y. \*p < .05, \*\*p < .01. Kategorien mit nicht akzeptabler Beurteilerübereinstimmung sind grau hinterlegt. Nicht dargestellt sind Lesefertigkeit und Wahrnehmungsstörungen.

Tabelle 131

Beurteilerübereinstimmung für alle Kategorien zu Frage 2: „Was wird getan?“

		T1 n = 21	T2 n = 45	T3 n = 39	T4 n = 45	T5 n = 57	T6 n = 68	T7 n = 21	
<b>Organisatorische innerschulische Aspekte</b>									
Förderung / Unterstützung in der Klasse	K	.83**	.62**	.53**	.59**	.61**	.53**	.57**	
	PÜ	90	78	72	76	83	78	71	
	K	1.00**	.85**	.71**	.55**	.51**	.50**	.78**	
Lehrkraft fördert selbst	PÜ	100	91	85	73	84	79	90	
	K	.00	.40**	-.05	.17*	.28*	.18*	.34*	
	PÜ	86	85	59	71	84	67	76	
BFZ-Kraft fördert	K	.22*	-.02	.03	.00	.00	.09	.00	
	PÜ	86	89	62	67	70	70	57	
	K		.76**	.71**	1.00**	.83**	.92**	.73**	
Förderunterricht	PÜ		93	82	100	91	96	86	
	K	.00	.00	.65**	.48**	.00	.37**	.00	
	PÜ	95	96	95	96	88	91	67	
Einzelförderung	K	.00	.00	.00	.00	.05	.20**	.22*	
	PÜ	95	91	72	53	88	69	67	
	K	.00	.69**	.18**	.08	.44**	.37**	.32**	
Sonstige innerschulische Förderung	PÜ	91	93	80	80	83	84	71	
	<b>Organisatorische Aspekte der Zusammenarbeit</b>								
	Elterngespräche <sup>+</sup>	K		.73**	.77**	.90**	.79**	.81**	
PÜ			93	95	98	98	94		
Y			1.00	.84	1.00	1.00	.87		
Außerschulische Hilfe	K	.76**	.73**	.77**	.79**	.55**	.57**	.79**	
	PÜ	86	91	92	98	91	84	90	
	K	1.00**	.81**	.84**	.79**	.83**	.78**	.83**	
BFZ ist eingeschaltet <sup>+</sup>	PÜ	100	96	97	98	91	90	95	
	Y	1.00	1.00	1.00	1.00	.83	.85	1.00	
	<b>Inhaltliche Aspekte</b>								
Differenzierung	K	1.00**	1.00**	1.00**	.66**	.68**	.74**	.78**	
	PÜ	100	100	100	98	95	97	86	
	K		.65**		.65**	.74**			
Lob, positive Verstärkung	PÜ		96		96	98			

## Anhang B: Beurteilerübereinstimmungen

		T1 <i>n</i> = 21	T2 <i>n</i> = 45	T3 <i>n</i> = 39	T4 <i>n</i> = 45	T5 <i>n</i> = 57	T6 <i>n</i> = 68	T7 <i>n</i> = 21
Motivationsförderung	κ	.00	.00		.00	.66**		
	PÜ	81	84		96	98		
Förderplan	κ		.82**		1.00**	1.00**	1.00**	.73**
	PÜ		98		100	100	100	86
Spezifische Fördermethode genannt	κ	1.00**	.91**		1.00**	.50**	.50**	.64**
	PÜ	100	98		100	98	99	91
Sonstige Inhalte	κ	.07	.05	.11	.37**	.15	.36**	.58**
	PÜ	67	56	39	67	51	61	81
<b>Sonstiges</b>								
Klassenwiederholung <sup>+</sup>	κ		1.00**	.88**	1.00**	1.00**	1.00**	
	PÜ		100	98	100	100	100	
	Y		1.00	1.00	1.00	1.00	1.00	
ASF ist festgestellt <sup>+</sup>	κ		.88**	1.00**	.66**	.73**	1.00**	1.00**
	PÜ		98	100	98	97	100	100
	Y		1.00	1.00	1.00	1.00	1.00	1.00
Sonstige professionelle Förderung	κ	.68**	.74**	.31**	1.00**	.55**	.40**	.52**
	PÜ	81	93	80	100	91	79	81

*Anmerkungen.* *n* = Zahl der Kinder zu jedem Messzeitpunkt, zu denen Aussagen getätigt wurden. Leere Zellen bedeuten, keine Beurteilerin hat diese Kategorie gewählt. κ = Cohens Kappa; PÜ = Prozentuale Übereinstimmung; Y = Yules Y. \**p* < .05, \*\**p* < .01. <sup>+</sup>dichotome Variable. Kategorien mit nicht akzeptabler Beurteilerübereinstimmung sind grau hinterlegt.

## Anhang C: Berechnungen mit GPower

Test family	Statistical test	
t tests	Means: Difference between two independent means (two groups)	
Type of power analysis		
A priori: Compute required sample size - given $\alpha$ , power, and effect size		
Input Parameters		
Determine =>	Tail(s)	Two
	Effect size d	0.8
	$\alpha$ err prob	0.05
	Power (1- $\beta$ err prob)	0.80
	Allocation ratio N2/N1	1
Output Parameters		
Noncentrality parameter $\delta$	2.8844410	
Critical t	2.0085591	
Df	50	
Sample size group 1	26	
Sample size group 2	26	
Total sample size	52	
Actual power	0.8074866	

Test family	Statistical test	
t tests	Means: Difference between two independent means (two groups)	
Type of power analysis		
Post hoc: Compute achieved power - given $\alpha$ , sample size, and effect size		
Input Parameters		
Determine =>	Tail(s)	Two
	Effect size d	0.6
	$\alpha$ err prob	0.05
	Sample size group 1	9
	Sample size group 2	9
Output Parameters		
Noncentrality parameter $\delta$	1.2727922	
Critical t	2.1199053	
Df	16	
Power (1- $\beta$ err prob)	0.2236076	

Abbildung 13. A-priori-Berechnung des benötigten Stichprobenumfangs für die Gruppenvergleiche in Fragestellung 2 (oben) und Post-Hoc-Berechnung der Power des Gruppenvergleichs mit der größten Effektstärke von 0.60 in Fragestellung 2 (unten). Die Abbildungen finden sich in Originalgröße ebenfalls auf der Daten-CD.