

Ein Prozessmodell des Studieneinstiegs: Differentielle Aspekte studiumsbezogener Kognitionen und deren Effekte auf Studienerfolg und Studienabbruch

Zur Erlangung des akademischen Grades Dr. rer. nat. vorgelegte Dissertation

Pascale Stephanie Petri (geb. Bothe)

Justus-Liebig-Universität Gießen

Fachbereich Psychologie und Sportwissenschaften

Institut für Psychologische Diagnostik

Otto-Behaghel-Straße 10F

35394 Gießen

Einreichung der Dissertation:	17.12.2019
Betreuer der Arbeit:	Prof. Dr. Martin Kersting Prof. Dr. Alexander Minnaert
Gutachter*innen:	Prof. Dr. Martin Kersting Prof. Dr. Edith Braun Prof. Dr. Alexander Minnaert
Tag der Disputation:	22.07.2020
Mitglieder der Prüfungskommission:	Prof. Dr. Martin Kersting Prof. Dr. Edith Braun Prof. Dr. Alexander Minnaert Prof. Dr. Joachim Stiensmeier-Pelster

¹Anmerkung: Aus Gründen des Leseflusses wird das generische Maskulinum verwendet. Gleichwohl sind stets jegliche Geschlechter gemeint.

Inhaltsverzeichnis

1	Einleitung	15
1.1	Relevanz des Themas	15
1.1.1	Ausmaß des Studienabbruch in Deutschland und international	16
1.1.2	Methodische Herausforderungen bei der Quantifizierung von Studienabbruch	18
1.1.3	Gründe für einen Studienabbruch	19
1.1.4	Was geschieht aktuell auf dem Feld der Studienabbruch-Forschung?	20
1.1.5	Desiderate der Forschung zum Studienerfolg und -abbruch	25
2	Studienerfolg und -abbruch: Definition und empirische Befunde zu Prädiktoren	27
2.1	Empirische Befunde zum Studienerfolg	27
2.1.1	Definition von Studienerfolg	27
2.1.2	Noten und kognitive Fähigkeiten als Prädiktoren von Studienerfolg	29
2.1.3	Demografische und sozio-ökonomische Variablen als Prädiktoren von Studienerfolg	31
2.1.4	Interessen als Prädiktoren für Studienerfolg	33
2.1.5	Studienrelevante Persönlichkeitseigenschaften als Prädiktoren von Studienerfolg	33
2.1.6	Selbstwirksamkeit als Prädiktor von Studienerfolg	34
2.1.7	Emotionen im Leistungskontext als Prädiktoren von Studienerfolg	36
2.1.8	Motivation als Prädiktor von Studienerfolg	37
2.1.9	Soziale Integration und Unterstützung als Prädiktoren von Studienerfolg	38
2.1.10	Studentisches Engagement und Commitment als Prädiktoren von Studienerfolg	39
2.1.11	Selbstregulation als Prädiktor von Studienerfolg	40
2.1.12	Akademische Resilienz als Prädiktor von Studienerfolg	40
2.1.13	Institutionelle Faktoren als Prädiktoren von Studienerfolg	41
2.2	Zusammenfassende und vergleichende Erkenntnisse zu Prädiktoren von Studienerfolg	42
2.3	Empirische Befunde zum Studienabbruch	45
2.3.1	Definition von Studienabbruch	45
2.3.2	Noten als Prädiktoren von Studienabbruch	46
2.3.3	Selbstwirksamkeit als Prädiktor von Studienabbruch	46
2.3.4	Emotionen als Prädiktoren von Studienabbruch	47
2.3.5	Motivation als Prädiktor von Studienabbruch	47
2.4	Modelle und Theorien zum Studienabbruch	47
2.4.1	Unterschiedliche Perspektiven auf Studienerfolg versus -abbruch	48

2.4.2	Im Fokus: Vier etablierte Modelle	52
2.5	Zusammenfassung des aktuellen Forschungsstandes	60
2.6	Forschungslücke: Es fehlt ein integratives Prozessmodell	60
2.7	Ansatzpunkt der vorliegenden Arbeit	61
2.7.1	Intendierter Beitrag	61
3	Skalenentwicklung: Erfassung von Selbstwirksamkeit im Studieneinstieg	63
3.1	Überblick: Skalen zur Erfassung der Selbstwirksamkeit	63
3.2	Itementwicklung	64
3.2.1	Schritt 1: Akademische Herausforderungen im Studieneinstieg identifizieren	65
3.2.2	Schritt 2: Quantitativer und qualitativer Vortest	67
3.2.3	Die revidierte Version der SESW-Skala	68
3.3	Untersuchungen zur psychometrischen Qualität	69
3.3.1	Präanalysen	71
3.3.2	Itemanalyse nach KTT	72
3.3.3	Itemanalyse nach IRT	73
3.3.4	Untersuchungen zur Reliabilität	75
3.3.5	Untersuchungen zur Validität	78
3.4	Normierung	81
3.5	Weiterführende Analysen	81
3.5.1	Messinvarianz über die Jahrgänge hinweg	81
3.5.2	Entwicklung der SESW im Studieneinstieg	85
4	Empirische Studien	92
4.1	Studie 1 bis 3: Die Längsschnittstudien	92
4.1.1	Verwendete Instrumente	94
4.1.2	Repräsentativität der untersuchten Stichproben	104
4.1.3	Selektivität der untersuchten Stichproben	105
4.1.4	Dropout innerhalb der Studien	105
4.1.5	Einschlusskriterien	106
4.1.6	Fehlende Werte	106
4.1.7	Test auf <i>Missing Completely at Random</i> in den Längsschnittdaten	107
4.1.8	Imputationsverfahren	109
4.1.9	Modellfit testen	110
4.2	Studie 4: Sekundärdatenanalyse	111
4.3	Studie 5: Interventionsstudie	111
4.4	Verwendete Software	111
5	Studie 1	113
5.1	Messzeitpunkte bei Studie 1	113

5.2	Stichprobe bei Studie 1	113
5.3	Repräsentativität	114
5.4	Erhobene Variablen in Studie 1	115
5.5	Deskriptive Statistiken	115
5.6	Unterschiede zwischen den Fachclustern?	117
5.7	Korrelative Zusammenhänge	117
5.8	Modelltests	117
5.9	Zusammenfassung der Ergebnisse aus der ersten Längsschnittstudie	120
6	Entwicklung eines neuen, integrativen Modells	123
6.1	Aspekte, die im neuen Modell berücksichtigt werden sollten	123
6.2	Das EOS-Modell	125
6.2.1	Einordnung in die Literatur	126
6.3	Erster empirischer Test des <i>EOS</i> -Modells	126
6.4	Der Mehrwert des <i>EOS</i> -Modells	127
7	Studie 2	129
7.1	Messzeitpunkte bei Studie 2	129
7.2	Stichprobe bei Studie 2	129
7.3	Repräsentativität	130
7.4	Erhobene Variablen in Studie 2	131
7.5	Deskriptive Statistiken	131
7.6	Unterschiede zwischen den Fachclustern?	131
7.7	Korrelative Zusammenhänge	133
7.8	Modelltests	133
7.9	Zusammenfassung der Ergebnisse aus der zweiten Längsschnittstudie	137
8	Studie 3	139
8.1	Messzeitpunkte bei Studie 3	139
8.2	Stichprobe bei Studie 3	139
8.3	Repräsentativität	140
8.4	Erhobene Variablen in Studie 3	141
8.5	Deskriptive Statistiken	141
8.6	Unterschiede zwischen den Fachclustern?	141
8.7	Korrelative Zusammenhänge	143
8.8	Modelltests	143
8.9	Zusammenfassung der Ergebnisse aus der dritten Längsschnittstudie	146
9	Zusammenschau der Ergebnisse der drei Längsschnittstudien	148
9.1	Erkenntnisse aus den drei Längsschnittstudien	148

9.2	Unterschiede in der Höhe der Pfadkoeffizienten: Messfehler berücksichtigen	149
9.3	Messinvarianz über die drei Jahrgänge hinweg	150
9.3.1	Ergebnis der Messinvarianzprüfung	151
9.3.2	Bedeutung der metrischen Messinvarianz des <i>EOS</i> -Modells	151
10	Sekundärdatenanalyse	153
10.1	Vorteile großer bundesweiter Studierendenbefragungen	153
10.1.1	Forschungsfragen	153
10.2	Auswahl geeigneter Sekundärdaten	154
10.2.1	Die 21. Sozialerhebung	154
10.3	Methode	155
10.3.1	Auswahl eines Teildatensatzes	156
10.3.2	Art der Analysen	156
10.3.3	Analysestichprobe	157
10.3.4	Betrachtete Variablen	158
10.4	Ergebnisse	159
10.4.1	Deskriptive Statistiken	159
10.4.2	Präanalysen	160
10.4.3	Analysen	164
10.4.4	Die Frage der Generalisierbarkeit über die Fachcluster	167
10.5	Diskussion	168
10.5.1	Validierung des <i>EOS</i> -Modells	168
10.5.2	Vergleich der Varianzaufklärung zwischen den Fachclustern	169
10.5.3	Fazit	169
10.5.4	Limitationen	169
10.5.5	Ausblick	170
11	Interventionsstudie	171
11.1	Hintergrund	171
11.2	Metaanalytische Befunde zur Wirksamkeit von Interventionen im Studium	172
11.3	Fragestellung	173
11.4	Design: Prä-Post-Messung der Studieneinstiegsselbstwirksamkeit	174
11.5	Methode	174
11.5.1	Stichprobe	175
11.5.2	Vorgehen	175
11.6	Ergebnisse	176
11.6.1	Auswertung	176
11.6.2	Prä-Post-Vergleich	176
11.7	Diskussion	177
11.7.1	Limitationen	179

11.7.2 Praktische Relevanz	180
11.7.3 Ausblick	180
12 Diskussion	182
12.1 Zusammenfassung der Modellvergleiche auf theoretischer Ebene	182
12.2 Zusammenfassung der empirischen Befunde	183
12.2.1 Längsschnittstudien	184
12.2.2 Sekundärdatenanalyse	185
12.2.3 Interventionsstudie	185
12.2.4 Limitationen	186
12.3 Weiterführende Analysen	187
12.4 Skalenentwicklung	188
12.5 Fazit	188
12.6 Forschungsdesiderate	189
12.7 Praktische Implikationen	190
12.8 Gedanken zu <i>Open Science</i> und guter wissenschaftlicher Praxis	192
Abbildungsverzeichnis	195
Tabellenverzeichnis	197
Literatur	200
13 Anhang	I
13.1 Skalenentwicklung	I
13.2 Zu Studie 1	VI
13.3 Zu Studie 2	XII
13.4 Zu Studie 3	XVIII
13.5 Zusammenschau	XXIV
13.6 Sekundärdatenanalyse	XXXV
13.7 Interventionsstudie	XXXVI

Danksagung

An dieser Stelle möchte ich all' denjenigen danken, die mich auf vielfältige Weise bei meinem Promotionsprojekt unterstützt haben.

Zunächst gilt mein ganz besonderer Dank meinen Doktorvätern, Prof. Dr. Martin Kersting und Prof. Dr. Alexander E.M.G. Minnaert, die mich stets unterstützt und diese Arbeit begleitet haben. Unsere vielzahligen sehr konstruktiven fachlichen Diskussionen haben mir immer wieder neue Perspektiven eröffnet und nicht zuletzt auch meine Motivation stets aufrechterhalten. Weiterhin möchte ich meinen Kolleg*innen aus dem Team der *Psychologischen Diagnostik* dafür danken, dass sie immer „ein offenes Ohr“ für mich haben und wir uns in so vielen fachlichen und überfachlichen Zusammenhängen immer wieder gegenseitig unterstützen. Besonderer Dank gilt auch meinem Kollegen Dr. Dirk Lubbe, mit dem ich zahlreiche (methodische) Diskussionen führen durfte, die stets sehr bereichernd für mich waren.

Neben dieser vielfältigen Unterstützung im akademischen Bereich, habe ich das Glück, auch im privaten Bereich einen tollen Rückhalt zu haben. Hier gilt mein Dank in erster Linie meinem Mann Andreas Petri, der mich vom ersten Tag meines Promotionsprojektes an unterstützt und motiviert hat. Weiter sind es besonders meine Eltern Samira und Thomas Bothe sowie meine Schwester Anna Bothe, von denen ich stets Unterstützung erfahre. Zuletzt möchte ich namentlich noch meiner lieben Freundin Anna Voelske danken, in der ich eine wunderbare Gesprächspartnerin habe, die mich jederzeit auch moralisch unterstützt.

All' diesen Personen und auch denjenigen, die hier nicht namentlich genannt sind, mich aber ein Stück weit auf dem Weg zur Promotion begleitet haben, gilt mein herzlicher Dank.

Neben der persönlichen Unterstützung, wurde ich auch vom *Deutschen Akademischen Austauschdienst (DAAD)* gefördert. So erhielt ich ein Doktorandenstipendium für einen Forschungsaufenthalt an der *Rijksuniversiteit Groningen* (Kooperationspartner: Prof. Dr. Alexander Minnaert). Auch für diese Förderung möchte ich mich bedanken.

Zusammenfassung

Bildung eröffnet Perspektiven. Sie ermöglicht es jedem Einzelnen, seine Talente zu entfalten. Bildung ist der Schlüssel für gesellschaftliche Teilhabe, sozialen Aufstieg und ein erfülltes Leben. (BMBF, 2019)

Bildung prägt den Lebensweg. Insbesondere die Entscheidung, ein Hochschulstudium zu absolvieren, hat einen großen Einfluss auf den anschließenden beruflichen Werdegang inklusive des damit assoziierten sozioökonomischen Status. Insbesondere vor diesem Hintergrund erscheint der erfolgreiche Abschluss eines Studiums wünschenswert, ein Studienabbruch hingegen als ungünstig.

Aktuelle Untersuchungen zeigen jedoch auf, dass der Anteil an Studienabbrechern unter den Studierenden in den letzten Jahren innerhalb der OECD-Länder (*Organisation for Economic Cooperation and Development*) relativ konstant bei circa 30% liegt (OECD, 2016, 2018). Diese Quote gilt auch für Deutschland im Speziellen, gemittelt über alle Studienfächer und Hochschularten hinweg (Heublein et al., 2017; OECD, 2014). Dass somit durchschnittlich fast jeder dritte Studierende sein Hochschulstudium abbricht, wird unter anderem von der deutschen Bildungspolitik als inakzeptabel hoch bewertet (Heublein et al., 2017; Bloom, Hartley & Rosovsky, 2006).

Obwohl es bereits eine Vielzahl an institutionellen Angeboten (auch) zur Abbruchprävention gibt, basieren diese in der Regel aber nicht auf wissenschaftlichen Erkenntnissen zum Studienabbruch. Dies mag auch daran liegen, dass etablierte Modelle zur Erklärung von Studienabbruch entweder einen sehr engen theoretischen Fokus haben und oder noch nicht empirisch validiert sind; insbesondere noch nicht für die Population der Studierenden in Deutschland (D. Klein & Stocké, 2016). Vor dem Hintergrund, dass eine Reduktion der Abbruchquote angestrebt wird, besteht Handlungsbedarf, um die Diskrepanz zwischen wissenschaftlichen Untersuchungen im Bereich Studienabbruch und deren Erkenntnissen einerseits und den in der institutionellen Beratungspraxis vorzufindenden Angeboten andererseits zu minimieren.

Im Rahmen meines Promotionsprojektes habe ich daher zunächst ausgewählte bestehende Modelle des Studienabbruchs theoretisch und empirisch verglichen und anschließend auf Basis der so gewonnenen Erkenntnisse ein neues, integratives Modell des Studienabbruch entwickelt (das EOS-Modell), welches ich umfangreich validiert habe. Es kann zukünftig als Grundlage für institutionelles Handeln im Bereich Studienabbruchprophylaxe dienen. Im Folgenden möchte ich einen kurzen Überblick über das mehrschrittige Vorgehen geben:

1. Zunächst habe ich vier Modelle (Spady, 1971; Tinto, 1975; Neuville et al., 2007; Lent & Brown, 2013) als besonders relevant für das Verständnis von Studienabbruch identifiziert und auf theoretischer Ebene verglichen (Kapitel 2).

2. Anschließend habe ich sie empirisch hinsichtlich ihrer Nützlichkeit zur Beschreibung und Erklärung von Studienabbruch(-sintentionen) unter Studierenden in Deutschland verglichen. Hierfür habe ich drei unabhängige längsschnittliche Untersuchungen an Studienanfängern durchgeführt. Anhand der Daten der ersten Kohorte habe ich die vier etablierten Modelle hinsichtlich ihrer Modellfits verglichen (Kapitel 5).
3. Daraufhin entwickelte ich anhand dieser empirischen Erkenntnisse über die Stärken und Schwächen der vier Modelle und unter Berücksichtigung neuer metaanalytischer Erkenntnisse hinsichtlich einzelner valider Prädiktoren von Studienabbruch und Studienerfolg ein neues, integratives Modell (*EOS*-Modell, Kapitel 6).
4. Die Daten der zweiten und dritten Kohorte dienten zur Replikation des Modellvergleichs, den ich anhand der Daten der ersten Kohorte durchgeführt habe sowie zur Validierung des *EOS*-Modells (Kapitel 7 und 8).
5. Eine zusätzliche (partielle) Validierung (Kapitel 10) erfolgte anhand des Datensatzes aus der 21. Sozialerhebung (Middendorff et al., 2017).

Das *EOS*-Modell erwies sich mit einem guten Modellfit nicht nur als valide, sondern auch als überlegen gegenüber den getesteten etablierten Modellen. Zudem konnte ich die Messinvarianz des *EOS*-Modells über die drei Kohorten hinweg nachweisen. Dies erlaubt es, das *EOS*-Modell als zukünftiges Rahmenmodell für weitere Forschung zum Studienabbruch zu empfehlen.

Weiter kann es auch die Beratungspraxis unterstützen: Die zumeist limitierten Ressourcen für Beratungsangebote sollten auf die zentralen Aspekte des *EOS*-Modells konzentriert werden, um die Wirksamkeit der Interventionen zu erhöhen.

Als Beispiel für die Verwendung des *EOS*-Modells als Rahmenmodell für die Evaluation von Interventionen, führte ich eine erste Pilotstudie in Kooperation mit dem *Studiencoaching* der Justus-Liebig-Universität Gießen durch (Kapitel 11).

Neben der Modellentwicklung habe ich mich auch mit der Frage nach der Operationalisierung der im Fokus stehenden Konstrukte beschäftigt. Ganz konkret habe ich eine neue Skala entwickelt, die zur Erfassung der *Selbstwirksamkeit im Studieneinstieg* dient (Kapitel 3).

Abstract

Education has long been recognized as an essential element in life course. Especially the decision to participate in tertiary education has an important influence on the subsequent career path (OECD, 2016). Taking this into account, successful completion (and the avoidance of dropout) appears desirable. Unfortunately, the percentage of students who drop out remains at a relatively high level of about 30% in recent years not only in general across OECD (*Organisation for Economic Cooperation and Development*) countries (OECD, 2016, 2018), but also in specific in Germany (Heublein et al., 2017; OECD, 2014). This is considered to be unacceptably high (Heublein et al., 2017). Although there are institutional attempts to intervene, these are usually not based on scientific models of dropout. This might be due to the fact that existing models have either a very narrow theoretical focus or are not yet empirically validated, especially not for the German student population (D. Klein & Stocké, 2016).

The discrepancy between scientific insights gained in the field of higher education research on the one hand and counseling practices in higher education in Germany on the other hand calls for action. I therefore reviewed selected models of student dropout and compared them empirically in their fit to data collected from freshmen in Germany. Subsequently, I developed a new, integrative model (*EOS-Model*) of student dropout which has been subject to a comprehensive set of validation studies. Below, I would like to outline the stepwise approach I used in my PhD project:

1. I examined and compared four established models (Spady, 1971; Tinto, 1975; Neuville et al., 2007; Lent & Brown, 2013) theoretically (chapter 2) and empirically to determine their utility for the German freshmen population: I conducted three longitudinal studies (each over nine months) to track three cohorts of freshmen. Using data from the first cohort, I compared model fit of the established models.
2. I then developed a new process model – the *EOS-Model* – with an integrative broad theoretical perspective, based on existing models and including recent meta-analytic findings (Robbins et al., 2004; Richardson, Abraham & Bond, 2012) on single predictors of academic achievement and dropout (chapter 6).
3. Independent data from the second and third cohort served to validate it (chapters 7 and 8).
4. Additionally, a (partial) validation study (chapter 10) was performed using cross-sectional DZHW data (Middendorff et al., 2017).

The new model not only revealed a good model fit and proved its utility but also outperformed the other tested models. Furthermore, I could show its measurement invariance across the three cohorts. This allows to conclude that it could serve as a framework for future research on dropout (intentions) in tertiary education. In addition, it could serve as a framework for higher education counseling as it points out the most important aspects which are highly predictive for dropout intentions. Hence, tailoring interventions to the identified key elements may help concentrating limited counseling resources to the most promising aspects in order to maximize the impact of the interventions. Last but not least, the new model can serve as a guideline for evaluation studies for (institutional) interventions. To give an example for this application of the model, I conducted a first pilot study in cooperation with the *Student Coaching* at the Justus-Liebig-Universität Gießen.

Besides the development of the new model, an additional focus within my PhD project was the operationalization of the constructs under consideration. I would like to emphasize in particular the development of a new scale for the assessment of self-efficacy specifically tailored to the context of study entrance in the German higher education system (chapter 3).

1 Einleitung

Im Zentrum dieser Arbeit steht das Phänomen *Studienabbruch*. Im Folgenden soll zunächst ein Überblick über die Bedeutsamkeit des Studienabbruchs gegeben werden, bevor im Kapitel 2 dann empirische Befunde aus dem zugehörigen Forschungskontext dargelegt werden.

1.1 Relevanz des Themas

Die Relevanz der genaueren Untersuchung des Studienabbruchs (und meiner Ansicht nach damit auch des Studieneinstiegs als gesamten Prozess) lässt sich gemäß Isleib (2015) auf drei Ebenen ableiten: Auf der Makroebene führt eine hohe Studienabbruchquote zum Verlust an Innovationskraft und befördert den Fachkräftemangel. Das Bundesministerium für Bildung und Forschung (BMBF) veröffentlichte am 24.02.2016 im Bundesanzeiger eine Richtlinie zur Förderung von Forschung über „Studienerfolg und Studienabbruch“ (BMBF, 2016a). Darin wird erläutert:

Eine möglichst hohe Studienerfolgsquote bei gleichbleibend hohem akademischem Standard ist ein zentrales strategisches Ziel aller Hochschulen. Diese entsprechen damit ihrer gesellschaftlichen Verantwortung und kommen ihrer Fürsorgepflicht gegenüber den Studierenden nach. Vielfach wird Studienabbruch auch als finanzielle Fehlinvestition interpretiert, die es grundsätzlich zu vermeiden gilt. Neue und zusätzliche Relevanz erlangt das Thema „Studienerfolg und Studienabbruch“ vor dem Hintergrund des – nicht nur demografisch bedingten – steigenden Bedarfs an akademisch ausgebildeten Fachkräften.

Zudem erzeugen hohe Abbruchquoten hohe sozio-ökonomische Kosten (Isleib, 2015).

Auf der Mesoebene bringt das Prinzip der „performance-based fundings“ (Förderhöhe für Wissenschaftseinrichtungen wird auch anhand der Anzahl an Absolventen bemessen) die Hochschulen unter Druck (Isleib, 2015; Thomas & Hovdhaugen, 2014; Vossensteyn et al., 2015). Es gibt zwar erste kritische Auseinandersetzungen mit diesem Förderprinzip: So diskutiert Krempkow (2015) dieses Vorgehen und zeigt alternative Lösungsvorschläge auf. Aktuell ist jedoch festzuhalten, dass Hochschulen auch aufgrund dieses Förderprinzips großes Interesse an einer Reduktion der Abbruchquote haben. Zudem werden die Statistiken über Studienabbrecherquoten auf institutioneller Ebene auch zu Evaluationszwecken und zur Planung der Studierendenzahlen herangezogen (Tieben, 2016; D. Klein & Stocké, 2016).

Auf der Mikroebene schließlich birgt ein Studienabbruch finanzielle, zeitliche und emotionale Kosten

für das Individuum (Sarcletti & Müller, 2011). Zusammenfassend lässt sich festhalten, dass eine Reduktion der Abbruchquote wünschenswert ist. Eine Möglichkeit zu intervenieren sind institutionelle Beratungs- und Unterstützungsangebote. Grundlage für deren Konstruktion und Evaluation sollte ein evidenzbasiertes Verständnis für den Prozess des Studienabbruchs sein.

Heublein et al. (2017) beschreiben den Studienabbruch als multikausales Phänomen. Die letztendliche Entscheidung, das Studium endgültig abzubrechen, ist als Ergebnis eines Prozesses und nicht als kurzfristige Entscheidung zu verstehen. Um gezielt intervenieren zu können, sollte der Prozess des Studieneinstiegs und die Bewältigung von studienbezogenen Herausforderungen weiter untersucht werden. Erst ein umfassendes Verständnis des Prozesses des Studieneinstiegs wird es ermöglichen, gezielte Interventionen auf die Bedürfnisse der Zielgruppe abzustimmen und auftretenden Problemen möglichst früh entgegenzuwirken; sie möglicherweise sogar zu verhindern.

1.1.1 Ausmaß des Studienabbruch in Deutschland und international

Gemäß des OECD-Berichts aus dem Jahr 2017 zum Thema *Bildung*, liegt die durchschnittliche Erfolgsquote für ein Studium in den 22 in die Analysen einbezogenen Ländern bei ca. 69%, sofern als Bezugszeitraum die Regelstudienzeit plus drei Jahre zugrunde gelegt wird. Erfolg wird dabei definiert als Erreichen eines grundständigen Studienabschlusses (OECD, 2017). Diese Erfolgsquote findet sich mit nur minimalen Schwankungen (2019: 67%) auch in den neueren Ausgaben des OECD-Bildungsreports (OECD, 2018, 2019). Dabei ist zu beachten, dass die meisten der berücksichtigten Länder sogenannte *true cohort*-Daten bereitstellten, also Längsschnittdaten einer Kohorte an Studienanfängern. Diese Studienanfänger konnten mittels Panelstudien über die Studienzeit hinweg verfolgt werden, sodass tatsächlich deren Studienverlauf aufgezeichnet wurde. Einige Länder, darunter auch Deutschland, jedoch, lieferten für diese Analysen sogenannte *cross cohort*-Daten. Dies sind keine längsschnittlichen Daten. Mit *cross cohort*-Daten werden Kohortenvergleiche vollzogen. Das bedeutet, die Anzahl an Studienanfängern zu einem Zeitpunkt wird mit der Anzahl an Absolventen nach Ablauf der Regelstudienzeit in Beziehung gesetzt. Dass somit die gleichen Personen zu Studienbeginn und Studienabschluss erfasst werden, kann hierbei nicht sichergestellt werden. Diese Art der Daten ermöglicht durchaus auch eine Schätzung der Erfolgsquote, die *true cohort*-Daten sind jedoch methodisch gesehen verlässlicher und daher der bevorzugte *gold standard*. Zudem führen Schätzungen anhand der *cross cohort*-Daten in der Regel zu höheren Erfolgsquoten als Schätzungen anhand von *true cohort*-Daten (OECD, 2019).

Laut dem Statistischen Bundesamt lag die Erfolgsquote in Deutschland im Mittel bei den Anfängerjahrgängen 2002 bis 2006 zwischen 75% und 79% (Statistisches Bundesamt, 2016). Etwas weniger optimistischere Zahlen liefert das *Deutsche Zentrum für Hochschul- und Wissenschaftsforschung* (kurz: DZHW), das 2013 aus dem *Hochschul-Informationssystem* (kurz: HIS) entstand. Dort werden auf Basis von Stichprobendaten aus der Grundgesamtheit aller deutschen Hochschulen seit vielen Jahren regelmäßig Studien zum Thema Studienabbruch durchgeführt. Den Publikationen

zu diesen Studien ist zu entnehmen, dass sich die Abbruchquote unter Studierenden in Deutschland mit Einführung der Bologna-Reform 1998 verändert hat (Alesi, Bürger, Kehm & Teichler, 2005). Die Studienabbruchquote über alle Hochschulen und Fächer hinweg in Deutschland schwankt laut Heublein (2014) seit der Einführung der Bachelor- und Masterstudiengänge im Bereich zwischen 25% und 30%. Abbildung 1.1 zeigt die Abbruchquote unter deutschen Studierenden im Mittel über alle Fächer und Hochschulen hinweg für die Absolventenjahrgänge 2006, 2008, 2010, 2012 und 2014, worin die Studierenden der Anfängerjahrgänge 2002 bis 2006 enthalten sein müssten. Die Diskrepanz

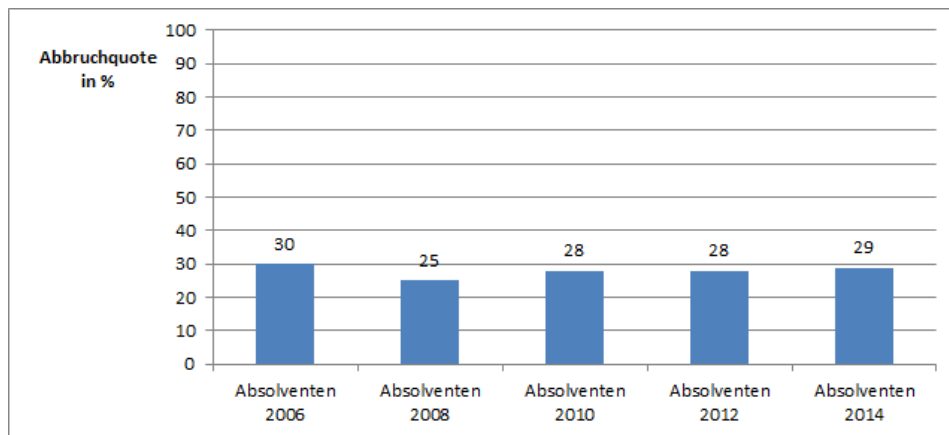


Abbildung 1.1: Abbruchquoten unter den Absolventenjahrgängen 2006 bis 2014 nach Heublein (2014) und Heublein et al. (2017)

zwischen den Angaben des Statistisches Bundesamt (2016) einerseits und des DZHW andererseits beruht auf unterschiedlichen Berechnungsverfahren, wie Heublein, Hutzsch, Schreiber, Sommer und Besuch (2010) und das Statistisches Bundesamt (2016) ausführlich erläutern. An dieser Stelle sei bereits angemerkt, dass es keine einheitliche Vorgehensweise für die Berechnung der Abbruchquoten gibt (Neugebauer, Heublein & Daniel, 2019) und darüber nach wie vor eine methodische Debatte geführt wird, welche beispielsweise von Heublein und Wolter (2011) ausführlich dargelegt wird. So ist die vom *Statistischen Bundesamt* verwendete Berechnungssystematik eine andere (die geringere Abbruchquoten erkennen lässt) als die vom DZHW verwendete. Beispielsweise Klöpping et al. (2017) verwendeten wieder eine etwas andere Berechnungssystematik. Hinzu kommt, dass neuere Ansätze wiederum Kritik an der vom DZHW verwendeten Herangehensweise üben und zu deutlich geringeren Schätzungen der Abbruchquoten gelangen (Tieben, 2019).

Ein Beispiel für einen Vergleich der Abbruchquote in Deutschland mit der anderer OECD-Länder ist Folgendes: Für den Absolventenjahrgang 2014 wird in Deutschland insgesamt von einer Quote von 29% an Studienabbrechern ausgegangen (Heublein et al., 2017). Das heißt gemäß der Berechnungssystematik des *DZHW* (Heublein, Richter, Schmelzer & Sommer, 2014), dass die zur Gruppe der Absolventen des Jahres 2014 korrespondierende Gruppe der Studienanfänger um 29% reduziert ist.

Da bei dieser Berechnung – sofern dies mit den amtlichen Daten möglich ist – Studienfachwechsler nicht als Abbrecher zählen (Heublein & Wolter, 2011), hat innerhalb dieser Referenzgruppe somit fast jeder dritte Studienanfänger die Hochschule ohne Abschluss verlassen. Damit lag Deutschland knapp unter dem OECD-Durchschnitt von ca. 30% (OECD, 2017).

Je nach Hochschulart und Art des angestrebten Abschlusses unterscheiden sich die Abbruchquoten allerdings deutlich, wie in Tabelle 1.1 aufgeführt. Weiterhin ergeben sich Unterschiede in den Abbruchquoten zwischen den Studienfächern. So konnten Fellenberg und Hannover (2006) zeigen, dass in den sogenannten *MINT*-Fächern (Akronym für Mathematik, Ingenieurwissenschaften, Naturwissenschaften und Technik), die Abbruchquoten höher sind als in den restlichen Fächergruppen. Besonders hohe Abbruchquoten finden sich insbesondere unter den Studierenden der Ingenieurwissenschaften (Heublein et al., 2017). Heublein et al. (2017) verzeichnen in ihrer Untersuchung eine

Tabelle 1.1: Abbruchquote nach Heublein, Richter, Schmelzer und Sommer (2014)

Hochschulart	Abbruchquote (Absolventen 2012)
Bachelor insgesamt	28%
Bachelor Universitäten	33%
Bachelor Fachhochschulen	23%
Staatsexamen	13%

gewisse Stabilisierung dieser bundesweit gemittelten Abbruchquote auf 29%. Der Wert liege dennoch immer noch über der laut einer Befragung von Fachbereichen vertretbaren Quote.

Ein weiterer wichtiger Aspekt, der sich in den Statistiken zum Studienabbruch erkennen lässt: Die mit Abstand größte Gruppe an Studienabbrechern ist die der frühen Abbrecher. Anders ausgedrückt: Die Studienabbruchquote ist in den ersten beiden Semester, dem sogenannten Studieneinstieg, maximal (Heublein et al., 2017; Stifterverband für die Deutsche Wissenschaft, 2015).

1.1.2 Methodische Herausforderungen bei der Quantifizierung von Studienabbruch

Die Quantifizierung von Studienabbruch ist grundsätzlich relativ schwierig. Dass es auch in Deutschland mit den aus der Administration der Hochschulen verfügbaren Daten schwierig bis unmöglich ist, Fehl kategorisierungen (im Sinne von *Studienabbrecher* versus *Studienfachwechsler*) auszuschließen, ist vielfach beschrieben worden. Beispielsweise kann quasi nicht sicher davon ausgegangen werden, dass eine Person, die ihr Erststudium (zunächst) abgebrochen hat und als „Abbrecher“ kategorisiert wurde, nicht zu einem späten Zeitpunkt ihr Studium wieder aufnimmt. Wie Tieben (2016) beschreibt, kann möglicherweise schon bei einem bloßen Fachwechsel innerhalb der gleichen Institution nicht immer sichergestellt werden, dass die betreffende Person nicht im zunächst gewählten

Fach als Abbrecher registriert wird. Es gibt allerdings auch optimistischere Einschätzungen bezüglich der zukünftigen Nutzung und Nutzbarkeit administrativer Daten: Hahm und Stock (2018) resümieren in ihrem „Werkstattbericht“ von der *Humboldt-Universität zu Berlin*, dass sie davon ausgehen, dass aufgrund der Novellierung des Hochschulstatistikgesetzes zukünftig mehr Daten zur Verfügung stehen werden und dass neue Analysemethoden (Stichwort *big data analysis*) dazu führen werden, dass administrative Daten häufiger zu Analysezwecken herangezogen werden.

Wenn es statt oder zusätzlich zu einem Fachwechsel auch zu einem Wechsel der Institution kommt, kann in der Regel nicht mehr nachverfolgt werden, ob eine Person, die ihr Studium (zunächst) abgebrochen oder unterbrochen hat, zu einem späteren Zeitpunkt noch einen Abschluss erwirbt. Zwar kann dies in Befragungen retrospektiv wie beispielsweise in den Abbrecher- bzw. Absolventenbefragungen des DZHW oder der *Humboldt-Universität zu Berlin* (Krempkow, 2017) aufgedeckt werden, ist aber im Allgemeinen sehr aufwändig und unter ausschließlicher Verwendung administrativer Daten nicht möglich.

Forschung zum Studienabbruch wird daher nicht selten auch unter Verwendung des „Proxys“ *Studienabbruchintentionen* durchgeführt (Neugebauer et al., 2019). Hierbei gilt es zu bedenken, dass man zwar einerseits davon ausgehen kann, dass einem Studienabbruch in der Regel Studienabbruchintentionen vorausgehen, dass man andererseits aber nicht damit rechnen kann, dass alle Personen, die zu einem gewissen Zeitpunkt Studienabbruchintentionen berichten, auch tatsächlich ihr Studium später abbrechen werden. Daher ist davon auszugehen, dass der Anteil derjenigen, die Studienabbruchintentionen berichten, höher ist, als der Anteil tatsächlicher Studienabbrecher an allen Studierenden.

1.1.3 Gründe für einen Studienabbruch

Blüthmann (2012) untersuchten $N = 439$ Studierende der freien Universität Berlin im Rahmen eines Querschnittsdesigns, um Studienabbruchgründe zu analysieren. Sie identifizierten zwei Cluster von Studienabbruchgründen: (i) Exmatrikulation aufgrund von Studienbedingungen und (ii) Exmatrikulation aufgrund von Studienanforderungen. Sie konnten zeigen, dass sich Studierende, die der einen versus der anderen Gruppe zuzuordnen waren, anhand verschiedener Attribute unterscheiden ließen: Während solche Studierende, die aufgrund der Studienbedingungen das Studium abbrachen, tendenziell hohes Fachinteresse und vergleichsweise bessere Noten in der Hochschulzugangsberechtigung aufwiesen, wiesen Studierende aus der Gruppe derjenigen, die sich aufgrund der Studienanforderungen exmatrikulierten, besonders häufig eine höhere Zahl an Wartesemestern auf, waren vergleichsweise leistungsschwächer im Studium, gaben an, mit Erkrankungen oder Krisen belastet zu sein und berichteten häufiger extrinsische Studienwahlmotive.

Einer Umfrage unter $N = 6029$ Exmatrikulierten im Zeitraum 2014–2015 zur Folge (Heublein et al., 2017) sind die vier am häufigsten genannten Gründe für den Studienabbruch:

1. Leistungsprobleme
2. mangelnde Studienmotivation
3. Wunsch nach praktischen Tätigkeiten
4. finanzielle Schwierigkeiten

Im Vergleich zur vorangegangenen Untersuchung der Arbeitsgruppe am DZHW im Jahr 2008 hat sich hier eine Verschiebung ergeben: Der Wunsch nach einer praktischen Tätigkeit anstelle des Studiums ist nunmehr unter den „Top 3“ der meistgenannten Abbruchgründe zu finden. Dahingegen werden finanzielle Gründe immer noch oft genannt, sind aber nicht mehr unten den „Top 3“.

In einer Pressemitteilung hat das BMBF über die Förderung neuer, zusätzlicher Informations- und Interventionsprogramme für Studierende, die über einen Studienabbruch nachdenken, berichtet. Erklärtes Ziel ist es, die Studienabbruchquoten zu senken (BMBF, 2016b).

1.1.4 Was geschieht aktuell auf dem Feld der Studienabbruch-Forschung?

Wie einleitend beschrieben, ist das Thema *Studienerfolg und Studienabbruch* von hoher (hochschul-)politischer Relevanz. Daher verwundert es kaum, dass das BMBF auch die Forschung in diesem Bereich fördert. Konkret wurde im Februar 2016 die Förderlinie *Studienerfolg und Studienabbruch* ins Leben gerufen. Ziel der finanziellen Förderung von Forschungsprojekten in diesem Feld ist einerseits der Gewinn eines umfassenden Verständnisses des Phänomens Studienabbruchs, sowie eine Quantifizierung der daraus resultierenden Kosten. Andererseits sollen bestehende Lösungsansätze zur Sicherung des Studienerfolges auf ihre Wirksamkeit hin untersucht werden (BMBF, 2018).

Auf der Internetseite zu vom BMBF geförderten Hochschul- und Wissenschaftsforschung¹ findet sich eine erschöpfende Liste aller 19 im Zuge der Förderlinie *Studienerfolg und Studienabbruch* geförderten Projekte.

Mit der folgenden Liste, die nur diejenigen Projekte umfasst, welche inhaltlich in direktem Bezug zu meinem Promotionsprojekt stehen, möchte ich einen kurzen Überblick geben. Dies soll – in Ergänzung zur bereits erfolgten und noch folgenden Darstellung publizierter bisheriger wissenschaftlicher Befunde – zu einem umfassenden Bild des Forschungsfeldes zum Thema *Studienerfolg und Studienabbruch* in Deutschland beitragen. Damit soll auch der Tatsache Rechnung getragen werden, dass wissenschaftliche Publikationen meist nicht tagesaktuelle Ergebnisse widerspiegeln, sondern der Publikationsprozess in der Regel einen Zeitverzug mit sich bringt. In diesem Sinne kann ein Überblick über aktuelle große Forschungsprojekte im Forschungsfeld die Aktualität des in dieser Arbeit gebotenen Überblicks steigern.

¹<https://www.wihoforschung.de/de/studienerfolg-und-studienabbruch-620.php>

Die wenigen, ersten Ergebnisse einiger dieser Projekte, die erst zum Zeitpunkt der Fertigstellung dieser Arbeit publiziert wurden, werden im Folgenden zwecks Übersichtlichkeit direkt im Anschluss an die entsprechende Projektbeschreibung berichtet.

- **CASSIS: Chemie, Sozialwissenschaften und Ingenieurwissenschaften: Studienerfolg und Studienabbruch**
Universität Duisburg/Essen
Im Zentrum des Projektes steht die Tatsache, dass häufig Diskrepanzen zwischen der Erwartung der Hochschulen in Bezug auf die Vorkenntnisse von Studienanfänger und deren tatsächlichen Vorkenntnissen vorliegt. Im Zuge des Projektes werden Messinstrumente entwickelt, anhand derer einerseits Lehrveranstaltungen evaluiert werden können und andererseits der Vorwissensstand der Studienanfänger abgeprüft werden kann. Anhand der Messergebnisse sollen die Studienanfänger beispielsweise gemäß ihrem Kenntnisstand in unterschiedliche Kurse eingeteilt oder an Vorbereitungskurse verwiesen werden.
- **DMPS: Determinanten und Modelle zur Prognose von Studienabbrüchen**
Universität Duisburg/Essen
In diesem Projekt wird anhand umfangreicher (administrativer) Datensätze unter Verwendung von Data Mining Methoden nach einem Prognosemodell für den Studienabbruch gesucht. Diese Methode, die aus der Bildungsökonomik stammt, soll das Modellieren der Multikausalität des Studienabbruchs ermöglichen. Die Modelle sollen anschließend den Hochschulen zur Identifikation sogenannter Risikogruppen dienen, um diesen dann institutionelle Interventionen zukommen zulassen.
- **FragSte: Früherkennung abbruchgefährdeter Studierender und experimentelle Studien zur Wirksamkeit von Maßnahmen**
Bergische Universität Wuppertal
Unter ausschließlicher Nutzung administrativer Daten soll ein Frühwarnsystem in Bezug auf potentielle Studienabbrecher erstellt werden, das in der Lage sein soll, automatische Feedbacks zu geben. Zudem sollen bestehende Interventionen auf ihre Wirksamkeit hin überprüft werden und das auch im Vergleich verschiedener Gruppen von Studierenden. Dieses nur auf administrativen Daten beruhende System soll auf andere Hochschulen übertragbar sein. In einer ersten Publikation zu diesem Projekt vergleichen Berens und Schneider (2019) drei verschiedene Frühwarnsysteme hinsichtlich der mit ihnen zu erzielenden Prognosegüte. Dabei verwendeten sie Daten von einer privaten und einer staatlichen Hochschule. Alle drei Systeme eigneten sich gut, um abbruchgefährdete Studierende frühzeitig zu identifizieren. Die Prognose konnte sogar noch verbessert werden, wenn neben den Daten aktueller Studierender auch auf Daten ehemaliger Studierender zugegriffen werden konnte, um den Algorithmus zu trainieren.
- **LAST: Lebenslaufansatz und Studienabbruch. Eine multikontextuelle Analyse zu den Ursachen und Folgen des Studienabbruchs**

Carl-von-Ossietzky-Universität Oldenburg

Der Fokus dieses Projekts liegt auf der längsschnittlichen Analyse von Studienverlaufsdaten aus der Perspektive von Lebenslauftheorien. Es sollen Erkenntnisse hinsichtlich Prädiktoren von Studienabbruch(-neigung) und Ansatzpunkte zur Reduktion des Studienabbruchs identifiziert werden. Neben verschiedenen anderen theoretischen Perspektiven, die integriert werden sollen, liegt ein Augenmerk auf dem privaten sozialen Umfeld der Studierenden und der Rolle dieses Umfeldes beim Gelingen oder Scheitern institutioneller Interventionen. Erste Ergebnisse aus diesem Projekt belegen, dass die Einbindung der Eltern und die Kommunikation mit ihnen über das Studium, signifikante korrelative Zusammenhänge mit Studienerfolg und Abbruchintention aufweisen (Feldhaus & Baalman, 2019).

- PIRanHa: Projekt zur Identifikation und Reduktion angehender Hochschulabbrecher und -abbrecherinnen

IZA - Forschungsinstitut zur Zukunft der Arbeit GmbH

Unter Verwendung der Daten aus dem *National Educational Panel Survey (NEPS)* werden Ursachen für einen Studienabbruch untersucht, die wiederum in die Erstellung eines Frühwarnsystems einfließen sollen. Die Methodik des *machine learnings* wird herangezogen, um ein Prognosemodell zu erstellen, welches die Wahrscheinlichkeit eines Studienabbruchs auf Individualebene kalkulieren können soll.

- ProkRASt: Prokrastination als Risikofaktor für den Abbruch des Studiums, Teilprojekt: Die Rolle der Studienmotivation

Universität Bielefeld; Universität Augsburg; Ruhr-Universität Bochum

In diesem Verbundprojekt wird die Intention zum Studienabbruch in Zusammenhang mit dem Risikofaktor Prokrastination untersucht. Unter Prozessperspektive werden der Zusammenhang, die Bedingungen und die Entwicklung von Prokrastination und Studienabbruch im Studienverlauf erforscht.

- ReSt@MINT: Der Einfluss der Resilienz auf den Studienerfolg von MINT-Studierenden

Ludwig-Maximilians-Universität München; Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung

Mit Fokus auf die für hohe Abbruchquoten bekannten sog. *MINT*-Fächer (Mathematik, Ingenieurwissenschaften, Naturwissenschaften und Technik) werden drei Konstrukte und deren Zusammenhang untersucht: Studienbelastung, Resilienz und Studienerfolg. Hofmann, Datzer, Razinkas und Högl (2019) berichten erste Erkenntnisse aus einer qualitativen Vorstudie, die der großen Längsschnittuntersuchung vorangestellt war. In Letzterer werden MINT-Studierenden mit Studierenden in den Wirtschaftswissenschaften verglichen. Ein Fokus liegt dabei auf der Frage, welchen Effekt die Resilienz auf das im Verlauf des Studieneinstiegs entwickelte Commitment hat und ob sie Einfluss auf verschiedene Studienerfolgskriterien (wie z.B. Studienleistung, Abbruchintention) hat.

- SAM: Studienabbruch in der Mathematik
Eberhard-Karls-Universität Tübingen; Universität Stuttgart
Dieses Projekt untersucht den Prozess des Studienabbruchs mit Augenmerk auf dessen Multikausalität speziell für das Studienfach Mathematik. Dabei stehen die Variablen Kompetenz und Motivation im Vordergrund. Ein Experimental-Kontrollgruppen-Design der angedachten Intervention soll helfen, Studienabbrüche zu reduzieren.
- STHAGE: Studienabbruch, Habitus und Gesellschaftsbild
Universität Duisburg-Essen; Hochschule Hannover
Dieses Projekt untersucht Studienabbrüche vor dem Hintergrund des jeweiligen sozialen Herkunftsmilieus der Studierenden. Unter Verwendung von strukturierten Interviews wird dabei an jedem Projektstandort ein Hochschultyp unter die Lupe genommen (Universität und Fachhochschule).

Begleitforschung zum *Qualitätspakt Lehre*

Seit dem Jahr 2011 unterstützt das BMBF unter dem Namen *Qualitätspakt Lehre (QPL)* ² „die Verbesserung der Studienbedingungen und der Lehrqualität an deutschen Hochschulen“. Dies geschieht vornehmlich durch die finanzielle Förderung umfangreicher Forschungs- und Lehrprojekte, die an verschiedenen Standorten in ganz Deutschland angesiedelt sind. Im Folgenden möchte ich die jüngeren im Rahmen des *Qualitätspakt Lehre* vom BMBF finanzierten (Begleitforschungs-)Projekte kurz vorstellen.

Ein großes vom BMBF gefördertes Projekt ist das *StuFHe (Studierfähigkeit - institutionelle Förderung und studienrelevante Heterogenität)*-Projekt in Hamburg. Das als Begleitforschung zum *Qualitätspakt Lehre* aufgesetzte Projekt fokussiert auf die Untersuchung der Heterogenität Studierender und der Bedingungen für gelingendes Studieren. Hierbei wird Heterogenität auf individueller, sozialer und organisationaler Ebene betrachtet und mit dem Konzept der Studierfähigkeit in Zusammenhang gesetzt. Unter Studierfähigkeit werden Kompetenzen verstanden, die sich unter anderem in vorangegangenen psychologischen Studien als Prädiktoren für Studienerfolg erwiesen, wie beispielsweise Lernstrategien und epistemische Neugier.

Einerseits wurden basierend auf einer qualitativen Vorstudie Instrumente zur Erfassung von Studienanforderungen und Studienzielen entwickelt. Andererseits wurde eine umfangreiche Längsschnittuntersuchung (Erhebungen im 1., 3. und 5. Semester) mit Fragebogen unter Studierenden an vier Partneruniversitäten durchgeführt. Im zugehörigen Projektbericht wird ein Modell zu „studienrelevanten

² <https://www.qualitaetspakt-lehre.de/de/qualitat-von-hochschullehre-und-studienbedingungen-verbessern-1764.php>

Heterogenitätsmerkmalen“ in Bezug auf verschiedene Kriterien (wie beispielsweise Studienzufriedenheit oder das Erreichen formaler Studienziele) vorgestellt. Die quantitativen Untersuchungen hierzu sind bisher vornehmlich in einer Publikation ausführlich dargestellt (Barnat, Abelha Faria & Bosse, 2017): Sie basieren auf Varianzanalysen und multiplen Regressionsanalysen der Daten vom ersten Erhebungszeitpunkt, mit denen der Zusammenhang verschiedener Heterogenitätsmerkmale mit zwei ausgewählten Kompetenzaspekten (Lernstrategien und epistemische Neugier als Facetten von Studierfähigkeit) untersucht wurde. Diese ersten Befunde basieren folglich auf Daten querschnittlicher Natur mit einem beachtlichen Stichprobenumfang von $N = 1942$ Studierenden. Für die Analysen der Daten aus allen drei Messzeitpunkten (dann längsschnittlich) sei auch die Modellierung mit Strukturgleichungsmodellen angedacht (Barnat, Abelha Faria & Bosse, 2017).

Das Modell, das bisher in dem das Projekt zusammenfassenden Bericht dargestellt ist, scheint jedoch weniger ein Prozessmodell im engeren Sinne zu sein, da es keine detaillierten Zusammenhänge (Interaktionen, Mediationen, Moderationen) wichtiger Prädiktoren aufzeigt, sondern diese eher im Sinne eines Schemas auflistet (Bosse, Mergner, Wallis, Jänsch & Kunow, 2019). Darüber hinaus wurden im *StuFHe*-Projekt auf institutioneller Ebene vorhandene Angebote gesichtet. Diese Schwerpunktlegung im Projekt zeigt auch auf, dass es durchaus eine Vielzahl an institutionellen Unterstützungsangeboten für Studieneinsteiger gibt, diese aber wenig systematisch sind, und - wie Barnat, Bosse und Mergner (2017) berichten - auch selten theoriebasiert entwickelt oder ausgerichtet wurden.

In einer weiteren Projektpublikation führen Barnat, Bosse und Mergner (2017) auf, dass man an den Partnerhochschulen Angebote für Studierende in der Studieneingangsphase gesichtet und kategorisiert habe. Die vorgeschlagenen neun Kategorien spannen einen Rahmen auf, in den sich die sehr unterschiedlichen Angebote (verschiedene inhaltliche Schwerpunkte, verschiedene didaktische Formate, unterschiedliche Zielgruppen innerhalb der Studierenden) einordnen lassen.

Forschungsprojekte auf EU-Ebene

Zwei weitere große (EU-geförderte) Forschungsprojekte zum Thema Studienabbruch sind das *Prev Drop - Detecting and Preventing Drop out from Higher Education or Supporting Students to Switch Successfully to VET* ³ und das *SUnStAR - Supporting UNiversity STudents At Risk of dropping out* ⁴. Im Rahmen des letztgenannten Projektes entstand auch die Dissertation von Nolden (2019). Er hat sich ebenfalls mit der Bildung eines multi-perspektivischen Modells des Studienabbruchs beschäftigt. Hierbei hat er sich vornehmlich aus soziologischer Perspektive dem Problem der Modellbildung genähert und sein Modell ebenfalls empirisch getestet. Obgleich er sowohl umfangreiche Stichproben rekrutiert hat, als auch eine Validierung seiner Befunde mittel Modelltests anhand der Daten einer zweiten unabhängigen Stichprobe vollzogen hat, bleibt ein wichtiger beispielsweise auch von Richardson et al. (2012) als Desiderat genannter Punkt (vgl. Kapitel 2) unberücksichtigt: Nolden

³<http://www.prevdrop.eu/de>

⁴<https://sunstar-eu.org/>

(2019) arbeitete hierbei nicht mit Längsschnittdaten, sondern mit Querschnittdaten. Zudem liefert er zwar wichtige Antworten hinsichtlich Fragen der Invarianz seiner Teilmodelle über verschiedene Gruppen (verschiedene Hochschularten, Geschlechter, MINT- versus Nicht-MINT-Fächer), jedoch testete er stets nur Teilmodelle, zu keinem Zeitpunkt jedoch ein Gesamtmodell, das alle aufgestellten Teilmodelle gleichzeitig berücksichtigt.

1.1.5 Desiderate der Forschung zum Studienerfolg und -abbruch

Robbins et al. (2004) halten im Zuge ihrer Metaanalyse fest, dass die Forschung zu Studienerfolg und Studienabbruch vornehmlich aus zwei Traditionen erfolgt: Zum einen sei die Entwicklung von Theorien zu Übergang und Anpassung der Studienanfänger an die Hochschule und das Testen dieser Theorien mittels Longitudinalstudien die Stärke der Bildungsforschung. Robbins et al. (2004) bemängeln allerdings, dass die herangezogenen Konstrukte oft atheoretisch seien und die Messinstrumente manchmal nur ein Item umfassen würden. Zum anderen läge die Stärke der psychologischen Forschung in diesem Feld eben in der Verwendung valider Konstrukte und Messinstrumente. Jedoch fehle es der psychologischen Forschung an Systematik; kurzum an der Einbettung in einen theoretischen Rahmen, der den Prozess des Studieneinstiegs beziehungsweise akademischen Werdeganges beschreibt. Aus diesem Vergleich der beiden Forschungstraditionen leiten Robbins et al. (2004) ab, dass eine Symbiose dieser Forschungstraditionen und eine Integration der diversen bisherige Befunde sinnvoll ist. Sie resümieren weiter, dass das Feld der Bildungsforschung durch Berücksichtigung psychologischer Konstrukte profitieren würde.

Was Robbins et al. (2004) somit metaanalytisch auf internationaler Ebene aufzeigen, kann ergänzt werden um noch offene Fragen und Herausforderungen im Kontext der deutschen Hochschulbildung: Neugebauer et al. (2019) werfen in ihrem Überblicksartikel vier Fragen zum Thema *Studienabbruch* auf: (i) Wie viele Studierende brechen ihr Studium ab?, (ii) Was sind die Ursachen von Studienabbruch?, (iii) Welche Folgen hat Studienabbruch und schlussendlich? (iv) Wie kann Studienabbruch verhindert werden? Neben der Zusammenschau der wichtigsten empirischen Befunde zur Beantwortung dieser Fragen, werfen sie Desiderate für zukünftige Forschung auf. So verweisen sie darauf, dass neben institutionellen auch individuelle Faktoren in Vorhersagemodelle des Studienabbruchs Berücksichtigung finden sollten. Die für die Ableitung von Handlungsanleitungen für die Bildungspolitik zwingend notwendigen empirischen Überprüfungen sollten längsschnittlich erfolgen, um kausale Schlüsse überhaupt zu ermöglichen. Weiterhin betonen sie die Notwendigkeit der systematischen Evaluation der Wirksamkeit bereits vorhandener institutioneller Interventionen. Zwar gebe es vereinzelte Evaluationsstudien, diese sollten jedoch wenn möglich metaanalytisch zusammengefasst werden. Zudem sei die „differenzierte Wirksamkeit“ (S. 17, ebd.) der Interventionen zu berücksichtigen. Sie schließen mit der Feststellung, dass „die Konzeption und Implementation erfolgreicher Maßnahmen

[...] folglich untrennbar mit der Diagnose individueller Abbruchursachen verknüpft“ (S. 17) sein sollte.

Dies zeigt den Stellenwert der Betrachtung des Phänomens *Studienabbruch* auf individueller Ebene besonders deutlich. So spiegelt sich in diesem Desiderat klar das Prinzip der Förderdiagnostik wider: Auf Basis zielgerichteter Diagnostik sollten letztendlich besonders Abbruchgefährdete identifiziert und auf zu ihren Bedürfnissen passende Förderangebote hingewiesen werden. Interventionen sollten entsprechend hinsichtlich ihrer Wirksamkeit (Ziel: Reduktion der Abbruchquote) evaluiert werden.

2 Studienerfolg und -abbruch: Definition und empirische Befunde zu Prädiktoren

2.1 Empirische Befunde zum Studienerfolg

In diesem Kapitel stelle ich dar, welche empirischen Befunde zur Prädiktion von *Studienerfolg versus -abbruch* bereits vorliegen. Dazu werde ich neben Befunden zur Validität einzelner Prädiktoren auch Modelle und Theorien zur Erklärung von *Studienerfolg versus -abbruch* darstellen. Darauf aufbauend gebe ich einen Überblick über die empirischen Befunde, die bereits zu den verschiedenen Theorieansätzen vorliegen. Hierbei werden Forschungsergebnisse aus verschiedenen Disziplinen, insbesondere den Sozial- und Bildungswissenschaften einerseits und der Psychologie andererseits, berücksichtigt. Wie Hüther und Krücken (2018) in ihrem Buch zum tertiären Bildungswesen in Deutschland veranschaulichen, lassen sich die Perspektiven dieser beiden Disziplinen zwar einerseits klar differenzieren, jedoch enthüllt eine genauere Betrachtung Schnittmengen zwischen diesen. Daher gehe ich in diesem Kapitel neben der Gegenüberstellung der Ansätze abschließend darauf ein, warum deren Integration ein vielversprechender Ansatz für ein umfassendes Verständnis von *Studienerfolg versus -abbruch* zu sein scheint.

2.1.1 Definition von Studienerfolg

Bevor ich im Folgenden auf die Erkenntnisse bisheriger Forschung eingehe, möchte ich das Konstrukt *Studienerfolg* genauer betrachten. Schlägt man den Begriff *Erfolg* im Duden nach, so findet man als Erklärung: „positives Ergebnis einer Bemühung; Eintreten einer beabsichtigten, erstrebten Wirkung“ (Bibliographisches Institut, 2019). Umgangssprachlich bezieht sich der Begriff *Erfolg* meist auf Leistung. Im akademischen Kontext liegt es nahe, gute Noten oder gar den Studienabschluss insgesamt als Operationalisierung von Erfolg zu verstehen. Dies kann als Erfolg im „engeren Sinne“ bezeichnet werden. Gemäß Trapmann, Hell, Weigand und Schuler (2007) gibt es im akademischen Kontext nur wenige Studien, in denen Erfolg nicht nur als Studiennote operationalisiert wurde.

Robbins et al. (2004) berichten in ihrer Metaanalyse, dass der Notendurchschnitt im Studium das am weitesten verbreitete Maß für akademische Leistung ist. Auch wird ein besonders zügiger Studienverlauf, also eine kurze Studiendauer mit anschließendem Studienabschluss, zuweilen als Erfolg betrachtet, insbesondere wenn von institutioneller Seite Erfolgsmaße definiert werden. Die Studiendauer bis zum Studienabschluss wird in empirischen Untersuchungen zum Studienerfolg jedoch eher

selten als Kriterium verwendet (Robbins et al., 2004). Eine der wenigen Ausnahmen im deutschen Hochschulsystem bildet eine Untersuchung von Krempkow (in Druck).

Heublein et al. (2015) führten im Jahr 2015 eine Befragung an Fakultäten und Fachbereichen (231 Fakultäten, darunter 121 Universitäten) durch, um zu ermitteln, welche Definition von Erfolg von institutioneller Seite her verwendet wird. Eine Faktorenanalyse der Befragungsdaten ergab zwei grundlegende Ansichten. Gemäß der einen bedeutet Studienerfolg sehr gute fachliche Fähigkeiten, „gute Vorbereitungen auf die Anforderungen des Arbeitsmarktes“ sowie „hohe Studienzufriedenheit“ (S.15). Gemäß der anderen Ansicht kann Studienerfolg verstanden werden als „gute Prüfungsnoten“ sowie „niedrigen Studienabbruch“ (S.15). Diese beiden Ansichten unterscheiden sich vornehmlich darin, dass gemäß der ersten Erfolg auf Individuumsebene eher in Form subjektiver Erfolgsmaße betrachtet wird, wohingegen bei der zweiten Erfolg an den Noten und Absolventenstatistiken, also eher „objektiven“ Kriterien festgemacht wird.

Bezieht man die Perspektive der Studierenden mit ein, so sind die erstgenannten subjektiven Erfolgsmaße auch von Bedeutung (Bosse et al., 2019). Zu nennen wären da unter anderem Zufriedenheit während des Studiums, ein während des Studiums aufrechterhaltenes oder idealerweise gesteigertes Interesse am Fach oder ein subjektiv akzeptables Maß an Stress im Studium (wobei anzumerken ist, dass wiederum eine deutliche Unterforderung aus Studierendensicht ebenfalls unerwünscht ist, Krempkow, 2005). Diese Erfolgsmaße sind einerseits zwar nur von den einzelnen Studierenden selbst wahrnehmbar und somit zunächst nur für das Individuum relevant, andererseits wiederum beeinflussen diese subjektiven Einschätzungen aber im weiteren Verlauf das Handeln der Studierenden, sodass sie auf lange Sicht auch Einfluss auf die „objektiven“ Erfolgsmaße haben können, die beispielsweise von institutioneller Seite her betrachtet werden.

Ein besonders „einfaches“ Maß für Erfolg im akademischen Kontext wäre das Erlangen eines Abschlusses an sich: Ungeachtet der vorangehenden Studiendauer und ohne genaue Betrachtung der Noten, könnte man Erfolg mit dem Erlangen eines Abschlusses gleichsetzen und folgerichtig Misserfolg definieren als Studienabbruch. In dieser Dichotomisierung wird deutlich, wie die beiden im Fokus dieser Arbeit stehenden Kriterien (*Studienerfolg versus -abbruch*) miteinander in Verbindung stehen. Aus dieser Perspektive erscheint es logisch konsequent, dass Prädiktoren des einen Kriteriums auch Prädiktoren des anderen Kriteriums sind. Das sei an dieser Stelle bereits angemerkt. Es ist die Erklärung dafür, dass in diesem Kapitel zwar zunächst empirische Befunde zur Prädiktion von Studienerfolg und anschließend separat zur Prädiktion von Studienabbruch aufgeführt werden, dass allerdings, um Redundanzen zu vermeiden, im zweiten Teil des Kapitels nur noch auf Befunde, die sich explizit nur auf Studienabbruch beziehen, eingegangen wird.

Durch die Berücksichtigung von Erfolg und Abbruch(-intentionen) zugleich soll der Prozess der Studieneinstiegs besonders umfassend beleuchtet werden. Außerdem erscheint es sinnvoll, sowohl breitere als auch engere Definitionen von Erfolg mit einzubeziehen, da von Prädiktorensseite her ebenfalls enger und weiter gefasste Variablen einbezogen werden sollen. Dies entspricht dem Grundprinzip der

Prädiktor-Kriterium-Symmetrie (Brunswik, 1955; Wittmann, 1990).

Robbins et al. (2004) verweisen in der Diskussion ihrer metaanalytischen Befunde darauf, dass die Symmetrie von Prädiktor und Kriterium Einfluss auf die Höhe des Zusammenhangs hat: Prädiktoren, die vergleichsweise eng gefasste Definitionen haben (z.B. akademische Selbstwirksamkeit), weisen höchstwahrscheinlich stärkere Zusammenhänge mit eng gefassten Kriterien wie der Durchschnittsnote im Studium auf, als Prädiktoren, die auf breitere Konstruktdefinitionen fußen (z.B. allgemeines Selbstkonzept). Im Rahmen dieser Arbeit möchte ich folgende relativ breite Definition von Erfolg im Studium verwenden.

Definition Studienerfolg:

Im Rahmen dieser Arbeit wird Studienerfolg mehrdimensional betrachtet. Dazu zählt neben der Studienleistung in Noten, auch die berichtete Studienzufriedenheit und die Abwesenheit von empfundenem Stress.

Ausgehend von dieser Definition von Studienerfolg möchte ich nun einen Überblick zu empirischen Befunden zu einzelnen Prädiktoren von Studienerfolg geben.

2.1.2 Noten und kognitive Fähigkeiten als Prädiktoren von Studienerfolg

Insbesondere von der Abiturnote ist bekannt, dass sie hohe prädiktive Validität bei der Vorhersage von Studienerfolg ausweist, operationalisiert beispielsweise als Studienabschlussnote (Trapmann et al., 2007). Dies erscheint insbesondere auch aus methodischer Sicht wenig überraschend, da in diesem Fall eine hohe, generell wünschenswerte, Prädiktor-Kriterium-Symmetrie vorliegt.

Auch Brandstätter, Grillich und Farthofer (2006) konnten den bekanntermaßen hohen Zusammenhang zwischen Schul- bzw. Testleistungen und durchschnittlicher Studienleistung replizieren. Dabei fanden sie, dass der Effekt durch die Studienleistung im ersten Studiensemester vermittelt wird. Robbins et al. (2004) bezeichnen in ihrer Metaanalyse ebenfalls die Highschool-Noten und Ergebnisse standardisierter Leistungstest als klassische Prädiktoren. Von diesen wisse man, dass sie im Mittel etwa 25% der Varianz der Studiennoten im ersten Studienjahr aufklären.

Trapmann et al. (2007) stellten für den deutschsprachigen Raum fest, dass die Schulabschlussnoten noch immer das am häufigsten verwendete Kriterium für die Vergabe von Studienplätzen waren. Dies liegt zum einen an der leichten Verfügbarkeit dieser Prädiktorvariablen, zum anderen aber vermutlich auch daran, dass die Validität der Schulnoten für die Vorhersage von Studienerfolg bereits vielfach gezeigt werden konnte (Trapmann et al., 2007). Obgleich sich die Regularien zur Vergabe von Medizinstudienplätzen nun aufgrund eines Urteils des Bundesverfassungsgerichts (BVerfG, 2017-12-19) ändern werden, gelten die Befunde zur Validität der Schulabschlussnoten natürlich weiterhin.

Kognitive Fähigkeiten gelten allgemein als guter Prädiktor für Erfolg in der Ausbildung und dem Beruf, wie Salgado et al. (2003) für den europäischen Raum und Kramer (2009) für Deutschland bereits metaanalytisch zeigen konnten. Inwiefern innerhalb der Population der Studierenden Maße der kognitiven Fähigkeiten (beispielsweise operationalisiert als Ergebnisse eines Intelligenztests) hohe prädiktive Validität in Bezug auf das Kriterium Studienerfolg aufweisen, ist jedoch weniger eindeutig. Einzelne Untersuchungen unter Studierenden liefern unterschiedliche Ergebnisse. So untersuchte Schmidt-Atzert (2005) $N = 79$ Psychologiestudierende (die als sehr selektive Gruppe einzuordnen sind) und konnte keinen signifikanten Zusammenhang zwischen dem in einem Intelligenztest erzielten Ergebnis und der Note im Vordiplom feststellen. Dem entgegengesetzt sind die Ergebnisse einer Studie von Farsides und Woodfield (2003), die $N = 432$ Studierende diverser Studienfelder an einer britischen Universität untersuchten. Hierbei erwies ein Maß für verbale Intelligenz einen bedeutsamen Zusammenhang mit der Abschlussnote des dreijährigen Studienprogramms auf ($r = .20, p < .01$). In diese Richtung deuten auch die Befunde von Minnaert (1996): In seiner Longitudinalstudie nahmen $N = 517$ Studienanfänger aus den Studienfeldern Medizin, Bildungswissenschaften und Psychologie teil. Für den eingesetzte Intelligenztest (neben Maßen für Metakognition, siehe unten) zeigte sich in einer ANOVA mit dem Kriterium *Prozent erreichter Leistungspunkte* ein signifikanter Haupteffekt ($F(2, 530) = 13.96, p < .01$).

In der Metaanalyse von Richardson et al. (2012) wird angeführt, dass die Population der Studierenden beispielsweise aufgrund der Zulassungsbeschränkung für viele Studiengänge eine vorselektierte Gruppe ist, innerhalb derer die Varianz der kognitiven Fähigkeiten eingeschränkt ist. Sie plädieren daher dafür, den Fokus bei der Prädiktion von Studienerfolg eher auf nicht-kognitive Faktoren zu legen.

Hierbei ist allerdings zu bedenken, dass es weithin den Konsens gibt, dass Schulnoten und Studiennoten als „Proxy“ für kognitive Fähigkeiten betrachtet werden können. Dies beruht darauf, dass deren enger Zusammenhang vielfach empirisch (Gottfredson & Crouse, 1986; Hunter, 1986), sogar metaanalytisch gezeigt werden konnte (Roth et al., 2015). Für Studiendesigns insbesondere im Hochschulkontext ist das Erfragen der Abiturnote oder der aktuellen Studiennote demzufolge oft ein guter Kompromiss, um zwar einen Schätzer für die kognitiven Fähigkeiten einer Person zu erhalten, aber die Datenerhebung nicht durch den Einsatz eines zusätzlichen kognitiven Tests massiv zu verlängern. Vor diesem Hintergrund ist es gegebenenfalls auch eher verständlich, warum in den später in diesem Kapitel vorgestellten Modellen und Theorien zum *Studienerfolg versus -abbruch* selten die kognitiven Fähigkeiten einer Person als Prädiktor aufgeführt werden. Durch die Aufnahme der Schulleistung (meist operationalisiert durch die Abiturnote) in die Modelle sind die kognitiven Fähigkeiten dann indirekt (im Sinne einer Mediation) berücksichtigt.

2.1.3 Demografische und sozio-ökonomische Variablen als Prädiktoren von Studienerfolg

In diesem Abschnitt möchte ich exemplarisch Befunde zum Zusammenhang demografischer sowie sozio-ökonomischer Variablen mit Studienerfolg aufführen.

Zum Merkmal *Geschlecht* lässt sich im deutschen Hochschulkontext festhalten, dass die Gruppe der Frauen im Vergleich zur Gruppe der Männer die höhere Studienerfolgsquote (Anzahl Absolventen im Vergleich zur Anzahl der Studienanfänger im korrespondierenden Anfängerjahrgang) aufweist. International belegen die umfangreichen Statistiken der OECD ebenfalls, dass in fast allen der 22 in diese Analyse einbezogenen Staaten die Gruppe der Frauen im Durchschnitt eine höhere Studienerfolgsquote (im Sinne eines Studienabschlusses) aufweist als die Gruppe der Männer (OECD, 2016, 2018).

Insbesondere aus soziologischer Sicht ist die Frage nach dem Einfluss des Herkunftsmilieus auf den Bildungserfolg traditionell ein Kernthema der Bildungsforschung. Als besonders einflussreich sind hier die Arbeiten von Bourdieu und Passeron (1971) und Bourdieu (1983) und darauf aufbauende Theorien zu nennen. International und nationalen liegen diverse empirische Untersuchungen zum Einfluss der sozialen Herkunft auf den individuellen Bildungsweg vor. Hierbei ist anzumerken, dass es im Ländervergleich durchaus Unterschiede gibt.

Marks (2007) verwendeten Längsschnittdaten aus einem australischem Studenten-Panel. Sie untersuchten den Zusammenhang verschiedener demografischer und sozio-ökonomische Variablen mit Daten zum (Nicht-)Absolvieren einzelner Kurse und zum Studienabschluss insgesamt. Hinsichtlich der demografischen Prädiktoren von Studienerfolg (im Sinne des Abschließen eines Studiums) zeigten sich unter anderem Unterschiede in der Erfolgsquote zwischen Studierenden verschiedener Ethnien. Sie berichten aber darüber hinaus, dass sie keinen Effekt des sozio-ökonomischen Status' der Herkunftsfamilie der Studierenden fanden, was sie so erklären, dass der sozio-ökonomische Status vermutlich den Zugang zur tertiären Bildung erschwere; wer jedoch bereits studiere, den beeinflusse seine Herkunft nicht mehr bedeutsam.

Ein anderes Bild zeigen Befunde für den primären und sekundären Bildungsweg in Deutschland: Insgesamt ist dem deutschen Bildungssystem insbesondere auf Ebene der primären und sekundären Bildung deutliche Chancenungleichheit für Personen unterschiedlicher sozialer Herkunft zu attestieren (Artelt, Demmrich & Baumert, 2001). Dies erfuhr insbesondere Anfang der 2000er Jahre im Zuge der PISA-Studien große mediale Aufmerksamkeit. Kristen (1999) gibt einen guten Überblick über theoretische Ansätze zur Erklärung unterschiedlicher Bildungschancen in unterschiedlichen sozialen Milieus und fasst zudem bisherige empirische Befunde, insbesondere für den Übergang zur Sekundarstufe, zusammen.

In Bezug auf den tertiären Bildungsweg sind die Befunde weniger eindeutig. So ist Büchler (2012) zur Folge für Personen aus Familien mit vergleichsweise niedrigem sozio-ökonomischem Status nicht nur die Aufnahme eines Studiums weniger wahrscheinlich. Auch die Erfolgsquote sei für diese Gruppe von Studierenden im Allgemeinen niedriger. Heublein und Wolter (2011) jedoch resümieren in ihrem

Überblicksartikel, dass es sowohl Befunde gibt, die eine erhöhte Studienabbruchswahrscheinlichkeit bei niedrigerem Herkunftstatus attestieren, als auch Befunde, die das Gegenteil belegen. Das bedeutet, dass die Befundlage diesbezüglich nicht eindeutig ist. Neuere Untersuchungen im deutschen Hochschulkontext legen jedoch nahe, dass sich mittlerweile doch auch ein Einfluss des Herkunftsmilieus auf Erfolgsquoten im Hochschulkontext abzeichnet (Stifterverband für die Deutsche Wissenschaft & McKinsey & Company, 2017). Dabei geht ein höherer sozialer Herkunftstatus tendenziell mit einer höheren Erfolgsquote einher.

Erste Befunde zu beruflich-qualifizierten Studierenden liegen auch bereits vor. Damit sind Studierende gemeint, die keine allgemeine Hochschulzugangsberechtigung haben, sondern durch eine Berufsausbildung und entsprechende Berufserfahrung oder eine Aufstiegsfortbildung eine als äquivalent anerkannte Zugangsberechtigung erlangt haben. Die Gruppe der so genannten nicht-traditionell Studierenden weist eine geringere Erfolgsquote auf als die der traditionell Studierenden (Dahm, Becker & Bornkessel, 2018). Darüber hinaus konnten Dahm et al. (2018) mit ihrer Analyse eines Teildatensatzes ($N = 621$) der 21. Sozialerhebung (Middendorff et al., 2016) zeigen, dass bei nicht-traditionell Studierenden neben mangelhafter sozialer Integration auch Problemen bei der Studienfinanzierung und der Vereinbarkeit von Studium und Beruf prädiktiv für Studienabbruchintentionen sind.

Brandstätter und Farthofer (2003) untersuchten den Zusammenhang von Erwerbstätigkeit und Studienleistung bei $N = 361$ Studierenden der Universität Linz. In dieser Longitudinalstudie wurde über drei Jahre hinweg jedes Semester die durchschnittliche Stundenzahl, die mit einer Erwerbstätigkeit verbracht wurde (Prädiktor), erfasst. Das Kriterium *Studienerfolg* operationalisierten sie sowohl als Anzahl der bis inklusive des fünften Semesters abgelegten Prüfungen und dem dabei erzielten Notendurchschnitt, als auch als Studienzufriedenheit, Beständigkeit der Studienwahl und Studienabbruch(-gedanken). Studienwahl bezeichnet dabei die retrospektive Einschätzung, ob man eine gute Studienwahl getroffen hat oder gegebenenfalls über einen Studienabbruch nachdenkt. Der Studienabbruch wiederum als hartes Kriterium wurde operationalisiert als vorzeitiges Ende beziehungsweise Ausbleiben weiterer elektronisch erfasster Prüfungsnachweise, wobei bei diesem Kriterium Studierende, die einen Wechsel an eine andere Universität berichteten, ausgeschlossen werden mussten (Brandstätter et al., 2006). Alle diese Kriterien wiesen signifikante Zusammenhänge mit Erwerbstätigkeit neben dem Studium auf. Dabei war die Anzahl abgelegter Prüfungen niedriger, die Durchschnittsnote schlechter, die Studienzufriedenheit und die Einschätzung der Richtigkeit der Studienwahl geringer, je höher die durchschnittliche Stundenzahl für Erwerbstätigkeit war. Dabei betonen Brandstätter et al. (2006) explizit, dass eine Erwerbstätigkeit, die 19 Stunden oder mehr pro Woche in Anspruch nimmt, den Studienerfolg gefährdet.

2.1.4 Interessen als Prädiktoren für Studienerfolg

Das Konzept des Interesses am Studienfach wurde von Schiefele, Kapp, Wild und Winteler (1993) genau untersucht. Die Autoren entwickelten den *Fragebogen zum Studieninteresse (FSI)*, der dieses auf drei Dimensionen abbilden soll: (i) gefühlsbezogene Valenzen, (ii) persönliche wertbezogene Valenzen und (iii) intrinsischer Charakter. Neben umfangreichen Untersuchungen zur Dimensionalität und Reliabilität liegen auch Erkenntnisse zur prädiktiven Validität der Skala hinsichtlich verschiedener Operationalisierungen von Studienerfolg vor. So konnten beispielsweise der Zeit- und Anstrengungsaufwand, das Flow-Erleben und die Studienleistung bei $N = 298$ Studierenden bedeutsam vorhergesagt werden (Schiefele et al., 1993). In einer weiteren Untersuchung Schiefele und Jacob-Ebbinghaus (2006) mit $N = 101$ wurde Studienzufriedenheit mit der Skala von Hiemisch, Westermann und Michael (2005) erfasst. In dieser Skala werden drei Dimensionen von Studienzufriedenheit abgebildet: (i) Zufriedenheit mit den Inhalten, (ii) mit den Studienbedingungen und (iii) mit der Bewältigung von Belastung. Die beiden Dimensionen Zufriedenheit mit den Inhalten und Zufriedenheit mit der Bewältigung von Belastung konnten mithilfe des Fachinteresses vorhergesagt werden. Die Autoren erklären ersteres so, dass Studierende mit hohem fachlichem Interesse besonders dann hohe Zufriedenheit mit den Studieninhalten berichten, wenn auch berufsbezogene Inhalte behandelt werden. Dass die Zufriedenheit mit der Bewältigung von Belastung ebenfalls durch Fachinteresse (neben dem in ihrer Untersuchung wichtigen Prädiktor: der wahrgenommenen Lehrqualität) vorhergesagt werden kann, stufen Schiefele und Jacob-Ebbinghaus (2006) als konform mit der vorherigen Literatur ein. Blüthmann (2012) konnte anhand einer fächerheterogenen Stichprobe ($N = 2591$) ebenfalls zeigen, dass das Fachinteresse die Studienzufriedenheit (allgemein erfasst) vorhersagen kann.

Im Studien- und Berufswahlkontext wird oft das in der Psychologie gängige *RIASEC*-Modell von Holland (1997) zur Kategorisierung von Interessen herangezogen. Das individuelle Interessensprofil an sich scheint sich jedoch weniger gut als Prädiktor für Studienerfolg zu eignen (De Fruyt & Mervielde, 1996): Rounds und Su (2014) fassen in ihrem Überblicksartikel zusammen, dass man anhand des Interesses die akademische Leistung besonders dann gut vorhersagen kann, wenn der Fokus weniger auf dem Interesse am Studienfach als solches liegt, sondern vielmehr darauf, dass ein möglichst guter „Fit“ (im Sinne geringer Differenzen) zwischen dem individuellen Interessensprofil und dem interessenbezogenen Anforderungsprofil des jeweiligen Studiengangs vorliegt.

So verwundert es kaum, dass die Erfassung von Interessen bei der Studienwahlberatung von Studieninteressierten ein weit verbreiteter Ansatz ist (Hasenberg, 2012). Beispielsweise entwickelte Stoll (2013) einen Interessenstest zur Berufs- und Studienwahl, der im Rahmen von Online Self-Assessments an deutschen Hochschulen eingesetzt wird.

2.1.5 Studienrelevante Persönlichkeitseigenschaften als Prädiktoren von Studienerfolg

Ähnlich wie bei der Vorhersage von Berufserfolg, gelten auch für die Prädiktion von Studienerfolg einige Persönlichkeitseigenschaften als vielversprechend. Um die Ausführungen an dieser Stelle kurz zu

halten, möchte ich die Befundlage auf die von Costa und McCrae (1992) postulierten *Big 5* der Persönlichkeit beschränken. Farsides und Woodfield (2003) analysierten in der oben bereits genannten Untersuchung anhand einer Stichprobe von $N = 432$ Studierenden neben dem Zusammenhang von Intelligenz und Studienleistung (operationalisiert als Abschlussnote) auch den prädiktiven Wert dieser fünf Persönlichkeitsdimensionen. In ihrer Untersuchung bestand nur zwischen der *Offenheit* und der Abschlussnote im Studium ein signifikanter positiver Zusammenhang. Andere Untersuchungen allerdings belegten wiederholt den prädiktiven Wert der Persönlichkeitsdimension *Gewissenhaftigkeit*, wie beispielsweise die Studie von De Fruyt und Mervielde (1996) mit $N = 714$ belgischen Studierenden. Auf Zusammenschau einer Vielzahl solcher Untersuchungen, liegen mittlerweile auch metaanalytische Befunde vor, die unter den *Big 5* die *Gewissenhaftigkeit* als den besten Prädiktor für Studienerfolg ausweisen (M. Schneider & Preckel, 2017; Richardson et al., 2012). Das bedeutet, dass ein hohes Maß an Gewissenhaftigkeit, was sich beispielsweise im fristgerechten und vollständigen Einreichen von Studienleistungen, der Teilnahme an Veranstaltungen ohne Anwesenheitspflicht oder im Lesen zusätzlicher Literatur manifestieren kann, dem Erfolg im Studium zuträglich ist.

2.1.6 Selbstwirksamkeit als Prädiktor von Studienerfolg

Das Konzept der Selbstwirksamkeit wurde 1977 von Bandura eingeführt. Selbstwirksamkeit ist die Überzeugung, dass man in einer bestimmten Situation ein bestimmtes Verhalten erfolgreich ausführen kann (Bandura, 1977). Im Folgenden werde ich die Begriffe *Selbstwirksamkeit* und *Selbstwirksamkeitsüberzeugungen* und *Selbstwirksamkeitserwartungen* äquivalent verwenden, wie es auch in der umfangreichen Literatur zu diesem Thema der Fall ist.

Wichtig ist die Unterscheidung zwischen allgemeiner und spezifischer Selbstwirksamkeit. Mit allgemeiner Selbstwirksamkeit ist die grundlegende Überzeugung, Herausforderungen erfolgreich bewältigen zu können, gemeint. Spezifische Selbstwirksamkeit hingegen meint die Überzeugung, in einem bestimmten Kontext oder beim Ausführen spezifischer Handlungen erfolgreich sein zu können.

Das Konzept der Selbstwirksamkeit ist gemäß Bandura (1977) ein dynamisches Konzept. Das bedeutet, im Gegensatz zu Persönlichkeitseigenschaften im engeren Sinne, beispielsweise den *Big 5* nach Costa und McCrae (1992), die als relativ zeitstabil angesehen werden (traits), unterliegen Selbstwirksamkeitseinschätzungen ständigem Wandel. Dies ergibt sich auch daraus, dass sie sich aus vier Quellen speisen (Bandura, 1977), die fortlaufend neue Informationen liefern. Diese sind *i* eigene Erfolgserfahrungen, *ii* Beobachtungslernen, *iii* verbale Überzeugungen und *iv* physiologisches Arousal in der Leistungssituation selbst. Auf diese vier Quellen möchte ich im Folgenden eingehen.

1. Eigene Erfolgserfahrungen: Macht eine Person die Erfahrung, eine Situation oder Aufgabe erfolgreich bewältigen zu können, so erhöht dies ihre Selbstwirksamkeit in Bezug auf die Bewältigung vergleichbarer Situationen oder Aufgaben. Umgekehrt führen Misserfolgserfahrungen zur Reduktion der situations- oder aufgabenspezifischen Selbstwirksamkeit. Außerdem kann ein Transfer der positiven Erfahrungen stattfinden: Erfolgserfahrungen in einer Situation oder

bezüglich einer Aufgabe können sich positiv auf Selbstwirksamkeitserwartungen in anderen, ähnlichen Situationen auswirken (Bandura, 1977).

2. Beobachtungslernen: In Anlehnung an die sozial-kognitive Lerntheorie geht Bandura (2001) davon aus, dass sowohl tatsächlich im eigenen Umfeld vorhandene als auch symbolische Rollenvorbilder, die eine Aufgabe erfolgreich meistern, die aufgabenspezifische Selbstwirksamkeit des Beobachters stärken können. Dies erklärt er so, dass der Beobachter sich selbst überzeugt, dass - wenn andere eine Situation erfolgreich bewältigen können- er es zu einem gewissen Grad ebenfalls können müsse. Soziale Vergleiche werden dabei herangezogen, um einzuschätzen, inwiefern die beobachtete Person mit der eigenen Person vergleichbar ist. Das bedeutet: Je höher der Beobachter die Ähnlichkeit zwischen sich und der beobachteten Person bewertet, desto relevanter ist die Beobachtung für die eigene Selbstwirksamkeitserwartung. In Studien mit Phobikern zeigte sich zudem, dass das Beobachten einer Person, die eine für den Phobiker beängstigende Situation erfolgreich bewältigt, dann besonders günstigen Einfluss auf die Selbstwirksamkeitserwartung des Phobikers hat, wenn deutlich zu beobachten war, dass das erfolgreiche Bewältigen mit Durchhaltevermögen zu tun hat (Kazdin, 1973). Umgekehrt kann das Beobachten des Scheiterns eines Rollenvorbildes die aufgabenspezifische Selbstwirksamkeitserwartung reduzieren.
3. verbale Überzeugungen: Von einer anderen Person überzeugt zu werden, eine Situation oder Aufgabe bewältigen zu können, beeinflusst die Selbstwirksamkeitserwartung ebenfalls bis zu einem gewissen Grad. Überzeugungen sind laut Bandura (1977) ein weit verbreitetes Mittel, um die Selbstwirksamkeitserwartungen zu erhöhen. Allerdings seien eigene Erfahrungen wesentlich einflussreicher, sodass beispielsweise nach einer selbstwirksamkeitssteigernden verbalen Instruktion und anschließendem Misserfolgserfahrungen, der Einfluss der Misserfolgserfahrung im Vordergrund stehe (Bandura, 1977).
4. Emotionales und physiologisches Arousal: Das Beobachten der eigenen Gefühlslage kann ebenfalls Informationen liefern, die in die Bildung der Selbstwirksamkeitserwartung einfließen. Das physiologische Arousal, das Empfindungen von Angst und Stress zugrunde liegt, wird hierbei als Indikator verwendet. Da ein hohes Arousal zumeist die Leistung beeinträchtigt, sei die Abwesenheit von Angst und Stress oftmals aus Sicht des Individuums ein Grund für Erfolgszuversicht (Bandura, 1977). Im klinischen Anwendungsbereich spielt diese Annahme ebenfalls eine Rolle. Auf sie fußen beispielsweise auch solche Therapieansätze bei Phobikern, die unter anderem zum Ziel haben, die Aufmerksamkeit vom eigenen Arousal wegzulenken.

Da aus diesen vier Quellen fortlaufend neue Informationen eintreffen, unterliegen insbesondere situations- bzw. kontextspezifische Selbstwirksamkeitserwartungen dem ständigen Wandel. Sehr vereinfacht lässt sich festhalten, dass eigene und beobachtete Erfolgserlebnisse die eigene Selbstwirksamkeitserwartung in Bezug auf die Bewältigung einer spezifischen Situation stärken, wohingegen das Beobachten

und selbst Durchleben von Misserfolg die spezifische Selbstwirksamkeitserwartung reduziert. Vergleicht man die Effekte, die Einflüsse aus diesen vier Quellen haben können, so hat sich gezeigt, dass eigene (Miss-)Erfolgserfahrungen den stärksten Einfluss haben. Dies gilt insbesondere auch für Selbstwirksamkeitserwartungen Studierender im Hochschulkontext (van Dinther, Dochy & Segers, 2011). Ebenfalls einflussreich sind das Beobachtungslernen.

In zahlreichen Studien im klinischen Bereich zeigte sich immer wieder, dass verbale Überzeugung oft nur wenig Einfluss hat. Dabei ist auch oft nicht genau auszumachen, ob verbale Überzeugungen überhaupt die Selbstwirksamkeit erhöhen können oder bloß die Ergebniserwartungen (die Überzeugung, dass eine bestimmte Handlung zu einem bestimmten Ergebnis führe; eine Bewertung ohne Berücksichtigung persönlicher Erfolgswahrscheinlichkeit).

Zudem scheinen verbale Überzeugungen in Kombination mit konkreten Verhaltensanweisungen effektiver darin zu sein, Personen dazu zu bewegen, sich anzustrengen, als verbale Überzeugungen alleine. Dies kann besonders dann hilfreich sein, wenn es darum geht, korrektive Erfahrungen zu ermöglichen. Das bedeutet, wenn eine Person bereits Misserfolgserfahrungen gemacht hat, nun aber verbal überzeugt wird, eine kommende Situation erfolgreich bewältigen zu können und zusätzlich konkrete Handlungsanweisungen bekommt, dann kann dies gezielt helfen, eine Situation zu bewältigen und somit eine korrektive Erfahrung zu machen. Allerdings können Misserfolgserfahrungen, die eintreten, nach dem eine Person verbal überzeugt wurde, eine Situation bewältigen zu können, die Glaubwürdigkeit der Person, die diese Überzeugungen geleistet hat, reduzieren. Somit können Misserfolgserfahrungen den Effekt der Informationen aus dieser Quelle reduzieren.

In einer Längsschnittuntersuchung über das erste Studienjahr hinweg mit $N = 819$ Studierenden unterschiedlicher Fächer konnten Brahm, Jenert und Wagner (2014) zeigen, dass studienbezogene Angst und empfundene Herausforderungen anhand des Niveaus an studienbezogener Selbstwirksamkeit zum Zeitpunkt des Studienbeginns vorhergesagt werden können.

In diverse weiteren Studien konnte gezeigt werden, dass Selbstwirksamkeit Einfluss auf Zielsetzung, Durchhaltevermögen und Leistung hat (van Dinther et al., 2011; Bandura, 1977). Auch in der Längsschnittstudie von De Clercq, Galand, Dupont und Frenay (2013) mit $N = 317$ Studierenden einer belgischen Universität zeigte sich, dass Selbstwirksamkeit einer der stärksten Prädiktoren für Studienerfolg ist.

2.1.7 Emotionen im Leistungskontext als Prädiktoren von Studienerfolg

Emotionen im Leistungskontext sind vor allem im Bereich der sekundären, aber auch im Bereich der tertiären Bildung bereits untersucht worden. Besonders einflussreich sind hier die Arbeiten zur sogenannten *Kontroll-Wert-Theorie der Leistungsempfindungen* (Pekrun, 2000). Zusammenfassend lässt sich festhalten, dass positiv konnotierte Emotionen (wie beispielsweise Freude) im Allgemeinen einen

leistungsförderlichen und negativ konnotierte Emotionen (wie beispielsweise Angst) einen leistungsmindernden Effekt haben (Pekrun, 2006).

Respondek, Seufert, Stupnisky und Nett (2017) untersuchten $N = 883$ Studienanfänger (erstes Semester) und Studierende im dritten Studiensemester, die vornehmlich in den *MINT*-Fächern studierten. In dieser Studie wurden Fragebogendaten, die zu Beginn des Wintersemesters erhoben wurden, mit den administrativen Prüfungsdaten der Studierenden zum Ende des Wintersemesters in Beziehung gesetzt. Ausgehend von der Hypothese, Kontrollüberzeugungen beeinflussten die entstehenden Emotionen (im Fokus der Studie: Freude, Langeweile und Angst) im akademischen Kontext, welche wiederum in direktem Zusammenhang mit Leistung und Abbruchintentionen stünden, berechneten die Autoren bivariate Korrelationen und ein Strukturgleichungsmodell. Während keine der drei genannten Emotionen einen direkten, bedeutsamen Pfad zum Studienerfolg (operationalisiert als Studiennoten) aufwies, zeigte sich eine bedeutsame (negative) Assoziation zwischen studienbezogener *Angst* und Studienleistung.

Emotionen im Studienkontext scheinen daher ebenfalls ein wichtiger Prädiktor für die subjektiven und objektiven Maße von Studienerfolg zu sein. Zudem beeinflussen die Leistungsemotionen, wie Pekrun (2000) sie auch nennt, wiederum die Lernmotivation (Pekrun, 2006), welche wiederum die folgenden Leistungen beeinflussen kann.

Neben verschiedenen distalen Antezedenzen werden Emotionen in Leistungssituationen gemäß der *Control-Value Theory of Achievement Emotions* maßgeblich durch die beiden proximalen Faktoren (i) Wert (welcher Wert wird der Situation oder dem Ergebnis der Situation zugeschrieben) und (ii) Kontrolle (welche Kontrollüberzeugungen hinsichtlich dem Bewältigen der konkreten Situation liegen vor) determiniert. Die erlebten Emotionen in Leistungssituationen wiederum sind in dieser Theorie über Feedback-Schleifen mit den proximalen und distalen Antezedenzen verbunden und bestimmen so zukünftige Emotionen in solchen Situationen.

Dabei weist das Konstrukt *Kontrollüberzeugungen* große Überlappungen mit dem Konstrukt *Selbstwirksamkeit* auf, weshalb Pekrun (2006) selbst die Äquivalenz dieser beiden Konstrukte in diesem Zusammenhang aufgreift. Sofern sich Selbstwirksamkeitserwartungen auf konkrete Handlungen im Studium bezögen, seien sie als äquivalent zu den Kontrollüberzeugungen in Bezug auf diese Handlungen zu sehen.

2.1.8 Motivation als Prädiktor von Studienerfolg

Die Bedeutung von Zielsetzung als Richtungsweiser und Antreiber (menschlichen) Handelns wird von Locke (1968) ausführlich dargelegt. Unter anderem aufbauend auf Locke (1968) diskutieren Deci und Ryan (2000) die Bedeutung intrinsischer Motivation und Selbstbestimmung im Rahmen ihrer *Self-Determination Theory*: Laut ihrer Theorie spielt die Befriedigung (basaler) psychologischer Bedürfnisse eine entscheidende Rolle bei der Verfolgung persönlicher Ziele.

Hinsichtlich des Setzen von Zielen im Leistungsbereich ist eine viel rezipierte Differenzierung die von

Dweck (1986) vorgeschlagene Dichotomisierung in *Lernziele* (Fokus auf Kompetenzentwicklung) und *Leistungsziele* (Fokus auf Kompetenzdemonstration). Deci und Ryan (2000) heben diesbezüglich hervor, dass *Leistungsziele* mit einem starken Fokus auf der Demonstration des eigenen Könnens im sozialen Vergleich assoziiert sind und dass diese Art von Zielen eher mit extrinsischer Motivation und Vermeidungszielen (das Vermeiden einer schlechten Bewertung) assoziiert sind, wohingegen *Lernziele* eher mit intrinsischer Motivation beim Lernen assoziiert sind. Mit *Lernzielen* sei somit eher eine Annäherungsmotivation verbunden, denn Personen mit solchen Zielen würden herausfordernde Situationen eher als Chance zur persönlichen Weiterentwicklung, denn als Gefahr des Scheitern bewerten. Phillips und Gully (1997) führten eine Studie mit $N = 405$ US-amerikanischen Studierenden, die an Psychologie- oder Management-Kursen teilnahmen, durch. Dabei konnten sie unter anderem zeigen, dass das Level eines selbstgesetzten Ziels (auch unter Kontrolle der individuellen bisherigen Studienleistungen) prädiktiv war für den Prüfungserfolg (operationalisiert als Note in der den Kurs abschließenden Prüfung).

Ranellucci, Hall und Goetz (2015) untersuchten $N = 138$ US-amerikanische Studierende diverser Studienfächer in einem Längsschnittsdesign. Via Online-Fragebogen wurden die Studienteilnehmer einmal zu Studienbeginn und dann ein zweites Mal sechs Monate später befragt. Neben der Zielsetzung wurden Emotionen, selbstreguliertes Lernen und die akademische Leistung erfasst. Sie analysierten sowohl bivariate Korrelationen als auch sogenannte *Cross-Lagged-Modelle*. Hierbei wurden Zusammenhänge zwischen wiederholten Messungen eines Konstrukts ebenso berücksichtigt wie die Zusammenhänge verschiedener Konstrukte zum selben Messzeitpunkt. Ranellucci et al. (2015) heben besonders die signifikanten Zusammenhänge von Lernzielen und positiven Emotionen sowie Leistung hervor. Hierbei gilt wie schon von Pekrun (2006) betont, dass die Beziehungen dieser Konstrukte bidirektional sind und diese in komplexen Interaktionen stehen, welche beispielsweise als Feedback-Schleifen vorstellbar sind.

M. Schneider und Preckel (2017) fassen zusammen, dass insbesondere spezifische, klar definierte und von den Studierenden selbst gesetzte Ziele positiv mit Studienerfolg zusammenhängen. Sie heben hervor, dass sich erfolgreiche Studierende auch dadurch auszeichnen, sich selbst Notenziele zu setzen.

2.1.9 Soziale Integration und Unterstützung als Prädiktoren von Studienerfolg

Diverse Studien belegen, dass soziale Interaktion im Studium großen Einfluss auf den Studienerfolg hat: So führten DeBerard, Spielmans und Deana (2004) eine Längsschnittstudie mit $N = 204$ Soziologie- und Psychologiestudierenden an einer US-amerikanischen Privathochschule durch. Unter Berücksichtigung von 10 Prädiktoren, darunter Schulnoten und Gesundheitsverhalten sowie soziale Unterstützung, konnten sie 56% der Varianz in der Durchschnittsnote nach einem Studienjahr aufklären, wobei die (erhaltene) soziale Unterstützung einen signifikanten Beitrag zur Varianzaufklärung

leistete.

Bereits sogar metaanalytisch belegt ist, dass soziale Integration (Robbins et al., 2004; Richardson et al., 2012) und soziale Unterstützung (Robbins et al., 2004) positive Zusammenhänge mit Studienerfolg aufweisen: Je besser Studierende sozial involviert sind und je mehr soziale Unterstützung sie im Studium erleben, desto wahrscheinlicher ist der Studienerfolg bzw. desto höher die aktuelle Durchschnittsnote.

Eine weitere Differenzierung des Integrationsbegriffs schlagen Trautwein und Bosse (2017) vor: Sie untersuchten, welchen Herausforderungen sich Studieneinsteiger gegenüber sehen. In ihrer qualitativen Studien mit $N = 25$ Studierenden sammelten sie eine große Bandbreite an Herausforderungen. Innerhalb des gewonnenen Datensatzes identifizierten sie vier Dimensionen: personelle, organisationale, inhaltsbezogene und soziale Herausforderungen. Eine zusätzliche Analyse der Häufigkeit der Nennung von zu den jeweiligen Dimensionen gehörigen Herausforderungen ergab, dass die personellen und die organisationalen Herausforderungen am häufigsten berichtet wurden. Basierend auf diesen Erkenntnissen schlagen sie vor, den von Tinto (1987) (siehe unten, Abschnitt 2.4.2) erwähnten Integrationsbegriff zu differenzieren in die Bereiche inhaltsbezogen, personell und organisational.

2.1.10 Studentisches Engagement und Commitment als Prädiktoren von Studienerfolg

Das Konstrukt *Student Engagement* hat – ausgehend von der US-amerikanischen (Hoch-)Schulforschung zunächst in den Bildungswissenschaften – große Aufmerksamkeit erfahren. Unter *Student Engagement* ist die Mühe und die Zeit zu verstehen, die ein Studierender in studienbezogene Aktivitäten investiert (Kuh, 2009). Eine ausführliche Darstellung des Konstrukts inklusive der vier postulierten Facetten (*behavioral, emotional, kognitiv und sozial*), findet sich beispielsweise bei Reeve und Tseng (2011) und Bowden, Tickle und Naumann (2019). Der prädiktive Wert des studentischen Engagement für verschiedene Operationalisierungen von Studienerfolg konnte vielfach gezeigt werden (Kuh, 2009; Bowden et al., 2019).

Lawson und Lawson (2013) schlagen eine Revision des Konzepts des *Student Engagement* vor, indem es als dynamisches Zusammenspiel psychologischer und sozialer Aspekte verstanden wird. Hierbei betonen sie, dass auch Aspekte wie Motivation, vorangegangene Erfahrungen und Interessen das letztendlich gezeigte studentische Engagement prägen.

An der Stelle bietet sich ein Verweis auf ein verwandtes Konstrukt, welches eher der psychologischen Forschung zuzuordnen ist, an: das *University Citizenship Behavior* (Gehring, 2006). Hierbei handelt es sich um ein Konstrukt, welches an das Konzept des *Organizational Citizenship Behavior* (Organ, 1988) angelehnt ist. Übertragen auf Studierende als Mitglieder der Hochschule (im Vergleich zu Arbeitnehmern als Mitglieder einer Organisation), geht es um umfeldbezogene Leistungen, die Studierende im Hochschulkontext erbringen können. Dies kann das freiwillige Unterstützen von Kommilitonen betreffen, die Teilnahme an extracurriculären Veranstaltungen, hochschulpolitisches

Engagement oder auch die gewissenhafte Erfüllung studentischer Pflichten, obgleich dies weder kontrolliert noch explizit honoriert (im Sinne von gut benotet) wird, wie beispielsweise die Teilnahme an Veranstaltungen ohne Anwesenheitspflicht. Für das Konstrukt *University Citizenship Behavior (UCB)* konnte bereits eine bedeutsame, logisch positive Assoziation mit dem individuellen akademischen Potential (beurteilt durch Lehrende) (Schwager et al., 2014) sowie eine logisch negative Assoziation mit *counterproductive academic behavior* (im weitesten Sinne: der Hochschule schadendem Verhalten, operationalisiert beispielsweise als Täuschen in Prüfungen) nachgewiesen werden (Zettler, 2011).

2.1.11 Selbstregulation als Prädiktor von Studienerfolg

Selbstregulatorische Fähigkeiten im Studium beziehungsweise beim Lernen, manchmal auch als *Metakognition* bezeichnet, erwiesen sich ebenfalls als gute Prädiktoren für Studienerfolg. So konnte Minnaert (1996) anhand einer Stichprobe von $N = 517$ Studienanfängern zeigen, dass das Wissen über und der Gebrauch von metakognitiven Strategien im Studium inkrementelle Varianzaufklärung hinsichtlich der erreichten Leistungspunkte am Ende des ersten Studienjahres leisten konnte, über den prädiktiven Wert von Intelligenz (erfasst mit einem adaptierten Intelligenztest) hinaus (gemeinsam 15% Varianzaufklärung, davon 8% durch die metakognitiven Variablen). Darüber hinaus liefert er den Beleg, dass Metakognition im Studium gegebenenfalls vorhandene Intelligenz-Defizite in gewissem Rahmen kompensieren kann. Er resümiert, dass Intelligenz und Metakognition nur teilweise überlappende Konstrukte zu sein scheinen. Der Metakognition sollte allerdings wegen ihrer Erlernbarkeit und Trainierbarkeit besondere Aufmerksamkeit im Bereich der Förderung Studierender zukommen.

Mit dem Fokus auf eine andere Art von selbstregulatorischem Verhalten konnte Britton und Tesser (1991) zeigen, dass „Zeitmanagement“-Fähigkeit inkrementelle Validität über Highschool-Noten hinaus haben, was die Vorhersage des Notendurchschnitts vier Jahre später im Studium betrifft. An dieser Studie nahmen $N = 90$ US-amerikanische Studierende teil. Zwei der drei Subskalen des Zeitmanagement-Fragebogens konnten 21% ($R^2 = .21$) der Varianz im Notendurchschnitt nach vier Jahren Studium aufklären. Die Hinzunahme der Highschool-Noten als Prädiktor konnte dann bloß weitere 5% der Varianz aufklären ($R^2 = .05$). An dieser Stelle sei angemerkt, dass der Begriff „Zeitmanagement“ oft auch als ein Aspekt des *Selbstmanagement* verstanden wird; ein Begriff, der sprachlich passender zu sein scheint.

2.1.12 Akademische Resilienz als Prädiktor von Studienerfolg

Auf Basis einer Zusammenschau von diversen motivationstheoretischen Prozessen definierte Martin (2002) *akademischen Resilienz (academic resilience)* als Fähigkeit, mit Rückschlägen, Stress und Druck im Studium umzugehen. Wie diese Beschreibung schon vermuten lässt, ist das Konstrukt nicht sehr scharf definiert. Folglich finden sich in der Literatur dazu verschiedene Auslegungen. Ihnen gemein ist, dass in der Regel mehrere Facetten untersucht werden. Somit scheint *akademische*

Resilienz weniger ein Konstrukt im engeren Sinne zu sein, sondern vielmehr ein Sammelbegriff für Personeneigenschaften, von denen man annimmt, sie seien hilfreich beim erfolgreichen Umgang mit Rückschlägen und ähnlichen Herausforderungen im Studium. Exemplarisch möchte ich zwei Studien benennen, in denen Resilienz unter Studierenden untersucht wurde, deren Ergebnisse allerdings in unterschiedliche Richtungen weisen: Stack-Cutler, Parrila und Torppa (2015) untersuchten anhand einer Stichprobe von $N = 120$ Studierenden aus verschiedenen Disziplinen den Zusammenhang zwischen Resilienz, Lebenszufriedenheit, Durchhaltevermögen im Studium und Studienerfolg (operationalisiert als Durchschnittsnote). Die Besonderheit bei dieser Stichprobe lag darin, dass alle untersuchten Studierenden eine *Leseschwäche* hatten. Innerhalb dieser Gruppe zeigten – entgegen der Erwartung der Autoren – die individuumsbezogenen Resilienzfacetten (auch als *intrapersonelle protektive Faktoren* bezeichnet; Eigenschaften der Studierende, wie Selbstvertrauen und soziale Kompetenz) keinen bedeutsamen Zusammenhang mit dem Maß für Studienerfolg. Bajaj und Pande (2016) dagegen konnte anhand einer Stichprobe von $N = 327$ indischen Studierenden zeigen, dass Resilienz als Mediator zwischen *Mindfulness* und Zufriedenheit sowie (hohem) positiven und niedrigem (negativen) Affekt (mögliche Operationalisierungen von Studienerfolg) fungiert.

Zusammenfassen kann festgehalten werden, dass *akademische Resilienz* per se möglicherweise zu unscharf gefasst ist, als dass man sie als Prädiktor für Studienerfolg heranziehen sollte. Vielmehr scheint die detaillierte Betrachtung der Konstrukte, die unter diesem Sammelbegriff subsummiert werden, lohnenswert. Dies führt dann wiederum vornehmlich zu den Prädiktoren, die in diesem Kapitel bereits aufgeführt wurden.

2.1.13 Institutionelle Faktoren als Prädiktoren von Studienerfolg

Neben den oben aufgeführten individuellen Faktoren, welche in Zusammenhang mit Studienerfolg (und -abbruch) stehen, möchte ich an dieser Stelle darauf eingehen, dass sich ein weiterer Zweig der Hochschulforschung damit beschäftigt, welche Merkmale der Hochschule in bedeutsamem Zusammenhang mit diesen Kriterien stehen. Beispielsweise findet sich bereits bei Krempkow (2008), der Analysen in $k = 150$ Studiengängen in Sachsen durchführte, dass neben individuellen Prädiktoren des Studienerfolgs auch institutionelle Faktoren wie die Hochschulart, die Anzahl (Mit-)Studierender sowie die Qualität der zur Verfügung stehenden Bibliothek und die Auslastung eines Studiengangs (als Maß für die Überfüllung) geeignet sind, die Abschlussnote im Studium vorherzusagen.

Weiterhin konnte Georg (2009) anhand von Daten aus dem *Konstanz Student Survey* zeigen, dass das Commitment der Studierenden ihrer Hochschule gegenüber einen entscheidenden Einfluss auf den Verbleib im Studium (versus Studienabbruch) hat. Er leitet daraus ab, dass Hochschulen indirekt das Commitment fördern könnten, indem sie sich um eine hohe Lehrqualität bemühen.

Oseguera und Rhee (2009) analysierten Fragebogendaten (Fragebogen an Studierende und Institutionen) und administrative Daten von 170 Hochschulen/Instituten und insgesamt $N > 37.000$ US-amerikanischen Studierenden, um mit Hilfe hierarchischer Regressionsmodelle auf (i) institutioneller

und (ii) individueller Ebene die Bedeutsamkeit verschiedener potentieller Prädiktoren des Studienabbruch zu untersuchen. Sie berichten, dass das *institutionelle Klima* (operationalisiert als durchschnittliche Höhe der Ausprägung von Abbruchintentionen aller Studierenden am Institut) ein signifikanter Prädiktor für Studienabbruch sei (sogar unter Kontrolle etablierter Prädiktoren wie beispielsweise Schulnoten und sozio-ökonomischer Status), wohingegen das an der Fakultät vorherrschende Klima aus Sicht der Fakultät (hohe Priorisierung für Lehren und Lernen, multikulturelle Umgebung und aktive Lernsettings) keinen signifikanten Beitrag zur Varianzaufklärung im Kriterium leistet. Sie interpretieren dies so, dass das Klima innerhalb der Studierendenpeergroups großen Einfluss auf die individuellen Studienabbruchsentscheidungen habe, weshalb die Institute diesem Klima große Beachtung schenken sollten.

Aktuellere Forschung von Schaeper (2019) zeigt in eine ähnliche Richtung. Hierzu ist allerdings zu sagen, dass in dieser Studie *Soziale Integration* das vorherzusagende Kriterium war. Auf Basis von Daten aus der *National Educational Panel Study (NEPS)* (Blossfeld, Roßbach & Maurice, 2016) mit $N = 10.697$ Studierenden in Deutschland, konnte sie anhand eines Multi-Level-Modells zeigen, dass neben Maßen für Selbstwirksamkeit, kognitive Fähigkeiten, Gewissenhaftigkeit sowie Zielsetzung auf individueller Ebene, auf institutioneller Ebene aktivierende Lehrmethoden (im Vergleich zu Lehrmethoden, bei denen Studierende eher passive Rollen einnehmen) ebenfalls einen signifikanten Beitrag zur Varianzaufklärung im Kriterium *soziale Integration* leisteten. Wie im Abschnitt 2.1.9 beschrieben, ist soziale Integration ein bedeutsamer Prädiktor von Studienerfolg. Somit ist anzunehmen, dass die von Schaeper (2019) identifizierten Prädiktoren zumindest indirekt (mediert über die soziale Integration) mit Studienerfolg assoziiert sein könnten.

Zusammenfassend lässt sich sagen, dass auch institutionelle Faktoren, neben individuellen Faktoren, geeignet sind, Studienerfolg vorherzusagen. Hierbei ist jedoch auch die Interaktion der beiden Ebenen zu berücksichtigen, wie auch Schaeper (2019) festhält: Einerseits gibt es empirische Belege für differenzielle Wirkungen verschiedener Lehrmethoden und weiterer institutioneller Attribute. Andererseits ist deren subjektive Beurteilung durch den einzelnen Studierenden auch Ausschlag gebend dafür, wie sein individueller akademischer Werdegang verläuft (Studienabschluss, Verbleib an der Hochschule, Hochschulwechsel, Fachwechsel, Studienabbruch).

2.2 Zusammenfassende und vergleichende Erkenntnisse zu Prädiktoren von Studienerfolg

Über den Erkenntnisgehalt von Einzelstudien hinaus gehend, geben Metaanalysen Informationen zu Effektstärken beobachteter Effekte sowie zur Generalisierbarkeit von Befunden und potentiell bieten sie auch die Möglichkeit, Effekte einzelner Prädiktoren hinsichtlich ihrer Bedeutsamkeit zu vergleichen. Obgleich einige metaanalytischen Befunde in den vorherigen Ausführungen teilweise am Rande bereits erwähnt wurden, werde ich im Folgenden ein Überblick über die für diese Arbeit bedeutsamen Metaanalysen gegeben, um insbesondere die relative Bedeutsamkeit einzelner Prädiktoren (im

Vergleich zu anderen) zu beleuchten.

Aus der bereits oben erwähnten Metaanalyse von Trapmann et al. (2007) zur Validität der Vorhersage von Studienerfolg mittels Schulnoten geht hervor, dass der Zusammenhang von Schulnoten und Studienerfolg im Allgemeinen als bedeutsam einzustufen ist. In diese Analyse wurden insgesamt 83 Koeffizienten aus Studien, die zwischen 1980 und 2005 publiziert wurden, einbezogen. Darunter Studien aus Deutschland, Großbritannien, Österreich, Tschechien und Norwegen. Als Prädiktoren wurden teilweise einzelne Fachnoten verwendet oder die Durchschnittsnote in der Schule bzw. die Abiturnote. Kriterien in den Primärstudien waren durchschnittliche Studiennoten in diversen Studiengängen, wie Wirtschaftswissenschaften, Psychologie, Veterinär-, Zahn, Humanmedizin, Mathematik, diverse Naturwissenschaften, Ingenieurwissenschaften, Maschinenbau und verschiedene Fächer aus den Sprach- und Kulturwissenschaften. Es ergab sich unter Einbezug von 53 Korrelationskoeffizienten und einer Gesamtstichprobe von $N = 48.178$ ein Zusammenhang zwischen der Durchschnittsnote in der Schule und Durchschnittsnote im Studium (zu unterschiedlichen Zeitpunkten im Studium erhoben) von $\rho = .46$ (nach Korrektur der Kriteriumsreliabilität), der sogar noch etwas höher lag, wenn man nur die Studien mit deutscher Stichprobe betrachtet.

Ähnlich hohe Zusammenhänge wie für die Durchschnittsnote in der Schule fanden sie auch für die Mathematiknote, wobei diese als besonders geeigneter Prädiktor der Bachelor-Abschlussnote in den *MINT*-Fächern erscheint, da hier die Zusammenhänge wesentlich höher waren als in den Sprach- und Kulturwissenschaften, sowie der Medizin. In die Analysen in Bezug auf die Mathematiknote wurden allerdings nur deutsche Studien einbezogen. Somit erwies sich insgesamt allerdings die Durchschnittsnote in der Schule als besserer Prädiktor für die Studiennote als einzelne Fachnoten. Trapmann et al. (2007) berichten weiter von einem Trend, dem zufolge Studien neueren Datums höhere Validitäten aufzeigten: Sie setzen ihre Befunde in Bezug zu den Ergebnissen der Metaanalyse, die Baron-Boldt (1989) zur Vorhersage von Studienerfolg publizierten. Da Trapmann et al. (2007) im Durchschnitt etwas höhere Werte für die Validität fanden, resümieren sie, dass die deutschen Schulnoten über die Zeit hinweg an Validität gewonnen hätten.

Kuncel, Hezlett und Ones (2004) fassten Studien zur Validität des *MAT* (*Miller Analogies Test*) mit akademischer sowie beruflicher Leistung zusammen. Beispielsweise bezogen sie alleine für das Studienerfolgskriterium *Durchschnittsnote im Studienabschluss* $k = 70$ Studien mit Rohdaten von $N = 11.368$ Studierenden heran. Sie kommen zu dem Schluss, dass die in diesem Test erzielten Ergebnisse als Maß für kognitive Fähigkeiten in signifikantem Zusammenhang mit akademischer und beruflicher Leistung stehen. Für den Zusammenhang mit der Durchschnittsnote im Studienabschluss berichten sie eine nach Stichprobengröße gewichtete mittlere Korrelation in Höhe von $\rho = .27$.

Robbins et al. (2004) fassten in ihrer Metaanalyse die Ergebnisse von 109 Studien, die zwischen

1973 und 2001 mit US-amerikanischen Studierenden durchgeführt wurden, zusammen. Darunter war neben publizierten Studien auch nicht-publizierte. Untersucht wurde der Zusammenhang von psychosozialen sowie Study-Skills-Faktoren mit Studienerfolg (operationalisiert als Notendurchschnitt) und Fortführung des Studiums (im Sinne des Gegenteils von Studienabbruch). Dies ist besonders zu betonen, da es meiner Kenntnis nach zum Zeitpunkt der Fertigstellung dieser Arbeit die einzige Metaanalyse in diesem Kontext ist, die zugleich zwei Kriterien (eine Operationalisierung von Studienerfolg und eine Operationalisierung des Gegenteils von Studienabbruch) berücksichtigt. Die einzelnen Stichprobengrößen lagen hinsichtlich des Kriteriums *Fortführung des Studiums* zwischen $n = 40$ und $n = 3369$, hinsichtlich des Kriteriums *Studienerfolg* zwischen $n = 24$ und $n = 4805$. Die Fortführung des Studiums wies durchschnittlich moderate Zusammenhänge mit akademischen Zielen, akademischer Selbstwirksamkeit und studienbezogenen Fertigkeiten auf ($\rho = .340$, $\rho = .359$ und $\rho = .366$). Akademische Selbstwirksamkeit und Leistungsmotivation waren darüber hinaus die besten Prädiktoren für Studienerfolg ($\rho = .496$ und $\rho = .303$).

Richardson et al. (2012) fassten mit ihrer Metaanalyse 7.167 Studien zu 50 potentiellen Prädiktoren von Studienerfolg (operationalisiert als Notendurchschnitt) zusammen. Sie berichten, dass demografische und psychosoziale Kontextfaktoren im Allgemeinen niedrige Korrelationen mit dem Studienerfolgskriterium aufweisen (bspw. $\rho = .15$ für sozioökonomischen Status und $\rho = .13$ für akademische Integration), wohingegen sich für die Schulnoten Korrelationen mittlerer Höhe zeigten ($\rho = .41$). Unter den nicht-kognitiven Prädiktoren wiesen wie oben bereits aufgeführt die Maße für Selbstwirksamkeitserwartungen („performance self-efficacy“: $\rho = .67$) neben der Zielsetzung (Notenziel: $\rho = .49$) und der Anstrengungsregulation (eine Facette von selbstregulatorischen Fähigkeiten: $\rho = .35$) eine mittlere bis hohe Zusammenhangshöhe mit Studienerfolg auf. Richardson et al. (2012) heben besonders hervor, dass performanzbezogene Selbstwirksamkeitserwartungen (abzugrenzen von allgemeinen Selbstwirksamkeitserwartungen, siehe auch Kapitel 3) von allen 50 einbezogenen Prädiktoren den höchsten Zusammenhang mit dem Kriterium aufwiesen.

Zu guter Letzt sei an dieser Stelle noch auf eine Meta-Metaanalyse verwiesen: M. Schneider und Preckel (2017) fassten 18 Metaanalysen (Primärdaten von nahezu 2.000.000 Studierenden weltweit) zur Vorhersagen von Studienerfolg zusammen und präsentieren eine Rangliste von 105 Prädiktoren, geordnet nach Effektstärke. Ihr Fokus lag einerseits auf Attributen der Studierenden und andererseits auf Attributen der Instruktionen im Lehr-Lern-Kontext. Die „Top 5“ der Prädiktoren sind: (i) Einschätzung der akademischen Leistung durch studentische Peergroup, (ii) leistungsbezogene Selbstwirksamkeit, (iii) Güte der Unterrichtsvorbereitung des Lehrenden, (iv) Klarheit und Verständlichkeit der Instruktionen durch den Dozenten und (v) von dem Studenten selbst gesetztes (Minimal-)Notenziel. Anhand dieses umfassenden Ansatzes in der Meta-Metaanalyse wird ein Vergleich der Bedeutsamkeit individueller versus institutioneller Faktoren möglich. Dass unter den „Top 5“ der Prädiktoren eben auch institutionelle Faktoren zu finden sind, unterstreicht somit deren Bedeutung.

Zusammenfassend möchte ich in einer sehr groben Kategorisierung tabellarisch festhalten, welche Prädiktorgruppen hier vorgestellt wurden und welche relative Bedeutsamkeit ihnen zukommt (Tabelle 2.1).

Tabelle 2.1: Überblick über die relative Bedeutsamkeit vers. Prädiktorgruppen

Prädiktorgruppe	Kriterium	
	Studienerfolg	Studienabbruch
soziale Herkunft	•	• •
Schulleistung/ Intelligenz	• • (•)	• •
Integration/ soz. Unterstützung	• (•)	• (•)
studienbezog. Persönlichkeitseigenschaften	• (•)	-
Motivation (u.a. Zielsetzung, Selbstwirksamkeit)	• • •	• •
Lehr/-Instruktionsqualität	• • •	-
Attribute der Hochschule (u.a. Größe, Selektivität)	-	(•)

Anm. je mehr Punkte, desto stärker der Zusammenhang, - = keine Aussage

2.3 Empirische Befunde zum Studienabbruch

2.3.1 Definition von Studienabbruch

Bevor es im Folgenden um die Vorhersage von Studienabbruch geht, soll zunächst eine Definition von Studienabbruch gegeben werden, die im Rahmen dieser Arbeit verwendet wird.

*Definition **Studienabbruch**: Im Rahmen dieser Arbeit wird Studienabbruch als das Verlassen der Hochschule ohne Studienabschluss verstanden. Das bedeutet, dass Fachwechsler und Hochschulwechsler nicht als Studienabbrecher verstanden werden.*

Diese entspricht der von Heublein et al. (2010) und von Tieben (2016) verwendeten Definition.

Wie bereits oben erwähnt, werden im Folgenden nur Befunde, die sich explizit auf *Studienabbruch* oder *Studienabbruchintentionen* beziehen, berichtet. Das bedeutet, es werden zwecks Vermeidung von Redundanzen nicht die Befunde zur Vorhersage von *Studienerfolg* wiederholt, obgleich bis zu einem gewissen Grad nichts gegen eine Übertragbarkeit der gewonnenen Erkenntnisse zur Prädiktion von *Studienerfolg* auf die Prädiktion von *Studienabbruch* spricht, insbesondere wenn jedes dieser beiden Kriterien jeweils als das Gegenteil des anderen verstanden wird.

Zwecks Übersichtlichkeit ist daher zwar die Reihenfolge der Auflistung die gleiche wie beim Kriterium *Studienerfolg*, ich gehe allerdings an dieser Stelle nur auf die Prädiktoren *Noten*, *Selbstwirksamkeit*, *Emotionen* und *Motivation* ein, da zu diesen Befunde mit expliziten Bezug zu Abbruch vorliegen.

2.3.2 Noten als Prädiktoren von Studienabbruch

Gold und Souvignier (2005) führten anhand längsschnittlicher Daten einen Vergleich zwischen Studienabbrechern und Absolventen eines Jahrgangs in den Fächern Ingenieurwissenschaften, Wirtschaftswissenschaften und Jura durch ($N = 402$). Innerhalb der Fächer Wirtschaftswissenschaften und Jura fanden sie signifikante Unterschiede (zu Ungunsten der Studienabbrecher) in der Abiturnote und fachspezifischen Schulnoten. Im Fach Ingenieurwissenschaften jedoch zeigte sich in keinem der von ihnen ausgewählten Prädiktoren von Studienefolg ein bedeutsamer Unterschied zwischen Abbrechern und Absolventen. Daraus lässt sich ableiten, dass die prädiktive Validität der Abiturnote für die Vorhersage von Studienabbruch gegebenenfalls nicht fachübergreifend gegeben ist. Spezifische Befunde zum Zusammenhang von Leistungstest und Studieabbruch, die nicht bereits unter 2.1.2 aufgeführt sind, sind mir nicht bekannt.

2.3.3 Selbstwirksamkeit als Prädiktor von Studienabbruch

Chemers und Kollegen untersuchten an $N = 373$ US-amerikanischen Studienanfängern, inwiefern akademische Selbstwirksamkeit und Optimismus (erfasst zu Studienbeginn) Leistung, empfundenen Stress und Gesundheit am Ende des ersten Studienjahres vorhersagen können. Dabei zeigte sich, dass akademische Selbstwirksamkeit sowohl direkte als auch indirekte Effekte auf die Kriterien hatte. Zu betonen ist dabei, dass die Highschool-Abschlussnote, welche als etablierter Prädiktor von Studienerfolg gilt, ebenfalls erhoben und auspartialisiert wurde, sodass die Ergebnisse darauf hindeuten, dass akademische Selbstwirksamkeit und Optimismus bei der Vorhersage von Studienleistung inkrementelle Validität über die Highschool-Abschlussnote hinaus aufweisen (Chemers, Hu & Garcia, 2001). Fellenberg und Hannover (2006) untersuchten anhand einer Stichprobe von $N = 210$ Erstsemesterstudierenden in Deutschland, welche psychologischen Variablen einerseits Studienabbruchstendenzen und andererseits Fachwechseltendenzen vorhersagen können. An ihrer Studie nahmen Studierende aus den MINT-Fächern und Studierende der Sozial- und Sprachwissenschaften teil. Von den Studienteilnehmern berichteten 12,9% ernsthaft einen Fachwechsel und 11% einen Studienabbruch zu erwägen. In den Regressionsanalysen erweisen sich das *Fachinteresse* und die *Selbstwirksamkeitserwartungen* als die besten Prädiktoren für Fachwechsel- und Abbruchintentionen, neben den ebenfalls bedeutsamen Variablen *fachbezogenes Fähigkeitsselbstkonzept* und *Bereitschaft zur Mobilisierung sozialer Unterstützung bei Studienschwierigkeiten*, die geringere Beta-Gewichte aufwiesen.

Brandstätter et al. (2006) vermuteten, dass aufgaben- bzw. studienfachspezifische Selbstwirksamkeit über etablierte Maße für kognitive Leistungsfähigkeit hinaus inkrementelle Validität hinsichtlich der Vorhersage von Studienabbruch aufweisen könnte. Sie konnten dies anhand der Daten einer Stichprobe $N = 948$ österreichische Maturanten, die zwischen 1991 und 1998 ihren Schulabschluss erworben hatten und an der Universität Linz an einer Studienberatung teilgenommen hatten (Studierende, die im Laufe der Untersuchung die Universität wechselten wurden nicht berücksichtigt) nachweisen.

2.3.4 Emotionen als Prädiktoren von Studienabbruch

In der bereits oben erwähnte Studie von Respondek et al. (2017) wies die erlebte Angst in Leistungssituationen die stärkste Assoziation zu Studienabbruchintentionen auf. Dieser Befund verdeutlicht, dass insbesondere bei der Vorhersage des Kriterium *Studienabbruch(-sintention)*, welches zweifelsohne ein multikausales Phänomen ist (Heublein et al., 2017), auch nicht-kognitiven Prädiktoren eine wichtige Rolle zukommt. Zu diesem Schluss kommt auch Pekrun (2018), der die große Bedeutung von leistungsbezogenen Emotionen auch für den akademischen Kontext betont.

2.3.5 Motivation als Prädiktor von Studienabbruch

Aus der Perspektive der *Self-Determination Theory* (Deci & Ryan, 2000) betrachtet, ist das Konstrukt der Motivation im engeren Sinne ebenfalls eines, das bei der Prädiktion von *Studienerfolg versus -abbruch* betrachtet werden sollte. In einer Längsschnittstudie mit $N = 1002$ Studienanfängern aus MINT-Fächern an neun verschiedenen deutschen Hochschulen untersuchten Rump, Esdar und Wild (2017), inwiefern die jeweils vorherrschende Art der Studienmotivation, Studienabbruch vorhersagen kann. Dabei unterschieden sie verschiedene Motivationsmuster, vorstellbar auf einem Kontinuum von der reinen intrinsischen Motivation bis zur reinen extrinsischen Motivation. Hypothesenkonform stellte sich heraus, dass je höher die intrinsische Motivation ausgeprägt ist, desto geringer die Intention zum Studienabbruch.

Zusammenfassend lässt sich festhalten, dass es bereits eine Fülle an Einzelstudien zu Prädiktoren von *Studienerfolg und Studienabbruch* gibt. Die aufgeführten Metaanalysen ermöglichen einen Überblick über die relative Bedeutsamkeit verschiedener Prädiktoren. Im Folgenden sollen nun Prädiktorkombinationen in Form von Modellen zum *Studienerfolg und Studienabbruch* thematisiert werden.

2.4 Modelle und Theorien zum Studienabbruch

Neben empirischen Befunden zu einzelnen Prädiktoren von *Studienerfolg versus -abbruch* gibt es auch diverse, teilweise konkurrierende theoretische Erklärungsansätze. Zur Zeit liegt kein Modell zur Erklärung von Studienabbruch in Deutschland vor, das im Sinne eines allgemeinen Konsenses in diesem Forschungsfeld einheitlich verwendet wird (Heublein et al., 2017).

Noch schwieriger wird es, wenn man nach einem international gültigen Modell zur Erklärung von Studienabbruch sucht: Dies ist insbesondere deshalb schwierig, da nationale Besonderheiten in den Bildungssystemen maßgeblichen Einfluss auf Studienabbruch(-quoten) haben können (Heublein, 2014;

Hüther & Krücken, 2018). Es gibt allerdings einige Erklärungsansätze zum Studienerfolg und -abbruch, die international große Beachtung fanden und das Forschungsfeld maßgeblich prägten. Einige davon werde ich im Folgenden vorstellen, wobei ich auf vier Modelle im Detail eingehen werde (die Modelle, die ich auch empirisch getestet habe). Zuvor jedoch möchte ich verschiedene Perspektiven aufzeigen.

2.4.1 Unterschiedliche Perspektiven auf Studienerfolg versus -abbruch

Heublein (2014) fassen bei der Betrachtung des Forschungsfeldes zusammen, dass psychologische Modelle immer mehr Beachtung finden: So werden Einflussfaktoren wie Persönlichkeitseigenschaften, Motivation und Erwartungen miteinbezogen. Auch Ansätze in Form von Erwartungs-Wert-Theorien werden erprobt, wie beispielsweise bei Neuville et al. (2007).

Sarceletti und Müller (2011) beschreiben den Blickwinkel von vier verschiedenen Disziplinen auf das Phänomen *Studienabbruch* auf: Pädagogik, Psychologie, Soziologie und Ökonomie. Sie fassen die in den jeweiligen Disziplinen als zentral identifizierten Variablen zusammen. So seien aus der pädagogischen Perspektive die soziale und akademische Integration zentral, aus der individuumszentrierten psychologischen Perspektive stünden kognitive und motivationale Aspekte im Vordergrund. Aus der soziologischen Perspektive werden in Zusammenhang mit den Schlagwörtern *kulturelles Kapital* und *institutioneller Habitus* einerseits der Einfluss des sozio-ökonomischen Status' der Familie, aus der ein Studierender kommt, und andererseits das Wertesystem innerhalb des akademischen Kontextes in der Vordergrund gerückt. Dahinter steht die maßgeblich durch die Arbeiten von Bourdieu und Passeron (1977) geprägte Annahme, dass Studierende, die aus bildungsferneren Elternhäusern kommen, größere Adaptionsleistungen erbringen müssen, um im akademischen Umfeld zu reussieren. Ihnen falle es weniger leicht, den institutionellen Habitus (damit sind Werte und Normen gemeint) anzunehmen, da sich dieser stärker von dem Habitus, den sie aus ihrem Elternhaus kennen, unterscheide (vgl. Sarceletti und Müller, 2011; Thomas, 2002). Das bedeutet, dass die Übernahme des institutionellen Habitus für diese Gruppe von Studierenden eine größere Adaptionsleistung erfordert. Die ökonomische Perspektive betrachte schlussendlich die Entscheidung für oder gegen einen Studienabbruch als rationale Abwägung von Kosten und Nutzen, die mit der Weiterführung eines Studiums assoziiert sind.

Die Theorien und Modelle, die bislang zur Vorhersage von Studienerfolg und oder Studienabbruch genutzt werden, fokussieren in der Regel nur auf einen Aspekt (z.B. Integration) (Weber, Daniel, Becker & Bornkessel, 2018) und vernachlässigen dabei die Multikausalität der Entstehung von Studienerfolg (Heublein et al., 2017) oder auch der Entscheidung, ein Studium abzubrechen.

Ein wirklich umfangreiches Modell der Entstehung von Studienabbruch dagegen stellten Heublein et al. (2017) vor. Basierend auf den regelmäßig durchgeführten Studienabbrecher-Studien des DZ-HW entwickelten sie ein Rahmenmodell, das in Listen-ähnlicher Weise nahezu alle Kontexte berücksichtigt, aus denen Faktoren auf eine Studienabbruchsentscheidung hin einwirken können (siehe

Abbildung 2.1). Zwar birgt das Modell von Heublein et al. (2017) eine Fülle an Informationen, aus unten genauer dargelegten Gründen eignet es sich allerdings weniger für empirische Untersuchungen. Nichtsdestotrotz möchte ich es an dieser Stelle vorstellen, zumal es meiner Kenntnis nach bislang im deutschen Hochschulkontext das meist rezipierte Modell ist. In diesem Modell finden sich nahezu alle in vorangegangenen Studien als separate Prädiktoren betrachtete Variablen wieder. Der Prozess der Entscheidung für oder gegen einen Studienabbruch wird unter Berücksichtigung der vorangegangenen Erkenntnisse hinsichtlich relevanter Einflussfaktoren dargestellt. Zentral ist in diesem Modell die aktuelle Studiensituation des Individuums, wobei die vorangegangene Phase des Studieneinsteigs und die kommende Phase der Entscheidung für oder gegen den Studienabbruch ebenfalls im Modell widergespiegelt wird.

Heublein et al. (2017) unterscheiden folglich zwischen den drei Stadien: Studienvorphase, aktuelle Studiensituation und Entscheidungsphase. Dabei sind Prädiktoren der Studienabbruch-Entscheidung in der Studienvorphase und der aktuellen Studiensituation verortet, während die Entscheidungsphase dadurch gekennzeichnet ist, dass sich die individuelle Studienabbruchmotivation herauskristallisiert hat, die wiederum letztendlich zur terminalen Entscheidung (Studienabbruch oder Verbleib im Studium) führt.

Eine genauere Betrachtung der Studieneingangsphase in diesem Modell, zeigt, dass hier tendenziell stabile Merkmale des Individuums aufgeführt werden. Dazu zählen Herkunft (soziale Herkunft, Bildungsherkunft, Migrationshintergrund), Persönlichkeit (im Sinne der *Big 5* sensu Costa und McCrae (1992) mit besonderem Augenmerk auf Offenheit, Extraversion und Gewissenhaftigkeit) sowie die Bildungssozialisation (besuchte Schularten, Berufsausbildung, Übergangstätigkeit und fachliche Studienvoraussetzungen). Herkunft und Persönlichkeit beeinflussen gemäß dem Modell die Bildungssozialisation. Alle drei beeinflussen dann wiederum gemeinsam die Studienentscheidung. Im Modell ist die Studienentscheidung an der Schnittstelle zwischen Studienvorphase und aktueller Studiensituation verortet.

Die aktuelle Studiensituation ist das Herzstück des Modells. Hier findet sich eine umfangreiche Sammlung relevanter Variablen: Die vier übergeordneten Faktoren Studienbedingungen (Betreuung, Lehrqualität und Anforderungen), Lebensbedingungen (Finanzierung, Familie, Wohnsituation), Information (institutionelle Beratung, Eltern, Peergroups) und Alternativen (Fachwechsel, Ausbildung, Übergangstätigkeit) beeinflussen, was Heublein et al. (2017) als *individuellen Studienprozess* zusammenfassen. Hierunter fallen Studienverhalten (Integration, Lernstil, Zeitmanagement), Studienmotivation (Nutzen des Studiums, Fachidentifikation, Berufsperspektive), Studienleistung (Leistungsfähigkeit und -bereitschaft) und physische sowie psychische Ressourcen (Eigenaktivität, Stressresistenz, Gesundheit). Alle diese Faktoren in der aktuellen Studiensituation beeinflussen gemäß Heublein et al. (2017) individuelle Studienabbruchmotive und führen letztendlich zur terminalen Entscheidung für oder gegen einen Studienabbruch.

Diesem Modell ist als Rahmenmodell für die Studienabbruchforschung in Deutschland sicherlich ein

Modell des Studienabbruchprozesses

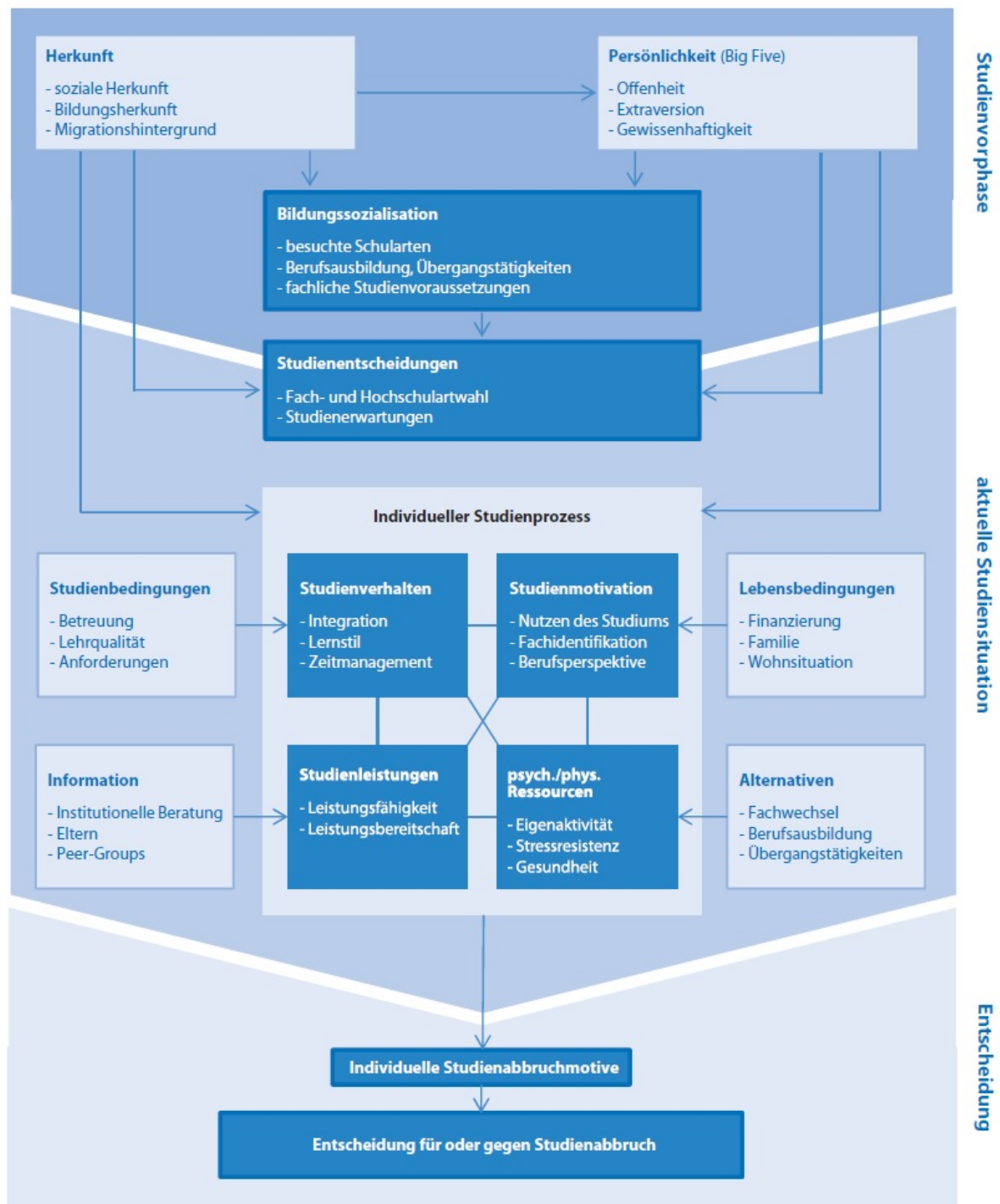


Abbildung 2.1: Modell des Studienabbruchprozesses, entnommen aus Heublein et al., 2017 (S.12)

großer Wert zuzuschreiben. Es ist allerdings in zweierlei Hinsicht wenig geeignet als Prozessmodell zur Modellierung dessen, was in der Studieneingangsphase (auf sozial-kognitiver Ebene) passiert: Zum einen ist es mit seinem Listen-ähnlichen Charakter zwar ausführlich, weist allerdings keinerlei Zusammenhänge zwischen den aufgeführten Faktoren (= Variablen) auf. Es wird somit keine Annahme über mögliche Interaktionen, also mögliche Moderationseffekte oder Mediationseffekte zwischen den Faktoren getroffen. Solche Zusammenhänge sind jedoch bedeutsam, wenn man auf Mikroebene (individuumszentriert) verstehen möchte, was bei den Studierenden geschieht, um dort gegebenenfalls zu intervenieren. Zum anderen ist das nahezu erschöpfende Modell viel zu komplex, um es mit ausreichend statistischer Power empirisch zu testen. Dafür bedürfte es mindestens solcher Stichprobenumfänge wie in der 21. Sozialerhebung (Middendorff et al., 2016). In diesen großen Panelstudien werden jedoch in der Regel keine psychometrisch hochwertigen, ausführlichen Fragebogeninstrumente eingesetzt, sondern oft so genannte „single-shot-Messungen“ durchgeführt. Das bedeutet, dass einzelne Variablen oft nur mit einem einzigen Item erfasst werden, was insbesondere aus psychodiagnostischer Sicht Zweifel an der psychometrischen Güte der Messungen aufwirft.

Zum Zeitpunkt der Fertigstellung dieser Arbeit erschien die erste Publikation, die sich mit einer empirischen Testung des Modells beschäftigt: Isleib, Woisch und Heublein (2019) verwendeten Daten aus der DZHW-Exmatrikuliertenbefragung aus dem Jahr 2014. Anhand dieses Querschnittsdatensatzes untersuchten sie die im Modell postulierten Zusammenhänge mittels multipler logistischer Regression (Vorhersage des dichotomen Kriteriums *Abbruch*). Letztendlich flossen in die Analysen Daten von $N = 5.580$ Personen ein, 55% davon wurden als *Studienabbrecher* kategorisiert. Zwar wurden nicht nur single-shot-Messungen zur Erfassung der Variablen verwendet, aber zumeist nur sehr wenige Items eingesetzt, um ein Konstrukt abzubilden. Die Ergebnisse, die Isleib et al. (2019) berichten, bestätigen vorangegangene Befunde zu bekannten Prädiktoren (z.B. Abiturnote, Integration). Neu ist, dass sie mit ihren Ergebnissen erste Belege liefern, dass auch die Erwerbstätigkeit und die Fahrtdauer zwischen Wohnort und Hochschule in signifikantem Zusammenhang mit dem Kriterium stünden.

Somit liegt nunmehr ein erster empirischer Test des Modells von Heublein et al. (2017) vor, was zumindest die Kritik, dass es bisher nicht empirisch überprüft sei, obsolet macht. Die beiden oben genannten Punkte (Modell trifft keine Aussagen über Interaktionseffekte und ist so komplex, dass es schwierig ist, einen ausreichend großen Datensatz mit psychometrisch zufriedenstellenden Messungen zur Modelltestung heranzuziehen) jedoch bleiben bestehen.

Solch detaillierte Einzelstudien mit ausführlichen Itembatterien, wie ich sie im Rahmen des vorliegenden Promotionsprojektes durchgeführt habe, erzielen jedoch meist nicht annähernd so große Stichprobenumfänge wie die Panelstudien, sodass diese Datensätze wiederum zu geringe Umfänge haben, um das von Heublein et al. (2017) vorgeschlagene Modell umfassender empirisch zu testen. Wegen des Mangels an Einblicken in die Prozesse im Studieneinstieg und im Abgleich des für empirische Testungen dieses Modells notwendigen Stichprobenumfangs mit den im Rahmen meines Promotionsprojektes realisierbaren Stichprobenumfängen, habe ich mich daher dagegen entschieden, dieses Modell in meine empirischen Analysen einzubeziehen. Es gehört daher nicht zu den vier Modellen, die ich im Folgenden ausführlicher beleuchten werde, obgleich dadurch die Bedeutung des

Modells von Heublein et al. (2017) als „Landkarte“ in der Forschung zum Studienabbruch keinesfalls in Frage gestellt werden soll.

2.4.2 Im Fokus: Vier etablierte Modelle

Die vier Modelle, die im Rahmen dieser Arbeit genauer beleuchtet werden, habe ich ausgewählt, weil sie entweder zu den meist rezipierten Modellen in der Forschung zum Studienabbruch gehören (Spady, 1971; Tinto, 1975) oder repräsentativ für theoretische Perspektiven sind (Erwartungs-Wert-Modell: Neuville et al. (2007); sozial-kognitives Modell: Lent und Brown (2013)). Zudem haben die ausgewählten Modelle gemeinsam, dass sie die Vorhersage sowohl von *Studienleistung* als auch von *Studienabbruch(-intention)* thematisieren. Darüber hinaus enthalten alle vier Modelle zwei weitere Variablen (Schulleistung und sozio-ökonomischer Status der Herkunftsfamilie), die man dem Bereich der Studienvoraussetzungen bzw. des sozio-ökonomischen Hintergrundes zuordnen kann.

Weitere Gemeinsamkeiten zwischen einigen der Modellen und insbesondere deren Unterschiede möchte ich im Folgenden aufzeigen, nachdem ich zunächst jedes Modell kurz skizziere.

Dabei ist anzumerken, dass alle vier Modelle individuumszentriert sind. Das bedeutet, dass die enthaltenen Variablen individuelle Merkmale sind, also Merkmale der Hochschule oder der Lehre (institutionelle Faktoren) keine explizite Berücksichtigung finden. Im Rahmen der vorliegenden Arbeit steht insgesamt die individuumszentrierte Perspektive im Fokus. Gleichwohl sollen Einflüsse der Hochschule nicht unberücksichtigt bleiben, weshalb bereits in Abschnitt 2.1.13 Befunde zu derartigen Prädiktoren aufgeführt wurden und auch in Kapitel 12 noch weiter auf diese Einflussfaktoren eingegangen wird.

Modell 1: Spady's Dropout Model

Das erste Modell, das ich betrachten möchte, ist das *Dropout Model* von Spady (1971). Es ist das älteste unter den vier hier getesteten Modelle und hat die Forschung zum Thema *Studienabbruch* geprägt. Spady (1971) betrachtet die Entscheidung, ein Studium abzubrechen, als Entscheidung, ein soziales System, die Hochschule, zu verlassen. Diese Idee lehnte er an die zum Ende des 19. Jahrhunderts entstandene Arbeit *Suicide* von Durkheim (1951) an, welche als erste soziologische, empirische Arbeit unter Verwendung einer Fallstudie Maßstäbe für nachfolgende soziologische Werke setzte und als einen zentralen Aspekt bei der Entstehung der terminalen Entscheidung zum Selbstmord die *soziale Integration* nennt. Spady überträgt dies auf den Hochschulkontext als soziales System, den Studienabbrecher final verlassen möchten. Entsprechend sieht er unter den Antezedenzen neben der erbrachten akademischen Leistung insbesondere soziale Integration, Zufriedenheit und Commitment.

Zentrale Idee in Spadys Modell ist, dass jeder Studienanfänger mit seinem individuellen Muster an sozialen und akademischen (im Sinne von Schulleistung) Erfahrungen die Hochschule betritt und eine Anpassungsleistung vollbringen muss. Je höher dabei die normative Kongruenz sei (Passung des Wertesystems des Individuums zu dem des akademischen Umfeldes = Person-University-Fit), desto einfacher und erfolgreicher die soziale Integration.

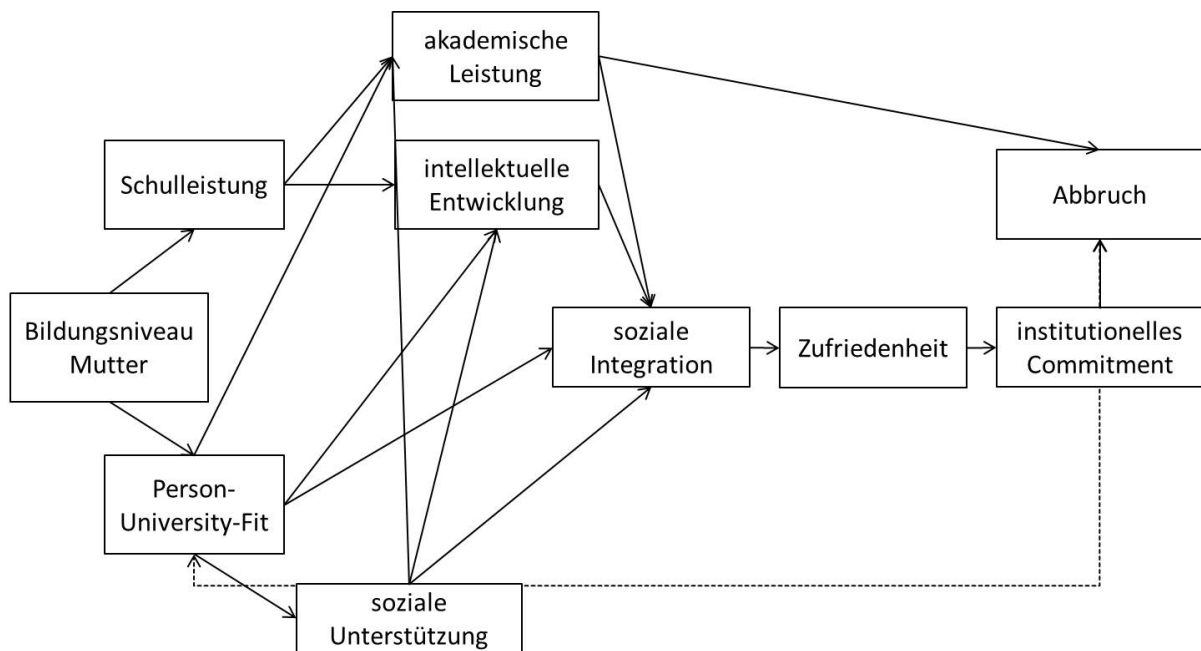


Abbildung 2.2: Spadys Dropout Modell (Spady, 1971, S. 58, eigene Abbildung)

Da die sozialen und akademischen Anforderungen und Rollen im Kontext Hochschule großflächig überlappten, würden Spady (1971) zu Folge der Person-University-Fit und die soziale Unterstützung sowie die (kognitiven) Fähigkeiten und die Vorbildung zur intellektuellen Entwicklung und den formalen akademischen Leistungen beitragen. Voll integriert seien Studierende dann, wenn sie die sozialen und akademischen Anforderungen erfüllten. Er geht weiter davon aus, dass die intellektuelle Entwicklung und die soziale Integration im Studium beeinflussen, in welchem Ausmaß Zufriedenheit erlebt wird. Diese wiederum beeinflusst, wie stark der jeweilige Studierende Commitment zu seinem Studium zeigt, was sich direkt auf Studienabbruch auswirken soll (Spady, 1971). Abbildung 2.2 zeigt *Spady's Dropout Model*.

Anhand einer Längsschnittstudie mit $N = 683$ Studienanfängern einer US-amerikanischen Hochschule untersuchte er mittels multipler Regressionsanalysen die empirische Passung seines Modells. Die in seinem Modell aufgeführten Prädiktoren konnten circa 30% der Varianz in den Abbruchintentionen aufklären. Basierend auf der Betrachtung der Beta-Gewichte der einzelnen Prädiktoren schlägt er kleinere Revisionen seines Modells vor. Er limitiert die Aussagekraft dieser empirischen Befunde allerdings mit dem Verweis darauf, dass viele zuvor nicht erprobte Instrumente zur Erfassung der Variablen zum Einsatz gekommen seien. Das berge zwei Schwierigkeiten: (i) Einerseits lägen für diese Instrumente keine Referenzkennwerte vor. (ii) Andererseits beobachtete er deutliche Unterschiede in den Kennwerten für Frauen versus Männer, was genauerer Untersuchung bedürfe. Beides seien Probleme, die in nachfolgender Forschung adressiert werden sollten. Trotz dieser Limitationen sei sein Modell und dessen erster empirischer Test ein guter Ausgangspunkt für weitere Forschung.

Zusammenfassend resümiert Spady (1971), dass die empirischen Befunde zeigen, dass sowohl für Frauen als auch für Männer der Person-University-Fit eng mit dem institutionellen Commitment zusammenhängt. Während jedoch innerhalb der Subgruppe weiblicher Studierender zusätzlich die soziale Eingebundenheit vornehmlich eng mit dem institutionellen Commitment assoziiert sei, sei es innerhalb der Subgruppe männlicher Studierender der familiäre Hintergrund, der eng mit dem institutionellen Commitment assoziiert sei. Schlussendlich seien das Commitment und die akademische Leistung jedoch in beiden Subgruppen für den Verbleib an der Hochschule prädiktiv.

Modell 2: Tinto's Integration Model

Das zweite Modell, welches ich genauer beleuchten möchte, *Tinto's Integration Model* (Tinto, 1975), ist nur wenig später veröffentlicht worden, als Spady's Dropout Model. Tinto (1975) kritisierte die durchaus umfangreiche Studienlage zum Thema *Studienabbruch*, die bereits bis Mitte der 1970er Jahre vorlag. Er nennt zwei Hauptkritikpunkte: Zum einen fehle – trotz zahlreicher Untersuchungen zum Studienabbruch – eine detaillierte Definition. Besonders betont Tinto, dass differenziert werden sollte zwischen Studienabbruch aufgrund von Leistungsproblemen („rausgeprüft“) und freiwilligem Studienabbruch, beispielsweise aufgrund motivationaler Probleme. Weiterhin sei zu unterscheiden zwischen permanentem und temporärem Studienabbruch und (Fach-) bzw. Institutionswechsel. Zum anderen seien viele Modelle deskriptiv ausgerichtet; würden folglich wenig zur Erklärung des Zustandekommens von Studienabbruch beitragen. Bis auf wenige Ausnahmen seien die deskriptiven Ansätze wenig geeignet, um abzubilden, wie die individuelle Fähigkeit und der soziale Status die Abbruchwahrscheinlichkeit über die Zeit beeinflussen (Tinto, 1975). Hierfür seien longitudinale Untersuchungen und entsprechende Modelle notwendig.

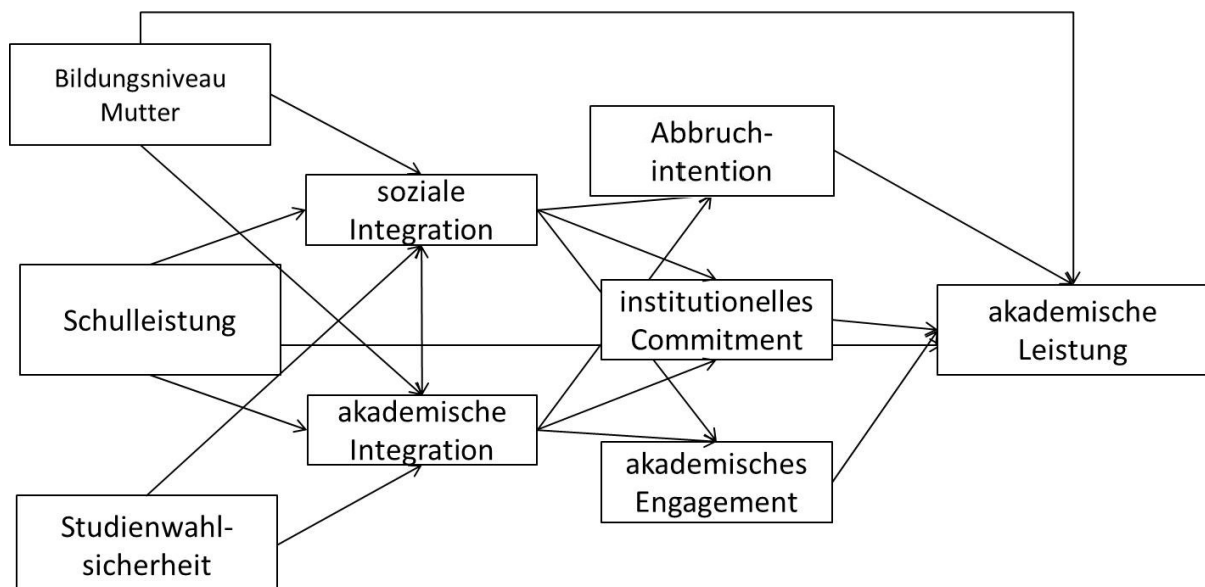


Abbildung 2.3: Tintos Integration Modell (Tinto, 1975, S. 95, eigene Abbildung)

Ebenso wie Spady (1971), sieht auch Tinto in der sozialen und akademischen Integration den Schlüssel zur Erklärung von Studienabbruch (siehe Abbildung 2.3). Er sieht den Prozess, der zu Studienabbruch führen kann, als fortlaufende Interaktion zwischen dem Studierenden und dem akademischen sowie sozialen System. Die von dem Individuum gemachten Erfahrungen in dieser Interaktion resultieren in einem gewissen Maß in Integration, welche das Commitment zu individuellen (Leistungs-)Zielen und zur Hochschule so beeinflusst, dass letztendlich gegebenenfalls die Entscheidung für das Verlassen der Hochschule resultiert. Damit lehnt er sein Modell wie Spady (1971) an Durkheim (1951) an.

Tinto beschreibt ähnlich wie Spady (1971) die Zusammenhänge so, dass durch soziale Interaktion im Hochschulkontext eine soziale Integration entstehe. Tinto (1975) unterscheidet auch im Modell, nicht nur in der Beschreibung (wie Spady, 1971) zwischen sozialer und akademischer Integration. Dabei versteht er unter sozialer Integration vornehmlich das Eingebundensein in die relevante Peergroup der Mitstudierenden, aber auch die Teilnahme an extra-curriculären Aktivitäten und die Interaktionen mit Dozenten und administrativem Hochschulpersonal. Unter akademischer Integration wiederum versteht er die Adaption an die Leistungskultur im Hochschulkontext: die intellektuelle Entwicklung im Studium sowie die akademische Leistung (in Noten) selbst (Tinto, 1975, 1975, 1997).

Der Abbruchintention vorgeschaltet umfasst sein Modell zwei weitere Variablen: Zum einen handelt es sich um das institutionelle Commitment, was seinen Ausführungen nach als Teil des unter 2.1.10 beschriebenen *University Citizenship Behavior* verstanden werden kann. Zum anderen handelt es sich um *goal commitment*, was seinen und insbesondere den Ausführungen von Neuville et al. (2007) zur Folge allerdings eher als akademische Engagement verstanden werden kann und sich beispielsweise in Form der durchschnittlichen Wochenstundenzahl, die für Studienzwecke zusätzlich zum Besuch von Präsenzveranstaltungen aufgebracht wird, manifestiert.

Zudem enthält sein Modell zusätzlich die (sehr grob gefasste) Variable der *individuellen Attribute*. Damit meint er „individuelle Charakteristiken und Dispositionen“ (S. 6), welche für den Verbleib an der Hochschule von Bedeutung sind. Dazu zählen neben demografischen Variablen (wie Geschlecht) auch (kognitive) Fähigkeiten. Er empfiehlt allerdings die Schulleistung als Prädiktor für Studienabbruch zu verwenden, anstatt einem Testwert in einem kognitiven Test, da Schulnoten in engerer Beziehung zu Studiennoten stehen (siehe dazu auch Abschnitt 2.1.2). Zudem fließen hier auch Aspekte ein, die in Zusammenhang mit der Studienwahl stehen. Wie Neuville et al. (2007) ausführt, kann beispielsweise die Studienwahlsicherheit ein solcher Aspekt sein.

Tinto (1975) untermauert sein theoretisches Modell durch eine Vielzahl an zu dem Zeitpunkt vorliegenden empirischen Befunden, die die einzelnen von ihm vorgeschlagenen Pfade in seinem Modell erklären. Er greift zwar in großen Teilen das Modell von Spady (1971) auf, erweitert dieses jedoch, sodass sich im Rahmen der vorliegenden Arbeit eine Differenzierung der beiden Modelle lohnt, obgleich sie sich beide als Integrationsmodelle kategorisieren lassen.

In einer Intervention-Kontrollgruppen-Studie untersuchte Tinto (1997) Studierende eines US-amerikanischen Colleges. Während die Interventionsgruppe im ersten Studienquartal an einem *Coordinated Studies*

Program teilnahm, besuchten die Studierenden in der Kontrollgruppe regulär die von ihnen individuell gewählten Kurse. Das *Coordinated Studies Program* zielte darauf ab, den Studierenden eine Lerngemeinschaft zu bieten, indem sie als Gruppe alle gemeinsam die verschiedenen Kurse besuchten und an von Programmseite her organisierten wöchentlichen Gruppenlernaktivitäten teilnahmen. Die Studie bestand aus einem quantitativen und einem qualitativen Teil. Im qualitativen Teil wurden im Sinne von Fallstudien Interviews mit Studierenden der Interventionsgruppen geführt. Ziel war es, aus Perspektive der Studierenden zu erfahren, wie die Teilnahme an einem solchem kollaborativem Lernprogramm ihre Lernerfahrungen und somit ihre Erfahrungen im ersten Studienjahr im Allgemeinen beeinflusst hat. Zu drei Zeitpunkten (Beginn und Ende des ersten Studienquartals und im folgenden Quartal nach Programmende) wurden Interviews geführt. Im quantitativen Teil wurde eine längsschnittliche Fragebogenstudie durchgeführt: Die Teilnehmer bearbeiteten zu Beginn und zu Ende des ersten Studienquartals ($N = 287$) einen Fragebogen und diese Daten wurden in Zusammenhang gebracht mit den nach einem Jahr erreichten Creditpoints und den Noten. Diese objektiven Leistungsdaten zusammen mit subjektiven Einschätzungen des Lernzuwachses waren die Kriterien im quantitativen Teil der Studie. Hinsichtlich nahezu allen Kriterien waren die Teilnehmer der Interventionsgruppen besser als die der Kontrollgruppe.

Als Resümee seiner Arbeit nennt Tinto (1997) verschiedene konkrete Forschungsdesiderate. Aspekte, die seiner Meinung nach weiterer Untersuchung bedürfen sind z.B. eine differenzierte Betrachtung des Studienabbruchs in verschiedenen Subgruppen oder der Einfluss von Subkulturen in den Institutionen auf Verbleib im Studium. Darüber hinaus verweist er auf übergeordneter Ebene darauf, dass es notwendig sei, den Prozess des Studienabbruchs auch als Prozess, sprich mit Longitudinaldaten zu untersuchen (Tinto, 1997).

Aufbauend auf Tintos Arbeiten wurden verwandte Modelle entwickelt. Exemplarisch sei an dieser Stelle auf das Modell von Bean (1985) verweisen, der das Modell von Tinto (1975) nur leicht veränderte und dann einer umfangreichen empirischen Testung anhand einer Stichprobe US-amerikanischer Studierender unterzog. Bean (1985) betont wie Tinto auf Basis seiner empirischen Erkenntnisse die Bedeutung der Sozialisation im Studium, die sich innerhalb der studentischen Peergroup vollziehe. Kerby (2015) wiederum resümiert 40 Jahre nach der Veröffentlichung von Tintos Modell des Studienabbruchs: Tintos Perspektive, dass Studienabbruch oder Verbleib in der Hochschule als Prozess zu betrachten sei, bei dem Fähigkeit und soziale wie institutionelle Integration die bedeutenden Prädiktoren sind, sollte erweitert werden. Hochschulbildung bewirke einen Prozess des sozialen Wandels, im Zuge dessen nicht bloß Wissen vermittelt werde, sondern auch Sozialisation und Selbstentdeckung stattfinde. Da sich das (bildungs-)politische Klima beispielsweise in den USA in den vergangenen Jahren geändert habe und dies Studienabbruchentscheidungen beeinflussen kann, sollte es im Modell berücksichtigt werden.

Modell 3: Nevilles Erwartungs-Wert-Modell

Neuvill et al. (2007) würdigen das von Tinto (1975) vorgeschlagene Modell als wichtigen Wendepunkt in der Literatur. Sie diskutierten die Stärken und Schwächen von Tintos Modell: Sie heben positiv hervor, dass Tinto einen integrativen Ansatz wählte. Damit meinen sie, dass sowohl Hintergrundvariablen (z.B. sozioökonomischer Status, Abiturnote), als auch Intentionen zu Studienbeginn im Modell als Antezedenzen der Integration im Studium verstanden werden. Allerdings würden Theorien aus dem Bereich der Leistungsmotivation nicht berücksichtigt (Neuvill et al., 2007). Als eine Kategorie der viel beforschten Leistungsmotivationsparadigmen verweisen Neuvill et al. (2007) auf Modelle mit einer Erwartungs-Wert-Struktur, die auf die Vorhersage von *Studienerfolg versus -abbruch* angewendet werden sollten.

Modelle dieser Art sind grundsätzlich so aufgebaut, dass die beiden Faktoren *Erwartung* und *Wert* zur Erklärung von Verhalten herangezogen werden (Wigfield & Eccles, 2000; Eccles & Wigfield, 2016). Dabei ist mit *Erwartung* die subjektive Einschätzung der eigenen Erfolgchance gemeint, während unter dem Begriff *Wert* die subjektive Einschätzung der Wichtigkeit oder persönlichen Bedeutsamkeit der antizipierten Verhaltenskonsequenzen zu verstehen ist. Neuvill et al. (2007) veranschaulichen, was diese beiden Faktoren inhaltlich bedeuten, indem sie sie auf zwei Fragen zurückführen: „Kann ich das tun?“ (Erwartung, S.2) und „Warum sollte ich das tun?“ (Wert, S.2). Das Konstrukt *Erwartung*, welches sie beschreiben, deckt sich im vorliegenden Zusammenhang mit der Definition von *Selbstwirksamkeitserwartung*. Daher werden im Folgenden im Bezug auf das Modell von Neuvill et al. (2007) diese beiden Begriffe synonym verwendet.

Gemäß Neuvill et al. (2007) sei die Erwartungs-Wert-Perspektive vielversprechend, da der Zusam-

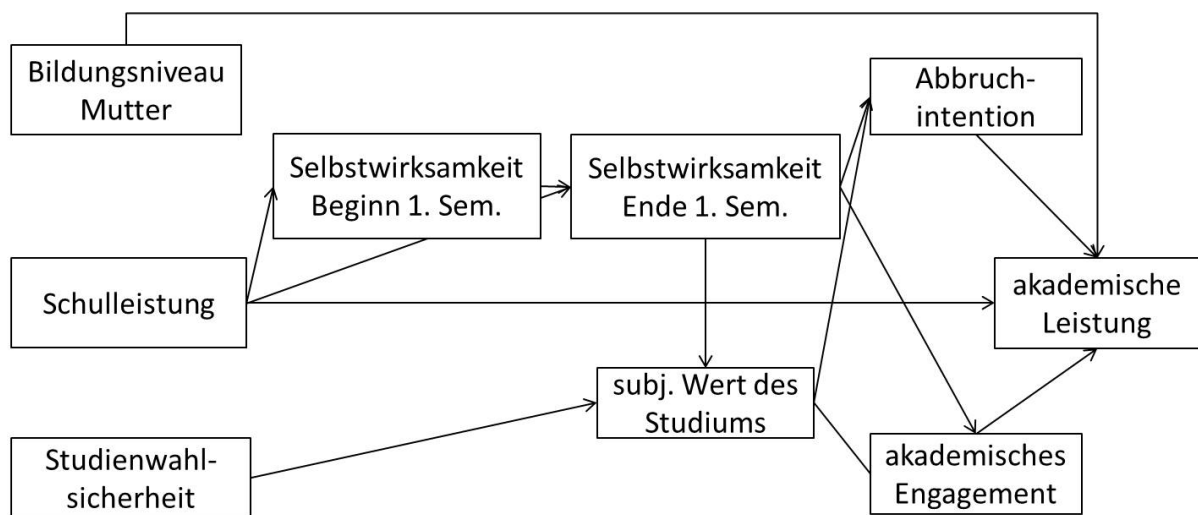


Abbildung 2.4: Erwartungs-Wert-Modell (Neuvill et al., 2007, S. 36, eigene Abbildung)

menhang der kognitiven Konstrukte *Erwartung* und *Wert* mit verschiedenen Varianten von leistungsbezogenen Verhalten, wie Persistenz und Performanz, vielfach belegt sei.

Unter anderem konnten Eccles und Wigfield (2016) für den Bereich der sekundären Bildung zeigen,

dass Kontroll- und Wert-Überzeugungen sowohl Schulnoten als auch Intentionen zum weiteren Besuch bestimmter Lernangebote vorhersagen.

Respondek et al. (2017) wiederum zeigten wie oben bereits aufgeführt, dass Erwartung und Wert interindividuelle Unterschiede im Durchhaltevermögen im akademischen Kontext vorhersagen konnten. Vor dem Hintergrund zwei nebeneinander existierender Erklärungsansätze für leistungsbezogenes Verhalten – (i) die soziale und akademische Integration sensu Spady (1971) und Tinto (1975) versus (ii) die Erwartungs-Wert-Perspektive – war es das übergeordnete Ziel von Neuville et al. (2007), diese beiden zu einem Modell zu integrieren und die beiden ursprünglichen Modelle sowie das integrierte Modell gegeneinander zu testen, um herauszufinden, welches der drei Modelle am besten zu den Daten aus einer längsschnittlichen Untersuchung an Studienanfängern einer belgischen Universität passt.

Die Prädiktoren, die in alle drei Modelle mit einbezogen wurden, waren das Bildungsniveau der Mutter, die Schulabschlussnote der Studierenden und der Grad an Sicherheit, mit dem sie sich für ihr Studium entschieden hatten. Diese Variablen wurden stets berücksichtigt, weil sie sich in vorangegangener Forschung als bedeutsame Prädiktoren von leistungsbezogenen Verhalten erwiesen haben. Die Kriterien waren stets die akademische Leistung (operationalisiert als Note) sowie die Intention, das Studium fortzuführen (als Gegenteil zu Abbruchintentionen). Die drei Modelle unterscheiden sich darin, welche intermittierenden Variablen sie zwischen den drei oben genannten allgemeinen Prädiktoren und den Kriterien postulieren. Abbildung 2.4 zeigt das von Neuville et al. (2007) postulierte Erwartungs-Wert-Modell für den Studieneinstieg.

In ihrer Studie verglichen Neuville et al. (2007) anhand der Daten von Studienanfängern ($N = 2637$) einer belgischen Hochschule drei verschiedene Modellvarianten: Das auf Integration fokussierte Modell von Tinto (1975), ein Erwartungs-Wert-Modell, das sie selbst aufgestellt hatten und eine Kombination der beiden erstgenannten. Sie konnten dabei zeigen, dass das von ihnen vorgeschlagene Modell (vgl. Abbildung 2.4) mit der Erwartungs-Wert-Struktur die Daten ihrer Stichprobe am besten wiedergeben konnte. In diesem Erwartungs-Wert-Modell, das eingangs ebenfalls die besagten Hintergrundvariablen enthält, stehen Erwartungen (respektive die Selbstwirksamkeit) zu Studienbeginn und im späteren Verlauf und der Wert, der dem Studium beigemessen wird, im Mittelpunkt. Gemäß diesem Modell beeinflussen die Erwartungshaltung und der dem Studium beigemessene Wert, in welchem Ausmaß Studieneinsteiger Persistenz und akademisches Engagement (auch beeinflusst durch die akademische und soziale Integration) zeigen. Persistenz kann in diesem Zusammenhang als das Gegenteil von Studienabbruch verstanden werden. Das Ausmaß an Studienleistung, das gezeigt wird, werde wiederum direkt von der Persistenz und dem akademischen Engagement beeinflusst.

Modell 4: Lent & Browns CSM Model

Das vierte hier betrachtete Modell (siehe Abbildung 2.5) ist das *Career Self-Management Model* (kurz: *CSM-Modell*) von Lent und Brown (2013). Ursprünglich stammt das Modell aus dem Bereich der Laufbahnforschung. Es soll den Prozess der Adaption an neue Umstände in Stadien des Umbruchs im Verlauf der (beruflichen) Entwicklung abbilden. Die Grundidee ist, dass es in Phasen des Um-

bruchs immer die gleichen relevanten Mechanismen gibt, anhand derer sich vorhersagen lässt, ob ein Individuum einen Umbruch mehr oder weniger erfolgreich bewältigen wird. Dabei ist das Modell eher allgemeingültig formuliert: Als Kriterien am Ende des Modells sind nicht konkrete Erfolgs- oder Misserfolgsmaße angegeben, sondern lediglich der Platzhalter „outcomes“ vorgemerkt. Das Modell kann folglich kontextspezifisch angepasst werden, indem beispielsweise hier im Kontext des Studieneinstiegs (Umbruch von Schule zu Hochschule) als Kriterien *Studienerfolg versus -abbruch* eingesetzt werden (vgl. Abbildung 2.5).

Grundsätzlich lässt es sich als sozial-kognitives Modell kategorisieren, da Lent und Brown (2013) im Rückgriff auf die Arbeiten von Bandura (1977) und die vom ihm postulierte *sozial-kognitive Lerntheorie* das Rahmenmodell (*Social Cognitive Model of Career Self-Management*) aufstellten. Hierbei geht es vornehmlich darum zu modellieren, wie Individuen Ziele setzen und diese erreichen unter Berücksichtigung kontextspezifischer Kognitionen und der steten Interaktion mit ihrem Umfeld. Wie

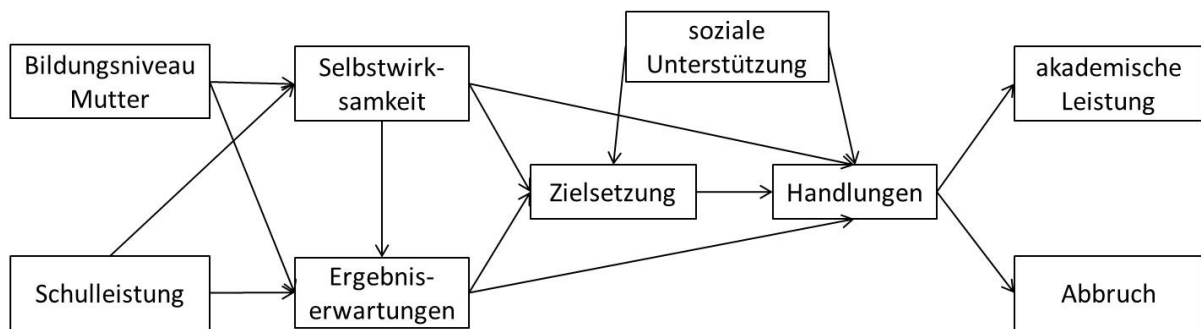


Abbildung 2.5: CSM-Modell (adaptiert nach Lent und Brown, 2013, S. 562)

die drei oben beschriebenen Modelle berücksichtigt auch dieses Modell Hintergrundvariablen. Ähnlich wie im Modell von Neuville et al. (2007) werden Erwartungen einbezogen, allerdings nicht in der klassischen Erwartungs-Wert-Struktur, wie im Modell von Neuville et al. (2007), sondern es wird (eng angelehnt an Banduras Arbeiten) zwischen Selbstwirksamkeitserwartungen und Ergebniserwartungen differenziert. Erstere bezeichnen, wie oben bereits aufgeführt, die Überzeugung, spezifische Herausforderungen erfolgreich meistern zu können. Letztere dagegen bezeichnen die Überzeugung, dass bestimmte Handlungen spezifische Ergebnisse zur Folge haben (im Sinne von: „intensive Klausurvorbereitungen führen zu guten Prüfungsleistungen“), ungeachtet persönlicher Erfolgswahrscheinlichkeit. Die Grundidee in diesem Modell – übertragen auf den Studieneinstieg – ist, dass Selbstwirksamkeitserwartung und Ergebniserwartung gemeinsam beeinflussen, welche Ziele ein Studieneinsteiger sich setzt und wie stark die Zielsetzung ausgeprägt ist (Kahn & Nauta, 2001). Dies wiederum beeinflusst, welche Handlungen im Zusammenhang mit dem Studieneinstieg tatsächlich ausgeführt werden. Sowohl auf die Zielsetzung und die dann tatsächlich ausgeführten Handlungen als auch auf die finalen „outcomes“ (wie Studienleistung oder Studienabbruch) wirken gemäß dem CSM-Modell weitere Faktoren als Moderatorvariablen: Zusammengefasst als Persönlichkeitsfaktoren und kontextuelle Faktoren führen Lent und Brown (2013) beispielsweise die soziale Unterstützung an: Personen, die ein hohes Maß

an sozialer Unterstützung erleben, sind eher dazu in der Lage, ihre Ziel auch umzusetzen.

Sowohl Lent, Brown und Larkin (1986) als auch Kahn und Nauta (2001) konnten regressionsanalytisch bereits zeigen, dass die zentralen Variablen des *CSM-Modells* valide Prädiktoren von Persistenz im Studieneinstieg sind. Lent et al. (1986) untersuchten $N = 105$ Studierende der Ingenieurs- und Naturwissenschaften. Kahn und Nauta (2001) hingegen zogen eine Zufallsstichprobe und untersuchten $N = 400$ Studierende verschiedener Fächer an einer mittelgroßen amerikanischen Universität.

2.5 Zusammenfassung des aktuellen Forschungsstandes

Eine Vielzahl an Studien untersuchte bereits verschiedene Prädiktoren von Studienerfolg. Immer wieder wird dabei deutlich, dass die Kombination verschiedener Prädiktoren besonders dann aussagekräftig ist, wenn die Prädiktoren aus unterschiedlichen Bereichen stammen. So ist offensichtlich die Kombination soziodemografischer, motivationaler und kognitiver Variablen besonders vielversprechend (De Clercq et al., 2013; Robbins et al., 2004; Richardson et al., 2012). Neben der Auswahl relevanter Prädiktoren ist deren Kombination im Sinne eines Prozessmodells erstrebenswert, um ein Verständnis vom Prozess der Studieneinstiegs zu gewinnen. Modelltests haben im Vergleich zu multiplen Regressionen den Mehrwert, dass vermutete Zusammenhänge der Prädiktoren überprüft werden können.

Mit dem Blick auf die Phase des Studieneinstiegs lässt sich sagen, dass bereits eine Reihe an Modellen vorgeschlagen wurde. Die meisten dieser Modelle sind entweder gar nicht in Gänze empirisch validiert oder nicht an der im Fokus dieser Arbeit stehenden Zielpopulation der Studienanfänger in Deutschland. Ein solcher Modelltest etablierter Modelle bietet sich folglich an. Darüber hinaus argumentierten bereits Neuville et al. (2007), dass es sich anbiete, Modelle, die in anderen Kontexten erfolgreich Verhalten erklären konnten, testweise auf den Studieneinstieg zu übertragen.

2.6 Forschungslücke: Es fehlt ein integratives Prozessmodell

Weber et al. (2018) stellten ein - wie sie es nennen - exploratives Modell im Sinne eines „Proof of Concepts“ (S. 74) anhand bisheriger Befunde zu bivariaten Zusammenhängen einzelner Prädiktoren und unterschiedlicher subjektiver und objektiver Studienerfolgskriterien auf. Dieses prüften sie empirisch anhand der Daten aus der 21. Sozialerhebung (Middendorff et al., 2016). Der große Mehrwert ihrer Untersuchung liegt einerseits darin, dass sie mit diesem großen repräsentativen Datensatz gearbeitet haben und diesen zur empirischen Prüfung eines Modells, statt „nur“ einzelner, von einander scheinbar unabhängiger bivariater Beziehungen verwendet haben. Andererseits liegt der Mehrwert auch darin, dass sie sowohl subjektive (z.B. Studienzufriedenheit), als auch objektive Studienerfolgskriterien (Studiennoten) berücksichtigt haben. Obgleich der von Weber et al. (2018) gelieferte Literaturüberblick zum Thema Studienerfolg und auch ihre empirischen Erkenntnisse zweifellos ein wertvoller Beitrag

für dieses Forschungsfeld sind, bleiben dennoch die unter 1.1.5 genannten Desiderate, insbesondere aus psychometrischer Sicht, bestehen.

2.7 Ansatzpunkt der vorliegenden Arbeit

Der vorliegenden Arbeit liegt die Annahme zugrunde, dass die Studienabbruchquote nur dann nachhaltig gesenkt werden kann, wenn ein Grundverständnis des Prozesses des Studieneinstiegs und der Entstehung von Studienabbruchintentionen vorliegt. Basierend auf einem solchen Grundverständnis sollten dann die zur Verfügung stehenden Ressourcen zur Unterstützung potentiell Abbruch gefährdeter Studierender gezielt eingesetzt werden.

Richtet man den Blick auf den Studieneinstieg, wenn die Studienwahl also bereits getroffen ist und das Studium angetreten wurde, so lassen sich wiederum verschiedene Ansatzpunkte für institutionelles Handeln ausmachen.

Dabei erscheint es sinnvoll, sich auf jene Aspekte im Prozess des Studieneinstiegs zu konzentrieren, die potentiell durch institutionelle Interventionen veränderbar sind. Das bedeutet, dass Interventionen an motivationalen (und sozialen) Aspekten ansetzen sollten, da diese am ehesten veränderbar erscheinen (eher als solche (Leistungs-) Defizite, die auf einen Mangel an Fähigkeiten zurückzuführen sind).

Während beispielsweise bei finanziellen Schwierigkeiten im Studium die von institutioneller Seite her möglichen Hilfestellungen recht klar vorgezeichnet sind (v.a. Beratungsangebote), bedarf es vornehmlich für die Aufrechterhaltung aller unter dem Oberbegriff *Studienmotivation* subsummierten Faktoren eines tieferen Verständnisses, um Studienanfänger gezielt im Rahmen institutioneller Möglichkeiten zu unterstützen.

2.7.1 Intendierter Beitrag

Diese Arbeit hat daher das Ziel, auf Basis bisheriger Befunde zum Thema *Studienerfolg versus Studienabbruch* ein geeignetes Modell zur Abbildung des Prozesses des Studieneinstieg im deutschen Hochschulkontext zu identifizieren. Da *Studienerfolg und Studienabbruch* eng miteinander zusammenhängen, werden sie im Rahmen dieser Arbeit auch gemeinsam thematisiert. Der Schwerpunkt jedoch liegt auf dem *Studienabbruch*: Angestrebt wird ein tieferes Verständnis des Entstehungsprozesses einer terminalen Studienabbruchsentscheidung. Hierbei konzentrieren sich meine Erhebungen auf die Studieneingangsphase, sprich die ersten beiden Studiensemester, zumal die Studienabbruchquoten in dieser Phase ihr Maximum innerhalb des grundständigen Studiums erreichen (Stifterverband für die Deutsche Wissenschaft, 2015; Heublein et al., 2017).

Konkret ist diese Arbeit der Beantwortung folgender Fragestellungen gewidmet:

- Welches der vier hier fokussierten etablierten Modelle zum Studienerfolg versus Studienabbruch kann empirische Daten, erhoben in der Population der Studierenden in Deutschland, am besten erklären?
- Wie könnte ein neues integratives Modell aussehen, dass die Stärken der bereits vorhandenen Modelle kombiniert und zusätzlich neuste metaanalytische Erkenntnisse berücksichtigt?
- Welche sind die Konstrukte, die gemäß dem neuen Modell eine Schlüsselfunktion im Prozess des Studieneinstiegs haben und daher in der Hochschulpraxis beispielsweise in Form eines Screenings besonders beachtet werden sollten?

3 Skalenentwicklung: Erfassung von Selbstwirksamkeit im Studieneinstieg

Die hohe prädiktive Validität von Selbstwirksamkeit bei der Vorhersage von Studienerfolg und Studienabbruch (-intentionen) wurde bereits im vorherigen Kapitel ausführlich dargelegt (vgl. 2.1.6 und 2.3.3). Da Selbstwirksamkeit somit einerseits als einzelner Prädiktor eine wichtige Rolle bei der Vorhersage dieser Kriterien zu spielen scheint und andererseits sowohl in dem Modell von Neuville et al. (2007) als auch in dem Modell von Lent und Brown (2013) aufgeführt ist, stellte sich nunmehr die Frage, wie sie operationalisiert werden sollte.

Wann immer möglich, habe ich versucht, bereits etablierte Skalen zu verwenden. Dies war jedoch nicht für alle Konstrukte, die erfasst werden sollten, möglich. Teilweise waren die vorhandenen Instrumente nicht ganz passend für den Kontext meiner Erhebungen und teilweise fehlte es gänzlich an geeigneten Skalen für meine Messintentionen. Im erstgenannten Fall habe ich bereits vorhandene Instrumente leicht adaptiert. Sofern kein deutschsprachiges in Frage kommendes Instrument vorlag, habe ich auch selbst Items beziehungsweise Skalen entwickelt (vgl. Kapitel 4).

In diesem Kapitel möchte ich die Entwicklung einer Skala zur Erfassung von *Selbstwirksamkeit im Studieneinstieg* im Detail darlegen. Dazu werde ich zunächst kurz erläutern, warum die Entwicklung einer neuen Skala notwendig war. Anschließend möchte ich sowohl die Itementwicklung als auch den durchgeführten qualitativen und quantitativen Vortest sowie die Revision der Skala beschreiben. Zudem werde ich die ersten Befunde zur psychometrischen Güte der Skala (Itemstatistiken, Reliabilität, Validität) aufführen. Darüber hinaus stelle ich zum Abschluss dieses Kapitels weiterführende Analysen vor: eine Analyse zur Messinvarianz der neuen Skala über drei Jahrgänge von Studieneinsteigern hinweg sowie ein *Latent Growth Model*, anhand dessen unterschiedliche Entwicklungsverläufe in der Selbstwirksamkeit im Studieneinstieg bei Studierenden mit *versus* ohne Abbruchintentionen deutlich werden. Der besseren Lesbarkeit wegen werde ich im Folgenden das Akronym *SESW* statt *StudienEinstiegsSelbstWirksamkeit* verwenden.

3.1 Überblick: Skalen zur Erfassung der Selbstwirksamkeit

Zwar gibt es Skalen zur Erfassung der sogenannten „akademischen Selbstwirksamkeit“ (zum Beispiel von Chemers et al., 2001), allerdings ergab meine Recherche, dass es bis heute keine spezifische Skala zur Erfassung der Selbstwirksamkeit im Studieneinstieg gibt, die fachübergreifend anwendbar und auf die Population der Studienanfänger in Deutschland zugeschnitten ist. So sind die meisten

publizierten Skalen zur Erfassung von Selbstwirksamkeit im Studium (*nicht* im Studieneinstieg) für andere Bildungskontexte (andere Länder und Bildungssysteme) konzipiert und liegen oftmals auf Englisch vor. Aus psychodiagnostischer Sicht kann in diesem Fall keineswegs per se davon ausgegangen werden, dass diese Skalen ebenso für die Verwendung im deutschen Hochschulkontext geeignet sind. Darüber hinaus ist in der Regel durch eine einfache Übersetzung einer englischsprachigen Skala ins Deutsche nicht gesichert, dass das deutschsprachige Instrument dann äquivalent zum englischsprachigen ist (International Test Commission, 2016). Beispielsweise muss die kulturelle Äquivalenz erst untersucht werden und kann nicht per se als gegeben angesehen werden (Rammstedt, Harkness & Mohler, 2011).

Eine Ausnahme bildet die Skala von Peiffer, Preckel und Ellwart (2018), welche auf Deutsch vorliegt. Sie bezieht sich allerdings explizit auf Psychologiestudierende (es werden konkrete Fächer innerhalb der Psychologie thematisiert), weshalb auch diese Skala für die fächerübergreifenden Erhebungen im Rahmen des vorliegenden Promotionsprojektes als nicht geeignet erschien.

Daher entwickelte ich eine neue Skala, die diese „Lücke“ füllen soll: Eine Skala zur Erfassung von Selbstwirksamkeit im Studieneinstieg (SESW), zugeschnitten auf den Hochschulkontext in Deutschland. Dabei folgte ich der vielfach zu findenden Empfehlung, Selbstwirksamkeit domänenspezifisch zu messen (Betz & Hackett, 2006; Bandura, 1977). Das bedeutet, dass ich zunächst die typischen Herausforderungen im Studieneinstieg identifizieren musste, um die Skala so zu gestalten, dass sie das Maß an Zutrauen in die eigene Bewältigungskompetenz hinsichtlich eben dieser Herausforderungen erfasst.

Als Kurzform des Skalennamens *Skala zur Erfassung der studieneinstiegsspezifischen Selbstwirksamkeit* möchte ich im Folgenden den Begriff *SESW-Skala* verwenden.

3.2 Itementwicklung

Die Skalenentwicklung erfolgte in einem zweistufigen Vorgehen, wie es auch in gängigen Testkonstruktionslehrbüchern empfohlen wird (z.B. Eid und Schmidt, 2014). Im ersten Schritt entwickelte ich Items, die die identifizierten Herausforderungen im Studieneinstieg abbilden.

Im zweiten Schritt unterzog ich diese erste Version der Skala einem quantitativen und einem qualitativen Vortest. Ziel der quantitativen Vortests war es, die psychometrische Qualität der Skala zu untersuchen. Der qualitative Vortest sollte darüber hinaus Informationen liefern, inwiefern alle wichtigen Herausforderungen im Studieneinstieg von der Skala abgedeckt werden und ob die Befragten die Items so verstehen, wie sie gemeint sind. Die so gewonnenen Informationen habe ich dann zur Revision des ersten Entwurfs der Skala verwendet.

3.2.1 Schritt 1: Akademische Herausforderungen im Studieneinstieg identifizieren

Da die Zielgruppe meiner Untersuchungen die Population der Studienanfänger in Deutschland ist, liegt es nahe, explizit die Herausforderungen für Studienanfänger im deutschen Hochschulkontext als Ausgangsbasis für die Itemgenerierung zu verstehen.

Um ein möglichst umfassendes Bild der akademischen Herausforderungen, denen die Studienanfänger gegenüber stehen, zu erhalten, wurden drei Quellen herangezogen. Das waren (i) die Studierendenbefragung an der Justus-Liebig-Universität Gießen im Jahr 2016 (Ehrlich, 2016) und (ii) der zwölfte Studierendensurvey des BMBF (Ramm, Multrus, Barge & Schmidt, 2014). Beide Studierendenbefragungen zeichnen sich dadurch aus, dass sie einen großen Stichprobenumfang aufweisen ($N = 5.263$ Studierende und $N = 647$ Studienanfänger) und dass über alle Fächergruppen hinweg befragt wurde. Zusätzlich wurde (iii) noch die Sicht der Lehrenden berücksichtigt, in dem fachübergreifenden Erkenntnisse aus der CHE-Professorenbefragung (Horstmann & Hachmeister, 2016) ebenfalls einbezogen wurden.

In der Studierendenbefragung an der Justus-Liebig-Universität Gießen (Ehrlich, 2016) wurde Studierenden aller Semester (neben anderen Themenbereichen) eine Liste von Kompetenzen (beispielsweise „Fähigkeit, wissenschaftliche Methoden anwenden zu können“) vorgelegt und sie sollten sich selbst hinsichtlich dieser Kompetenzen einschätzen. Zudem wurde auch nach einer Selbsteinschätzung des Studienerfolgs gefragt („Ich bin mir sicher, dass ich das Studium erfolgreich bewältigen werde.“). Diejenigen Kompetenzen, welche fächerübergreifend einen signifikanten Zusammenhang mit diesem Erfolgskriterium aufwiesen, wurden als relevant für die hier beschriebene Skalenentwicklung betrachtet. Dabei können die *Kompetenzen oder Fähigkeiten*, die mit Studienerfolg assoziiert sind, durch einfache Umformulierung so ausgedrückt werden, dass sie eine Herausforderung beschreiben. So wird aus der oben exemplarisch genannten Kompetenz die „*Herausforderung, wissenschaftliche Methoden anwenden zu können*“.

Die so gesammelten Herausforderungen wiesen zwar einen signifikanten Zusammenhang mit dem selbstberichteten Studienerfolg auf, waren allerdings nicht spezifisch von Studienanfänger berichtet worden. Daher zog ich zusätzlich eine Befragung eben dieser speziellen Subgruppe heran. Im zwölften Studierendensurvey (Ramm et al., 2014) wurden unter anderem eben auch Studienanfänger gezielt zur Studieneingangsphase befragt. Bargel (2015) bereitete die Daten der Teilstichprobe der Studienanfänger auf. Er berichtet, dass diese Teilstichprobe hinsichtlich diverser Merkmale (Verteilung auf Hochschulen versus Fachhochschulen, Alter, Verteilung auf die Geschlechter, Verteilung auf die Studienfächer) repräsentativ für die Population der Studienanfänger in Deutschland sei. In dieser Befragung wurde die Zielgruppe nach Schwierigkeiten im Studieneinstieg befragt. Dies kann hier als Synonym für Herausforderungen verstanden werden. Die drei großen Bereiche, in denen Schwierigkeiten berichtet wurden, waren: Leistungsanforderungen, Prüfungsvorbereitungen und Studienplanung

(S. IV und S. 26). Die von Bargel (2015) aufgelisteten erlebten Schwierigkeiten wurden ebenfalls als relevant für die Skalenentwicklung erachtet.

Als weitere Quelle bei der Sammlung der akademischen Herausforderungen im Studieneinstieg bezog ich die Sicht der Lehrenden ein. Dazu verwendete ich die im Rahmen der CHE-Professorenbefragung von Horstmann und Hachmeister (2016) gewonnenen Erkenntnisse. Sie gingen explizit der Frage nach „Welche Fähigkeiten und Voraussetzungen sollten Studierende je nach Studienfach mitbringen?“. Zwischen 2013 und 2015 wurden insgesamt circa 9.500 Professoren aus 32 verschiedenen Fächern befragt. Horstmann und Hachmeister (2016) berichten neben fachspezifischen Anforderungsprofilen auch solche Anforderungen, die fachübergreifend häufig genannt wurden: Abstraktes/ logisches/ analytisches Denkvermögen, selbstständiges/ selbstorganisiertes und diszipliniertes Lernen und Arbeiten, Selbstmanagement/ Bereitschaft zum Selbststudium sowie Lernbereitschaft/ Einsatz- und Leistungsbereitschaft (S.6). Dabei sind die genannten Anforderungen wieder direkt in Herausforderungen übersetzbar. Diese aus Sicht der Lehrenden fachübergreifend relevanten Aspekte wurde ebenfalls als relevant für die Skalenentwicklung betrachtet.

Somit lag schließlich aus diesen drei Quellen eine Liste an akademischen Herausforderungen vor. Tabelle 3.1 zeigt die aus den drei aufgeführten Quellen extrahierten Herausforderungen. Daraus habe ich insgesamt zehn Items entwickelt, welche zur Abfrage der Selbstwirksamkeit bei der Bewältigung dieser dienen soll. Aus abstrakten Begriffen wie „Zeitplanung“ entwickelte ich Items, die konkrete Handlungen aufführen, beispielsweise „die zur Verfügung stehende Zeit zur Prüfungsvorbereitung sinnvoll zum Lernen zu nutzen“.

Die aufgelisteten Herausforderungen zeigen hohe Überlappungen mit den Ergebnissen der qualitativen Interview-Studie von Trautwein und Bosse (2017), die 25 Studierende (11 Studienanfänger und 14 Fortgeschrittene) der Universität Hamburg zum Thema *Herausforderungen im Studium* befragten.

Den Empfehlungen von Urdan und Pajares (2006) folgend, konzipierte ich die SESW-Skala so, dass die oben genannten akademischen Herausforderungen aufgelistet sind und für jede auf einer fünfstufigen Likertskala angegeben werden soll, inwiefern man sich zutraut, diese zu bewältigen (1 = *gar nicht* bis 5 = *voll und ganz*). Ein Item (inklusive der vorgelagerten Instruktion) lautet beispielsweise: *Im Folgenden geht es um Ihren Studieneinstieg. Bitte geben Sie an, inwieweit Sie sich zutrauen, Ihren Stundenplan für die kommenden Semester selbstständig zusammenzustellen.*

Ein übergeordnetes Ziel bei der Skalenentwicklung war es, ein Screening-Instrument zu entwickeln. Das bedeutet, dass die Skala insbesondere im Bereich niedriger Selbstwirksamkeit gut differenzieren soll. Dies soll es ermöglichen, Studienanfänger mit „auffällig“ niedriger Selbstwirksamkeit zuverlässig zu identifizieren.

Tabelle 3.1: Sammlung der Herausforderungen im Studieneinstieg aus drei Quellen

Quelle	Herausforderungen
Ehrlich (2016)	Beherrschung eigenes Fach, Arbeiten unter Druck, effizient auf Ziel hinarbeiten, Organisationsfähigkeit, analytische Fähigkeiten, Anwendung wissenschaftl. Methoden, eigene Wissenslücken erkennen und schließen
Ramm, Multrus, Barge und Schmidt (2014)	hohe Leistungsanforderungen, Prüfungsvorbereitungen (organisieren), Zeitplanung, aktive Teilnahme an Diskussionen, Kontakt zu Mitstudierenden finden
Horstmann und Hachmeister (2016)	logisches Denken, Selbstmanagement, Disziplin beim Studieren, Durchhaltevermögen zeigen, Leistungsmotivation aufbringen

3.2.2 Schritt 2: Quantitativer und qualitativer Vortest

Quantitativer Vortest

Insgesamt nahmen $N = 148$ Studieneinsteiger im Fach Psychologie am quantitativen Vortest teil. Davon gaben 18.2% an, männlich zu sein, 80.4% gaben an weiblich zu sein und 1.4% machten keine Angabe zu ihrem Geschlecht. Im Mittel waren sie $M = 20.78$ Jahre alt ($SD = 2.92$ Jahre). Neben der ersten Version der SESW-Skala bearbeiteten sie auch die Skala zur Erfassung der *allgemeinen Selbstwirksamkeit* (Schwarzer & Jerusalem, 1999). Dies sollte Daten zur (konvergenten) Konstruktvalidierung liefern. Tatsächlich zeigte sich eine erwartungsgemäß positive und hohe Korrelation zwischen den Skalen ($r = .618$, $p = .001$). Die Itemschwierigkeiten waren relativ hoch ($p = .72$ bis $.79$) und die Itemtrennschärfen ($r_{tt} = .31$ bis $.49$) lagen im erwartungsgemäßen Bereich für ein Screening-Instrument. Darüber hinaus wies die erste Version der SESW-Skala eine befriedigende interne Konsistenz auf: Cronbachs $\alpha = .719$.

Qualitativer Vortest

Während die Erkenntnisse aus dem quantitativen Vortest folglich recht vielversprechend waren, deckten die Erkenntnisse aus dem qualitativen Vortest noch Verbesserungsbedarf auf. Der qualitative Vortest umfasste fünf strukturierte Interviews mit Studierenden. Hierbei kam die Technik des *kognitiven Interviews* (De Leeuw, Hox & Dillman, 2012) zum Einsatz; vornehmlich die *think-aloud-Technik*. Dabei wurden die Interviewpartner zunächst gebeten, die Skala zu bearbeiten und dabei alle Gedanken zu verbalisieren. Dies dient dazu festzustellen, ob beispielsweise einzelne Formulierungen missverständlich sind, beziehungsweise schwierig zu verstehen sind. Nach der Beantwortung aller Items wurde zudem explizit gefragt, ob irgendwelche Schwierigkeiten bei der Bearbeitung aufgefallen sind. Anschließend wurden die Interviewpartner gefragt, ob aus ihrer persönlichen Erfahrung noch weitere typische Herausforderungen im Studieneinstieg auftreten, die in dieser ersten Version der Skala noch nicht enthalten waren.

Wichtig war außerdem zu erfahren, wie die Formulierung „auch bei kleineren Rückschlägen im Studium nicht den Mut zu verlieren“ wirkt. Die Formulierung sollte einerseits thematisieren, dass man im Studieneinstieg auch Misserfolge erleben kann, sollte aber andererseits nicht „verschreckend“ wirken. Die Befragten äußerten, dass sie diese Formulierung als verständlich und unproblematisch ansehen würden.

Ich erhielt allerdings weitere wertvolle Hinweise, welche dazu führten, dass ich die Skala um weitere drei Items ergänzte. Als zusätzliche akademische Herausforderungen kam das *Erlernen fachspezifischer wissenschaftlicher Methoden*, das *Verstehen komplexer Zusammenhänge* und das *Lernen großer Mengen an neuem Stoff* dazu. Dies deckt sich mit den Befunden der drei zuvor genannten Quellen, allerdings war ich bei der Itementwicklung davon ausgegangen, dass das Item *...Aufgaben aus Ihrem Studienfach durch logisches Denken zu lösen* die kognitiven Herausforderungen bereits ausreichend abbildet. Da sich die zusätzlich von den Befragten genannten Aspekte in deren Augen noch nicht ausreichend differenziert in der Skala widerspiegeln, habe ich im Zuge der Revision diese drei Aspekte in Form von zusätzlichen Items aufgenommen, wodurch sich das relative Gewicht der Items, die kognitive Leistungen thematisieren, erhöht hat.

3.2.3 Die revidierte Version der SESW-Skala

Nachdem die oben genannte Aspekte zur Revision des ersten Entwurfs der SESW-Skala geführt haben, lag nun die (vorerst) finale Version der Skala vor. Die Skala ist dabei so aufgebaut, dass einleitend der Kontext (Studieneinstieg) genannt wird und allen Items gemeinsam der Satzanfang „Bitte geben Sie an, inwieweit Sie sich zutrauen,...“ vorausgeht. Die Skala umfasst nunmehr folgende 13 Items, die in Abbildung 3.1 dargestellt sind:

Im Folgenden geht es um Ihren Studieneinstieg. Bitte geben Sie an, inwieweit Sie sich zutrauen,...

	gar nicht	eher nicht	teils teils	etwas	voll und ganz
1. ...Ihren Stundenplan für die kommenden Semester selbstständig zusammenzustellen.	☐	☐	☐	☐	☐
2. ...die zur Verfügung stehende Zeit zur Prüfungsvorbereitung sinnvoll zum Lernen zu nutzen.	☐	☐	☐	☐	☐
3. ...auch unter Zeitdruck gute Studienleistungen zu erbringen.	☐	☐	☐	☐	☐
4. ...Ihre Prüfungsvorbereitungen selbstverantwortlich zu organisieren.	☐	☐	☐	☐	☐
5. ...eventuell auftretende Wissenslücken zu erkennen.	☐	☐	☐	☐	☐
6. ...eventuell auftretende Wissenslücken zu schließen.	☐	☐	☐	☐	☐
7. ...Fristen einzuhalten (zum Beispiel zur Abgabe von Hausarbeiten oder zur Anmeldung zu Prüfungen).	☐	☐	☐	☐	☐
8. ...auch bei kleineren Rückschlägen im Studium nicht den Mut zu verlieren.	☐	☐	☐	☐	☐
9. ...Ihr Vorwissen aus der Schule im Studium zu nutzen.	☐	☐	☐	☐	☐
10. ...Aufgaben aus Ihrem Studienfach durch logisches Denken zu lösen.	☐	☐	☐	☐	☐
11. ...die fachspezifischen Methoden in Ihrem Studienfach zu lernen.	☐	☐	☐	☐	☐
12. ...komplexe fachliche Zusammenhänge zu verstehen.	☐	☐	☐	☐	☐
13. ...große Mengen an Studieninhalten zu lernen.	☐	☐	☐	☐	☐

Abbildung 3.1: SESW-Skala

3.3 Untersuchungen zur psychometrischen Qualität

In diesem Abschnitt berichte ich empirische Befunde, die auf drei verschiedenen Stichproben beruhen. Generell wurden in alle berichteten Analysen nur Daten von Personen einbezogen, die jeweils alle 13 Items der SESW-Skala bearbeitet hatten. Dieses Ausschlusskriterium wurde angelegt, da es bei einer Item- und Skalenanalyse eines neuen Messinstruments nicht in Frage kommt, die Antworten auf die Items zu imputieren. Somit ergaben sich leicht reduzierte Stichprobenumfänge, was an der jeweiligen Stelle kenntlich gemacht wird. Die besagten drei Datensätze, die herangezogen wurden, sind:

1. Normierungsstichprobe: Die Normierungsstichprobe umfasst $n_1 = 1842$ Studienanfänger der Justus-Liebig-Universität Gießen, die an der Studieneingangsbefragung im Jahr 2018 teilnahmen. Im Mittel waren sie $M = 20.8$ Jahre ($SD = 3.4$ Jahre) alt. 29% gaben an, männlich zu

sein, 64% weiblich, und weniger als 1% ordneten sich der Geschlechtskategorie „inter“ zu, die restlichen Teilnehmer machten keine Angaben zu ihrem Geschlecht. Verglichen mit den Daten der Zielpopulation (Statistisches Bundesamt, 2018) sind die weiblichen Studierenden in dieser Stichprobe somit überrepräsentiert.

Der Übersichtlichkeit wegen waren in der Studieneingangsbefragung die Studienfächer in Fachcluster aufgeteilt. In dieser Normierungsstichprobe fielen auf die verschiedenen Fachcluster folgende Stichprobenanteile ab: 19.8% Geisteswissenschaften, 28.4% Naturwissenschaften, 11.8% Medizin, 24.2% Lehramtstudiengänge, 9.5% Recht und Wirtschaft und 5% Psychologie. Sieht man davon ab, dass (i) an der Justus-Liebig-Universität Gießen kein Studium der Ingenieurwissenschaften angeboten wird, weshalb keine Studierenden der Ingenieurwissenschaften in der Stichprobe enthalten sind und (ii) dass die Gruppe der Agrarwissenschaftenstudierenden dagegen leicht überrepräsentiert ist, ist die Normierungsstichprobe immerhin in gewissem Maße mit der beim Statistisches Bundesamt (2018) verzeichneten Verteilung der Studienanfänger auf die Fächer vergleichbar. Somit liegt mit dieser Normierungsstichprobe zwar keine gänzlich für die Population der Studienanfänger in Deutschland repräsentative Stichprobe vor, jedoch kann davon ausgegangen werden, dass aufgrund der großen Fächerheterogenität nicht mit bedeutsamen Verzerrungen aufgrund einer starken Überrepräsentierung einzelner Fächer bzw. Fachkulturen zu rechnen ist.

Aus diesem Datensatz mussten 276 Fälle (13.03%) wegen einzelner nicht beantworteter SESW-Items für die Analysen ausgeschlossen werden, woraufhin der oben genannte Umfang der Normierungsstichprobe von $N = 1842$ Personen resultierte. In Anbetracht des Erhebungskontextes (freiwillige Teilnahme, keine Vergütung) und der Tatsache, dass die SESW-Skala weit hinten in der Umfrage angeordnet war, ist dieser Anteil als akzeptabel einzustufen.

2. Längsschnittliche Validierungsstichprobe: Diese Stichprobe besteht insgesamt aus $n_{2Start} = 1087$ Studienanfängern verschiedener Universitäten in Deutschland. Es handelt sich um den aggregierten Datensatz aus den drei Längsschnittstudien (Studienanfänger der Wintersemester 16/17, 17/18 und 18/19), die in den folgenden Kapiteln ausführlich dargestellt werden (Kapitel 5, 7 und 8). Wie weiter unten im Detail gezeigt (siehe 9.3.1) erwies sich die SESW-Skala als messinvariant über diese drei Jahrgänge, was die Aggregation der Daten legitimiert.

In dieser Gesamtstichprobe gaben 19.4% männlich zu sein, 78.3% gaben an, weiblich zu sein, die restlichen Teilnehmer machten keine Angabe zu ihrem Geschlecht. Das Alter lag im Mittel bei $M_{Alter} = 20.79$ Jahren ($SD_{Alter} = 4.01$ Jahre).

Wie in Longitudinalstudien nicht unüblich (Gold & Souvignier, 2005) ist auch in diesen Studien über die Laufzeitdauer von jeweils neun Monaten eine Panelmortalität von ca. 50% zu verzeichnen (circa 40% Panelmortalität nach sechs Monaten). Für die Analyse der prognostischen Validität der Skala (Vorhersage von Abbruchintention zum Ende der Studieneingangsphase) konnten somit nur diejenigen herangezogen werden, die auch beim jeweils dritten Messzeitpunkt noch teilnahmen. Daraus resultiert nach der Bereinigung der Datensätze (siehe Kapitel

5, 7 und 8) ein reduzierter Stichprobenumfang für diese Analysen von $n_{2final} = 424$ (17.7% männlich, 80.4% weiblich, $M_{Alter} = 20.85$ Jahre, $SD_{Alter} = 3.84$ Jahre).

Hinsichtlich der Verteilung auf die Studienfächer ist diese Stichprobe sehr heterogen, aber ebenfalls nicht gänzlich repräsentativ für die Population der Studienanfänger in Deutschland: 21.3% Naturwissenschaften, 24.8% Medizin oder Psychologie, 13.7% Sprachen/Literatur/Kultur, 22.2% Erziehungswissenschaften, 18.0% Wirtschaft, Recht oder sonstige Fächer.

Für die spätere Interpretation der Ergebnisse ist Folgendes zu berücksichtigen: Diejenigen, die bis inklusive des letzten Messzeitpunktes an den Studien teilnahmen, sprich deren Daten in die Analysen der prädiktiven Validität einbezogen werden konnten, unterschieden sich in ihrer SESW signifikant von denen, die während der neunmonatigen Studienlaufzeit die Teilnahme abbrachen. Diejenigen, die nicht an der kompletten Studie teilnahmen, berichteten im Mittel eine niedrigere SESW: $t(1078) = -.298$, $p = .036$.

Aus diesem Datensatz mussten keine Fälle wegen nicht beantworteter SESW-Items für die Analysen ausgeschlossen werden. Dies mag daher rühren, dass die Datenerhebung online durchgeführt wurde und ein Hinweis angezeigt wurde, falls jemand (aus Versehen) ein Item nicht beantwortet hatte. Die Person wurde dann darauf hingewiesen, dass es wünschenswert ist, dass alle Items beantwortet werden (technisch „erzwungen“ wurde die Antwort jedoch nicht).

3. Querschnittliche Validierungsstudie: An dieser Validierungsstichprobe nahmen insgesamt $n_3 = 148$ Personen teil. 28 davon (23%) mussten für die Analysen ausgeschlossen werden, weil sie nicht alle 13 SESW-Items beantwortet hatten. Die verbleibende Stichprobe umfasst $n_3 = 120$ Studienanfänger der Fächer Chemie/Lebensmittelchemie (72.5%) und Psychologie. 40.8% gaben an, weiblich zu sein, 41.7% männlich und 17.5% machten keine Angabe hinsichtlich ihres Geschlechts. Aus Gründen des Datenschutzes wurde das Alter in Kategorien abgefragt: weniger als 2% waren jünger als 18 Jahre, 51.7% waren zwischen 18 und 19 Jahre alt, 17.5% zwischen 20 und 21 Jahre alt, 6.7% waren zwischen 22 und 23 Jahre alt, 2.5% zwischen 24 und 25 Jahre alt und 5% waren älter als 25 Jahre. Die restlichen 15% der Teilnehmer machten keine Angaben zu ihrem Alter.

3.3.1 Präanalysen

Zunächst wurde die Verteilung des SESW-Skalenwerts in der Normierungsstichprobe auf Normalität geprüft: Sowohl der durchgeführte Kolmogorov-Smirnov-Test als auch der Shapiro-Wilk-Test wurden signifikant: $D(1842) = .989$, $p < .01$ und $W(1842) = .979$, $p < .01$. Das bedeutet, dass keine Normalverteilung vorliegt. Schaut man sich Abbildung 3.2 an, so wird auch durch visuelle Inspektion des Histogramms deutlich, dass der Mittelwert der Verteilung oberhalb der Mitte der Antwortskala (1 bis 5) liegt: $M = 4.02$ ($SD = 0.55$). Zudem ist die Verteilung linksschief und rechtssteil. Man kann hier von einem *Deckeneffekt* sprechen. Als Konsequenz dieser Verteilungsform mussten für die folgenden Analysen –sofern möglich– nicht-parametrische Verfahren verwendet werden. Konkret

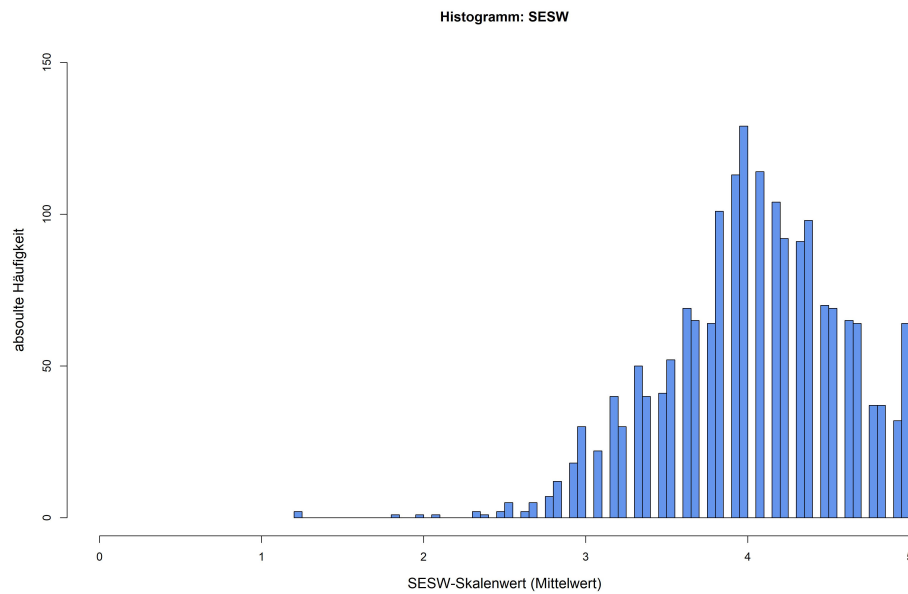


Abbildung 3.2: Verteilung des SESW-Skalenwerts in der Normstichprobe ($N = 1842$)

bedeutet das in erster Linie, dass die Spearman- statt Pearson-Korrelationen berichtet werden. Die Daten dieser Normierungsstichprobe wurden für die Analysen sämtlicher im Folgenden berichten Itemstatistiken herangezogen. Zur Berechnung der Reliabilitäts- und Validitätskoeffizienten wurden sie teilweise ebenfalls verwendet, teilweise aber auch die Daten aus den beiden anderen aufgeführten Stichproben. Es wird für jede der berichteten Analysen jeweils angegeben, auf welcher Stichprobe bzw. welchem Stichprobenumfang sie basiert.

3.3.2 Itemanalyse nach KTT

In diesem Abschnitt werden die Itemstatistiken aus der Perspektive der *klassischen Testtheorie (KTT)* dargestellt. Es handelt sich hierbei um die Itemschwierigkeit und die Itemtrennschärfe. Tabelle 3.2 zeigt für jedes Item den Mittelwert, die Standardabweichung sowie die beiden genannten Itemkennwerte. Weiterhin ist angegeben, wie die interne Konsistenz (gemessen als Cronbachs α) ausfiele, würde das jeweilige Item weggelassen werden. Die Itemschwierigkeiten liegen etwas über dem im Allgemeinen empfohlenen Bereich um .50 (Schmidt-Atzert & Amelang, 2012). Bedenkt man allerdings, dass (i) die SESW-Skala kein Leistungstest, sondern im weitesten Sinne eher ein Persönlichkeitsfragebogen ist und (ii) dass die SESW-Skala als Screening-Instrument insbesondere im Bereich niedriger Ausprägung der SESW gut differenzieren können soll, können die beobachteten Itemschwierigkeiten als zufriedenstellend eingestuft werden (Moosbrugger & Kelava, 2012). Besser wäre es allerdings, wenn es zumindest auch ein Item in der Skala gäbe, welches eine geringere Popularität aufweist und somit auch im Bereich höherer Ausprägung der SESW besser differenzieren könnte.

Tabelle 3.2: Itemstatistiken gemäß KTT für die SESW-Skala ($N = 1842$)

Item	M	SD	Schwierigkeit	Trennschärfe	α (ohne Item)
SESW 01	4.08	0.95	0.82	0.36	0.89
SESW 02	3.97	0.82	0.79	0.56	0.87
SESW 03	3.8	0.85	0.76	0.61	0.87
SESW 04	4.1	0.82	0.82	0.63	0.87
SESW 05	3.99	0.82	0.80	0.59	0.87
SESW 06	3.98	0.79	0.80	0.67	0.87
SESW 07	4.37	0.8	0.87	0.48	0.88
SESW 08	3.89	0.94	0.78	0.57	0.87
SESW 09	3.92	1.03	0.78	0.45	0.88
SESW10	4.11	0.79	0.82	0.60	0.87
SESW 11	4.19	0.73	0.84	0.65	0.87
SESW 12	4.03	0.78	0.81	0.68	0.87
SESW 13	3.88	0.86	0.78	0.63	0.87

Für alle Items außer Item 01 liegt die Trennschärfe im empfohlen Bereich [.40; .70] (Moosbrugger & Kelava, 2012). Fast alle Items weisen sogar eine Trennschärfe über .50 auf, was im Allgemeinen als *gut* eingestuft wird.

Insgesamt zeigt sich zudem das, was auf Skaleneben bereits oben beschrieben ist: Die befragten Studienanfänger tendierten zu relativ hohen Zustimmungsraten bei allen Items.

3.3.3 Itemanalyse nach IRT

In Ergänzung zur Darstellung der Itemstatistiken aus der Perspektive der KTT, sollen an dieser Stellen nun die entsprechenden Itemkennwerte aus der Perspektive der *Item Response Theory* (kurz: *IRT*) dargelegt werden. Für deren Berechnung wurde das *Graded Response Model* (*GRM*) nach Samejima (1969) zu Grunde gelegt, da dieses explizit für polytome Items mit geordneten Antwortkategorien postuliert wurde und bei der vorliegenden Anzahl an Items sowie den vorliegenden Stichprobenumfängen zuverlässig angewendet werden kann (Lubbe & Schuster, 2019). Tabelle 3.3 zeigt die Itemdiskrimination α_i sowie die vier Schwellenparameter β_{ik} ($k=1$ bis 4 bei fünf Antwortkategorien) für alle 13 Items.

Passend zu den Ergebnissen der Analysen nach KTT, kann auch aus Perspektive der IRT für alle Items außer Item 01 ($\alpha_1 = .792$) festgehalten werden, dass der Steigungsparameter α im gewünsch-

Tabelle 3.3: Itemstatistiken gemäß IRT für die 13 Items der SESW-Skala ($N = 1842$)

Item	α_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}
SESW 01	0.792	-5.523	-3.720	-1.601	0.642
SESW 02	1.516	-4.290	-2.750	-0.882	0.891
SESW 03	1.792	-3.580	-2.173	-0.508	1.119
SESW 04	1.871	-4.477	-2.457	-0.968	0.51
SESW 05	1.719	-4.181	-2.415	-0.899	0.801
SESW 06	2.329	-3.578	-2.321	-0.774	0.761
SESW 07	1.236	-5.252	-3.339	-1.791	-0.125
SESW 08	1.560	-3.618	-2.076	-0.692	0.807
SESW 09	1.145	-3.850	-2.271	-0.871	0.671
SESW 10	1.887	-3.982	-2.591	-1.085	0.572
SESW 11	2.307	-4.312	-2.739	-1.182	0.424
SESW 12	2.568	-3.382	-2.258	-0.875	0.686
SESW 13	2.050	-3.334	-2.010	-0.64	0.893

α_i = Itemdiskrimination, die β_{ik} stehen für den Punkt auf dem latenten Kontinuum, an dem für Item i der Cut-off-Wert zwischen Kategorie k und $k+1$ liegt.

ten Bereich von $[0.80; 2.50]$ (de Ayala, 2013) liegt.

Wie erwartet erwies sich für alle Items die Annahme der geordneten Kategorie als haltbar, was daran ersichtlich wird, dass die β_{i1} bis β_{i4} für alle Items in aufsteigender Reihenfolge geordnet sind.

Abbildung 3.3 zeigt die sogenannten *Item-Tracelines* für die Items, deren Inspektion weitere Details offenlegt. Item-Tracelines sind eine Darstellung der Kategoriewahrscheinlichkeiten, also der Wahrscheinlichkeiten, mit der eine bestimmte Antwortkategorie beim Vorliegen einer bestimmten Ausprägung der latenten Variable, angekreuzt wird. Hier ist für jede der fünf Antwortkategorien die jeweilige Kurve in einer anderen Farbe abgetragen. Betrachtet man die Kurven für jedes einzelne Item, so wird deutlich, dass für alle Items wie gewünscht alle fünf Kurven eingipfelig sind. Zudem überragt bei allen Items (außer bei Item 01) jede Kurve am Punkt ihres Maximums auch alle vier anderen Kurven. Das bedeutet, dass bei Item 02 bis Item 13 jede Antwortkategorie bei einer bestimmten Ausprägung der SESW am wahrscheinlichsten gewählt wird. Bei Item 01 hingegen ist die Kurve für die höchste Antwortkategorie (gelb) stark verbreitert: Bereits bei einer mittleren Ausprägung der latenten Variable SESW ist es am wahrscheinlichsten, dass Personen die höchste Antwortkategorie wählen. Dies ist also ein weiterer Beleg dafür, dass Item 01 eine hohe Schwierigkeit und eine niedrige Trennschärfe aufweist bzw. die Antworten zu Item 01 einen starken Deckeneffekt aufweisen. Relativ gesehen erweist sich dieses Item somit als das mit der niedrigsten psychometrischen Qualität innerhalb der SESW-Skala. Gleichwohl würde die Reliabilität der Skala nicht bedeutsam verbessert durch eine Entfernung des Items (vgl. Tabelle 3.2). Bei genauerer Inspektion von Tabelle 3.3, wird deutlich,

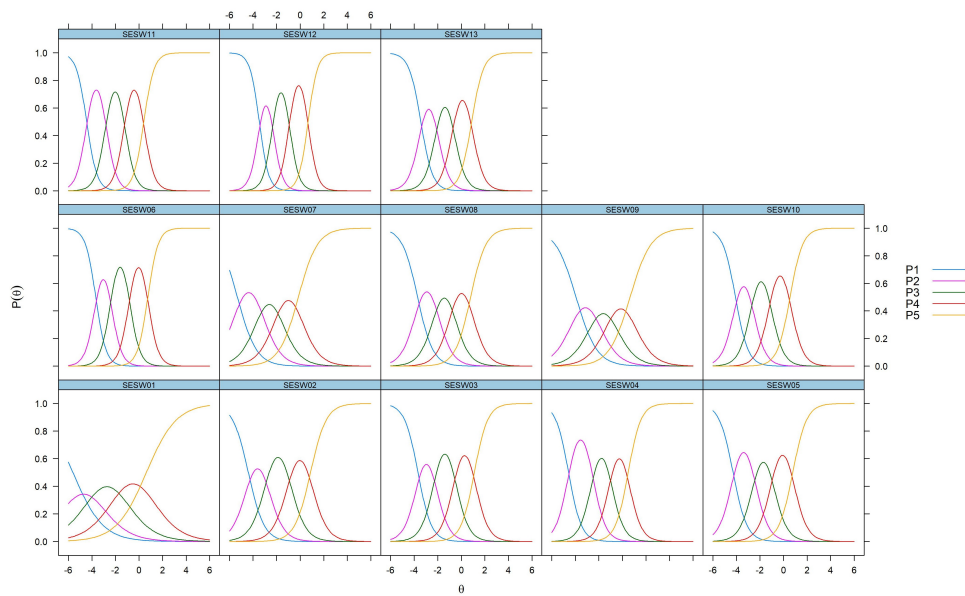


Abbildung 3.3: SESW: Tracelines

dass sich der oben beschriebene Deckeneffekt auch darin zeigt, dass für alle Items die Kennwerte β_{i1} bis β_{i3} negativ sind. Das bedeutet, dass sogar Personen mit relativ niedrig ausgeprägter SESW nicht nur in die untersten Antwortkategorien ankreuzten, sondern häufig die ersten vier Antwortkategorien genutzt haben. Anders ausgedrückt: Auch Personen mit vergleichsweise niedriger SESW trauen sich oftmals zumindest *teilweise* (Wortlaut der mittleren Antwortkategorie) zu, die in der SESW-Skala aufgeführten Herausforderungen zu bewältigen.

Weiterhin zeigt die Testinformationskurve (Abbildung 3.4), dass die SESW-Skala bei niedrigen Ausprägungen der latenten Variable SESW am besten differenziert, da hier die Maxima der Kurve liegen (links der Null auf der x-Achse).

Betrachtet man die Iteminformationskurven (Abbildung 3.5), so erkennt man ganz ähnliche Kurvenverläufe auch separat für fast alle Items. Auffällig ist hierbei jedoch, dass insbesondere Item 01 diese Kurve sehr flach verläuft: Entsprechend der bereits berichteten geringen Trennschärfe ist auch der Informationsgehalt quasi über die verschiedenen möglichen Ausprägungen der latenten Variable SESW hinweg (abgetragen auf der x-Achse in Abbildung 3.5) relativ gering.

3.3.4 Untersuchungen zur Reliabilität

Der erste Schritt zur Untersuchung der Reliabilität der SESW-Skala im Sinne der internen Konsistenz ist, dass ich zunächst auf die Korrelationen nullter Ordnung zwischen den Items eingehe, die Abbildung 3.6 und Tabelle 13.2 (im Anhang) entnommen werden können. Aufgrund der nicht-parametrischen

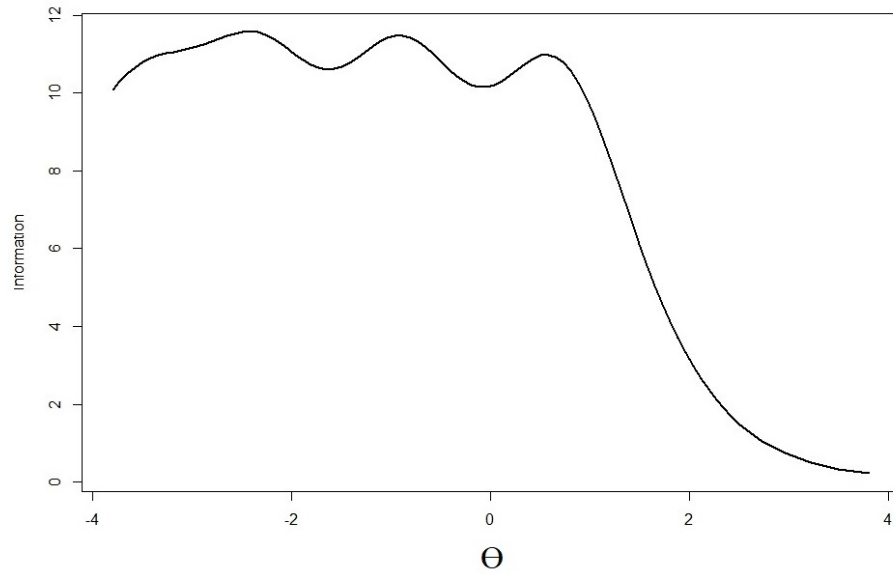


Abbildung 3.4: SESW: Testinformation

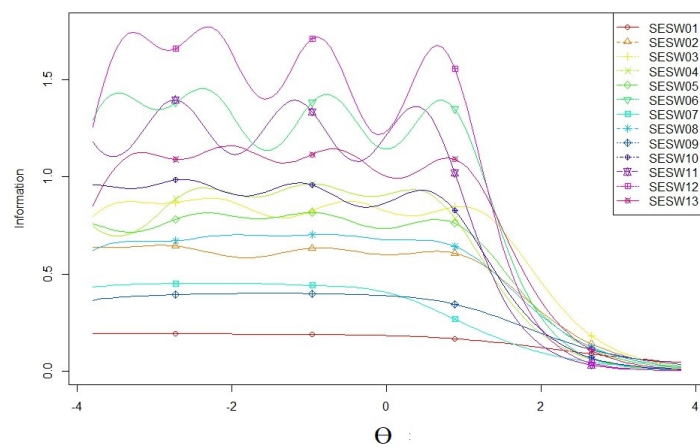


Abbildung 3.5: SESW: *Item Information Curves*

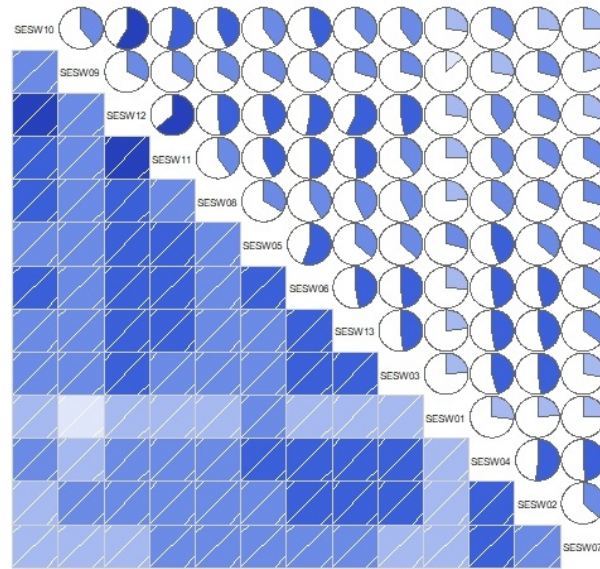


Abbildung 3.6: Korrelogramm zur SESW-Skala ($N = 1842$)

Verteilung der SESW-Skalenwerte wurden nicht-parametrisch Spearman-Korrelationen (r_s) verwendet.

Wie die Korrelationstabelle 13.2 zeigt, weisen die Items untereinander teilweise enge Zusammenhänge auf (basierend auf den Daten der Normierungsstichprobe). Dies zeigt auch die visuelle Inspektion des in Abbildung 3.6 dargestellten Korrelogramms. Hierbei wird die Höhe des Zusammenhangs farblich dargestellt: je dunkler das Blau, desto höher korrelieren die jeweiligen Items. Der Grad der Füllung des jeweiligen Kreises oberhalb der Diagonalen repräsentiert analog die Höhe des Zusammenhangs: Ein leerer Kreis stünde für $r = .00$, ein komplett gefüllter Kreis für $r = 1$ oder $r = -1$.

Interne Konsistenz

Vor diesem Hintergrund sind die Befunde zur internen Konsistenz wenig überraschend. Für die SESW-Skala ergab sich anhand der Daten der Normierungsstichprobe: Cronbachs $\alpha = .88$ ($KI_{95\%} [.88, .89]$). Das verwandte Maß McDonalds ω fiel sogar noch etwas höher aus ($\omega_{total} = .90$), wobei dieser Schätzer analog zu Cronbachs α interpretiert werden kann. Dies ist als hohe interne Konsistenz zu beurteilen.

Test-Retest-Reliabilität

Als zusätzliches Maß für die Reliabilität der SESW-Skala wurde die Test-Retest-Reliabilität berechnet. Gemäß der Definition von (kontextspezifischer) Selbstwirksamkeit als dynamisches Konstrukt (vgl. 2.1.6), welches der ständigen Veränderung durch neue Erfahrungen unterliegt, ist aus theoretischer Sicht ein Stabilitätskoeffizient mittlerer Höhe für die SESW-Skala zu erwarten. Genau das zeigte sich auch: Die Test-Retest-Reliabilität belief sich über die ersten vier Monate der Längsschnittstudien auf $r_{t1t2} = .476$ ($n = 589$, $p < .01$) und über die Gesamtlaufzeit der Längsschnittstudien (9 Monate) auf $r_{t1t3} = .346$ ($p < .01$, $n = 485$).

Somit konnte für die SESW-Skala eine sehr gute interne Konsistenz sowie eine erwartungsgemäß mittlere zeitliche Stabilität im Sinne der Test-Retest-Reliabilität festgestellt werden.

3.3.5 Untersuchungen zur Validität

Im Folgenden werden die Ergebnisse erster Untersuchungen zur Konstrukt- sowie Kriteriumsvalidität berichtet. Darüber hinaus werde ich auf die Augenscheinvalidität eingehen.

Konstruktvalidität

Wie bereits bei der Analyse der Reliabilität, wurden auch bei der Validitätsanalyse aufgrund der non-parametrischen Verteilung der SESW-Skalenwerte stets nicht-parametrisch Spearman-Korrelationen (r_s) berechnet.

Anhand der oben beschriebenen kumulierten Stichprobe des jeweils ersten Messzeitpunktes in allen drei Jahrgängen konnte einen signifikanter, positiver Zusammenhang zwischen der SESW-Skala und der Skala zur Erfassung der allgemeinen Selbstwirksamkeit (Schwarzer und Jerusalem, 1999, 10 Items, Cronbachs $\alpha = .848$) in Höhe von $r_s = .569$, $p < .001$, $N = 1079$) beobachtet werden.

Weiterhin erwies sich erwartungsgemäß der Zusammenhang zwischen dem SESW-Skalenwert und dem Testwert in einem kognitiven Fähigkeitstest als nicht bedeutsam ($r_s = .153$ $p = .096$). Diese Prüfung entspricht der Strategie der diskriminanten Konstruktvalidierung. Zum Einsatz kam bei der oben beschriebenen dritten Stichprobe der *Gießener kognitive Kompetenzentest (GkKT)* mit 66 Items (Ulfert, Ott, Bothe & Kersting, 2017; Petri, Weingardt & Kersting, 2019), der eine Reliabilität von Cronbachs $\alpha = .88$ ($N = 120$) aufwies.

Als weitere Untersuchung zur Konstruktvalidität der SESW-Skala sollte deren interne Struktur getestet werden. Folgende hierarchische Struktur möchte ich auf Basis der Iteminhalte postulieren: Auf der obersten Ebene findet sich ein Generalfaktor: *SESW*. Auf der darunter liegenden Ebene gibt es drei Faktoren erster Ordnung: *Fristen und Strategien (FuS)*, *Motivation und Kognition (MuK)* und *kognitive Herausforderungen (kH)*. Hierbei laden unterschiedliche viele Items (manifeste Variablen)

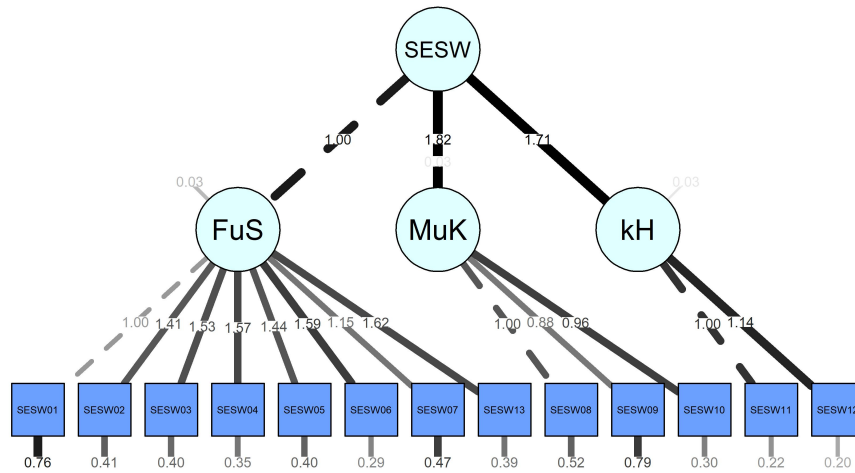


Abbildung 3.7: Hierarchisches Modell der SESW

auf den drei Faktoren. Auf dem Faktor *FuS* laden acht Items (Item 01 bis Item 07 und Item 13), darunter Item 03 „die zur Verfügung stehende Zeit zur Prüfungsvorbereitung sinnvoll zum Lernen zu nutzen“. Auf dem Faktor *MuK* laden drei Items (Item 08 bis Item 10), wie zum Beispiel Item 08 „auch bei kleineren Rückschlägen im Studium nicht den Mut zu verlieren“. Auf dem dritten Faktor *kH* laden die restlichen zwei Items (Item 11 und 12), beispielsweise Item 12 „komplexe fachliche Zusammenhänge zu verstehen“.

Empirisch getestet wurde dieses hierarchische Modell anhand der Normierungsstichprobe von $N = 1842$ Studienanfänger:innen. Abbildung 3.7 zeigt das Strukturgleichungsmodell, welches die hierarchische Struktur der SESW-Skala darstellt. Hierbei erwies sich das Modell als sehr gut geeignet zur Beschreibung der Daten aus dieser Stichprobe: SRMR = .041, CFI = .928, RMSEA = .075 [.070; .080]. Gemäß weithin verbreiteten Cut-off-Werten für diese Modellfitindizes (Hu & Bentler, 1998, 1999; Hooper, Coughlan & Mullen, 2008) kann also die postulierte hierarchische Struktur der Studieneinsteigsselbstwirksamkeit als belegt gelten.

Im Anwendungskontext ist die Verwendung des Generalfaktor (manifestiert im SESW-Skalenwert) sinnvoll. Für weitere faktoranalytische Untersuchungen der latenten Variable SESW sowie für die Einordnung des Konstrukts in ein nomologisches Netz inklusive der Abgrenzung gegenüber verwandter Konstrukte wie dem allgemeinen Selbstwert oder Ähnlichem sollten zwecks Gewinnung detaillierter Einblicke auch die drei Faktoren erster Ordnung berücksichtigt werden.

Kriteriumsvalidität

Da im Rahmen dieser Arbeit die Vorhersage der *Abbruchintention* eine große Rolle spielt, ist es wichtig zu prüfen, welchen Vorhersagebeitrag die SESW-Skala diesbezüglich leisten kann. Wie bereits in Kapitel 2 dargelegt, kann davon ausgegangen werden, dass die Intention, das Studium abzubrechen, mit der Selbstwirksamkeit im Studienkontext assoziiert ist.

Konkret wurde im Rahmen der drei Längsschnittstudien (vgl. Kapitel 5, 7, 8) am ersten Befragungszeitpunkt (neben diversen anderen Skalen) auch die SESW-Skala eingesetzt. Am dritten und letzten Befragungszeitpunkt wurden die Abbruchintentionen erhoben (siehe auch 4.1.1). Die Items dafür, welche an vorherige Studien angelehnt waren (Respondek et al., 2017; Treppesch & Ehrlich, 2018), lauteten: „Ich habe vor, das Studium zu einem erfolgreichen Abschluss zu bringen.“ und „Ich habe vor, das Studium abzubrechen.“. Auf einer siebenstufigen Likertskala sollte die Zustimmung angegeben werden (*stimme gar nicht zu* = 1 bis *stimme voll und ganz zu* = 7). Die Antworten auf das erste Item wurden invertiert und die Items dann durch Aufsummieren der beiden Antwortscores zu einem Skalenwert zusammengefasst.

Hierbei gingen etwas weniger Datensätze in die Berechnung ein, als für die Analyse der Test-Retest-Reliabilität herangezogen werden konnten. Dies resultiert daraus, dass zwar $n = 485$ Studienanfänger zum dritten Messzeitpunkt Angaben zur SESW machten, einige von ihnen aber keine Angaben zur Abbruchintention machten.

Der zum ersten Messzeitpunkt (Beginn erstes Semester) erfasste SESW-Skalenwerte wies einen bedeutsamen Zusammenhang mit der am Ende der Studieneinstiegsphase berichteten Abbruchintention auf: $r_s = -.163$, $p < .001$ ($N = 424$). Dies ist als Beleg für die prädiktive Validität der SESW-Skala zu werten.

Augenscheinvalidität

Kersting (2003) definiert: „Einem Verfahren wird eine hohe Augenscheinvalidität zugesprochen, wenn die subjektiv wahrgenommene Messintention mit der seitens der Testperson vermuteten oder ihr explizit genannten diagnostischen Fragestellung übereinstimmt, zu deren Beantwortung das betreffende Verfahren eingesetzt wird. [...] Die Augenscheinvalidität eines Verfahrens wird begünstigt, wenn das Verfahren Ähnlichkeiten mit „Real-life“-Situationen [...] aufweist, die in Zusammenhang mit der diagnostischen Fragestellung stehen.“ (S.54, Kersting, 2003).

Aufgrund der Einbeziehung der Sicht der Studierenden bei der Skalenentwicklung (siehe oben) ist davon auszugehen, dass die Skala eine gute Augenscheinvalidität aufweist. Dies ist selbstverständlich zukünftig in weiteren Studien zu prüfen, beispielsweise unter Einsatz des *Akzept-P/-*-Fragebogens (Beermann, Kersting, Stegt & Zimmerhofer, 2013), mit dem die Akzeptanz der Diagnostizierten gegenüber Persönlichkeitsfragebogen gemessen werden kann. Im *Akzept-P/-*-Fragebogen gibt es eine Subskala, die explizit die Augenscheinvalidität erfasst.

3.4 Normierung

Eine Normtabelle mit T-Werten zur SESW-Skala findet sich im Anhang 13.1. Sie basiert auf den im Wintersemester 2018 erhobenen Daten der oben beschriebenen Normierungsstichprobe ($N = 1842$ Studienanfänger). Ein Blick auf die Normtabelle unterstreicht, was oben bereits ausgeführt ist: Die SESW-Skala ist als Screening-Instrument konzipiert und kann daher im unteren Ausprägungsbereich der SESW am besten differenzieren, eignet sich aber nicht für eine Differenzierung im oberen Bereich.

3.5 Weiterführende Analysen

Abschließend möchte ich an dieser Stelle noch zwei weiterführende Analysen vorstellen: eine Analyse der Messinvarianz der SESW-Skala über die drei Jahrgänge hinweg sowie ein *Latent Growth Model*, anhand dessen unterschiedliche Entwicklungen in der Selbstwirksamkeit im Studieneinstieg für Studierende mit *versus* ohne Abbruchintentionen deutlich werden.

3.5.1 Messinvarianz über die Jahrgänge hinweg

Wie bereits oben erwähnt, wurden Daten aus den Längsschnittstudien verwendet, um die psychometrische Güte der SESW-Skala zu untersuchen. Das Vorliegen von Daten aus drei unabhängigen Jahrgängen ermöglicht auch die Exploration der Messinvarianz der SESW-Skala über die Jahrgänge hinweg.

Konkret wurde geprüft, ob das unter 3.7 dargestellte hierarchische Modell der SESW die Daten aus allen drei Jahrgängen gleichermaßen gut abbildet. Zunächst sind im Folgenden die Modellfitindizes pro Jahrgang aufgeführt:

- Jahrgang 1: SRMR = .057, CFI = .869, RMSEA = .080 [.063; .106]
- Jahrgang 2: SRMR = .074, CFI = .863, RMSEA = .076 [.049; .100]
- Jahrgang 3: SRMR = .071, CFI = .844, RMSEA = .095 [.760; .114]

Das Modell passt relativ gut zu den Daten aus allen drei Jahrgängen, wobei die Fitindizes rein nominell etwas hinter denen, die anhand der Normierungsstichprobe ermittelt wurden (vgl. 3.3.5), zurück bleiben. Neben dieser visuellen Inspektion der Fitindizes je Jahrgang, empfiehlt sich eine umfassende Invarianztestung, mit der die Frage beantwortet werden soll: Ist in allen drei Jahrgängen die gleiche (latente) Struktur der SESW zu finden und die SESW-Skala für alle drei (und damit ggf. auch für viele andere) Jahrgänge geeignet, um die SESW zu erfassen? Mit anderen Worten: Ist die Skala invariant über verschiedene Jahrgänge?

Nun mag es zunächst plausibel erscheinen, schlicht den Modellfit (beispielsweise anhand des χ^2 -Wertes) über die drei Jahrgänge hinweg zu vergleichen. Dafür würde sich der χ^2 -Differenztest eigentlich geradezu anbieten. Bei genauerem Betrachten allerdings wird klar, dass diese Vorgehen ungeeignet ist, zumal es sich hier um eine wiederholte Schätzung ein und desselben Modells mit verschiedenen Stichproben handelt, und nicht um verschiedene Modelle, die auch noch ineinander genested sind.

Die Formel des χ^2 -Differenztests, in deren Nenner die Differenz der Freiheitsgrade der zu vergleichenden zwei Modelle steht, kann hier folglich nicht zum Einsatz kommen, da das Modell in jedem Jahrgang mit gleich vielen Freiheitsgraden einher geht, spricht im Nenner des Terms eine Null stünde. Neben diesem rein praktischen Aspekt ist es weiterhin so, dass die Prüfung des Messinvarianz über verschiedene Gruppen (hier die drei Jahrgänge hinweg) weitaus differenzierter betrachtet werden muss.

Bevor ich auf das konkrete Ergebnis der hier durchgeführten Messinvarianztestungen für die SESW-Skala eingehe, möchte ich kurz das Konzept der Messinvarianztestungen erläutern. Dabei ist zwischen verschiedenen Stufen von Invarianz zu unterscheiden, die unterschiedlich strenge Annahmen umfassen. Zugrunde legen möchte ich die folgende im *Dorsch-Lexikon der Psychologie* zu findende Definition von Messvarianz (M.) gemäß Eid (2017):

M. ist gegeben, wenn ein Set von Items [...] in versch. Populationen und/oder unter versch. Bedingungen (z. B. Messzeitpunkten) dasselbe Konstrukt erfasst. M. ist notwendig, um Gruppenunterschiede (z. B. Unterschiede zw. versch. Kulturen) bzw. Veränderungen in den Verteilungskennwerten (z. B. Mittelwerten, Varianzen) oder indiv. Werten als Gruppenunterschiede in Konstruktausprägungen bzw. als Veränderungen in Konstruktausprägungen interpretieren zu können. [...] Je nach Skalenniveau der Items und der Konstrukte (latente Variablen) können zur Analyse der M. Modelle der Klassischen Testtheorie, Item-Response-Theorie (IRT)), Latenten Klassenanalyse oder der latenten Profilanalyse herangezogen werden. M. ist gegeben, wenn die Items in versch. Populationen bzw. unter versch. Bedingungen in derselben Weise mit dem Konstrukt (Variable, latente) verknüpft sind. I. R. der Faktorenanalyse unterscheidet man drei Arten von M.

- *Schwache M. liegt vor, wenn sich die Faktorladungen zw. Populationen bzw. Bedingungen nicht unterscheiden.*
- *Bei starker M. wird zusätzlich angenommen, dass auch die Achsenabschnitte gleich sind.*
- *Strikte M. setzt zusätzlich Gleichheit der Messfehlervarianzen voraus.*

Dabei ist anzumerken, dass die Begrifflichkeiten in der Literatur nicht ganz einheitlich sind. So werden oft *metrische* und *schwache* Messinvarianz sowie *skalare* und *starke* Messinvarianz äquivalent verwendet.

Vandenberg und Lance (2000) erläutern in ihrem Artikel, welche Bedeutung Messinvarianz für die

Generalisierbarkeit empirischer Ergebnisse hat und sprechen konkrete Empfehlungen für die Durchführung von Messinvarianzprüfungen aus.

Mit Hilfe des R-Pakets *semTools* von Jorgensen, Pornprasertmanit, Schoemann und Rosseel (2018) habe ich die Messinvarianz des neuen Modells über die drei Jahrgänge hinweg geprüft. Hierbei wird entsprechend ein schrittweises Vorgehen angewendet. Zunächst wird geprüft, ob das Modell (hier das hierarchische Modell der SESW) in allen drei Jahrgängen einen guten Fit aufweist. Dann werden in den nächsten Schritten jeweils immer mehr Restriktionen zur Modellspezifikation hinzugefügt, die – bei weiterhin guten Modellfit – die entsprechende Messinvarianzstufe indizieren. Das bedeutet, dass dann mit den unten ausgeführten zusätzlichen Restriktionen jeweils erneut der Modellfit beurteilt wird. Verschlechtert sich dieser nicht signifikant, so gilt die jeweilige Invarianzstufe als belegt. Dabei sind die Stufen aufeinander aufbauend, das bedeutet, sobald auf einer Stufe kein guter Fit mehr vorliegt, ist das auf der vorherigen Stufe bestätigte Messinvarianz-Niveau das maximal anzunehmende. Es werden dann keine weiteren (strikteren) Restriktionen mehr getestet.

Das schrittweise Vorgehen läuft im Detail nach folgenden Schema ab: Begonnen wird mit der einfachsten Variante, der konfiguralen Invarianz, bei der lediglich spezifiziert wird, dass der Modellfit über alle drei Jahrgänge hinweg gleich gut ist. Das bedeutet, im ersten Schritt wird im Fall der SESW-Skala getestet, ob die Modellstruktur als Ganzes zu den Daten des jeweiligen Jahrgangs passt. Methodisch geht es im Rahmen dieser Arbeit also um die Frage, ob für jeden Jahrgang die Faktorstruktur in den Daten durch das hierarchische Modell adäquat abgebildet werden kann.

Im nächsten Schritt wird dann spezifiziert, dass auch die Faktorladungen über alle Jahrgänge hinweg gleich sein sollen. Trifft das zu, so kann die schwache (also die metrische) Invarianz als gegeben angenommen werden.

Im dritten Schritt wird zudem spezifiziert, dass die *Intercepts*, also die Achsenabschnitte der manifesten Variablen, über alle Jahrgänge hinweg gleich sind. Passt das so restringierte Modell weiterhin gut, kann die starke (also skalare) Messinvarianz als gegeben angenommen werden.

Im nächsten Schritt wird weiter spezifiziert, dass ebenfalls die Fehlervarianzen über alle drei Jahrgänge hinweg gleich sein sollen. Würde der Modellfit weiterhin gut sein, so kann von strikter Messinvarianz ausgegangen werden.

Es besteht auch die Möglichkeit, im Sinne von „Zwischenstufen“ partielle Messinvarianz zu testen. Dazu werden – nach zunächst nicht mehr guten Modellfit, nach Hinzunahme weiterer Restriktionen – wieder für einige Parameter die Restriktionen gelockert. Erzielt man dann damit wieder einen guten Modellfit, so kann man von partieller Messinvarianz auf der jeweiligen Stufe (also beispielsweise partieller skalarer Messinvarianz) sprechen. Vandenberg und Lance (2000) warnen allerdings explizit davor, dass damit keinesfalls gemeint sei, dass man so lange verschiedene Lockerungen der Restriktionen probiert, bis der Modellfit wieder gut ist, sondern fundierte theoretische Begründungen angeführt werden müssten, welche Parameter im Detail von den Restriktionen ausgenommen werden. Da ich mich gegen das Testen auf partielle Messinvarianz entschieden habe, soll die kurze Skizzierung der partiellen Messinvarianz an dieser Stelle genügen.

Tabelle 3.4 zeigt die Modellfitindizes für das Strukturmodell der SESW-Skala für die verschiedenen, aufeinander folgenden Messinvarianzstufen. Hierbei wurde neben den bereits oben berichteten Fitindizes (CFI, SRMR und RMSEA) auch der korrigierte χ^2_{SB} -Wert herangezogen, der bei nicht-normalverteilten Daten statt dem unkorrigierten empfohlen wird (Satorra & Bentler, 2001, 2010).

Tabelle 3.4: SESW-Skala: Test auf Messinvarianz über die drei Jahrgänge hinweg

Invarianz-Stufe	χ^2_{SB}	df	χ^2_{SB}/df	CFI	RMSEA	SRMR
konfigurale	381.34	186	2.05	.848	.086	.067
metrisch	403.10	210	1.92	.850	.081	.078
skalar	416.075	228	1.82	.854	.076	.079
strikt	438.44	254	1.73	.856	.072	.083

Vergleicht man die Fitindizes für das Modell unter den Restriktionen der konfiguralen Invarianz (erste Zeile in Tabelle 3.4) mit denen für das Modell unter den Restriktionen für metrische Invarianz (zweite Zeile in Tabelle 3.4), so zeigt sich kein bedeutsamer Unterschied im Fit, was bedeutet, dass sowohl die erste als auch die zweite Stufe der Invarianztestung als erfolgreich absolviert verstanden werden kann. Die Differenzen liegen jeweils (bis auf eine Ausnahme im Vergleich der ersten mit der zweiten Zeile in Bezug auf den SRMR: $\Delta_{SRMR} = .011$) innerhalb gängiger Toleranzgrenzen (F. F. Chen, 2007) für nicht signifikante Änderungen im Modellfit: SOLL für $\Delta_{CFI} < -.010$, SOLL für $\Delta_{RMSEA} < .015$ und SOLL für $\Delta_{SRMR} < .010$. Vergleicht man weiter die zweite Zeile in Tabelle 3.4 mit der dritten, in der die Fitindizes für das Modell unter den Restriktionen skalarer (starker) Invarianz aufgeführt sind, so zeigt sich das gleiche Muster: keine signifikante Verschlechterung des Modellfits. Gleiches gilt für den Vergleich der dritten mit der vierten Zeile in Tabelle 3.4.

Daher kann angenommen werden, dass die SESW-Skala strikte Messinvarianz über die drei Jahrgänge hinweg aufweist. Das bedeutet, dass das hierarchische Modell über die drei Jahrgänge hinweg gleiche Faktorladungen und gleiche Achsenabschnitte für die manifesten Variablen (die beobachteten Items) aufweist. Zudem kann von äquivalenten Fehlervarianzen ausgegangen werden.

An dieser Stelle möchte ich explizit anmerken, dass sich (betrachtet man den CFI und den RMSEA) tatsächlich der Modellfit durch die Hinzunahme weiterer Restriktionen verbessert hat, wobei die Höhe des CFI insgesamt nicht im guten Bereich liegt, die Höhe des RMSEA und des SRMR jedoch innerhalb der gängigen Grenzen für befriedigenden bzw. guten Modellfit liegen.

Auf Basis der somit nachgewiesenen strikten Messinvarianz der Skala kann diese für den Einsatz in anderen Jahrgängen empfohlen werden, da davon auszugehen ist, dass ihr Nutzen zur Erfassung der SESW über Jahrgänge hinweg generalisierbar ist.

3.5.2 Entwicklung der SESW im Studieneinstieg

Es ist davon auszugehen, dass das, was mit der SESW-Skala erfasst werden kann, als Manifestation der Ausprägung eines latenten Konstrukts, nämlich der SESW, verstanden werden kann. Wie bereits dargestellt, unterliegt die Selbstwirksamkeit per Definition einem ständigen Wandel durch neue Erfahrungen. Im Abschnitt 3.3.4 habe ich bereits berichtet, dass die Skala erwartungsgemäß eine mittelhohe Test-Retest-Reliabilität aufweist. Für die Berechnung dieses Kennwertes wurden die Daten aus den drei Längsschnittstudien herangezogen (die oben beschriebene Stichprobe 2). Hierbei wurde noch nicht zwischen verschiedenen Personengruppen differenziert.

An dieser Stelle jedoch wollte ich der Frage nachgehen, ob sich Studienanfänger, die eine relativ hohe Ausprägung von Abbruchintentionen berichten, von den anderen Studienanfängern hinsichtlich ihrer SESW unterscheiden. Konkret sollte folgende Frage beantwortet werden: Weisen Personen mit vergleichsweise hoher Abbruchintention am Ende der Studieneinstiegsphase im Mittel einen anderen Verlauf in der Entwicklung der SESW über die drei Messzeitpunkte der Längsschnittstudien auf?

Um dies zu untersuchen, bildete ich zunächst zwei Gruppen anhand der berichteten Ausprägung der *Abbruchintention* (vgl. 3.3.5): Alle Studienanfänger, deren Skalenwert in Höhe der Mitte der Antwortskala oder höher lag, wurden der *Dropout-Gruppe* ($DG, n_{DG} = 77$) zugeordnet. Dies sind also die Studienanfänger, die zum Ende der Studieneinstiegsphase mindestens eine mittelstarke Zustimmung zu den Aussagen zur Studienabbruchintention angaben. Alle anderen Studienteilnehmer wurde der *Persistenz-Gruppe* ($PG, n_{PG} = 347$) zugeordnet.

Vergleicht man die Entwicklung des mittleren SESW-Skalenwerts der Studierenden, die in der Dropout-Gruppe sind, mit dem der Studierenden in der Persistenz-Gruppe, so liefert bereits die visuelle Inspektion der Boxplots in Abbildung 3.8 Indizien dafür, dass die Gruppen sich hierin unterscheiden. Weiterhin zeigt Abbildung 3.9 die individuelle Entwicklung der Studieneinsteigsselbstwirksamkeit über die drei Messzeitpunkte hinweg. Hierbei wird jeder Studienanfänger durch ein Symbol ($DG =$ Dreiecke, $PG =$ Kreise) im dreidimensionalen Koordinatensystem repräsentiert. Auf der x-Achse ist der Skalenwert zum ersten, auf der y-Achse der zum zweiten und auf der z-Achse zum dritten Messzeitpunkt angegeben. Der SESW-Skalenwert zum dritten Messzeitpunkt ist zusätzlich farblich kodiert: je höher, desto „wärmer“ (d.h. in den Gelb- bis Rotbereich gehend). Man erkennt in dieser Abbildung also einerseits, dass die Studienanfänger in der DG (Dreiecke) im Mittel eher in Blau- und Grüntönen dargestellt sind, wohingegen die Studienanfänger in der PG (Kreise) im Mittel eher in Gelb- und Rottönen symbolisiert werden und dass insbesondere zum zweiten und dritten Messzeitpunkt (y- und z-Achse) die Lage der jeweiligen Symbole deutliche Gruppenunterschiede zeigt. Zusammenfassend lässt sich also durch visuelle Inspektion bereits ein Unterschied zugunsten der PG feststellen. Die deskriptiven Statistiken für die SESW-Skalenwerte der beiden Gruppen über die drei Messzeitpunkte hinweg sind Tabelle 3.5 zu entnehmen. Ebenfalls dieser Tabelle zu entnehmen sind die Ergebnisse der *t*-Tests pro Messzeitpunkt. Die Gruppen unterscheiden sich auch statistisch signifikant hinsicht-

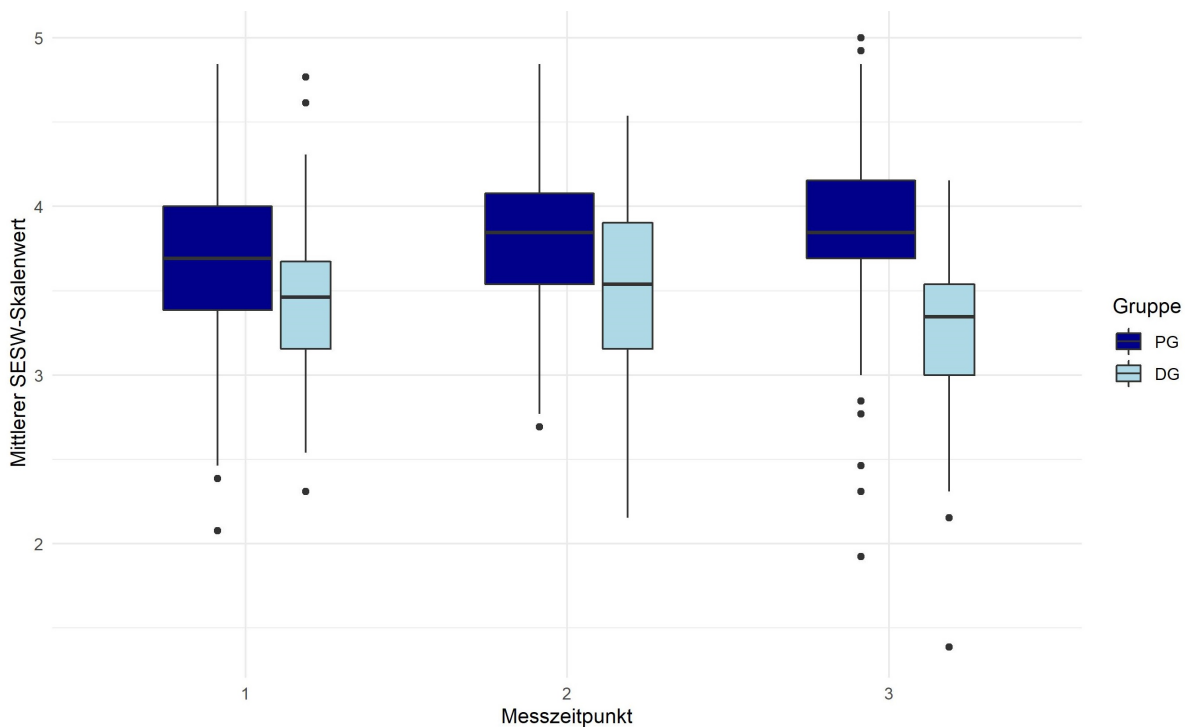


Abbildung 3.8: Verlauf der SESW im Gruppenvergleich (Persistenzgruppe (PG) = dunkelblau, Dropout-Gruppe (DG) = hellblau)

lich des mittleren SESW-Skalenwerts zu allen drei Messzeitpunkten. Zudem wächst die Differenz zwischen den mittleren SESW-Skalenwerten der Gruppen im Verlauf des Studieneinstiegs. Der deskriptive Vergleich sowie die t -Tests werden durch das Ergebnis der zweifaktoriellen ANOVA (within subject: Messwiederholungsfaktor für die SESW-Skalenwerte; between subjects: Gruppenzugehörigkeit), bestätigt: $F(1, 399) = 48.24, p < .001$.

Weiter zeigen die Tabellen 13.3 und 13.4 im Anhang die Korrelationen (nach Spearman) nullter Ordnung zwischen den SESW-Skalenwerten, die an den drei Messzeitpunkten erhoben wurden, separat für die beiden Gruppen. Obgleich das grundsätzliche Muster gleich ist, lässt sich festhalten, dass die Korrelationen in der DG durchweg nominell höher sind.

Um die Entwicklung der SESW über den Studieneinstieg hinweg weiter zu explorieren, lag es nahe, ein *Latent Growth Model* (Muthén & Curran, 1997) aufzusetzen. Damit sollte untersucht werden, wie die Wachstumskurven der SESW über die drei Messzeitpunkte hinweg gruppenspezifisch möglichst optimal abgebildet werden können.

McArdle, Hamagami, Elias und Robbins (1991) und Duncan und Duncan (2009) beschreiben die Idee hinter einer *multiple group*-Analyse. Diese habe ich aufgegriffen und werde im Folgenden den Transfer dieser Idee auf die Messwiederholungsdaten des SESW-Skalenwerts beschreiben.

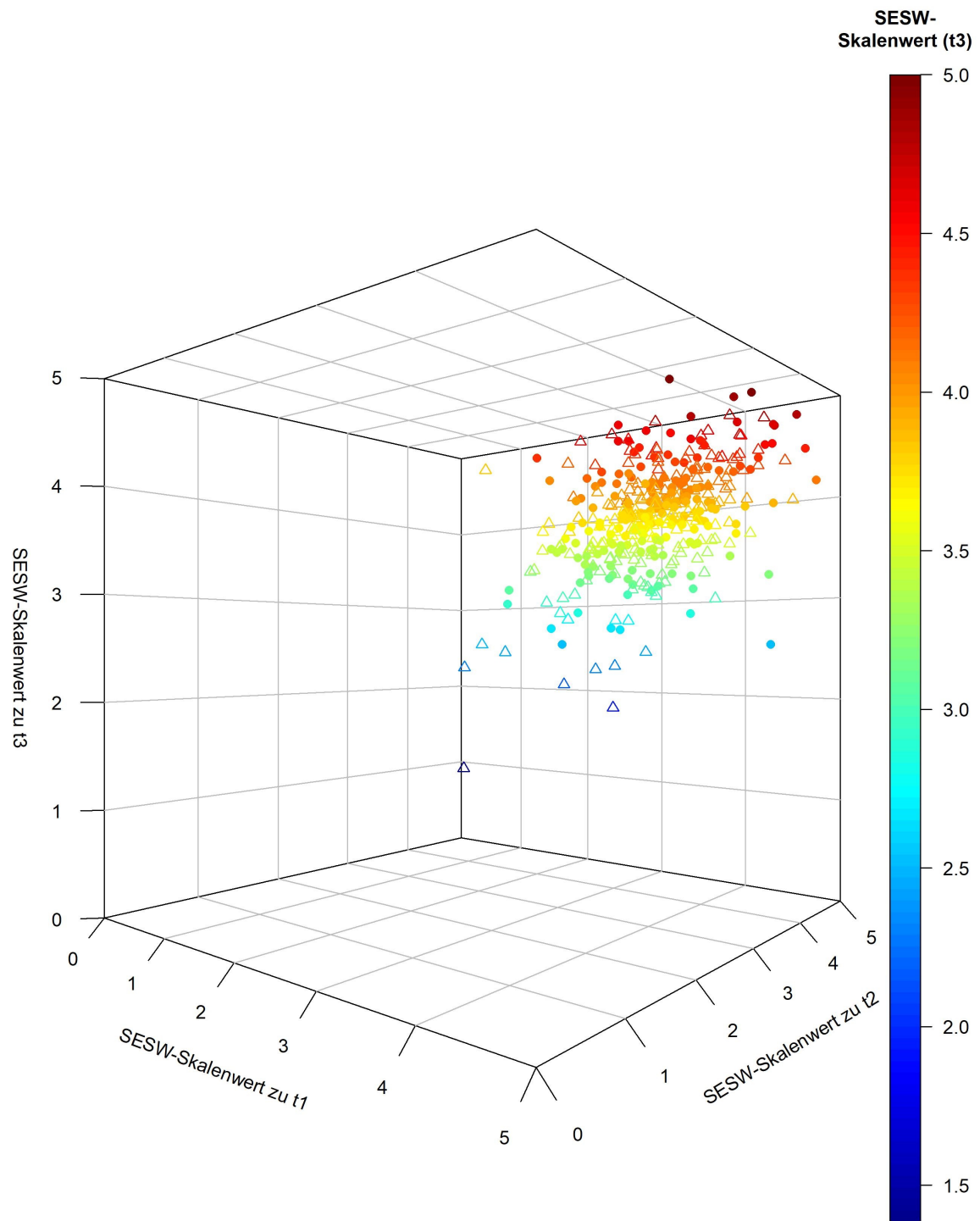


Abbildung 3.9: Entwicklung der SESW (DG = Dreiecke, PG = Kreise, $N = 424$)

Tabelle 3.5: SESW-Skala: Mittelwerte und Standardabweichung für die beiden Gruppen an den drei Messzeitpunkten (t1, t2, t3)

Messzeitpunkt	M (SD)	t -Test	
		Dropout-Gruppe ($n_{DG} = 77$)	Persistenz-Gruppe ($n_{PG} = 347$)
t1	3.44 (0.45)	3.64 (0.54)	$t(92.47) = 3.00$ ($p = .004$)
t2	3.58 (0.52)	3.79 (0.45)	$t(93.77) = 3.27$ ($p = .002$)
t3	3.33 (0.56)	3.88 (0.45)	$t(90.52) = 7.89$ ($p < .001$)

Latent Growth Model der SESW

Muthén und Curran (1997) schlagen vor, gruppenspezifische *Latent Growth Curves* zu modellieren, um Gruppenunterschiede auf latenter Ebene zu untersuchen. Im Fall der SESW-Skala bietet es sich an, zwischen DG und PG zu differenzieren. Die inhaltliche Begründung dafür ist, dass anzunehmen ist, dass die Studienanfänger, die am Ende des zweiten Semesters vergleichsweise starke Abbruchintentionen berichten (die DG), sich systematisch in ihrem Verlauf der Entwicklung der SESW von den Studienanfängern, die keine Abbruchintentionen berichten (die PG), unterscheiden.

Anzunehmen ist konkret, dass die DG im Mittel einen niedrigeren SESW-Skalenwert aufweist (was bereits im vorherigen Absatz 3.5.2 dargelegt wurde) und dass sich dieser im Mittel auch anders über den Verlauf des Studieneinstiegs hinweg entwickelt (im Sinne von einer abnehmenden SESW in der DG).

Isiordia und Ferrer (2018) plädieren in ihrem Artikel dafür, zur Modellierung von Wachstumskurven insbesondere im Bildungskontext nicht „bloß“ *First Order Latent Growth Models (1LGM)* zu verwenden, sondern *Curve of Factors Models (CUFF)* zu analysieren. Der Unterschied zwischen den 1LGM und den CUFF bestehe darin, dass bei letzteren eine weitere Faktorhierarchiestufe mit modelliert werde. So könne statt einem Summenscore oder einem Skalenmittelwert, der bei den 1LGM die beobachtete Variable darstelle, jedes einzelne Item in die Modellierung aufgenommen werden. Bei den CUFF ist folglich die unterste Ebene die der beobachteten einzelnen Itemantworten, die dann auf der mittleren Ebene auf den latenten Faktoren (statt Summenscores oder Mittelwerten) laden, welche wiederum dann auf der obersten Ebene auf den latenten Intercept- und Slope-Variablen laden. Das ermögliche die Trennung zwischen reliabler, zeitspezifischer Varianz und Residualvarianz. Genau das sei der Mehrwert der CUFF: eine akkuratere Schätzung der Wachstumskurve sei damit möglich (Isiordia & Ferrer, 2018). Voraussetzung für die Analyse eines CUFF ist, dass das Modell über die Messzeitpunkte hinweg mindestens starke (skalare) Messinvarianz aufweist.

Über die drei Jahrgänge hinweg erwies sich die SESW-Skala (bzw. das zugehörige Strukturmodell)

als stark messinvariant, wie im vorherigen Abschnitt beschrieben. Dies musste nun auch noch hinsichtlich der Messinvarianz über die drei Messzeitpunkte innerhalb der Längsschnittstudien gezeigt werden.

Da bereits in Abschnitt 3.5.1 ausführlich beschrieben wurde, wie die stufenweise Testung auf Messinvarianz erfolgt, möchte ich an dieser Stelle nur die Ergebnisse der Testung auf Messinvarianz über die drei Messzeitpunkte aufführen. Tabelle 3.6 zeigt die Modellfitindizes.

Tabelle 3.6: SESW-Skala: Test auf Messinvarianz über die drei Messzeitpunkte ($N = 1842$)

Invarianz-Stufe	χ^2_{SB}	df	χ^2_{SB}/df	CFI	RMSEA	SRMR
konfigural	656.186	186	3.53	.892	.079	.051
metrisch	703.084	210	3.35	.887	.076	.062
skalar	926.108	228	4.06	.840	.087	.075
strikt	997.800	254	3.93	.829	.086	.084

Wie schon beim Test auf Messinvarianz über die Jahrgänge hinweg, ist es auch beim Test auf Invarianz über die Messzeitpunkte hinweg so, dass der Modellfit (anders als anhand der Normierungsdaten beobachtet, siehe 3.3.5) insgesamt nur als gut bis befriedigend zu bewerten ist.

Vergleicht man wieder die Zeilen der Ergebnistabelle (hier: Tabelle 3.6) miteinander, so lässt sich festhalten, dass die Differenzen der Modellfitindizes nur bis zur Stufe der *metrischen Invarianz* jeweils innerhalb gängiger Toleranzgrenzen (F. F. Chen, 2007) für *nicht* signifikante Änderungen im Modellfit liegen: SOLL für $\Delta_{CFI} < -.010$, SOLL für $\Delta_{RMSEA} < .015$ und SOLL für $\Delta_{SRMR} < .010$, beziehungsweise das letztgenannte SOLL sogar knapp verfehlt wird.

Daher kann angenommen werden, dass die SESW-Skala metrische, nicht aber skalare Messinvarianz über die drei Messzeitpunkte hinweg aufweist. Das bedeutet, dass das hierarchische Modell über die drei Messzeitpunkte hinweg gleiche Faktorladungen für die manifesten Variablen (die beobachteten Items) aufweist. Für die Modellierung eines CUFF ist allerdings starke Messinvarianz eine wichtige Voraussetzung (Isiordia & Ferrer, 2018).

Ich entschied mich trotz der Überschreitung der oben genannten Toleranzgrenzen für die Differenz der Modellfitindizes dafür, auf rein explorativer Basis eine erste *Latent Growth Analysis* durchzuführen. Selbstverständlich kann aber beim Nicht-Vorliegen starker Messinvarianz nicht davon ausgegangen werden, dass ein solches Modell dann einen guten Fit aufweist.

Um die Komplexität des zu fittenden Modells zu reduzieren und zugleich die Idee der CUFFs aufzugreifen, entschied ich mich für eine Minimalvariante des Strukturmodells der SESW: Statt alle 13 Items in das Modell einzubeziehen, wie ich es in der oben dargestellten Analyse zur latenten Struktur getan habe (vgl. Abbildung 3.7), modellierte ich nur die mittlere Ebene und den übergeordneten latenten Faktor SESW.

Konkret spezifizierte ich für das *Latent Growth Model* einen Unterschied in den latenten Achsenabschnitten (latent Intercepts) zwischen der DG und der PG, ähnlich wie beispielsweise auch H. Chen,

Tabelle 3.7: SESW-Entwicklung im Gruppenvergleich: *Latent Growth Analysis* (standardisierte Koeffizienten)

Parameter	Dropout-Gruppe $n_{DG} = 71$	Persistenz-Gruppe $n_{PG} = 330$
First-Order Factor Loadings		
t1: SESW		
FuS1	1.493 (0.138)	1.745 (0.150)
MuK1	1.264 (0.193)	1.373 (0.121)
kH1	1.485 (0.193)	1.721 (0.147)
t2: SESW		
FuS2	1.589 (0.218)	1.745 (0.150)
MuK2	1.191 (0.086)	1.373 (0.121)
kH2	1.575 (0.265)	1.721 (0.147)
t3: SESW		
FuS3	1.604 (0.134)	1.745 (0.150)
MuK3	1.266 (0.176)	1.373 (0.121)
kH3	1.354 (0.155)	1.721 (0.147)
Latent Intercept Means		
t1	1.755 (0.132)	1.870 (0.135)
t2	1.843 (0.162)	1.951 (0.182)
t3	1.791 (0.146)	2.258 (0.236)

Dai und Gao (2019) Unterschiede in der Selbstwirksamkeit zwischen verschiedenen Schülergruppen untersuchten. Für den vorliegenden Untersuchungsanlass bedeutet diese Spezifikation, eine Testung der Annahme, dass der latente SESW-Faktor einen gruppenspezifischen Einfluss auf die mithilfe der SESW-Skala erfasste manifestierte SESW hat.

Der Unausbalanciertheit der Gruppengrößen habe ich durch eine Gleichgewichtung Rechnung getragen. Das mit der *Maximum-Likelihood*-Methode geschätzte Modell offenbarte dann die vermuteten Gruppenunterschiede in der latenten Intercept-Variable für jeden der drei Zeitpunkte (vgl. Tabelle 3.7). Im Anhang in Abbildung 13.1 findet sich zudem eine grafisch Darstellung der Ergebnisse.

Da sich die SESW-Skala als „nur“ metrisch invariant über die drei Messzeitpunkte hinweg erwiesen hat, sind diese Kennwerte nur als erste Anhaltspunkte zu verstehen. Ein Test auf Signifikanz der Differenz der Ladungen ist nur beim Vorliegen von mindestens starker, besser noch strikter Invarianz (Isiordia & Ferrer, 2018; Vandenberg & Lance, 2000) sinnvoll.

Die Ergebnisse dieser explorativen Analyse können zusammenfassend als Indiz dafür verstanden werden, dass dem signifikanten Gruppenunterschied in den manifesten Variablen (den drei SESW-

Skalenwerte von Beginn, Mitte und Ende der jeweiligen Längsschnittstudie) ein übergeordneter, latenter SESW-Faktor zugrunde liegt. Dieser beeinflusst die beobachtbare Höhe der SESW in den betrachteten Gruppen unterschiedlich: In der PG ist der Anstieg des latenten Intercepts nominell deutlich höher als in der DG. Man spricht in diesem Fall auch von einer *Latent Means Difference*.

Wünschenswert wäre, den Aspekt der gruppenspezifischen *Latent Growth Curves* weiter zu untersuchen. Bei noch größeren Stichprobenumfängen dürfte auch die statistische Power bei der Modellierung höher sein. Denkbar wäre auch, dann in Konkurrenz zum hier modellierten Ansatz mit einer Differenz im latenten Intercept, zusätzlich nach Unterschieden in latenten Slopes (Steigungskoeffizienten) zu suchen. Dann wären es –wörtlich genommen– auch tatsächlich *Latent Growth Curves*.

4 Empirische Studien

In diesem Kapitel möchte ich einen Überblick über alle fünf Studien, die ich im Rahmen meines Promotionsprojektes durchgeführt habe, geben. Zu Verbesserung der Nachvollziehbarkeit stelle ich im Folgenden die Studien nacheinander dar. Dabei beschreibe ich zunächst jeweils die Forschungsfrage, die mit der jeweiligen Studie beantwortet werden sollte. Anschließend skizziere ich das Studiendesign. Für die Längsschnittstudien beschreibe ich in diesem Kapitel einmal zentral alle verwendeten Instrumente.

Zudem werde ich auf einige studienübergreifende Aspekte eingehen. Insbesondere werde ich methodische Aspekte, die den drei Längsschnittstudien gemein sind, in diesem Kapitel einmal zentral erläutern, um Redundanzen in den Kapitel, die den einzelnen Studien gewidmet sind, zu vermeiden. Abschließend berichte ich, welche Software-Pakete für die Datenanalysen verwendet wurden.

4.1 Studie 1 bis 3: Die Längsschnittstudien

Insgesamt führte ich drei Längsschnittstudien durch. Die somit gewonnenen Daten sollten der empirischen Testung der in Kapitel 2.4.2 beschriebenen Modelle dienen. Abbildung 4.1 zeigt die Erhebungszeiträume der drei Längsschnittstudien.

Diese drei Studien, die alle nach dem selben Design durchgeführt wurden, liefen jeweils über die ersten beiden Semester (neun Monate) des jeweiligen untersuchten Jahrgangs. Somit liegen drei Längsschnitte von Stichproben aus den Populationen der Studienanfänger der Jahrgänge mit Studienbeginn zum Wintersemester 2016/2017¹, 2017/2018 und 2018/2019 vor.

¹Die Daten aus Studie 1 stellte ich für folgende studentische Qualifikationsarbeiten zur Verfügung, die wertvolle Erkenntnisse zum Zusammenhang von Selbstwirksamkeit, Stress, Emotionen und Zufriedenheit lieferten.

- Davood, A. (2017). Zum Zusammenhang von sozialer Unterstützung und Studienzufriedenheit im Studieneinstieg. Justus-Liebig-Universität Gießen: Unveröffentlichte Bachelor-Thesis.
- Elkenhans, L. (2017). Wie fühlen sich Studienanfänger? Eine Studie zu Emotionen von Studierenden im ersten Semester anhand der Kontroll-Wert-Theorie. Justus-Liebig-Universität Gießen: Unveröffentlichte Master-Thesis.
- Keil, J. (2017). Studieneinsteiger im Stress – Der Einfluss von Selbstwirksamkeitserwartungen und Copingstrategien auf das Stressempfinden von Studieneinsteigern. Justus-Liebig-Universität Gießen: Unveröffentlichte Master-Thesis.
- Krüger, L.M. (2017). Stress im Studieneinstieg: Eine Longitudinalstudie zum Einfluss von Selbstwirksamkeitserwartung und sozialer Unterstützung. Justus-Liebig-Universität Gießen: Unveröffentlichte Master-Thesis.

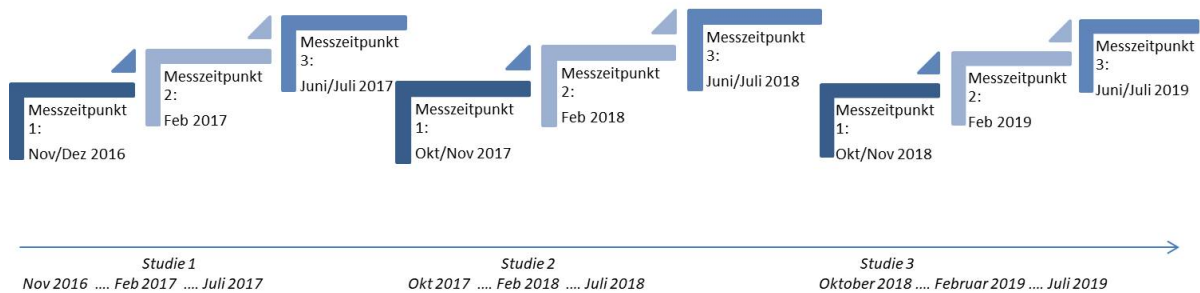


Abbildung 4.1: Überblick Längsschnittstudien

Jede der Studien diente zur Erfassung der in den Modellen beschriebenen Variablen, verteilt über drei Messzeitpunkte. Die Messzeitpunkte lagen jeweils zu Beginn des ersten Semesters (Oktober bzw. November/Dezember), Ende des ersten Semesters (Februar) und Ende des zweiten Semesters (Juni/Juli).

Die Datenerhebungen erfolgten stets in Form einer Online-Befragung unter Nutzung des frei zugänglichen Befragungstools *SoSciSurvey* (Leiner, 2016). Dazu waren die Befragungen immer jeweils ca. 30 Tage online geschaltet. Zur anfänglichen Rekrutierung nutzte ich drei verschiedene Kanäle: (i) Rundmails an der Justus-Liebig-Universität Gießen und anderen Universitäten in Deutschland, an denen ich die Möglichkeit hatte, Rundmails über die Uni-Mailverteiler zu versenden. (ii) An der Justus-Liebig-Universität Gießen habe ich zudem mit der Unterstützung einiger Studierender Flyer vor und nach einigen Vorlesungen der Erstsemesterstudierenden in diversen Fachbereichen verteilt und kleine Werbeplakate mit dem Aufruf zur Teilnahme an den jeweiligen Studien ausgehängt. (iii) Über das soziale Netzwerk *facebook* konnte ich jedes Jahr diversen *facebook*-Gruppen für Erstsemesterstudierende an verschiedenen Universitäten in Deutschland beitreten und einen Aufruf zur Teilnahme an den Untersuchungen inklusive Zugangslink posten. Über diese drei Kanäle habe ich jeweils eine Kurzbeschreibung der Studie sowie den Link zur Online-Befragung und teilweise auch einen QR-Code, der diesen Link enthielt, verbreitet.

Die Teilnahme an den online durchgeführten Längsschnittstudien erfolgte auf freiwilliger Basis. Als Anreiz für die Teilnahme wurde sowohl die Möglichkeit der Teilnahme an einer Verlosung von Amazon-Gutscheinen (und bei Jahrgang 1 auch eines Tablets) ausgelobt als auch auf das Ziel der Studien verwiesen: ein besseres Verständnis des Prozesses des Studieneinstiegs erhalten, um Studieneinsteiger besonders zielgerichtet unterstützen zu können. Dass das Ausloben von Gewinnen beziehungsweise der Teilnahme an einem Gewinnspiel die Teilnahmequoten in Online-Studien erhöhen kann, wurde bereits metaanalytisch belegt (Göritz, 2006). Insgesamt ist davon auszugehen, dass insbesondere Personen, die Interesse an der Verlosung hatten und oder das Ziel der Studien als wertvoll erachteten und daher diese Forschung unterstützen wollten, teilgenommen haben.

Diese Rekrutierungsmaßnahmen wurden für jeden Jahrgang nur im Oktober/November durchgeführt. Nach Beendigung des jeweils ersten Messzeitpunktes war somit festgelegt, wer zum jeweiligen Stichprobenumfang gehört. Die Teilnehmer wurden gebeten, ihre Emailadresse zu hinterlassen, falls sie an den beiden weiteren Messzeitpunkten teilnehmen wollten. Die so zusammengestellte Liste an Emailadressen entspricht dem Stichprobenrahmen aus dem jeweiligen Jahrgang.

Dass nur Studienanfänger an den Studien teilnahmen, wurde auf zwei Wegen sichergestellt. Zum einen rekrutierte ich sofern möglich explizit nur innerhalb der Zielpopulation. Zum anderen wurde auf der Startseite der Online-Befragung dann nochmals erläutert, dass nur Studienanfänger befragt werden sollen und wer angab, nicht im ersten Semester zu studieren, wurde mittels Filterfunktion aus der Online-Befragung rausgeleitet.

SoSciSurvey (Leiner, 2016) bietet die Möglichkeit einer pseudonymisierten Befragung, die ich nutzte: Ich hinterlegte die Emailadressen in einer Adressliste (Panel) und vom System wurde jeder Eintragung (resp. jeder Person) eine zehnstellige Seriennummer zugeordnet. Die Einladungen zum zweiten und dritten Messzeitpunkt wurden dann vom System zu den von mir gewählten Zeitpunkten versendet. Dabei erhielt jeder einen individuellen Zugangslink, anhand dessen die Umfragedaten aus dem zweiten und dritten Messzeitpunkt jeweils eindeutig mit den Angaben vom ersten Messzeitpunkt gematched werden konnten.

Nach einer initialen Einladung zum jeweils zweiten und dritten Messzeitpunkt, versendete ich bei Bedarf weitere Remindermails, falls innerhalb der ersten fünf bis sieben Tage nicht an der nächsten Befragung teilgenommen wurde (auch eine Funktion in *SoSciSurvey*: bisherige Non-Responder erneut anschreiben). Insbesondere ist zu betonen, dass selbst bei ausbleibender Teilnahme am zweiten Messzeitpunkt, Einladungen zur Teilnahme am dritten Messzeitpunkt an die betreffenden Personen gesendet wurden. Somit ist es möglich, dass Personen nur am ersten und am dritten Messzeitpunkt teilgenommen haben.

4.1.1 Verwendete Instrumente

Die drei hier vorgestellten Längsschnittstudien waren exakt gleich aufgebaut. Daher wurden auch die gleichen Instrumente verwendet. Diese werde ich im vorliegenden Abschnitt vorstellen und in den einzelnen Kapiteln zu den drei Längsschnittstudien dann darauf verzichten, die Beschreibung der Instrumente zu wiederholen.

Der Empfehlung zur guten wissenschaftlichen Praxis im Sinne von *Open Science* (Simmons, Nelson & Simonsohn, 2011) folgend, findet sich in Tabelle 4.1 eine vollständige Auflistung der verwendeten Itemgruppen aus den Längsschnittstudien. Dies soll der vollständig transparenten Darstellung meiner Datenerhebungen im Zuge der Längsschnittstudien dienen, indem ich auch die Itemgruppen aufgelistet habe, die nicht in Analysen im Rahmen meines Promotionsprojektes verwendet wurden. Für jede Itemgruppe ist jeweils angegeben, an welchen Messzeitpunkten sie eingesetzt wurde (markiert durch ein entsprechendes Kreuz in der jeweiligen Spalte). Die Variablen, die im Rahmen der Modelltests

Berücksichtigung fanden, sind mit einem * markiert.

Alle Daten wurden im Selbstbericht im Zuge der Online-Befragung gesammelt. In seltenen Fällen wurden die Variablen mit single-Items operationalisiert, meist aber mit Skalen (Fragebogen, bestehend aus mehreren Items) erfasst. Dabei kamen neben etablierten Fragebogen auch selbst entwickelte oder angepasste Skalen zum Einsatz. In der folgenden Auflistung ist für jede Variable, die im Zuge der Modelltests Verwendung fand, angegeben, wie ich sie operationalisiert habe. Dabei ist die Auflistung so gestaltet, dass zunächst die am ersten, dann die am zweiten und schlussendlich die am dritten Messzeitpunkt erhobenen Variablen aufgeführt sind. Variablen, die wiederholt erhoben wurden, sind entsprechend kenntlich gemacht. Falls es sich um bereits publizierte Instrumente handelte und psychometrische Kennwerte zu diesen vorliegen, sind diese aufgeführt. Welche Maße sich für die Reliabilität der verwendeten Skalen in meinen Untersuchungen ergaben, ist im Detail im jeweiligen Kapitel (5, 7, 8) zu den Längsschnittstudien aufgeführt.

Soziale Herkunft

Die soziale Herkunft von Studierenden, also die Bildungsnähe oder Bildungsferne ihrer Ursprungsfamilie, wird –wie bereits in Kapitel 2 ausgeführt– als wichtiger Prädiktor für akademischen Erfolg diskutiert (Bornkessel, 2018). Insbesondere in Anlehnung an die Arbeiten von Tinto (1975) und Spady (1971) sowie auch angelehnt an die Arbeit von Neuville et al. (2007) habe ich das Bildungsniveau der Mutter der befragten Personen erhoben, um damit die Bildungsnähe der Ursprungsfamilie zu operationalisieren. Die konkrete Art der Erfassung ist dabei an die Vorgehensweise des Statistischen Bundesamts angelehnt. Konkret lautete die Frage: *Welchen höchsten allgemeinbildenden Schulabschluss hat Ihre Mutter?*. Geantwortet werden sollte, in dem eine der folgenden Kategorien ausgewählt wurde: Hauptschulabschluss, Mittlere Reife, Realschulabschluss, Fachschulreife, Fachhochschulreife/Abschluss einer Fachoberschule, Abitur, allgemeine oder fachgebundene Hochschulreife, einen anderen Schulabschluss, einen anderen Abschluss oder kein Abschluss. An dieser Stelle ist anzumerken, dass im selben Format auch das Bildungsniveau des Vaters abgefragt wurde, was allerdings bei den Modelltests keine Berücksichtigung fand.

Schulleistung

Die Schulleistung wurde als Abiturnote operationalisiert. Die Studienanfänger wurden gebeten, ihre Abiturnote in Dezimalzahlen anzugeben, wobei die an deutschen Schulen in der Regel verwendete Notenskala von 1 bis 6 verwendet wurde. Das Item lautete: *Bitte geben Sie hier Ihre Abiturnote bzw. die Note Ihrer Hochschulzugangsberechtigung an (Dezimalzahl, z.B. 2,3)*.

Tabelle 4.1: Überblick über die verwendeten Instrumente (* verwendet für Modelltests)

	t1	t2	t3
Begrüßung, Infos zur Studie	x	x	x
Studieneinstiegsselbstwirksamkeit*	x	x	x
Allgemeine Selbstwirksamkeit (Schwarzer & Jerusalem, 1999)	x		
Zielsetzung* + Notenziel + minimales Notenziel	x		
Ergebniserwartungen*	x		
Studienzufriedenheit* (Westermann, Heise, Spies & Trautwein, 1996)	x	x	x
Alternativen zum Studium	x		x
Stress (E. M. Klein et al., 2016)	x	x	x
Symptome + Coping (Satow, 2012)		x	
Person-University-Fit* (Cable & DeRue, 2002)	x		x
soziale Unterstützung* (Schulz & Schwarzer, 2003) + Integration*	x	x	x
Informiertheit bei Studienwahl*	x		
Demografie*	x		
Einverständnis zur Verknüpfung der Daten + VP-Code	x		
Ablaufbeschreibung + Gewinnspielteilnahme, Abfrage Emailadresse	x		
Abfrage, ob in Klausurenphase + ob Hausarbeiten geschrieben werden		x	
Handlungen*		x	
Persönlichkeit (Rammstedt, Kemper, Klein, Beierlein & Kovaleva, 2012)		x	
Studieninteresse (Schiefele, Kapp, Wild & Winteler, 1993)		x	
Erwartungen*		x	
Kontrolle* (Perry, Hladkyj, Pekrun & Pelletier, 2001)		x	
Wert* (Pintrich, Smith, Garcia & McKeachie, 1991)		x	
Emotionen (Pekrun, Goetz, Frenzel, Barchfeld & Perry, 2011)		x	
akademische Integration* (Ramm, Multrus, Barge & Schmidt, 2014)		x	x
Ausblick: Ablaufbeschreibung	x	x	x
aktuelle Studiensituation			x
Filter: Studiensituation (falls Abbruch, Sprung zu „Kurzversion“)			x
Kurzversion: falls Abbruch -> Frage nach Gründen			x
Wochenarbeitsstunden für Studium* etc.			x
Anzahl bisher erh. Noten + aktueller Notenschnitt*			x
Selbsteinschätzung Leistung + intell. Entwicklung			x
UCB* (Gehring, 2006)			x
Studienabbruchgedanken + Abbruchintention*			x
Abbruchgründe			x
Bereitschaft zu Nachbefragung			x
Erläuterung Gewinnspiel			x
Abschluss + Dank	x	x	x

Selbstwirksamkeit

Da die kontextspezifische Selbstwirksamkeit eine wichtige Rolle als Prädiktor von Studienerfolg und Studienabbruch spielt, war eine valide und kontextspezifische Erfassung ein wichtiges Anliegen im Zuge der Skalenauswahl. Die in Kapitel 3 ausführlich beschriebene SESW-Skala kam hier zum Einsatz. Sie umfasst 13 Items und soll erfassen, in welchem Ausmaß sich Studieneinsteiger zutrauen, typische Hürden und Herausforderungen im Studieneinstieg zu meistern. Als Skalenwert wurde der Mittelwert über alle 13 Antworten gebildet. Detaillierte Informationen zur Skala finden sich in Kapitel 3, in dem die psychometrische Güte der Skala ausführlich dargestellt wird. Die Selbstwirksamkeit wurde zu allen drei Erhebungszeitpunkten mit dieser Skala erfragt.

Ergebniserwartungen

In Anlehnung an Lent und Brown (2013) habe ich selbst eine Skala zusammengestellt, die die Erwartungen der Studienanfänger hinsichtlich dessen, was bestimmte Verhaltensweisen im Studieneinstieg bewirken, abfragt. Ein Itembeispiel ist *Gutes Zeitmanagement bei der Prüfungsvorbereitung reduziert den Stress im Studium*. Für insgesamt sechs solcher Aussagen sollte auf einer fünfstufigen Likertskala (*trifft nicht zu* = 1 bis *trifft zu* = 5) angegeben werden, inwieweit man der Aussage zustimmt. Auch hier wurde der Skalenwert als mittlere Zustimmung (Mittelwert über die Items) gebildet.

Soziale Unterstützung

Die wahrgenommene soziale Unterstützung wurde mittels zwei Subskalen aus den *Berlin Social Support Scales* von Schulz und Schwarzer (2003) erfasst. Es handelte sich um die Subskala zur wahrgenommenen emotionalen Unterstützung (bspw. *Wenn ich Trost und Zuspruch brauche, ist jemand für mich da.*) und zur wahrgenommenen instrumentalen Unterstützung (bspw. *Es gibt Menschen, die mir ihre Hilfe anbieten, wenn ich sie brauche.*). Beide Subskalen umfassen jeweils vier Items zu denen auf einer vierstufigen Likertskala (*stimme gar nicht zu* = 1 bis *stimme voll und ganz zu* = 4) das persönliche Ausmaß an Zustimmung angegeben werden sollte. Hierbei ist zu erwähnen, dass die Itemformulierungen sehr allgemein gehalten sind, sodass diese Skala, obgleich sie nicht explizit für den Studienkontext entwickelt wurde, eingesetzt werden konnten. Verwendet wurde sie sowohl zum ersten als auch zum zweiten Messzeitpunkt. Der Skalenwert wurde durch Aufsummieren der Itemantworten berechnet. Schulz und Schwarzer (2003) berichteten für die beiden Skalen zusammen (also über acht Items) eine gute interne Konsistenz in Höhe von Cronbachs $\alpha = .83$ ($N = 457$).

Zielsetzung

Ebenfalls angelehnt an Lent und Brown (2013) habe ich das individuelle Maß an Zielsetzung und Zielbindung zu quantifizieren versucht. Nicht selten wird im Studienkontext als recht grobes Maß schlicht nach einem Notenziel gefragt (im Sinne von: *Welche Abschlussnote möchten Sie erreichen?*) oder nach einer Note, die das „Mindestziel“ ist. Zu Vergleichszwecken habe ich diese Operationalisierung ebenfalls mit erhoben, sie allerdings nicht für die Analysen verwendet. Hierfür habe ich angelehnt an die SESW-Skala die 13 dort gelisteten Herausforderungen als Ausgangsbasis verwendet. Die Instruktion der verwendeten Zielsetzungs-Skala lautete: *In diesem Block geht es um Ihre Ziele im Studieneinstieg. In diesem Zusammenhang sind Handlungen aufgeführt, für die Sie bereits zuvor angegeben haben, inwiefern Sie sich diese zutrauen. Bitte geben Sie nun für jedes Ziel an, wie wichtig es für Sie ist, dieses zu erreichen.*

Ziele: Ich werde...

Es sollte dann beispielsweise angegeben werden, wie wichtig es einem ist (*sehr unwichtig* = 1 bis *sehr wichtig* = 5), ... *die zur Verfügung stehende Zeit zur Prüfungsvorbereitung sinnvoll zum Lernen zu nutzen*. Der Skalenwert wurde als Mittel über alle Items gebildet.

Studienwahlsicherheit/Informiertheit

Die Modelle von Tinto (1975) und Neuville et al. (2007) beinhalten die Variable *Studienwahlsicherheit*. Ich entschied mich, ein quantitatives Maß an Informiertheit bei der Studienwahl zu erfassen. Konkret lautet die von mir verwendete Instruktion: *In diesem Block geht es um Ihre Entscheidung für Ihr Studienfach. Bitte geben Sie an, ob Sie die unten aufgeführten Informationsquellen genutzt haben*. Folgende Informationsquellen standen zur Wahl (Mehrfachwahl möglich):

- Haben Sie im Internet zum Thema Studienwahl recherchiert?
- Haben Sie vor Studienbeginn ein oder mehrere Online Self-Assessments genutzt, um sich über Ihr Studienfach zu informieren?
- Haben Sie in der Schule an einer oder mehreren Informationsveranstaltungen zum Thema Studium teilgenommen?
- Haben Sie im Rahmen eines Hochschulinformationstages (HIT) eine oder mehrere Hochschulen besucht?
- Haben Sie eine Studienberatung an einer Hochschule in Anspruch genommen?
- Haben Sie eine Studien- und/oder Berufsberatung bei der Agentur für Arbeit in Anspruch genommen?
- Haben Sie an einem Studieneinstiegsprogramm (z.B. Studieneinstiegswoche) teilgenommen?

- Haben Sie an einem Vorkurs für Ihr Studium teilgenommen (z.B. in Mathematik)?

Für jede Informationsquelle sollte dann angegeben werden, ob diese genutzt wurde. Einen Skalenwert habe ich gebildet, indem ich für jede genutzte Quelle einen Punkt vergeben habe und die Punkte dann aufsummiert habe.

Person-University-Fit

Um die subjektive Passung der Studierenden zu der Hochschule, an der sie studieren, zu erfassen, verwendete ich eine deutsche Übersetzung der *Perceived Fit*-Skala (Cable & DeRue, 2002), die an den Studienkontext adaptiert wurde. Diese Skala habe ich vornehmlich verwendet, um ein Maß für das zu haben, was Spady (1971) „normative Kongruenz“ nennt.

Ursprünglich kommt das der Skala zugrunde liegende Konzept aus dem arbeits- und organisationspsychologischen Kontext und betrifft die subjektive Einschätzung von der Passung einer Person zu dem Unternehmen, in dem sie tätig ist. Es geht hierbei darum, ob die Werte des Unternehmens (Studienkontext: Werte der Hochschule bzw. der Fachkultur) kompatibel sind mit den eigenen Werthaltungen. Konkret umfasst die Skala neun Items, die sich gleichmäßig auf die drei Subskalen aufteilen: *person-organization fit* (P-O-Fit), *needs-supplies fit* und *demands-abilities fit*. Alle drei Subskalen wiesen in der Untersuchung von Cable und DeRue (2002) mit $N_1 = 215$ und $N_2 = 1501$ Angestellten gute bis sehr gute Reliabilitäten auf (Cronbachs $\alpha = .84$ bis $.92$).

Die neun Items wurden an den Studienkontext angepasst, beispielsweise: *Meine persönlichen Werte stimmen mit den Werten und der Kultur meines Faches überein*. Auf einer siebenstufigen Likertskala (*stimme gar nicht zu* = 1 bis *stimme voll und ganz zu* = 7) sollten die Studieneinsteiger ihre Zustimmung angeben. Der Skalenwert wurde durch Aufsummieren der Itemantworten gebildet.

Stress

Zur Erfassung des empfundenen Stresses setzte ich die deutsche Übersetzung (E. M. Klein et al., 2016) der *Perceived Stress Scale* (kurz PSS-10) (S. Cohen, Kamarck & Mermelstein, 1983) zu allen drei Messzeitpunkten ein. Wie der Name schon vermuten lässt, umfasst die Skala zehn Items, die nach potentiell aufgetretenen stressigen Situationen fragen, zum Beispiel: *Wie oft wurden Sie im letzten Monat von unerwarteten Ereignissen überrascht?* Auf einer fünfstufigen Likertskala sollten die Studieneinsteiger angeben, ob beziehungsweise wie oft sie das jeweilige Ereignis in den letzten 30 Tagen erlebt haben (*nie* = 1 bis *sehr oft* = 5). Diese Skala kommt aus dem (sub-)klinischen Diagnostikbereich und dient einer ersten Erfassung von Stressbelastung im Alltag. Sie ist daher zwar nicht explizit auf den Studienkontext zugeschnitten, allerdings so allgemein gehalten, dass sie sich für die Erfassung des Stressempfindens von Studierenden problemlos verwenden lässt. Die zehn Items lassen sich auf die Subskalen *wahrgenommene Hilflosigkeit* (sechs Items, je größer die Hilflosigkeit,

desto größer der Stress) und *wahrgenommen Selbstwirksamkeit* (je höher die Selbstwirksamkeit, desto weniger Stress) aufteilen. Die Items der letztgenannten Subskala wurden entsprechend umgepolt, bevor die Subskalenwerte als Summe der Itemantworten berechnet wurden, sodass höhere Werte für ein höheres Stresslevel stehen. Im Zuge der Längsschnittstudien habe ich für die Modelltest nur den Skalenwert der Subskala *wahrgenommene Hilflosigkeit* verwendet. In ihrer umfangreichen Validierungsstudie für die deutsche Übersetzung der Skala, berichten E. M. Klein et al. (2016) eine interne Konsistenz der Gesamtskala von Cronbachs $\alpha = .84$ sowie diverse gut bis sehr gute Validitätskennwerte. Diese Kennwerte ermittelten sie an einer für die deutsche Gesamtbevölkerung repräsentativen Stichprobe von $N = 2527$ Erwachsenen.

Handlungen

Analog zur Abfrage der Zielsetzung, angelehnt an die SESW-Skala, erfragte ich am zweiten Messzeitpunkt, inwiefern die aus den Herausforderungen direkt abzuleitenden Handlungen tatsächlich ausgeführt wurden. Die Instruktion lautete:

Bitte denken Sie jetzt über Ihr erstes Studiensemester, das Sie gerade abschließen, nach. Im Folgenden sind Handlungen aufgeführt, die Sie während Ihres ersten Semesters ausgeführt haben könnten. Bitte geben Sie jeweils an, inwieweit Sie diese tatsächlich (!) ausgeführt haben.

Ich habe...

Wieder folgte die Liste mit den 13 Herausforderungen, diesmal als Handlungen formuliert, also beispielsweise: *Ich habe selbstständig meinen Studienplan zusammengestellt*. Beantwortet wurden die 13 Items auf einer fünfstufigen Likertskala (*gar nicht* = 1 bis *voll und ganz* = 5). Der Skalenwert wurde als Mittel über die Items gebildet.

An dieser Stelle möchte ich bereits anmerken, dass für die Interpretation der Höhe der Zusammenhänge der Skalenwerte in der SESW-Skala und dieser Skala selbstverständlich zu bedenken ist, dass aufgrund des analogen Aufbaus gegebenenfalls die Korrelationen zwischen den Skalen erhöht sein könnten. Dem wollte ich zumindest in gewissem Maße entgegen wirken, in dem ich in der Instruktion zu dieser Skala explizit (wie oben aufgeführt) darauf hingewiesen habe, was genau erfragt wird, damit Personen, welche unaufmerksam lesen, nicht etwa denken, es seien die gleichen Fragen wie in der SESW-Skala.

Wert

Um den subjektiven Wert des Studiums, der im Modell von Neuville et al. (2007) eine Rolle spielt, zu erfassen, verwendete ich eine deutsche Übersetzung der Subskala *task value* (fünf Items) des *Motivated Strategies for Learning Questionnaire* von Pintrich, Smith, Garcia und McKeachie (1991).

Ein Beispielitem lautet: *Es ist mir sehr wichtig, den Inhalt meines Studiums zu verstehen*. Als Antwortformat kam eine siebenstufige Likertskala (*trifft gar nicht zu* = 1 bis *trifft voll und ganz zu* = 7) zum Einsatz. Der Skalenwert ist die Summe der Itemantworten. Für die englische Version der Skala berichten Pintrich et al. (1991) eine interne Konsistenz von Cronbachs $\alpha = .90$ ($N = 380$).

Zufriedenheit

Um die Zufriedenheit spezifisch im Studienkontext zu erfassen, verwendete ich die von Westermann, Heise, Spies und Trautwein (1996) vorgeschlagene Skala zur Erfassung der Studienzufriedenheit. Diese Skala besteht insgesamt aus neun Items, die in gleichen Anteilen auf die drei Subskalen *Zufriedenheit mit den Inhalten*, *Zufriedenheit mit den Studienbedingungen* und *Zufriedenheit mit der Bewältigung von Studienbelastungen* verteilt sind. Bei den beiden letztgenannten Subskalen sind die Items im Sinne des Konstrukts logisch negativ formuliert, weshalb sie invertiert wurden. Da im Rahmen meiner Studien der Fokus auf personenbezogenen, im weitesten Sinne motivationalen Aspekten lag, entschied ich mich zwar dafür die gesamte Skala jeweils zum zweiten und dritten Messzeitpunkt der Längsschnittstudien einzusetzen, jedoch für die Modelltests nur die Subskala *Zufriedenheit mit der Bewältigung von Studienbelastungen* zu verwenden. Dies erscheint mir am passendsten im vorliegenden Kontext und daher dem Symmetrieprinzip (Wittmann, 1990) am ehesten gerecht zu werden. Ein Beispiel für die Items dieser Subskala lautet: *Das Studium frisst mich auf*. Hasenberg (2012) berichtet in ihrer Untersuchung mit $N = 102$ Studierenden eine interne Konsistenz dieser Subskala von Cronbachs $\alpha = .63$.

Anders als Hiemisch et al. (2005), die eine Antwortskala von 0 bis 100 bei dieser Skala einsetzten, verwendete ich der Konsistenz innerhalb der Befragung wegen, wieder die fünfstufige Likertskala (*trifft gar nicht zu* = 1 bis *trifft voll und ganz zu* = 5), die ich bereits bei vielen anderen Instrumenten im Rahmen dieser Erhebungen verwendete.

Soziale Integration

Zur Erfassung der sozialen Integration verwendete ich drei selbstformulierte Items:

- Ich habe bereits guten Anschluss zu meinen Kommilitonen/Kommilitoninnen gefunden.
- Sollte ich Verständnisprobleme beim Vorlesungsstoff haben, weiß ich, an wen ich mich wenden muss, um Unterstützung zu erhalten.
- Wenn ich beim Lernen fürs Studium Hilfe benötige, kann ich meine Kommilitonen/Kommilitoninnen fragen.

Wichtig ist hierbei die klare Abgrenzung von (allgemeiner) sozialer Unterstützung (siehe oben). Die drei oben genannten Items dienen gezielt der Erfassung der Integration in die akademische Peergroup. In Form eines dichotomen Antwortformates konnten die Studieneinsteiger den Aussagen zustimmen oder sie ablehnen. Der Skalenwert ist die Summe der Itemantworten.

Commitment

Das Commitment gegenüber der Hochschule, habe ich mit Hilfe des *University Citizenship Behavior (UCB)*-Fragebogens (Gehring, 2006) erfasst. Angelehnt an das Konzept des *Organizational Citizenship Behavior* (Organ, 1988) aus dem arbeits- und organisationspsychologischen Kontext hat Gehring (2006) diesen Fragebogen entwickelt, „der es erlaubt, umfeldbezogene [in Abgrenzung zu aufgabenbezogener] Leistung an Hochschulen als Kriterium des Studienerfolges zu erfassen“ (S.42).

Hierbei handelt es sich im Studienkontext um solche Leistungen, die nicht in direktem Zusammenhang mit den zu benotenden Studienleistungen stehen, also nicht direkt „belohnt“ werden (nicht zum *formal reward system* gehörig). Dieser insgesamt 27 Items umfassende Fragebogen ist in vier Subskalen aufgeteilt, die im Folgenden jeweils mit einem Beispielitem aufgelistet sind.

- Hilfsbereitschaft (8 Items): Ich helfe KommilitonInnen, denen der Lernstoff Verständnisprobleme bereitet.
- Gewissenhaftigkeit (4 Items): Ich komme so oft zu den Lehrveranstaltungen wie ich kann.
- Unkompliziertheit (8 Items): Ich akzeptiere ungünstige Rahmenbedingungen (z.B. Veranstaltungszeiten oder -orte), wenn mir die Veranstaltung wichtig ist.
- Eigeninitiative (7): Ich informiere mich über Entwicklungen in der Hochschulpolitik.

Genau wie von Gehring (2006) beschrieben, habe ich als Antwortformat eine sechsstufige Likertskala (*trifft überhaupt nicht zu* = 1 bis *trifft vollständig zu* = 6) verwendet. In seiner Studie mit $N = 282$ Studierenden ergaben sich gute bis sehr gute Reliabilitätskennwerte von Cronbachs $\alpha = .70$ bis $.85$. Entgegen der Vorgehensweise von Gehring (2006) habe ich einerseits den (Sub-)Skalenwert jeweils als Mittelwert über die Items berechnet (statt als Summe) und habe andererseits die vier Subskalen gleich gewichtet und somit als Gesamt-Skalenwert einen gleich-gewichteten Mittelwert über die Subskalen verwendet. Den Subskalenwert für *Unkompliziertheit* habe ich im Zuge der Testung der Modelle von Spady (1971) und Tinto (1975) als Operationalisierung der Variable *institutionelles Commitment* herangezogen, weil diese Subskala das von Spady (1971) und Tinto (1975) beschriebene Konstrukt (vgl. Kapitel 2.4.2) gut abbildet. In der Untersuchung von Gehring (2006) wies diese Subskala eine sehr gute Reliabilität auf: $\alpha = .85$.

Akademische Integration

Ebenfalls an die Studie von Ramm et al. (2014) angelehnt, wurde die Teilnehmer gebeten, die Lehrveranstaltungen, an denen sie teilnehmen, insgesamt zu bewerten. Die Instruktion lautete: *In diesem Block geht es um Ihren Studienalltag. Auf wie viele Lehrveranstaltungen, die Sie im zweiten Semester besuchen/besuchten, trifft Folgendes zu?* Es folgte eine Aufzählung von Aspekten, die die akademische Integration, wie Tinto (1975) sie beschreibt, fördern können. Für jeden Aspekt sollte auf einer fünfstufigen Likertskala (*keine* = 1 bis *alle* = 5) angegeben werden sollte, auf wie viele Lehrveranstaltungen der jeweilige Aspekt zutrifft. Ein Beispiel ist: *Der Dozent/die Dozentin schafft es, für das Gebiet/den Stoff zu interessieren und zu motivieren..* Alle aufgeführten Aspekte waren positive Beschreibungen. Der Skalenwert wurde gebildet, indem die Summe der Itemantworten gebildet wurde. Diese Variable ist also weniger eine Personeneigenschaft im engeren Sinne sondern eine (subjektive) Einschätzung der Studienanfänger hinsichtlich dessen, was die Lehrenden tun (können), um die Einbindung der Studierenden zu fördern.

Intellektuelle Entwicklung

Die Teilnehmer wurden zudem gebeten, ihre eigene bisherige intellektuelle Entwicklung im Studium einzuschätzen. Die dazu verwendeten Items sind an die Erhebung von Ramm et al. (2014) angelehnt. Die Instruktion lautete: *Bitte geben Sie an, inwieweit Sie sich Ihrer Einschätzung nach in den folgenden Bereichen im Rahmen Ihres Studiums weiterentwickelt haben.* Die aufgeführten Bereiche waren: (i) *Fähigkeit zum selbstständigen Weiterlernen* und (ii) *Teamfähigkeit/Zusammenarbeit und Aufgabenlösung mit anderen*. Auf einer siebenstufigen Likertskala (*gar nicht* = 1 bis *sehr viel* = 7) sollten die Teilnehmer ihre Einschätzung abgeben. Das erste Item wurde als Maß für die intellektuelle Entwicklung verwendet.

Akademisches Engagement

Um das akademische Engagement im Sinne von Durchhaltevermögen und Fleiß im Studium zu bestimmen, wählte ich ein quantitatives Maß: *In diesem Block geht es darum, wie viel Zeit Sie pro Woche für unterschiedliche Tätigkeiten aufwenden. Schätzen Sie bitte möglichst genau, wie viel Zeit (Stunden pro Woche) Sie durchschnittlich in den letzten Semestern und Ferien aufgewendet haben für...:*

- ...die Teilnahme an Veranstaltungen im Rahmen Ihres Studiums (Gemeint sind damit die Stunden, in denen Sie an Vorlesungen, Seminaren etc. teilnehmen oder Kurse im Labor oder im Feld haben. Hierzu zählen NICHT die Stunden, die Sie für die Vor- und Nachbereitung der Veranstaltungen investieren.)

- ...die Vor- und Nachbereitung von Veranstaltungen, Lernen für Prüfungen, erstellen von Hausarbeiten etc. (Hierzu zählen NICHT die Stunden, in denen Sie Veranstaltungen besuchen.)

In ein Freitextfeld war entsprechend eine Stundenzahl einzutragen. Diese Variable ist sowohl im Modell von Tinto (1975) als auch im Modell von Neuville et al. (2007) zu finden (siehe Kapitel 2.4.2).

Abbruchintention

Die Abbruchintention der Studieneinsteiger wurde mittels zweier Items erfasst, für die auf einer siebenstufigen Likertskala die Zustimmung angegeben werden sollte (*stimme gar nicht zu* = 1 bis *stimme voll und ganz zu* = 7). Die Items lauteten: *Ich habe vor, das Studium zu einem erfolgreichen Abschluss zu bringen.* und *Ich habe vor, das Studium abubrechen.* Die Antworten auf das erste Item wurden invertiert und die Items dann durch Aufsummieren der beiden Antwortscores zu einem Skalenwert zusammengefasst

Ähnlich wurde die Abbruchintention beispielsweise auch in der 21. Sozialerhebung (Middendorff et al., 2016, siehe Kapitel 10) oder bei Nolden (2019) sowie Respondek et al. (2017) und Treppesch und Ehrlich (2018) operationalisiert. Zusätzlich wurde der Vollständigkeit halber noch nach der Intention zum Fach- und Hochschulwechsel gefragt, sowie danach, ob eine temporäre Unterbrechung des Studiums in Erwägung gezogen wird. Letztere wird in der Literatur in Abgrenzung zu *dropout* auch als *stopout* bezeichnet (Thomas & Hovdhaugen, 2014).

Aktuelle Studienleistung

Um die aktuelle Studienleistung zu erfassen, wurden die gängigste Operationalisierung (Brandstätter et al., 2006) verwendet: Die Teilnehmer wurden darum gebeten, ihre aktuelle Durchschnittsnote anzugeben. Verwendet wurde die an deutschen Hochschulen in der Regel verwendete Notenskala von 0 bis 15 Punkten.

4.1.2 Repräsentativität der untersuchten Stichproben

Wie in den meisten Untersuchungen, war auch in diesen Längsschnittstudien damit zu rechnen, dass die untersuchte Stichprobe nicht repräsentativ für die Zielpopulation ist. Die Zielpopulation der hier vorgestellten Studien ist die Gruppe der Studieneinsteiger in Deutschland. Wenn es um Repräsentativität geht, stellt sich immer die Frage: "Repräsentativ hinsichtlich welcher Eigenschaften?". Die Repräsentativität hinsichtlich demografischer Variablen wie Alter, Geschlecht, Verteilung auf Studienfächer beispielsweise lässt sich problemlos prüfen, indem man die Kennwerte der Stichprobe mit

den beim Statistischen Bundesamt hinterlegten Kennwerten der Zielpopulation vergleicht. Dies werde ich für jede der drei Längsschnittstichproben jeweils im entsprechenden Kapitel berichten.

4.1.3 Selektivität der untersuchten Stichproben

Die Repräsentativität der Stichproben hängt eng zusammen mit der (Selbst-)Selektion der Teilnehmer. Dies ist bedingt durch die Rahmenbedingungen der Datenerhebung, die ich im Abschnitt 4.1 ausführlich dargestellt habe. Wie bisherige Forschung zum Thema *Selbstselektion zur Studienteilnahme* gezeigt hat, sind Frauen im Mittel eher geneigt, an Studien teilzunehmen (von Maurer & Jandura, 2009).

Zudem liegt die Vermutung nahe, dass besonders hoch gewissenhafte, offene und verträgliche Studienanfänger eher die Entscheidung treffen, an einer solchen Studie teilzunehmen. In Bezug auf die Allgemeinbevölkerung konnte dies bereits gezeigt werden (Nestler, Thielsch, Vasilev & Back, 2015). Daher ist im Folgenden davon auszugehen, dass die Ausgangsstichproben (und damit folglich auch die Analysestichproben) durchaus einem Selbstselektionsbias unterliegen. Dieser wiederum kann einerseits zu Varianzeinschränkungen führen und andererseits die Generalisierbarkeit der mit diesen Daten erzielten Befunde (externe Validität) limitieren. Auf diese beiden Aspekte werde ich bei der Interpretation der Ergebnisse eingehen.

4.1.4 Dropout innerhalb der Studien

An dieser Stelle möchte ich ein klassisches Problem längsschnittlichen Studiendesigns thematisieren (Asendorpf, van de Schoot, Denissen & Hutteman, 2014): Dropout der Teilnehmer über die Erhebungsdauer hinweg (in der englischsprachigen Literatur auch als „attrition“ und im Zusammenhang mit Panelstudien auch als „Panelmortalität“ bezeichnet). Dropout ist immer problematisch, besonders ungünstig ist es, wenn der Dropout systematisch ist. Im vorliegenden Fall geht es um den Studienabbruch. Es ist zu befürchten, dass insbesondere solche Befragten bei den späteren Messzeitpunkten nicht mehr teilnehmen, die abbruchgefährdet sind (Tieben, 2016). Das bedeutet, dass nicht nur im Allgemeinen mit einem Verlust an Stichprobenumfang über die Dauer der Untersuchung hinweg zu rechnen ist, sondern auch davon auszugehen ist, dass ein selektiver Dropout vorliegt: Wer (von den Teilnehmern am ersten Messzeitpunkt) sich mit dem Gedanken trägt, das Studium abubrechen, ist vermutlich auch weniger gewillt, im Verlauf des Studieneinstiegs noch an zwei Folgebefragungen zum Thema Studium teilzunehmen. Dies ist nur eine Annahme, die ich aufgrund der beschriebenen Dropout-Problematik hier nicht empirisch testen kann, jedoch sollte diese mögliche (zusätzliche) Selektivität der Analysedatensätze bei der Interpretation der Ergebnisse berücksichtigt werden.

4.1.5 Einschlusskriterien

Im Rahmen der getesteten Modell sollen die beiden Kriterien *Studiennote* und *Abbruchintention* vorhergesagt werden. Daher legte ich als Einschlusskriterium für die Analysen fest: Nur wer am dritten Messzeitpunkt teilgenommen und sowohl seine aktuelle Studiennote als auch seine Abbruchintention angab, wurde in die Analysen einbezogen. In Kombination mit dem beobachteten Dropout über die Gesamtlaufzeit der Längsschnittstudien hinweg (9 Monate), ergab sich für alle drei Jahrgänge eine Reduktion des ursprünglichen Stichprobenumfangs von circa 50%. Dies ist nahezu vollständig auf das beschriebene Phänomen des Dropouts zurückzuführen. Studienanfänger, die am jeweils dritten Messzeitpunkt noch an der Studie teilnahmen, machten in der Regel auch Angaben zu diesen beiden Variablen. Die genaue Beschreibung der Analysedatensätze findet sich in den separaten Kapiteln zu den Studien (Kapitel 5, 7, 8), die Angabe, wie viele Fälle jeweils ausgeschlossen wurden findet sich in Abschnitt 4.1.8.

4.1.6 Fehlende Werte

Wie in Studien mit Selbstberichtsdaten nicht unüblich (Lüdtke, Robitzsch, Trautwein & Köller, 2007), sind auch in den dargestellten Längsschnittstudien fehlende Werte zu finden. Diese resultieren teilweise aus Unit-Nonresponse (eine Person hat an einem der drei Befragungszeitpunkte gar nicht teilgenommen, siehe oben) oder aus Item-Nonresponse (einzelne Items oder Subskalen wurden nicht beantwortet). Selbst nach dem Anlegen des oben beschriebenen Einschlusskriteriums gab es fehlende Werte in den Analysedatensätzen, da teilweise Item-Nonresponse vorlag und zudem Personen, die nur am zweiten Messzeitpunkt nicht teilgenommen hatten, durchaus in der Stichprobe verblieben, weshalb in deren Daten dann fehlende Werte in allen am zweiten Messzeitpunkt erhobenen Variablen vorlagen.

In der Klassifikation nach Rubin (1976) werden drei Arten von fehlenden Werten unterschieden: Missing Completely At Random (MCAR), Missing At Random (MAR) und Missing Not At Random (MNAR) (Rubin, 1976; Lüdtke et al., 2007).

Diese drei Arten unterscheiden sich hinsichtlich der Frage, ob die Annahme, dass die fehlenden Werte zufällig aufgetreten sind, haltbar ist. Bei MCAR gilt diese Annahme. Das bedeutet, dass man in diesem Fall davon ausgeht, dass das Fehlen eines Wertes weder mit der Ausprägung des Konstruktes (und dem theoretisch resultierenden Wert in der entsprechenden Variable) bei der betreffenden Person, noch mit der Ausprägung anderer Variablen bei dieser Person zusammenhängt.

Die Zwischenkategorie MAR, die von Lüdtke et al. (2007) auch als *Conditionally Random* bezeichnet wird, geht mit der Annahme einher, dass unter Kontrolle gewisser anderer Variablen, das Fehlen der Werte in der Variablen, die man betrachtet, zufällig ist.

Zuletzt ist die Kategorie MNAR zu nennen, bei der die Annahme gemacht wird, dass das Fehlen von Werten in einer Variablen auf die Ausprägung dieser fehlenden Werte zurückzuführen ist. Das bedeutet, auch unter Kontrolle gewisser anderer Variablen, kann man –anders als bei MAR – nicht

davon ausgehen, dass das Fehlen dieser Werte zufällig ist. Im Allgemeinen ist, wenn fehlende Werte vorliegen, MCAR am wenigsten problematisch.

Der geringen Anteil an fehlenden Werten in den Analysedatensätzen der drei Jahrgänge (Jahrgang 1: unter 1% , Jahrgang 2: unter 2% und Jahrgang 3: ebenfalls unter 2%), ist so gering, dass dieser als unproblematisch eingestuft werden. Beispielsweise Schafer (1999) postuliert, dass bis zu einem Anteil von 5% fehlender Werte, nicht mit ernststen Verzerrungen in den anschließenden Analysen zu rechnen ist. Andere Autoren gehen sogar von höheren, noch akzeptablen Anteilen von fehlenden Daten aus (Dong & Peng, 2013). Dies ist allerdings als Richtwert zu verstehen. Wie Dong und Peng (2013) festhalten, ist im Einzelfall auch entscheidend, ob und wenn ja welchem Muster die fehlenden Wert folgen. Sie betonen, dass es wichtig ist, dass das Thema *fehlende Werte* berichtet und ausführlich dargestellt wird und dass im weiteren Verlauf geeignete Imputationsverfahren verwendet werden, um diese fehlenden Werte zu ersetzen.

Diesen Empfehlungen folgend, stelle ich im nächsten Abschnitt dar, welche Erkenntnisse hinsichtlich des (nicht) zufälligen Auftretens fehlender Werte ich in Bezug auf die Analysedatensätze aus den drei Jahrgängen gewonnen habe und wie ich mit diesen umgehe.

4.1.7 Test auf *Missing Completely at Random* in den Längsschnittdaten

Zunächst führte ich für jeden Jahrgang einen Test auf *missing completely at random* (MCAR) nach Little (1988) durch, also darauf, ob die fehlenden Werte in den zu analysierenden Variablen im Analysedatensatz rein zufällig sind und keinem Muster folgen. Der MCAR-Test ist ein Schritt in Vorbereitung auf die Entscheidung ist, ob und wie fehlende Werte imputiert werden können. Dieser χ^2 -Test hat die Nullhypothese, dass die fehlenden Werte rein zufällig über den Datensatz verteilt sind. Dementsprechend wären hier nicht-signifikante Ergebnisse im *MCAR-Test* wünschenswert. Im Folgenden berichte ich die Ergebnisse der durchgeführten MCAR-Tests für die Analysedatensätze der drei Jahrgänge separat. Anzumerken ist hierbei, dass die unterschiedliche Anzahl an Freiheitsgraden aus den unterschiedlichen Fallzahlen pro Variable in den Datensätze aus den drei Jahrgängen resultiert, obwohl in allen drei Datensätzen die gleiche Anzahl an Variablen berücksichtigt wurde.

- Jahrgang 1: $\chi^2(117) = 212.091, p < .001$.
- Jahrgang 2: $\chi^2(46) = 52.561, p = .235$.
- Jahrgang 3: $\chi^2(71) = 103.556, p = .007$.

Die Ergebnisse belegen, dass nur für den Analysedatensatz aus dem zweiten Jahrgang die Annahme, dass MCAR vorliegt, beibehalten werden kann. Für die Analysedatensätze aus Jahrgang 1 und 3 muss die Annahme, dass MCAR vorliegt, verworfen werden.

Somit kann im Folgenden nicht davon ausgegangen werden, dass die fehlenden Werte in den Analysedatensätzen aus Jahrgang 1 und 3 zufällig über die Variablen verteilt sind. Das bedeutet weiterhin, dass mit einer systematischen Verzerrung aufgrund von systematischen *Missings* zu rechnen ist. Dabei ist die Systematik der fehlenden Werte ziemlich deutlich erkennbar: In Jahrgang 1 nahm eine, in Jahrgang 2 neun und in Jahrgang 3 sogar dreizehn Personen nur zum ersten und dann wieder zum dritten Messzeitpunkt an der jeweiligen Studie teil. Da ich mich entschieden habe, dass nur die Teilnahme am dritten Messzeitpunkt und dabei insbesondere das Beantworten der Items hinsichtlich der beiden Kriterienvariablen *Abbruchintentionen* und *aktuelle (Studien-) Note* die Einschlusskriterien sind, verblieben diese Personen in den Analysedatensätzen, obgleich sie fehlende Werte für alle am zweiten Messzeitpunkt erhobenen Variablen aufweisen. Bei diesen Personen nun ist allerdings aufgrund der fehlenden Werte in einer zum zweiten Messzeitpunkt erhobenen Variable durchaus auf fehlende Werte in den anderen zum zweiten Messzeitpunkt erhobenen Variablen zu schließen. Das bedeutet, dass hierbei eine Systematik vorliegt, was die signifikanten Ergebnisse in den MCAR-Tests erklären könnte. Jedoch müsste diese Erklärung dann für alle drei Jahrgänge gleichermaßen gelten. Da der MCAR-Test für den Analysedatensatz von Jahrgang 2 nicht signifikant wurde, muss es eine zusätzliche Erklärung geben, die nur für Jahrgang 1 und 3 zutrifft. Dies könnte gemäß der Definition von MCAR folgende Erklärung sein: Nur in Jahrgang 1 und 3 lassen sich die fehlenden Werte (in den Variablen, die am zweiten Messzeitpunkt erhoben wurden), aufgrund der Ausprägung der vorliegenden Werte in den zum ersten und zum dritten Messzeitpunkt erhobenen Variablen vorhersagen. Dies lässt sich tatsächlich erkennen: Alle Personen, die in Jahrgang 1 am zweiten Messzeitpunkt nicht teilnahmen, haben auffällig niedrige Werte in der Variable *Integration* zum dritten Messzeitpunkt berichtet. Gleiches gilt für die Personen aus Jahrgang 3, die am zweiten Messzeitpunkt nicht teilnahmen, am dritten aber schon. Bei Jahrgang 2 allerdings liegen für drei der neun Personen, die nicht am zweiten Messzeitpunkt teilnahmen, neben auffällig niedrigen Werten in der Variable *Integration* auch auffällig hohe Werte in der Variablen *Abbruchintention* (Messzeitpunkt 3) vor, für die anderen sechs nicht. Hier ist das Muster augenscheinlich weniger systematisch, was das nicht-signifikante Ergebnis des MCAR-Tests in Bezug auf den Analysedatensatz von Jahrgang 2 erklären könnte.

Vor dem Hintergrund der im Abschnitt 4.1.6 beschriebenen sehr niedrigen Anteile von fehlenden Daten in den Analysedatensätzen, mag dies jedoch weniger schwer wiegen (Dong & Peng, 2013). Allerdings ist dabei zu bedenken, dass ich, wie oben beschrieben, nur diejenigen Fälle in die Analysen einbezogen habe, für die zum einen überhaupt Angaben vom dritten Messzeitpunkt und zum anderen auf jeden Fall Daten zu den beiden Kriterienvariablen *Abbruchintention* und *aktuelle Studiennote* vorliegen. Dass im Gesamtdatensatz die fehlenden Werte durchaus einer Systematik unterliegen, ist anzunehmen, was sich aus der bereits angesprochenen Dropout-Problematik in Längsschnittstudien im Allgemeinen (Demirtas, 2004) und Untersuchungen zum Studienabbruch im Speziellen (Tieben, 2016) ergibt.

Wie im folgenden Abschnitt ausführlich beschrieben, habe ich mich dafür entschieden, die fehlenden Wert im jeweiligen Analysedatensatz mithilfe der *Full Information Maximum Likelihood*-Methode zu imputieren.

4.1.8 Imputationsverfahren

Es gibt verschiedene Möglichkeiten, mit fehlenden Werten umzugehen. Lüdtke et al. (2007) teilen die verschiedenen Verfahren in zwei Gruppen ein: klassische, imputationsbasierte (erst fehlende Werte schätzen, dann Modelle berechnen) und modellbasierte (fehlende Werte schätzen und Modelle berechnen in einem Schritt) Verfahren. Im Folgenden stelle ich die Verfahren etwas ausführlicher dar. Zu den klassischen Verfahren im Umgang mit fehlenden Werten gehört der sogenannte fallweise Ausschluss. Dabei werden Fälle, bei denen Werte fehlen, komplett ausgeschlossen. Dieses Vorgehen wird generell eher kritisch beurteilt. Graham (2009) empfiehlt dieses Vorgehen nur, falls weniger als 5% der Fälle dadurch ausgeschlossen werden. So würden meist nur geringe Fehler in der Parameterschätzung auftreten. Allerdings führt eine Reduktion der Stichprobe aufgrund fehlender Werte zwangsläufig zum Verlust statistischer Power (Asendorpf et al., 2014) und es gehen Informationen verloren (Lüdtke et al., 2007).

Ein weiteres klassisches Vorgehen ist der paarweise Ausschluss. Dabei werden alle Fälle exakt so berücksichtigt, wie sie vorliegen. Das bedeutet, dass ausschließlich die vorliegenden Informationen ohne zusätzlichen Schätzungen berücksichtigt werden, aber auch keine Fälle ausgeschlossen werden, „nur“ weil eine oder mehrere Variablen fehlende Werte aufweisen. Lüdtke et al. (2007) raten deutlich von der Verwendung dieses Vorgehens ab. Unter anderem nennen sie als Nachteil, dass durch dieses Vorgehen das Problem auftreten kann, dass eine Korrelationsmatrix zwischen den betrachteten Variablen nicht positiv-definit ist, was ein Problem für weitere Analysen wie beispielsweise Modellfitanalysen darstellt.

Eine dritte Möglichkeit, mit fehlenden Werten umzugehen, ist die *Imputation*. Hierbei werden fehlende Werte geschätzt, sodass Analysen mit einem vollständigen Datensatz gerechnet werden können, ohne dass Fälle dafür ausgeschlossen werden müssen und somit die statistische Power reduziert würde.

Verschiedene Imputationsverfahren unterscheiden sich darin, wie ein fehlender Wert geschätzt wird. So kann eine einfache Ersetzung erfolgen (vgl. single imputation, Lüdtke et al., 2007) oder eine multiple. Letztere ist dadurch charakterisiert, dass ein fehlender Wert nicht durch eine einzelne Schätzung ersetzt wird, sondern durch mehrerer in Frage kommende Werte. So ergeben sich mehrere vollständige Datensätze, die alle auf dieselbe Art und Weise mit weiteren Analysen untersucht werden und abschließend werden die Ergebnisse dieser Analysen zu einem Gesamtergebnis zusammengefasst; hier sind Imputation und Analyse demnach separiert (Lüdtke et al., 2007).

Die vierte Möglichkeit, mit fehlenden Werten umzugehen, ist die *Full Information Maximum Likelihood*-Methode (*FIML*-Methode), die zu den modellbasierten Verfahren gehört. Dabei werden nicht, wie bei der multiplen Imputation, das Ersetzen und das Analysieren separat nacheinander durchgeführt, sondern in einem gemeinsamen Schritt.

Gemäß Lüdtke et al. (2007) werden „keine fehlenden Werte imputiert oder geschätzt, vielmehr wird eine Schätzung der Populationsparameter und ihrer Standardfehler auf der Basis der beobachteten Daten unter der Annahme MAR vorgenommen.“ (S.112). Zusätzlich wird die multivariate Normal-

verteilung der beteiligten Daten vorausgesetzt.

Simulationsstudien, die sich mit den Auswirkungen von Verletzungen dieser beiden Annahmen beschäftigen, zeigen, dass die *FIML*-Methode selbst beim Vorliegen von Verletzungen dieser Annahmen verlässlichere Ergebnisse erzielt als die klassischen Verfahren zum Umgang mit fehlenden Werten (Enders, 2001; Newman, 2016).

Lüdtke et al. (2007) empfehlen zusammenfassend den fallweisen Ausschluss, sofern dadurch weniger als 5% der Fälle ausgeschlossen werden und die *FIML*-Methode insbesondere dann, wenn man mit Strukturgleichungsmodellen arbeiten möchte. Auch andere Autoren empfehlen die Verwendung der *FIML*-Methode explizit für die Analyse von Längsschnittdaten (Demirtas, 2004) beziehungsweise von Strukturgleichungsmodellen (Asendorpf et al., 2014).

Nachdem ich verschiedene Methoden des Umgangs mit fehlenden Werten und verschiedene Imputationsverfahren vorgestellt habe, möchte ich an dieser Stelle zusammenfassen, wie ich in Bezug auf die Daten der drei Längsschnittstudien vorgegangen bin: Ich entschied mich insgesamt für ein mehrschrittiges Vorgehen. Zunächst wurden die Personen ausgeschlossen, welche nicht am dritten Messzeitpunkt teilgenommen hatten (siehe oben). Die in der Stichprobe verbleibenden Personen wurden anschließend nochmals selektiert: Wer die beiden Kriterien nicht berichtet hat, wurde ausgeschlossen. Das führte in Jahrgang 1 zum Ausschluss 6 Fällen (etwa 4%), bei Jahrgang 2 zum Ausschluss von 29 Fällen (etwa 20%) und bei Jahrgang 3 zum Ausschluss von 26 Fällen (etwa 14%). Dies mag zwar prozentual über der von Lüdtke et al. (2007) empfohlenen Grenze für fallweisen Ausschluss liegen, jedoch erscheint es mir verlässlicher, die Werte in den letztendlich vorherzusagenden Kriterienvariablen nicht zu imputieren. Fehlende Werte in allen anderen für die Modelltests verwendeten Variablen habe ich anschließend mittels *Full Information Maximum Likelihood*-Methode imputiert.

4.1.9 Modellfit testen

Um die Modellfits zu testen, führte ich pro Jahrgang für jedes Modell einen Modelltest durch. Dafür spezifizierte ich die Modelle jeweils als Pfadmodelle, also als Modelle, die ausschließlich mit manifesten Variablen aufgesetzt werden. Die Analyse noch komplexerer Strukturgleichungsmodelle (manifestes Messmodell kombiniert mit einem latenten Strukturmodell) hätte bei den realisierten Stichprobenumfängen eine zu geringe statistische Power gehabt. Entsprechend der Empfehlung von Hu und Bentler (1999) zog ich zur Beurteilung der Güte der Modellfits eine Kombination von Modellfitindizes heran. Konkret berichte ich in den entsprechenden Kapitel folgende Modellfitindizes (die schon aus Kapitel 3 bekannt sind): χ^2 , CFI, RMSEA, SRMR und AGFI.

4.2 Studie 4: Sekundärdatenanalyse

Neben dem Vergleich etablierter Modelle, wurde im Zuge dieses Promotionsprojektes auch ein neues Modell entwickelt und empirisch getestet (vgl. Kapitel 6). Im Anschluss an die drei Längsschnittstudien sollte dieses zusätzlich mit externen Daten validiert werden. Zudem sollte anhand eines umfangreichen Datensatz mit ausreichender Besetzung der Fachcluster untersucht werden, ob die im neuen Modell spezifizierten Prädiktoren in den verschiedenen Fachclustern vergleichbar gut zur Vorhersage der beiden Kriterien geeignet sind.

Um diese Forschungsfragen beantworten zu können, wurden die Daten der 21. Sozialerhebung herangezogen. Aufgrund der Datenstruktur entschied ich, multiple Regressionen statt Pfadmodellanalysen durchzuführen, wie im Kapitel 10 ausführlich dargestellt. Zum Umgang mit fehlenden Werten wurde, wie auch bei den Analysen der Längsschnittdaten, die *FIML*-Methode herangezogen.

4.3 Studie 5: Interventionsstudie

Die fünfte und letzte Studie im Rahmen meines Promotionsprojekts ist eine Interventionsstudie zur Untersuchung der Effekte institutioneller Interventionen auf die im neuen Modell spezifizierten Variablen. Exemplarisch wurde untersucht, ob sich bei zwei Workshops für Studieneinsteiger im Prä-Post-Vergleich ein Effekt auf die Selbstwirksamkeit, erfasst mit der SESW-Skala, zeigt. Mit dieser Pilotstudie sollte exemplarisch aufgezeigt werden, wie auf Basis des neuen Modells die Wirksamkeit institutioneller Interventionen evaluiert werden kann.

4.4 Verwendete Software

Für die verschiedenen Analysen, die ich im Rahmen dieses Promotionsprojektes durchgeführt habe, kamen verschiedene Software-Pakete zum Einsatz. Für die Analyse der Längsschnittdaten verwendete ich einerseits *SPSS* (IBM Corp, 2012) und andererseits *SAS software* (*SAS 9.4 und SAS Studio 3.8*)².

Zur Analyse der psychometrischen Güte der SESW-Skala und zur Auswertung der Daten aus der 21. Sozialerhebung und aus der Interventionsstudie verwendete ich zusätzlich zu *SPSS* noch *R* (R Core Team, 2017). Einige Analysen und das Erstellen mancher Tabellen und Abbildungen wurde mit Hilfe verschiedener R-Pakete durchgeführt, darunter das *lavaan*-Paket (Rosseel, 2012), das *sjPlot*-Paket (Lüdtke, 2018), das *ggplot2*-Paket (Wickham, 2016), das *psych*-Paket (Revelle, 2018), das *cNORM*-Paket (Lenhard, Lenhard, Suggate & Segerer, 2018) und das *stargazer*-Paket (Hlavac,

²Copyright © 2019 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

2018).

Nachdem in diesem Kapitel ein Überblick über die durchgeführten empirischen Studien gegeben wurde und studienübergreifend relevante methodische Aspekte dargelegt wurden, sind in den folgenden Kapiteln die einzelnen Studien separat dargestellt. Die für die Sekundärdatenanalyse (siehe Kapitel 10) und für die Interventionsstudie (siehe Kapitel 11) verwendeten Instrumente werden in den jeweiligen Kapiteln vorgestellt. Für alle fünf Studien werden die jeweils realisierten Stichproben ebenso wie die Ergebnisse der jeweiligen Analysen in den einzelnen Kapiteln beschrieben.

5 Studie 1

5.1 Messzeitpunkte bei Studie 1

Studie 1 ist die erste der drei Längsschnittstudien im Rahmen dieses Promotionsprojektes. Sie umfasste drei Messzeitpunkte und lief über das erste Studienjahr (9 Monate) hinweg. Messzeitpunkt 1 war von November bis Dezember 2016 (Beginn erstes Semester), Messzeitpunkt 2 war im Februar 2017 und Messzeitpunkt 3 von Juni bis Juli 2017. Dabei waren die jeweiligen Online-Befragungen immer etwa einen Monat lang zugänglich.

Zum ersten Messzeitpunkt nahmen $N = 311$ Personen teil. Zu Messzeitpunkt 2 waren es noch $N = 157$ Personen, was einem Dropout von ca. 50% entspricht. Zu Messzeitpunkt 3 waren noch 151 Personen dabei. Eine ähnliche Dropout-Quote verzeichneten beispielsweise Ehlert et al. (2017) in ihrem *Berliner Studienberechtigten-Panel*, im Rahmen dessen Oberstufenschüler vom letzten Schuljahr bis zum dritten/vierten Semester begleitet wurden.

5.2 Stichprobe bei Studie 1

Wie im Kapitel 4 beschrieben, wurde das Einschlusskriterium (es müssen Angaben zur aktuellen Studiennote und zur Abbruchintention vorliegen) angelegt. Nur 6 Fälle erfüllten dieses Einschlusskriterium nicht, sodass eine Analysestichprobe aus dem ersten Jahrgang im Umfang von $N = 145$ vorlag.

Im Mittel waren die Studienanfänger in der Analysestichprobe $M = 20.3$ Jahre ($SD = 2.5$) alt. 79.3% von ihnen gaben an, weiblich zu sein, 18.6% gaben an, männlich zu sein, die restlichen 2.1% machten keine Angabe zu ihrem Geschlecht. Zwecks Übersichtlichkeit berichte ich die Verteilung auf die Studienfächer zusammengefasst als Verteilung auf Fachcluster. In dieser Studie sah die Verteilung wie folgt aus: 18.6% Naturwissenschaften, 29.0% Medizin oder Psychologie, 14.5% Sprachen, Kulturen, Soziales, 30.3% Erziehungswissenschaften, 7.6% Wirtschaftswissenschaften, Jura oder sonstige Fächer. Die Verteilung der Studierenden auf die Fachcluster ist in Abbildung 5.1 grafisch dargestellt. Vergleicht man die Ausgangsstichprobe von $N = 311$ Studienanfängern mit der Analysestichprobe von $N = 145$, so zeigt sich ein signifikanter Unterschied hinsichtlich der Verteilung auf die Studienfachcluster ($\chi^2(4) = 16.26, p < .01$). Dieser Unterschied manifestiert sich in erster Linie darin, dass prozentual einerseits deutlich weniger Teilnehmer aus den Fachclustern *Sprachen, Kulturen, Soziales* sowie *Naturwissenschaften* und andererseits deutlich mehr Teilnehmer aus dem Fachcluster *Erziehungswissenschaften* in der Analysestichprobe im Vergleich zur Ausgangsstichprobe waren. Anders

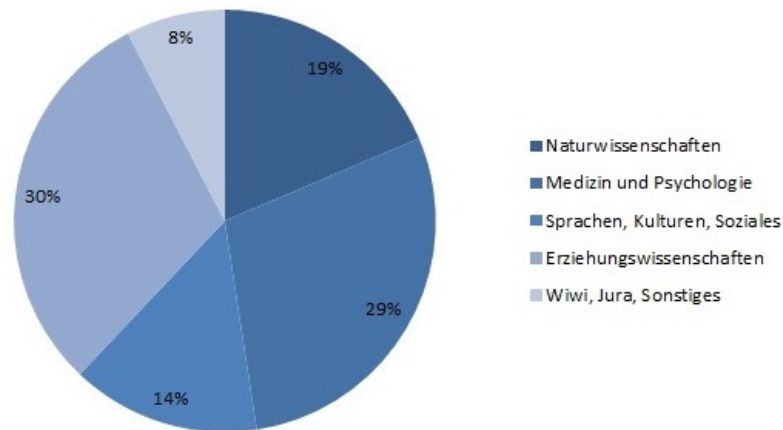


Abbildung 5.1: Verteilung auf die Fachcluster [Jahrgang 1, $N = 145$]

formuliert: Der Dropout über den Erhebungszeitraum hinweg war in den Fachclustern *Sprachen, Kulturen, Soziales* sowie *Naturwissenschaften* höher als in dem Fachcluster *Erziehungswissenschaften*. Hinsichtlich des mittleren Alters unterscheiden sich die Ausgangs- und die Analytestichprobe nicht ($t(144) = -1.93$, $p = .056$) und die Verteilung auf die Geschlechter ist ebenfalls nicht bedeutsam verschieden ($t(144) = 0.69$, $p = .489$).

5.3 Repräsentativität

Hinsichtlich des Alters, der Verteilung auf die Geschlechter und über die Fachcluster hinweg, habe ich die Repräsentativität der untersuchten Stichprobe für die Population der Studienanfänger im Wintersemester 2016/2017 geprüft. Dazu habe ich die beim statistischen Bundesamt verzeichneten Populationsstatistiken (Statistisches Bundesamt, 2019) als Referenz verwendet. Als Durchschnittsalter ist dort $M = 21.6$ Jahre verzeichnet (ohne Angabe der Standardabweichung). Das Durchschnittsalter der Analytestichprobe aus Jahrgang 1 liegt darunter, wobei diese Differenz signifikant ist ($t(144) = -6.14$, $p < .01$). Weiterhin ist beim statistischen Bundesamt für die Studienanfänger im Jahr 2016 ein Frauenanteil von 50.5% verzeichnet. In der vorliegenden Stichprobe liegt dieser weit höher. Was bereits rein nominell erkennbar ist, wird durch einen t -Test bestätigt: $t(144) = 9.33$, $p < .01$. Hinsichtlich des mittleren Alters und der Verteilung auf die Geschlechter ist die vorliegende Stichprobe aus Jahrgang 1 folglich nicht repräsentativ für die Zielpopulation. Dies ist vor dem Hintergrund, dass auch innerhalb der Population der Studierenden Frauen eher als Männer geneigt sind, an Studien teilzunehmen, wenig verwunderlich. So konnten Sax, Gilmartin und Bryant (2003) innerhalb eines großen, US-weiten Befragungsprojektes unter Studienanfängern zeigen, dass die Response-Rate unter Frauen sogar über verschiedene Erhebungsmodi (online versus Paper-Pencil) hinweg durchgängig fast doppelt so hoch war wie die Response-Rate der Männer.

An dieser Stelle möchte ich anmerken, dass in der ersten (und auch der zweiten) Längsschnittstichprobe das Studienfach der Teilnehmer mittels eines Freitextitems (*Welches Fach studieren Sie aktuell?*) erfragt wurde. Die Clusterung habe ich später manuell vorgenommen. Die oben beschriebene Clusterung entspricht nicht der, die beim Statistischen Bundesamt verwendet wird. Aufgrund der Tatsache, dass mir aber die Angaben zu den einzelnen Fächern vorlagen, konnte ich auch eine Clusterung nach dem Schema, welches in den offiziellen Statistiken verwendet wird, erstellen und dann einen χ^2 -Test zum Vergleich der vorliegenden Verteilung mit der Verteilung, die beim Statistischen Bundesamt für den Jahrgang 2016/2017 verzeichnet ist, berechnen. Das Ergebnis dieses Tests ist, dass ein signifikanter Unterschied vorliegt: $\chi^2(8) = 225.16$, $p < .01$.

Dies ist schon anhand der Tatsache, dass in der Stichprobe aus Studie 1 keine Studienanfänger der Ingenieurwissenschaften enthalten sind, zu erwarten gewesen. Die Gruppe der Studienanfänger in den Ingenieurwissenschaften macht nämlich einen beträchtlichen Teil der Population der Studienanfänger in Deutschland aus: 28.1%. Es war keineswegs beabsichtigt, diese Gruppe von Studienanfängern nicht in die Stichprobe aufzunehmen, sondern ist vermutlich dadurch zu Stande gekommen, dass ich bei der Rekrutierung nicht explizit darauf geachtet habe, diese Gruppe zu adressieren und somit vornehmlich zwar fächerübergreifend, aber offensichtlich nicht in ausreichendem Maße gezielt an Hochschulen, an denen das Studium der Ingenieurwissenschaften möglich ist, rekrutiert habe. An dieser Stelle möchte ich bereits anmerken, dass dies für alle drei Längsschnittstichproben gilt.

Insgesamt ist die Stichprobe aus Jahrgang 1 folglich nicht repräsentativ für die Zielpopulation in Bezug auf die oben genannten Aspekte. Sie kann aufgrund der Vielzahl an vertretenen Fächern jedoch durchaus einen fächerübergreifenden Einblick in die Zielpopulation geben. Somit ist die vorliegende Stichprobe ähnlich wie die in der quantitativen Teilstudie des *StufHe*-Projektes (Bosse et al., 2019) als eingeschränkt repräsentativ einzustufen. Nicht gänzlich repräsentative Stichproben finden sich in den meisten der bisher durchgeführten Studien in diesem Kontext, wobei der Großteil deutlich problematischere Abweichungen von der Repräsentativität aufweist.

5.4 Erhobene Variablen in Studie 1

Die erhobenen Variablen wurden bereits im vorherigen Kapitel einmal an zentraler Stelle für die drei Längsschnittstudien dargestellt, weshalb an dieser Stelle auf Kapitel 4.1.1 verwiesen sei.

5.5 Deskriptive Statistiken

Tabelle 5.1 zeigt die deskriptiven Statistiken aller im Rahmen der Modellfittests verwendeten Variablen (Mittelwert, Standardabweichung, Schiefe und Kurtosis jeweils mit Standardfehler). Sofern eine Skala bestehend aus mehreren Items verwendet wurde, wird auch ein Reliabilitätsschätzer (Cronbachs α) für die jeweilige Skala berichtet.

Tabelle 5.1: Mittelwert, Standardabweichung sowie Schiefe, Kurtosis und ggf. Reliabilität aller verwendeten Variablen in Studie 1 ($N = 145$)

Variable (Anzahl Items)	M	SD	Schiefe (SE)	Kurtosis (SE)	Cronbachs α
Bildungsniveau d. Mutter (1)	3.09	1.36	.59 (.20)	.60 (.40)	-
Abiturnote (1)	2.22	0.63	.12 (.20)	-.84 (.40)	-
Selbstwirksamkeit (13)	3.44	.47	-.39 (.20)	.71 (.40)	.80
Informiertheit (1)	4.19	1.53	.16 (.20)	-.09 (.40)	-
Integration (3)	5.38	0.85	-1.37 (.20)	1.22 (.40)	.53
Zielsetzung (13)	49.01	6.44	-.12 (.20)	.24 (.40)	.83
soziale Unterstützung (8)	29.34	3.62	-1.57 (.20)	1.71 (.40)	.92
Ergebniserwartungen (6)	26.87	3.03	-1.35 (.20)	3.15 (.40)	.72
P-U-Fit (9)	44.79	8.25	-.78 (.20)	1.06 (.40)	.88
Selbstwirksamkeit (13)	3.74	0.49	-1.20 (0.20)	5.89 (0.40)	.84
Stress (6)	12.22	4.54	.12 (.20)	.05 (.40)	.82
Zufriedenheit (3)	3.44	.95	-.48 (.20)	-.40 (.40)	.81
Wert (5)	30.37	3.43	-.20 (.29)	-1.04 (.56)	.83
Lernklima (8)	27.11	4.28	-.21 (.20)	.40 (.40)	.77
Handlungen (13)	48.33	6.25	-.12 (.20)	6.23 (.40)	.80
Selbstwirksamkeit (13)	3.83	0.48	-.36 (.20)	.45 (.40)	.85
Persistenz (1)	12.98	8.21	1.05 (.20)	.82 (.40)	-
Integration (3)	3.86	.44	-.32 (.20)	9.45 (.40)	.76
Zufriedenheit (3)	3.50	1.00	-.38 (.20)	-.62 (.40)	.84
Abbruchvorhaben (2)	2.90	2.04	3.40 (.20)	13.23 (.40)	-
intellektuelle Entw. (1)	5.07	1.10	-.65 (.20)	.95 (.40)	-
institutionelles Commitment (8)	4.31	.73	.05 (.20)	-.38 (.40)	.66
UCB (27)	3.97	0.54	.25 (.20)	.38 (.40)	.82
aktuelle Note (1)	10.02	2.81	-.97 (.20)	.40 (.40)	-

5.6 Unterschiede zwischen den Fachclustern?

Im Hinblick auf die beiden vorherzusagenden Kriterien (aktuelle Studiennote und Abbruchintention) habe ich Mittelwertsvergleiche zwischen den Fachclustern berechnet. Ein Blick auf die Literatur zeigt deutlich, dass sich die Abbruchquoten zwischen den Fachclustern unterscheiden (z.B. Heublein et al., 2017; Wissenschaftsrat, 2012). Ähnliche Befunde liegen auch für die Studiennoten vor (Dahm & Kerst, 2016). Wäre das auch in der vorliegenden Stichprobe der Fall, wäre eine Zentrierung der Variablen pro Fachcluster sinnvoll, um die im Fokus stehenden Zusammenhänge bereinigt um Unterschiede, die aus den (Benotungs-)Kulturen in den einzelnen Fächern resultieren, analysieren zu können.

Die ANOVA zum Kriterium *Abbruchintention* wurde ebenso wenig signifikant wie die ANOVA zum Kriterium *aktuelle Studiennote*: $F_{Note}(4) = 1.005$ ($p = .407$) und $F_{Abbruch}(4) = .601$ ($p = .663$). Folglich war eine Zentrierung dieser Variablen pro Fachcluster nicht notwendig.

5.7 Korrelative Zusammenhänge

Bevor die Ergebnisse der Pfadanalysen dargestellt werden, möchte ich zunächst auf die Tabellen im Anhang verweisen (13.5 ff.), in denen die bivariaten Korrelationen nullter Ordnung aufgeführt sind. Da eine Tabelle mit allen Variablen, die in mindestens einem der Modelle enthalten sind, so groß wäre, dass sie nicht mehr übersichtlich wäre, sind jeweils nur die Korrelationen aller in einem Modell berücksichtigten Variablen in einer separaten Tabelle aufgeführt.

5.8 Modelltests

Die folgenden Abbildungen (5.2 bis 5.5) zeigen die Pfadmodelle zu den vier Modellen mit standardisierten Pfadkoeffizienten, empirisch getestet mit den Daten des ersten Jahrgangs. Insgesamt möchte ich an dieser Stelle anmerken, dass signifikante Pfade kenntlich gemacht sind. Teilweise liegen die p -Werte mancher Pfade jedoch auch nur marginal über dem Signifikanzniveau von $\alpha = .05$. Dies gilt insbesondere für das Modell von Neuville et al. (2007). Das erklärt den teilweise guten Fit der Modelle trotz der geringen Anzahl an als signifikant markierten Pfaden. Der Vollständigkeit halber finden sich im Anhang (Tabelle 13.20 ff.) für alle drei Längsschnittstudien für alle Modelle Tabellen, in denen die Pfadkoeffizienten mit zugehörigem Standardfehler der Schätzung und p -Wert berichtet werden.

Eine genau Inspektion von Abbildung 5.2 zeigt, dass das Modell von Spady (1971) zwar zwei signifikante Pfade aufweist, insgesamt aber sehr viele nicht-signifikante Pfade hat. Bedeutsam wurde im vorderen Teil des modellierten Prozesses lediglich der Zusammenhang des Person-University-Fit

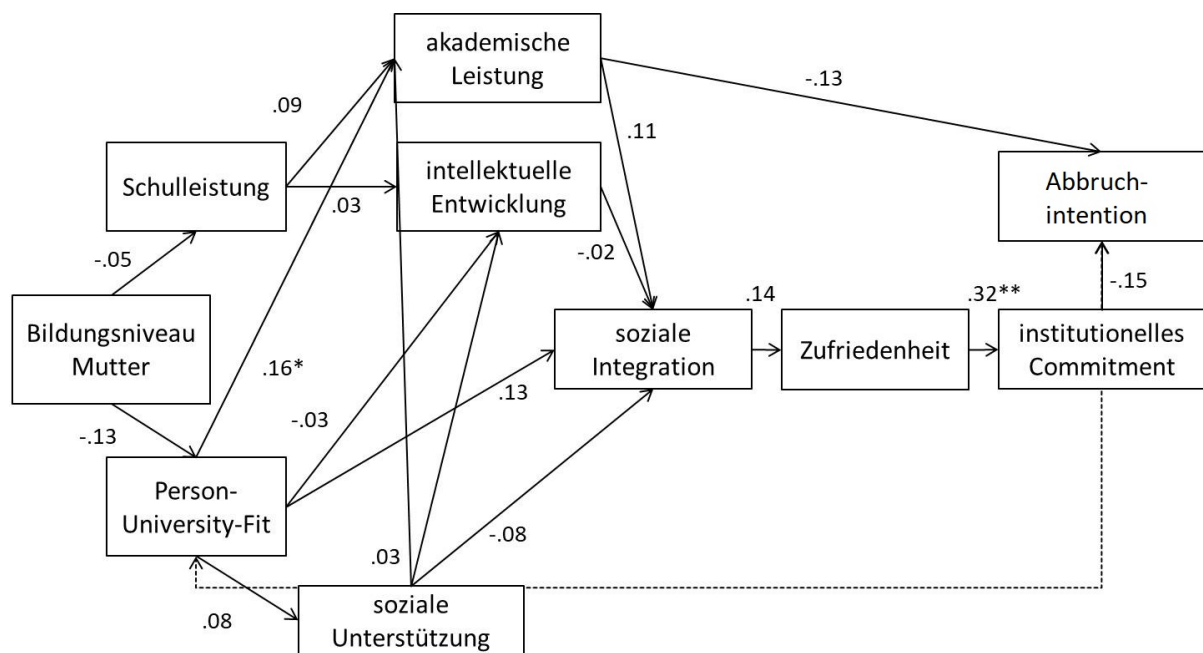


Abbildung 5.2: Spadys Modell mit standardisierten Pfadkoeffizienten, * $p < .05$, ** $p < .01$.

mit der akademischen Leistung (= aktuelle Studiennote) und im hinteren Teil des Prozesses der Zusammenhang der Studienzufriedenheit mit dem institutionellen Commitment.

Tabelle 5.2: Modellfitindizes für die vier getesteten Modelle [Jahrgang 1]

	df	χ^2	χ^2/df (p -value)	CFI	AGFI	RMSEA [$CI_{90\%}$]	SRMR
Spadys Modell	29	78.35	2.70 ($p < .01$)	.33	.82	.11 [.08;.14]	.10
Tintos Modell	15	23.13	1.54 ($p = .08$)	.72	.91	.06 [.00;.11]	.06
Neuville et al.'s Modell	20	18.13	0.91 ($p = .58$)	1.00	.94	.00 [.00;.06]	.05
Lent & Browns Modell	19	35.79	1.88 ($p < .01$)	.89	.89	.08 [.04;.12]	.07

Insbesondere das Kernstück des Modells, die Pfade zwischen sozialer Unterstützung, akademischer Leistung und sozialer Integration und die daran anschließenden Zusammenhänge mit der Studienzufriedenheit und der Abbruchintention wurden nicht signifikant. Wie Tabelle 5.2 zeigt, sind die Modellfitindizes eindeutig so zu interpretieren, dass das Modell die Daten aus dieser Stichprobe nicht gut abbildet. Besonders auffällig ist der niedrige CFI, der gemäß gängigen Cut-off-Werten (Hu & Bentler, 1998, 1999; Hooper et al., 2008) für einen guten Modellfit deutlich zu niedrig ist.

Ein ähnliches Bild zeigt sich hinsichtlich des Modells von Tinto (1975), welches in Abbildung 5.3 dargestellt ist. Viele Pfade sind nicht signifikant, nur drei weisen Signifikanz auf. In diesem Modell ist der Pfad von der akademischen Integration zum institutionellen Commitment signifikant. Weiter

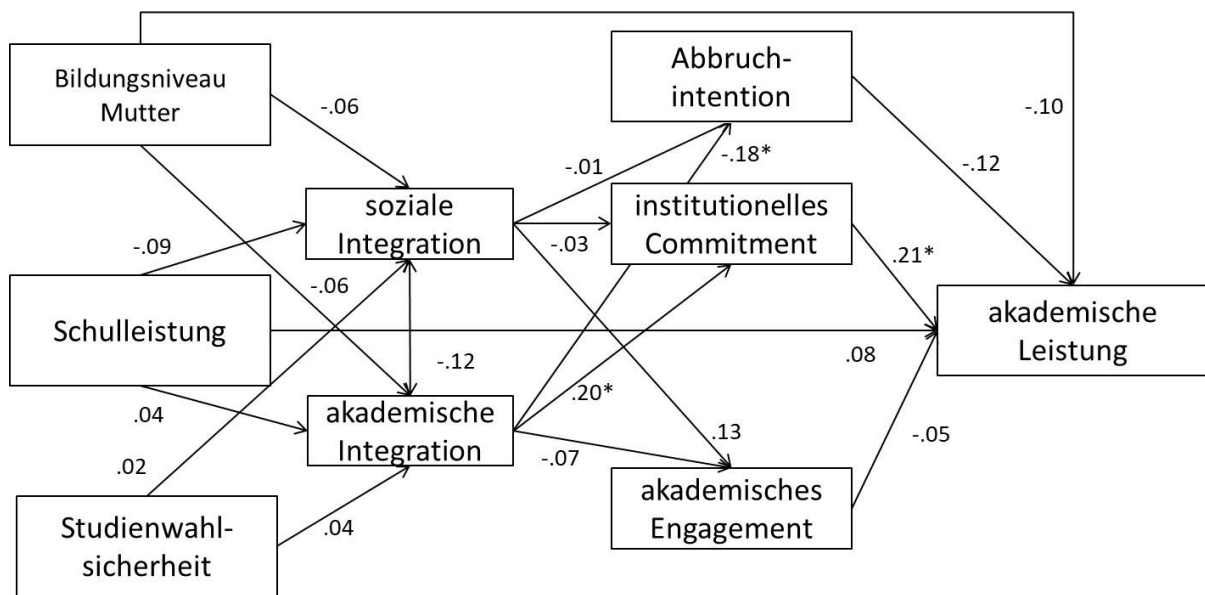


Abbildung 5.3: Tintos Modell mit standardisierten Pfadkoeffizienten, * $p < .05$.

weist das institutionelle Commitment dann einen bedeutsamen Pfad zur akademischen Leistung und der Abbruchintention auf. Insgesamt ist festzuhalten, dass auch das Modell von Tinto (1975) die

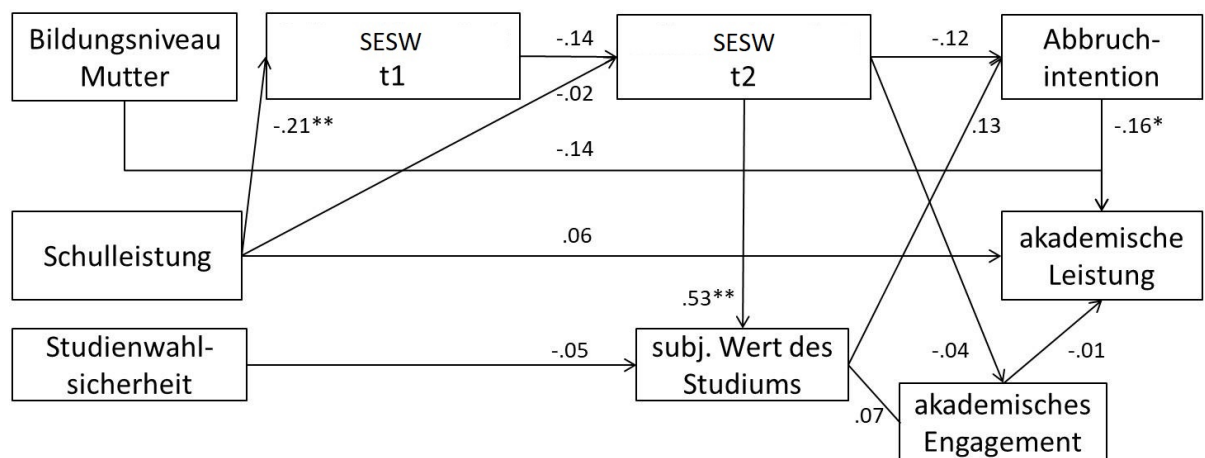


Abbildung 5.4: Neuville et al.'s Modell mit standardisierten Pfadkoeffizienten, * $p < .05$, ** $p < .01$.

Daten der Stichprobe nicht adäquat abbilden kann (vgl. Tabelle 5.2).

Im Modell von Neuville et al. (2007) sind zwar auch nur drei Pfade signifikant geworden (vgl. Abbildung 5.4), zwei davon jedoch sehr signifikant ($p < .01$). Der Pfad von der Schulleistung auf die Selbstwirksamkeit (SESW), erhoben zum ersten Messzeitpunkt, ist sehr signifikant. Weiter ist der

Pfad zwischen der SESW, die am zweiten Messzeitpunkt berichtet wurde, und dem Wert, der dem Studium beigemessen wird, sehr signifikant. Zudem ist der Pfad von der Abbruchintention auf die akademische Leistung signifikant.

Auffällig ist dabei, dass insbesondere das Kernstück dieses Modells, die Erwartungs-Wert-Modellstruktur (SESW = Selbstwirksamkeit(-serwartung) und Wert) bedeutsame Pfadkoeffizienten ausweist (vgl. Abbildung 5.4). Der erste beschriebene Pfad spiegelt zudem das wieder, was in den Kapiteln 2.1.6 und 2.3.3 bereits beschrieben wurde: Selbstwirksamkeit wird beeinflusst von vorangegangenen Erfolgs- und Misserfolgserfahrungen. Die kontextspezifische SESW wird folglich durch vorangegangene Erfahrungen im korrespondierenden Leistungskontext der Schule beeinflusst. Leistungserfolge in der Schule wiederum dürften sich in der Schulleistung (operationalisiert als Abiturnote) widerspiegeln. Betrachtet man die Modellfitindizes, so wird deutlich, dass das Modell von Neuville et al. (2007) einen sehr guten Modellfit aufweist (vgl. Tabelle 5.2), gemessen an gängigen Cut-off-Werten für guten Modellfit.

Das vierte hier getestete Modell schließlich, das CSM-Modell nach Lent und Brown (2013), weist zwei sehr signifikante Pfade auf (vgl. Abbildung 5.5). Wie auch im Modell von Neuville et al. (2007) ist der

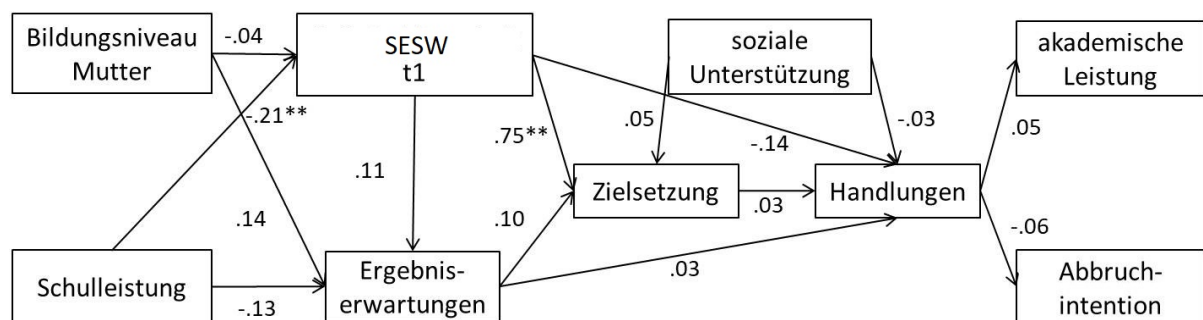


Abbildung 5.5: CSM-Modell nach Lent und Brown (2013) mit standardisierten Pfadkoeffizienten, $**p < .01$.

Pfad von der Schulleistung zur SESW signifikant. Von der SESW wiederum ausgehend ist der Pfad zur Zielsetzung signifikant. Hier zeigt sich wie schon beim Modell von Neuville et al. (2007), dass die Zusammenhänge der SESW mit anderen Konstrukten aus dem motivationalen Bereich bedeutsam sind. Insgesamt ist der Modellfit des Modells nach Lent und Brown (2013) jedoch nur als befriedigend einzustufen (vgl. Tabelle 5.2).

5.9 Zusammenfassung der Ergebnisse aus der ersten Längsschnittstudie

Die Modellfitanalysen anhand der Daten aus der ersten Längsschnittstudie zeigen, dass die vier Modelle die Zusammenhänge in den Daten aus dem ersten Jahrgang unterschiedlich gut abbilden. Den

Tabelle 5.3: Aikies Informations Criterion (AIC) für die vier getesteten Modelle [Jahrgang 1]

Modell	AIC
Spadys Modell	5445.26
Tintos Modell	5216.73
Neuville et al.'s Modell	4957.21
Lent & Browns Modell	6493.78

schlechtesten absoluten Modellfit wies das Modell von Spady (1971) auf, den besten wies das Modell von Neuville et al. (2007) auf. Auffällig ist, dass insbesondere nur die Modelle, in denen die SESW berücksichtigt wird (Modelle von Neuville et al., 2007 und Lent und Brown, 2013), gute bis befriedigende Modellfits aufweisen.

Man könnte auch sagen: Die (historisch) neueren Modelle weisen bessere Modellfitindizes auf bei der Beschreibung der vorliegenden Daten der Studienanfänger in Deutschland aus dem Jahrgang 2016/2017. Diese beiden Modelle zeichnen sich auch dadurch aus, dass zumindest einige Prädiktoren, die gemäß aktuellen Metaanalysen (vgl. Kapitel 2.2) gut geeignet sind, um Studienerfolg und Studienabbruch vorherzusagen, in diesen Modellen Berücksichtigung finden. Zu nennen wären in erster Linie die Selbstwirksamkeit und die Zielsetzung.

Ein direkter, numerischer Vergleich der berichteten absoluten Modellfitindizes unterschiedlich komplexer Modelle anhand der in Tabelle 5.2 aufgeführten Modellfitindizes ist nicht anzuraten.

Ein solcher Vergleich sollte nur mit Modellfitindizes erfolgen, die dafür geeignet sind (Kadane & Lazar, 2004; Vrieze, 2012). Hierfür kommen die *information criteria*, die die Modellkomplexität berücksichtigen, wie beispielsweise das AIC oder BIC in Frage (ein niedrigerer Wert indiziert einen besseren Modellfit). Ein solcher Vergleich ergibt hier (vgl. Tabelle 5.3), dass das Modell von Neuville et al. (2007) den anderen drei Modellen hinsichtlich des Modellfits überlegen ist. Inwiefern dies über verschiedene Jahrgänge von Studienanfängern generalisierbar ist, sollten die zweite und dritte Längsschnittstudie zeigen (vgl. Kapitel 7 und 8).

Zusammenfassend lässt sich sagen, dass neben sozialen Aspekten, wie sie in den Modellen von Spady (1971) und Tinto (1975) aufgeführt sind, auch im weitesten Sinne motivationale Aspekte wichtig zu sein scheinen. Das belegt die Bedeutung der Theorien zu Erwartungs-Wert-Modellen und sozial-kognitiven Modellen zur Vorhersage von aktueller Studienleistung und Abbruchintention.

Aufgrund der Tatsache, dass sich in den verschiedenen Modellen unterschiedliche Pfade als bedeutsam erwiesen, lag es nahe, ein neues Modell zu entwickeln, das die „gut funktionierenden“ Zusammenhänge in den hier beschriebenen vier etablierten Modellen aufgreift und zusätzlich weitere Prädiktoren, deren

Validität bei der Vorhersage von Studienerfolg und Studienabbruch vielfach belegt wurde (vgl. Kapitel 2.2), einbezieht.

6 Entwicklung eines neuen, integrativen Modells

Anknüpfend an die im vorherigen Kapitel dargestellten Befunde lag es nahe, ein neues Modell zu entwickeln, welches einerseits aktuelle (metaanalytisch gesicherte) Befunde zu Prädiktoren von *Studienerfolg und Studienabbruch* und andererseits die in Studie 1 gewonnenen empirischen Erkenntnisse berücksichtigt.

Im Folgenden stelle ich dar, wie ich ein solches, integratives (= verschiedene Erkenntnisse integrierendes) Modell entwickelt habe. Zudem berichte ich, welche Modellfitindizes sich beim empirischen Testen des neuen Modells anhand der Daten aus Studie 1 ergaben.

6.1 Aspekte, die im neuen Modell berücksichtigt werden sollten

Wie bereits durch die Ergebnisse aus Studie 1 erkennbar wurde, schienen folgende Punkte für die Entwicklung des neuen Modells relevant:

- *Zufriedenheit* sollte berücksichtigt werden (siehe Spady, 1971)
- *Commitment* zum Studium bzw. zur Hochschule sollte einbezogen werden, wie im Modell von Tinto (1975)
- *SESW* sollte in dem neuen Modell enthalten sein (siehe Modell von Neuville et al., 2007), gegebenenfalls sogar eine Messwiederholung für die SESW
- weitere (im weitesten Sinne) motivationale Aspekte sollten abgebildet werden (siehe *Zielsetzung* im Modell nach Lent und Brown, 2013)

Dass die Studienzufriedenheit, wie sie im Modell von Spady (1971) berücksichtigt wird, im neuen Modell aufgeführt werden soll, ist aus meiner Sicht deshalb besonders wichtig, weil dadurch im Prinzip ein vorgelagertes Kriterium des Studienerfolgs thematisiert wird. Ich würde an dieser Stelle von einem „Zwischen-Outcome“ sprechen: Satt nur am Ende der Studieneingangsphase nach Studienerfolg und Abbruchintention zu fragen, sollte auch bereits im Verlauf der ersten beiden Semester ein Augenmerk darauf gelegt werden, wie es den Studierenden geht. Die Zufriedenheit im Studium erwies sich in vorangegangenen Studien als prädiktiv für Studienerfolg und Studienabbruch (Hiemisch et al., 2005; Hasenberg, 2012). Die Höhe der Pfadkoeffizienten im Modell von Spady (1971), die anhand der Daten aus Studie 1 ermittelt wurden, deutet ebenfalls darauf hin, dass die Studienzufriedenheit berücksichtigt werden sollte.

Den zweite Aspekt in der oben stehenden Aufzählung, die Berücksichtigung des *Commitments*, welches angelehnt an das Modell von Tinto (1975) aufgenommen finden soll, wollte ich allerdings gerne besonders umfassend abbilden, indem ich das Konstrukt des UCB (University Citizenship Behavior) einbeziehe, welches erst lange nach der Veröffentlichung des Modells von Tinto (1975) durch die Arbeit von Gehring (2006) im Hochschulkontext Einzug gehalten hat.

Der dritte Aspekt in der oben stehenden Auflistung spiegelt auch diverse metaanalytische Befunde wider, die bereits in Kapitel 2.2 dargestellt wurden. Dass die SESW sogar zweimal anstatt nur einmal im Laufe des Studieneinstieg erfasst und ins Modell einbezogen werden sollte, ist einerseits sinnvoll, da dieses Konstrukt, wie in Kapitel 3 ausgeführt, einem ständigen Wandel unterliegt und andererseits die Pfadkoeffizienten, die anhand der Daten aus Studie 1 für das Modell von Neuville et al. (2007) ermittelt wurden, zeigen, dass damit – zumindest in den Daten aus Studie 1– vorhandene Zusammenhänge sinnvoll abgebildet werden können.

Anknüpfend an das Modell nach Lent und Brown (2013) lässt sich zudem festhalten, dass die Zielsetzung, welche als Variable aus dem motivationalen Kontext zu klassifizieren ist, ebenfalls berücksichtigt werden soll. Dies wird unterstrichen durch eine Vielzahl an vorherigen Untersuchungen und metaanalytischen Studien (vgl. Kapitel 2.1.8 und 2.2).

Ausgehend von diesen Punkten konzipierte ich das neue Modell, welche in Abbildung 6.1 dargestellt ist. Dieses Modell berücksichtigt die aufgezeigten Punkte, indem es individuumszentriert den Prozess des Studieneinstiegs skizziert mit Fokus auf das individuelle Erleben in dieser Phase. Aus diesem Grund möchte ich das neue Modell im Folgenden als das *Erlebens-Orientierte-Studieneinstiegs-Modell* (das EOS-Modell) bezeichnen.

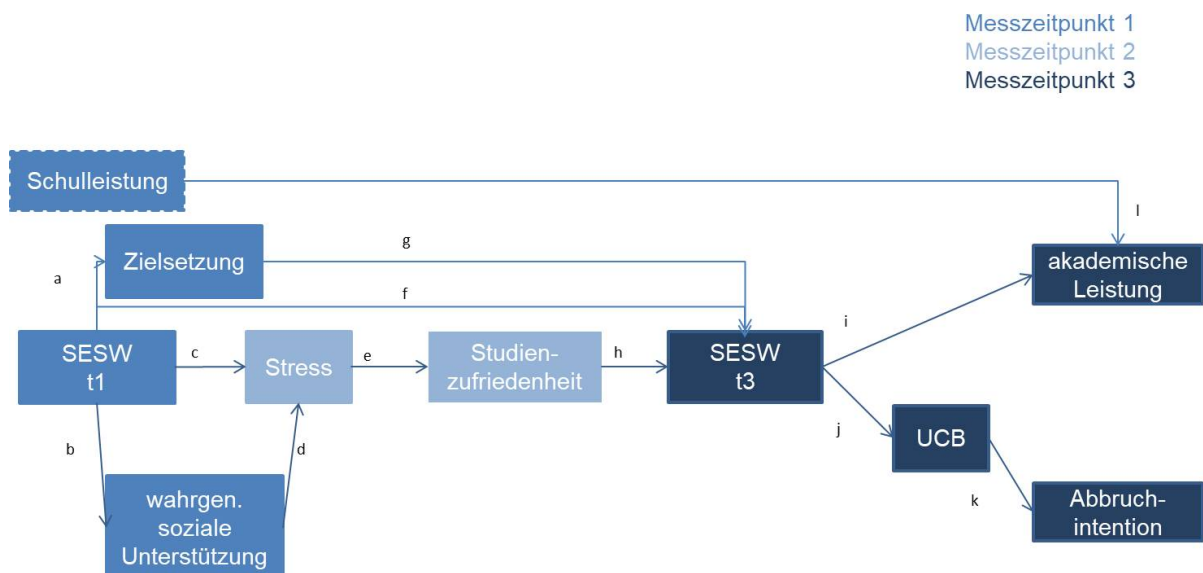


Abbildung 6.1: Das EOS-Modell

6.2 Das EOS-Modell

Dieses neue, integrative Modell möchte ich in diesem Abschnitt im Detail vorstellen. Dazu werde ich den abgebildeten Prozess einmal von Beginn bis Ende der Studieneinstiegsphase „entlang gehen“ und die spezifizierten Pfade erläutern. Der Übersichtlichkeit wegen werde ich dabei die in Abbildung 6.1 aufgeführten Bezeichnungen der Pfade verwenden (Pfad a, b, c, etc.) und zu jedem Pfad kurz zu erläutern, welche Bedeutung er hat.

An dieser Stelle möchte ich anmerken, dass hier und im Folgenden die grafische Darstellung des EOS-Modells stets farbig erfolgt, um die Bedeutung dieses Modells im Rahmen der vorliegenden Arbeit zu unterstreichen. Die Farbabstufungen kodieren zudem, zu welchem Zeitpunkt innerhalb des Studieneinstiegs das jeweilige Konstrukt erfasst werden sollte: Messzeitpunkt t1 ist zu Beginn des ersten Semesters, Messzeitpunkt t2 ist am Ende ersten Semesters und Messzeitpunkt t3 ist am Ende des zweiten Semesters.

- Pfad a: Wer eine höhere SESW hat, setzt sich höhere Ziele.
- Pfad b, c und d: Zudem führt eine höhere SESW auch zu einer höheren wahrgenommenen sozialen Unterstützung. Je höher die SESW zu Studienbeginn, desto niedriger der empfundene Stress. Dies ist der direkte Zusammenhang dieser beiden Konstrukte. Die wahrgenommene soziale Unterstützung sollte ebenfalls negativ mit dem wahrgenommenen Stress assoziiert sein. Somit stellt die Verknüpfung von Pfad b und d eine Mediation des direkten Effekts der SESW auf den Stress (Pfad c) dar: Eine höhere SESW führt zu einer höheren wahrgenommenen sozialen Unterstützung, welche Stress reduzieren kann.
- Pfad e: Ein niedrigeres Stresslevel geht mit einer höheren Studienzufriedenheit einher.
- Pfad f: Die SESW zu Beginn ist mit der SESW zum Ende der Studieneinstiegsphase positiv assoziiert. Dies ergibt sich bereits aus der Definition der SESW als dynamisches Konstrukt, welchen dem ständigen Wandel unterliegt. Im Kontext der Studieneinstiegsphase ist es das Erleben des Studienalltags, welches die fortlaufende Aktualisierung der SESW maßgeblich beeinflusst.
- Pfad g: Die Zielsetzung zu Studienbeginn beeinflusst die SESW zum Ende der Studieneinstiegsphase: Wer sich bessere (höhere) Ziele setzt, dessen SESW ist zum Ende der Studieneinstiegsphase im Allgemeinen auch höher.
- Pfad h: Die Studienzufriedenheit ist positiv mit der SESW zum Ende des Studieneinstiegs assoziiert. Wer zufriedener im Studium ist, traut sich auch die Bewältigung der auftretenden akademischen Herausforderungen eher zu.
- Pfad i, j und k: Von der SESW geht ein direkter Pfad zur akademischen Leistung und zum UCB, von welchem aus ein Pfad zur Abbruchintention führt. Das bedeutet, dass Personen

mit höherer SESW zum Ende des Studieneinstiegs in der Regel nicht nur die besseren akademischen Leistungen erbringen, sondern im Mittel auch ein höheres UCB und eine niedrigere Abbruchintention aufweisen.

- Pfad I: Zuletzt ist der klassische Prädiktor für akademische Leistung, die Schulleistung, ebenfalls im Modell integriert.

Im Hinblick auf die in Kapitel 2 dargestellte (grobe) Kategorisierung von Prädiktorgruppen, die gemäß (metaanalytischen) empirischen Befunden bedeutsam bei der Vorhersage von *Studienerfolg und Studienabbruch* sind (vgl. Tabelle 2.1), lässt sich festhalten, dass das EOS-Modell Variablen aus allen Kategorien, die sich auf Attribute des Individuums (d.h. des Studienanfängers) beziehen, umfasst.

6.2.1 Einordnung in die Literatur

Wie in Kapitel 2 beschrieben, kann das Rahmenmodell von Heublein et al. (2017) dazu dienen, verschiedene empirisch testbare Modelle im Forschungsfeld einzuordnen. So möchte ich an dieser Stelle das EOS-Modell in diesem Rahmenmodell verorten. Von den drei zeitlichen Phasen, die Heublein et al. (2017) einteilen, ist das EOS-Modell in die mittlere Phase, die der *aktuellen Studiensituation* einzuordnen. Im Rahmenmodell steht dann der individuelle Studienprozess in der Mitte des Schemas. Zu diesem Prozess gehören gemäß Heublein et al. (2017) vier große Bereiche: Studienverhalten, Studienmotivation, Studienleistung und psychische/ physischer Ressourcen (S.12, Heublein et al., 2017). Alle vier Bereiche sind im EOS-Modell berücksichtigt. Daher lässt sich festhalten, dass das EOS-Modell als individuumszentriertes Modell alle gemäß Heublein et al. (2017) für diese Perspektive wichtigen Aspekte umfasst.

6.3 Erster empirischer Test des EOS-Modells

Abbildung 6.2 zeigt das neue Modell mit standardisierten Pfadkoeffizienten, die anhand der Daten aus Studie 1 ermittelt wurden. Wie die absoluten Modellfitindizes zeigen, weist es einen sehr guten Fit auf (vgl. Tabelle 6.1). In Ergänzung zu Tabelle 6.1 möchte ich die AICs für alle vier etablierten Modelle und das EOS-Modell an dieser Stelle gesammelt darstellen (vgl. Tabelle 6.2). Ein Vergleich des AICs des EOS-Modells mit denen der anderen Modelle zeigt deutlich, dass das EOS-Modell den Modellen von Spady (1971), Tinto (1975) und Lent und Brown (2013) vorzuziehen ist. Lediglich das Modell von Neuville et al. (2007) ist ein geringfügig niedrigerer AIC zu verzeichnen.

Ob diese nominell kleine Differenz statistisch bedeutsam ist, kann ein χ^2 -Differenz-Test für die Modellfits dieser beiden Modelle zeigen. Dieser fällt jedoch nicht signifikant aus: $\delta_{\chi^2} = 17.22$, $p = .24$.

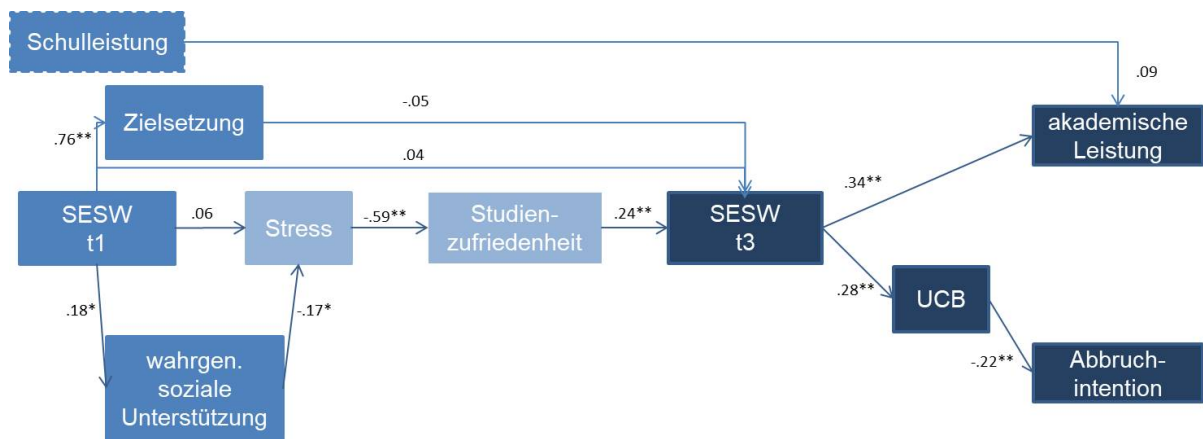


Abbildung 6.2: Neues Modell J1

Tabelle 6.1: Modellfitindizes für die vier getesteten Modelle [Jahrgang 1]

	df	χ^2	χ^2/df (p -value)	CFI	AGFI	RMSEA [$CI_{90\%}$]	sRMR
Spadys Modell	29	78.35	2.70 ($p < .01$)	.33	.82	.11 [.08;.14]	.10
Tintos Modell	15	23.13	1.54 ($p = .08$)	.72	.91	.06 [.00;.11]	.06
Neuville et al.'s Modell	20	18.13	0.91 ($p = .58$)	1.00	.94	.00 [.00;.06]	.05
Lent & Browns Modell	19	35.79	1.88 ($p < .01$)	.89	.89	.08 [.04;.12]	.07
EOS-Modell	32	35.35	1.10 ($p = .31$)	.99	.93	.03 [.00;.07]	.06

Daraus kann man schließen, dass das Modelle von Neuville et al. (2007) und das EOS-Modell die Datenstruktur aus Jahrgang 1 gleich gut abbilden können.

6.4 Der Mehrwert des EOS-Modells

Wie bereits in der Einleitung dieses Kapitels beschrieben, ist der Mehrwert des EOS-Modells gegenüber den vier hier beleuchteten etablierten Modellen, dass es auf einer breiteren theoretischen Basis steht: Im EOS-Modell werden sowohl Aspekte der sozialen Herkunft, der Schulleistung und der SESW zu Studienbeginn, sowie das Zielsetzungsverhalten und die wahrgenommene soziale Unterstützung berücksichtigt. Weiter umfasst es auch den erlebten Stress und die Studienzufriedenheit in der Prozessmitte. Zum Abschluss des Studieneinstiegs schließlich spielen die dann vorliegende SESW sowie das UCB die entscheidenden Rollen als Vermittler des Einflusses der vorherigen Variablen auf die beiden Kriterien *Studienerfolg* und *Studienabbruch*. Damit berücksichtigt es alle empirisch im Rahmen von Studie 1 als relevant identifizierten Aspekte sowie zusätzlich die Idee, die SESW wie-

Tabelle 6.2: Aikies Informations Criterion (AIC) für die vier getesteten Modelle [Jahrgang 1]

Modell	AIC
Spadys Modell	5445.26
Tintos Modell	5216.73
Neuville et al.'s Modell	4957.21
Lent & Browns Modell	6493.78
EOS-Modell	4962.40

derholt aufzuführen und das UCB als sehr umfassendes Konzept von Commitment einzubinden.

Anhand der Daten aus Studie 1 konnte bereits gezeigt werden, dass das *EOS*-Modell einen guten Modellfit aufweist. Zumal eben diese empirischen Daten beziehungsweise die auf ihnen basierenden Ergebnisse der Modellfitanalysen für die vier etablierten Modelle jedoch die Modellbildung maßgeblich prägten, war eine empirische Testung des *EOS*-Modells mit weiteren Daten (unabhängig von Studie 1) dringend notwendig, um die Güte des Modells zu prüfen. Zu diesem Zweck führte ich entsprechend die beiden weiteren Längsschnittstudie – analog zu Studie 1 aufgebaut – durch.

Im Zuge dieser beiden Folgestudien (dargestellt in den nächsten beiden Kapiteln) wurde also das exakt gleiche Design verwendet und mit den gleichen Syntaxen der Modellfit für die vier etablierten und das neue *EOS*-Modell ermittelt. Damit sollten Studie 2 und 3 im Sinne einer Replikation der Ergebnisse aus Studie 1 auch dazu dienen, dass Ergebnissen, die potentiell in allen drei Studien gleichermaßen erzielt wurden, eine hohe Generalisierbarkeit zugesprochen werden kann.

7 Studie 2

7.1 Messzeitpunkte bei Studie 2

Studie 2 ist die zweite Längsschnittstudie mit drei Messzeitpunkten über das erste Studienjahr hinweg. Im Gegensatz zu Studie 1 haben die Teilnehmer in Studie 2 alle ihr Studium im Wintersemester 2017/2018 begonnen, gehören also zum Folgejahrgang. Messzeitpunkt 1 war von Oktober bis November 2017 (Beginn erstes Semester), Messzeitpunkt 2 war im Februar 2018 und Messzeitpunkt 3 von Juni bis Juli 2018. Dabei waren die jeweiligen Online-Befragungen immer etwa einen Monat lang zugänglich.

Am ersten Messzeitpunkt nahmen $N = 323$ Personen teil. An Messzeitpunkt 2 waren es noch $N = 190$ Personen, was einem Dropout von ca. 41% entspricht. Zu Messzeitpunkt 3 waren noch 151 Personen dabei, womit wie in Studie 1 ein Dropout von etwas über 50% vorliegt.

7.2 Stichprobe bei Studie 2

Wie im Kapitel 4 beschrieben, wurde das Einschlusskriterium (es müssen Angaben zur aktuellen Studiennote und zur Abbruchintention vorliegen) angelegt. Ganze 32 Fälle erfüllten dieses Einschlusskriterium nicht, sodass eine Analysestichprobe aus dem ersten Jahrgang im Umfang von $N = 119$ vorlag. Dieser auffällig hohe Anteil an aussortierten Datensätzen resultiert vornehmlich daraus, dass aufgrund einer technischen Panne zum dritten Messzeitpunkt die Variable *aktuelle Studiennote* nicht abgefragt wurde. Ich musste folglich eine zusätzliche Befragung direkt anschließen, in der diese Variable dann nacherhoben wurde. Diese Nachbefragung lies ich besonders lange (30.07.-08.09) laufen, um noch von möglichst vielen Teilnehmern diese Information zu erhalten, da ihre Daten andernfalls gemäß des Einschlusskriteriums sowieso nicht in die Analysen hätten einbezogen werden können. So gesehen kann von einem sehr guten Rücklauf in dieser Nachbefragung gesprochen werden (angeschrieben wurden nur diejenigen, die bereits am regulären dritten Messzeitpunkt teilgenommen hatten).

Im Mittel waren die Personen, deren Daten letztendlich in die Analysestichprobe aus Jahrgang 2 einfließen, $M = 21.0$ ($SD = 4.4$) Jahre alt. 82.4% von ihnen gaben an, weiblich zu sein, 16.0% gaben an, männlich zu sein, die restlichen 1.7% machten keine Angabe zu ihrem Geschlecht. In dieser Studie sah die Verteilung auf die Fachcluster wie folgt aus: 29.4% Naturwissenschaften, 16.8% Medizin oder Psychologie, 14.3% Sprachen, Kulturen, Soziales, 26.1% Erziehungswissenschaften, 13.4%

Wirtschaftswissenschaften, Jura oder sonstige Fächer. Die Verteilung der Studierenden aus Jahrgang 2 auf die Fachcluster ist Abbildung 7.1 zu entnehmen. Vergleicht man die Ausgangsstichprobe

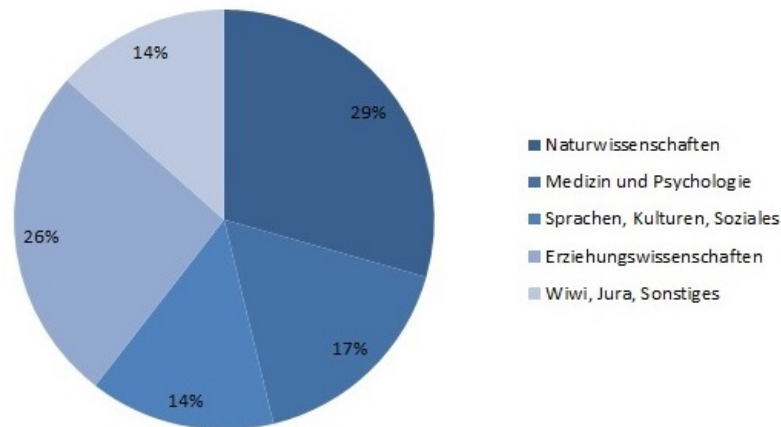


Abbildung 7.1: Verteilung auf die Fachcluster [Jahrgang 2, $N = 119$]

von $N = 323$ Studierenden mit der Analysestichprobe von $N = 119$, so zeigt sich kein signifikanter Unterschied hinsichtlich der Verteilung auf die Studienfachcluster ($\chi^2(4) = 1.51$, $p = .825$). Die Teilnehmer in der Analysestichprobe unterschieden sich auch nicht im Alter ($t(144) = 1.00$, $p = .321$) und in der Verteilung auf die Geschlechter ($t(118) = 0.75$, $p = .457$) bedeutsam von der Ausgangsstichprobe.

7.3 Repräsentativität

Hinsichtlich des Alters, der Verteilung auf die Geschlechter und über die Fachcluster hinweg, habe ich die Repräsentativität der untersuchten Stichprobe für die Population der Studienanfänger im Wintersemester 2017/2018 geprüft. Dazu habe ich die beim Statistischen Bundesamt verzeichneten Populationsstatistiken als Referenz verwendet (Statistisches Bundesamt, 2019). Als Durchschnittsalter (ohne Angabe der Standardabweichung) sind 21.6 Jahre verzeichnet. Das Durchschnittsalter der Analysestichprobe aus Jahrgang 2 liegt nur knapp darunter mit $M = 21.0$ Jahren ($SD = 4.4$), wobei diese Differenz nicht signifikant ist ($t(117) = -1.51$, $p = .134$). Die vorliegende Stichprobe aus Jahrgang 2 ist somit hinsichtlich des mittleren Alters mit der Zielpopulation vergleichbar. Beim statistischen Bundesamt ist für die Studienanfänger im Jahr 2017 ein Frauenanteil von 50.8% verzeichnet. In der vorliegenden Stichprobe liegt dieser weit höher: 82.4%. Was bereits rein nominell erkennbar ist, wird durch einen t -Test bestätigt: $t(118) = 9.601$, $p < .01$. Hinsichtlich der Verteilung auf die Geschlechter ist die vorliegende Stichprobe aus Jahrgang 2 folglich genau wie die aus Jahrgang 1 nicht repräsentativ für die Zielpopulation. Dies ist vor dem Hintergrund, dass Frauen im Allgemeinen eher als Männer geneigt sind, an Studien teilzunehmen, wenig verwunderlich (siehe die

Erläuterung dazu in Kapitel 5).

Da in der zweiten Längsschnittstichprobe (wie auch in der ersten) das Studienfach der Teilnehmer mittels eines Freitextitems (*Welches Fach studieren Sie aktuell?*) erfragt wurde, war es mir wieder möglich, auch eine Clusterung nach dem Schema, welches in den offiziellen Statistiken (Statistisches Bundesamt, 2018) verwendet wird, zu erstellen und dann einen χ^2 -Test zum Vergleich der vorliegenden Verteilung mit der Verteilung, die beim Statistischen Bundesamt für den Jahrgang 2017/2018 verzeichnet ist, zu berechnen. Das Ergebnis dieses Tests ist (wie auch in der ersten Längsschnittstichprobe), dass ein signifikanter Unterschied vorliegt: $\chi^2(8) = 306.49$, $p < .01$). Wie bereits im Kapitel zuvor beschrieben, führe ich dies vornehmlich darauf zurück, dass keine Studienanfänger der Ingenieurwissenschaften in der Stichprobe enthalten sind. Diese Gruppe von Studienanfänger machte allerdings wie auch im Jahrgang zuvor, so auch im Jahrgang 2017/2018, einen großen Teil der Population der Studienanfänger in Deutschland aus: 27.5%.

Insgesamt ist die Stichprobe aus Jahrgang 2 folglich nicht repräsentativ für die Zielpopulation in Bezug auf die oben genannten Aspekte. Sie kann aufgrund der Vielzahl an vertretenen Fächern jedoch durchaus einen fächerübergreifenden Einblick in die Zielpopulation geben und somit wie die Stichprobe aus Jahrgang 1 als eingeschränkt repräsentativ eingestuft werden.

7.4 Erhobene Variablen in Studie 2

Die erhobenen Variablen wurden bereits einmal an zentraler Stelle für die drei Längsschnittstudien dargestellt, weshalb an dieser Stelle auf Kapitel 4.1.1 verwiesen sei.

7.5 Deskriptive Statistiken

Tabelle 7.1 zeigt die deskriptiven Statistiken aller im Rahmen der Modellfit verwendeten Variablen (Mittelwert, Standardabweichung, Schiefe und Kurtosis jeweils mit Standardfehler).

7.6 Unterschiede zwischen den Fachclustern?

Im Hinblick auf die beiden vorherzusagenden Kriterien (aktuelle Studiennote und Abbruchintention) habe ich Mittelwertsvergleiche zwischen den Fachclustern berechnet. Wie schon für die erste Längsschnittstudie (Kapitel 5) sollte damit geprüft werden, ob sich die Fachcluster hinsichtlich der beiden Kriterien unterscheiden. Weder die ANOVA zum Kriterium *aktuelle Studiennote* noch die ANOVA zum Kriterium *Abbruchintention* wurde signifikant: $F_{Note}(4) = 2.164$ ($p = .077$) und $F_{Abbruch}(4) = .723$ ($p = .578$). Das bedeutet, auch für die Stichprobe aus Jahrgang 2 war keine Zentrierung der Variablen *Abbruchintention* und *aktuelle Studiennote* vorzunehmen.

Tabelle 7.1: Mittelwert, Standardabweichung sowie Schiefe, Kurtosis und ggf. Reliabilität aller verwendeten Variablen in Studie 2 ($N = 119$)

Variable (Anzahl Items)	M	SD	Schiefe (SE)	Kurtosis (SE)	Cronbachs α
Bildungslevel d. Mutter (1)	3.18	1.71	0.83 (0.22)	0.29 (0.44)	-
Abiturnote (1)	2.19	0.61	0.05 (0.22)	-0.76 (0.44)	-
Informiertheit (1)	4.19	1.35	-0.03 (0.22)	-0.54 (0.44)	-
Selbstwirksamkeit (13)	3.66	0.44	-0.29(0.22)	0.30(0.44)	.81
Integration (3)	5.46	0.86	-1.38 (0.22)	0.68 (0.44)	.61
Zielsetzung (13)	55.49	5.57	-0.13 (0.22)	-0.62 (0.44)	.82
Ergebniserwartungen (6)	26.54	2.29	-0.21 (0.22)	-0.29 (0.44)	.46
P-U-Fit (9)	46.35	7.57	-0.68 (0.22)	0.81(0.44)	.86
soziale Unterstützung (8)	28.87	4.44	-2.38 (0.22)	6.57 (0.44)	.94
Lernklima (8)	26.69	4.24	-0.41 (0.23)	-0.14 (0.46)	.78
Wert (5)	30.18	4.03	-1.04 (0.23)	1.12 (0.46)	.86
Stress (6)	11.85	4.26	0.39 (0.23)	0.15 (0.46)	.80
Selbstwirksamkeit (13)	3.74	0.45	-0.41 (0.23)	1.03 (0.46)	.84
Handlungen (13)	48.28	6.14	-0.52 (0.23)	1.12 (0.46)	.79
Zufriedenheit (3)	3.42	0.94	-.48 (0.23)	-.25 (0.46)	.82
Selbstwirksamkeit (13)	3.70	0.60	-0.96 (0.22)	1.85 (0.44)	.91
intellektuelle Entw. (2)	5.03	1.18	-0.75 (0.22)	0.55 (0.44)	-
Persistenz (1)	13.72	8.90	1.04 (0.22)	0.66 (0.44)	-
institutionelles Commitment (8)	4.40	0.64	-0.30 (0. 22)	0.11 (0.44)	.61
Integration	3.80	0.53	-2.60 (0.22)	5.67 (0.44)	.79
UCB (27)	3.99	0.56	-0.29 (0.22)	0.05 (0.44)	.84
aktuelle Note (1)	9.83	2.89	-1.05 (0.22)	1.09 (0.44)	-
Abbruchvorhaben (2)	2.92	1.80	2.19 (0.22)	4.14 (0.44)	-
Zufriedenheit (3)	3.28	0.89	-0.45 (0.22)	0.02 (0.44)	.76

7.7 Korrelative Zusammenhänge

Bevor die Ergebnisse der Pfadanalysen dargestellt werden, möchte ich zunächst auf die Tabellen im Anhang verweisen (13.10 ff.), in denen die bivariaten Korrelationen nullter Ordnung aufgeführt sind. Zwecks Übersichtlichkeit sind jeweils nur die Korrelationen aller in einem Modell berücksichtigten Variablen in pro Tabelle aufgeführt.

7.8 Modelltests

Analog zur Darstellung des vorherigen Kapitels, werde ich im Folgenden die Ergebnisse der empirischen Testung der vier etablierten Modelle anhand der Daten aus der zweiten Längsschnittstudie berichten. Anschließend werde ich zusätzlich den Modellfit des *EOS*-Modells berichten.

Die Abbildungen 7.2 zeigt das Pfadmodell nach Spady (1971). Auf den ersten Blick lässt sich erkennen, dass deutlich mehr Pfade signifikant wurden als beim Modelltest anhand der Daten aus dem ersten Jahrgang. Ganze sieben Pfade wiesen bedeutsam hohe Koeffizienten auf. Zu betonen ist hierbei, dass anders als bei Studie 1, im Kernstück des Modells viele Pfade signifikant wurde. So erwies sich der Pfad vom Person-University-Fit zur sozialen Integration und zur intellektuellen Entwicklung als bedeutsam, genau wie der Pfad vom Bildungsniveau der Mutter auf die Schulleistung und weiter zur akademischen Leistung (das negative Vorzeichen des Koeffizienten resultiert aus der entgegengesetzten Benotungsskala in der Schule versus der Hochschule; logisch ist der Zusammenhang positiv). Weiter wurden die Pfade von der sozialen Integration, über die Studienzufriedenheit auf das institutionelle Commitment signifikant.

Hier zeigte sich folglich, dass das Modell von Spady (1971) die Daten aus dem zweiten Jahrgang zumindest in der Mitte des abgebildeten Prozesses recht gut darstellen konnte. Die Modellfitindizes (vgl. Tabelle 7.2) dagegen bescheinigen dem Modell insgesamt keinen guten Fit. Das spricht dafür, dass wichtige Zusammenhänge, die sich im Datensatz finden lassen, in diesem Modell nicht berücksichtigt sind, sodass trotzdem sieben der modellierten Pfade signifikante Koeffizienten aufweisen, die Datenstruktur insgesamt nicht umfassend abgebildet wird.

Beim Modell von Tinto (1975) ist zunächst bei visueller Inspektion (Abbildung 7.3) des Pfadmodells ebenfalls festzustellen, dass in Studie 2 mehr Pfade signifikant wurden als in Studie 1. Immerhin fünf Pfade erwiesen sich als statistisch bedeutsam. So wurde der (bidirektionale) Pfad zwischen sozialer und akademischer Integration signifikant.

Weiter wurden die Pfade von der sozialen Integration und vom institutionellen Commitment zur Abbruchintention ebenso signifikant wie der Pfad von der akademischen Integration zum institutionellen Commitment und weiter zur akademischen Leistung. Somit ist auch für das Modell von Tinto (1975) festzuhalten, dass das Kernstück bedeutsame Zusammenhänge in der Datenstruktur abbilden kann.

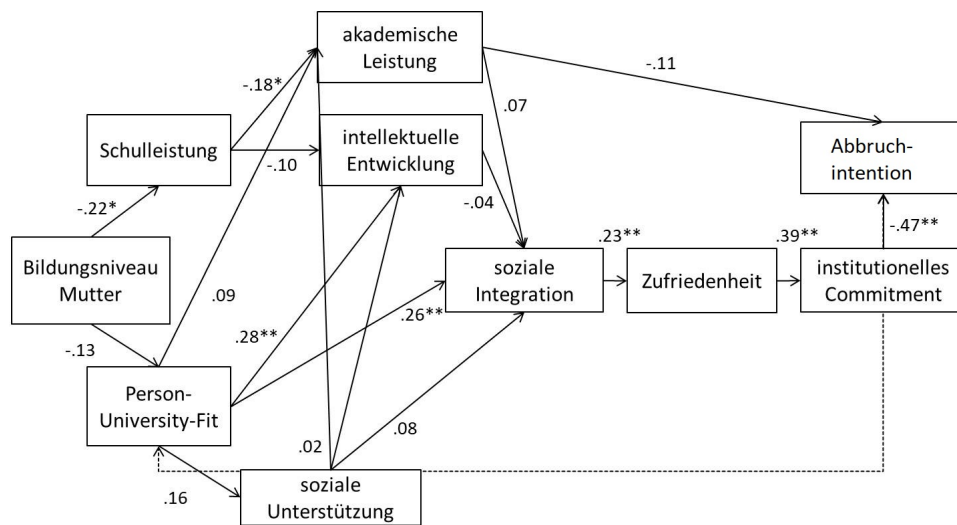


Abbildung 7.2: Spadys Modell mit standardisierten Pfadkoeffizienten *Anm.* * $p < .05$, ** $p < .01$.

Tabelle 7.2: Modellfitindizes für die fünf getesteten Modelle [Jahrgang 2]

	df	χ^2	χ^2/df (p -value)	CFI	AGFI	RMSEA [$CI_{90\%}$]	SRMR
Spadys Modell	29	99.42	3.43 ($p < .01$)	.55	.74	.14 [.11;.17]	.14
Tintos Modell	15	64.17	4.28 ($p < .01$)	.44	.70	.16 [.12;.21]	.10
Neuville et al.'s Modell	20	37.61	1.72 ($p < .01$)	.91	.86	.08 [.04;.12]	.07
Lent & Browns Modell	19	46.51	2.48 ($p < .01$)	.84	.83	.11 [.07;.14]	.08
EOS-Modell	32	58.26	1.83 ($p < .01$)	.91	.86	.08 [.05;.11]	.08

Ein Blick auf die Modellfitindizes jedoch ernüchtert: der Fit kann weiterhin nicht als gut beurteilt werden (vgl. Tabelle 7.2).

Das Modell von Neuville et al. (2007) wiederum zeichnete sich wie schon in Studie 1 durch einen sehr guten Fit aus (vgl. Tabelle 7.2). Wie Abbildung 7.4 zeigt, wurden für alle bis auf vier Pfade die Koeffizienten signifikant. Insgesamt kann auch zu diesem Modell festgehalten werden, dass beim Test mit den Daten aus dem zweiten Jahrgang mehr Pfade signifikant wurden. Der Modellfit jedoch ist nicht unbedingt besser, sondern als über beide Jahrgänge hinweg (sehr) gut zu beurteilen. Im Detail: Die Schulleistung war über einen direkten, signifikanten Pfad mit der akademischen Leistung verbunden. Von der SESW zum ersten Messzeitpunkt ging ein signifikanter Pfad zur SESW zum zweiten Messzeitpunkt. Von dort aus wiederum ging ein signifikanter Pfad direkt zur Abbruchintention und zum subjektiven Wert, der den Studium beigemessen wird. Vom subjektiven Wert aus gingen sowohl zur Abbruchintention als auch zum akademischen Engagement signifikante Pfade ab. Zudem war der Pfad von der Abbruchintention zur akademischen Leistung mit einem bedeutsamen Koeffizienten

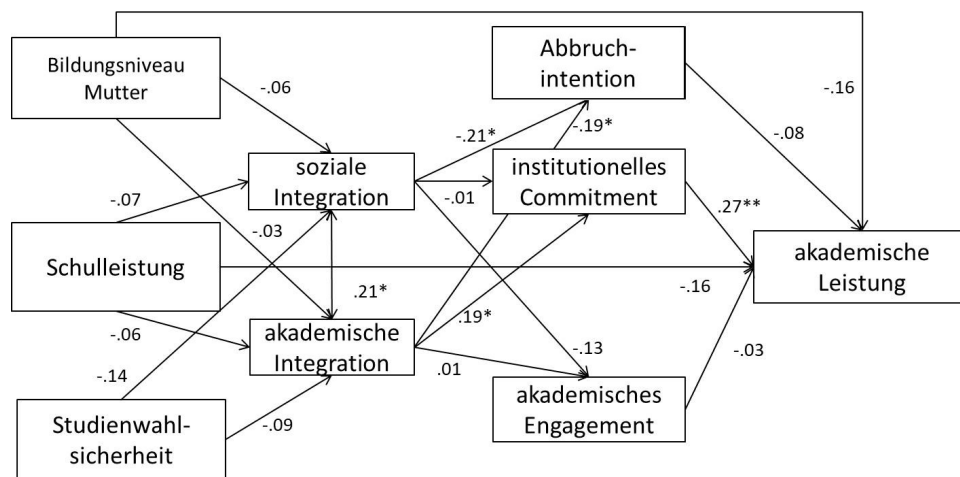


Abbildung 7.3: Tintos Modell mit standardisierten Pfadkoeffizienten *Anm.* * $p < .05$, ** $p < .01$.

versehen.

Insgesamt lässt sich festhalten, dass die im Modell von Neuville et al. (2007) spezifizierten Zusam-

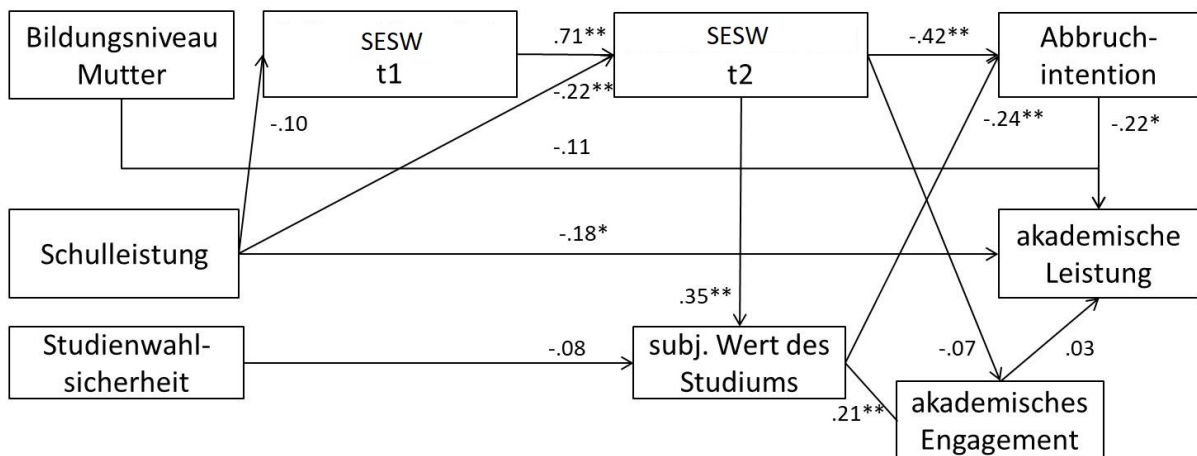


Abbildung 7.4: Neuville et al.'s Modell mit standardisierten Pfadkoeffizienten *Anm.* * $p < .05$, ** $p < .01$.

menhänge also durchaus in den Daten aus Studie 2 zu finden waren.

Das Modell nach Lent und Brown (2013) ist in Abbildung 7.5 dargestellt. Zwar wurden bei diesem Modell mehr Pfade beim Testen mit den Daten aus Jahrgang 2 signifikant als beim Testen mit den Daten aus Jahrgang 1 (sechs versus zwei), es wies aber insgesamt wieder nur einen maximal als befriedigend zu beurteilenden Fit auf (vgl. Tabelle 7.2). Konkret erwiesen sich die Pfade von der SESW zu den Ergebniserwartungen, zur Zielsetzung und zu den Handlungen als signifikant. Darüber hinaus wurden die Pfade von den Ergebniserwartungen zur Zielsetzung und von den Handlungen zur

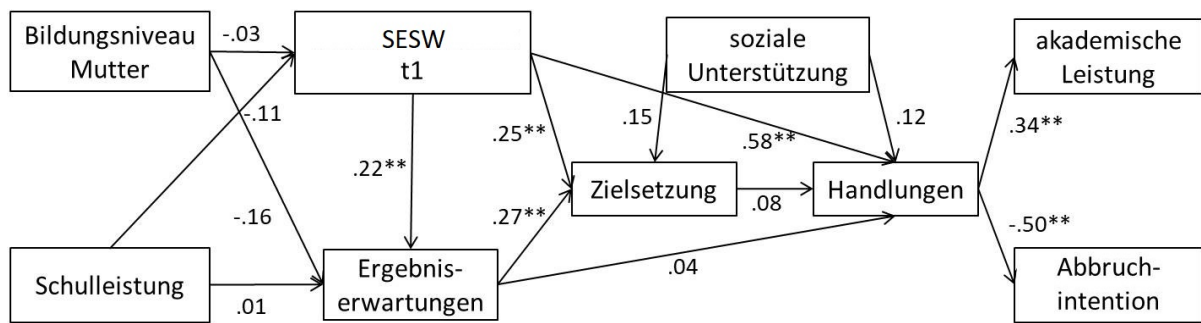


Abbildung 7.5: Lent und Browns Modell mit standardisierten Pfadkoeffizienten *Anm.* ** $p < .01$.

Abbruchintention und zur akademischen Leistung signifikant.

Nachdem anhand der Erkenntnisse aus der ersten Längsschnittstudie (Kapitel 5) das EOS-Modell entwickelt wurde (Kapitel 6), war im Rahmen von Studie 2 dessen erstmalige Erprobung anhand unabhängiger Daten von besonderer Bedeutung. Unabhängig meint in diesem Zusammenhang, dass das EOS-Modell wie bereits beschrieben, basierend auf den Erkenntnissen zum Modellfit der vier etablierten Modelle anhand der Daten aus Jahrgang 1 und unter Berücksichtigung aktueller meta-analytischer Befunde entwickelt wurde. Daher ist die Berechnung des Modellfits des EOS-Modells anhand eben dieser Daten aus der ersten Studie weniger erkenntnisreich, als der Modellfittest anhand des nächsten Jahrgangs.

Abbildung 7.6 zeigt, welche Pfade des EOS-Modells signifikant wurden. Bis auf den Pfad von der

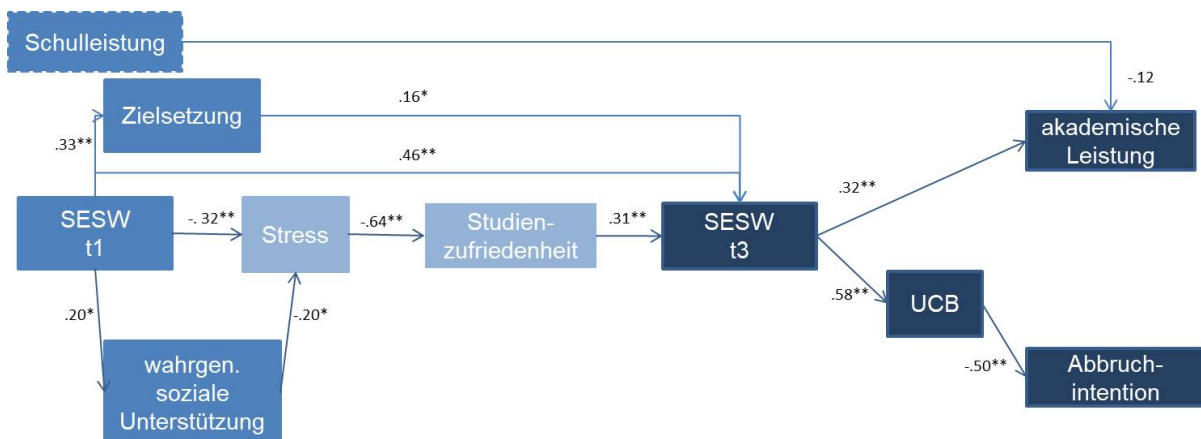


Abbildung 7.6: EOS-Modell mit standardisierten Pfadkoeffizienten, * $p < .05$, ** $p < .01$.

Schulleistung zur aktuellen Studiennote, wurden alle Pfade signifikant oder sehr signifikant. Die Pfade von der SESW zum ersten Messzeitpunkt zum (empfundene) Stress sowie der direkte Pfad zur SESW zum dritten Messzeitpunkt (und der indirekte via Zielsetzung) wurden signifikant. Weiter wur-

Tabelle 7.3: Aikies Informations Criterion (AIC) für die fünf getesteten Modelle [Jahrgang 2]

Modell	AIC
Spadys Modell	4563.15
Tintos Modell	4313.74
Neuville et al.'s Modell	4129.30
Lent & Browns Modell	5552.57
EOS-Modell	4170.63

de der Pfad von der SESW (erster Messzeitpunkt) zur wahrgenommenen sozialen Unterstützung und weiter zum (empfundenen) Stress und von dort aus zur Studienzufriedenheit signifikant. Letztere wies wiederum einen signifikanten Pfad zur SESW (dritter Messzeitpunkt) auf. Die SESW zum dritten Messzeitpunkt war über signifikante Pfade mit der akademischen Leistung und dem UCB verbunden, von welchem aus ein weiterer signifikanter Pfad zur Abbruchintention führte.

Das bedeutet, dass das EOS-Modell die Zusammenhänge im Datensatz aus Studie 2 (sehr) gut wiedergeben kann. Ein Blick auf Tabelle 7.2 belegt das: Genau wie das Modell von Neuville et al. (2007) weist auch das EOS-Modell gemäß gängigen Cut-off-Kriterien für guten Modellfit (Hu & Bentler, 1998, 1999; Hooper et al., 2008) einen guten bis sehr guten Fit zu den Daten aus Jahrgang 2 auf.

7.9 Zusammenfassung der Ergebnisse aus der zweiten Längsschnittstudie

Ein absoluter Vergleich der Modellfits kann wieder anhand des Modellfitkriteriums AIC erfolgen. Tabelle 7.3 zeigt die entsprechenden Kennwerte anhand der Daten aus Studie 2: das Modell von Neuville et al. (2007) und das EOS-Modell sind im Vergleich die beiden besten, wobei das Modell von Neuville et al. (2007) nominell sogar noch einen etwas besseren Fit aufweist als das EOS-Modell. Wie bereits im vorherigen Kapitel beschrieben, lohnt es sich für den Vergleich dieser beide Modelle die Differenz im Fit auch statistisch abzusichern. Der χ^2 -Differenz-Test für diese beiden Modelle wurde jedoch – wie schon anhand der Daten aus Studie 1 – nicht signifikant: $\delta\chi^2 = 20.65$, $p = .11$. Das bedeutet, dass das Modell von Neuville et al. (2007) und das EOS-Modell die Daten aus Studie 2 gleich gut abbilden können.

Als Erkenntnis aus der zweite Längsschnittstudie lässt sich zunächst festhalten, dass das Modell von Neuville et al. (2007) wieder einen (sehr) guten Fit aufwies. Das EOS-Modell konnte die Datenstruktur in Studie 2 ebenfalls gut abbilden. Daraus lässt sich – wie bereits im vorherigen Kapitel angedeutet – ableiten, dass Modelle, die die Variable SESW an zentraler Stelle haben, die Daten besser erklären können als Modelle, die diese Variable nicht enthalten. Im Vergleich des Modells von Lent und Brown (2013) mit dem von Neuville et al. (2007) und dem EOS-Modell, kommt – wie

bereits im vorherige Kapitel beschrieben – die Vermutung auf, dass Modelle, die wiederholte Messungen der SESW aufweisen, besonders gut geeignet sind, um die Datenstruktur abzubilden.

Im folgenden Kapitel ist die dritte Längsschnittstudie beschrieben, anhand derer die Modelle erneut empirisch getestet wurden.

8 Studie 3

8.1 Messzeitpunkte bei Studie 3

Studie 3 ist die dritte und letzte Längsschnittstudie im Rahmen dieses Promotionsprojektes. Sie umfasste ebenfalls drei Messzeitpunkte und lief über das erste Studienjahr hinweg. Messzeitpunkt 1 war von November bis Dezember 2018 (Beginn erstes Semester), Messzeitpunkt 2 war im Februar 2019 und Messzeitpunkt 3 von Juni bis Juli 2019.

Zum ersten Messzeitpunkt nahmen $N = 462$ Personen teil. Zu Messzeitpunkt 2 waren es noch $N = 248$ Personen, was einem Dropout von ca. 46% entspricht. Zu Messzeitpunkt 3 waren noch 186 Personen dabei. Somit liegt der Dropout über die Gesamtlaufzeit von Studie 3 (9 Monate) sogar bei fast 60%. Im Vergleich der drei Längsschnittstudien ist folglich Studie 3 die mit dem größten absoluten Stichprobenumfang und mit dem größten Verlust an Stichprobenumfang über die Dauer der Studie hinweg.

8.2 Stichprobe bei Studie 3

Durch das Anlegen des bereits benannten Einschlusskriteriums (vgl. Kapitel 4, 5, 7) wurde der Stichprobenumfang dann um weitere 26 Personen reduziert, sodass in Studie 3 letztendlich ein Stichprobenumfang von $N = 160$ vorlag.

Im Mittel waren die Personen, deren Daten in die Analysestichprobe aus Jahrgang 3 einfließen $M = 20.7$ ($SD = 3.4$) Jahre alt. 80.0% von ihnen gaben an, weiblich zu sein, 18.1% gaben an, männlich zu sein, die restlichen 1.9% machten keine Angabe zu ihrem Geschlecht. In dieser Studie sah die Verteilung auf die Fachcluster wie folgt aus: 17.5% Naturwissenschaften, 26.9% Medizin oder Psychologie, 12.5% Sprachen, Kulturen, Soziales, 11.9% Erziehungswissenschaften, 31.3% Wirtschaftswissenschaften, Jura oder sonstige Fächer. Die Verteilung der Studierenden auf die Fachcluster ist in Abbildung 8.1 grafisch dargestellt. Vergleicht man die Ausgangsstichprobe von $N = 462$ Studierenden mit der Analysestichprobe von $N = 160$, so zeigt sich kein signifikanter Unterschied hinsichtlich der Verteilung auf die Studienfachcluster ($\chi^2(4) = 5.59$, $p = .323$). Die Analysestichprobe unterschied sich auch hinsichtlich des Alters ($t(159) = 0.96$, $p = .339$) und der Verteilung auf die Geschlechter ($t(159) = 0.53$, $p = .597$) nicht bedeutsam von der Ausgangsstichprobe.

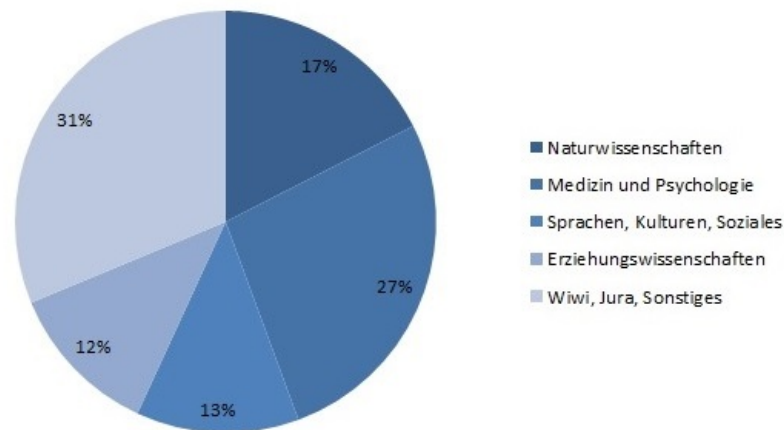


Abbildung 8.1: Verteilung auf die Fachcluster [Jahrgang 3, $N = 160$]

8.3 Repräsentativität

Hinsichtlich des Alters, der Verteilung auf die Geschlechter und über die Fachcluster hinweg, wollte ich (wie schon in Studie 1 und 2) die Repräsentativität der untersuchten Stichprobe für die Population der Studienanfänger im Wintersemester 2018/2019 prüfen. Dazu habe ich die beim Statistischen Bundesamt verzeichneten Populationsstatistiken als Referenz herangezogen. Für die Studienanfänger aus dem Jahr 2018 sind noch keine Angaben zum Durchschnittsalter auf den Seiten des Statistischen Bundesamtes verfügbar. Dies ist zwar etwas verwunderlich, da beispielsweise die Daten zur Verteilung auf die Geschlechter und die Studienfächer vorliegen (Statistisches Bundesamt, 2019), allerdings lässt sich die Repräsentativität der Analysestichprobe aus Studie 3 vermutlich trotzdem einschätzen: Das Durchschnittsalter der Studienanfänger scheint in den letzten Jahren relativ stabil zwischen 21.5 Jahren und 22.5 Jahren zu liegen, weshalb für die Analysestichprobe aus Jahrgang 3 schätzungsweise ebenso wie für die Analysestichproben aus Jahrgang 1 und 2 (siehe Kapitel 5 und 7) angenommen werden kann, dass das Durchschnittsalter mit $M = 20.7$ Jahren ($SD = 3.4$) (knapp) unter dem der Zielpopulation liegt.

Für den Anteil der Frauen an der Population der Studienanfänger im Jahr 2018 liegen Kennwerte vor: ein Frauenanteil von 51.3% wurde verzeichnet. In der vorliegenden Stichprobe liegt dieser weit höher: 80.0%. Auch hier ist bereits nominell erkennbar, dass die Stichprobe aus Jahrgang 3 hinsichtlich der Verteilung auf die Geschlechter nicht repräsentativ ist für die Zielpopulation, was ein t -Test bestätigen konnte: $t(159) = 9.821$, $p < .01$.

Anders als in Studie 1 und 2 wurde in Studie 3 das Studienfach nicht mittels eines Freitextitems abgefragt, sondern es sollte angegeben werden, zu welchem der bereits oben benannten Fachcluster das eigene Studienfach zählt (Single-Choice-Item). Diese kleine Änderung im Fragebogen zum ersten Messzeitpunkt war der Tatsache geschuldet, dass dadurch der Arbeitsschritt der manuellen

Clusterung der Freitext-Angaben entfiel. Allerdings lässt sich dadurch kein rechnerischer Vergleich mit der Verteilung in der Population durchführen, da die von mir verwendeten Cluster nicht denen des Statistischen Bundesamts entsprechen. Eine „Umsortierung“ ist deshalb quasi nicht möglich, weil mit der Darstellung in Clustern automatisch ein Informationsverlust einhergeht und daher nicht mehr differenziert werden kann, welche Anteile auf welches Fach abfallen. Diese Information wäre allerdings notwendig, um die verschiedenen Fächer dann wiederum nach der anderen Cluster-Methode zusammenzufassen.

Klar erkennbar ist jedoch wie bei Studie 1 und 2, dass eine zufriedenstellende Repräsentativität für die gesamte Population nicht vorliegt, zumal die Stichprobe keine Studierende der Ingenieurwissenschaften umfasst.

Somit ist die Stichprobe aus Jahrgang 3 ebenfalls nicht repräsentativ für die Zielpopulation in Bezug auf die oben genannten Aspekte. Wie die Stichproben aus Jahrgang 1 und 2 kann auch diese Stichprobe aus Jahrgang 3 aufgrund der Vielzahl an vertretenen Fächern jedoch durchaus einen fächerübergreifenden Einblick in die Zielpopulation geben und die Stichprobe zumindest als eingeschränkt repräsentativ eingestuft werden.

8.4 Erhobene Variablen in Studie 3

Die erhobenen Variablen wurden bereits einmal an zentraler Stelle für die drei Längsschnittstudien dargestellt, weshalb an dieser Stelle auf Kapitel 4.1.1 verwiesen sei.

8.5 Deskriptive Statistiken

Tabelle 8.1 zeigt die deskriptiven Statistiken aller im Rahmen der Modellfit verwendeten Variablen (Mittelwert, Standardabweichung, Schiefe und Kurtosis jeweils mit Standardfehler).

8.6 Unterschiede zwischen den Fachclustern?

Im Hinblick auf die beiden vorherzusagenden Kriterien (aktuelle Studiennote und Abbruchintention) habe ich Mittelwertsvergleiche zwischen den Fachclustern berechnet. Wie schon für die ersten beiden Längsschnittstudien (Kapitel 5 und 7) sollt dieser Vergleich klären, ob eine Zentrierung der beiden Kriterienvariablen notwendig ist.

Die ANOVA zum Kriterium *aktuelle Studiennote* wurde nicht signifikant, die ANOVA zum Kriterium *Abbruchintention* ebenfalls nicht: $F_{Note}(4) = 1.734$ ($p = .145$) und $F_{Abbruch}(4) = .953$ ($p = .435$). Somit war auch für diese Stichprobe keine Zentrierung der Variablen *Abbruchintention* und *aktuelle Studiennote* notwendig.

Tabelle 8.1: Mittelwert, Standardabweichung sowie Schiefe, Kurtosis und ggf. Reliabilität aller verwendeten Variablen in Studie 3 ($N = 160$)

Variable (Anzahl Items)	M	SD	Schiefe (SE)	Kurtosis (SE)	Cronbachs α
Bildungslevel d. Mutter (1)	3.13	1.56	0.94 (0.19)	0.98 (0.38)	-
Abiturnote (1)	2.10	0.63	0.22 (0.19)	-0.78 (0.38)	-
Informiertheit (1)	4.06	1.53	-0.18 (0.19)	-0.42 (0.38)	-
Selbstwirksamkeit (13)	3.65	0.50	-0.30 (0.19)	0.16 (0.38)	.83
Integration (3)	5.35	0.97	-1.38 (0.19)	0.70 (0.38)	.70
Zielsetzung (13)	50.14	6.05	-0.25 (0.19)	0.17 (0.38)	.82
Ergebniserwartungen (6)	26.82	2.38	-0.47 (0.19)	-0.51 (0.38)	.57
P-U-Fit (9)	45.88	7.68	-0.60 (0.19)	0.32 (0.38)	.86
soziale Unterstützung (8)	28.39	4.53	-1.62 (0.19)	2.89 (0.38)	.93
Lernklima (8)	26.72	4.88	-0.01 (0.20)	0.01 (0.40)	.83
Wert (5)	29.76	3.94	-0.80 (0.20)	0.78 (0.40)	.83
Stress (6)	13.16	3.91	0.22 (0.20)	-0.12 (0.40)	.77
Selbstwirksamkeit (13)	3.78	0.45	-0.16 (0.20)	-0.01 (0.40)	.83
Handlungen (13)	49.14	6.31	-0.49 (0.20)	0.70 (0.40)	.81
Zufriedenheit (3)	3.39	0.94	-0.36 (0.20)	-0.46 (0.40)	.81
Selbstwirksamkeit (13)	3.78	0.49	-0.20 (0.19)	0.28 (0.38)	.84
intellektuelle Entw. (2)	5.22	1.20	-0.63 (0.19)	0.78 (0.38)	-
Persistenz (1)	15.17	9.87	1.18 (0.19)	2.04 (0.38)	-
institutionelles Commitment (8)	4.34	0.70	-0.69 (0.19)	0.18 (0.38)	.59
Integration (3)	3.78	0.50	-2.25 (0.19)	4.33 (0.38)	.57
UCB (27)	4.06	0.53	-0.40 (0.19)	0.31 (0.38)	.80
aktuelle Note (1)	10.01	2.87	-0.95 (0.19)	0.85 (0.38)	-
Abbruchvorhaben (2)	2.81	1.66	3.09 (0.19)	10.84 (0.38)	-
Zufriedenheit (3)	3.15	0.99	-0.28 (0.19)	-0.44 (0.38)	.78

8.7 Korrelative Zusammenhänge

Bevor die Ergebnisse der Pfadanalysen dargestellt werden, möchte ich zunächst auf die Tabellen im Anhang verweisen (13.15 ff.), in denen die bivariaten Korrelationen nullter Ordnung aufgeführt sind. Wie schon für Studie 1 und 2 sind zwecks Übersichtlichkeit pro Tabelle jeweils nur die Korrelationen aller in einem Modell berücksichtigten Variablen aufgeführt.

8.8 Modelltests

Abbildung 8.2 zeigt das Pfadmodell nach Spady (1971) mit standardisierten Pfadkoeffizienten. Sieben Pfade wurden beim Modelltest mit den Daten aus Jahrgang 3 signifikant: Der Pfad von der Schulleistung zur akademischen Leistung wies einen signifikanten Koeffizienten auf. Zudem wurde der Pfad vom Person-University-Fit sowohl zur intellektuellen Entwicklung als auch zur sozialen Unterstützung signifikant. Von der sozialen Unterstützung ging ein bedeutsamer Pfad zur sozialen Integration, von dort aus weiter zur Studienzufriedenheit und weiter zum institutionellen Commitment und schlussendlich zur Abbruchintention. Im Vergleich der Modellfitergebnisse aus den drei Studien konnte das Modell von Spady (1971) somit die Daten aus Studie 3 am besten erklären (vgl. auch Tabelle 8.2). Anders formuliert: Nur in Studie 3 wurden die Pfade im Kern dieses Modells tatsächlich signifikant.

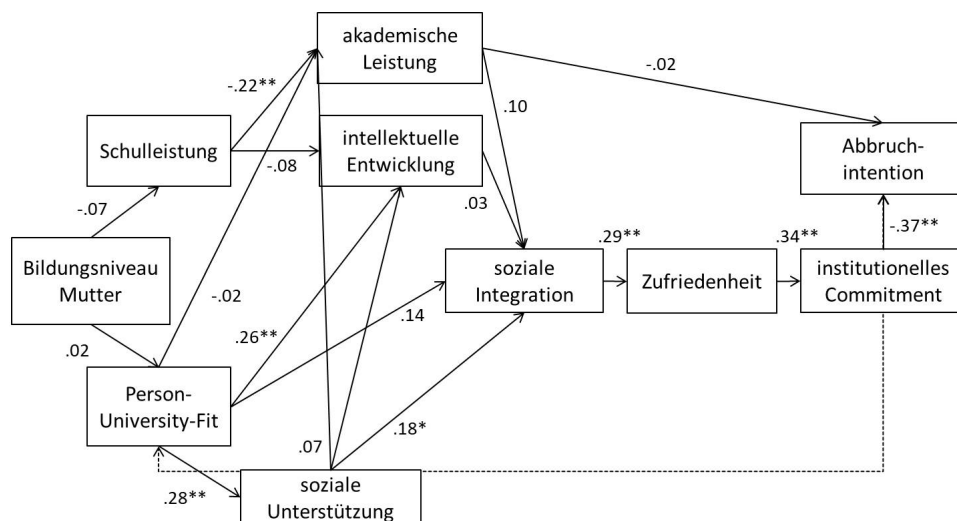
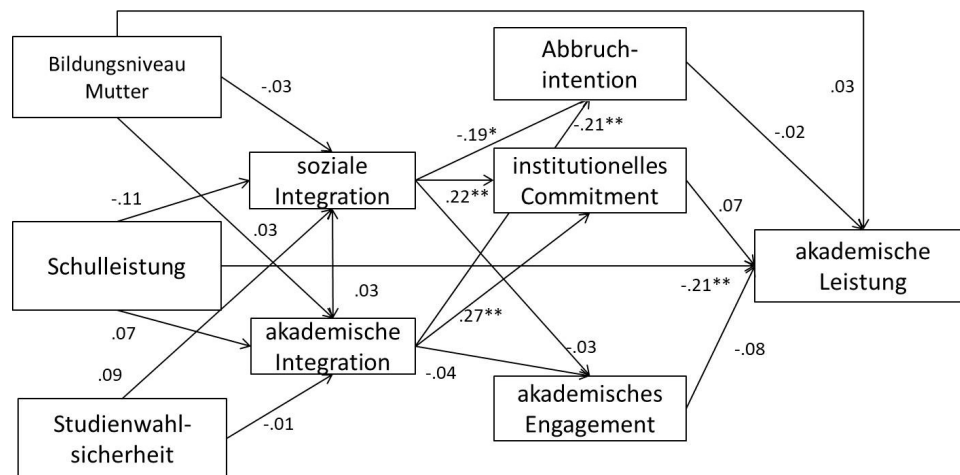


Abbildung 8.2: Spadys Modell mit standardisierten Pfadkoeffizienten *Anm.* $^*p < .05$, $^{**}p < .01$.

Auch das Modell von Tinto (1975) weist im Vergleich mit den Ergebnissen aus Studie 1 und 2 in dieser Studie den besten Fit auf: Es kann die Daten aus Jahrgang 3 besser erklären als die Daten aus den beiden vorherigen Jahrgängen. Konkret wiesen wie in Jahrgang 2 fünf Pfade signifikante Koeffizienten aus (vgl. Abbildung 8.3). Die Pfade von der sozialen Integration auf die Abbruchintention und das institutionelle Commitment wurden ebenso signifikant wie der direkte Zusammenhang

Tabelle 8.2: Modellfitindizes für die fünf getesteten Modelle [Jahrgang 3]

	df	χ^2	χ^2/df (p -value)	CFI	AGFI	RMSEA [$CI_{90\%}$]	SRMR
Spadys Modell	29	89.65	3.01 ($p < .01$)	.61	.82	.11 [.09;.14]	.12
Tintos Modell	15	38.01	2.53 ($p < .01$)	.61	.86	.10 [.06;.14]	.07
Neuville et al.'s Modell	20	13.91	0.70 ($p = .83$)	1.00	.96	.00 [.00;.04]	.05
Lent & Browns Modell	19	31.54	1.66 ($p = .04$)	.96	.91	.06 [.02;.10]	.07
EOS-Modell	32	58.29	1.82 ($p < .01$)	.94	.89	.07 [.04;.10]	.08

Abbildung 8.3: Tintos Modell mit standardisierten Pfadkoeffizienten *Anm.* * $p < .05$, ** $p < .01$.

von der Schulleistung mit der akademischen Leistung. Zudem verlief ein signifikanter Pfad von der sozialen Integration zur Abbruchintention. Somit wurden die meisten Pfade im Kern dieses Modell signifikant. Der Modellfit (vgl. Tabelle 8.2) jedoch ist maximal als befriedigend einzustufen.

Das Modell von Neuville et al. (2007) erwies sich als sehr gut geeignet, um die Datenstruktur aus Studie 3 zu erklären (vgl. Tabelle 8.2). Zwar wurden etwas weniger Pfade signifikant als in Studie 2, aber mit immerhin fünf signifikanten Pfadkoeffizienten war erneut der Kern des Modells mit signifikanten Pfaden versehen (vgl. Abbildung 8.5). Konkret wurden die Pfade von der Schulleistung zur akademischen Leistung und zur SESW am ersten Messzeitpunkt, sowie der Zusammenhang der SESW zum ersten mit der SESW zum zweiten Messzeitpunkt signifikant. Weiter erwies sich der Pfad von der SESW (zweiter Messzeitpunkt) zum subjektiven Wert des Studiums und von dort aus zur Abbruchintention als bedeutsam. Somit zeigte sich erneut, dass die Erwartungs-Wert-Struktur die Daten gut abbilden konnte.

Das Modell nach Lent und Brown (2013) passte im Kontrast zu den Ergebnissen aus Studie 2, gut

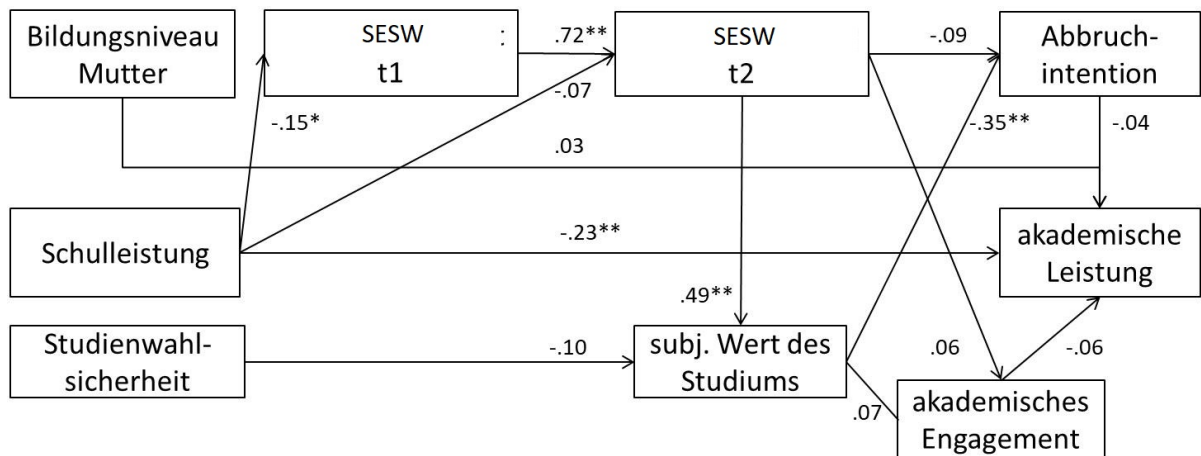


Abbildung 8.4: Neuville et al.'s Modell mit standardisierten Pfadkoeffizienten Anm. * $p < .05$, ** $p < .01$.

zu den Daten aus Studie 3. Das zeigt nicht nur der sehr gute Modellfit, sondern auch die visuelle Inspektion: Sechs Pfade wurden signifikant, sodass man sagen kann, dass ein Großteil der vom Modell spezifizierten Zusammenhänge in den Daten aus Studie 3 wiederzufinden war.

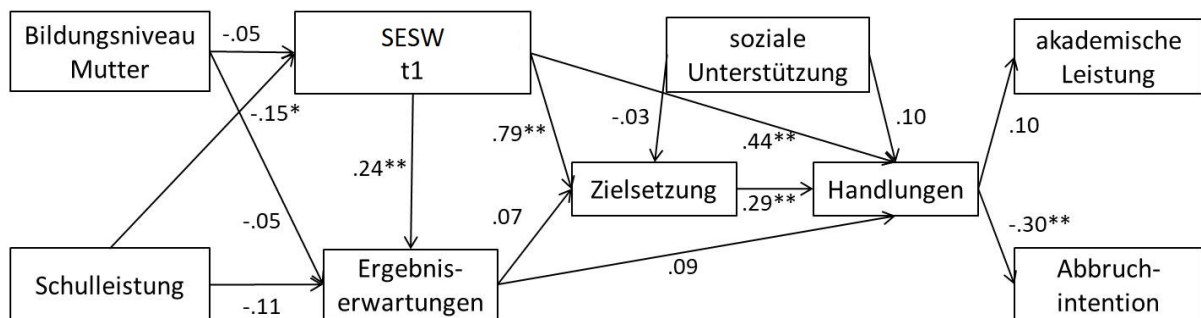


Abbildung 8.5: Lent und Browns Modell mit standardisierten Pfadkoeffizienten Anm. * $p < .05$, ** $p < .01$.

Wie schon in Studie 1 erwies sich der Pfad von der Schulleistung auf die SESW (zum ersten Messzeitpunkt) als bedeutsam. Weiterhin führten von der SESW bedeutsame Pfade sowohl zu den Ergebniserwartungen als auch zur Zielsetzung und zu den Handlungen. Von der Zielsetzung aus verlief ebenfalls ein signifikanter Pfad zu den Handlungen, von wo aus ein weiterer signifikanter Pfad zur Abbruchintention führte.

Zu guter Letzt bleibt noch der Fit des EOS-Modells zu beschreiben: Wie schon bei der Modellfitanalyse mit den Daten aus Studie 1 und 2 erwies es sich auch in Bezug auf die Daten aus Studie 3 als (sehr) gut passend (vgl. Tabelle 8.2). Konkret wurden bis auf zwei Pfade alle signifikant (Abbildung

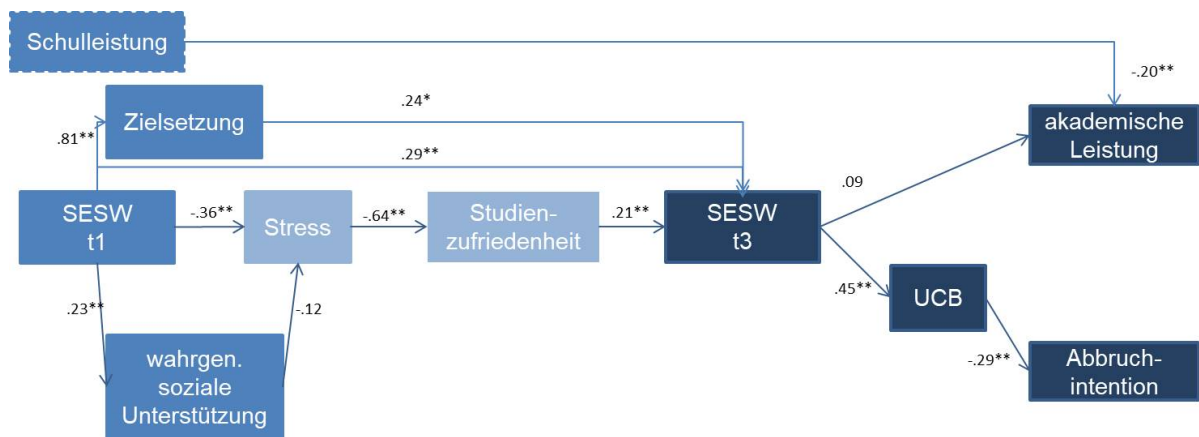


Abbildung 8.6: EOS-Modell mit standardisierten Pfadkoeffizienten, * $p < .05$, ** $p < .01$.

Tabelle 8.3: Aikies Informations Criterion (AIC) für die fünf getesteten Modelle [Jahrgang 3]

Modell	AIC
Spadys Modell	6122.39
Tintos Modell	5833.26
Neuville et al.'s Modell	5593.17
Lent & Browns Modell	7281.47
EOS-Modell	5454.53

8.6): Von der SESW zum ersten Messzeitpunkt führten signifikante Pfade zur Zielsetzung, zur SESW (gemessen am dritten Messzeitpunkt), zum erlebten Stress und zur wahrgenommenen Unterstützung. Von letzterer aus ging ein signifikanter Pfad zur Studienzufriedenheit und weiter zur SESW (dritter Messzeitpunkt). Der Pfad von der SESW (dritter Messzeitpunkt) zur akademischen Leistung wurde nicht signifikant, aber der zum UCB und weiter zur Abbruchintention. Von der Schulleistung ging wie bereits vorher in diesem Kapitel beschrieben einen bedeutsamer Pfad zur akademischen Leistung. Zusammenfassend möchte ich an dieser Stelle festhalten, dass sich nahezu die komplette im EOS-Modell spezifizierte Struktur in den Daten aus Studie 3 wiederfinden ließ.

8.9 Zusammenfassung der Ergebnisse aus der dritten Längsschnittstudie

Wie schon in den beiden vorherigen Kapiteln möchte ich auch zum Abschluss dieses Kapitels den direkten Vergleich der Modelle anhand des AICs durchführen. Tabelle 8.3 zeigt die entsprechenden Kennwerte. Erneut zeigt sich, dass das Modell von Neuville et al. (2007) und das EOS-Modell den

anderen drei Modellen im Fit überlegen sind. Diesmal ist es so, dass das *EOS*-Modell rein nominell betrachtet, gegenüber dem von Neuville et al. (2007) zu bevorzugen ist. Anhand der Daten aus Studie 1 und Studie 2 war das AIC der beiden Modelle fast gleich (das Modell von Neuville et al., 2007 hatte ein nur minimal niedrigeres AIC, siehe Tabellen 6.2 und 7.3, der Unterschied war gemäß χ^2 -Differenz-Test beide Male nicht signifikant). In Studie 3 allerdings wurde der χ^2 -Differenz-Test signifikant: Das *EOS*-Modell bildete die Daten besser ab ($\delta_{\chi^2} = 44.38$, $p < .01$).

Deutlich wird auf jeden Fall wieder, was schon anhand der Daten aus Studie 1 und 2 zu beobachten war: Modelle, die die Variable SESW enthalten und besonders die Modelle, die die SESW an mehr als einem Zeitpunkt aufführen, weisen einen besseren Fit zu den in den drei Längsschnittstudien erhobenen Daten auf als solche Modelle, die die SESW nicht berücksichtigen.

Im folgenden Kapitel möchte ich die Erkenntnisse aus Studie 1 bis 3 zusammenfassen und eine weiterführende, explorative Analyse anhand der Daten aus diesen Studien vorstellen, die auf diese Erkenntnisse fußt.

9 Zusammenschau der Ergebnisse der drei Längsschnittstudien

In diesem Kapitel möchte ich die Ergebnisse aus den drei Längsschnittstudien zusammenfassen, um abschließend ein vorläufiges Gesamtfazit zu ziehen. Zudem werde ich die Ergebnisse einer Analyse der Messinvarianz des *EOS*-Modells über die drei Jahrgänge hinweg berichten.

9.1 Erkenntnisse aus den drei Längsschnittstudien

Die drei Längsschnittstudien haben deutlich aufgezeigt, dass das Modell von Neuville et al. (2007) und das *EOS*-Modell die Daten aus allen drei Jahrgängen mit Abstand am besten aufklären konnten (vgl. Tabelle 9.1). Wie schon in den Kapiteln zu den drei Studien angerissen, legt diese Erkenntnis die Vermutung nahe, dass die Berücksichtigung der SESW von entscheidender Bedeutsamkeit ist, um den Prozess der Studieneinstiegs gut abbilden zu können. Die Tatsache, dass die Daten dieser drei Längsschnittstudien unabhängig voneinander sind und doch immer wieder diese Quintessenz der Modellfitvergleiche deutlich wurde, untermauert diese Vermutung.

Dies kann insofern als konform mit der Literatur eingestuft werden, als dass es die Ergebnisse neuerer Metaanalysen zu einzelnen Prädiktoren von Studienerfolg und Studienabbruch widerspiegelt. Insbesondere die Ergebnisse der Meta-Metaanalyse von M. Schneider und Preckel (2017) möchte ich hier erwähnen, in der die Selbstwirksamkeit Platz zwei in der Rangreihe der bedeutendsten Prädiktoren erzielte (vgl. Kapitel 2.2).

Was in den Daten aus Studie 1 auch deutlich wird: Der in der Literatur vielfach aufgezeigte und metaanalytisch belegte Zusammenhang zwischen der Schulleistung (operationalisiert als die Abiturnote) und der akademischen Leistung (operationalisiert als die aktuelle Studiennote) ist in der untersuchten Stichproben aus dem Jahrgang 2016/2017 deutlich weniger stark, sogar nicht statistisch bedeutsam. Dies könnte der beschriebenen Selektivität der Stichproben geschuldet sein. Konkret steht die Annahme im Raum, dass aufgrund der Freiwilligkeit der Teilnahme an den Studien insbesondere jene Studienanfänger, die an Untersuchungen im Studienkontext interessiert und besonders gewissenhaft sind, teilgenommen haben (vgl. Kapitel 4). Letztere Eigenschaft hat sich bereits in vorangegangenen Studien als geeigneter Prädiktor für (Studien-)Leistung herausgestellt. Daher lag die Vermutung nahe, dass die Selektivität der Stichprobe zu einer Varianzeinschränkung entweder in einer oder in beiden Variablen geführt haben könnte. Der Vergleich der deskriptiven Statistiken zur Abiturnote und zur aktuellen Studiennote mit den Daten aus einer aktuellen großen, repräsentativen Umfrage innerhalb der Zielpopulation (Middendorff et al., 2016, Kapitel 10) jedoch zeigt, dass die Abiturnote (Mittelwert und Standardabweichung) im vergleichbaren Bereich liegt. Die aktuelle Studiennote,

Tabelle 9.1: Aikies Informations Criterion für alle fünf Modelle in den drei Längsschnittstudien

Model	AIC
Spadys Modell	5445.26
Tintos Modell	5216.73
Neuville et al.'s Modell	4957.21
Lent & Browns Modell	6493.78
EOS-Modell	4962.40
Spadys Modell	4563.11
Tintos Modell	4313.74
Neuville et al.'s Modell	4129.30
Lent & Browns Modell	5552.57
EOS-Modell	4170.63
Spadys Modell	6122.27
Tintos Modell	5833.26
Neuilles Modell	5593.17
Lent & Browns Modell	7281.47
EOS-Modell	5454.53

die in diesem Referenzdatensatz auf der Schulnotenskala (1 bis 6) erhoben wurde und in den drei Längsschnittstudien auf der Skala von 1 bis 15 Notenpunkten, lässt ebenfalls keine klare Varianzeinschränkung erkennen. Diese Erklärung entfällt somit. Da sich nur in einer der drei Studien (Studie 1) die Diskrepanz zu den etablierten Befunden zeigte, ist davon auszugehen, dass es sich um ein Artefakt handelt.

Dies zeigt einmal mehr, dass die Durchführung von drei (statt nur einer) Längsschnittstudien den großen Mehrwert birgt, die eigenen Ergebnisse potentiell replizieren zu können. Zusammenhänge und Beobachtungen, die wiederholt über die drei Studien auftauchen, sind dann als besonders verlässlich einzustufen, so, wie beispielsweise die Erkenntnisse hinsichtlich der Modellfitvergleiche.

9.2 Unterschiede in der Höhe der Pfadkoeffizienten: Messfehler berücksichtigen

Im Anhang (vgl. Tabelle 13.5 ff.) sind die Korrelationen nullter Ordnung für die in den Modellen berücksichtigten Variablen für alle drei Jahrgänge, separat für jedes Modell dargestellt. Hier, wie auch im Vergleich der Pfadkoeffizienten eines Modells über die drei Studien hinweg, zeigt sich, dass die Datenstruktur in den Jahrgängen durchaus nicht exakt äquivalent ist. Dies kann zum einen daran liegen, dass sicherlich trotz der umfangreichen Datenerhebung möglicherweise weitere Variablen, die potentiell als Kovariaten betrachtet werden könnten, Einfluss auf die Höhe der bivariaten Zusam-

menhänge haben können.

Zum anderen, und dies sollte stets bei der Interpretation der Koeffizienten berücksichtigt werden, sind die berichteten Kennwerte Schätzer. Das bedeutet, dass sie stets mit einem Messfehler versehen sind. Dies entspricht der Grundannahme der KTT. Um dies zu verdeutlichen, habe ich im Anhang für jedes Modell für jede Studie die Pfade sowie die zugehörigen Standardfehler und die p -Werte aufgeführt (vgl. Tabelle 5.2). Die soll auch verdeutlichen, dass bei den Abbildungen in Kapitel 5, 7 und 8 zwar signifikante Pfade ($p < .05$) kenntlich gemacht wurden, aber in den Abbildungen nicht ersichtlich ist, ob ein nicht als signifikant markierter Pfad ggf. einen p -Wert nur knapp über der Grenze von $\alpha = .05$ aufweist oder der Koeffizient weit entfernt davon ist, bedeutsam zu sein.

An dieser Stelle sei nochmals auf die *FIML*-Methode verwiesen, die ich als Imputationsmethode verwendet habe (vgl. Kapitel 4.1.8): Gemäß Lütke et al. (2007) ist die Verwendung dieser Methode als *state of the art* zu bezeichnen, unter anderem, weil durch den Einsatz eines solchen multiplen Imputationsverfahrens der Standardfehler reduziert wird. Das bedeutet, dass zwar naturgemäß auch bei den hier berichteten Modellfitanalysen *Schätzer* und nicht „wahre Werte“ ermittelt wurden, dass das verwendete Verfahren jedoch dasjenige ist, welches die Schätzfehler möglichst gering hält. Dies führt letztendlich zu einer höheren Verlässlichkeit der erzielten Ergebnisse, im Vergleich zur Verwendung alternativer Methoden.

9.3 Messinvarianz über die drei Jahrgänge hinweg

Ein wichtiger Aspekt für die Generalisierbarkeit des *EOS*-Modells ist die Frage der Messinvarianz über die drei Jahrgänge hinweg. Da ich das Konzept der Messinvarianz und der Messinvarianzanalysen bereits in Kapitel 3 ausführlich beschrieben habe, möchte in diesem Kapitel direkt zu den Ergebnissen der Messinvarianzprüfung übergehen.

Vandenberg und Lance (2000) erläutern in ihrem Artikel, welche Bedeutung Messinvarianz für die Generalisierbarkeit empirischer Ergebnisse hat und sprechen konkrete Empfehlungen für die Durchführung von Messinvarianzprüfungen aus. Dabei beziehen sich die Ausführungen im Allgemeinen auf Faktoranalysen oder Strukturgleichungsmodelle, also solche Modelle, die ein Messmodell (Zusammenhang latenter und manifester Variablen) und ein Strukturmodell (Zusammenhang zwischen den latenten Variablen) umfassen. In meinen Modellanalysen habe ich mit Pfadmodellen gerechnet. Damit ist (gängigen Bezeichnungen zur Folge, siehe Cole und Preacher, 2014) ein Modell mit ausschließlich manifesten Variablen gemeint. Daher ist der Begriff *Faktorladungen* im Zusammenhang mit *Messinvarianz* hier als *Pfadkoeffizienten* beziehungsweise *Regressionskoeffizienten* zu übersetzen. Weiter bedeutet das, dass im Gegensatz zur Testung eines Modells mit latenten und manifesten Variablen, wie es bei dem hierarchischen Modell der SESW-Skala der Fall ist, die Spezifikationen für konfigurale und metrische Messinvarianz gleich sind, da nicht zwischen den Ladungen manifester und latenter Faktoren differenziert werden kann, da das Pfadmodell keine latenten Variablen umfasst.

9.3.1 Ergebnis der Messinvarianzprüfung

Vergleicht man die Fitindizes für das Modell unter den Restriktionen der konfiguralen Invarianz (erste Zeile in Tabelle 9.2) mit denen für das Modell unter den Restriktionen für schwache (metrische) Invarianz (zweite Zeile in Tabelle 9.2), so zeigt sich kein Unterschied im Fit (weil die Modelle exakt gleich sind, da keine latente Ebene vorhanden ist), was bedeutet, dass die sowohl die erste Stufe der Invarianztestung als auch die zweite als erfolgreich absolviert verstanden werden kann.

Tabelle 9.2: EOS-Modell: Test auf Messinvarianz über die drei Jahrgänge

Invarianz-Stufe	χ^2_{SB}	χ^2/df (p -value)	CFI	RMSEA	SRMR
konfigurale	131.118	1.22 (.006)	.957	.056	.064
metrische	131.118	1.22 (.006)	.957	.056	.064
skalare	222.994	2.05 (<.001)	.867	.090	.132

Vergleicht man weiter die zweite Zeile in Tabelle 9.2 mit der dritten, in der die Fitindizes für das Modell unter den Restriktionen starker (skalarer) Invarianz aufgeführt sind, so erkennt man eine deutliche Differenz in den Fitindizes. Diese Differenz liegt außerhalb gängiger Toleranzgrenzen für nicht signifikante Änderungen im Modellfit (F. F. Chen, 2007). Daher kann die nächste Stufe der Invarianz, die starke Invarianz, nicht als gegeben angesehen werden. Es bleibt folglich festzuhalten, dass sich das Pfadmodell zum EOS-Modell als metrisch invariant über die drei untersuchten Jahrgänge hinweg erwiesen hat.

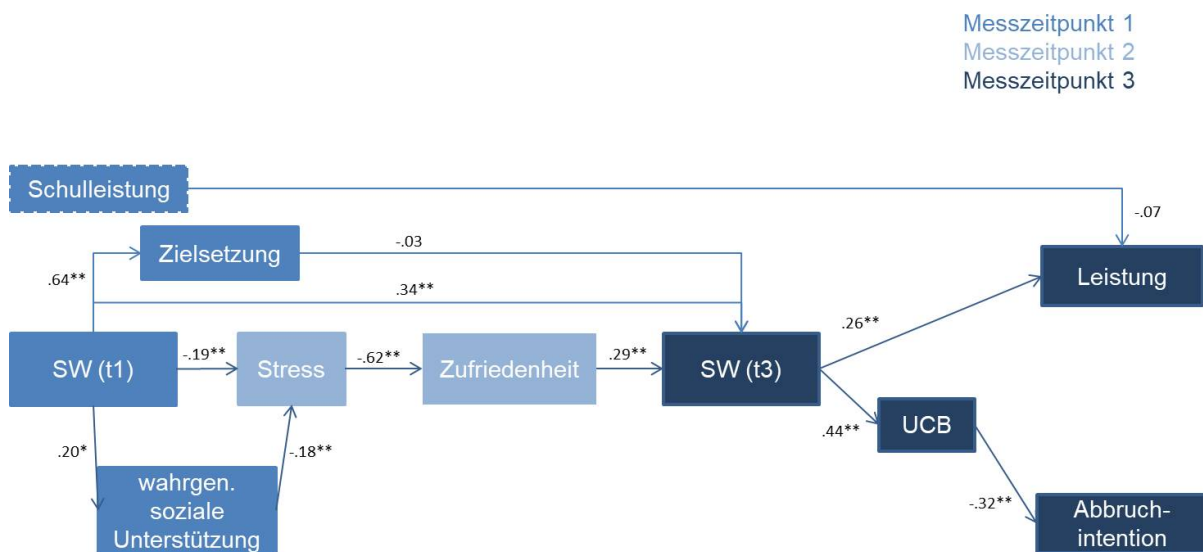
Abschließend lässt sich also festhalten, dass das Pfadmodell zum EOS-Modell zwar hinsichtlich konfiguraler und metrischer Aspekte invariant über die drei Jahrgänge ist, also eine schwache Messinvarianz durchaus aufweist, jedoch keiner strengeren Messinvarianzprüfung Stand hielt. Aufgrund der Verwendung der üblichen Nomenklatur („schwache Messinvarianz“ gemäß Meredith, 1993) mag dieses Ergebnis als unbefriedigend erscheinen. Dies jedoch ist immer vor dem Hintergrund des Anwendungskontextes zu beurteilen (Schmitt & Kuljanin, 2008). Tatsächlich ist dieses Ergebnis für den vorliegenden Anwendungskontext, namentlich die Übertragung des EOS-Modells auf weitere Jahrgänge von Studienanfängern vollkommen ausreichend.

9.3.2 Bedeutung der metrischen Messinvarianz des EOS-Modells

Das Ergebnis der Messinvarianz-Prüfung ist so zu interpretieren, dass sich das EOS-Modell als übertragbar auf verschiedene Jahrgänge von Studieneinsteigern erwiesen hat. Dies attestiert dem EOS-Modell eine gewisse Generalisierbarkeit der im Modell ausgewiesenen Zusammenhänge. Somit ist die

Voraussetzung dafür, dass das EOS-Modell für zukünftige Jahrgänge von Studienanfängern herangezogen wird, gegeben. Zudem ist somit die Voraussetzung erfüllt, um die Daten der drei Jahrgänge zu aggregieren und den Modellfit abschließend nochmal anhand dieses aggregierten Datensatzes zu prüfen.

Abbildung 9.1 zeigt das Pfadmodell für das EOS-Modell (mit standardisierten Pfadkoeffizienten), empirisch geprüft mit dem Gesamtdatensatz. In diesem aggregierten Gesamtdatensatz über die drei Jahrgänge hinweg sind also alle drei Analysestichproben zusammengefasst, sodass ein Gesamtstichprobenumfang von $N_{gesamt} = 424$ resultiert.



Neues Modell mit Daten aus allen drei Jahrgängen ($N = 424$), UCB = University Citizenship Behavior, drei Messzeitpunkte pro Jahrgang (farblich kodiert), ** $p < .01$.

Abbildung 9.1: EOS-Modell mit standardisierten Pfadkoeffizienten, * $p < .05$, ** $p < .01$, ($N = 424$)

Dabei ergaben sich die in Tabelle 9.3 aufgeführten Werte für die Fit-Indizes. Dass die Modellfit-Schätzungen mit einem größeren Stichprobenumfang genauer werden, ist hierbei auch daran zu erkennen, dass das Konfidenzintervall für den RMSEA-Index deutlich kleiner ist, als bei den Modellfit-Schätzungen auf Basis nur einer der Jahrgänge (vgl. Kapitel 6, 7, 8).

Tabelle 9.3: Fitindizes für das EOS-Modell [aggregierter Datensatz, $N = 424$]

χ^2	df	χ^2/df (p -value)	CFI	AGFI	RMSEA [$CI_{90\%}$]	sRMR
75.35	32	2.35 ($p < .01$)	.95	.95	.05 [.04;.07]	.06

10 Sekundärdatenanalyse

Zusätzlich zur Durchführung eigener Studien sollte die Validierung des *EOS*-Modells auch anhand von externen Daten erfolgen.

10.1 Vorteile großer bundesweiter Studierendenbefragungen

Wie in Kapitel 9 beschrieben, bieten longitudinale Daten die Möglichkeit, Prozesse wie die im Fokus dieser Arbeit stehenden, abzubilden. In den drei im Rahmen dieser Arbeit vorgestellten Längsschnittstudien kamen zudem zumeist ausführliche, psychometrisch hochwertige Messinstrumente zum Einsatz, um die interessierenden Konstrukte möglichst valide zu erfassen. Neben diesen Vorteilen der Daten aus den drei Längsschnittstudien ist jedoch zu bedenken, dass die Stichproben zwar von ausreichender Größe sind, jedoch nicht repräsentativ (hinsichtlich demografischer Aspekte wie Verteilung auf die Geschlechter oder Verteilung auf die Studienfachcluster) für die Zielpopulation der Studienanfänger in Deutschland.

Die Ergebnisse der Längsschnittstudien zeigen, dass das *EOS*-Modell Invarianz über verschiedene Kohorten von Studienanfängern aufweist. Eine Überprüfung der Generalisierbarkeit über verschiedene Studienfächer und Studienfachcluster hinweg ist mit der relativ kleinen Fallzahl in manchen Fachclustern in den Längsschnittstudien jedoch in Ermangelung ausreichender statistischer Power nicht verlässlich möglich: Die Längsschnittstichproben weisen zwar eine große Heterogenität hinsichtlich der vertretenen Studienfächer auf, jedoch ist ein Vergleich zwischen den Fachclustern nicht ratsam, da einige Zellen für diese Analysen zu dünn besetzt wären. Eine Überprüfung der Invarianz über verschiedene Fachclustern hinweg ist folglich mit diesen Datensätzen nicht durchführbar.

Große Stichprobenumfänge und repräsentative Stichproben bieten solche großen Datensätze, wie beispielsweise das Forschungsdatenzentrum des DZHW oder des Leibniz-Instituts für Bildungsverläufe (FDZ-LIfBi) sie erhebt und für Sekundärdatenanalysen zur Verfügung stellt.

Daher lag es nahe, einen solche Datensatz für eine weitere Validierungsstudie zum *EOS*-Modell in Form einer Sekundärdatenanalyse heranzuziehen.

10.1.1 Forschungsfragen

Mit Hilfe der Sekundärdatenanalyse sollten folglich zwei Forschungsfragen beantwortet werden:

1. Ist das EOS-Modell geeignet, um die Zusammenhänge zwischen den Modellvariablen auch anhand eines externen Datensatzes abzubilden?
2. Ist das EOS-Modell geeignet, um in verschiedenen Fachclustern die aktuelle Studienleistung sowie Abbruchintentionen vorherzusagen?

Um diese Forschungsfragen beantworten zu können, suchte ich nach einen geeigneten, großen und repräsentativen, fächerheterogenen Datensatz im Angebot der beiden oben genannten Forschungsdatenzentren.

10.2 Auswahl geeigneter Sekundärdaten

Die Suche nach geeigneten Sekundärdaten führte mich zu zwei für die zu überprüfende Fragestellung passenden Datensätzen. Einerseits die Daten der *National Educational Panel Study* (kurz: NEPS) und andererseits die Daten der regelmäßig durchgeführten Sozialerhebung unter den in Deutschland Studierenden. Letztere liegen in der aktuellsten Auflage aus der 21. Sozialerhebung aus dem Jahr 2016 vor. Ich registrierte mich bei beiden Datenportalen und beantragte den Zugang zu den entsprechenden Datensätzen: das Scientific Use File (Version 10.0.0) der Startkohorte Studierende der NEPS (Blossfeld, Roßbach & von Maurice, 2011) und das Scientific Use File der 21. Sozialerhebung des deutschen Studentenwerks durchgeführt vom DZHW (Middendorff et al., 2016).

Von diesen beiden in Erwägung gezogenen Datensätzen umfasste nur der letztgenannte die meisten der im EOS-Modell enthaltenen Variablen, sodass nur dieser für die Testung des Modells in Frage kam. Zwar sind die Daten aus der 21. Sozialerhebung querschnittlicher Natur, jedoch sind zumindest alle für die Analyse der zweiten Hälfte des abgebildeten Prozesses notwendigen Variablen wenigstens mit einem Item abgebildet.

10.2.1 Die 21. Sozialerhebung

An dieser Stelle möchte ich kurz den Hintergrund der Sozialerhebung skizzieren: Das Deutsche Studentenwerk führt regelmäßig die sogenannten Sozialerhebungen durch. Hierbei handelt es sich um großflächige Befragungen innerhalb der Population der Studierenden in Deutschland. Etwa alle drei Jahre wird bundesweit circa jeder sechste Studierende eingeladen, an der Befragung teilzunehmen, in der es um die wirtschaftliche und soziale Lage der Studierenden geht und deren Erleben im Studium. Praktisch durchgeführt werden die Erhebungen seit 1982 vom HIS, heute DZHW, finanziell unterstützt vom BMBF.

Der zum Zeitpunkt der Entstehung dieser Arbeit aktuellste Datensatz aus den Sozialerhebungen ist der der 21. Sozialerhebung (Middendorff et al., 2016). Eine kurze Beschreibung soll die Herkunft der Daten beleuchten: Diese umfangreiche Datenerhebung unter Studierenden in Deutschland fand zwischen Mai 2016 und August 2016 statt. Im Vergleich zu vorherigen Sozialerhebungen unter den

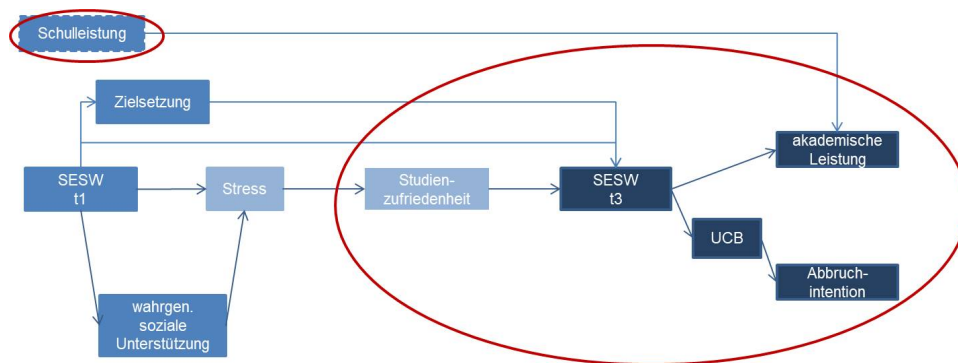


Abbildung 10.1: Teilbereich des EOS-Modells, der mit Daten der 21. Sozialerhebung analysiert wurde (rot markiert)

Studierenden fand diese Erhebung erstmals online statt: Einladungen zur Befragung wurden per Email versendet. Die Grundgesamtheit waren alle im Sommersemester 2016 an deutschen Hochschulen immatrikulierten Studierenden, wobei Studierende an Verwaltungsfachhochschulen, Hochschulen des Fernstudiums und an Universitäten der Bundeswehr ausgenommen wurden (Middendorff et al., 2017). Jeder sechste Studierende (entspricht circa 16,7%) wurde in die Stichprobe aufgenommen, was dazu führte, dass circa 400.000 Studierende zur theoretischen Stichprobe gehörten. Die realisierte Stichprobe umfasst letztendlich 67.007 deutsche und bildungsinländische Studierende von 248 Hochschulen. Damit beträgt die Brutto-Rücklaufquote 19,6%.

Die Daten werden vom DZHW nicht nur gesammelt, sondern auch aufbereitet, bevor sie jedermann zu Forschungszwecken auf Anfrage zur Verfügung gestellt werden: Nach der Datenbereinigung auf Basis von Plausibilitätsprüfungen verblieben 55.219 Fälle im Datensatz, was einer Netto-Rücklaufquote von 16,2% entspricht und somit unterhalb der Quote der vorherigen Sozialerhebung bleibt (Middendorff et al., 2016, S. 11).

Konkret habe ich aus dem Datensatz der Befragung der Bildungsinländer aus dem Jahr 2016 die Daten der Teilstichprobe der Studienanfänger verwendet, welche unter 10.3.3 detailliert beschrieben wird.

Abbildung 10.1 zeigt, welcher Teil des EOS-Modells mit den Daten der 21. Sozialerhebung analysiert werden kann. Konkret konnte ich aus dem Datensatz folgende Variablen verwenden: Abiturnote, (Studien-)Zufriedenheit, Selbstwirksamkeit, (ein Maß für) UCB, Studienleistung (aktuelle Durchschnittsnote) und Abbruchintention.

10.3 Methode

In diesem Abschnitt wird zunächst auf den ausgewählten Teildatensatz eingegangen. Anschließend soll dargelegt werden, für welche Art der Analysen ich mich entschieden habe. Zuletzt wird die analysierte Stichprobe genauer beschrieben sowie die erhobenen Variablen detailliert dargestellt.

10.3.1 Auswahl eines Teildatensatzes

Da im Datensatz der 21. Sozialerhebung Daten von Studierenden aller Semester und aller Studiengänge gesammelt sind, habe ich im ersten Schritt eine Substichprobe zusammengestellt. Konkret wurden in die folgenden Analysen nur Personen einbezogen, die im ersten oder zweiten Studiensemester waren und entweder in einem Bachelor- oder einem Lehramts- oder einem zum Staatsexamen führenden Studiengang eingeschrieben waren. Das bedeutet, alle Daten von Studierenden, die nicht mehr in der Studieneingangsphase waren und damit auch von solchen, die einen Masterstudiengang belegten, wurden ausgeschlossen. Damit ergab sich ein deutlich reduzierter Umfang für die Substichprobe von $N = 8713$. Da wie erwähnt Gruppenunterschiede zwischen den Fachclustern untersucht werden sollten, wurden weitere 45 Personen ausgeschlossen, weil sie keine Angaben zu ihrem Studienfach gemacht hatten.

Danach wurde das auch für die Längsschnittstudien verwendete Einschlusskriterium angelegt: Für die beiden vorherzusagenden Variablen *aktuelle Studiennote* und *Abbruchintention* mussten Daten vorliegen, damit der jeweilige Fall in den Analysen berücksichtigt wird. Zwar führte das nur zu einer geringfügigen Reduktion des Stichprobenumfangs, jedoch zeigte die weitere Inspektion der Daten, dass mehr als die Hälfte der verbleibenden Datensätze *item-nonresponse* in den Items zur Erfassung der Selbstwirksamkeit und zur Erfassung des UCB aufwiesen. Hinzu kam, dass nicht von *missing at random* ausgegangen werden kann, da Personen, die Items zur Erfassung der Selbstwirksamkeit nicht beantwortet hatten, zumeist auch die Items zur Erfassung des UCB nicht beantwortet hatten. Dies könnte darauf zurückzuführen sein, dass die beiden Itemgruppen im Fragebogen nahe bei einander angeordnet waren: beide Itemgruppen gehören zum Block *Psychologische Merkmale*, welcher weit hinten in der Umfrage angeordnet war (Itemgruppe 157 und 161 von 168). Im Gesamtdatensatz zeigt sich, dass Personen, die Items zur Erfassung dieser beiden Konstrukte nicht beantworteten, oft den ganzen Block nicht beantworteten. Aufgrund dieses Musters bei den fehlenden Antworten, entschied ich mich gegen die Verwendung der *full information maximum likelihood Methode* (siehe 4) zur Imputation fehlender Werte und schloss stattdessen alle Personen mit diesem Muster von fehlenden Werten von der Analyse aus. Dies führte letztendlich zu einer Analysestichprobe vom Umfang $N = 2523$.

Zwar ist somit der ursprüngliche Datensatz deutlich reduziert, allerdings dient diese Beschränkung der Qualitätssicherung im Sinne der Steigerung der Verlässlichkeit der Ergebnisse, zumal eine Imputation dieser beiden Prädiktoren eine Imputation der Hälfte der berücksichtigten Prädiktoren in den betroffenen Datenreihen bedeuten würde.

10.3.2 Art der Analysen

Aufgrund der Tatsache, dass (i) dieser Datensatz querschnittliche Daten liefert und dass (ii) nur die zweiten Hälfte des abgebildeten Prozesses geprüft werden kann (da die Variablen, die in der

ersten Hälfte des im *EOS*-Modell abgebildeten Prozesses aufgeführt sind, nicht im Datensatz enthalten sind), entschied ich statt einer Pfadanalyse multiple Regressionsanalysen durchzuführen (eine für jedes der beiden Kriterien). Der Grund dafür ist, dass im Modell wesentlich mehr Pfade spezifiziert sind, als in dieser Sekundärdatenanalyse berücksichtigt werden können. Daher scheint ein Vergleich des Modellfits auf Basis der Längsschnittdaten mit dem Modellfit auf Basis der Daten der 21. Sozialerhebung wenig sinnvoll. Die Prüfung, ob zumindest die im Datensatz der 21. Sozialerhebung vorhandenen im *EOS*-Modell spezifizierten Variablen einen bedeutsamen Anteil der Varianz in den beiden Kriterien aufklären können jedoch, stellt allerdings eine sinnvolle (Teil-)Validierung des *EOS*-Modells dar.

10.3.3 Analysestichprobe

Die analysierte Teilstichprobe (= Analysestichprobe) hatte einen Umfang von $N = 2523$. 54.2% (1368 Personen) der Befragten gaben an, weiblich zu sein, 45.1% (1138 Personen) gaben an männlich zu sein, 17 Personen machen dazu keine Angaben. Das Alter wurde in Kategorien abgefragt: 35.0% (884) Personen waren 19 Jahre alt oder jünger, 45.7% (1152) waren zwischen 20 und 22 Jahren alt, 11,7% (295) waren zwischen 23 und 25 Jahre alt und 7.4% (187) Personen waren 26 Jahre alt oder älter. Die Verteilung auf die Fachcluster ist Tabelle 10.1 zu entnehmen.

Tabelle 10.1: Verteilung auf die Fachcluster

Fachcluster	absolute Häufigkeit	Prozent
Geisteswiss.	298	11.8
Sport	28	1.1
Rechts-, Wirtschafts-, Sozialwiss.	914	36.2
Mathematik, Naturwiss.	322	12.8
Medizin, Gesundheitswiss.	115	4.6
Agrar-, Ernährungswiss.	76	3.0
Ingenieurwiss.	732	29.0
Kunst	38	1.5

Um die gewünschte Repräsentativität der folgenden Analyseergebnisse für die Population der Studieneinsteiger in Deutschland zu erzielen, entschied ich, die Fälle so zu gewichten, dass die Analysestichprobe in Bezug auf die Besetzung der Fächercluster repräsentativ ist. Selbstverständlich konnten für diese Gewichtung nicht die komfortabler Weise im Datensatz der 21. Sozialerhebung mitgelieferten Gewichtungen verwendet werden, da sich diese auf eine Adjustierung des Gesamtdatensatzes beziehen. Folglich erstellte ich selbst eine Gewichtungsvariable und zog dafür die Daten zur Beset-

zung der Fachcluster innerhalb der Studieneinsteiger heran, die das deutsche statistische Bundesamt ¹ zur Verfügung stellt. Da hier ausschließlich Daten aus den jeweiligen Wintersemestern berichtet werden, wählte ich die Kohorte der Studieneinsteiger im Wintersemester 2015/2016 als Referenz. Indem ich dann die prozentuale Fächerbesetzung in meiner Analytestichprobe in Beziehung setzte zur prozentualen Fächerbesetzung der offiziellen Statistik des deutschen statistischen Bundesamtes, berechnete ich die Gewichtungen. Diese Gewichtungen verwendete ich für die folgenden Analysen. Hierbei möchte ich anmerken, dass die Adjustierungen von geringen Ausmaß waren, also die Analytestichprobe in Bezug auf die Besetzung der Fächer nur wenig von der Verteilung in der Population abwich.

10.3.4 Betrachtete Variablen

Wie Abbildung 10.1 zeigt, sollte mit dieser Sekundärdatenanalyse nur ein Teil des EOS-Modells getestet werden. Konkret zog ich für die Analysen sechs Variablen heran: Abiturnote, Zufriedenheit, Selbstwirksamkeit, ein Maß für UCB und die beiden Kriterienvariablen *aktuelle Studiennote* und *Abbruchintention*. Folgendermaßen wurden diese Variablen erfasst:

- Abiturnote (einzelnes Item): Welche Durchschnittsnote hatten Sie in dem Abschlusszeugnis, das Sie zur Aufnahme eines Hochschulstudiums berechtigt? (Schulnotenskala; Freitextitem, eine Dezimalstelle)
- Zufriedenheit (einzelnes Item): Wie zufrieden sind Sie insgesamt mit den bisher von Ihnen erbrachten Studienleistungen in Ihrem aktuellen Studium? (1 = *sehr unzufrieden* bis 5 = *sehr zufrieden*)
- Selbstwirksamkeit (drei Items): *Kurzskala zur Erfassung allgemeiner Selbstwirksamkeitserwartungen (ASKU)* von Beierlein, Kovaleva, Kemper und Rammstedt (2013), Bsp.: Auch anstrengende und komplizierte Aufgaben kann ich in der Regel gut lösen. (1 = *trifft gar nicht zu* bis 5 = *trifft voll und ganz zu*)
- UCB (sechs Items): Items zur *akademischen Passung* die zwar meiner Kenntnis nach keiner publizierten Skala entnommen sind, allerdings in dieser Befragung als Itembatterie präsentiert wurden. Aufgrund der inhaltlichen Homogenität der Items (siehe unten), entschied ich, diese zu einem Skalenwert zusammenzufassen. Bsp.: Alles in allem bin ich gerne Student(in). (1 = *trifft gar nicht zu* bis 5 = *trifft voll und ganz zu*)
- aktuelle Studiennote (einzelnes Item): Wie wurden Ihre bisherigen Studienleistungen in Ihrem aktuellen Studiengang im Durchschnitt bewertet? (Schulnotenskala; Freitextitem, eine Dezimalstelle)

¹<https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/studierende-erstes-fs-faechergruppen.html>

- Abbruchintention (einzelnes Item): Beabsichtigen Sie zur Zeit, Ihr aktuelles Studium abzubrechen? (1 = *nein, auf keinen Fall* bis 5 = *ja, auf jeden Fall*)

Sowohl die drei Items zur Erfassung der Selbstwirksamkeit (im Folgenden: ASKU-Skala) als auch die sechs Items zur Erfassung der akademischen Passung, welche in diesem Zusammenhang ein verkürztes Äquivalent zur UCB-Skala nach Gehring (2006) bilden sollen (im Folgenden: UCB-Skala), fasste ich jeweils durch eine Mittelwertbildung zu einem Skalenwert zusammen. Beide Skalen wiesen eine gute Reliabilität auf: $\alpha = .845$ bei der ASKU-Skala und $\alpha = .801$ bei der UCB-Skala ($N = 2523$). Es bleibt allerdings zu bedenken, dass die ASKU-Skala in keiner Weise die Selbstwirksamkeit kontextspezifisch erfasst, wie es die in meinen Längsschnittstudien verwendete SESW-Skala tut (vgl. Kapitel 3). Dass allerdings die kontextspezifisch erfasste Selbstwirksamkeit in der Regel eine höhere (prädiktive) Validität aufweist als die allgemein erfasste, ist hinlänglich bekannt (Betz & Hackett, 2006). Daher ist es davon auszugehen, dass das sehr spezifische Kriterium *Abbruchintention* hier in keinem engen Zusammenhang mit diesem allgemeinen Maß für Selbstwirksamkeit steht.

Mit der aus sechs Items zusammengestellten UCB-Skala verhält es sich ähnlich. Es ist nicht davon auszugehen, dass sie eine so umfangreiche Erfassung des *University Citizenship Behavior* leistet wie die Skala von Gehring (2006). Insbesondere fehlen in den hier berücksichtigten Items die Aspekte des UCB, die sich auf Gewissenhaftigkeit beziehen. Da allerdings bekannt ist, dass unter anderem eben diese Eigenschaft einen bedeutsamen Zusammenhang mit akademischer Leistung aufweist (siehe 2.1.5), ist ebenfalls zu vermuten, dass die hier verwendete Operationalisierung keinen bedeutsamen Beitrag zur Varianzaufklärung im Kriterium *aktuelle Studiennote* leisten kann.

Die Verwendung der ASKU-Skala und der UCB-Skala aus dem Datensatz der 21. Sozialerhebung erfolgt demnach versuchsweise und unter Berücksichtigung der möglicherweise eingeschränkten Validität dieser Operationalisierungen.

10.4 Ergebnisse

In diesem Abschnitt werden zunächst die deskriptiven Statistiken vorgestellt. Anschließend berichte ich die durchgeführten Präanalysen (Prüfungen der Voraussetzungen für die multiplen Regressionsanalysen) und die Ergebnisse der multiplen Regressionsanalysen.

10.4.1 Deskriptive Statistiken

Tabelle 10.2 zeigt die Mittelwerte und Standardabweichungen der für die multiplen Regressionsanalysen berücksichtigten Variablen.

Den beiden vorherzusagenden Kriterien kommt dabei unter den berücksichtigten Variablen eine besondere Rolle zu, sodass deren deskriptive Darstellung an dieser Stelle etwas detaillierter erfolgen soll.

Tabelle 10.2: Mittelwert und Standardabweichung der berücksichtigten Variablen ($N = 2.523$)

	M	SD
Abitur	2.163	0.630
Zufriedenheit	3.479	1.103
Selbstwirksamkeit	3.890	0.679
UCB	4.171	0.628
aktuelle Note	2.262	0.680
Abbruchintention	1.533	0.946

Die Verteilung der Abbruchintention nach Fachclustern ist in Abbildung 10.2 und die der mittleren aktuellen Studiennote nach Fachclustern in Abbildung 10.3 dargestellt.

Abbildung 13.2 ermöglicht, in Form eines Korrelogramms, einen Überblick über die bivariaten Zusammenhänge der analysierten Variablen zu erhalten.

Eine ANOVA zeigte, dass sich die mittlere aktuelle Studiennote zwischen den Fachclustern unterschied: $F_{aktNote}(1, 2521) = 77.31$ ($p < .001$). Daher entschied ich mich, die aktuelle Studiennote für die weiteren Analysen pro Fachcluster zu zentrieren. Das mittlere Niveau an Abbruchintentionen hingegen unterschied sich nicht signifikant zwischen den Fachclustern ($F_{Abb}(1, 2521) = 0.70$, $p = .404$).

10.4.2 Präanalysen

Nachdem ich somit die Analysestichprobe erstellt und die Gewichtung vorbereitet hatte, führte ich die Präanalysen zur Prüfung der Erfüllung der Voraussetzungen für die multiplen Regressionen durch. Für die geplanten multiplen Regressionen sind grundsätzlich drei Voraussetzungen zu prüfen (Bortz & Schuster, 2010; Osborne & Waters, 2002).

- Linearität: Der Erwartungswert des Kriteriums in der Population lässt sich als Linearkombination der Prädiktorwerte darstellen (vgl. Bortz und Schuster, 2010, S. 348).
- Normalität: „Die Verteilung der y -Werte für jede Kombination von Prädiktorwerten ist eine Normalverteilung“ (Bortz und Schuster, 2010, S. 348). Das bedeutet, dass die Residuen für jede Kombination von Prädiktorwerten normalverteilt sein sollen (Bortz und Schuster, 2010, S. 193).
- Homoskedastizität: „Die Varianz der y -Werte, welche für eine bestimmte Kombination von Prädiktorwerten vorliegt, ist über alle Kombinationen von Prädiktorwerten hinweg konstant“

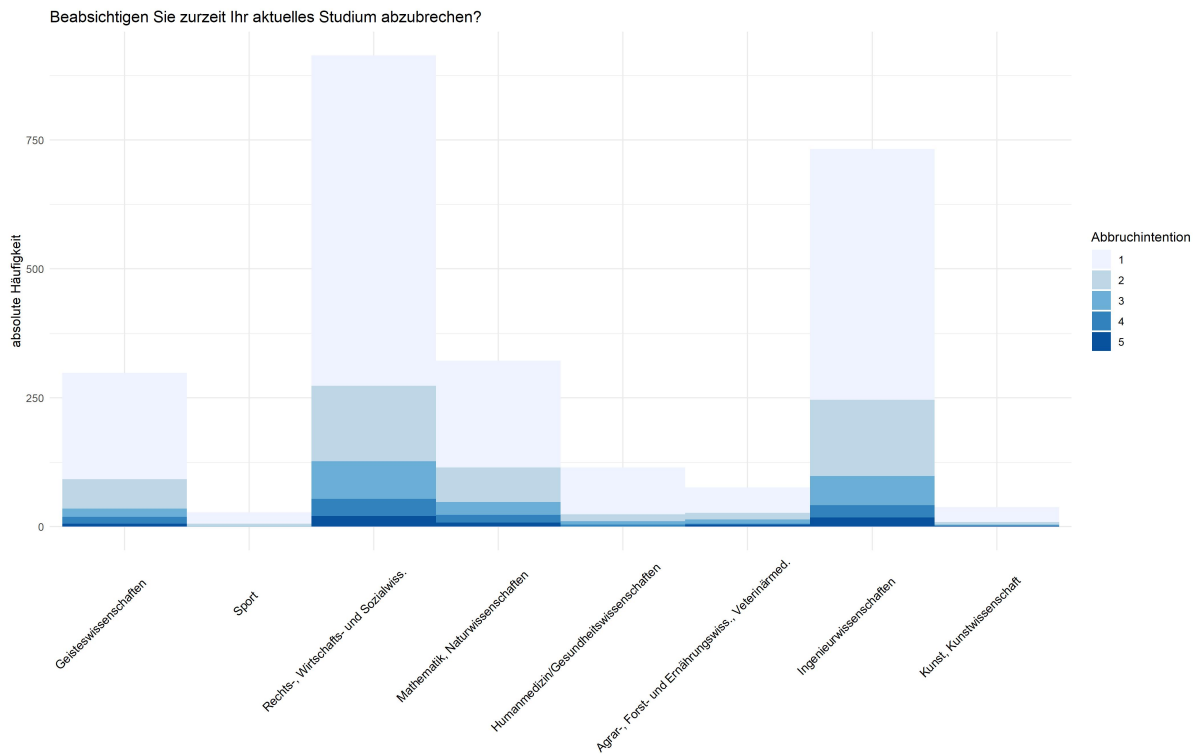


Abbildung 10.2: Abbruchintention nach Fachclustern (1 = *nein, auf keinen Fall* bis 5 = *ja, auf jeden Fall*)

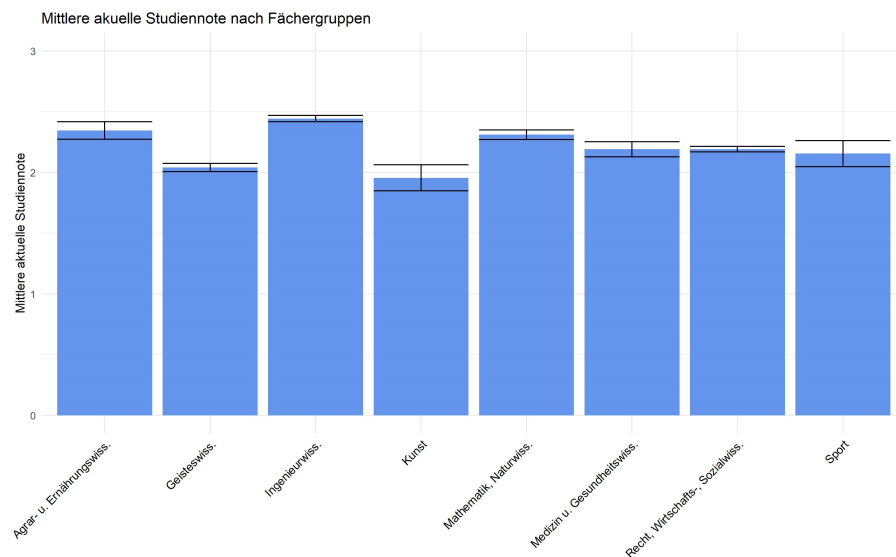


Abbildung 10.3: Durchschnittliche aktuelle Studiennote nach Fachclustern

(Bortz und Schuster, 2010, S. 348). Das bedeutet, dass die Residuen für alle Kombinationen von Prädiktorwerten die gleiche Varianz aufweisen sollen (Bortz und Schuster, 2010, S. 193).

Gemäß Bortz und Schuster (2010) lassen sich die beiden letztgenannten Voraussetzungen auch folgendermaßen zusammenfassen: Die Verteilung der Residuen sollen für jede Kombination von Prädiktoren eine Normalverteilung mit gleicher Varianz sein. Dies lässt sich sowohl durch eine visuelle Inspektion der Residuenverteilungen als auch beispielsweise mit Hilfe eines Kolmogorov-Smirnov-Tests prüfen (Baltes-Götz, 2019).

Abbruchintention

Für die Analysen hinsichtlich der *Abbruchintention* lässt sich dazu Folgendes feststellen:

- (i) Dass eine lineare Beziehung zwischen den Regressoren und dem Kriterium besteht, ist letztendlich inhaltlich zu begründen. Wie bereits beschrieben, gibt es für die Regressoren zahlreiche Befunde aus vorangegangenen Studien, die lineare Zusammenhänge, beispielsweise in Form von Korrelationen oder Regressionen nachweisen konnten (Robbins et al., 2004; Richardson et al., 2012; M. Schneider & Preckel, 2017). Daher nahm ich für die folgenden Analysen an, dass dieser lineare Zusammenhang besteht.
- (ii) Die Verteilung der Residuen kann beispielsweise mit Hilfe des Kolmogorov-Smirnov-Tests überprüft werden. Hierbei ergab sich für das Kriterium *Abbruchintention*, dass die Residuen nicht normalverteilt sind: $D(2,523) = .151, p < .001$.
- (iii) Weiterhin offenbarte die visuelle Inspektion der Verteilung der Residuen (siehe Abbildung 10.4), dass keine Homoskedastizität angenommen werden kann. Baltes-Götz (2019) empfiehlt im Falle einer multiplen linearen Regression die genauere Inspektion der Residuen separat für alle Regressoren. Abbildung 10.5 zeigt die entsprechenden Diagramme, die wie das gemeinsame Diagramm zeigen, dass keine Homoskedastizität angenommen werden kann.

Aktuelle Studiennote

Für die Analysen hinsichtlich der *aktuellen Studiennote* verhält es sich ähnlich.

- (i) Dass eine lineare Beziehung zwischen den Regressoren und dem Kriterium besteht, ist letztendlich inhaltlich zu begründen. Wie bereits beschrieben, gibt es für die Regressoren zahlreiche Befunde aus vorangegangenen Studien, die lineare Zusammenhänge, beispielsweise in Form von Korrelationen oder Regressionen nachweisen konnten (Robbins et al., 2004; Richardson et al., 2012; M. Schneider & Preckel, 2017). Daher nahm ich auch für das Kriterium *aktuelle Studiennote* an, dass dieser lineare Zusammenhang besteht.
- (ii) Die Verteilung der Residuen habe ich ebenfalls mit Hilfe des Kolmogorov-Smirnov-Tests überprüft. Hierbei ergab sich für das Kriterium *aktuelle Studiennote*, dass die Residuen nicht normalverteilt sind: $D(2,523) = .029, p < .001$.

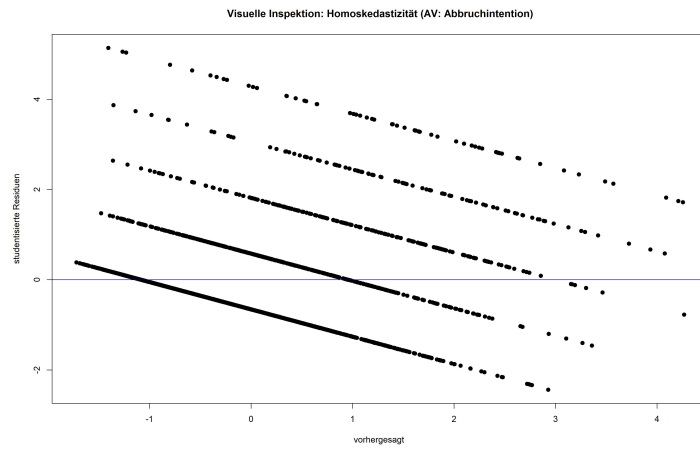


Abbildung 10.4: Residuenverteilung (AV: Abbruchintention)

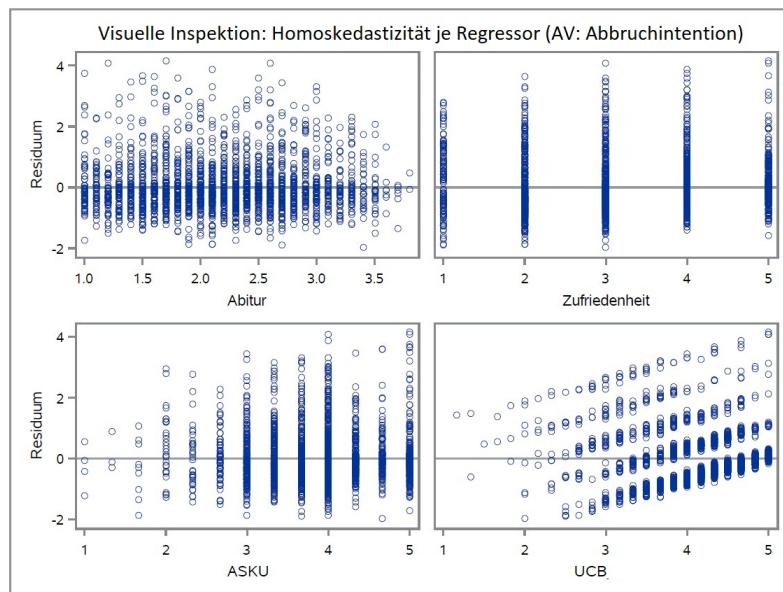


Abbildung 10.5: Residuen separat für die vier Regressoren (AV: Abbruchintention)

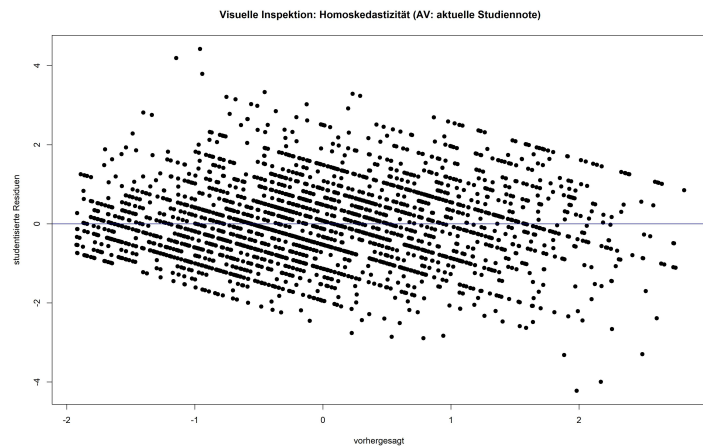


Abbildung 10.6: Residuenverteilung (AV: aktuelle Studiennote)

(iii) Für das Kriterium *aktuelle Studiennote* jedoch offenbart die visuelle Inspektion der Verteilung der Residuen (siehe Abbildung 10.6), dass in diesem Fall Homoskedastizität angenommen werden kann. Analog zur obigen Beschreibung finden sich in Abbildung 10.7 auch die Diagramme für die Residuen, separat für jeden Regressor.

In beiden Fällen empfiehlt es sich aufgrund der Verletzung der Voraussetzungen, die Analysen nicht nur einfach durchzuführen, sondern stattdessen Bootstrapping zu verwenden (Baltes-Götz, 2019). Dies habe ich getan und $k = 10.000$ Bootstrapsamples automatisch erzeugen lassen. Dadurch ergibt sich die Möglichkeit, die Ergebnisse mit Angabe eines Konfidenzintervalls (für die Regressorenengewichte) zu versehen und somit eine höhere Verlässlichkeit (Verbesserung der Schätzung der Standardfehler) dieser zu erreichen.

10.4.3 Analysen

Wie bereits beschrieben, sollte anhand der Daten der 21. Sozialerhebung nur ein Auszug aus dem EOS-Modell untersucht werden. Die zwei multiplen Regressionen, die berechnet wurden, sollten zur Testung zweier Hypothesen dienen, die sich auf die erste unter 10.1.1 genannte Forschungsfrage beziehen:

- H1: Die vom Modell spezifizierten Regressoren (Abiturnote, Selbstwirksamkeit, Studienzufriedenheit und Integration) können einen substantiellen Anteil der Varianz in der Abbruchintention der Studienanfänger aufklären.
- H2: Die vom Modell spezifizierten Regressoren (Abiturnote, Selbstwirksamkeit, Studienzufriedenheit und Integration) können einen substantiellen Anteil der Varianz in der aktuellen Studiennote der Studienanfänger aufklären.

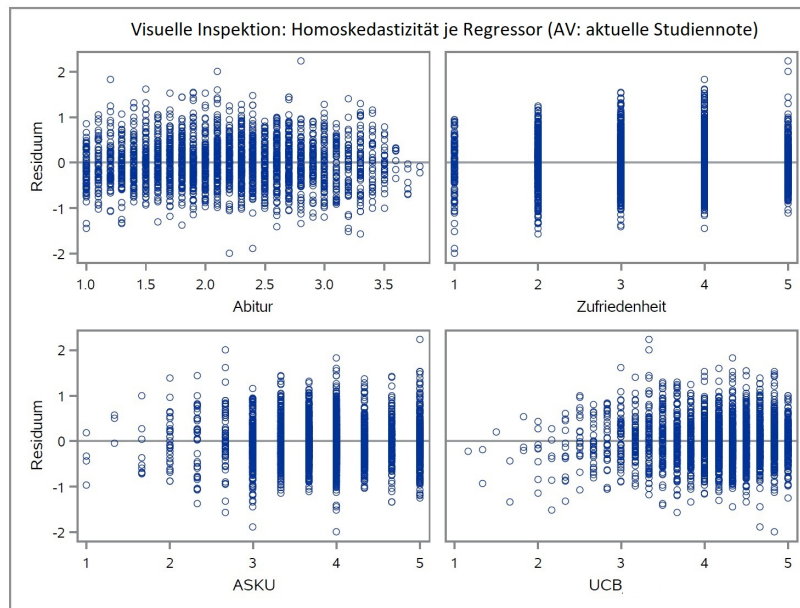


Abbildung 10.7: Residuen separat für die vier Regressoren (AV: aktuelle Studiennote)

Neben der Berechnung der beiden multiplen Regressionen für die gesamte Analytestichprobe, bot sich aufgrund des Stichprobenumfangs auch die Möglichkeit, die Analysen zusätzlich pro Fachcluster durchzuführen. Der Vergleich der so erzielten Ergebnisse sollte zeigen, ob ähnlich große Varianzanteile in den Kriterien in den verschiedenen Fachclustern aufgeklärt werden können. Somit sollen die Analysen pro Fachcluster dazu dienen, die zweite unter 10.1.1 genannte Fragestellung zu beantworten.

In Tabelle 10.3 und Tabelle 10.4 sind die Ergebnisse der beiden multiplen Regressionsanalysen aufgeführt.

Abbruchintention

Das Modell zur Vorhersage der *Abbruchintention* wurde signifikant: $F(4, 2518) = 230.66, p < .001$. Alle Regressoren bis auf die *Selbstwirksamkeit* erwiesen sich als sehr signifikant. Insgesamt konnten ganze 27% der Varianz im Kriterium *Abbruchintention* aufgeklärt werden. Dies kann als großer Effekt eingestuft werden (J. Cohen, 1988).

Tabelle 10.3: Regression der Abbruchintention auf die Modellvariablen

	<i>AV: Abbruchintention</i>		
	B	SE	β
Abitur	-0.113*** [-0.16, -0.06]	0.026	-0.07***
Zufriedenheit	-0.092*** [-0.12, -0.06]	0.016	-0.07***
Selbstwirksamkeit	-0.014 [-0.06, 0.04]	0.026	-.01
UCB	-0.704*** [-0.76, -0.65]	0.029	-.47***
Konstante	5.088***	0.147	
R^2	.267		
R^2_{corr}	.266		

Anm. *** $p < .01$ $N = 2523$.

Aktuelle Studiennote

Das Modell zur Vorhersage der *aktuellen Studiennote* wurde ebenfalls signifikant: $F(4, 2518) = 548.56$, $p < .001$. Alle Regressoren bis auf das *UCB* erwiesen sich als sehr signifikant. Insgesamt konnten ganze 46% der Varianz innerhalb der Variable *aktuelle Studiennote* aufgeklärt werden; ebenfalls ein großer Effekt (J. Cohen, 1988). Tabelle 10.4 zeigt die Ergebnisse im Detail. Diese Ergebnisse beziehen sich auf die gesamte Analytestichprobe.

Tabelle 10.4: Regression der aktuellen Studiennote auf die Modellvariablen

	<i>AV: aktuelle Studiennote</i>		
	B	SE	β
Abitur	0.261*** [0.22, 0.29]	0.016	0.25***
Zufriedenheit	-0.356*** [-0.39, -0.35]	0.010	-0.59***
Selbstwirksamkeit	-0.043*** [-0.07, -0.01]	0.015	-0.04***
UCB	0.012 [-0.03, 0.04]	0.018	0.01
Konstante	0.793***	0.089	
R^2	.462		
R^2_{corr}	.462		

Anm. *** $p < .01$ $N = 2523$.

Tabelle 10.5: Aufgeklärt Varianz im Kriterium *Abbruchintention* nach Studienfachcluster

Studienfachcluster	R_{corr}^2
Geisteswissenschaften (n = 298)	0.25
Sport (n = 28)	0.26
Rechts-, Wirtschafts- und Sozialwissenschaften (n = 914)	0.25
Mathematik und Naturwissenschaften (n = 322)	0.26
Medizin (n = 115)	0.28
Agrar- und Ernährungswissenschaften (n = 76)	0.28
Ingenieurwissenschaften (n = 732)	0.31
Kunst (n = 38)	0.33

Tabelle 10.6: Aufgeklärt Varianz im Kriterium *aktuelle Studiennote* nach Studienfachcluster

Studienfachcluster	R_{corr}^2
Geisteswissenschaften (n = 298)	0.45
Sport (n = 28)	0.69
Rechts-, Wirtschafts- und Sozialwissenschaften (n = 914)	0.46
Mathematik und Naturwissenschaften (n = 322)	0.41
Medizin (n = 115)	0.43
Agrar- und Ernährungswissenschaften (n = 76)	0.53
Ingenieurwissenschaften (n = 732)	0.53
Kunst (n = 38)	0.24

10.4.4 Die Frage der Generalisierbarkeit über die Fachcluster

Anschließend an die Analysen auf Basis der gesamten Analysestichprobe sollte verglichen werden, ob jeweils der Anteil aufgeklärter Varianz in den beiden Kriterien über die verschiedenen Fachcluster hinweg ähnlich groß ist. Die Tabellen 10.5 und 10.6 zeigen die Ergebnisse der fachclusterspezifischen Analysen.

Vergleicht man den Anteil aufgeklärter Varianz im Kriterium *Abbruchintention* zwischen den Fachclustern, so zeigt sich, dass in allen Fachclustern über 25% der Varianz aufgeklärt werden konnte. Die größte Varianzaufklärung (33%) liegt für das Fachcluster *Kunst* vor, wobei insbesondere bei diesem Fachcluster und beim Fachcluster *Sport* die Ergebnisse nur mit Vorsicht zu interpretieren sind, da jeweils nur $N = 38$ bzw. $N = 28$ Personen in den Subgruppen sind. Alle anderen Subgruppen weisen deutlich größere Stichprobenumfänge auf, sodass die Ergebnisse hinsichtlich der anderen Subgruppen mit weniger Bedenken interpretiert werden können. Insgesamt wird bei allen Fachclustern in

diesem Kriterium circa ein Viertel bis ein Drittel der Varianz durch die vom *EOS*-Modell spezifizierten Prädiktoren aufgeklärt. Hierbei ist zu bedenken, dass für diese Analysen nur ein Auszug aus dem *EOS*-Modell betrachtet wurde, also nicht alle im Modell spezifizierten Prädiktoren einfließen.

Was das Kriterium *aktuelle Studiennote* betrifft, so lässt sich festhalten, dass der Range an aufgeklärter Varianz deutlich größer ist als beim Kriterium *Abbruchintention*. In den verschiedenen Fachclustern konnte zwischen 24% (Kunst) und 69% (Sport) der Varianz in der *aktuellen Studiennote* aufgeklärt werden. Wie schon bezüglich des Kriteriums *Abbruchintention* angemerkt, sind eben diese beiden Fachcluster allerdings die beiden mit kritisch kleinem Stichprobenumfang. Lässt man sie außen vor, so ergibt sich ein Range von 41% bis 53% aufgeklärte Varianz. Das bedeutet, dass in den anderen Fachclustern (ohne Kunst und Sport) ein relativ vergleichbarer und beträchtlich hoher Anteil an Varianz in der *aktuellen Studiennote* aufgeklärt werden kann.

10.5 Diskussion

Mit der vorliegenden Sekundärdatenanalyse sollten zwei Forschungsfragen beantwortet werden.

10.5.1 Validierung des *EOS*-Modells

Zum einen sollte der Frage nachgegangen werden, ob die im *EOS*-Modell spezifizierten Prädiktoren auch in einem externen Datensatz einen bedeutsamen Beitrag zur Varianzaufklärung in den beiden Kriterien liefern können. Die zugehörigen Hypothesen 1 und 2 (siehe 10.4.3) konnten bestätigt werden: Die in die Analysen einbezogenen Prädiktoren konnten sowohl im Kriterium *Abbruchintention* als auch im Kriterium *aktuelle Studiennote* einen bedeutsamen, als groß einzustufenden Anteil der Varianz aufklären. Dies kann als partielle Validierung des *EOS*-Modells gewertet werden.

Es ist jedoch anzumerken, dass die ASKU-Skala bei der Varianzaufklärung im Kriterium *Abbruchintention* keinen signifikante Beitrag leistet. Das Gleiche gilt für den Beitrag der UCB-Skala bei der Varianzaufklärung in der *aktuellen Studiennote*. Hierzu lässt sich anmerken, dass, wie eingangs erwähnt, im Zuge dieser Sekundärdatenanalyse auf vorhandene Operationalisierungen zurückgegriffen werden musste.

Im Vergleich zu den Ergebnissen der drei Längsschnittstudien ist ein weiterer Aspekt auffällig. In dieser Sekundärdatenanalyse leistete die Abiturnote bei der Varianzaufklärung in beiden Kriterien einen bedeutsamen Beitrag, was in den Längsschnittstudien so deutlich nicht zu beobachten war. Damit deckt sich der Befund anhand der Daten der 21. Sozialerhebung mit vorangegangenen Befunden, wie denen von Trapmann et al. (2007), Robbins et al. (2004) und Richardson et al. (2012).

Vergleicht man den Anteil aufgeklärter Varianz im Kriterium *Abbruchintention*, der in dieser Untersuchung durch einige der im *EOS*-Modell spezifizierte Prädiktoren geleistet wurde mit dem Anteil an Varianzaufklärung, die Isleib et al. (2019) für die erste empirische Testung ihres Modells berichten

($Pseudo-R^2 = .34$, $N = 5580$), so lässt sich festhalten: Die im *EOS*-Modell spezifizierten Prädiktoren haben in der vorliegenden Untersuchung einen ähnlich großen Varianzanteil aufgeklärt. Selbstverständlich kann kein belastbarer Vergleich zwischen der vorliegenden Untersuchung und der von Isleib et al. (2019) gezogen werden, da unterschiedliche Datensätze mit unterschiedlichen Operationalisierungen (von Studienabbruch(-intentionen) und anderen Variablen) verwendet wurden, jedoch gibt der rein nominelle Vergleich der Höhe der Varianzaufklärung Anlass, in weitergehenden Untersuchungen einen verlässlichen, inferenzstatistisch abgesicherten Vergleich der Modelle zu vollziehen.

10.5.2 Vergleich der Varianzaufklärung zwischen den Fachclustern

Zum anderen wurden auf Basis des umfangreichen Datensatzes aus der 21. Sozialerhebung die Anteile an aufgeklärter Varianz in den beiden Kriterien zwischen den Fachclustern verglichen. Zwar sind die Fachcluster *Kunst* und *Sport* nur dünn besetzt, sodass die Ergebnisse für diese beiden Fachcluster mit Vorsicht zu interpretieren sind (geringe statistische Power), lässt man diese beiden Fachcluster jedoch außen vor, so kann man zusammenfassend sagen, dass in den verbleibenden sechs Fachclustern (über 97% der Stichprobe) jeweils ähnlich große Anteile an Varianz in den beiden Kriterien aufgeklärt werden konnten.

10.5.3 Fazit

Abschließend lässt sich festhalten, dass mit der vorliegenden Sekundärdatenanalyse eine partielle Validierung des *EOS*-Modells an einer großen, hinsichtlich der Fachclusterbesetzung repräsentativen Stichprobe von Studienanfänger in Deutschland durchgeführt werden konnte. Obwohl nur ein Teil des *EOS*-Modells berücksichtigt werden konnte, konnten die einbezogenen Variablen sogar einen großen Beitrag zur Varianzaufklärung in den beiden Kriterien *Abbruchintention* und *aktuelle Studiennote* leisten.

Im Vergleich zwischen den Fachclustern zeigte sich, dass sich der Erklärungswert des *EOS*-Modells über die Fachcluster hinweg generalisieren lässt.

10.5.4 Limitationen

Die beschriebenen Limitationen, die mit der Erfassung von Konstrukten anhand einzelner Items oder sehr kurzen Skalen einhergehen, sind bereits benannt worden: Es ist davon auszugehen, dass die verwendeten Operationalisierungen die betrachteten Konstrukte nicht umfassend abbilden können.

10.5.5 Ausblick

Um die Daten aus der Sozialerhebung zukünftig noch besser zur Beantwortung differentiell-diagnostischer Fragestellungen nutzbar zu machen, wären einige Desiderate zu adressieren: Wünschenswert wäre die Inkludierung zumindest einiger etwas umfangreicherer Skalen zur Selbstwirksamkeit (im Studieneinstieg), Studienzufriedenheit und zur soziale Integration (und idealerweise auch zum UCB).

Das DZHW, welches die Sozialerhebungen durchführt, schreibt auf seiner Projektseite² zumindest, dass in der Sozialerhebung neben den Kernthemen auch immer wieder zusätzliche Themen berücksichtigt werden. Gegebenenfalls wäre es im Rahmen eines solche zusätzlichen Themenblocks möglich, einige ausführliche Skalen zu psychosozialen Themenbereichen zu inkludieren.

Gleiches gilt für andere bundesweite Befragungen innerhalb der Population der Studienanfänger beziehungsweise Studierenden. So wäre die Berücksichtigung dieser Konstrukte beispielsweise in den NEPS-Befragungen ebenfalls wünschenswert, wodurch sogar die Möglichkeit geschaffen würde, Zusammenhänge der im *EOS*-Modell spezifizierten Variablen mit longitudinal gesammelten Studienerfolgsmaßen zu analysieren.

²https://www.dzhw.eu/forschung/projekt?pr_id=460

11 Interventionsstudie

Wie in Kapitel 1 erwähnt, ist einer der intendierten Beiträge dieser Arbeit, anhand eines Beispiels auszuloten, inwieweit sich die im Rahmen der Modelltests gewonnenen Erkenntnisse auf die Hochschulpraxis übertragen lassen. Vor dem Hintergrund, dass im Fokus dieser Arbeit insbesondere solche Aspekte stehen, die potentiell geeignete Ansatzpunkte institutioneller Interventionen sind, möchte ich in diesem Kapitel von einer ersten Pilotstudie berichten, mit der getestet wurde, ob eine exemplarisch ausgewählte Variable aus dem EOS-Modell im Rahmen eines Kurz-Workshops bedeutsam verändert werden kann.

11.1 Hintergrund

In der besagten Pilotstudie soll zunächst exemplarisch der Effekt der Teilnahme an einem Kurz-Workshop auf eine der im EOS-Modell zentralen Variablen, untersucht werden. Hierbei fiel die Wahl auf die *Selbstwirksamkeit im Studieneinstieg*, unter anderem, weil eine Skala zur Erfassung dieser (die SESW-Skala, siehe Kapitel 3) im Rahmen dieses Promotionsprojektes entwickelt wurde und aktuelle, umfangreiche Normierungsdaten aus der Population der Studienanfänger in Deutschland für diese Skala vorliegen.

Es ist anzunehmen, dass die Selbstwirksamkeit im Studieneinstieg beispielsweise durch Interventionen gezielt positiv beeinflusst werden kann. Diese Annahme fußt einerseits auf der Definition von kontextspezifischer Selbstwirksamkeit als Eigenschaft, die sich über die Zeit hinweg verändern kann, unter anderem aufgrund von Erfolgs- und Misserfolgserfahrungen, die eine Person in einem bestimmten Kontext macht (Bandura, 1977). Andererseits wird die Annahme durch die Befunde von van Dinther et al. (2011) gestützt (siehe 2.1.6, wie unten stehend aufgeführt).

Im klinischen Bereich, beispielsweise bei der Behandlung von Phobien, konnte gezeigt werden, dass in Therapiesettings gemachte Erfolgserfahrungen, tatsächlich die Selbstwirksamkeit der Patienten im jeweiligen Kontext steigerten (Bandura, 1980). Es liegt daher nahe anzunehmen, dass die Auseinandersetzung mit dem erfolgreichen Umgang mit Herausforderungen im Studieneinstieg beispielsweise in einem Workshop auch positive Effekte auf die Selbstwirksamkeit der Teilnehmer im Hinblick auf die Bewältigung dieser Herausforderungen haben könnte.

11.2 Metaanalytische Befunde zur Wirksamkeit von Interventionen im Studium

Evans und Burck (1992) führten eine Metaanalyse zur Wirksamkeit von Interventionen im US-amerikanischen Bildungskontext durch. Dabei berücksichtigten sie 67 Studien, die zwischen 1966 und 1986 durchgeführt wurden. Insgesamt kamen sie auf eine Effektstärke von Cohen's $d = .16$, was im Allgemeinen als kleiner Effekt einzustufen ist. Robbins et al. (2004) würdigen diese – wenn auch kleine – Effektstärke als praktisch bedeutsam: Könnte man anhand von durchschnittlichen Schulnoten und psychosozialer Faktoren sowie studienbezogener Fertigkeiten voraussagen, welche Studierenden gegebenenfalls Leistungsschwierigkeiten bekommen werden, so könnte man diesen dann gezielt Interventionsprogramme anbieten, von denen man nun auf Basis der Metaanalyse von Evans und Burck (1992) sagen könnte, dass sie akademische Leistung positiv beeinflussen.

Hattie, Biggs und Purdie (1996) veröffentlichten vier Jahre später eine weitere Metaanalyse, in die sie 51 Studien (respektive 270 Effektstärken) ebenfalls vornehmlich aus dem US-amerikanischen Bildungskontext einbezogen. Sie kategorisierten die Interventionen anhand der Komplexität der mit ihnen intendierten Lernergebnisse. Die „Komplexitätsstufen“ lauten: (1) unistructural (intendierte Lernergebnisse haben nur eine Dimension, beispielsweise wird nur eine Strategie vermittelt), (2) multistructural (mehrere Strategien werden unabhängig voneinander erlernt; keine Verknüpfung dieser), (3) relational (mehrere Strategie und deren kontextbezogenes Zusammenspiel werden erlernt) und (4) extended abstract (die komplexe Kombination verschiedener Strategien wird kontextspezifisch erlernt und dann noch im Zuge der Intervention in einen anderen Kontext übertragen). Die so gebildeten vier „Stufen“ der unabhängigen Variable „strukturelle Komplexität der Intervention“ wurde metaanalytisch in Beziehung gesetzt zu den drei Gruppen von abhängigen Variablen: akademische Leistung, Study Skills und Affekt. Selbstwirksamkeit wird dabei explizit als der letzten Kategorie zugehörig genannt.

Insgesamt lässt sich festhalten, dass Interventionen mittlerer struktureller Komplexität die größte Effektstärken im Hinblick auf alle drei Gruppen von abhängigen Variablen ausweisen. Ein Beispiel für eine Intervention mit mittlerer Komplexitätsstufe wäre laut Hattie et al. (1996) ein typischer Study Skills Kurs, bei dem unterschiedliche Strategien oder Vorgehensweisen thematisiert und geübt werden, ohne jedoch einen metakognitiven oder konditionalen Aspekt (Wie hängen verschiedene Strategien zusammen? Wann ist welche Herangehensweise erfolgsversprechend?) zu berücksichtigen. Über alle betrachteten Studien hinweg berichten die Autoren eine Effektstärke auf Studienebene von Cohen's $d = 0.63$ (respektive $d = 0.45$ auf Ebene der Einzeleffektstärken), was als gute, mittlere Effektstärke zu werten ist. Hinzu kommt, dass die meisten berücksichtigten Interventionen Kurz-Interventionen von ein bis zwei Tagen Dauer waren, was ihren Ergebnissen zur Folge auch die effektivste Interventionsdauer zu sein scheint.

Aus einer weiteren Metaanalyse zur Effektivität von institutionellen Angeboten, diesmal zur Stressre-

duktion, in die 24 Studien mit Daten von insgesamt $N = 1431$ Studierenden wieder vornehmlich aus dem US-amerikanischen Raum einbezogen wurden, geht ebenfalls hervor, dass bereits mit relativ kurzen Interventionen statistisch bedeutsame Effekte erzielt werden können: So berichten Regehr, Glancy und Pitts (2013) von einem mittleren standardisierten Mittelwertsunterschied zwischen Kontroll- und Interventionsgruppe von $\Delta = -.077$ im – in allen berücksichtigten Studien verwendeten – Fragebogen zur Erfassung von erlebtem Stress. Regehr et al. (2013) fassen zusammen, dass sich Interventionen mit unterschiedlichen Ansätzen (kognitiv oder behavioral oder auf *Mindfulness* abzielend) als wirksam erwiesen haben.

Eine Zusammenfassung der Forschung zur Wirksamkeit von Interventionen, die die Selbstwirksamkeit im Hochschulkontext positiv beeinflussen sollen, findet sich bei van Dinther et al. (2011). Insgesamt bezogen sie 39 Studien aus den Jahren 1993 bis 2010 ein und berichten, dass in 80% der Studien tatsächlich ein Effekt beobachtet wurde. Dabei erwiesen sich insbesondere solche Interventionsprogramme als effektiv, die auf Basis der sozial-kognitiven (Lern-)Theorie entwickelt wurden. Sie listeten die in den Studien beobachteten Effekte detailliert auf und extrahierten, welche konkreten Aspekte in den jeweiligen Studien einerseits theoretisch als Einflussfaktoren genannt und welche andererseits auch tatsächlich empirisch untersucht wurden. Hierbei klassifizierten sie unter anderem nach den vier von Bandura (1977) postulierten Quellen, aus denen sich die Selbstwirksamkeit speist.

Ihre Empfehlung lautet, dass von institutioneller Seite her der Entwicklung der Selbstwirksamkeit unter den Studierenden Beachtung geschenkt werden sollte. Neben verschiedenen Vorschlägen, wie Aufgaben und Lernumgebungen mit Bedacht gestaltet werden sollten, zeigt ihre Metaanalyse auch, dass einzelne Interventionen, die nicht Teil des jeweiligen Curriculums sind, effektiv sein können. Das bedeutet, dass auch studiengangübergreifend beispielsweise von der Studierendenberatung oder dem Studiencoaching angebotene Interventionen positive Effekte auf die Selbstwirksamkeit im Studium haben können, ohne fachspezifische Inhalte zu berücksichtigen.

11.3 Fragestellung

Anknüpfend an die aufgeführten Befunde, sollte diese Pilotstudie die Frage beantworten, ob bereits durch eine relativ kleine Intervention (Workshopdauer: 3 h) auf Gruppenebene die Studieneinstiegsselbstwirksamkeit erhöht werden kann. Darüber sollte die intraindividuellen Veränderung der Studieneinstiegsselbstwirksamkeit bei den Workshopteilnehmern mit einer wiederholten Messung mit Hilfe der SESW-Skala Aufschluss geben.

11.4 Design: Prä-Post-Messung der Studieneinstiegsselbstwirksamkeit

Das *Zentrum für fremdsprachliche und berufsorientierte Kompetenzen (ZfbK)* der Justus-Liebig-Universität bietet Study Skills Kurse zur überfachlichen Weiterqualifikation für Studierende an. In Kooperation mit den beiden dort tätigen Kolleginnen Debora Peine und Dr. Gabi Wotschke vom Studiencoaching der Justus-Liebig-Universität Gießen, erfolgte die Prä-Post-Messung der Studieneinstiegsselbstwirksamkeit (erfasst mit der SESW-Skala) im Rahmen zweier von ihnen angebotenen Workshops.

11.5 Methode

Im Sommersemester 2019 fanden unter anderem die zwei ausgewählten Workshops des Studiencoachings statt; einer unter der Leitung von Debora Peine und einer unter der Leitung von Dr. Gabi Wotschke. Eine Kurzbeschreibung der Workshops (der Homepage des Studiencoachings entnommen: <https://www.uni-giessen.de/fbz/zentren/zfbk/afk/coa>) soll einen Einblick in die Inhalte der dreistündigen Workshops geben.

- Kurzworkshop I: Mehr Zeit – weniger Stress (17.05.2019)
Kursleitung: Debora Peine
Wer kennt das nicht? Die Klausuren stehen an, Hausarbeiten sollen geschrieben werden und eigentlich wollte man mal wieder zum Sport oder Zeit zum Ausspannen haben. Im Kurz-Workshop können Sie Techniken zum Thema Zeitmanagement ausprobieren, die Ihnen dabei helfen Prioritäten zu setzen, Ihre Aufgaben zu bewältigen und auch noch Raum für Freizeit zu haben.
- Kurzworkshop II: Effektiv lernen (28.06.2019)
Kursleitung: Dr. Gabi Wotschke
Leichter und schneller lernen – geht das? Ja! Denn wer weiß, wie das menschliche Gedächtnis funktioniert, kann mit wenig Aufwand sein eigenes Lernen effektiver und effizienter gestalten.

Diese beiden Workshops lassen sich somit einerseits gemäß der Kategorisierung von Hattie et al. (1996) als Kurz-Workshops beschreiben. Zum anderen sind die im Fokus stehenden Themen typische von Hattie et al. (1996) beschriebene Study Skills Inhalte. Zu guter Letzt lässt sich das im Rahmen dieser Pilotstudie betrachtete Kriterium, die Studieneinstiegsselbstwirksamkeit, gemäß Hattie et al. (1996) als Affekt bezogenes Kriterium einordnen. Eine andere Kategorisierung wäre die von Bosse et al. (2019) (Projekt *StuFHe*, vgl. Kapitel 1) vorgeschlagene Einteilung. Hier würden diese beiden Kurz-Workshops in die Kategorie „Vermittlung überfachlicher Kompetenzen und wissenschaftliche Arbeitsweisen“ (S. 41) fallen. Sie sind somit den Workshops zu Lern- und Arbeitstechniken zuzuordnen.

Tabelle 11.1: Demografische Angaben der Teilnehmer

	Geschlecht		Fachsemester			Studiengang		
	m	w	1+2	3+4	>4	BA	MA	StEx
Anzahl	7	20	18	7	3	18	2	8
%	24.1	69.0	62.0	24.1	10.3	62.1	6.9	27.6

11.5.1 Stichprobe

Insgesamt liegen aus beiden Workshops zusammen $N_{total} = 28$ Datensätze vor. Im Hinblick auf diesen kleinen (wenn auch für Workshops erwartungsgemäßen) Stichprobenumfang werden im Folgenden die Daten immer für beide Workshops zusammen dargestellt. 11.1 zeigt die Angaben der Teilnehmer hinsichtlich ihres Geschlechts (eine Person machte hierzu keine Angabe), des aktuellen Fachsemesters und ihres Studiengangs. Die Stichprobe bestand größtenteils aus Frauen (69%) und die meisten Teilnehmer entstammten der Zielgruppe der Studienanfänger (62%), wobei fast zwei Drittel von Ihnen angaben, in einem Bachelor-Studiengang (BA) eingeschrieben zu sein. Die Teilnehmer studierten in verschiedenen Fächern, welche sie in ein Freitextfeld eintrugen. Dabei wurde jeder der folgenden Nennungen maximal von zwei Personen gemacht: Agrarwissenschaften, Anglistik, Bewegung und Gesundheit, Biologie, Chemie, Ernährungswissenschaften, Erziehungswissenschaften, außerschulische Bildung, Soziologie, Geographie, Geschichte, Französisch, Völkerrecht, GuK, Humanmedizin, Jura, L5 Kunst, Förderschulschwerpunkte Lernhilfe und Sprachheilpädagogik, L1 Grundschullehramt (Deutsch, Mathe, Sport), L2 Deutsch, katholische Religion, Lehramt - Englisch und Geschichte, Lehramt (Deutsch, Mathe, Sport), Psychologie, Rechtswissenschaften, SLK, Social Sciences, Umweltmanagement, Wirtschaftswissenschaften. Im Mittel berichteten sie eine Abiturnote von $M = 2.29$ ($SD = 0.60$), welche keinen signifikanten Zusammenhang mit den beiden SESW-Scores aufwies.

11.5.2 Vorgehen

In beiden Workshops wurde zu Beginn ein einblättriger, doppelseitig bedruckter Fragebogen an die Teilnehmer ausgegeben. Sie wurden darauf hingewiesen, dass eine kleine Begleitstudie durchgeführt wird, an der sie auf freiwilliger Basis anonym teilnehmen konnten (siehe Abbildung 11.1). Sie wurden gebeten, die Vorderseite direkt zu Workshopbeginn und die Rückseite zum Abschluss auszufüllen, ohne auf die Vorderseite zurückzublättern.

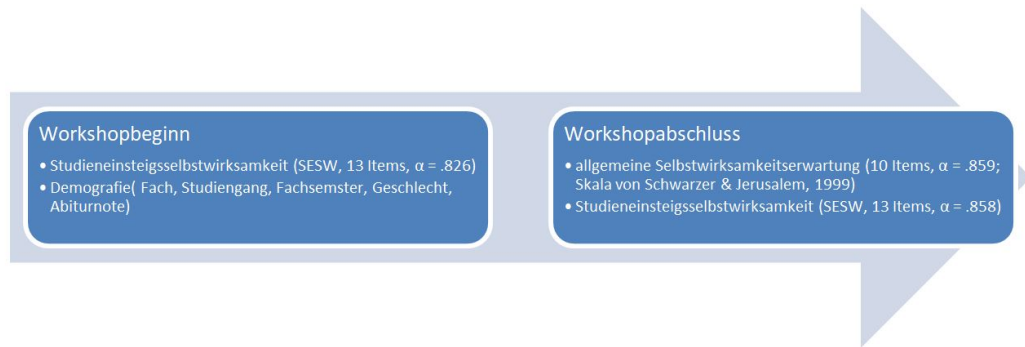


Abbildung 11.1: Ablauf der Datenerhebung

11.6 Ergebnisse

Die Auswertung der Ergebnisse erfolgte sowohl im Sinne intraindividuelle Vergleiche als auch im Vergleich mit den Kennwerten der in Kapitel 3 beschriebenen Normierungsstichprobe.

11.6.1 Auswertung

Die SESW-Skala, die im Fokus dieser Studie stand, ist so aufgebaut, dass die Teilnehmer gebeten werden, für dreizehn Herausforderungen im Studieneinstieg anzugeben, inwiefern sie sich zutrauen, diese zu bewältigen (vgl. Kapitel 3). Im Anhang (13.3 und 13.4) befindet sich eine Übersicht der Antworthäufigkeiten für jedes einzelne Item der SESW-Skala sowohl für die Prä- als auch für die Post-Messung.

Als Antwortformat ist eine fünfstufige Likert-Skala vorgegeben (*gar nicht bis voll und ganz*). Pro Item wird entsprechend ein Score von 1 bis 5 vergeben. Der Skalenwert (SESW-Score) wird gebildet, indem der Mittelwert über alle 13 Items berechnet wird.

Generell ist anzumerken, dass nicht alle Teilnehmer alle Items bearbeitet haben, jedoch in die Analysen der SESW-Scores nur diejenigen einbezogen wurden, die jeweils alle dreizehn Items beantwortet haben. Daraus resultiert die minimal reduzierte Fallzahl für manche der folgenden Angaben.

11.6.2 Prä-Post-Vergleich

Im Mittel erzielten die Teilnehmer zu Beginn des Workshops einen SESW-Score von $M = 3.45$ ($SD = 0.55$) und zum Abschluss des Workshops einen SESW Score von $M = 4.00$ ($SD = 0.48$). Beide Scores sind gemäß Shapiro-Wilk-Test normalverteilt (prä: $W(29) = .94$, $p = .1036$, post:

$W(29) = .97, p = .6431$). Die Abbildungen 11.2 und 11.3 zeigen die Verteilungen in Form von Histogrammen. Im Vergleich zur Normierungsstichprobe der SESW-Skala mit Daten von $N = 1842$ Studieneinsteigern verschiedener Studienfächer ($M = 4.02, SD = 0.55$), lässt sich festhalten, dass diese Stichprobe im Mittel vor den Workshops einen geringeren ($t(27) = -5.527, p < .001$) und nach dem Workshops einen vergleichbaren ($t(25) = -0.213, p = .833$) SESW-Score erzielte.

Abbildung 11.4 zeigt die intraindividuelle Veränderung: Im zweidimensionalen Koordinatensystem ist jeder Teilnehmer als ein Punkt repräsentiert. Dabei ist auf der x-Achse der individuelle SESW-Score vor dem Workshop ("prä") und auf der y-Achse der individuelle SESW-Score nach dem Workshop ("post") abgetragen. Die Workshop-Zugehörigkeit ist farblich kodiert: Durch hellblaue Punkte werden Teilnehmer des Workshops „Mehr Zeit – weniger Stress“ repräsentiert und durch dunkelblaue Punkte werden Teilnehmer des Workshops „Effektiv lernen“ repräsentiert. Die eingezeichnete Winkelhalbierende ($x = y$) stellt die Gerade dar, auf der alle Punkte liegen müssten, wenn bei allen Teilnehmern der SESW-Score nach dem Workshop exakt dem vor dem Workshop entspräche, sprich wenn der Effekt exakt Null wäre. Punkte, die unter der Winkelhalbierenden liegen, stehen für Teilnehmer, bei denen der Score nach dem Workshop unter dem Skalenwert vor dem Workshop liegt. Das ist konkret bei einer Person der Fall gewesen. Alle anderen Punkte liegen über der Winkelhalbierenden, was bedeutet, dass zumindest rein deskriptiv bei allen anderen Teilnehmern die selbst berichtete Studieneinstiegsselbstwirksamkeit nominell nach dem Workshop höher ist als zuvor.

Dieses zunächst deskriptive Ergebnis kann durch einen t -Test untermauert werden: $t(25) = 6.35, p < .001$. Das bedeutet, dass im Mittel eine signifikante Erhöhung der selbstberichteten Studieneinstiegsselbstwirksamkeit im Vergleich von *vor* versus *nach* dem Workshop festgestellt werden kann. Das entspricht einer Effektstärke von Cohen's $d = 1.015$. Diese Effektstärke ist allgemeinen Beurteilungsstandards (J. Cohen, 1988) zur Folge als *groß* einzustufen.

11.7 Diskussion

Mit dieser Pilotstudie sollte exemplarisch überprüft werden, ob im Rahmen zweier Kurz-Workshops eine der im EOS-Modell zentralen Variablen bedeutsam verändert werden kann. Hierfür wurde die Studieneinstiegsselbstwirksamkeit als abhängige Variable für diese Interventionsstudie ausgewählt. Im Vergleich der Prä- mit der Post-Messung (intraindividueller Vergleich) wird ein signifikanter Unterschied deutlich: Im Mittel war der SESW-Score der Teilnehmer nach dem Workshop bedeutsam höher. Das heißt, dass sie nach dem Workshop eine höhere Studieneinstiegsselbstwirksamkeit berichteten. Erst nach dem Workshop lag der mittlere SESW-Score der Teilnehmer auf dem Niveau, welches in der Normierungsstichprobe dem Mittelwert entspricht.

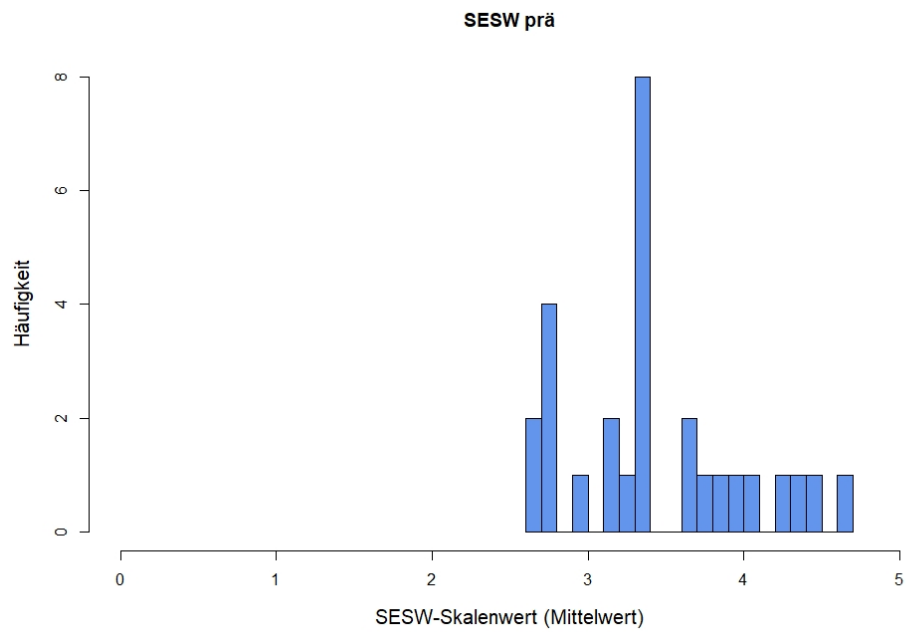


Abbildung 11.2: Verteilung der SESW-Skalenwerts vor dem Workshop

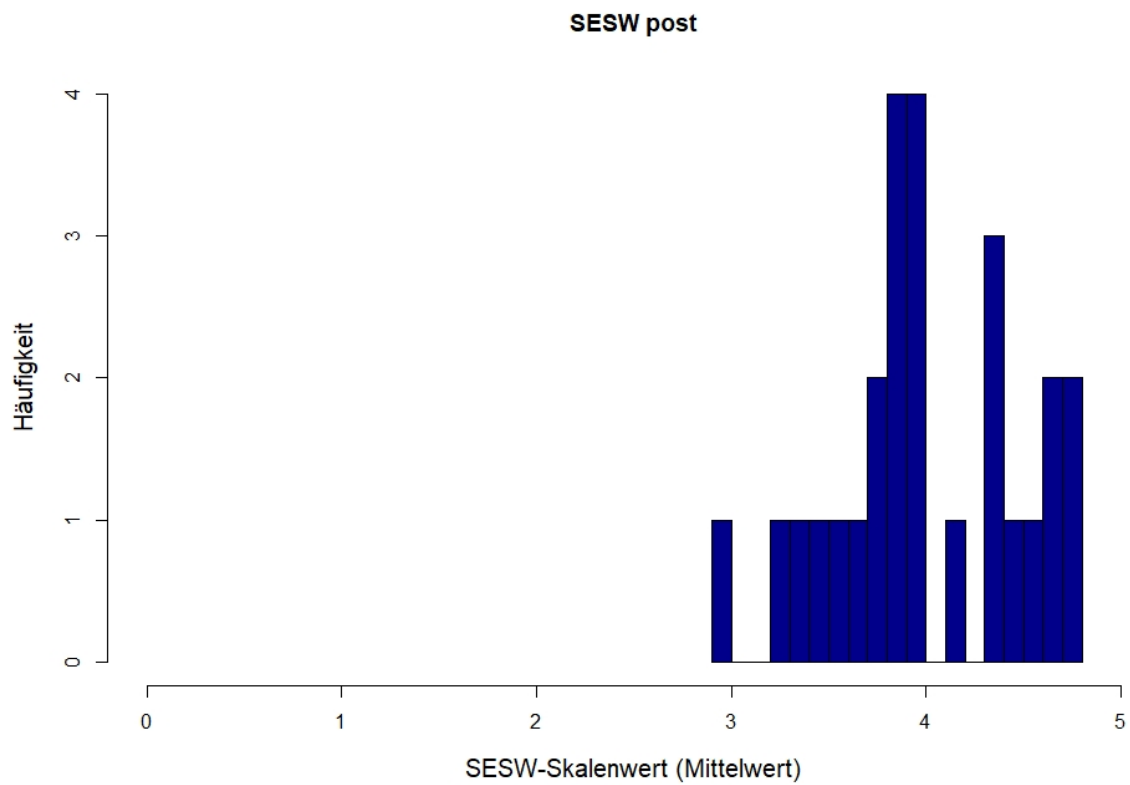


Abbildung 11.3: Verteilung der SESW-Skalenwerts nach dem Workshop

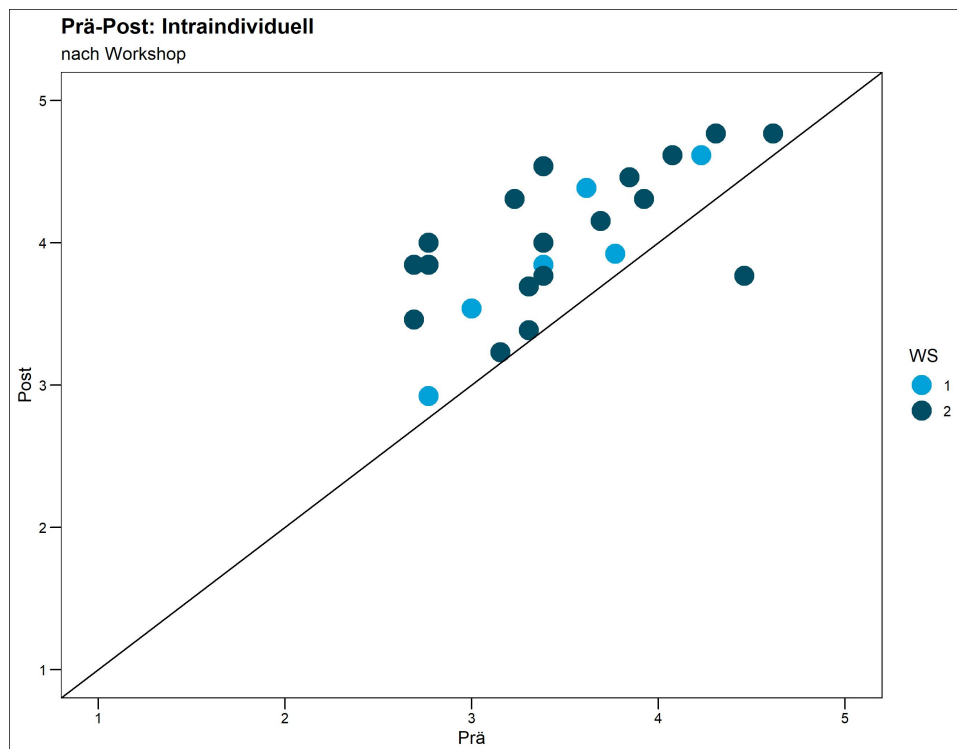


Abbildung 11.4: Intraindividuelle Veränderung des SESW-Scores, $N = 25$

11.7.1 Limitationen

Dabei ist jedoch zur Vorsicht bei der Interpretation der Effektstärke zu raten. Zwar ergab eine mit dem Programm *G*Power* (Moshagen & Erdfelder, 2016) post-hoc durchgeführte Poweranalyse, dass hier mit einer Power von .99 durchaus eine verlässliche Statistik (trotz des kleinen Stichprobenumfangs) vorliegt. Der Post-Test, der hier aus organisatorischen Gründen direkt im Anschluss an den Workshop stattfinden musste, kann aber keinen Aufschluss über die Nachhaltigkeit des Effekts geben und es liegt nahe zu vermuten, dass dieser große Effekt, der weit über denen der Befunde aus vorherigen, umfangreichen Evaluationsstudien zu vergleichbaren Interventionen liegt, den tatsächlichen Effekt stark überschätzt.

Ein Vergleich mit den vorangegangenen Befunde ist auch daher nur sehr begrenzt möglich, da für diese anderen Studien in der Regel größere Prä-Post-Test-Intervalle bzw. Prä-Follow-up-Test-Intervalle (Hattie et al., 1996) gewählt wurden, was sicherlich für die praktische Anwendung auch wesentlich mehr Informationsgehalt bietet.

Weiterhin gab es in dieser Untersuchung auch keine Kontrollgruppe, die den gleichen Fragebogen ausgefüllt, aber nicht an diesen Workshops teilgenommen hat. Dies wäre allerdings wünschenswert, um nicht nur intraindividuelle Veränderungen abbilden zu können, sondern auch einen Gruppenvergleich (interindividuell) zu ermöglichen. Würde man dann einen Gruppenunterschied (bei Konstanzhaltung möglichst vieler potentieller Kovariaten) beobachten, würde dies die Verlässlichkeit der Zuschreibung

des beobachteten Effekts auf die Intervention steigern können.

11.7.2 Praktische Relevanz

Unabhängig davon bleibt jedoch festzuhalten, dass die in dieser Pilotstudie gewonnenen Erkenntnisse darauf hindeuten, dass tatsächlich bereits mit Kurz-Workshops *praktisch* bedeutsame Effekte erzielt werden können. Damit ist gemeint, dass neben der statistischen Signifikanz immer auch zu bedenken ist, ob der beobachtete Effekt für den Anwendungskontext relevant ist. In der Literatur zur quantitativen Bildungsforschung findet sich häufig die Aussage, dass Effekte, die mindestens in der Höhe von $d = 0.33$ liegen, als praktisch bedeutsam gewertet werden können (Gall, Gall & Borg, 2002; Kirk, 1996).

11.7.3 Ausblick

Diese Prä-Post-Messung erlaubt es, einen ersten Eindruck zu bekommen, ob die Studieneinstiegsselbstwirksamkeit bereits durch relativ kleine Interventionen effektiv gesteigert werden kann. Der beobachtete Effekt ist vielversprechend. Praktisch relevant ist jedoch auch die Stabilität dieses Effektes. Darüber kann diese erste Pilotstudie keine Aussage machen. Untersuchungen mit wiederholten Messungen beispielsweise einen Monat später oder in der nächsten Klausurvorbereitungsphase wären wünschenswert, um die Stabilität des gefundenen Effektes zu prüfen. Zudem ist aus methodischer Sicht auch der relativ kleine Stichprobenumfang zu bedenken, der zur Vorsicht bei der Generalisierung der Befunde veranlassen sollte.

In Anknüpfung an die genannten Limitationen sind weitere Studien wünschenswert. Diese sollten zunächst die Replizierbarkeit der Ergebnisse prüfen und idealerweise die statistische Robustheit der Ergebnisse durch größere Stichprobenumfänge (beispielsweise durch Aggregation über mehrere Workshopgruppen) sichern. Darüber hinaus sollte in Bezug auf die Steigerung der Studieneinstiegsselbstwirksamkeit auch exploriert werden, die Bearbeitung welcher Themenbereiche im Workshop einen möglichst großen Effekt erzielt.

Zusammenfassend lässt sich als Forschungsdesiderat festhalten: Ideal wäre ein Prä-Post-Follow-up-Design mit einer Interventions- und einer Kontrollgruppe. So würde die interne Validität der Untersuchung – sprich die Sicherheit der Rückführung des beobachteten Effekts auf die Intervention – gesteigert.

In Erweiterung dieser ersten Untersuchung wäre es zudem sinnvoll, auch andere bekannte Prädiktoren von Studienabbruch in den Fokus zu nehmen. Beispielsweise könnten Interventionen zur Stärkung der Peergroup-Interaktion oder zur (schnellen) Bildung von Peergroups (die die wichtige Ressource *soziale Unterstützung* liefern können) zunächst gezielt konzipiert und dann evaluiert werden. Weiterhin wäre es auch gemäß den Ergebnissen der Modelltests und dem EOS-Modell zur Folge wünschenswert, das *University Citizenship Behavior* der Studienanfänger zu fördern. Das bedeutet, dass auch

Workshops, die auf eine Steigerung des UCB abzielen, konzipiert, durchgeführt und mit ähnlicher Begleitforschung wie in dieser Pilotstudie skizziert, evaluiert werden sollten.

Abschließend lässt sich festhalten, dass mit dieser ersten Pilotstudie ein Vorstoß in Richtung Übertrag der im Rahmen dieser Arbeit gewonnenen Erkenntnisse in die Beratungspraxis an Hochschulen gemacht wurde. Die Ergebnisse sind vielversprechend und veranlassen zu weiterer, umfangreicherer und auf dem *EOS*-Modell basierender Begleitforschung zu institutionellen Interventionen.

12 Diskussion

Zum Abschluss der vorliegenden Arbeit möchte ich die theoretische Auseinandersetzung mit der Vorhersage von *Studienerfolg und Studienabbruch*, die der Startpunkt meines Promotionsprojektes war, nochmals zusammenfassend aufgreifen. Daran schließe ich die Zusammenfassung der Ergebnisse der durchgeführten Untersuchungen an. Dies möchte ich im Hinblick auf die Zielstellung dieser Arbeit reflektieren. Dabei werde ich sowohl auf die Limitationen der durchgeführten Studien eingehen, als auch aufzeigen, in welchem Rahmen die Ergebnisse generalisierbar sind. Ein besonderes Augenmerk soll auf dem neu entwickelten EOS-Modell liegen, das auf Basis der Auseinandersetzung mit Theorien und Befunden zum Thema *Studienerfolg und Studienabbruch* entwickelt wurde.

Abschließend werde ich einen Ausblick geben. Dieser soll einerseits Anknüpfungspunkte für weitere Forschung aufzeigen und andererseits anhand eines konkreten Beispiels den Transfer der Erkenntnisse in die Hochschulpraxis skizzieren.

12.1 Zusammenfassung der Modellvergleiche auf theoretischer Ebene

In Kapitel 2 habe ich vier ausgewählte, etablierte Modelle vorgestellt und miteinander verglichen. Konkret waren dies zwei Integrationsmodelle (von Spady, 1971 und Tinto, 1975), ein Modell mit einer Erwartungs-Wert-Struktur (Modell von Neuville et al., 2007) und ein sozial-kognitives Modell aus dem Bereich der Laufbahnforschung, welche ich auf den Kontext des Studieneinstiegs übertragen habe (Modell nach Lent und Brown, 2013).

In der Gegenüberstellung der beiden Integrationsmodelle zeigte sich deren große Schnittmenge. Bei den liegt die Annahme zugrunde, dass Integration im Studienkontext das Commitment zum Studium und zur Hochschule fördert. Je besser die Integration und das Commitment, desto weniger wahrscheinlich sei ein Studienabbruch. Im Modell von Tinto (1975) wird dabei noch zwischen akademischer und sozialer Integration unterschieden. Insgesamt zählen diese beiden Modelle zu den meist rezipierten im Bereich der Studienabbruchsforschung. Zumal beide Modelle vor über 40 Jahren im Kontext des US-Bildungssystems entwickelt wurden, stellte sich die dringliche Frage, inwiefern sie überhaupt (heute noch) den Prozess, der ggf. zum Studienabbruch führt, innerhalb der Population der Studierenden in Deutschland abbilden können.

Das deutlich neuere Modell von Neuville et al. (2007), welches im belgischen Hochschulkontext entwickelt und erprobt wurde, bezieht sich explizit auf das in der Psychologie beheimatete Erwartungs-Wert-Paradigma. Im Zentrum der Vorhersage von Studienabbruch stehen daher Erwartungen der

Studierenden und der Wert, den diese ihrem Studium beimessen. Neuville et al. (2007) hatten bereits selbst gezeigt, dass ihr Modell in einer empirischen Testung unter belgischen Studierenden dem von Tinto (1975) überlegen war.

Das vierte Modell, das ich im Rahmen dieser Arbeit genauer betrachtet habe, ist das CSM-Modell (Lent & Brown, 2013). Dieses Modell, welches der theoretischen Strömung der sozial-kognitiven (Lern-)Theorie (Bandura, 1977) zuzuordnen ist, ist sehr allgemein gehalten formuliert. Es ist so konzipiert, dass man es auf verschiedene Anwendungskontexte übertragen kann, in dem man übergeordnete Variablen wie *Selbstwirksamkeit* oder *Ergebniserwartungen* und *Handlungen* mit konkret zum Kontext passenden Variablen und zugehörigen Skalen operationalisiert. Insbesondere auch die Berücksichtigung von *Selbstwirksamkeit* und *Zielsetzung* zeichnen dieses Modell aus und grenzen es damit auch vom zuvor genannten Modell ab.

Die theoretische Auseinandersetzung zeigte somit klar die Gemeinsamkeiten und Unterschiede der Modelle auf. Im Rahmen meines Promotionsprojektes wollte ich nun folgende weiterführende Fragen beantworten:

1. Welches der vier hier fokussierten etablierten Modelle zum Studienerfolg versus Studienabbruch kann empirische Daten, erhoben in der Population der Studierenden in Deutschland, am besten erklären?
2. Wie könnte ein neues integratives Modell aussehen, das die Stärken der etablierten Modelle kombiniert und zusätzlich neuste metaanalytische Erkenntnisse berücksichtigt?
3. Welche sind die Konstrukte, die gemäß einem solchen neuen Modell eine Schlüsselfunktion im Prozess des Studieneinstiegs haben und daher in der Hochschulpraxis beispielsweise in Form eines Screenings besonders beachtet werden sollten?

Zur Beantwortung der ersten Frage führte ich drei Längsschnittstudien durch, deren Ergebnisse ich im folgenden Abschnitt zusammenfassen möchte. Dem kommt auch deshalb eine besondere Bedeutung zu, da vielmals auf theoretischer Ebene Modelle, die beispielsweise im US-amerikanischen Raum aufgestellt (und zumindest teilweise empirisch erprobt) wurden, auf den deutschen Hochschulkontext übertragen werden. Allerdings wurde bislang aber in der Regel noch nicht empirisch aufgezeigt, dass diese Modelle ohne Weiteres auf ein anderes Hochschulsystem übertragen werden können, was man nicht als selbstverständlich voraussetzen kann (Hüther & Krücken, 2018).

12.2 Zusammenfassung der empirischen Befunde

Nach der Zusammenfassung des theoretischen Modellvergleichs möchte ich im Folgenden die im Rahmen dieses Promotionsprojektes durchgeführten empirischen Studien reflektieren.

12.2.1 Längsschnittstudien

Der empirische Vergleich der vier ausgewählten Modelle erfolgte anhand von drei Längsschnittstudien (Kapitel 5, 7, 8). Hierbei wurde aus drei aufeinanderfolgenden Kohorten von Studienanfängern in Deutschland jeweils eine große Stichprobe gezogen. Diese Studienanfänger wurden dann über die Dauer der Studieneingangsphase (9 Monate) an insgesamt drei Messzeitpunkten wiederholt befragt. Der Fokus lag deshalb auf der Studieneingangsphase, weil bekannt ist, dass hier der Peak in der Studienabbruchquote zeitlich lokalisiert ist (Stifterverband für die Deutsche Wissenschaft, 2015; Heublein et al., 2017).

Die Ergebnisse der empirischen Modelltests aus Studie 1 (Studienanfänger im Wintersemester 2016/2017) zeigte deutlich, dass das Modell von Neuville et al. (2007) den anderen überlegen war das den Fit zu den erhobenen Daten betrifft. Zudem zeigte sich, dass die beiden Modelle in denen die SESW berücksichtigt wird (von Neuville et al., 2007 und Lent und Brown, 2013), den anderen beiden im Modellfit überlegen waren. Insbesondere wurden vornehmlich die Pfade in den besagten Modellen signifikant, die mit der SESW verknüpft waren.

Aufgrund dieser Erkenntnisse aus Studie 1 und in Anbetracht der Tatsache, dass in den vergangenen zwei Dekaden mehrere Metaanalysen zur Vorhersage von *Studienerfolg* und *Studienabbruch* publiziert wurden (siehe 2.2), eröffnet sich die Möglichkeit, ein neues Modell aufzusetzen, welches die Erkenntnisse dieser beiden Quellen berücksichtigt.

Daher entwickelte ich nach der ersten Längsschnittstudie ein neues Modell: das *EOS*-Modell (Kapitel 6). Anhand der Daten aus der ersten Studie wies dieses Modell einen guten Fit auf und lag in der „Rangreihe“ nur knapp hinter dem Modell von Neuville et al. (2007).

In der zweiten und dritten Längsschnittstudie teste ich dann die vier etablierten Modelle und das neue *EOS*-Modell. Studie 2 und 3 sind exakte Replikationen von Studie 1. Dadurch sind die Beobachtungen, die ich wiederholt machte, als besonders verlässlich einzustufen.

Zusammenfassend zeigte sich das oben beschriebene Muster in allen drei Studien: Modelle, in denen die SESW berücksichtigt wird, konnten die Daten der Studienanfänger aus allen drei Jahrgängen besonders gut erklären. Im Detail zeigte sich im Vergleich des Modells nach Lent und Brown (2013) mit dem von Neuville et al. (2007) und dem *EOS*-Modell, dass insbesondere die Berücksichtigung von wiederholten Messungen der SESW den Modellfit positiv beeinflussen. So waren die beiden letztgenannten Modelle stets den anderen Modellen im absoluten Modellfit und im Vergleich der Modellfits anhand des AICs überlegen.

Wie im Kapitel 9 aufgezeigt, war das AIC für das *EOS*-Modell und das von Neuville et al. (2007) stets nur marginal verschieden. Zweimal war das AIC des Modells von Neuville et al. (2007) etwas kleiner (Studie 1 und 2), einmal das AIC des *EOS*-Modells. Eine inferenzstatistische Absicherung der Fit-Differenz zeigte jedoch, dass diese nur in Studie 3, in welcher das *EOS*-Modell überlegen war, statistisch bedeutsam war.

In der Gruppe der etablierten Modelle erwies sich das Modell von Neuville et al. (2007) folglich als das beste. Nun stellte sich die Frage, warum dann das *EOS*-Modell überhaupt gebraucht wird.

Wie in Kapitel 6 ausführlich dargelegt, birgt das *EOS*-Modell neben einem vergleichbar guten Fit wie das Modell von Neuville et al. (2007) einen entscheidenden Vorteil: Neben der Tatsache, dass die SESW auch im *EOS*-Modell eine wichtige Rolle spielt, sind auch weitere motivationale, emotionale und soziale Aspekte integriert. Im Modell von Neuville et al. (2007) liegt der Fokus recht eng auf der Erwartungs-Wert-Struktur. Das *EOS*-Modell ist deutlich breiter aufgestellt: Es trägt den Befunden zur *Zielsetzung* als Prädiktor von Studienerfolg und Abbruchintentionen Rechnung, greift den Aspekt der sozialen Unterstützung auf, beinhaltet sogenannten „Zwischen-Outcomes“ (Stress, Zufriedenheit) und berücksichtigt als einziges mir bekanntes Modell explizit das Konzept des UCB, welches den in Kapitel 2 aufgeführten vorangegangenen empirischen Befunden zur Folge ein sehr guter Prädiktor für Studienerfolg und Abbruchintentionen ist.

Zusammenfassend lässt sich folglich festhalten, dass das *EOS*-Modell auf einer breiteren theoretischen Basis steht und damit auch mehr Einblick gewährt in den Prozess, welcher zu Studienerfolg oder einem Studienabbruch führen kann. Somit lassen sich anhand der *EOS*-Modells mehr Ansatzpunkte für institutionelle Interventionen aufzeigen, wie ich in Abschnitt 12.7 weiter ausführen werde.

12.2.2 Sekundärdatenanalyse

Um das *EOS*-Modell zusätzlich anhand einer repräsentativen Stichprobe erneut zu validieren, führte ich in Ergänzung zu den drei Längsschnittstudien eine Validierung mit externen Daten in Form einer Sekundärdatenanalyse mit einem Teildatensatz der 21. Sozialerhebung (Middendorff et al., 2016) durch. Mit diesem sehr umfangreichen Datensatz war es mir möglich zu untersuchen, ob das *EOS*-Modell in verschiedenen Fachclustern einen vergleichbar guten Fit aufweist. Da dieser Datensatz aber einerseits querschnittlicher Natur war und andererseits nicht alle im Modell spezifizierten Variablen enthält, entschied ich, nur einen Teil des *EOS*-Modells zu testen. Konkret ging ich so vor, dass ich zwei multiple Regressionen (eine für das Kriterium *Studienerfolg* und eine für das Kriterium *Abbruchintention*) berechnete. In beiden wurden bis auf einen Regressor alle signifikant und es ergaben sich beachtliche Anteile an aufgeklärter Varianz. Dadurch, dass die Größe der „Zellen“ ausreichend groß war, um Fachcluster-spezifische Regressionen zu berechnen, konnte ich anschließend auch noch untersuchen, ob der Anteil aufgeklärter Varianz in den verschiedenen Fachclustern vergleichbar groß ist. Das war für beide Kriterien der Fall, weshalb das *EOS*-Modell fachübergreifend als geeignet zur Varianzaufklärung in der *Abbruchintention* und dem *Studienerfolg* bezeichnet werden kann.

12.2.3 Interventionsstudie

Weiterhin sollten Möglichkeiten des Transfers der im Rahmen dieses Promotionsprojektes gewonnenen Erkenntnisse auf die Hochschulpraxis dargestellt werden. Dazu führte ich in Kooperation mit dem

Studiencoaching der Justus-Liebig-Universität Gießen eine Pilotstudie durch. Im Rahmen dieser Studie wurde exemplarisch für die im *EOS*-Modell als relevant herausgestellten Variablen die *SESW* mit einem Prä-Post-Vergleich im Kontext zweier Workshops für Studieneinsteiger erfasst. Trotz der Limitationen dieser Pilotstudie (kleine Stichprobe, keine Follow-up-Messung und keine Kontrollgruppe) birgt sie doch einen wichtigen Erkenntnisgewinn: Bereits mit einer nur dreistündigen institutionellen Intervention auf Gruppenebene konnte die *SESW* im intraindividuellen Vergleich signifikant gesteigert werden. Dabei lag die *SESW* der Workshopteilnehmer vor dem Workshop signifikant unterhalb dem Mittelwert in der Normierungstichprobe und nach dem Workshop dann auf einem vergleichbaren (statistisch nicht verschiedenen) Niveau.

Somit konnte in dieser Pilotstudie aufgezeigt werden, wie das *EOS*-Modell in der Hochschulpraxis Anwendung finden kann: Die im Modell spezifizierten Variablen (hier wurde exemplarisch die *SESW* untersucht) sollten als Ansatzpunkt für Interventionen verstanden werden. Damit ist gemeint, dass institutionelle Interventionen darauf ausgerichtet werden sollten, diese Variablen, die gemäß den hier berichteten empirischen Befunden bedeutsame Prädiktoren von *Abbruchintention* und *Studienerfolg* sind, positiv zu beeinflussen.

Das *EOS*-Modell kann insofern handlungsanleitend sein, als dass es aufzeigt, in welchem Zusammenhang die aufgeführten Variablen im Prozess des Studieneinstiegs zusammenwirken. Man kann anhand der Anordnung der Variablen innerhalb des Prozesses beispielsweise ableiten, dass der Stress, den Studienanfänger etwa nach dem ersten Semester berichten, bedeutsam reduziert werden kann, in dem frühzeitig, zu Beginn des ersten Semesters die *SESW* der Studienanfänger gestärkt wird und sie darin gestärkt werden, soziale Unterstützung wahrzunehmen.

Zusammenfassend lässt sich festhalten, dass mit den empirischen Befunden auch die oben genannte zweite und dritte Forschungsfrage beantwortet werden konnten.

12.2.4 Limitationen

Als Limitationen der durchgeführten empirischen Untersuchungen sind in erster Linie vier Punkte zu nennen. (i) Zum einen sind die Stichproben in den Längsschnittstudien nicht repräsentativ für die Zielpopulation, insbesondere im Hinblick auf die Verteilung auf die Fachcluster, da vor allem keine Studierenden der Ingenieurwissenschaften miteinbezogen wurden. Wie in Kapitel 5 beschrieben, war dies keineswegs intendiert, sondern ein Planungsfehler bei der Rekrutierung. Zumal ich in allen drei Studien die gleichen Rekrutierungskanäle nutzte und ich das Fehlen dieser Studierendengruppe erst nach Abschluss der letzten Datenerhebung realisierte, gilt diese Limitation für alle drei Studien. Die in Studie 1 bis 3 realisierten Stichproben können hinsichtlich der Verteilung auf die Fachcluster somit zwar nicht als repräsentativ für die Zielpopulation bezeichnet werden, jedoch weisen alle drei eine große Fächerheterogenität auf, sodass ich sie in dieser Hinsicht als „eingeschränkt“ repräsentativ bezeichnen möchte, da sie doch ein wesentlich breiteten Einblick bieten als Studien, in denen ausschließlich Studierende eines Faches untersucht wurden.

Für zukünftige Studien kann ich an dieser Stelle empfehlen, bei der Rekrutierung bereits großen Wert darauf legen, möglich eine repräsentative Stichprobe zu ziehen. Hierfür empfiehlt es sich, eine gut ausgearbeitete Sampling-Prozedur anzuwenden, wie sie in der Sozialforschung wesentlich verbreiteter sind als in der Psychologie (De Leeuw et al., 2012).

Ein weiterer Punkt ist, dass (ii) nur Studienanfänger an Universitäten befragt wurden, um den Typus der Hochschule innerhalb der Stichproben konstant zu halten. Die Übertragbarkeit der Ergebnisse und besonders des *EOS*-Modells auf den Kontext der Fachhochschulen ist demnach noch nicht untersucht.

(iii) Weiter liegt in allen drei Längsschnittstudien ein Dropout über die Gesamtlaufzeit von circa 50% vor. Zwar ist dies nicht unüblich (vgl. z.B. Ehlert et al., 2017), jedoch muss man bei der Ergebnisinterpretation stets bedenken, dass die Analysestichproben (die nur die Personen umfassten, die auch am dritten Messzeitpunkt teilgenommen und zu den beiden Kriterien Angaben gemacht haben) somit vermutlich einem gewissen Grad an Selektivität unterliegen. Wie bereits von Tieben (2016) beschrieben ist insbesondere bei der Untersuchung von *Studienabbruch* anzunehmen, dass solche Personen, die sich mit Abbruchgedanken tragen, gegebenenfalls wenig geneigt sind, wiederholt an Befragungen zum Thema *Studium* teilzunehmen. Weiterhin ist aus Untersuchungen zur freiwilligen Teilnahme an Studien (wie es hier der Fall war) bekannt, dass tendenziell Frauen eher als Männer geneigt sind, zu partizipieren und hoch gewissenhafte Personen ebenfalls eher mitmachen. Der erstgenannte Punkt scheint sich auch in Studie 1 bis 3 widerzuspiegeln, da deutlich mehr Frauen teilnahmen als Männer. Somit ist bei der Ergebnisinterpretation auch der Aspekt der Selektivität zu bedenken.

Zuletzt sollte nicht unerwähnt bleiben, dass (iv) im Rahmen der vorliegenden Arbeit „nur“ *Abbruchintentionen* und nicht tatsächlicher Studienabbruch erhoben wurde. Sicherlich ist die Prädiktion der Intention zum Studienabbruch ebenso wertvoll, jedoch ist eine Intention zum Studienabbruch keinesfalls mit einem tatsächlichen Abbruch gleichzusetzen, auch wenn von einem starken Zusammenhang ausgegangen wird und daher *Abbruchintentionen* oftmals anstelle des tatsächlichen Abbruchs im Fokus der Forschung liegen (vgl. Kapitel 1). Dies ist oft auch der Tatsache geschuldet, dass diese einfacher zu erfassen sind (Neugebauer et al., 2019).

12.3 Weiterführende Analysen

Anknüpfend an die Modellentwicklung und empirische Testung, habe ich weitere Analysen zum *EOS*-Modell durchgeführt. So konnte ich zeigen, dass es metrische Messinvarianz über die drei untersuchten Jahrgänge aufweist (vgl. Kapitel 9.3.1). Dies belegt zusätzlich zum Ergebnis der Sekundärdatenanalyse die Generalisierbarkeit des Befundes, dass das *EOS*-Modell gut geeignet ist, um den Prozess des Studieneinstiegs in der Population der Studienanfänger in Deutschland abzubilden. Zusammenfassend liegen folglich verschiedene empirische Befunde vor, auf deren Basis die Generalisierbarkeit der Erkenntnisse zum *EOS*-Modell gerechtfertigt ist.

Berücksichtigt man die Messinvarianz des *EOS*-Modells, so ist eine Aggregation der Daten aus allen

drei Jahrgängen legitim. Die Modellfitindizes für das EOS-Modell berechnet mit diesem aggregierten Datensatz des Umfangs $N = 424$ wiesen erneut den guten Fit des Modells aus. Mit diesem großen (aggregierten) Stichprobenumfang sind die Standardfehler der Schätzung der Pfadkoeffizienten geringer als in den Einzelstudien, sprich das Ergebnis noch belastbarer.

12.4 Skalenentwicklung

Im Vorfeld zur Erhebung der Längsschnittdaten, galt es, die Operationalisierung der im Fokus stehenden Konstrukte festzulegen. In Ermangelung einer kontextspezifischen Skala zur Erfassung der Selbstwirksamkeit im deutschen Hochschulkontext, entwickelte ich die SESW-Skala (vgl. Kapitel 3). Zu dieser berichtete ich umfangreiche Analysen der Item- sowie Skalengütekriterien. Die SESW-Skala weist eine hohe psychometrische Qualität auf. Zudem liegen Normdaten von $N = 1842$ Studienanfängern aus dem Jahr 2018 vor. Entwickelt als Screening-Instrument, also zur Identifikation von Personen mit vergleichsweise niedriger SESW, kann die Skala im Zuge der Förderdiagnostik Anwendung finden: Studienanfänger, deren SESW-Skalenwert bedeutsam niedriger ist als der Mittelwert der Normgruppe, können gezielt auf zu ihren Bedürfnissen zugeschnittene institutionelle Angebote verwiesen werden. Wie das praktisch umsetzbar wäre, beschreibe ich im folgenden Abschnitt, indem ich einen Ausblick auf die praktische Anwendung der Erkenntnisse aus diesem Promotionsprojekt gebe.

Zusätzlich konnte ich in einem Gruppenvergleich zwischen Studienanfängern, die eine vergleichsweise stark ausgeprägte Abbruchintention berichten *versus* solchen, die eine eher niedrig ausgeprägt Abbruchintention berichten, zeigen, dass sie sich nicht nur auf manifester Ebene signifikant in der Höhe des SESW-Skalenwerts unterscheiden und dieser Unterschied über die drei Messzeitpunkte größer wird, sondern dass es auch einen Gruppenunterschied auf latenter Ebene (latenter Faktor SESW) gibt. Man spricht hier von einer *Latent Means Difference*.

12.5 Fazit

Drei Punkte möchte ich als Fazit zu den Erkenntnissen, die im Rahmen der vorliegenden Arbeit gewonnen werden konnten, aufführen.

1. Wie intendiert liefert diese Arbeit Erkenntnisse zur Nützlichkeit vier etablierter Modell zur Erklärung von *Studienerfolg und Studienabbruch* in der Population der Studienanfänger in Deutschland: Die beiden auf Integration fokussierten Modelle aus den 1970er Jahren (von Spady, 1971 und Tinto, 1975) passten nicht gut zu den Daten der Stichproben aus drei Jahrgängen von Studieneinsteigern in Deutschland. Das sozial-kognitive Modell nach Lent und

Brown (2013) wies teilweise einen befriedigenden Fit auf. Wirklich gut bis sehr gut passte das Modell von Neuville et al. (2007) mit seiner Erwartungs-Wert-Struktur zu den empirischen Daten.

2. Das neu entwickelte *EOS*-Modell wies in zwei der drei Längsschnittstudien einen ebenso guten Modellfit auf wie das Modell von Neuville et al. (2007), in der dritten Studie war es diesem sogar signifikant überlegen, was den Modellfit betrifft. Das *EOS*-Modell zeichnet sich insbesondere dadurch aus, dass es relevante Aspekte der vier etablierten Modelle berücksichtigt und zudem mit seiner breiten Aufstellung auch alle individuumsbezogenen Variablen, welche gemäß aktuellen Metaanalysen gute Prädiktoren für *Studienerfolg* und *Studienabbruch* sind, umfasst. Es ist daher den etablierten Modellen insofern vorzuziehen, als dass es den Prozess umfassender beleuchtet, da verschiedene theoretische Perspektiven integriert sind.
3. Mit den insgesamt fünf durchgeführten Studien liefert diese Arbeit eine umfassende empirische Basis einerseits zur Frage der Modellvergleiche und andererseits zum *EOS*-Modell, wodurch sich relativ verlässliche Ergebnisse festhalten lassen und konkrete Anknüpfungspunkte für weitere Forschung und die Hochschulpraxis deutlich werden. Meiner Kenntnis nach liegt zum Zeitpunkt der Fertigstellung dieser Arbeit kein anderes Modell zur Darstellung der Prozesse im Studieneinstieg, die auf Individuumsebene zu *Studienerfolg* oder *Studienabbruch* führen, vor. Auch ist die Fülle an empirischen Befunden ein Alleinstellungsmerkmal dieser Arbeit.

12.6 Forschungsdesiderate

An dieser Stelle möchte ich die Punkte aufzeigen, an die zukünftige Untersuchungen anknüpfen können.

In Anlehnung an die oben aufgeführten Limitationen sollte in zukünftigen Studien darauf geachtet werden, dass eine insbesondere im Hinblick auf die Fachcluster repräsentative Stichprobe gezogen wird. Zudem könnte eine stärkere Incentivierung (z.B. Gewinnspiel mit kleineren und größeren Gewinnen, aber einem garantierten Gewinn für jeden) oder möglicherweise eine Begleitstudie zur Evaluation eines grundständigen Studiengangs, an der verpflichtend, aber in pseudonymisierter Weise, teilgenommen werden muss, das Problem mit der Selektivität der Stichproben verringern oder umgehen.

Im Rahmen einer solchen Studie, die in Kooperation mit einem oder idealerweise mehreren Fachbereichen durchgeführt würde, könnten dann möglicherweise auch die Angaben in der Befragung mit administrativen Daten verknüpft werden. So könnte eine „objektive“ Erfassung der akademischen Leistung und auch bspw. der Studiengeschwindigkeit (operationalisiert als Anzahl absolvierter Creditpoints pro Semester) erfolgen, genauso wie erfasst würde, bei welchen Personen gegebenenfalls tatsächlich bis zum Ende der Datenerhebung eine Exmatrikulation vorliegt. Dies würde wertvolle Einblicke liefern, die die aktuelle Diskussion um die „wahre“ Höhe der Abbruchquote (vgl. z.B. Tieben, 2019) mit einer soliden empirischen Basis bereichern könnten. Zudem würde dadurch die oben

aufgeführte Limitation, dass in dem vorliegenden Promotionsprojekt „nur“ *Abbruchintentionen* und nicht tatsächlicher Abbruch erfasst wurde, adressiert werden.

Selbstverständlich kann eine solchen Studie nur durchgeführt werden, sofern das datenschutzkonform realisierbar ist.

Eine erneute Validierung des *EOS*-Modells mit einer Kombination aus Selbstberichtsdaten und administrativen Daten wäre ideal, da die sinnvolle Kombination verschiedener Datenquellen die Robustheit der Ergebnisse steigern kann. Dies könnte gegebenenfalls mit der im Folgenden beschriebenen praktischen Implikation verknüpft werden (siehe Abschnitt 12.7).

Auch der Einbezug von Merkmalen der Hochschule wäre in weiteren Studien wünschenswert. So ist bislang noch ungeklärt, ob das *EOS*-Modell nur im universitären Kontext gut geeignet ist, oder gleichermaßen auch an Fachhochschulen den Prozess des Studieneinstiegs gut abbilden kann. Auch beispielsweise der Einfluss der Größe der Hochschule oder des Fachbereichs und des „Betreuungs-schlüssel“ (Anzahl Studierende pro Lehrendem) könnte die Prognosegüte des Modells weiter steigern. Beispielsweise Thomas (2002) untersuchte in England, welche institutionellen Charakteristiken mit Studienabbruchraten assoziiert sind. Sie nahm dabei insbesondere den institutionellen Habitus in den Fokus und stütze ihre Überlegungen auf Theorien des sozialen bzw. Bildungskapitals sensu Bourdieu und Passeron (1977). Einige weitere Autoren betonen ebenfalls, dass Eigenschaften der Hochschule und auch des konkreten Lernsettings bei der Vorhersage von Studienerfolg und Studienabbruch berücksichtigt werden sollten (Georg, 2009; Oseguera & Rhee, 2009; Schaeper, 2019). Dies könnte mit dem *EOS*-Modell verknüpft werden, indem es zu einem *Multi-Level*-Modell erweitert wird. Auf der oberen Ebene wären dann Merkmale der Hochschule angesiedelt und auf der unteren Ebene die individuumszentrierte Perspektive, die im Rahmen dieser Arbeit im Vordergrund stand.

12.7 Praktische Implikationen

Wie kann das *EOS*-Modell in der Hochschulpraxis gezielt genutzt werden? Diese Frage möchte ich abschließend im Sinne der praktischen Implikationen beantworten. Ein möglicher Anwendungsbereich ist das Studiencoaching (Angebot von Beratungs- und Unterstützungsmöglichkeiten, Kursen etc.). Wie in Kapitel 11 dargestellt, kann die Wirksamkeit von Interventionen daran gemessen werden, inwiefern sie die im *EOS*-Modell als bedeutsam aufgezeigten Variablen positiv beeinflussen können. In diesem Sinne ist die dargestellte Pilotstudie auch als Evaluation der Workshops zu verstehen.

Einen Schritt weiter vorne ansetzen kann man, in dem man gezielt neue Interventionen, Kurse etc. entwickelt oder bereits vorhandene kritisch prüft und gegebenenfalls anpasst, sodass mit ihnen möglichst zielgerichtet die im *EOS*-Modell spezifizierten Variablen in den Fokus genommen werden können. Somit kann das *EOS*-Modell als „Landkarte“ dienen und helfen, die limitierten institutionellen Ressourcen zur Unterstützung der Studienanfänger auf die vielversprechenden (weil hochgradig für Abbruchintention und Studienerfolg prädiktiven) Variablen auszurichten.

Eine andere Möglichkeit der Anwendung wäre ein „flächendeckendes“ Screening von Studienanfängern. Beispielsweise könnte dies im Rahmen eines Online Self-Assessments (OSAs) umgesetzt werden. Eine Vielzahl an deutsche Universitäten bietet solche Test- und Informationsportale für die Zielgruppe der Studieninteressierten an. Ein äquivalentes Angebot für Studieneinsteiger könnte als Instrument der Förderdiagnostik im Studieneinstieg dienen.

Im OSA könnten verschiedene Skalen zu den im EOS-Modell aufgeführten Variablen zu den entsprechenden Zeitpunkten im Studieneinstieg dargeboten werden. Die Studienanfänger könnten anonymisiert oder pseudonymisiert teilnehmen und würden individuelles Feedback zu den Angaben, die sie gemacht haben, erhalten. Beispielsweise anhand einer normorientierten Rückmeldung könnten sie gezielt zu ihren Angaben individuell passend eine Liste von Interventionen und Beratungsangeboten angezeigt bekommen, wie bereits von Robbins et al. (2004) vorgeschlagen.

So könnte ein Feedback-Text beispielsweise zur SESW-Skala lauten:

„Im Vergleich zur Normstichprobe von 1842 Studienanfängern verschiedenster Fächer liegt die Ausprägung Ihrer Selbstwirksamkeit, welche aus Ihren Angaben berechnet wurde, im unterdurchschnittlichen Bereich. Selbstwirksamkeit bezeichnet....“

Ein solches Feedback könnte zu jeder Skala, die ausgefüllt werden kann, dargeboten werden. Letztendlich könnte abschließend eine Liste an passenden Angeboten angezeigt werden:

„Im Folgenden finden Sie passend zu Ihren Angaben, die Sie im OSA gemacht haben, eine Liste an Angeboten unserer Hochschule, von denen das eine oder andere möglicherweise für Sie interessant sein könnte...“

Mit einem solchen Tool könnten Studienanfänger niederschwellig (pseudonymisiert oder anonymisiert online, man muss sich nicht „überwinden“ und persönlich die Beratung aufsuchen) auf freiwilliger Basis gescreent werden und gezielt auf Angebote hingewiesen werden, von deren Existenz sie andernfalls möglicherweise gar nicht gewusst hätten. Die Studienanfänger könnten das erhaltene Feedback nutzen, um ihre Angaben im OSA zu reflektieren und sofern sie das möchten, anschließend an entsprechenden Workshops teilzunehmen oder bei Bedarf ein persönliches Beratungsgespräch zu suchen. Ein ähnlicher Ansatz findet sich beispielsweise bei den beiden großen Projekten *Prev Drop - Detecting and Preventing Drop out from Higher Education or Supporting Students to Switch Successfully to VET* (<http://www.prevdrop.eu/de>) und *SUnStAR - Supporting UNiversity STudents At Risk of dropping out* (<https://sunstar-eu.org/>).

Ein solches OSA könnte mit einem „Ampelsystem“ verknüpft werden, bei dem ein Koordinatensystem mit den im EOS-Modell spezifizierten Prädiktoren als Achsen erstellt wird. Anhand der im OSA gemachten Angaben wird jede Person dann grafisch in dem mehrdimensionalen Raum verortet.

Dazu sollten für jede im OSA verwendete Skala Normwerte vorliegen und entsprechende Cut-off-Werte, bei deren Unter- bzw. Überschreitung eine Person als potentiell „überdurchschnittlich abbruchgefährdet“ eingestuft werden kann. Liegt eine Person dann beispielsweise hinsichtlich mindestens zwei oder drei (das wäre noch weiter zu elaborieren) dieser Variablen im „gefährdeten“ Bereich,

so wird sie im Ampelsystem rot dargestellt und bekommt zusätzlich zum individuellen Feedback noch gezielt den Hinweis, dass sie bei Bedarf auch die Möglichkeit hat, das erhaltene Feedback in einem persönlichen Gespräch im Rahmen des Studiencoachings bzw. der Studienberatung zu reflektieren. Allen anderen stünde der Weg zur persönlichen Beratung selbstverständlich auch offen, nur diese als „besonders abbruchgefährdet“ identifizierten Studienanfänger würden einen expliziten Hinweis auf die Möglichkeit der Nachbesprechung der OSA-Ergebnisse erhalten, um den Handlungsbedarf, der sich aus ihren Angaben ableiten lässt, zu verdeutlichen. Diese Idee ist selbstverständlich ebenfalls als Förderdiagnostik zu verstehen: Studienanfänger bekommen gezieltes Feedback und Hinweise zu individuell passenden Angeboten, aber es wird niemand zur Teilnahme an Workshops oder Beratungsgesprächen verpflichtet.

Anhand der Abbildung 12.8 möchte ich exemplarisch aufzeigen, wie eine solche Verortung von Personen anhand ihrer Angaben grafisch dargestellt werden könnte. Dabei habe ich hier exemplarisch aus dem EOS-Modell drei Variablen ausgewählt (SESW, Studienzufriedenheit und UCB), um anhand einer dreidimensionalen Darstellung diese Idee zu visualisieren. Zu diesem Zweck habe ich die drei Variablen z-standardisiert, sodass beispielsweise als Cut-off-Wert stets „-2 und kleiner“ (also zwei Standardabweichungen unter dem Durchschnitt) infrage kommen könnte.

Ein solches Ampelsystem könnte idealerweise mit den in Kapitel 1.1.4 erwähnten Frühwarnsystemen (vgl. Berens und Schneider, 2019; K. Schneider, Berens und Burghoff, 2019), die ausschließlich auf Basis administrativer Daten arbeiten, kombiniert werden, um Synergien bei der Vorhersage von Studienabbruch zu schaffen, die wiederum die Prognosegüte steigern würden. Die Verknüpfung von Befragungsdaten mit administrativen Daten ist somit nicht nur für zukünftige Forschung, sondern auch im Ausblick auf anknüpfende Implementierungen in der Hochschulpraxis wünschenswert.

12.8 Gedanken zu *Open Science* und guter wissenschaftlicher Praxis

Zum Abschluss der vorliegenden Arbeit möchte ich zum Forschungsethos an der Justus-Liebig-Universität Gießen und zum Thema *Open Science* Stellung nehmen, um zu verdeutlichen, dass ich Anknüpfungspunkte an mein Promotionsprojekt nicht nur sehe, sondern das Aufgreifen dieser aktiv unterstützen möchte.

In den letzten Jahren sind die Themen *Open Science* und *gute wissenschaftliche Praxis* immer mehr in den Fokus sowohl öffentlicher als auch disziplin-interner Diskussionen gerückt. Insbesondere auch in der Psychologie wird die *Replikationskrise* oft thematisiert. Die Justus-Liebig-Universität Gießen verpflichtet alle ihr angehörigen Forschenden zur Einhaltung guter wissenschaftlicher Praxis. Hierunter wird in erster Linie Folgendes verstanden:

- Untersuchungen sollen nach dem neusten Stand der Erkenntnis durchgeführt werden (Kenntnis des Forschungsstandes und Auswahl angemessener Methodik)

- genaue Protokollierung und Dokumentation (auch zum Zwecke der Replizierbarkeit)
- Infragestellen der Ergebnisse, bis sie eine plausible Erklärung darstellen
- die Erkenntnisse sollen öffentlich zur Verfügung gestellt werden (Publikation)
- disziplin-interne Grundsätze wissenschaftlichen Arbeitens sind einzuhalten

Dem fühle ich mich selbstverständlich auch verpflichtet. Ich möchte allerdings weiter gehen und die im Zuge meines Promotionsprojektes von mir selbst erhobenen Daten (datenschutzkonform in anonymisierter Form) öffentlich zur Verfügung stellen. Weiterhin habe ich mich dazu entschlossen, zusätzlich exemplarisch zumindest einige Syntaxen zu den von mir durchgeführten Analysen bereitzustellen, um eine Überprüfung meiner Analysen und der von mir berichteten Ergebnisse problemlos zu ermöglichen. Dies soll zur Transparenz meiner Forschung beitragen.

Die Daten aus den drei Längsschnittstichproben sowie einige SAS- und R-Syntaxen finden sich auf meinem Profil des *Open Science Frameworks* (OSF): osf.io/6w425. Dort befinden sich auch ausgewählte Konferenzbeiträge (Foliensätze von Vorträgen und Poster). Dies soll ebenfalls dazu beitragen, meine im Rahmen dieses Promotionsprojektes gewonnenen Erkenntnisse einerseits der Öffentlichkeit zugänglich und andererseits potentiell replizierbar zu machen.

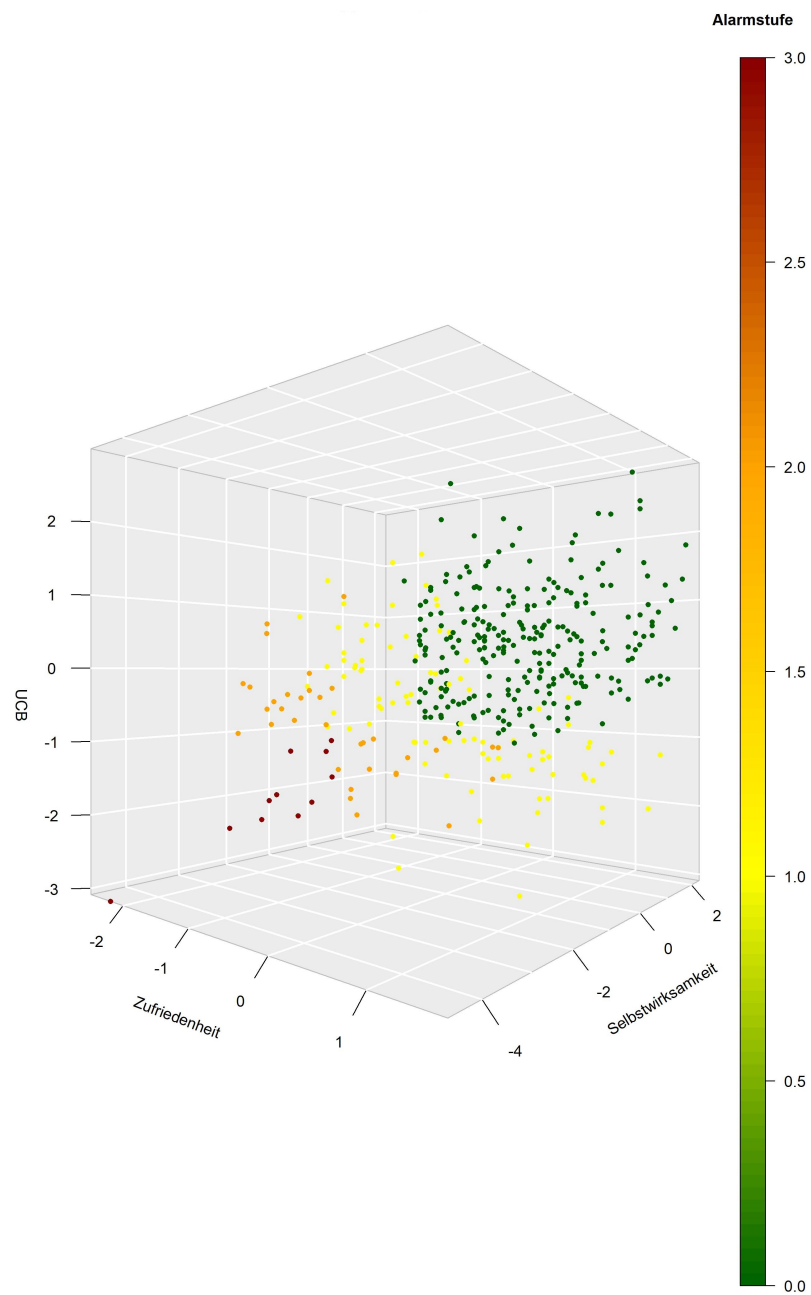


Abbildung 12.1: Darstellung eines Ampelsystems anhand von drei exemplarisch ausgewählten Prädiktoren

Abbildungsverzeichnis

1.1	Abbruchquoten 2006 bis 2014	17
2.1	DZHW Modell	50
2.2	Spadys Dropout Modell (1971)	53
2.3	Tintos Integration Modell (1975)	54
2.4	Erwartungs-Wert-Modell nach Neuville et al. 2007	57
2.5	CSM-Modell nach Lent und Brown (2013)	59
3.1	SESW-Skala	69
3.2	Verteilung des SESW-Skalenwerts	72
3.3	SESW: Tracelines	75
3.4	SESW: Testinformation	76
3.5	SESW: Item Information Curves	76
3.6	SESW Korrelogramm	77
3.7	Hierarchisches Modell der SESW	79
3.8	SESW im Gruppenvergleich (Boxplot)	86
3.9	SESW im Gruppenvergleich (3D)	87
4.1	Überblick Längsschnittstudien	93
5.1	Studie 1: Verteilung auf die Fachcluster	114
5.2	Studie 1: Spadys Modell	118
5.3	Studie 1: Tintos Modell	119
5.4	Studie 1: Neuville et al.'s Modell	119
5.5	Studie 1: CSM-Modell	120
6.1	EOS-Modell	124
6.2	Neues Modell J1	127
7.1	Studie 2: Verteilung auf die Fachcluster	130
7.2	Studie 2: Spadys Modell	134
7.3	Studie 2: Tintos Modell	135
7.4	Studie 2: Neuville et al.'s Modell	135
7.5	Studie 2: CSM-Modell	136
7.6	Studie 2: EOS-Modell	136

8.1	Studie 3: Verteilung auf die Fachcluster	140
8.2	Studie 3: Spadys Modell	143
8.3	Studie 3: Tintos Modell	144
8.4	Studie 3: Neuville et al.'s Modell	145
8.5	Studie 3: CSM-Modell	145
8.6	Studie 3: EOS-Modell	146
9.1	Aggregierter Datensatz: EOS-Modell	152
10.1	Auszug aus dem EOS-Modell	155
10.2	21. Sozialerhebung: Abbruchintention	161
10.3	21. Sozialerhebung: Studiennote	161
10.4	21. Sozialerhebung: Residuen (Abbruchintention)	163
10.5	21. Sozialerhebung: Residuen je Regressor (Abbruchintention)	163
10.6	21. Sozialerhebung: Residuen (Studiennote)	164
10.7	21. Sozialerhebung: Residuen je Regressor (Studiennote)	165
11.1	Ablauf der Datenerhebung	176
11.2	Verteilung der SESW-Skalenwerts vor dem Workshop	178
11.3	Verteilung der SESW-Skalenwerts nach dem Workshop	178
11.4	Intraindividuelle Veränderung des SESW-Scores, $N = 25$	179
12.1	Ampelsystem	194
13.1	SESW-Entwicklung im Gruppenvergleich	III
13.2	Sekundärdatenanalyse: Korrelogramm	XXXV
13.3	Interventionsstudie: SESW (prä)	XXXVI
13.4	Interventionsstudie: SESW (post)	XXXVII

Tabellenverzeichnis

1.1	Abbruchquote 2012 nach Heublein et al. 2014	18
2.1	Überblick: Bedeutsamkeit vers. Prädiktorguppen	45
3.1	Herausforderungen im Studieneinstieg	67
3.2	Deskriptive Statistiken gemäß KTT	73
3.3	Deskriptive Statistiken gemäß IRT	74
3.4	SESW: Messinvarianz über Jahrgänge	84
3.5	SESW: Deskriptive Statistiken im Gruppenvergleich	88
3.6	SESW: Messinvarianz über Messzeitpunkte	89
3.7	SESW: Latent Growth Analysis	90
4.1	Längsschnittstudien: Überblick über die verwendeten Instrumente	96
5.1	Studie 1: Deskriptive Statistiken	116
5.2	Studie 1: Modellfitindizes	118
5.3	Studie 1: AICs	121
6.1	Studie 1: Modellfitindizes (inklusive EOS-Modell)	127
6.2	Studie 1: AICs (inklusive EOS-Modell)	128
7.1	Studie 2: Deskriptiven Statistiken	132
7.2	Studie 2: Modellfitindizes für die fünf getesteten Modelle	134
7.3	Studie 2: AICs	137
8.1	Studie 3: Deskriptive Statistiken	142
8.2	Studie 3: Modellfitindizes	144
8.3	Studie 3: AICs	146
9.1	AICs für alle fünf Modelle (Studie 1-3)	149
9.2	EOS-Modell: Messinvarianz über die Jahrgänge	151
9.3	Aggregierter Datensatz; Modellfit EOS-Modell	152
10.1	21. Sozialerhebung: Verteilung auf Fachcluster	157
10.2	21. Sozialerhebung: Deskriptive Statistiken	160
10.3	21. Sozialerhebung: Regression der Abbruchintention	166

10.4 21. Sozialerhebung: Regression der Studiennote	166
10.5 21. Sozialerhebung: Varianzaufklärung Abbruchintention	167
10.6 21. Sozialerhebung: Varianzaufklärung Abbruchintention	167
11.1 Demografische Angaben der Teilnehmer	175
13.1 SESW-Skala: Normtabelle	II
13.2 Korrelationen SESW-Items	IV
13.3 Korrelationen SESW-Skalenwerte (PG)	V
13.4 Korrelationen SESW-Skalenwerte (DG)	V
13.5 Korrelationen nullter Ordnung für alle in Spadys Modell enthaltenen Variablen (Jahrgang 1)	VII
13.6 Korrelationen nullter Ordnung für alle in Tintos Modell enthaltenen Variablen (Jahrgang 1)	VIII
13.7 Korrelationen nullter Ordnung für alle in Neuville et al.s Modell enthaltenen Variablen (Jahrgang 1)	IX
13.8 Korrelationen nullter Ordnung für alle im CSM Modell enthaltenen Variablen (Jahrgang 1)	X
13.9 Korrelationen nullter Ordnung für alle im EOS-Modell enthaltenen Variablen (Jahrgang 1)	XI
13.10 Korrelationen nullter Ordnung für alle in Spadys Modell enthaltenen Variablen (Jahrgang 2)	XIII
13.11 Korrelationen nullter Ordnung für alle in Tintos Modell enthaltenen Variablen (Jahrgang 2)	XIV
13.12 Korrelationen nullter Ordnung für alle in Neuville et al.s Modell enthaltenen Variablen (Jahrgang 2)	XV
13.13 Korrelationen nullter Ordnung für alle im CSM Modell enthaltenen Variablen (Jahrgang 2)	XVI
13.14 Korrelationen nullter Ordnung für alle im EOS-Modell enthaltenen Variablen (Jahrgang 2)	XVII
13.15 Korrelationen nullter Ordnung für alle in Spadys Modell enthaltenen Variablen (Jahrgang 3)	XIX
13.16 Korrelationen nullter Ordnung für alle in Tintos Modell enthaltenen Variablen (Jahrgang 3)	XX
13.17 Korrelationen nullter Ordnung für alle in Neuville et al.'s Modell enthaltenen Variablen (Jahrgang 3)	XXI
13.18 Korrelationen nullter Ordnung für alle im CSM-Modell enthaltenen Variablen (Jahrgang 3)	XXII
13.19 Korrelationen nullter Ordnung für alle im EOS-Modell enthaltenen Variablen (Jahrgang 3)	XXIII

13.20 Spadys Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1	XXIV
13.21 Spadys Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2	XXV
13.22 Spadys Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3	XXVI
13.23 Tintos Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1	XXVII
13.24 Tintos Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2	XXVIII
13.25 Tintos Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3	XXIX
13.26 Neuville et al.'s Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1	XXX
13.27 Neuville et al.'s Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2	XXX
13.28 Neuville et al.'s Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3	XXXI
13.29 Modell nach Lent und Brown: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1	XXXI
13.30 Modell nach Lent und Brown: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2	XXXII
13.31 Modell nach Lent und Brown: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3	XXXIII
13.32 EOS-Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1	XXXIII
13.33 EOS-Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2	XXXIV
13.34 EOS-Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3	XXXIV

Literatur

- Alesi, B., Bürger, S., Kehm, B. M. & Teichler, U. (2005). *Stand der Einführung von Bachelor- und Master-Studiengängen in Bologna-Prozess sowie in ausgewählten Ländern Europas im Vergleich zu Deutschland*. Bonn.
- Artelt, C., Demmrich, A. & Baumert, J. (2001). Selbstreguliertes Lernen. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, ... M. Weiß (Hrsg.), *PISA 2000* (S. 271–298). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:\url{10.1007/978-3-322-83412-6_8}
- Asendorpf, J. B., van de Schoot, R., Denissen, J. J. A. & Hutteman, R. (2014). Reducing bias due to systematic attrition in longitudinal studies: The benefits of multiple imputation. *International Journal of Behavioral Development*, 38(5), 453–460. doi:10.1177/0165025414542713
- Bajaj, B. & Pande, N. (2016). Mediating role of resilience in the impact of mindfulness on life satisfaction and affect as indices of subjective well-being. *Personality and Individual Differences*, 93, 63–67. PII: S0191886915005784. doi:10.1016/j.paid.2015.09.005
- Baltes-Götz, B. (2019). *Lineare Regressionsanalyse mit SPSS*. Universität Trier: Zentrum für Informations-, Medien- und Kommunikationstechnologie (ZIMK).
- Bandura, A. (1977). Self-Efficacy. *Psychological Review*, 84, 191–215.
- Bandura, A. (1980). Tests of the generability of Self-Efficacy Theory. *Cognitive Therapy and Research*, 4, 39–66.
- Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Annual review of psychology*, 52, 1–26. Journal Article Research Support, Non-U.S. Gov't. doi:\url{10.1146/annurev.psych.52.1.1}. eprint: 11148297
- Bargel, T. (2015). *Studieneingangsphase und heterogene Studentenschaft - neue Angebote und ihr Nutzen. Befunde des 12. Studierenden surveys an Universitäten und Fachhochschulen: Hefte zur Bildungs- und Hochschulforschung*. Universität Konstanz: Arbeitsgruppe Hochschulforschung.
- Barnat, M., Abelha Faria, J. & Bosse, E. (2017). Heterogenität und Studierfähigkeit: Erste Ergebnisse einer Längsschnittbefragung. *Qualität in der Wissenschaft, Sonderheft zu Studienerfolgshforschung*, (1), 17–24.
- Barnat, M., Bosse, E. & Mergner, J. (2017). Forschungsbasierte Qualitätsentwicklung für die Studieneingangsphase. *Zeitschrift für Hochschulentwicklung*, 12 (2), 71–91.

- Baron-Boldt, J. (1989). *Die Validität von Schulabschlussnoten für die Prognose von Ausbildungs- und Studienerfolg: Eine Metaanalyse nach dem Prinzip der Validitätsgeneralisierung*. 06. Lang. Zugriff unter <https://books.google.de/books?id=JJvUAAAACAAJ%7D>
- Bean, J. P. (1985). Interaction Effects Based on Class Level in an Explanatory Model of College Student Dropout Syndrome. *American Educational Research Journal*, 22(1), 35–64. doi:[10.3102/00028312022001035](https://doi.org/10.3102/00028312022001035)
- Beermann, D., Kersting, M., Stegt, S. & Zimmerhofer, A. (2013). Vorurteile und Urteile zur Akzeptanz von Persönlichkeitsfragebogen. *PersonalQuarterly*, 65, 41–45.
- Beierlein, C., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2013). Kurzskala zur Erfassung allgemeiner Selbstwirksamkeitserwartungen (ASKU). *Methoden, Daten, Analysen*, 7(2), 251–278. doi:[10.12758/mda.2013.01](https://doi.org/10.12758/mda.2013.01)
- Berens, J. & Schneider, K. (2019). Drohender Studienabbruch: Wie gut sind Frühwarnsysteme? *Qualität in der Wissenschaft*, 102–107.
- Betz, N. E. & Hackett, G. (2006). Career Self-Efficacy Theory: Back to the Future. *Journal of Career Assessment*, 14, 3–11. doi:[10.1177/1069072705281347](https://doi.org/10.1177/1069072705281347)
- Bibliographisches Institut. (2019). Duden. Zugriff unter <https://www.duden.de/node/675633/revisions/1968046/view>
- Bloom, D. E., Hartley, M. & Rosovsky, H. (2006). Beyond Private Gain: The Public Benefits of Higher Education. In J. J. F. Forest & P. G. Altbach (Hrsg.), *International Handbook of Higher Education* (Bd. 18, 18, S. 293–308). Springer International Handbooks of Education. Dordrecht: Springer. doi:[10.1007/978-1-4020-4012-2_15](https://doi.org/10.1007/978-1-4020-4012-2_15)
- Blossfeld, H.-P., Roßbach, H.-G. & Maurice, J. (2016). (Eds.). Education as a lifelong process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft, Special Issue 14*. doi:[10.5157/NEPS:SC5:8.0.0](https://doi.org/10.5157/NEPS:SC5:8.0.0)
- Blossfeld, H.-P., Roßbach, H.-G. & von Maurice, J. (2011). Education as a Lifelong Process: The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft*.
- Blüthmann, I. (2012). Individuelle und studienbezogene Einflussfaktoren auf die Zufriedenheit von Bachelorstudierenden. *Zeitschrift für Erziehungswissenschaft*, 15 (2), 273–303. doi:[10.1007/s11618-012-0270-3](https://doi.org/10.1007/s11618-012-0270-3)
- BMBF. (2016a). Richtlinie zur Förderung von Forschung über "Studienerfolg und Studienabbruch". Zugriff unter <https://www.bmbf.de/foerderungen/bekanntmachung-1151.html%7D>
- Studienabbruch - und dann? (2016b, 19. Juli). Neues Onlineportal mit Hilfsangeboten für Studienzweifler/innen. Zugriff unter <https://www.bmbf.de/de/studienabbruch-und-dann-3131.html%7D>
- BMBF. (2018). Studienerfolg und Studienabbruch. Zugriff unter <https://www.wihoforschung.de/de/studienerfolg-und-studienabbruch-28.php%7D>

- BMBF. (2019). Aufstieg durch Bildung. BMBF. Zugriff 17. Februar 2019 unter <https://www.bmbf.de/de/aufstieg-durch-bildung-1240.html>
- Bornkessel, P. (Hrsg.). (2018). *Erfolg im Studium. Konzeptionen, Befunde und Desiderate*. doi:[10.3278/6004654w](https://doi.org/10.3278/6004654w)
- Bortz, J. & Schuster, C. [C.]. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollständig überarbeitete und erweiterte Auflage). Springer-Lehrbuch. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. doi:[10.1007/978-3-642-12770-0](https://doi.org/10.1007/978-3-642-12770-0)
- Bosse, E., Mergner, J., Wallis, M., Jänsch, V. K. & Kunow, L. (2019). *Gelingendes Studieren in der Studieneingangsphase. Ergebnisse und Anregungen für die Praxis aus der Begleitforschung zum Qualitätspakt Lehre im Projekt StuFHe*. (). Universität Hamburg. doi:[10.25592/StuFHe2019](https://doi.org/10.25592/StuFHe2019)
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten: Soziale Welt Sonderband 2* (S. 183–196). Göttingen.
- Bourdieu, P. & Passeron, J. C. (1977). *Reproduction in Education, Society and Culture*. London: Sage Publications.
- Bourdieu, P. & Passeron, J.-C. (1971). *Die Illusion der Chancengleichheit: Untersuchungen zur Soziologie des Bildungswesens am Beispiel Frankreichs* (1. Aufl.). Texte und Dokumente zur Bildungsforschung. Bourdieu, Pierre (VerfasserIn) Passeron, Jean-Claude (VerfasserIn). Stuttgart: Klett.
- Bowden, J. L.-H., Tickle, L. & Naumann, K. (2019). The four pillars of tertiary student engagement and success: a holistic measurement approach. *Studies in Higher Education*, 5(3), 1–18. doi:[10.1080/03075079.2019.1672647](https://doi.org/10.1080/03075079.2019.1672647)
- Brahm, T., Jenert, T. & Wagner, D. (2014). Nicht für alle gleich: subjektive Wahrnehmungen des Übergangs Schule - Hochschule. *Zeitschrift für Hochschulentwicklung*, 5, 63–82.
- Brandstätter, H. & Farthofer, A. (2003). Einfluss von Erwerbstätigkeit auf den Studienerfolg. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, 47, 134–145. doi:[10.1026/0932-4089.47.3.134](https://doi.org/10.1026/0932-4089.47.3.134)
- Brandstätter, H., Grillich, L. & Farthofer, A. (2006). Prognose des Studienabbruchs. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38, 121–131. doi:[10.1026/0049-8637.38.3.121](https://doi.org/10.1026/0049-8637.38.3.121)
- Britton, B. K. & Tesser, A. (1991). Effects of time-management practices on college grades. *Journal of Educational Psychology*, 83, 405–410. doi:[10.1037/0022-0663.83.3.405](https://doi.org/10.1037/0022-0663.83.3.405)
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62 (3), 193–217. doi:[10.1037/h0047470](https://doi.org/10.1037/h0047470)
- Bühler, T. (2012). *Studierende aus nichtakademischen Elternhäusern im Studium: Expertise im Rahmen des Projektes „Chancengleichheit in der Begabtenförderung“ der Hans-Böckler-Stiftung*. Düsseldorf: Hans-Böckler-Stiftung.
- BVerfG. (2017-12-19). Leitsätze zum Urteil des Bundesverfassungsgerichts. Zugriff unter https://www.bundesverfassungsgericht.de/e/lsg20171219_1bvl000314.html

- Cable, D. M. & DeRue, D. S. (2002). The convergent and discriminant validity of subjective fit perceptions. *Journal of Applied Psychology*, 87, 875–884. doi:[10.1037/0021-9010.87.5.875](https://doi.org/10.1037/0021-9010.87.5.875)
- Chemers, M. M., Hu, L. & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational Psychology*, 93, 55–64. doi:[10.1037/0022-0663.93.1.55](https://doi.org/10.1037/0022-0663.93.1.55)
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14 (3), 464–504. doi:[10.1080/10705510701301834](https://doi.org/10.1080/10705510701301834)
- Chen, H., Dai, J. & Gao, Y. (2019). Measurement invariance and latent mean differences of the Chinese version physical activity self-efficacy scale across gender and education levels. *Journal of sport and health science*, 8(1), 46–54. doi:[10.1016/j.jshs.2017.01.004](https://doi.org/10.1016/j.jshs.2017.01.004)
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, S., Kamarck, T. & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 385–396.
- Cole, D. A. & Preacher, K. J. (2014). Manifest variable path analysis: potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, 19 (2), 300–315. doi:[10.1037/a0033805](https://doi.org/10.1037/a0033805)
- Costa, P. T. & McCrae, R. R. (1992). Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory. *Psychological Assessment*, 4, 5–13.
- Dahm, G., Becker, K. & Bornkessel, P. (2018). Determinanten des Studienerfolgs nicht-traditioneller Studierender – zur Bedeutung der sozialen und akademischen Integration, der Lebensumstände und des Studienkontextes für die Studienabbruchneigung beruflich qualifizierter Studierender ohne Abitur. In P. Bornkessel (Hrsg.), *Erfolg im Studium. Konzeptionen, Befunde und Desiderate*.
- Dahm, G. & Kerst, C. (2016). Erfolgreich studieren ohne Abi? Ein mehrdimensionaler Vergleich des Studienerfolgs von nicht-traditionellen und traditionellen Studierenden. In A. Wolter, U. Banscherus & C. Kamm (Hrsg.), *Zielgruppen Lebenslangen Lernens an Hochschulen*. Waxmann Verlag.
- De Clercq, M., Galand, B., Dupont, S. & Frenay, M. (2013). Achievement among first-year university students: An integrated and contextualised approach. *European Journal of Psychology of Education*, 28, 641–662. doi:[10.1007/s10212-012-0133-6](https://doi.org/10.1007/s10212-012-0133-6)
- De Fruyt, F. & Mervielde, I. (1996). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality*, 10(5), 405–425. doi:[10.1002/\(SICI\)1099-0984\(199612\)10:5<405::AID-PER255>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1099-0984(199612)10:5<405::AID-PER255>3.0.CO;2-M)
- De Leeuw, E. D., Hox, J. & Dillman, D. (2012). *International handbook of survey methodology*. New York: Routledge.
- de Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

- DeBerard, M. S., Spielmans, G. I. & Deana, C. J. (2004). Predictors Of Academic Achievement And Retention Amongcollege Freshmen: A Longitudinal Study. *38*(1), 66–80.
- Deci, E. L. & Ryan, R. M. (2000). The What and Why of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), 227–268. doi:\url{10.1207/S15327965PLI1104_01}
- Demirtas, H. (2004). Modeling Incomplete Longitudinal Data. *Journal of Modern Applied Statistical Methods*, *3*(2), 305–321. doi:\url{10.22237/jmasm/1099267500}
- Dong, Y. & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. (2). Zugriff unter <http://www.springerplus.com/content/2/1/222>
- Duncan, T. E. & Duncan, S. C. (2009). The ABC's of LGM: An Introductory Guide to Latent Variable Growth Curve Modeling. *Social and Personality Psychology Compass*, *3* (6)(6), 979–991. Journal Article. doi:10.1111/j.1751-9004.2009.00224.x. eprint: 20577582
- Durkheim, E. (1951). *Suicide. Translated by L A. Spaulding & G. Simpson*. Glencoe, Ill.: The Free Press.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048. doi:10.1037/0003-066X.41.10.1040
- Eccles, J. S. & Wigfield, A. (2016). In the Mind of the Actor: The Structure of Adolescents' Achievement Task Values and Expectancy-Related Beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215–225. doi:\url{10.1177/0146167295213003}
- Ehlert, M., Peter, F., Finger, C., Rusconi, A., Solga, H., Spieß, C. K. & Zambre, V. (2017). The Berliner-Studienberechtigten-Panel (Best Up) – Methodological and Data Report. Berlin.
- Ehrlich, S. (2016). Studierendebefragung 2016: Ausgewählte Ergebnisse der zwölften Studierendenbefragung der Justus-Liebig-Universität Gießen. Zugriff unter <https://www.uni-giessen.de/org/admin/stab/stl/servicestelle/datenportal/datenstb>
- Eid, M. (2017). Messinvarianz. In M. A. Wirtz (Hrsg.), *Dorsch - Lexikon der Psychologie* (18., überarbeitete Auflage). Bern: Hogrefe. Zugriff 9. September 2019 unter <https://m.portal.hogrefe.com/dorsch/messinvarianz/>
- Eid, M. & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Bachelorstudium Psychologie. Eid, Michael (VerfasserIn) Schmidt, Katharina (VerfasserIn). Göttingen: Hogrefe. Zugriff unter http://sub-hh.ciando.com/book/?bok_id=1548490
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*(4), 352–370. doi:\url{10.1037/1082-989X.6.4.352}
- Evans, J. H. & Burck, H. D. (1992). The Effects of Career Education Interventions on Academic Achievement: A Meta-Analysis. *Journal of Counseling & Development*, *71*(1), 63–68. doi:10.1002/j.1556-6676.1992.tb02173.x
- Farsides, T. & Woodfield, R. (2003). Individual differences and undergraduate academic success: the roles of personality, intelligence, and application. *Personality and Individual Differences*, *34*(7), 1225–1243. doi:10.1016/S0191-8869(02)00111-3

- Feldhaus, M. & Baalmann, T. (2019). Soziales Kapital von Eltern und der Studienerfolg von Studierenden. *Qualität in der Wissenschaft*, 83–87.
- Fellenberg, F. & Hannover, B. (2006). Kaum begonnen, schon zerronnen? Psychologische Ursachenfaktoren für die Neigung von Studienanfängern, das Studium abzubrechen oder das Fach zu wechseln. *Empirische Pädagogik*, 20, 381–399.
- Gall, M. D., Gall, J. P. & Borg, W. R. (2002). *Educational research: An introduction*. Princeton, N.J.: Recording for the Blind & Dyslexic.
- Gehring, F. (2006). *Entwicklung und Validierung eines Fragebogens zu umfeldbezogener Leistung an Hochschulen: Unveröffentlichte Diplomarbeit*. Aachen.
- Georg, W. (2009). Individual and institutional factors in the tendency to drop out of higher education: a multilevel analysis using data from the Konstanz Student Survey. *Studies in Higher Education*, 34 (6)(6), 647–661. doi:[10.1080/03075070802592730](https://doi.org/10.1080/03075070802592730)
- Gold, A. & Souvignier, E. (2005). Prognose der Studierfähigkeit. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 37 (4), 214–222. doi:[10.1026/0049-8637.37.4.214](https://doi.org/10.1026/0049-8637.37.4.214)
- Göritz, A. S. (2006). Incentives in Web Studies: Methodological Issues and a Review. *International Journal of Internet Science*, (1 (1)), 58–70.
- Gottfredson, L. S. & Crouse, J. (1986). Validity versus utility of mental tests: Example of the SAT. *Journal of Vocational Behavior*, 29, 363–378.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, 60, 549–576. Journal Article Review. doi:[10.1146/annurev.psych.58.110405.085530](https://doi.org/10.1146/annurev.psych.58.110405.085530). eprint: [18652544](https://eprint.aps.org/abstract/18652544)
- Hahm, S. & Stock, J. (2018). Das Potenzial administrativer Daten für das Qualitätsmanagement an Hochschulen. *Zeitschrift für Hochschulentwicklung*, (1), 193–207.
- Hasenberg, S. (2012). *Zur prädiktiven Validität von Self-Assessments für die Studienzufriedenheit: unveröffentlichte Dissertation*. Marburg.
- Hattie, J., Biggs, J. & Purdie, N. (1996). Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis. *Review of Educational Research*, 66 (2), 99–136.
- Heublein, U. (2014). Student Drop-out from German Higher Education Institutions. *European Journal of Education*, 49, 497–513. doi:[10.1111/ejed.12097](https://doi.org/10.1111/ejed.12097)
- Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., König, R., Richter, J. & Woisch, A. (2017). *Zwischen Studiererwartungen und Studienwirklichkeit: Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen*. Hannover: DZHW.
- Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., Richter, J. & Schreiber, J. (2015). *Studienbereichsspezifische Qualitätssicherung im Bachelorstudium: Befragung der Fakultäts- und Fachbereichsleitungen zum Thema Studienerfolg und Studienabbruch*. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D. & Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen: Ergebnisse einer bun-*

- desweiten Befragung von Exmatrikulierten des Studienjahres 2007/08. Hannover: HIS: Hochschul-Informationen-System GmbH.
- Heublein, U., Richter, J., Schmelzer, R. & Sommer, D. (2014). *Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen: Statistische Berechnungen auf der Basis des Absolventenjahrgangs 2012*. Hannover.
- Heublein, U. & Wolter, A. (2011). Studienabbruch in Deutschland: Definition, Häufigkeit, Ursachen, Maßnahmen. *Zeitschrift für Pädagogik*, 57, 214–236.
- Hiemisch, A., Westermann, R. & Michael, A. (2005). Die Abhängigkeit der Zufriedenheit mit dem Medizinstudium von Studienzielen und ihrer Realisierbarkeit. *Zeitschrift für Psychologie / Journal of Psychology*, 213, 97–108. doi:\url{10.1026/0044-3409.213.2.97}
- Hlavac, M. (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. Zugriff unter <https://CRAN.R-project.org/package=stargazer>
- Hofmann, Y. E., Datzler, D., Razinskas, S. & Högl, M. (2019). Die Rolle von Resilienz bei Studienabbruchentscheidungen Stand der internationalen Forschung und erste Ergebnisse einer nationalen Studie. *Qualität in der Wissenschaft*, 77–82.
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa, FL: Psychological Assessment Resources.
- Hooper, D., Coughlan, J. & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, 6(1), 53–60.
- Horstmann, N. & Hachmeister, C.-D. (2016). *Anforderungsprofile für die Fächer im CHE Hochschulranking aus Professor(inn)ensicht*. Gütersloh: CHE gemeinnütziges Zentrum für Hochschulentwicklung.
- Hu, L. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3 (4), 424–453. doi:\url{10.1037/1082-989X.3.4.424}
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi:\url{10.1080/10705519909540118}
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29 (3), 340–362. PII: 0001879186900138. doi:10.1016/0001-8791(86)90013-8
- Hüther, O. & Krücken, G. (2018). *Higher Education in Germany-Recent Developments in an International Perspective*. Higher Education Dynamics. Cham: Springer International Publishing. Zugriff unter <http://dx.doi.org/10.1007/978-3-319-61479-3>
- IBM Corp. (2012). SPSS Statistics for Windows, Version 21.0. Armonk, NY.
- International Test Commission. (2016). *International Test Commission. (2016). The ITC Guidelines for Translating and Adapting Tests (Second edition)*. Zugriff unter www.InTestCom.org

- Isiordia, M. & Ferrer, E. (2018). Curve of Factors Model: A Latent Growth Modeling Approach for Educational Research. *Educational and Psychological Measurement*, 78 (2), 203–231. doi:[10.1177/0013164416677143](https://doi.org/10.1177/0013164416677143). eprint: [29795953](https://eprints.org/29795953)
- Isleib, S. (2015, 10. April). Neue Theorieströmungen zum Studienabbruch: Herkunft, Genese und Potenziale für die Studienabbruchs- und Hochschulforschung. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung. Kassel.
- Isleib, S., Woisch, A. & Heublein, U. (2019). Ursachen des Studienabbruchs: Theoretische Basis und empirische Faktoren. *Zeitschrift für Erziehungswissenschaft*, 22(5), 1047–1076. doi:[10.1007/s11618-019-00908-x](https://doi.org/10.1007/s11618-019-00908-x)
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M. & Rosseel, Y. (2018). semTools: Useful tools for structural equation modeling. R package version 0.5-1. Zugriff unter <https://CRAN.R-project.org/package=semTools>
- Kadane, J. B. & Lazar, N. A. (2004). Methods and Criteria for Model Selection. *Journal of the American Statistical Association*, 99(465), 279–290. doi:[10.1198/016214504000000269](https://doi.org/10.1198/016214504000000269)
- Kahn, J. H. & Nauta, M. M. (2001). Social-Cognitive Predictors of First-Year College Persistence: The Importance of Proximal Assessment. *Research in Higher Education*, 42(6), 633–652.
- Kazdin, A. E. (1973). Covert modeling and the reduction of avoidance behavior. *Journal of Abnormal Psychology*, 81, 87–95. doi:[10.1037/h0034036](https://doi.org/10.1037/h0034036)
- Kerby, M. B. (2015). Toward a New Predictive Model of Student Retention in Higher Education: An Application of Classical Sociological Theory, 1–24.
- Kersting, M. (2003). Augenscheinvalidität. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 54–55). Weinheim: Beltz, PVU.
- Kirk, R. E. (1996). Practical Significance: A Concept Whose Time Has Come. *Educational and Psychological Measurement*, 56(5), 746–759. doi:[10.1177/0013164496056005002](https://doi.org/10.1177/0013164496056005002)
- Klein, D. & Stocké, V. (2016). Studienabbruchquoten als Evaluationskriterium und Steuerungsinstrument der Qualitätssicherung im Hochschulbereich. In D. Großmann & T. Wolbring (Hrsg.), *Evaluation von Studium und Lehre*. Wiesbaden: Springer VS.
- Klein, E. M., Brahler, E., Dreier, M., Reinecke, L., Müller, K. W., Schmutzer, G., ... Beutel, M. E. (2016). The German version of the Perceived Stress Scale - psychometric characteristics in a representative German community sample. *BMC psychiatry*, 16, 159. Journal Article. doi:[10.1186/s12888-016-0875-9](https://doi.org/10.1186/s12888-016-0875-9). eprint: [27216151](https://eprints.org/27216151)
- Klöppling, S., Scherfer, M., Gokus, S., Dachsberger, S., Krieg, A., Wolter, A., ... Umbach, E. (2017). *Studienabbruch in den Ingenieurwissenschaften: Empirische Analyse und Best Practices zum Studienerfolg*. München: Herbert Utz Verlag.
- Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland. *Psychologische Rundschau*, 60 (2), 82–98. doi:[10.1026/0033-3042.60.2.82](https://doi.org/10.1026/0033-3042.60.2.82)

- Krempkow, R. (in Druck). Determinanten der Studiendauer – individuelle oder institutionelle Faktoren? Sekundärdatenanalyse einer bundesweiten Absolvent(inn)enbefragung. *Zeitschrift für Evaluation*.
- Krempkow, R. (2005). *Leistungsbewertung und Leistungsanreize in der Hochschullehre: Eine Untersuchung von Konzepten, Leistungskriterien und Bedingungen erfolgreicher Institutionalisierung*. Dresden: Technische Universität Dresden.
- Krempkow, R. (2008). Studienerfolg, Studienqualität und Studierfähigkeit. Eine Analyse zu Determinanten des Studienerfolgs in 150 sächsischen Studiengängen. *Die Hochschule*, (1), 91–107.
- Krempkow, R. (2015). Can Performance-Based Funding Enhance Diversity in Higher Education Institutions? In R. M. O. Pritchard, M. Klumpp & U. Teichler (Hrsg.), *Diversity and Excellence in Higher Education: Can the Challenges be Reconciled?* (Bd. 4, S. 231–244). Rotterdam: SensePublishers. doi:[10.1007/978-94-6300-172-4_13](https://doi.org/10.1007/978-94-6300-172-4_13)
- Krempkow, R. (2017). Welche Faktoren bestimmen die Übergänge nach dem Bachelor? Eine empirische Analyse anhand von Absolvent/innenstudien. *Qualität in der Wissenschaft*, (3+4), 86–97.
- Kristen, C. (1999). *Bildungsentscheidungen und Bildungsungleichheit – ein Überblick über den Forschungsstand: Arbeitspapiere, Nr. 5*. Mannheim: Mannheimer Zentrum für europäische Sozialforschung.
- Kuh, G. D. (2009). What Student Affairs Professionals Need to Know About Student Engagement. *Journal of College Student Development*, 50(6), 683–706. doi:[10.1353/csd.0.0099](https://doi.org/10.1353/csd.0.0099)
- Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: can one construct predict them all? *Journal of Personality and Social Psychology*, 86(1), 148–161. Journal Article Meta-Analysis Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. doi:[10.1037/0022-3514.86.1.148](https://doi.org/10.1037/0022-3514.86.1.148). eprint: [14717633](https://doi.org/10.1037/0022-3514.86.1.148)
- Lawson, M. A. & Lawson, H. A. (2013). New Conceptual Frameworks for Student Engagement Research, Policy, and Practice. *Review of Educational Research*, 83(3), 432–479. doi:[10.3102/0034654313480891](https://doi.org/10.3102/0034654313480891)
- Leiner, D. J. (2016). SoSci Survey (Version 2.6.00): [Computer software]. Zugriff unter [5Curl%7Bhttps://www.sosicurvey.de%7D](https://www.sosicurvey.de)
- Lenhard, A., Lenhard, W., Suggate, S. & Segerer, R. (2018). A Continuous Solution to the Norming Problem. *Assessment*, 25(1), 112–125. doi:[10.1177/1073191116656437](https://doi.org/10.1177/1073191116656437)
- Lent, R. W. & Brown, S. D. (2013). Social cognitive model of career self-management: toward a unifying view of adaptive career behavior across the life span. *Journal of counseling psychology*, 4, 557–568. Journal Article Review. doi:[10.1037/a0033446](https://doi.org/10.1037/a0033446). eprint: [23815631](https://doi.org/10.1037/a0033446)
- Lent, R. W., Brown, S. D. & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of counseling psychology*, 33(3), 265–269. doi:[10.1037//0022-0167.33.3.265](https://doi.org/10.1037//0022-0167.33.3.265)

- Little, R. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, (83(404)), 1198–1202.
- Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 3, 157–189. doi:\url{10.1016/0030-5073(68)90004-4}
- Lubbe, D. & Schuster, C. [Christof]. (2019). A Graded Response Model Framework for Questionnaires With Uniform Response Formats. *Applied Psychological Measurement*, 43(4), 290–302. doi:10.1177/0146621618789394
- Lüdtke, D. (2018). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.6.2. doi:\url{10.5281/zenodo.1308157}
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau*, 58(2), 103–117. doi:\url{10.1026/0033-3042.58.2.103}
- Marks, G. (2007). *Completing university: Characteristics and outcomes of completing and non completing students*. LSAY Research Reports. Longitudinal surveys of Australian youth.
- Martin, A. (2002). Motivation and Academic Resilience: Developing a Model for Student Enhancement. *Australian Journal of Education*, 46(1), 34–49. doi:\url{10.1177/000494410204600104}
- McArdle, J. J., Hamagami, F., Elias, M. F. & Robbins, M. A. (1991). Structural modeling of mixed longitudinal and cross-sectional data. *Experimental aging research*, 17(1), 29–52. Comparative Study Journal Article Research Support, U.S. Gov't, P.H.S. doi:10.1080/03610739108253884. eprint: 1936099
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58 (4), 525–543. PII: BF02294825. doi:10.1007/BF02294825
- Middendorff, E., Apolinarski, B., Becker, K., Bornkessel, P., Brandt, T., Heißenberg, S. & Poskowsky, J. (2016). 21. Sozialerhebung: Aufbereitet durch Baillet, F. & Weber, A., released 2018. Hannover: FDZ-DZHW. doi:\url{10.21249/DZHW:ssy21:1.0.}
- Middendorff, E., Apolinarski, B., Becker, K., Bornkessel, P., Brandt, T., Heißenberg, S. & Poskowsky, J. (2017). *The Economic and Social Situation of Students in Germany 2016. Summary of the 21st Social Survey of Deutsches Studentenwerk, conducted by the German Centre for Higher Education Research and Science Studies*. Berlin.
- Minnaert, A. (1996). Can metacognition compensate for intelligence in the first year of Belgian higher education? *Psychologica Belgica*, 36(4), 227–244.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage). Springer-Lehrbuch. Berlin: Springer.
- Moshagen, M. & Erdfelder, E. (2016). A New Strategy for Testing Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23.
- Muthén, B. O. & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2 (4), 371–402. doi:10.1037/1082-989X.2.4.371

- Nestler, S., Thielsch, M., Vasilev, E. & Back, M. (2015). Will They Stay or Will They Go? Personality Predictors of Dropout in an Online Study. *International Journal of Internet Science*, (10 (1)), 37–48.
- Neugebauer, M., Heublein, U. & Daniel, A. (2019). Studienabbruch in Deutschland: Ausmaß, Ursachen, Folgen, Präventionsmöglichkeiten. *Zeitschrift für Erziehungswissenschaft*, 22(5), 1025–1046. doi:[10.1007/s11618-019-00904-1](https://doi.org/10.1007/s11618-019-00904-1)
- Neuville, S., Frenay, M., Schmitz, J., Boudrenghien, G., Noël, B. & Wertz, V. (2007). Tintos Theoretical Perspective and Expectancy-Value Paradigm: A Confrontation to Explain Freshmen's Academic Achievement. *Psychologica Belgica*, 47, 31–50.
- Newman, D. A. (2016). Longitudinal Modeling with Randomly and Systematically Missing Data: A Simulation of Ad Hoc, Maximum Likelihood, and Multiple Imputation Techniques. *Organizational Research Methods*, 6(3), 328–362. doi:[10.1177/1094428103254673](https://doi.org/10.1177/1094428103254673)
- Nolden, P. (2019). *Student Experience and Drop-Out Intention. Development and Testing of a Multicausal and Multiperspective Explanatory Model in the University Context: Dissertation*. doi:[10.18154/RWTH-2019-00093](https://doi.org/10.18154/RWTH-2019-00093)
- OECD. (2014). *Education at a Glance 2014: OECD Indicators*. Paris: OECD Publishing.
- OECD. (2016). *Education at a Glance 2016: OECD Indicators*. Paris: OECD Publishing.
- OECD. (2017). *Education at a Glance 2017: OECD Indicators*. Paris: OECD Publishing.
- OECD. (2018). *Education at a Glance 2018: OECD Indicators*. Paris: OECD Publishing.
- OECD. (2019). *Education at a Glance 2019: OECD Indicators*. Paris: OECD Publishing. doi:[10.1787/f8d7880d-en](https://doi.org/10.1787/f8d7880d-en)
- Organ, D. W. (1988). *Organizational Citizenship Behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.
- Osborne, J. & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical assessment, research & evaluation*, (8(2)), 1–9.
- Oseguera, L. & Rhee, B. S. (2009). The Influence of Institutional Retention Climates on Student Persistence to Degree Completion: A Multilevel Approach. *Research in Higher Education*, 50 (6), 546–569. PII: 9134. doi:[10.1007/s11162-009-9134-y](https://doi.org/10.1007/s11162-009-9134-y)
- Peiffer, H., Preckel & Ellwart, T. (2018). Selbstwirksamkeitserwartung von Studierenden. *Diagnostica*, 64 (3), 133–144. doi:[10.1026/0012-1924/a000199](https://doi.org/10.1026/0012-1924/a000199)
- Pekrun, R. (2000). A social cognitive, control-value theory of achievement emotions. In J. Heckhausen (Hrsg.), *Motivational psychology of human development*. Oxford, UK: Elsevier.
- Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, 18, 315–341. doi:[10.1007/s10648-006-9029-9](https://doi.org/10.1007/s10648-006-9029-9)
- Pekrun, R. (2018). Emotion, Lernen und Leistung. In M. Huber & S. Krause (Hrsg.), *Bildung und Emotion* (S. 215–231). Wiesbaden: Springer VS. doi:[10.1007/978-3-658-18589-3_12](https://doi.org/10.1007/978-3-658-18589-3_12)
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P. & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ).

- Contemporary Educational Psychology*, 36, 36–48. doi:\url{10.1016/j.cedpsych.2010.10.002}
- Perry, R. P., Hladkyj, S., Pekrun, R. H. & Pelletier, S. (2001). Academic Control and Action Control in the Achievement of College Students: A Longitudinal Field Study. *Journal of Educational Psychology*, 93, 776–789.
- Petri, P. S., Weingardt, V. & Kersting, M. (2019). Let's get to the hard facts: Erfassung von Intelligenz im Rahmen von Online Self-Assessments. *Empirische Pädagogik*.
- Phillips, J. M. & Gully, S. M. (1997). Role of Goal Orientation, Ability, Need for Achievement, and Locus of Control in the Self-Efficacy and Goal-Setting Process. *Journal of Applied Psychology*, 82 (5), 792–802.
- Pintrich, P. R., Smith, D. A., Garcia, T. & McKeachie, W. J. (1991). *Motivated Strategies for Learning Questionnaire Manual*. Ann Arbor, MI.
- R Core Team. (2017). R: A language and environment for statistical computing. Wien, Österreich: R Foundation for Statistical Computing. Zugriff unter <https://www.R-project.org/>
- Ramm, M., Multrus, F., Barge, T. & Schmidt, M. (2014). *Studiensituation und studentische Orientierungen: 12. Studierendensurvey an Universitäten und Fachhochschulen*. Bonn/Berlin: Bundesministerium für Bildung und Forschung.
- Rammstedt, B., Harkness, J. & Mohler, P. P. (2011). Psychodiagnostische Verfahren im Kulturvergleich. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Methoden der Psychologischen Diagnostik, Enzyklopädie der Psychologie : Serie II, Psychologische Diagnostik Bd 2* (S. 415–471). Göttingen: Hogrefe.
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2012). *Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI-10)*. Mannheim.
- Ranellucci, J., Hall, N. C. & Goetz, T. (2015). Achievement goals, emotions, learning, and performance: A process model. *Motivation Science*, 1 (2), 98–120. doi:\url{10.1037/mot0000014}
- Reeve, J. & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257–267. doi:10.1016/j.cedpsych.2011.05.002
- Regehr, C., Glancy, D. & Pitts, A. (2013). Interventions to reduce stress in university students: a review and meta-analysis. *Journal of affective disorders*, 148(1), 1–11. doi:10.1016/j.jad.2012.11.026
- Respondek, L., Seufert, T., Stupnisky, R. & Nett, U. E. (2017). Perceived Academic Control and Academic Emotions Predict Undergraduate University Student Success: Examining Effects on Dropout Intention and Achievement. *Frontiers in psychology*, 8, 1–18. PMC5339277 Journal Article. doi:\url{10.3389/fpsyg.2017.00243}. eprint: 28326043
- Revelle, W. (2018). psych: Procedures for Personality and Psychological Research: Version = 1.8.12. Northwestern University, Evanston, Illinois, USA. Zugriff unter <https://CRAN.R-project.org/package=psych>

- Richardson, M., Abraham, C. & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387. doi:[10.1037/a0026838](https://doi.org/10.1037/a0026838)
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R. & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130, 261–288. doi:[10.1037/0033-2909.130.2.261](https://doi.org/10.1037/0033-2909.130.2.261)
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48 (2), 1–36.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F. & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118–137. doi:[10.1016/j.intell.2015.09.002](https://doi.org/10.1016/j.intell.2015.09.002)
- Rounds, J. & Su, R. (2014). The Nature and Power of Interests. *Current Directions in Psychological Science*, 23(2), 98–103. doi:[10.1177/0963721414522812](https://doi.org/10.1177/0963721414522812)
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63(3), 581–592.
- Rump, M., Esdar, W. & Wild, E. (2017). Individual differences in the effects of academic motivation on higher education students' intention to drop out. *European Journal of Higher Education*, 7(4), 341–355. doi:[10.1080/21568235.2017.1357481](https://doi.org/10.1080/21568235.2017.1357481)
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F. & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *The Journal of applied psychology*, 88(6), 1068–1081. Journal Article Meta-Analysis Research Support, Non-U.S. Gov't. doi:[10.1037/0021-9010.88.6.1068](https://doi.org/10.1037/0021-9010.88.6.1068). eprint: [14640817](https://doi.org/10.1037/0021-9010.88.6.1068)
- Samejima, F. (1969). *Estimation of Latent Ability Using a Response Pattern of Graded Scores* (Psychometric Monograph No. 17). Richmond, VA.
- Saracetti, A. & Müller, S. (2011). Zum Stand der Studienabbruchforschung. Theoretische Perspektiven, zentrale Ergebnisse und methodische Anforderungen an künftige Studien. *Zeitschrift für Bildungsforschung*, 1, 235–248. doi:[10.1007/s35834-011-0020-2](https://doi.org/10.1007/s35834-011-0020-2)
- Satorra, A. & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514. doi:[10.1007/BF02296192](https://doi.org/10.1007/BF02296192)
- Satorra, A. & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*, 75(2), 243–248. doi:[10.1007/s11336-009-9135-y](https://doi.org/10.1007/s11336-009-9135-y)
- Satow, L. (2012). Stress- und Coping-Inventar (SCI) - Skalendokumentation: Test- und Skalendokumentation. Zugriff unter [%5Curl%7Bhttp://www.drsatow.de.%7D](http://www.drsatow.de)
- Sax, L. J., Gilmartin, S. K. & Bryant, A. N. (2003). Assessing Response Rates and Nonresponse Bias in Web and Paper Surveys. *Research in Higher Education*, 44 (4), 409–432. doi:[10.1023/A:1024232915870](https://doi.org/10.1023/A:1024232915870)
- Schaeper, H. (2019). The first year in higher education: the role of individual factors and the learning environment for academic integration. *Higher Education*, 83 (3), 459. PII: 398. doi:[10.1007/s10734-019-00398-0](https://doi.org/10.1007/s10734-019-00398-0)

- Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, 8 (1), 3–15. Journal Article Research Support, U.S. Gov't, P.H.S. Review. doi:[10.1177/096228029900800102](#). eprint: [10347857](#)
- Schiefele, U. & Jacob-Ebbinghaus, L. (2006). Lernermerkmale und Lehrqualität als Bedingungen der Studienzufriedenheit. *Zeitschrift für Pädagogische Psychologie*, 20 (3), 199–212. doi:[\url{10.1024/1010-0652.20.3.199}](#)
- Schiefele, U., Kapp, A., Wild, K.-P. & Winteler, A. (1993). Fragebogen zum Studieninteresse (FSI). *Diagnostica*, 39 (4), 335–351.
- Schmidt-Atzert, L. (2005). Prädiktion von Studienerfolg bei Psychologiestudenten. *Psychologische Rundschau*, 56(2), 131–133. doi:[10.1026/0033-3042.56.2.131](#)
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik*. Berlin Heidelberg: Springer.
- Schmitt, N. & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18 (4), 210–222. PII: S1053482208000053. doi:[10.1016/j.hrmr.2008.03.003](#)
- Schneider, K., Berens, J. & Burghoff, J. (2019). Drohende Studienabbrüche durch Frühwarnsysteme erkennen: Welche Informationen sind relevant? *Zeitschrift für Erziehungswissenschaft*, 22(5), 1121–1146. doi:[10.1007/s11618-019-00912-1](#)
- Schneider, M. & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, 143 (6), 565–600. Journal Article Meta-Analysis Review Systematic Review. doi:[10.1037/bul0000098](#). eprint: [28333495](#)
- Schulz, U. & Schwarzer, R. (2003). Soziale Unterstützung bei der Krankheitsbewältigung: Die Berliner Social Support Skalen (BSSS). *Diagnostica*, 49, 73–82. doi:[\url{10.1026//0012-1924.49.2.73}](#)
- Schwager, I. T., Hülshager, U. R., Lang, J. W., Klieger, D. M., Bridgeman, B. & Wendler, C. (2014). Supervisor ratings of students' academic potential as predictors of citizenship and counterproductive behavior. *Learning and Individual Differences*, 35, 62–69. doi:[10.1016/j.lindif.2014.07.005](#)
- Schwarzer, R. & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin.
- Simmons, J. P., Nelson, L. D. & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22, 1359–1366. Journal Article. doi:[\url{10.1177/0956797611417632}](#). eprint: [22006061](#)
- Spady, W. G. (1971). Dropouts from higher education: Toward an empirical model. *Interchange*, 2, 38–62.

- Stack-Cutler, H. L., Parrila, R. K. & Torppa, M. (2015). Using a Multidimensional Measure of Resilience to Explain Life Satisfaction and Academic Achievement of Adults With Reading Difficulties. *Journal of learning disabilities*, 48(6), 646–657. Journal Article Research Support, Non-U.S. Gov't. doi:[10.1177/0022219414522705](https://doi.org/10.1177/0022219414522705). eprint: [24596110](https://doi.org/24596110)
- Statistisches Bundesamt. (2016). *Erfolgsquoten 2014 - Berechnung für die Studienanfängerjahrgänge 2002 bis 2006*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2018). Anzahl der Studienanfänger/-innen im ersten Hochschulse semester in Deutschland im Studienjahr 2017/2018* nach Fächergruppen. Zugriff 20. April 2018 unter [%5Curl%7Bhttps://de.statista.com/statistik/daten/studie/3247/umfrage/anzahl-der-studienanfaenger-nach-faechergruppen/%7D](https://de.statista.com/statistik/daten/studie/3247/umfrage/anzahl-der-studienanfaenger-nach-faechergruppen/%7D)
- Statistisches Bundesamt. (2019). Studienanfänger/-innen im 1. Hochschulse semester nach Fächergruppen. Zugriff unter <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/studierende-erstes-hs-faechergruppen.html>
- Stifterverband für die Deutsche Wissenschaft. (2015). *Handbuch Studienerfolg: Strategien und Maßnahmen: Wie Hochschulen Studierende erfolgreich zum Abschluss führen*. Positionen. Essen: Ed. Stifterverband.
- Stifterverband für die Deutsche Wissenschaft & McKinsey & Company. (2017). *Höhere Chancen durch höhere Bildung? Jahresbericht 2017/2018 - Halbzeitbilanz 2010 bis 2015* (Stand 30. Juni 2017). Hochschul-Bildungs-Report 2020. Essen: Edition Stifterverband.
- Stoll, G. (2013). *Entwicklung und Validierung eines Interessentests zur Berufs- und Studienfachwahl: Dissertation*. Saarbrücken.
- Thomas, L. (2002). Student Retention in Higher Education: The Role of Institutional Habitus. *Journal of Education Policy*, (17 (4)).
- Thomas, L. & Hovdhaugen, E. (2014). Complexities and Challenges of Researching Student Completion and Non-completion of HE Programmes in Europe: A comparative analysis between England and Norway. *European Journal of Education*, 49, 457–470. doi:[10.1111/ejed.12093](https://doi.org/10.1111/ejed.12093)
- Tieben, N. (2016). LEAD-Expertise - Studienverlauf, Verbleib und Berufsstatus von Studienabbrecherinnen und Studienabbrechern: Ergänzende Informationen zum Datenreport zum Berufsbildungsbericht 2016. doi:[10.5157/NEPS:SC6:1.0.0](https://doi.org/10.5157/NEPS:SC6:1.0.0)
- Tieben, N. (2019). Non-completion, Transfer, and Dropout of Traditional and Non-traditional Students in Germany. *Research in Higher Education*, 5(3), 231. doi:[10.1007/s11162-019-09553-z](https://doi.org/10.1007/s11162-019-09553-z)
- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 45, 89–125.
- Tinto, V. (1987). *Leaving college: rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.
- Tinto, V. (1997). Classrooms as Communities. *Journal of Higher Education*, 68, 599–623.

- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21, 11–27. doi:\url{10.1024/1010-0652.21.1.11}
- Trautwein, C. & Bosse, E. (2017). The first year in higher education—critical requirements from the student perspective. *Higher Education*, 73 (3), 371–387. doi:\url{10.1007/s10734-016-0098-5}
- Treppesch, C. & Ehrlich, S. (2018). *Studieneingangsbefragung WS 2018/19*. Servicestelle Lehr-evaluation der Justus-Liebig-Universität Gießen. Unveröffentlichte Rohdaten.
- Ulfert, A.-S., Ott, M., Bothe, P. S. & Kersting, M. (2017). Gießener kognitiver Kompetenztest (GkKT). Verfahrenshinweise. Version 2.0. Unveröffentlichtes Manuskript. Gießen.
- Urdan, T. & Pajares, F. (2006). *Self-efficacy beliefs of adolescents*. Greenwich, Connecticut: IAP.
- van Dinther, M., Dochy, F. & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, 6 (2), 95–108. doi:\url{10.1016/j.edurev.2010.10.003}
- Vandenberg, R. J. & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 3 (1), 4–70. doi:10.1177/109442810031002
- von Maurer, M. & Jandura, O. (2009). Masse statt Klasse? Einige kritische Anmerkungen zu Repräsentativität und Validität von Online-Befragungen. In N. Jakob, H. Schoen & T. Zerback (Hrsg.), *Sozialforschung im Internet* (Bd. 11, S. 61–73). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden. doi:10.1007/978-3-531-91791-7_4
- Vossensteyn, J. J., Kottmann, A., Jongbloed, B. W. A., Kaiser, F., Cremonini, L., Stensaker, B., ... Wollscheid, S. (2015). *Dropout and Completion in Higher Education in Europe: main report*. doi:\url{10.2766/826962}
- Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 17(2), 228–243. doi:10.1037/a0027127
- Weber, A., Daniel, A., Becker, K. & Bornkessel, P. (2018). Proximale Prädiktoren objektiver wie subjektiver Studienerfolgsindikatoren. In P. Bornkessel (Hrsg.), *Erfolg im Studium. Konzeptionen, Befunde und Desiderate* (S. 59–104).
- Westermann, R., Heise, E., Spies, K. & Trautwein, U. (1996). Identifikation und Erfassung von Komponenten der Studienzufriedenheit. *Psychologie, Erziehung, Unterricht*, 43, 1–22.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wigfield, A. & Eccles, J. S. (2000). Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25(1), 68–81. Journal Article. doi:\url{10.1006/ceps.1999.1015}. eprint: 10620382

- Wissenschaftsrat. (2012). Prüfungsnoten an Hochschulen im Prüfungsjahr 2010. Arbeitsbericht mit einem Wissenschaftspolitischen Kommentar des Wissenschaftsrates. Wissenschaftsrat. Hamburg.
- Wittmann, W. W. (1990). Brunswik-Symmetrie und die Konzeption der Fünf-Datenboxen—Ein Rahmenkonzept für umfassende Evaluationsforschung. *Zeitschrift für Pädagogische Psychologie*, 4, 241–251.
- Zettler, I. (2011). Self-control and academic performance: Two field studies on university citizenship behavior and counterproductive academic behavior. *Learning and Individual Differences*, 21(1), 119–123. doi:[10.1016/j.lindif.2010.11.002](https://doi.org/10.1016/j.lindif.2010.11.002)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt.

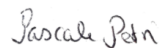
Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde.

Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren.

Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.

Gießen, 16.12.2019

Ort, Datum



Unterschrift

13 Anhang

13.1 Skalenentwicklung

Tabelle 13.1: Normtabelle zur SESW-Skala ($N = 1842$ Studienanfänger 2018)

Rohwert	T-Wert	Rohwert	T-Wert	Rohwert	T-Wert
1.00	1	3.15	35	4.92	68
1.00	2	3.22	36	4.96	69
1.00	3	3.28	37	4.99	70
1.00	4	3.35	38	5.00	71
1.00	5	3.41	39	5.00	72
1.00	6	3.48	40	5.00	73
1.00	7	3.54	41	5.00	74
1.04	8	3.60	42	5.00	75
1.13	9	3.66	43	5.00	76
1.22	10	3.72	44	5.00	77
1.31	11	3.78	45	5.00	78
1.39	12	3.84	46	5.00	79
1.48	13	3.90	47	5.00	80
1.56	14	3.95	48	5.00	81
1.65	15	4.01	49	5.00	82
1.73	16	4.07	50	5.00	83
1.81	17	4.12	51	5.00	84
1.89	18	4.17	52	5.00	85
1.97	19	4.23	53	5.00	86
2.05	20	4.28	54	5.00	87
2.13	21	4.33	55	5.00	88
2.21	22	4.38	56	5.00	89
2.29	23	4.43	57	5.00	90
2.37	24	4.48	58	5.00	91
2.44	25	4.52	59	5.00	92
2.52	26	4.57	60	5.00	93
2.59	27	4.62	61	5.00	94
2.66	28	4.66	62	5.00	95
2.74	29	4.71	63	5.00	96
2.81	30	4.75	64	5.00	97
2.88	31	4.79	65	5.00	98
2.95	32	4.83	66	5.00	99
3.02	33	4.88	67	5.00	100
3.08	34				

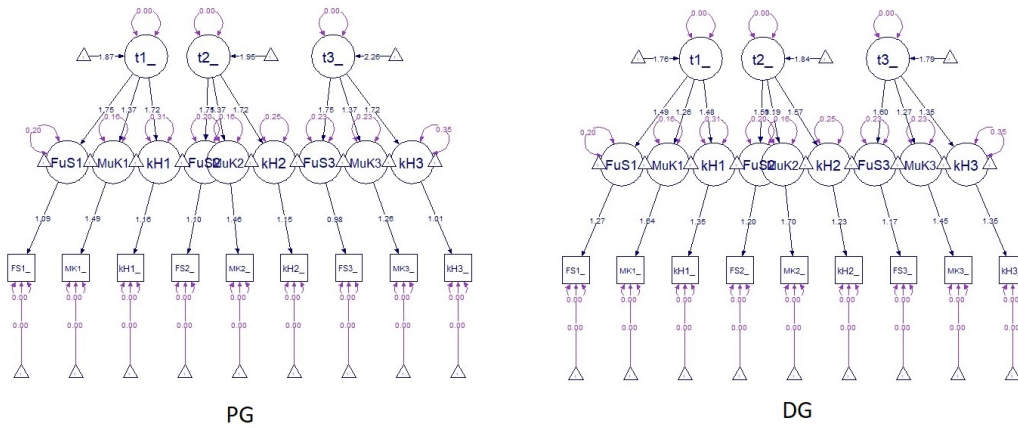
Abbildung 13.1: SESW-Entwicklung im Gruppenvergleich: *Latent Growth Analysis*

Tabelle 13.2: Korrelationen nullter Ordnung der dreizehn SESW-Items untereinander

	1	2	3	4	5	6	7	8	9	10	11	12
SESW 01												
SESW 02	-0.37											
SESW 03	-0.41	0.69**										
SESW 04	-0.20	0.92****	0.53									
SESW 05	0.36	0.30	0.32	0.34								
SESW 06	0.11	0.36	0.52	0.30	0.86***							
SESW 07	-0.09	0.79**	0.27	0.90****	0.26	0.21						
SESW 08	-0.14	-0.10	0.50	-0.16	0.24	0.39	-0.24					
SESW 09	-0.14	-0.36	-0.06	-0.42	0.36	0.27	-0.34	0.45				
SESW 10	0.15	-0.42	0.12	-0.45	0.47	0.48	-0.42	0.78**	0.78**			
SESW 11	0.12	-0.12	0.27	-0.09	0.63*	0.68*	-0.07	0.68*	0.60*	0.84***		
SESW 12	0.01	-0.13	0.38	-0.13	0.50	0.63*	-0.14	0.81***	0.55*	0.86***	0.97****	
SESW 13	-0.27	0.46	0.76**	0.41	0.44	0.68*	0.28	0.57*	0.07	0.35	0.68*	0.74**

* $p < .05$, ** $p < .01$, *** $p < .001$

Tabelle 13.3: Korrelationen nullter Ordnung zwischen den SESW-Skalenwerten für die PG, erhoben zu den drei Messzeitpunkten t1, t2, t3

	SESW t1	SESW t2	SESW t3
t1	1		
t2	0.36***	1	
t3	0.29***	0.44***	1

**** $p < .001$

Tabelle 13.4: Korrelationen nullter Ordnung zwischen den SESW-Skalenwerten für die DG, erhoben zu den drei Messzeitpunkten t1, t2, t3

	SESW t1	SESW t2	SESW t3
t1	1		
t2	0.53***	1	
t3	0.37***	0.49***	1

*** $p < .001$

13.2 Zu Studie 1

Tabelle 13.5: Korrelationen nullter Ordnung für alle in Spadys Modell enthaltenen Variablen (Jahrgang 1)

	1	2	3	4	5	6	7	8
Bildungsn. Mutter								
Schulleistung	-0.11							
soziale Unterstützung	0.15	0.04						
P-U-Fit	-0.15	-0.10	0.12					
Zufriedenheit (t3)	0.03	0.05	-0.04	-0.06				
Abbruchintention	-0.02	-0.05	-0.02	0.08	-0.34****			
akad. Leistung	-0.10	0.10	0.05	0.10	0.15	-0.13		
soziale Integration (t3)	-0.21*	-0.06	-0.04	0.16	0.10	-0.07	0.07	
inst. Commitment	-0.11	-0.06	-0.11	0.05	0.28***	-0.19*	0.23**	0.09

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.6: Korrelationen nullter Ordnung für alle in Tintos Modell enthaltenen Variablen (Jahrgang 1)

	1	2	3	4	5	6	7	8
Bildungsn. Mutter								
Schulleistung	-0.11							
soziale Integration	-0.01	-0.10						
Informiertheit	0.01	-0.23**	0.02					
akad. Engagement	0.12	-0.01	0.16	0.03				
Abbruchintention	-0.01	-0.05	-0.02	0.20*	0.00			
akad. Leistung	-0.10	0.10	-0.04	0.03	0.10	-0.12		
akad. Integration	-0.09	0.05	-0.13	0.02	-0.07	-0.25**	0.09	
inst. Commitment	-0.10	-0.06	-0.04	-0.02	0.23**	-0.19*	0.24**	0.21*

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.7: Korrelationen nullter Ordnung für alle in Neuville et al.s Modell enthaltenen Variablen (Jahrgang 1)

	1	2	3	4	5	6	7	8
Wert								
Bildungsn. Mutter	0.06							
Schulleistung	0.10	-0.10						
SESW (t1)	-0.01	-0.03	-0.21*					
Informiertheit	-0.11	0.00	-0.23**	-0.07				
SESW (t2)	0.50****	0.07	0.01	-0.10	-0.13			
akad. Engagement	0.04	0.13	-0.02	0.02	0.01	0.03		
Abbruchintention	0.07	-0.02	-0.06	-0.05	0.18*	-0.12	-0.03	
akad. Leistung	-0.17*	-0.08	0.10	-0.06	0.05	0.01	0.11	-0.10

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.8: Korrelationen nullter Ordnung für alle im CSM Modell enthaltenen Variablen (Jahrgang 1)

	1	2	3	4	5	6	7	8
Wert								
Bildungsn. Mutter	0.06							
Schulleistung	0.10	-0.10						
SESW (t1)	-0.01	-0.04	-0.21*					
Zielsetzung	-0.08	-0.03	-0.22**	0.72****				
soziale Unterstützung	-0.05	0.14	0.05	0.23**	0.27**			
Ergebniserwartung	0.02	0.18*	-0.17*	0.09	0.19*	0.28***		
Abbruchintention	0.07	-0.03	-0.06	-0.06	-0.07	-0.02	0.02	
akad. Leistung	-0.17*	-0.08	0.10	-0.06	0.01	0.06	0.12	-0.11

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.9: Korrelationen nullter Ordnung für alle im EOS-Modell enthaltenen Variablen (Jahrgang 1)

	1	2	3	4	5	6	7	8	9
Schulleistung									
SESW (t1)	-0.19*								
Zielsetzung	-0.22**	0.72****							
soziale Unterstützung	0.05	0.18*	0.24**						
Stress	0.00	0.06	0.04	-0.16					
Zufriedenheit (t2)	-0.04	-0.16	-0.06	-0.01	-0.58****				
SESW (t3)	-0.01	-0.05	-0.05	-0.05	-0.06	0.21*			
UCB	-0.05	0.03	0.02	-0.06	0.02	0.06	0.29***		
akad. Leistung	0.10	-0.06	0.01	0.05	-0.10	0.13	0.28***	0.24**	
Abbruchintention	-0.06	-0.06	-0.07	-0.03	0.00	-0.04	-0.35****	-0.24**	-0.11

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

13.3 Zu Studie 2

Tabelle 13.10: Korrelationen nullter Ordnung für alle in Spadys Modell enthaltenen Variablen (Jahrgang 2)

	1	2	3	4	5	6	7	8
Bildungsn. Mutter								
Schulleistung	-0.29**							
soziale Unterstützung	0.13	0.03						
P-U-Fit	-0.17	0.06	0.14					
Zufriedenheit (t3)	0.00	-0.03	0.03	0.38****				
Abbruchintention	0.20*	0.08	-0.11	-0.37****	-0.48****			
akad. Leistung	-0.04	-0.21*	0.02	0.09	0.30**	-0.29**		
soziale Integriertheit (t3)	-0.01	-0.03	0.05	0.20*	0.29**	-0.22*	0.16	
inst. Commitment	-0.06	-0.06	-0.01	0.12	0.28**	-0.42****	0.32****	0.11

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.11: Korrelationen nullter Ordnung für alle in Tintos Modell enthaltenen Variablen (Jahrgang 2)

	1	2	3	4	5	6	7	8
Bildungsn. Mutter								
Abiturnote	-0.29**							
soziale Integration (t1)	0.07	-0.03						
Informiertheit	-0.13	-0.03	-0.13					
akad. Engagement	0.00	0.05	-0.08	0.16				
Abbruchintention	0.20*	0.08	-0.18	-0.09	-0.16			
akad. Leistung	-0.04	-0.21*	-0.11	0.12	0.08	-0.29**		
akad. Integration	0.01	-0.02	0.19*	-0.08	0.05	-0.19*	0.10	
inst. Commitment	-0.06	-0.06	0.00	0.11	0.26**	-0.42***	0.32***	0.19*

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.12: Korrelationen nullter Ordnung für alle in Neuville et al.s Modell enthaltenen Variablen (Jahrgang 2)

	1	2	3	4	5	6	7	8
Wert								
Bildungsn. Mutter	-0.06							
Schulleistung	0.03	-0.25**						
SESW (t1)	0.23*	0.02	-0.12					
Informiertheit	-0.07	-0.23*	0.00	-0.07				
SESW (t2)	0.38****	0.08	-0.33****	0.64****	-0.03			
akad. Engagement	0.24*	-0.03	0.03	0.04	0.12	0.13		
Abbruchintention	-0.44****	0.25**	0.07	-0.24*	-0.09	-0.32****	-0.20*	
akad. Leistung	0.31**	-0.03	-0.22*	0.05	0.15	0.25**	0.09	-0.28**

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.13: Korrelationen nullter Ordnung für alle im CSM Modell enthaltenen Variablen (Jahrgang 2)

	1	2	3	4	5	6	7	8
Wert								
Bildungsn. Mutter	-0.06							
Schulleistung	0.03	-0.25**						
SES (t1)	0.23*	0.02	-0.12					
Zielsetzung	0.21*	-0.10	-0.02	0.30**				
soz. Unterstützung	-0.05	0.12	-0.02	0.10	0.07			
Ergebniserwartung	0.18	-0.15	0.03	0.22*	0.34***	0.11		
Abbruchintention	-0.44****	0.25**	0.07	-0.24*	-0.13	-0.10	-0.20*	
akad. Leistung	0.31**	-0.03	-0.22*	0.05	0.17	0.07	0.11	-0.28**

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.14: Korrelationen nullter Ordnung für alle im EOS-Modell enthaltenen Variablen (Jahrgang 2)

	1	2	3	4	5	6	7	8	9
Schulleistung									
SESW (t1)	-0.12								
Zielsetzung	-0.02	0.30**							
soziale Unterstützung	-0.04	0.15	0.10						
Stress	0.10	-0.29**	-0.02	-0.18					
Zufriedenheit (t2)	-0.07	0.21*	-0.01	0.10	-0.62****				
SESW (t3)	-0.17	0.50****	0.31**	0.27**	-0.32****	0.34****			
UCB	-0.01	0.20*	0.21*	0.10	-0.23*	0.17	0.54****		
akad. Leistung	-0.22*	0.05	0.17	0.13	-0.12	0.14	0.39****	0.30**	
Abbruchintention	0.07	-0.24*	-0.13	-0.09	0.24*	-0.25**	-0.49****	-0.45****	-0.28**

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

13.4 Zu Studie 3

Tabelle 13.15: Korrelationen nullter Ordnung für alle in Spadys Modell enthaltenen Variablen (Jahrgang 3)

	1	2	3	4	5	6	7	8
Bildungsn. Mutter								
Schulleistung	-0.12							
soziale Unterstützung	0.01	-0.03						
P-UF-it	-0.04	-0.14	0.27***					
Zufriedenheit (t3)	-0.06	-0.08	0.16*	0.37***				
Abbruchintention	0.09	0.02	-0.21**	-0.32***	-0.34***			
akad. Leistung	0.07	-0.26***	0.06	0.11	0.15	-0.14		
soziale Integration(t3)	0.23**	-0.05	0.19*	0.15	0.26***	-0.24**	0.10	
inst. Commitment	-0.02	-0.19*	-0.01	0.26***	0.32***	-0.28***	0.14	0.29***

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.16: Korrelationen nullter Ordnung für alle in Tintos Modell enthaltenen Variablen (Jahrgang 3)

	1	2	3	4	5	6	7	8
Bildungsn. Mutter								
Schulleistung	-0.12							
Integration (t1)	-0.05	-0.09						
Informiertheit	0.05	-0.17*	0.08					
akad. Engagement	0.06	-0.18*	-0.08	0.04				
Abbruchintention	0.09	0.02	-0.20*	0.10	-0.04			
akad. Leistung	0.08	-0.26***	0.03	0.02	0.00	-0.14		
akad. Integration	0.00	0.04	0.08	0.01	0.01	-0.20*	0.09	
inst. Commitment	-0.01	-0.19*	0.17*	0.19*	0.13	-0.28***	0.13	0.22**

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.17: Korrelationen nullter Ordnung für alle in Neuville et al.'s Modell enthaltenen Variablen (Jahrgang 3)

	1	2	3	4	5	6	7	8
Bildung.Mutter								
Schulleistung	-0.19*							
SESW (t1)	-0.05	-0.16						
Informiertheit	0.01	-0.20*	-0.04					
SESW (t2)	-0.05	-0.19*	0.72****	-0.05				
akad. Engagement	0.07	-0.19*	0.04	0.07	0.10			
Abbruchintention	0.08	0.05	-0.23**	0.08	-0.32****	-0.04		
akad. Leistung	0.07	-0.30***	0.17*	-0.01	0.20*	0.04	-0.16	
Wert	0.00	-0.17*	0.37****	-0.10	0.49****	0.13	-0.43****	0.09

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.18: Korrelationen nullter Ordnung für alle im CSM-Modell enthaltenen Variablen (Jahrgang 3)

	1	2	3	4	5	6	7	8
Bildungsn. Mutter								
Schulleistung	-0.19*							
SES (t1)	-0.06	-0.16						
Zielsetzung	-0.09	-0.16	0.80****					
soziale Unterstützung	0.00	-0.05	0.36****	0.23**				
Ergebniserwartung	0.01	-0.20*	0.27**	0.30***	0.20*			
Abbruchintention	0.08	0.05	-0.22**	-0.20*	-0.18*	-0.13		
akad. Leistung	0.06	-0.30***	0.18*	0.12	0.03	0.07	-0.15	
Wert	-0.01	-0.17*	0.39****	0.42****	0.28****	0.34****	-0.42****	0.10

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Tabelle 13.19: Korrelationen nullter Ordnung für alle im EOS-Modell enthaltenen Variablen (Jahrgang 3)

	1	2	3	4	5	6	7	8	9
Abiturnote									
SESW (t1)	-0.16								
Zielsetzung	-0.16	0.80****							
soziale Unterstützung	-0.04	0.34****	0.20*						
Stress	0.11	-0.45****	-0.38****	-0.19*					
Zufriedenheit	0.01	0.34****	0.24**	0.18*	-0.64****				
SESW (t3)	-0.22**	0.60****	0.55****	0.30***	-0.43****	0.38****			
UCB	-0.20*	0.22**	0.27***	0.04	-0.22**	0.21*	0.46****		
akt.Note	-0.30***	0.18*	0.12	0.03	-0.05	0.20*	0.25**	0.20*	
Abbruchvorhaben	0.05	-0.22**	-0.20*	-0.16	0.18*	-0.27**	-0.47****	-0.33****	-0.15

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

13.5 Zusammenschau

Tabelle 13.20: Spadys Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1

Pfad	Koeffizient	SE	p
Abiturnote $\leftarrow$ Bildungsn. Mutter	-0.05475	0.08236	0.5062
Person-University-Fit $\leftarrow$ Bildungsn. Mutter	-0.13045	0.08107	0.1076
akad. Leistung $\leftarrow$ Abiturnote	0.09332	0.08086	0.2485
intell. Entwicklung $\leftarrow$ Abiturnote	0.02583	0.08265	0.7546
akad. Leistung $\leftarrow$ Person-University-Fit	0.15940	0.08043	0.0475
intell. Entwicklung $\leftarrow$ Person-University-Fit	-0.02577	0.08265	0.7552
soz. Integration $\leftarrow$ Person-University-Fit	0.12760	0.08179	0.1187
soz. Unterstützung $\leftarrow$ Person-University-Fit	0.08776	0.08288	0.2897
akad. Leistung $\leftarrow$ soz. Unterstützung	0.03023	0.08133	0.7102
soz. Integration $\leftarrow$ soz. Unterstützung	-0.07890	0.08199	0.3359
soz. Integration $\leftarrow$ akad. Leistung	0.11443	0.08392	0.1727
soz. Integration $\leftarrow$ intell. Entwicklung	-0.01982	0.08424	0.8140
Zufriedenheit $\leftarrow$ soz. Integration	0.14365	0.08081	0.0755
inst. Commitment $\leftarrow$ Zufriedenheit	0.31600	0.07424	<.0001
Abbruchintention $\leftarrow$ inst. Commitment	-0.15063	0.08238	0.0675
Abbruchintention $\leftarrow$ akad. Leistung	-0.13302	0.08270	0.1077

Tabelle 13.21: Spadys Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2

Pfad	Koeffizient	SE	p
Abiturnote $\leftarrow$ Bildungsn. Mutter	-0.22264	0.08720	0.0107
Person-University-Fit $\leftarrow$ Bildungsn. Mutter	-0.12700	0.08944	0.1556
akad. Leistung $\leftarrow$ Abiturnote	0.18121	0.08807	0.0369
intell. Entwicklung $\leftarrow$ Abiturnote	-0.09935	0.08667	0.2517
akad. Leistung $\leftarrow$ Person-University-Fit	0.08693	0.09006	0.3344
intell. Entwicklung $\leftarrow$ Person-University-Fit	0.28184	0.08343	0.0007
soz. Integration $\leftarrow$ Person-University-Fit	0.25942	0.08873	0.0035
soz. Unterstützung $\leftarrow$ Person-University-Fit	0.16039	0.08857	0.0702
akad. Leistung $\leftarrow$ soz. Unterstützung	0.02162	0.08133	0.7102
soz. Integration $\leftarrow$ soz. Unterstützung	0.07991	0.08805	0.3641
soz. Integration $\leftarrow$ akad. Leistung	0.07179	0.08711	0.4099
soz. Integration $\leftarrow$ intell. Entwicklung	-0.03395	0.09071	0.7082
Zufriedenheit $\leftarrow$ soz. Integration	0.23309	0.08599	0.0067
inst. Commitment $\leftarrow$ Zufriedenheit	0.39073	0.07703	<.0001
Abbruchintention $\leftarrow$ inst. Commitment	-0.46802	0.07505	<.0001
Abbruchintention $\leftarrow$ akad. Leistung	-0.11805	0.08522	0.1660

Tabelle 13.22: Spadys Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3

Pfad	Koeffizient	SE	p
Abiturnote $\leftarrow$ Bildungsn. Mutter	-0.07292	0.07815	0.3508
Person-University-Fit $\leftarrow$ Bildungsn. Mutter	0.01913	0.07854	0.8075
akad. Leistung $\leftarrow$ Abiturnote	-0.21644	0.07559	0.0042
intell. Entwicklung $\leftarrow$ Abiturnote	-0.07686	0.07659	0.3156
akad. Leistung $\leftarrow$ Person-University-Fit	-0.01887	0.08038	0.8144
intell. Entwicklung $\leftarrow$ Person-University-Fit	0.25838	0.07440	0.0005
soz. Integration $\leftarrow$ Person-University-Fit	0.13685	0.08034	0.0885
soz. Unterstützung $\leftarrow$ Person-University-Fit	0.27963	0.07242	0.0001
akad. Leistung $\leftarrow$ soz. Unterstützung	0.06707	0.07964	0.3997
soz. Integration $\leftarrow$ soz. Unterstützung	0.18408	0.07750	0.0175
soz. Integration $\leftarrow$ akad. Leistung	0.10329	0.07533	0.1703
soz. Integration $\leftarrow$ intell. Entwicklung	0.02770	0.07903	0.7259
Zufriedenheit $\leftarrow$ soz. Integration	0.29486	0.07168	0.0001
inst. Commitment $\leftarrow$ Zufriedenheit	0.34179	0.06938	<.0001
Abbruchintention $\leftarrow$ inst. Commitment	-0.37043	0.06814	<.0001
Abbruchintention $\leftarrow$ akad. Leistung	-0.01923	0.07336	0.7933

Tabelle 13.23: Tintos Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1

Pfad	Koeffizient	SE	p
soziale Integration $\leftarrow$ Bildungn. Mutter	-0.0587	0.08224	0.4754
soziale Integration $\leftarrow$ Schulleistung	-0.08923	0.08468	0.2920
soziale Integration $\leftarrow$ Informiertheit	0.01883	0.08468	0.8241
akad. Integration $\leftarrow$ Bildungn. Mutter	-0.06364	0.08245	0.4402
akad. Integration $\leftarrow$ Schulleistung	0.03506	0.08521	0.6807
akad. Integration $\leftarrow$ Informiertheit	0.03814	0.08488	0.6532
akad. Integration $\leftarrow$ soziale Integration	-0.11589	0.08091	0.1520
Abbruchintention $\leftarrow$ soziale Integration	-0.00559	0.08198	0.9457
Abbruchintention $\leftarrow$ akad. Integration	-0.1783	0.08068	0.0271
inst. Commitment $\leftarrow$ soziale Integration	-0.02914	0.08152	0.7207
inst. Commitment $\leftarrow$ akad. Integration	0.19866	0.07995	0.013
akad. Engagement $\leftarrow$ soziale Integration	0.13433	0.08154	0.0995
akad. Engagement $\leftarrow$ akad. Integration	-0.06779	0.08209	0.4089
akad. Leistung $\leftarrow$ Bildungn. Mutter	-0.10495	0.08091	0.1946
akad. Leistung $\leftarrow$ Schulleistung	0.07696	0.08005	0.3364
akad. Leistung $\leftarrow$ Abbruchintention	-0.12262	0.08111	0.1306
akad. Leistung $\leftarrow$ akad. Engagement	-0.04793	0.08152	0.5566
akad. Leistung $\leftarrow$ inst. Commitment	0.20971	0.08168	0.0102

Tabelle 13.24: Tintos Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2

Pfad	Koeffizient	SE	p
soziale Integration $\leftarrow$ Bildungsn. Mutter	0.05875	0.0933	0.5289
soziale Integration $\leftarrow$ Schulleistung	-0.06559	0.09375	0.4841
soziale Integration $\leftarrow$ Informiertheit	-0.13922	0.09013	0.1224
akad. Integration $\leftarrow$ Bildungsn. Mutter	-0.02885	0.09445	0.76
akad. Integration $\leftarrow$ Schulleistung	-0.05972	0.09605	0.5341
akad. Integration $\leftarrow$ Informiertheit	-0.08632	0.09164	0.3462
akad. Integration $\leftarrow$ soziale Integration	0.2060	0.0853	0.0157
Abbruchintention $\leftarrow$ soziale Integration	-0.2054	0.08683	0.018
Abbruchintention $\leftarrow$ akad. Integration	-0.19367	0.08703	0.0261
inst. Commitment $\leftarrow$ soziale Integration	-0.00909	0.09154	0.9209
inst. Commitment $\leftarrow$ akad. Integration	0.18951	0.08996	0.0351
akad. Engagement $\leftarrow$ soziale Integration	-0.12605	0.09178	0.1696
akad. Engagement $\leftarrow$ akad. Integration	0.01184	0.09247	0.8981
akad. Leistung $\leftarrow$ Bildungsn. Mutter	-0.12345	0.09164	0.1779
akad. Leistung $\leftarrow$ Schulleistung	-0.16427	0.08927	0.0657
akad. Leistung $\leftarrow$ Abbruchintention	-0.08361	0.10427	0.4226
akad. Leistung $\leftarrow$ akad. Engagement	-0.03312	0.08868	0.7088
akad. Leistung $\leftarrow$ inst. Commitment	0.27268	0.09823	0.0055

Tabelle 13.25: Tintos Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3

Pfad	Koeffizient	SE	p
soziale Integration $\leftarrow$ Bildungsn. Mutter	-0.02787	0.07824	0.7217
soziale Integration $\leftarrow$ Schulleistung	-0.10651	0.07893	0.1772
soziale Integration $\leftarrow$ Informiertheit	0.08783	0.07905	0.2665
akad. Integration $\leftarrow$ Bildungsn. Mutter	0.0250	0.07892	0.7514
akad. Integration $\leftarrow$ Schulleistung	0.06799	0.07987	0.3946
akad. Integration $\leftarrow$ Informiertheit	-0.01348	0.08003	0.8662
akad. Integration $\leftarrow$ soziale Integration	0.03399	0.07762	0.6614
Abbruchintention $\leftarrow$ soziale Integration	-0.19018	0.07425	0.0104
Abbruchintention $\leftarrow$ akad. Integration	-0.20771	0.07398	0.005
inst. Commitment $\leftarrow$ soziale Integration	0.22489	0.07199	0.0018
inst. Commitment $\leftarrow$ akad. Integration	0.2684	0.07117	0.0002
akad. Engagement $\leftarrow$ soziale Integration	-0.02583	0.07875	0.743
akad. Engagement $\leftarrow$ akad. Integration	0.03997	0.07874	0.6117
akad. Leistung $\leftarrow$ Bildungsn. Mutter	0.03334	0.0768	0.6642
akad. Leistung $\leftarrow$ Schulleistung	-0.21467	0.07681	0.0052
akad. Leistung $\leftarrow$ Abbruchintention	-0.01879	0.08247	0.8198
akad. Leistung $\leftarrow$ akad. Engagement	-0.07518	0.07809	0.3357
akad. Leistung $\leftarrow$ inst. Commitment	0.06874	0.08395	0.4129

Tabelle 13.26: Neuville et al.'s Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1

Pfad	Koeffizient	SE	p
SESW (t1) $\leftarrow$ Schulleistung	-0.20587	0.07854	0.0088
SESW (t2) $\leftarrow$ SESW (t1)	-0.14209	0.08354	0.089
SESW (t2) $\leftarrow$ Schulleistung	-0.01734	0.08294	0.8344
Wert $\leftarrow$ Informiertheit	-0.04669	0.07100	0.5108
Wert $\leftarrow$ SESW (t2)	0.52461	0.06572	<.0001
Abbruchintention $\leftarrow$ Wert	0.13016	0.09615	0.1758
akad. Engagement $\leftarrow$ Wert	0.0700	0.10004	0.4841
Abbruchintention $\leftarrow$ SESW (t2)	-0.12209	0.09631	0.205
akad. Engagement $\leftarrow$ SESW (t2)	-0.03521	0.10013	0.7251
akad. Leistung $\leftarrow$ Bildungsn. Mutter	-0.13603	0.07996	0.0889
akad. Leistung $\leftarrow$ Schulleistung	0.05682	0.0805	0.4803
akad. Leistung $\leftarrow$ Abbruchintention	-0.16093	0.07949	0.0429
akad. Leistung $\leftarrow$ akad. Engagement	-0.00411	0.08062	0.9594

Tabelle 13.27: Neuville et al.'s Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2

Pfad	Koeffizient	SE	p
SESW (t1) $\leftarrow$ Schulleistung	-0.10221	0.08723	0.2413
SESW (t2) $\leftarrow$ SESW (t1)	0.71316	0.04317	<.0001
SESW (t2) $\leftarrow$ Schulleistung	-0.21743	0.05981	0.0003
Wert $\leftarrow$ Informiertheit	-0.08189	0.08542	0.3377
Wert $\leftarrow$ SESW (t2)	0.35069	0.08015	<.0001
Abbruchintention $\leftarrow$ Wert	-0.23492	0.08004	0.0033
akad. Engagement $\leftarrow$ Wert	0.21246	0.09486	0.0251
Abbruchintention $\leftarrow$ SESW (t2)	-0.4160	0.07458	<.0001
akad. Engagement $\leftarrow$ SESW (t2)	-0.07121	0.09395	0.4485
akad. Leistung $\leftarrow$ Bildungsn. Mutter	-0.11266	0.09001	0.2107
akad. Leistung $\leftarrow$ Schulleistung	-0.17799	0.08649	0.0396
akad. Leistung $\leftarrow$ Abbruchintention	-0.2161	0.08654	0.0125
akad. Leistung $\leftarrow$ akad. Engagement	0.02507	0.08404	0.7655

Tabelle 13.28: Neuville et al.'s Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3

Pfad	Koeffizient	SE	p
SESW (t1) $\leftarrow$ Schulleistung	-0.1502	0.0741	0.0427
SESW (t2) $\leftarrow$ SESW (t1)	0.72356	0.03723	<.0001
SESW (t2) $\leftarrow$ Schulleistung	-0.07359	0.05482	0.1794
Wert $\leftarrow$ Informiertheit	-0.10188	0.06974	0.1441
Wert $\leftarrow$ SESW (t2)	0.48995	0.0593	<.0001
Abbruchintention $\leftarrow$ Wert	-0.3461	0.0768	<.0001
akad. Engagement $\leftarrow$ Wert	0.0700	0.08868	0.4299
Abbruchintention $\leftarrow$ SESW (t2)	-0.09462	0.08046	0.2396
akad. Engagement $\leftarrow$ SESW (t2)	0.06219	0.08957	0.4875
akademische Leistung $\leftarrow$ Bildungsn. Mutter	0.02802	0.07387	0.7044
akad. Leistung $\leftarrow$ Schulleistung	-0.22652	0.07282	0.0019
akad. Leistung $\leftarrow$ Abbruchintention	-0.04382	0.07372	0.5522
akad. Leistung $\leftarrow$ akad. Engagement	-0.0631	0.07486	0.3993

Tabelle 13.29: Modell nach Lent und Brown: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1

Pfad	Koeffizient	SE	p
SESW (t1) $\leftarrow$ Bildungsn. Mutter	-0.04154	0.08018	0.6044
SESW (t1) $\leftarrow$ Schulleistung	-0.2062	0.07862	0.0087
Ergebniserwartungen $\leftarrow$ Bildungsn. Mutter	0.14451	0.07896	0.0672
Ergebniserwartungen $\leftarrow$ Schulleistung	-0.13094	0.08106	0.1063
Ergebniserwartungen $\leftarrow$ SESW (t1)	0.10754	0.08095	0.184
Zielsetzung $\leftarrow$ SESW (t1)	0.74902	0.03603	<.0001
Zielsetzung $\leftarrow$ Ergebniserwartungen	0.09678	0.05528	0.0800
Zielsetzung $\leftarrow$ soziale Unterstützung	0.04782	0.05638	0.3963
Handlungen $\leftarrow$ soziale Unterstützung	-0.02666	0.08791	0.7617
Handlungen $\leftarrow$ SESW (t1)	-0.14187	0.12632	0.2614
Handlungen $\leftarrow$ Ergebniserwartungen	0.02716	0.08721	0.7555
Handlungen $\leftarrow$ Zielsetzung	0.02461	0.12773	0.8472
akad. Leistung $\leftarrow$ Handlungen	0.04524	0.08295	0.5855
Abbruchintention $\leftarrow$ Handlungen	-0.05519	0.08236	0.5028

Tabelle 13.30: Modell nach Lent und Brown: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2

Pfad	Koeffizient	SE	p
SESW (t1) $\leftarrow$ Bildungsn. Mutter	-0.02706	0.08941	0.7621
SESW (t1) $\leftarrow$ Schulleistung	-0.10927	0.08943	0.2218
Ergebniserwartungen $\leftarrow$ Bildungsn. Mutter	-0.15559	0.08562	0.0692
Ergebniserwartungen $\leftarrow$ Schulleistung	0.01411	0.08794	0.8725
Ergebniserwartungen $\leftarrow$ SESW (t1)	0.22375	0.08278	0.0069
Zielsetzung $\leftarrow$ SESW (t1)	0.24517	0.08088	0.0024
Zielsetzung $\leftarrow$ Ergebniserwartungen	0.26643	0.07897	0.0007
Zielsetzung $\leftarrow$ soziale Unterstützung	0.14569	0.08048	0.0702
Handlungen $\leftarrow$ soziale Unterstützung	0.12413	0.07226	0.0858
Handlungen $\leftarrow$ SESW (t1)	0.57559	0.06367	<.0001
Handlungen $\leftarrow$ Ergebniserwartungen	0.03652	0.0747	0.6249
Handlungen $\leftarrow$ Zielsetzung	0.08113	0.07591	0.2852
akad. Leistung $\leftarrow$ Handlungen	0.34112	0.07731	<.0001
Abbruchintention $\leftarrow$ Handlungen	-0.49819	0.06651	<.0001

Tabelle 13.31: Modell nach Lent und Brown: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3

Pfad	Koeffizient	SE	p
SESW (t1) $\leftarrow$ Bildungsn. Mutter	-0.05314	0.07518	0.4797
SESW (t1) $\leftarrow$ Schulleistung	-0.15437	0.07439	0.038
Ergebniserwartungen $\leftarrow$ Bildungsn. Mutter	-0.05392	0.073	0.4601
Ergebniserwartungen $\leftarrow$ Schulleistung	-0.11263	0.07345	0.1251
Ergebniserwartungen $\leftarrow$ SESW (t1)	0.24458	0.07156	0.0006
Zielsetzung $\leftarrow$ SESW (t1)	0.79229	0.02992	<.0001
Zielsetzung $\leftarrow$ Ergebniserwartungen	0.07177	0.04587	0.1176
Zielsetzung $\leftarrow$ soziale Unterstützung	-0.02535	0.04517	0.5747
Handlungen $\leftarrow$ soziale Unterstützung	0.10278	0.05512	0.0622
Handlungen $\leftarrow$ SESW (t1)	0.44303	0.08792	<.0001
Handlungen $\leftarrow$ Ergebniserwartungen	0.08718	0.05625	0.1212
Handlungen $\leftarrow$ Zielsetzung	0.28958	0.09112	0.0015
akad. Leistung $\leftarrow$ Handlungen	0.10193	0.0755	0.177
Abbruchintention $\leftarrow$ Handlungen	-0.3035	0.06848	<.0001

Tabelle 13.32: EOS-Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 1

Pfad	Koeffizient	SE	p
akad. Leistung $\leftarrow$ Schulleistung	0.09161	0.07609	0.2286
soziale Unterstützung $\leftarrow$ SESW (t1)	0.18213	0.07903	0.0212
Zielsetzung $\leftarrow$ SESW (t1)	0.76179	0.03427	<.0001
SESW (t3) $\leftarrow$ Zielsetzung	-0.04531	0.12315	0.7129
akad. Leistung $\leftarrow$ SESW (t3)	0.3409	0.07189	<.0001
Stress $\leftarrow$ soziale Unterstützung	-0.1745	0.08168	0.0326
UCB $\leftarrow$ SESW (t3)	0.28374	0.07514	0.0002
Studienzufriedenheit $\leftarrow$ Stress	-0.58718	0.05401	<.0001
Stress $\leftarrow$ SESW (t1)	0.0564	0.08186	0.4908
SESW (t3) $\leftarrow$ Studienzufriedenheit	0.23531	0.07906	0.0029
SESW (t3) $\leftarrow$ SESW (t1)	0.0374	0.12419	0.7633
Abbruchintention $\leftarrow$ UCB	-0.22126	0.07766	0.0044

Tabelle 13.33: EOS-Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 2

Pfad	Koeffizient	SE	p
akad. Leistung $\leftarrow$ Schulleistung	-0.12003	0.08368	0.1515
soziale Unterstützung $\leftarrow$ SESW (t1)	0.19872	0.08424	0.0183
Zielsetzung $\leftarrow$ SESW (t1)	0.32654	0.07836	<.0001
SESW (t3) $\leftarrow$ Zielsetzung	0.15525	0.06868	0.0238
akad. Leistung $\leftarrow$ SESW (t3)	0.32407	0.07976	<.0001
Stress $\leftarrow$ soziale Unterstützung	-0.2013	0.08291	0.0152
UCB $\leftarrow$ SESW (t3)	0.58216	0.05807	<.0001
Studienzufriedenheit $\leftarrow$ Stress	-0.64241	0.0535	<.0001
Stress $\leftarrow$ SESW (t1)	-0.3243	0.08022	<.0001
SESW (t3) $\leftarrow$ Studienzufriedenheit	0.30772	0.06757	<.0001
SESW (t3) $\leftarrow$ SESW (t1)	0.46125	0.06681	<.0001
Abbruchintention $\leftarrow$ UCB	-0.49911	0.06594	<.0001

Tabelle 13.34: EOS-Modell: stand. Pfadkoeffizienten mit Standardfehler sowie zugehörigem p -Wert, Jhg. 3

Pfad	Koeffizient	SE	p
akad. Leistung $\leftarrow$ Schulleistung	-0.20407	0.07408	0.0059
soziale Unterstützung $\leftarrow$ SESW (t1)	0.23059	0.07199	0.0014
Zielsetzung $\leftarrow$ SESW (t1)	0.80908	0.02626	<.0001
SESW (t3) $\leftarrow$ Zielsetzung	0.23579	0.10435	0.0238
akad. Leistung $\leftarrow$ SESW (t3)	0.09037	0.0749	0.2276
Stress $\leftarrow$ soziale Unterstützung	-0.12276	0.07389	0.0966
UCB $\leftarrow$ SESW (t3)	0.4547	0.05992	<.0001
Studienzufriedenheit $\leftarrow$ Stress	-0.63751	0.04677	<.0001
Stress $\leftarrow$ SESW (t1)	-0.36419	0.06803	<.0001
SESW (t3) $\leftarrow$ Studienzufriedenheit	0.21124	0.06833	0.002
SESW (t3) $\leftarrow$ SESW (t1)	0.28491	0.10604	0.0072
Abbruchintention $\leftarrow$ UCB	-0.29301	0.06941	<.0001

13.6 Sekundärdatenanalyse

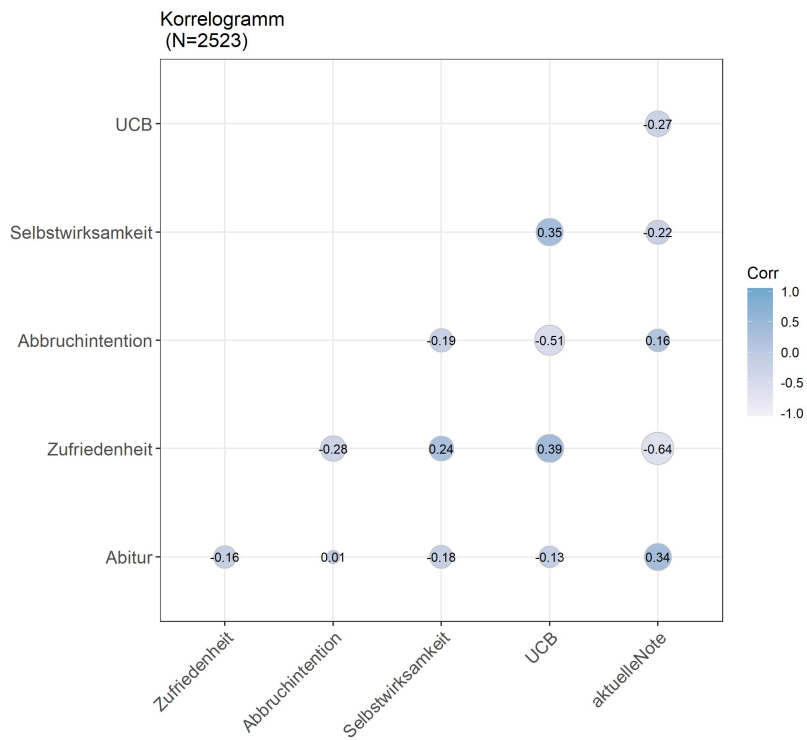


Abbildung 13.2: Korrelogramm der verwendeten Variablen aus dem (Teil-)Datensatz der 21. Sozialerhebung

13.7 Interventionsstudie

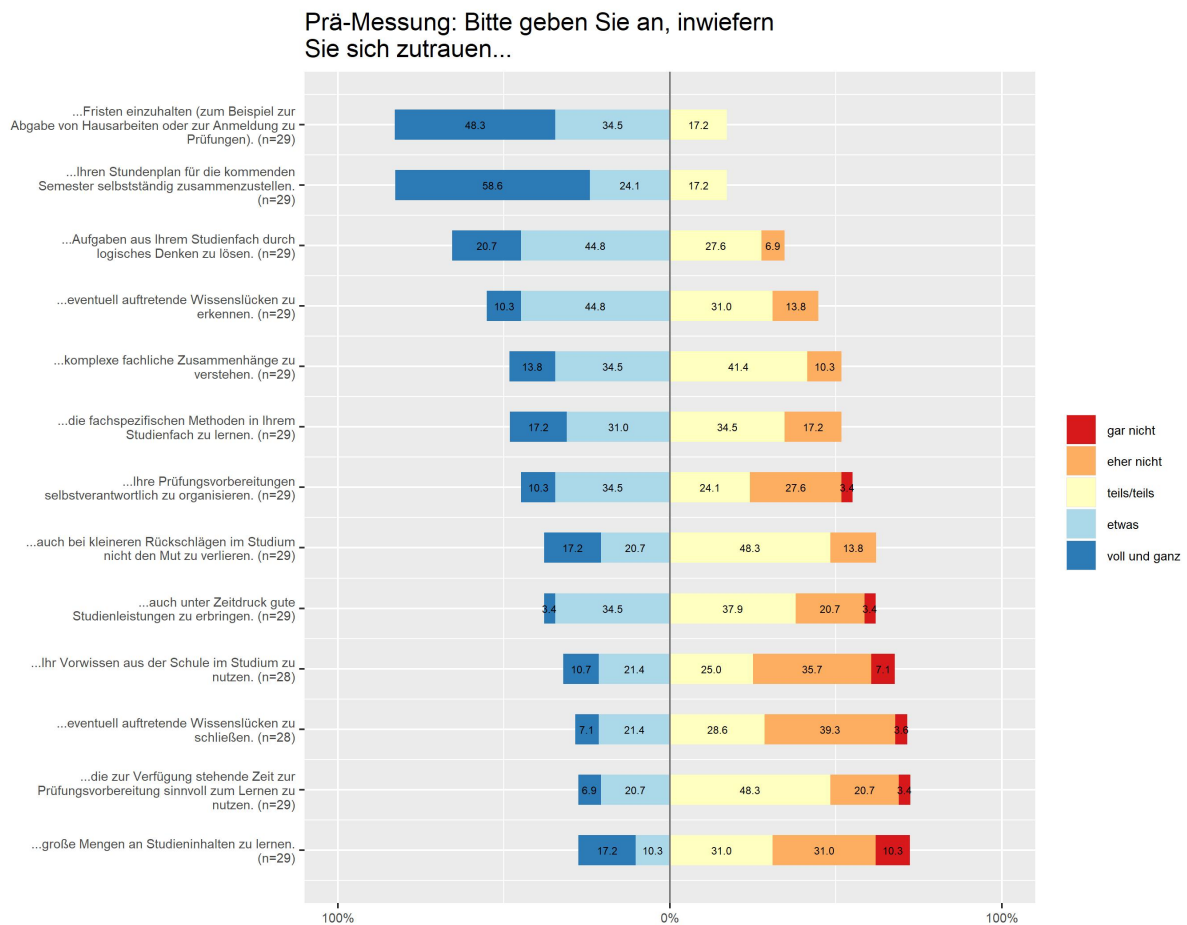


Abbildung 13.3: Interventionsstudie: SESW-Skalenwert-Verteilung (prä)

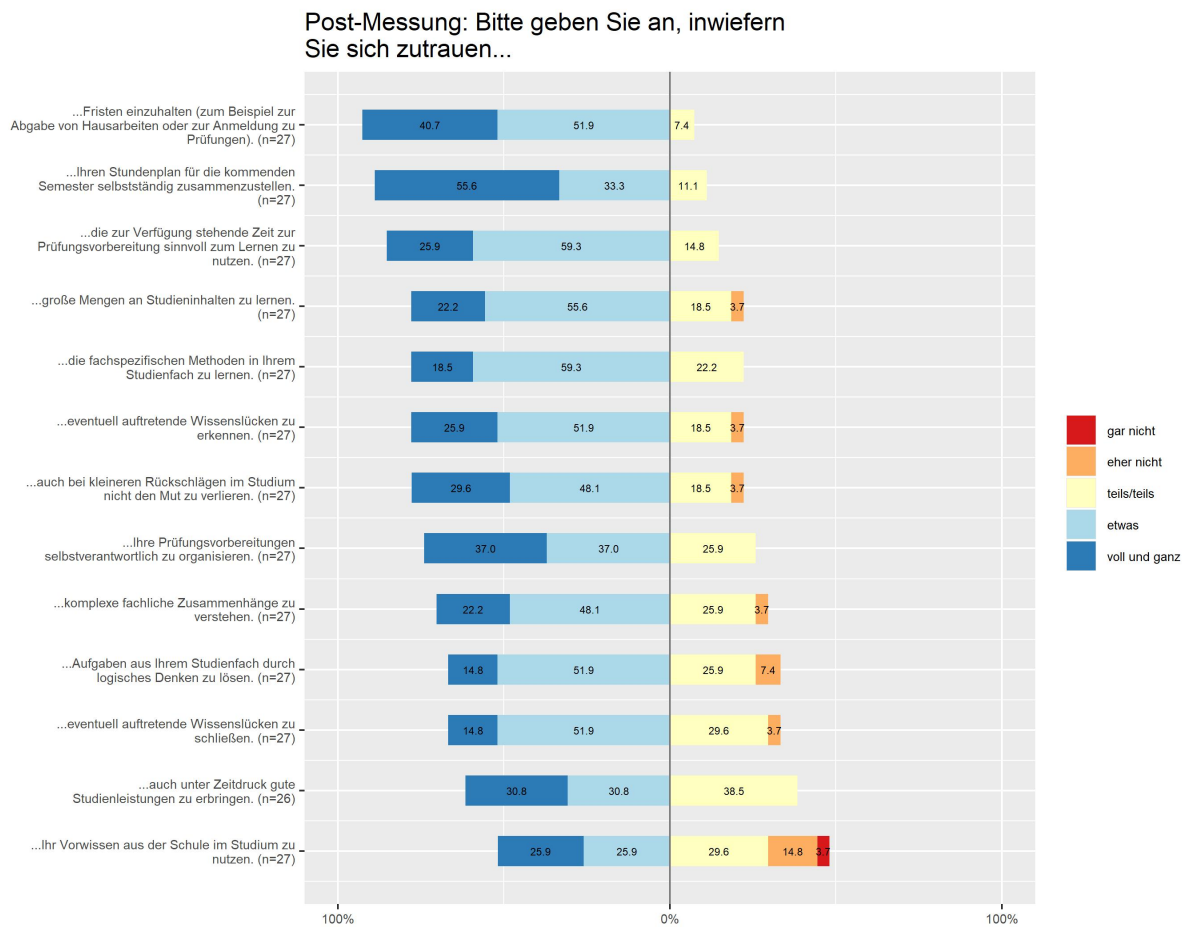


Abbildung 13.4: Interventionsstudie: SESW-Skalenwert-Verteilung (post)