Luise Annemarie Scholz

Was wir von den leistungsstärksten Schüler:innen lernen können

Eine mehrebenenanalytische Untersuchung von Voraussetzungen und Merkmalen (potenziell) leistungsstarker Schüler:innen

Hamburg University Press
Verlag der Staats- und Universitätsbibliothek Hamburg
Carl von Ossietzky

Impressum

Zugl.: Dissertation, Universität Hamburg, 2024 u. d. T. Was wir von den leistungsstärksten Schüler:innen lernen können. Eine mehrebenenanalytische Untersuchung von Voraussetzungen und Merkmalen (potenziell) leistungsstarker Schüler:innen

BIBLIOGRAPHISCHE INFORMATION DER DEUTSCHEN NATIONALBIBLIOTHEK Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://portal.dnb.de abrufbar.
LIZENZ Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Das Werk steht unter der Creative-Commons-Lizenz Namensnennung 4.0 International (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/legalcode.de). Ausgenommen von der oben genannten Lizenz sind Teile, Abbildungen und sonstiges Drittmaterial, wenn anders gekennzeichnet.



ONLINE-AUSGABE Die Online-Ausgabe dieses Werkes ist eine Open-Access-Publikation und ist auf den Verlagswebseiten frei verfügbar. Die Deutsche Nationalbibliothek hat die Online-Ausgabe archiviert. Diese ist dauerhaft auf dem Archivserver der Deutschen Nationalbibliothek (https://portal.dnb.de) verfügbar.
DOI https://doi.org/10.15460/hup.272.2138
GEDRUCKTE AUSGABE ISBN 978-3-910391-05-5
COVER UND SATZ Hamburg University Press
DRUCK UND BINDUNG Books on Demand – Norderstedt (Deutschland)
VERLAG Hamburg University Press, Verlag der Staats- und Universitätsbibliothek Hamburg Carl von Ossietzky, Hamburg (Deutschland), 2025
https://hup.sub.uni-hamburg.de

Inhalt

1 Einleitung 1

I Theoretischer Teil

2 Leistungsstarke Schüler:innen – Begriffsbestimmung und Abgrenzung 9

3 Erklärungsmodelle zur Entstehung von Schulleistungen 23

4 Prädiktoren von Schulleistungen 31

4.1 Merkmale der Schüler:innen 32
4.1.1 Fachbezogenes Selbstkonzept 41
4.1.2 Fachbezogene Einstellung zum Sachunterricht 46

4.2 Merkmale des Unterrichts 50
4.2.1 Unterrichtsqualität 51
4.2.2 Differenzierung 68

4.3 Merkmale der Lehrkraft 74
4.3.1 Berufszufriedenheit 79
4.3.2 Lehrkräfteselbstwirksamkeit 84

5 Forschungsfrage und Ableiten der Hypothesen 91

5.1 Hypothesen zu den Merkmalen der Schüler:innen 93
5.2 Hypothesen zu den Merkmalen des Unterrichts 94
5.3 Hypothesen zu den Merkmalen der Lehrkräfte 97
5.4 Zusammenfassung 99

II Methodischer Teil

6 Datengrundlage: TIMSS 2019 103

6.1 Hintergründe und Ziele der TIMS-Studie 104
6.2 Organisation der TIMS-Studie und Schlussfolgerungen 106
6.2.1 Zeitlicher Ablauf der Studienzyklen 107
6.2.2 Ablauf der Testung 107
6.2.3 Stichprobenziehung und Besonderheiten des Datensatzes 108

6.3 Beschreibung der Stichprobe 110
6.4 Beschreibung ausgewählter Instrumente 111
6.4.1 Kompetenztests 111
6.4.2 Schülerfragebogen 113
6.4.3 Lehrkräftefragebogen 115
6.4.4 Darstellung der relevanten Skalen 116

7 Methodisches Vorgehen 125

7.1 Datenaufbereitung und Voranalysen 126
7.1.1 Vorbereitung des Datensatzes 127
7.1.2 Fehlende Werte 128
7.1.3 Intraklassenkorrelation 132
7.1.4 Zusammenfassung und Schlussfolgerungen 134

7.2 Konfirmatorische Faktorenanalysen 134
7.2.1 Grundlagen der konfirmatorischen Faktorenanalyse 135
7.2.2 Überprüfung der Datengrundlage 136
7.2.3 Kriterien zur Beurteilung der Ergebnisse 136

7.3 Mehrebenenstrukturgleichungsmodelle 142
7.3.1 Strukturgleichungsmodelle 142
7.3.2 Mehrebenenmodelle 144
7.3.3 Zusammenfassung und Schlussfolgerungen für die Analysen 149

8 Ergebnisse 151

8.1 Prüfung der latenten Konstrukte und Ableitung der Messmodelle 152
8.2 Berechnung des Nullmodells als Referenzmodell 161
8.3 Berechnung der Vormodelle 162
8.3.1 Vormodell I: Lehrkräftemerkmale 163
8.3.2 Vormodell II: Unterrichtsmerkmale 164

8.4 Ableitung und Berechnung des Gesamtmodells 166
8.4.1 Hauptmodell I: Schülermerkmale 168
8.4.2 Hauptmodell II: Schüler- und Unterrichtsmerkmale 169
8.4.3 Hauptmodell III: Schüler-, Unterrichts- und Lehrkräftemerkmale 173
8.4.4 Hauptmodell IV: Gesamtmodell mit Kontrollvariablen 176

9 Zusammenfassung und Diskussion der Ergebnisse 183

9.1 Ergebnisse 184
9.2 Einordnung und Interpretation 188
9.2.1 Merkmale der Schüler:innen 188
9.2.2 Merkmale des Unterrichts 192
9.2.3 Merkmale der Lehrkraft 198
9.2.4 Zusammenführung vor dem Hintergrund des Rahmenmodells 201

9.3 Kritische Reflexion der vorliegenden Arbeit 202

10 Fazit 213

10.1 Ausblick für weitere Forschungen 214
10.2 Bedeutung für die pädagogische Praxis 223

Literaturverzeichnis 229
Anhang 275
Über die Autorin 293

Abbildungsverzeichnis

Abb.: 1	Zusammenfassende Systematisierung anhand des TAD-Frameworks nach Preckel et al. (2020), des Überblicksschemas nach Deiglmayr et al. (2017), des Talent-Development-Megamodells nach Subotnik et al. (2011) und des Differenzierten sowie Erweiterten Begabungs- und Talentmodell nach Gagné (2013); (eigene Darstellung)	20
Abb.: 2	Rahmenmodell zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen (Lankes et al., 2002)	25
Abb.: 3	Modell der fünf Kategorien nicht-kognitiver Merkmale nach Farrington et al. (2012)	34
Abb.: 4	Struktur des Selbstkonzeptes nach Marsh et al. (1988)	42
Abb.: 5	Drei-Komponenten-Modell der Einstellung nach Rosenberg und Hovland (1960)	47
Abb.: 6	Komponenten des Professionswissens nach Baumert und Kunter (2011) sowie Shulman (1987)	75
Abb.: 7	Kaskadenmodell des Professionswissens nach Krauss et al. (2020)	77
Abb.: 8	Heuristisches Modell der Selbstwirksamkeit nach Zee und Koomen (2016)	87
Abb.: 9	Hypothesen zu individuellen Merkmalen der Schüler:innen	94
Abb.: 10	Hypothesen zu Merkmalen des Unterrichts I	95
Abb.: 11	Hypothesen zu Merkmalen des Unterrichts II	96
Abb.: 12	Hypothesen zu den Merkmalen der Lehrkraft I	98
Abb.: 13	Hypothesen zu den Merkmalen der Lehrkraft II	99
Abb.: 14	Grafische Übersicht über die Hypothesen	100
Abb.: 15	Klassifikation des KMO-Kriteriums nach Kaiser (1974)	136
Abb.: 16	Übersicht über die in dieser Arbeit genutzten *Model-Fit*-Werte	141
Abb.: 17	Strukturgleichungsmodell in Anlehnung an Backhaus et al. (2015)	144
Abb.: 18	Zusammenhänge in MEAs in Anlehnung an Jäckle und Schärdel (2017)	146
Abb.: 19	Ergebnisse Vormodell I	164
Abb.: 20	Ergebnisse Vormodell II	166
Abb.: 21	Abgeleitetes Gesamtmodell	167
Abb.: 22	Ergebnisse Hauptmodell I	169
Abb.: 23	Ergebnisse Hauptmodell II	172
Abb.: 24	Ergebnisse Hauptmodell III	175
Abb.: 25	Ergebnisse Hauptmodell IV	180

Tabellenverzeichnis

Tab.: 1	Klassifikation der Ergebnisse konfirmatorischer Faktorenanalysen	135
Tab.: 2	Messmodell individuelle Förderung	154
Tab.: 3	Messmodell Klassenführung	154
Tab.: 4	Messmodell konstruktive Unterstützung	154
Tab.: 5	Messmodell Berufszufriedenheit	155
Tab.: 6	Messmodell I und II für fachbezogene Einstellung	156
Tab.: 7	Model-Fit-Werte für Messmodell I und II für fachbezogene Einstellung	156
Tab.: 8	Messmodell I und II für fachbezogenes Selbstkonzept	157
Tab.: 9	Model-Fit-Werte für Messmodell I und II für fachbezogenes Selbstkonzept	157
Tab.: 10	Messmodell I und II für kognitive Aktivierung	158
Tab.: 11	Model-Fit-Werte für Messmodell I und II für kognitive Aktivierung	158
Tab.: 12	Messmodell I und II für Lehrkräfteselbstwirksamkeit	160
Tab.: 13	Model-Fit-Werte für Messmodell I und II für Lehrkräfteselbstwirksamkeit	160
Tab.: 14	Model-Fit-Werte Nullmodell II	162
Tab.: 15	Model-Fit-Werte Vormodell I	163
Tab.: 16	Model-Fit-Werte Vormodell II	165
Tab.: 17	Model-Fit-Werte Hauptmodell I im Vergleich zu vorherigen Modellschritten	168
Tab.: 18	Model-Fit-Werte Hauptmodell II im Vergleich zu vorherigen Modellschritten	170
Tab.: 19	Model-Fit-Werte Hauptmodell III im Vergleich zu vorherigen Modellschritten	173
Tab.: 20	Model-Fit-Werte Hauptmodell VI im Vergleich zu vorherigen Modellschritten	177
Tab.: 21	Übersicht über die bestätigten Hypothesen	186
Tab.: 22	Übersicht über die abgelehnten Hypothesen	187

1

Einleitung

„Alle Schülerinnen und Schüler unabhängig von Herkunft, Geschlecht und sozialem Status so zu fördern, dass für alle Kinder und alle Jugendlichen ein bestmöglicher Lern- und Bildungserfolg gesichert ist – das ist Leitlinie einer auf Chancengleichheit und Bildungsgerechtigkeit zielenden Bildungspolitik. [...] Das gilt gleichermaßen für leistungsstarke wie potenziell besonders leistungsfähige Kinder und Jugendliche. Die Potenziale aller Kinder und Jugendlichen müssen möglichst frühzeitig erkannt werden. Alle Kinder und Jugendlichen benötigen geeignete Formen des Lehrens und Lernens sowie auf sie zugeschnittene und sie aktivierende Angebote der Beratung und Begleitung ihres Bildungsganges“ (KMK, 2016, o. S.).

Diese Forderung der *Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) an das deutsche Bildungssystem bietet den Auftakt zur ‚Gemeinsamen Initiative von Bund und Ländern zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler‘ (KMK, 2016). Bereits ein Jahr zuvor verständigten sich die Kultusminister:innen auf eine gemeinsame Förderstrategie für leistungsstarke Schüler:innen, in der die Beweggründe, diese Schülergruppen verstärkt in den Blick zu nehmen, erläutert werden:

> „Ein Blick auf den vergleichsweise geringen Anteil von Schülerinnen und Schülern auf den beiden oberen Kompetenzstufen der PISA-Studien bzw. der Ländervergleiche der Kultusministerkonferenz sowohl im Bereich der Naturwissenschaften/Mathematik als auch in Deutsch und Englisch verdeutlicht die Notwendigkeit, die Förderung von leistungsstarken und potenziell leistungsfähigen Schülerinnen und Schülern zu verbessern. Diese Zielgruppe umfasst Schülerinnen und Schüler, die bereits sehr gute beobachtbare Leistungen erbringen, ebenso wie Schülerinnen und Schüler, deren Potenziale es zu erkennen und durch gezielte Anregung und Förderung zu entfalten gilt“ (KMK, 2015, S. 3).

Die Befunde der *Trends in International Mathematics and Science Study* (TIMSS) 2019 für die Grundschule zeigen im Jahr 2020 erneut die Notwendigkeit, derartige Bemühungen fortzusetzen (Schwippert, Kasper, et al., 2020a). Weiterhin erreicht in Deutschland nur ein sehr geringer Anteil der Schüler:innen die höchste Kompetenzstufe in Mathematik und den Naturwissenschaften. Während in Mathematik über die letzten zwölf Jahre mit Anteilen zwischen 5 und 6 % auf der höchsten Kompetenzstufe kaum Veränderungen zu erkennen sind, ist für die naturwissenschaftlichen

Kompetenzen eine Abnahme des Anteils von Schüler:innen auf der höchsten Kompetenzstufe von knapp 10 % auf 6.9 % zu verzeichnen (Steffensky, Scholz, Kasper & Köller, 2020). Dies ist umso erstaunlicher, als dass seit dem ersten TIMSS-Zyklus im Jahr 2007 wiederholt eine bessere Förderung der leistungsstarken Schüler:innen gefordert wird.

Damit die Förderung bestimmter Schülergruppen erfolgreich sein kann, muss ausreichend Wissen über (relevante) Charakteristika der jeweiligen Schülergruppen vorhanden sein, um geeignete Ansatzpunkte für eine Förderung ausmachen zu können. In der bisherigen Forschung sind Untersuchungen, die differenzielle Ansätze verfolgen, um Zusammenhänge zwischen verschiedenen Merkmalen für unterschiedliche Gruppen zu untersuchen, jedoch eher selten. An dieser Stelle setzt die vorliegende Arbeit an. Um den Forderungen nach besserer Förderung (potenziell) leistungsstarker Schüler:innen nachkommen zu können, steht die Frage im Fokus, wodurch sich die leistungsstärksten Schüler:innen von potenziell leistungsstarken Schüler:innen unterscheiden und welche Merkmale dazu beitragen, dass Schüler:innen nicht nur gute Leistungen auf Kompetenzstufe IV erbringen, sondern auch die höchste Kompetenzstufe V erreichen. Anhand des Rahmenmodells zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen (Lankes et al., 2002) können verschiedene Bereiche relevanter Einflussfaktoren von Leistung identifiziert werden. Auf Basis der Befunde von Metaanalysen und umfangreichen Übersichtsarbeiten können dabei drei Bereiche als besonders bedeutsam eingestuft werden: individuelle Schülermerkmale, der Unterricht und die Lehrkraft (Hattie, 2009; Köller & Baumert, 2018; Lipowsky, 2006; Wang, Haertel & Walberg, 1993). Die wenigen differenziellen Untersuchungen, die bisher zu Prädiktoren von besonderer Leistungsstärke vorliegen, sowie weitere generelle Forschungen zu Einflussfaktoren von schulischen Leistungen konnten zeigen, dass insbesondere motivationale Merkmale der Schüler:innen wie das Selbstkonzept und die fachbezogene Einstellung (Kriegbaum, Becker & Spinath, 2018; Pitsia, 2022; Steinmayr & Spinath, 2009; Teig & Nilsen, 2022), Merkmale der Unterrichtsqualität (Klieme, Schümer & Knoll, 2001; Kunter & Voss, 2011; Praetorius, Klieme, Herbert & Pinger, 2018; Sanfo & Malgoubri, 2021; Schiepe-Tiska, Heine, Lüdtke, Seidel & Prenzel, 2016) und ausgewählte Komponenten des Professionswissens (Berufszufriedenheit und Lehrkräfteselbstwirksamkeit; Blömeke et al., 2014; Fauth et al., 2019; Rowen, Chiang & Miller, 1997) potenzielle Prädiktoren besonderer Leistungsstärke darstellen könnten. In dieser Arbeit wird deshalb mithilfe von (multinomial) logistischen Regressionen zur Bestimmung der Übergangswahr-

scheinlichkeiten von Kompetenzstufe IV auf Kompetenzstufe V sowie Mehrebenenstrukturgleichungsmodellen zur Identifikation von direkten und indirekten Effekten der ausgewählten Merkmale auf die Leistung die Frage nach relevanten Prädiktoren besonderer Leistungsstärke untersucht.

Da für Naturwissenschaften der Anteil leistungsstarker Schüler:innen in TIMSS über die letzten zwölf Jahre abgenommen hat und Mathematik als Domäne deutlich besser erforscht ist, werden die Untersuchungen in dieser Arbeit am Beispiel der Naturwissenschaften durchgeführt. Darüber hinaus werden Schüler:innen in der Grundschule in den Blick genommen. Zwar konnten beispielsweise mit dem *Programme for International Student Assessment* (PISA) ähnliche Verteilungen auch für die Sekundarstufe festgestellt werden, allerdings sind die untersuchten Schüler:innen in dieser Studie, in der 15-Jährige in den Blick genommen werden, bereits in fortgeschrittenen Phase der Sekundarschulzeit angekommen. Die Untersuchung von Schüler:innen zu diesem späten Zeitpunkt ihrer schulischen Laufbahn geht mit verschiedenen Nachteilen einher: Zum einen sind zu diesem Zeitpunkt zahlreiche Merkmale der Schüler:innen bereits in ihrer Ausprägung stabiler als dies im Grundschulalter der Fall ist (Deary, Whiteman, Starr, Whalley & Fox, 2004; Praetorius, Kastens, Hartig & Lipowsky, 2016; Stern & Grabner, 2014), zum anderen bleibt weniger Zeit, die Schüler:innen systematisch im schulischen Kontext zu fördern. *Large-Scale-Assessments* wie TIMSS oder die *Internationale Grundschul-Lese-Untersuchung* (IGLU; engl. *Progress in International Reading Literacy Study*; PIRLS) untersuchen hingegen Schüler:innen am Ende ihrer Grundschulzeit. Sie bieten somit den Vorteil, dass die Schüler:innen bereits einige Zeit im schulischen Kontext verbracht haben, diese aber noch in einem so frühen Stadium ihrer schulischen Laufbahn befinden, dass sich verschiedene Schülermerkmale noch einfacher und systematisch im weiteren Verlauf beeinflussen und fördern lassen. Da, wie eingangs beschrieben, im Rahmen von TIMSS 2019 erneut der geringe Anteil leistungsstarker Schüler:innen auf Kompetenzstufe V nachgewiesen werden konnte, bilden die Daten von TIMSS 2019 die Grundlage für die Untersuchungen in dieser Arbeit. Damit wird gleichzeitig der von der KMK im Eingangszitat formulierte Anspruch erfüllt, frühzeitig Potenziale zu identifizieren und zu fördern.

Die Nutzung der Daten von TIMSS 2019 bietet den großen Vorteil, dass ein bestehender Befund auf Basis der Daten näher untersucht werden kann. Gleichzeitig enthalten die umfangreichen Datensätze von TIMSS nicht nur Informationen über die befragten Schüler:innen, sondern auch über ihre Lehrkräfte, Schulleitungen und

Erziehungsberechtigten, sodass unterschiedlichste potenzielle Einflussfaktoren in den Blick genommen werden können. Darüber hinaus ermöglichen es die Daten der TIMS-Studie aufgrund ihres Designs, repräsentative Aussagen zu treffen. Sie bieten somit die ideale Grundlage, um relevante Befunde im Feld differenzieller Forschungen zu leistungsstarken Schüler:innen zu generieren.

Im ersten Teil der vorliegenden Arbeit werden die theoretischen Grundlagen für die Analysen erarbeitet. Dazu wird zunächst eine Begriffsbestimmung leistungsstarker Schüler:innen vorgenommen (Kapitel 2), bevor das Rahmenmodell zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen vorgestellt und auf Basis bisheriger Forschung relevante Merkmalsbereiche für die zu untersuchende Fragestellung abgeleitet werden (Kapitel 3). Anschließend werden die als bedeutsam identifizierten Bereiche theoretisch aufgearbeitet und konkrete Merkmale benannt (Kapitel 4). In drei Unterkapiteln werden die einzelnen potenziellen Prädiktoren mit relevanten theoretischen Grundlagen sowie Forschungsbefunden vorgestellt (Kapitel 4.1 bis 4.3). Darauffolgend werden aus den Erläuterungen der Kapitel 2 bis 4 die Forschungsfrage für diese Arbeit sowie die zu überprüfenden Hypothesen abgeleitet (Kapitel 5).

Im zweiten Teil der Arbeit werden die methodischen Aspekte dargestellt. Dazu erfolgt zunächst eine detaillierte Vorstellung der TIMS-Studie sowie der Datengrundlage und der Erhebungsinstrumente (Kapitel 6). Anschließend wird das methodische Vorgehen, das die Datenaufbereitung sowie die Grundlagen von konfirmatorischen Faktorenanalysen und Mehrebenenstrukturgleichungsmodellen umfasst, erläutert (Kapitel 7). Darauf aufbauend werden im nachfolgenden Kapitel die Ergebnisse der Analysen präsentiert (Kapitel 8). Dieses beinhaltet in vier Unterkapiteln die Prüfung der latenten Konstrukte und die Ableitung der konkreten Messmodelle (Kapitel 8.1), die Berechnung eines Referenzmodells (Kapitel 8.2), die Berechnung von Vormodellen (Kapitel 8.3) sowie die Ableitung und schrittweise Berechnung des Gesamtmodells (Kapitel 8.4). Anschließend werden die Ergebnisse der Analysen in den bisherigen Diskurs eingeordnet. Dazu werden die zentralen Befunde zunächst zusammengefasst (Kapitel 9.1) und anschließend entlang der drei Bereiche (individuelle Schülermerkmale, Unterricht und Lehrkraft) diskutiert (Kapitel 9.2). Darüber hianus erfolgt in diesem Kapitel auch eine kritische Reflexion der vorliegenden Arbeit (Kapitel 9.3). Den Abschluss dieser Arbeit bildet das Fazit (Kapitel 10), das sich aus einem Ausblick für weitere Forschungen (Kapitel 10.1) und der Darstellung der Bedeutung der Ergebnisse für die pädagogische Praxis zusammensetzt (10.2).

I Theoretischer Teil

2

Leistungsstarke Schüler:innen – Begriffsbestimmung und Abgrenzung

In der Diskussion um Schüler:innen, die hohe Leistungen in der Schule erbringen, existiert eine Vielzahl unterschiedlicher Bezeichnungen und Kategorisierungen. Einher gehen diese Umschreibungen häufig mit weiteren thematisch verwandten Begriffen wie Kompetenz, Intelligenz, Begabung oder Potenzial als mögliche Synonyme oder Quellen von Leistungsstärke, wobei die Verwendung und Systematik in der Literatur stark uneinheitlich ist. So werden beispielsweise in einigen Publikationen begabte oder auch hochbegabte sowie leistungsstarke und kompetente Schüler:innen gleichgesetzt, in anderen wiederum mit differenzierenden Maßstäben explizit unterschieden. Teilweise wird dabei ausschließlich auf den intellektuellen Bereich Bezug genommen, teilweise werden explizit die verschiedensten Domänen einbezogen und in wieder anderen Publikationen werden die Definitionen gar nicht oder nur sehr oberflächlich thematisiert – dies betrifft sowohl Alltagspublikationen (z. B. Elternratgeber), wissenschaftliche Literatur als auch politische Veröffentlichungen (Hüther & Hauser, 2012; KMK, 2009, 2015; Kyed, 2018; Neuendorf, 2022; Rost & Sparfeldt, 2008; Ziegler, 2017). Insbesondere die unterschiedlichen Intentionen verschiedener Disziplinen und ein variierender Gebrauch in den unterschiedlichen Sprachen, aber auch durch Bemühungen, mit Zusätzen wie ‚statisch', ‚dynamisch' oder ‚realisiert' die verschiedenen Konstrukte mit nur einem Begriff zu erklären, tragen zur Unschärfe der verschiedenen Begrifflichkeiten und ihrer Bedeutungen bei.

Publikationen, die sich mit einer umfassenden Systematisierung der verschiedenen Begriffe befassen, gibt es bisher nur wenige. In den meisten wird der Fokus nur auf einige wenige Konstrukte gelegt und deren Bedeutung und Verhältnis unter Zuhilfenahme nicht weiter definierter Begrifflichkeiten erklärt. In diesem Kapitel wird deshalb zunächst eine Systematik grundlegender Bezeichnungen (Kompetenz, Leistung, Performanz, Begabung, Potenzial, Hochbegabung, Talent und Expertise) erarbeitet. Da die Begrifflichkeiten und ihr Verständnis wie beschrieben zwischen verschiedenen Sprachen und Domänen variieren und die vorliegende Arbeit mit Fokus auf der Untersuchung von Schüler:innen in Deutschland im Feld der Erziehungswissenschaft verortet werden kann, wird in diesem Kapitel der Schwerpunkt auf deutschsprachige Forschungen aus den Bereichen Pädagogik und Pädagogische Psychologie gelegt, um den Begriff leistungsstarker Schüler:innen für die vorliegende Arbeit zu bestimmen.

Insbesondere durch großangelegte Schulvergleichsuntersuchungen wie TIMSS oder PISA ist in den letzten Jahrzehnten der Begriff der Kompetenzen stärker in den Fokus fachwissenschaftlicher Diskurse gerückt, wobei verschiedene Verständnisse von Kom-

petenz vorliegen.[1] Weinert, dessen Beschäftigung mit Kompetenzen als wegweisend für das aktuelle Verständnis angesehen werden kann, empfiehlt, sich an einer Definition zu orientieren, die auch den erwähnten *Large-Scale*-Untersuchungen zu Grunde liegt. Er beschreibt Kompetenzen als „funktional bestimmte, auf bestimmte Klassen von Situationen und Anforderungen bezogene kognitive Leistungsdispositionen, die sich psychologisch als Kenntnisse, Fertigkeiten, Strategien, Routinen oder auch bereichsspezifische Fähigkeiten beschreiben lassen" (Weinert, 1999 zitiert nach Klieme, 2004, S. 10). Unklar ist aktuell, inwiefern motivationale oder volitionale Faktoren als Teil des Kompetenzbegriffes angesehen werden können (Röhr-Sendlmeier & Käser, 2017). Während Weinert in einer späteren Definition diese Merkmale explizit mit in den Kompetenzbegriff einschließt[2], hebt Trültzsch-Wijnen (2020) zusammenfassend hervor, dass individuelle Faktoren wie Motivation nicht Teil des Kompetenzbegriffes sind, aber eng mit diesem zusammenhängen und beispielsweise eine vermittelnde Funktion bei der Umsetzung von Kompetenz in Leistung einnehmen (siehe auch Chomsky, 1980; Mandl & Krause, 2001).

Diese Perspektive zeigt zudem, dass Kompetenz nicht mit Leistung bzw. Performanz gleichzusetzen ist. Die Ursprünge dieses Verständnisses liegen im Kompetenz-Performanz-Modell nach Chomsky (1980). Das aus der Linguistik stammende Modell dient dazu, die Entwicklung von sprachlichen Fähigkeiten zu erklären und ist somit an der Schnittstelle zwischen Entwicklungspsychologie und Linguistik einzuordnen. Entsprechend fand es vielfach auch in psychologische oder erziehungswissenschaftliche Theorien und Publikationen Eingang (Mandl & Krause, 2001). Nach dem Kompetenz-Performanz-Modell kann davon ausgegangen werden, dass Sprachkompetenzen genetisch bedingt und dafür verantwortlich sind, dass Menschen Sprache erwerben können. Wird Sprache z. B. über das Verstehen von Gesprochenem oder das Bilden von Sätzen eingesetzt, spricht man nach dem Kompetenz-Performanz-Modell von Performanz (Chomsky, 1980), wobei nach Knoblauch (2009) die Performanz auch

1 Für eine Übersicht über die verschiedenen Definitionen von Kompetenz siehe beispielsweise Weinert (1999) oder Hartig und Klieme (2006).

2 „Dabei versteht man unter Kompetenzen die bei Individuen verfügbaren oder die erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können." (Weinert, 2001, S. 27–28)

als (objektivierte) Leistung verstanden werden kann. Performanz bzw. Leistung ist somit als Ausdrucksform von Kompetenzen zu verstehen, ein Verständnis, das auch in aktuellen erziehungswissenschaftlichen sowie psychologischen Publikationen vertreten wird. Verändert hat sich jedoch seit Chomskys Publikation die Auffassung über die Stabilität von Kompetenzen. So werden Kompetenzen inzwischen als entwickelbar angesehen (siehe beispielsweise Weinert, 2001), an die Stelle von unveränderbaren, genetisch bedingten Dispositionen rücken stattdessen andere Merkmale, die im Weiteren näher erläutert werden.

Diesen Entwicklungen folgend wird Leistung in dieser Arbeit synonym zu Performanz als manifestierter Ausdruck bestimmter Kompetenzen in konkreten Anforderungssituationen verstanden. Kompetenzen können wiederum definiert werden als erlernbare, bereichsspezifische Fähigkeiten, die eine notwendige, aber nicht hinreichende Voraussetzung für Leistung darstellen. Von gezeigten Leistungen kann dabei auf die dahinterliegenden spezifischen Kompetenzen im Sinne von latenten, also nicht direkt beobachtbaren Merkmalen, geschlossen werden, der Umkehrschluss ist jedoch nicht zulässig (Trültzsch-Wijnen, 2020).

Begrifflichkeiten, die insbesondere, wenn ein besonders hohes Maß an Leistung gezeigt wird, in den Diskurs mit einfließen, sind vor allem Intelligenz, Begabung, Potenzial, Hochbegabung, Talent und Expertise. Um eine eindeutige Eingrenzung des Forschungsgegenstandes für die vorliegende Arbeit zu erreichen, werden diese Begriffe im Folgenden definiert und wo nötig explizit von den Begriffen Leistung und Kompetenz abgegrenzt. Abschließend werden die verschiedenen Konstrukte in einer Übersicht systematisiert und eine Definition leistungsstarker Schüler:innen für die vorliegende Arbeit abgeleitet.

Zentral in der Diskussion um Leistungsstärke ist das Merkmal Intelligenz, dessen Definitionen sich über die letzten Jahrzehnte z. T. stark voneinander unterschieden und verschiedene Aspekte betonten. Während Boring (1923) in seiner eher pragmatischen, aber treffenden Definition Intelligenz als das, was ein Intelligenztest misst, definierte, fokussierten andere ältere Definitionen wie die von Binet und Simon (1916)[3]

[3] "It seems to us that in intelligence there is a fundamental faculty, the alteration or the lack of which, is of the utmost importance for practical life. This faculty is judgment, otherwise called good sense, practical sense, initiative, the faculty of adapting one's self to circumstances. To judge well, to comprehend well, to reason well, these are the essential activities of intelligence" (Binet & Simon, 1916, S. 42–43).

oder Wechsler (1944)[4] darauf, dass sich Intelligenz durch „gutes" oder „vernünftiges" Denken und Handeln äußert. Neuere Definitionen zeigen hingegen vor allem hohe Übereinstimmungen hinsichtlich der Problemlösefähigkeit und der Bewältigung neuer Anforderungen als zentrale Aspekte der Intelligenz. Eine Definition, die zusammenführend diese neueren Ansätze widerspiegelt und noch heute zu den am weitest akzeptierten Definitionen von Intelligenz zählt (Stern & Neubauer, 2016), stammt von Gottfredson (1997), die Intelligenz definiert als „very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings 'catching on,' 'making sense' of things, or 'figuring out' what to do" (S. 13).

Intelligenz ist somit eine grundlegende Voraussetzung zur Erbringung kognitiver Leistungen – in Abgrenzung zu sogenannten ‚nicht-kognitiven' Leistungen, z. B. im musischen oder sportlichen Bereich[5] – und entsprechend bedeutsam für das schulische Lernen (Vock & Gronostaj, 2017). Deshalb wurde in der Forschung häufig die Frage thematisiert, ob Intelligenz ein stabiles Persönlichkeitsmerkmal oder durch Umwelteinflüsse veränderbar ist. Wegweisend waren in diesem Zusammenhang die Arbeiten von Horn und Cattell (1966), die zwei Formen der Intelligenz, die fluide und die kristalline Intelligenz, unterscheiden. Unter der fluiden Intelligenz wird die genetische, also angeborene Komponente der Intelligenz verstanden, während die kristalline Intelligenz das Ergebnis von Sozialisation und Erfahrungen darstellt. In der jüngeren Vergangenheit entwickelte sich basierend auf früheren Erkenntnissen die These, dass Intelligenz sich vor allem durch das Zusammenspiel von Genen und Umwelteinflüssen entwickelt. Das bedeutet, dass die Intelligenz grundsätzlich durch genetische Dispositionen angelegt ist, es aber förderlicher Umweltmerkmale und Lerngelegenheiten bedarf, um diese zu entwickeln, auszubauen und zu erhöhen. Hierbei beeinflussen die genetischen Voraussetzungen, wie Lerngelegenheiten genutzt werden und wie individuelle Lernresultate

4 „Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (Wechsler, 1944, S. 3).

5 Hervorzuheben ist, dass der Begriff ‚nicht-kognitiv' nicht die Abwesenheit kognitiver Anteile in diesen Domänen meint. Stattdessen werden darunter Bereiche oder Merkmale verstanden, die nicht primär akademisch oder auf akademische Leistungen bezogen sind. Da der Begriff ‚nicht-kognitiv' in der Forschung jedoch bereits etabliert ist, wird er trotz dieser Unklarheit ebenfalls in dieser Arbeit verwendet.

ausfallen (Plomin & von Stumm, 2018; Stern & Grabner, 2014; Voland, 2006). Während in den frühen Jahren der Kindheit noch große Entwicklungssprünge zu verzeichnen sind, ist die Intelligenz ab dem späten Grundschulalter als zwar altersrelative, aber stabile Persönlichkeitseigenschaft anzusehen, die valide über Intelligenztests gemessen werden kann (Deary et al., 2004; Stern & Grabner, 2014).

Damit wird nach Hartig und Klieme (2006) auch eines der drei Kriterien zur Abgrenzung von Intelligenz und Kompetenz deutlich: Während Kompetenzen als erlernbar angesehen werden, ist Intelligenz ein eher stabiles Merkmal. Des Weiteren können beide Konzepte durch ihre Kontextspezifität (Intelligenz als generelles kognitives Potenzial vs. Kompetenz als bereichsspezifische Fähigkeit) und ihre Binnenstrukturen (grundlegende biologische und kognitive Prozesse als Grundlage der Intelligenz vs. anforderungsspezifische Strukturierung der Kompetenz) differenziert werden (Hartig & Klieme, 2006).

Abzugrenzen ist Intelligenz zusätzlich von den Konstrukten Begabung und Potenzial. Die Begriffe Potenzial und Begabung werden in der Literatur häufig synonym verwendet, wobei der Begriff Potenzial in zahlreichen Publikationen vornehmlich zur Erklärung von Begabung genutzt wird. Findet sich diese Definition, so wird Begabung als Möglichkeit (sprich als Potenzial) für die Erbringung von Leistung verstanden. Dabei ist zu betonen, dass die Bestimmung als Möglichkeit beinhaltet, dass Begabung bzw. Potenzial nicht zwangsläufig auch in Leistung umgesetzt werden muss (Rost, 2004). Begabung wird dabei zumeist als breiter Begriff verstanden, der verschiedene Domänen (kognitiv, musisch, sportlich etc.) umfasst und entsprechend für jeden Bereich unterschiedlich stark ausgeprägt sein kann (Fischer, 2015; iPEGE, 2009). Manche Autor:innen ergänzen diese Definition darüber hinaus mit der Klassifizierung von kognitiver Begabung als genetischer Komponente der Intelligenz (Kuhl, 2021; Preckel, Stumpf & Schneider, 2018). Da sich Intelligenz – abhängig von der Altersstufe und je nachdem wie Lernmöglichkeiten der individuellen Begabung entsprechend genutzt wurden – in unterschiedlichen Ausprägungen zeigt, kann sie als entwickelte und somit messbar gemachte kognitive Begabung verstanden werden.

In der Literatur lassen sich jedoch auch einige abweichende Positionen finden. So definieren beispielsweise Köster, Mehrtens und Müller (2022) in ihrem Modell naturwissenschaftlicher (Leistungs)Potenziale in Kita- und Grundschule Begabung als aus Potenzial entwickelte überdurchschnittliche Performanz, wobei sie damit eine eher seltene und an das Alltagsverständnis von Begabung angelehnte Definition vertreten. Heller (1976) kam zudem zu dem Schluss, dass Begabung in vielen Publikationen

mit Intelligenz gleichgesetzt wird. Wird in der Literatur zwischen beiden Begriffen differenziert, gibt es vor allem zwei Leitideen: Einerseits wird Intelligenz als generelle und Begabung als domänenspezifische Fähigkeit verstanden, andererseits wird entgegengesetzt Intelligenz als spezifischer und Begabung als weitgefasster Begriff verstanden. In den letzten Jahren scheint sich in dieser Debatte jedoch die zweite der beiden Leitideen durchzusetzen (siehe beispielsweise Deiglmayr, Schalk & Stern, 2017; Kuhl, 2021; Preckel et al., 2018).

Wird Begabung und dabei insbesondere eine hohe Begabung thematisiert, sollte auch der Begriff Hochbegabung näher beleuchtet werden. Wie man dieser Bezeichnung direkt entnehmen kann, handelt es sich bei Hochbegabung um eine hohe Ausprägung der Begabung, also einem besonders stark ausgeprägten Potenzial zu hohen Leistungen (Preckel et al., 2018). In diesem Punkt gibt es in der Literatur eine hohe Übereinstimmung, trotzdem existiert eine Vielzahl von Definitionen, mit denen eine abweichende (siehe beispielsweise Ziegler [2017], der Hochbegabung als besonders hohe Performanz definiert) oder explizitere Beschreibung angestrebt wird (Holling & Kanning, 1999). Auch deshalb gibt es Fragen in der Hochbegabungsforschung, die noch nicht abschließend geklärt sind, unter anderem auf welche Domänen der Begriff Hochbegabung angewendet werden kann und welche Kriterien erfüllt sein müssen, damit von einer Hochbegabung gesprochen werden kann.

Lange Zeit wurde mit dem Begriff Hochbegabung ausschließlich eine besonders hohe Ausprägung der kognitiven Begabung umschrieben. Insbesondere durch Modelle, die multiple Intelligenzen oder Begabungen postulieren und Bestrebungen hohe Begabungen in nicht-kognitiven Bereichen stärker anzuerkennen, sollte der Begriff zwischenzeitlich auch auf andere Domänen ausgeweitet werden. Dennoch wird Hochbegabung in den meisten Publikationen nach wie vor auf die kognitive Komponente bezogen, was neben der hohen Verbreitung dieses Verständnisses nicht zuletzt daran liegen dürfte, dass Modelle zu multiplen Intelligenzen inzwischen als überholt gelten (Stern & Neubauer, 2016). In dieser Arbeit wird der Begriff Hochbegabung deshalb ohne näheren Zusatz ausschließlich für eine hohe kognitive Begabung verwendet.

Als ein in Forschung und Praxis gängiges Kriterium für Hochbegabung kann der *Intelligenzquotient* (IQ) angesehen werden, wobei i. d. R. ein IQ von 130 oder höher als Grenzwert für Hochbegabung benannt wird. Zwar wird dieses IQ-Kriterium zunehmend als unidimensional und zu stark monokausal kritisiert, aber selbst führende Ansätze und Modelle zur multikausalen Beschreibung von Hochbegabung (wie z. B. das Modell von Mönks [1985] oder das Münchner Hochbegabungsmo-

dell von Heller und Perleth [2007]) schaffen es aufgrund fehlender Praktikabilität oder (partiellen) Unklarheiten (noch) nicht Hochbegabung ‚besser' diagnostizierbar zu machen (Arnold, Kleß & Prescher, 2014; Stern & Neubauer, 2016). Da sich das IQ-Kriterium in der Praxis zudem aus ökonomischen Gründen anbietet (etablierte Verfahren zur Diagnostik sowie klares Kriterium bspw. zur Aufnahme in spezielle Förderprogramme), wird in dieser Arbeit – unter Anerkennung der Multikausalität von Hochbegabung – der Intelligenzquotient als primäres Merkmal zur Diagnostik einer Hochbegabung verstanden.

Da Intelligenz und Hochbegabung nach den beschriebenen Verständnissen nur Facetten der kognitiven Domäne beschreiben, bedarf es zusätzlicher Begrifflichkeiten für andere Bereiche. Aktuell bietet sich dafür am ehesten der Begriff Talent an, der zwar in seinem Verständnis einem fast permanenten Wandel und stark differierenden Definitionen in verschiedenen Disziplinen unterliegt, insbesondere im deutschsprachigen Diskurs aber zur Beschreibung nicht-kognitiver Domänen wie Sport oder Kunst genutzt wird (Peter, 2018; Rost & Buch, 2018). Im Verhältnis zu Begabung definieren Gagné (1985) oder Deiglmayr et al. (2017) Talent als Pendant zur Intelligenz als durch Lernmöglichkeiten realisierte nicht-kognitive Begabung. Einen eigenständigen Begriff, der als Pendant zu Hochbegabung angesehen werden kann, scheint es hingegen nicht zu geben, weshalb zur Beschreibung besonders hoher nicht-kognitiver Begabungen am ehesten der Begriff Hochbegabung mit einem konkretisierenden Zusatz (z. B. künstlerisch) als geeignet angesehen werden kann.

Ein weiterer Begriff, der häufig in Debatten rund um Leistungsstärke oder Hochbegabung genutzt wird, ist der der Expertise. Die Expertiseforschung befasst sich anders als die Hochbegabungsforschung nicht damit, Merkmale zur Vorhersage möglicher zukünftiger Leistungsfähigkeit zu treffen, sondern untersucht, wie bereits gezeigte überdurchschnittliche Leistungen in der Rückschau zu erklären sind (Preckel et al., 2018). Die klassischen Anwendungsfelder der Expertiseforschung sind Schachspielen und Musizieren, beides Bereiche, in denen relativ eindeutig Menschen mit außergewöhnlichen Fähigkeiten zu beobachten sind. Gleichzeitig wird deutlich, dass es sich bei Expertise nicht um ein globales Fähigkeitsvermögen handelt, sondern um ein in hohem Maße domänenspezifisches und überdauerndes Merkmal. Dabei wird zwischen sogenannten Lai:innen (Personen ohne Kenntnisse), Noviz:innen (Personen mit durchschnittlichen Kenntnissen) und Expert:innen (Personen mit außergewöhnlichen Kenntnissen) unterschieden, die vor allem durch die investierte Übungszeit und aktive Beschäftigung mit dem Gegenstand und weniger durch Merkmale wie

Begabung oder Intelligenz klassifiziert werden. Als Richtwert für das Erreichen des Expertenstatus hat sich inzwischen eine Anzahl von ca. 10.000 Übungsstunden etabliert (de Groot, 1965; Glaser & Chi, 2009; Grabner & Meier, 2021). Da im Rahmen dieser Arbeit Grundschüler:innen betrachtet werden sollen, die allein aufgrund ihres Alters noch keinen Expertenstatus erreicht haben können, liegt für die klassische Expertiseforschung nur eine begrenzte Passung vor. In neueren Untersuchungen wird jedoch der Ansatz der Expertiseforschung mit Erkenntnissen zu weiteren Einflussfaktoren kombiniert, sodass diese für die vorliegende Arbeit Anhaltspunkte bieten können. Zu den prominentesten Modellen, die der integrativen Expertiseforschung zugeordnet werden können, zählen das Talent-Development-Megamodell nach Subotnik, Olszewski-Kublilius und Worrell (2011), das Differenzierte sowie das Erweiterte Begabungs- und Talentmodell (DMGT/EMGT) nach Gagné (1985, 2013) sowie das Talent-Development-In-Achievement-Domains-Framework (TAD) nach Preckel et al. (2020). Allen Modellen ist gemein, dass sie die Entwicklung von Begabung bzw. Potenzial hin zu entwickelten Talenten nachzeichnen und dabei verschiedene Stufen und Einflussfaktoren benennen.[6] Während sich das DMGT und EMGT darauf fokussieren, die verschiedenen Begabungs- und Talentfacetten sowie Einflussfaktoren differenziert darzustellen, betonen das Talent-Development-Megamodell und das TAD-Framework die verschiedenen, übereinstimmenden Stufen der Expertiseentwicklung: Potenzial, Kompetenz, Expertise und Eminenz bzw. außergewöhnliche Leistung. Dabei werden zur Reduktion der Komplexität die Umwelteinflüsse zwar in ihrer Bedeutsamkeit anerkannt, in den beiden Modellen selbst jedoch nachrangig behandelt. Das TAD-Framework zeichnet sich zudem gegenüber dem Talent-Development-Megamodell dadurch aus, dass es die verschiedenen Schritte inhaltlich näher bestimmt und die wichtigsten Prädiktoren und Indikatoren benennt. Das Talent-Development-Megamodell fokussiert hingegen stärker auf die Domänenspezifität sowie die Entwicklung verschiedener Fähigkeiten und ordnet diese Prozesse der jeweiligen Lebensphase zu (für Darstellungen der einzelnen Modelle siehe Anhang).

---

6 Da diese Modelle in der Tradition englischsprachiger Diskurse stehen, in der eine abweichenden Bedeutung des Begriffes *talent* vorherrschend ist, geht diese Benennung nicht mit den vorherigen Erläuterungen einher. Da der Begriff Talent jedoch nicht Teil des TAD-Frameworks und des Talent-Development-Megamodells selbst, sondern nur der Benennung ist und im Rahmen des DMGTs bzw. des EMGTs separat definiert wird, können diese Modelle trotzdem für eine Systematisierung und Rahmung herangezogen werden.

Damit erkennen diese Modelle stärker als die klassische Expertiseforschung die Schritte auf dem Weg zur Expertise (bzw. Eminenz als höchste Stufe) an und beziehen neben den Übungsprozessen weitere Einflussfaktoren mit ein, sodass das komplexe Zusammenspiel verschiedenster Faktoren betrachtet werden kann. Die Modelle der integrativen Expertiseforschung können zudem anders als die klassische Expertiseforschung nicht nur zur Erklärung von außergewöhnlichen Leistungen herangezogen werden, sondern auch für die Betrachtung der Zwischenschritte und verschiedenen Altersphasen und bieten somit für die vorliegende Arbeit wichtige Anhaltspunkte zur Erklärung hoher Leistungen auch in jüngeren Jahren.

Wie eingangs beschrieben, gibt es bisher kaum deutschsprachige Publikationen, die sich umfassend mit der Definition und Abgrenzung verschiedener Begrifflichkeiten aus diesem Forschungsfeld befassen. Ausnahme bildet eine Veröffentlichung von Deiglmayr et al. (2017), die in der Auswahl der systematisierten Begrifflichkeiten einige Überschneidungen mit dieser Arbeit zeigt. Im Folgenden werden die bisherigen Erkenntnisse zusammengeführt und unter Zuhilfenahme der Systematisierung nach Deiglmayr et al. (2017), des TAD-Frameworkes, des Talent-Development-Megamodells und des DMGTs bzw. EMGTs in eine für diese Arbeit grundlegende Übersicht überführt.

Die bisherigen Erkenntnisse lassen sich folgendermaßen zusammenfassen:

1. Unter Begabung bzw. Potenzial versteht man die Möglichkeiten einer Person Leistungen zu erbringen, wobei diese je nach Domäne unterschiedlich stark ausfallen können.
2. Als kognitive Komponente der Begabung, die durch erste Lerngelegenheiten geprägt wird und zu einem gewissen Anteil entwickelbar ist, kann Intelligenz verstanden werden, als Pendant für nicht-kognitive Komponenten das Talent.
3. Von Hochbegabung spricht man, wenn eine besonders hoch ausgeprägte kognitive Begabung (IQ ≥ 130) vorliegt.
4. Intelligenz und Talent können durch zunehmende Lerngelegenheiten zu domänenspezifischen Kompetenzen, also Fertigkeiten und Kenntnissen, entwickelt werden.
5. Werden Kompetenzen in konkreten Anforderungssituationen angewandt, spricht man von Leistungen oder Performanz.
6. Zeigt jemand auf einem bestimmten Gebiet außergewöhnliche Leistungen, kann von Expertise bzw. Eminenz gesprochen werden.

Einher gehen diese angenommenen Schritte mit einer Schärfung bzw. Einengung des eigenen Profils sowie einer Zunahme an Fertigkeiten und Kenntnissen in der gewählten Domäne. Dabei ist zu beachten, dass die Entwicklung von Kompetenzen bis hin zu Eminenz über einen längeren Zeitraum erfolgt, der verschiedene Altersstufen umfasst, wobei je nach individuellen Möglichkeiten, Anstrengungen, Interessen und äußeren Einflussfaktoren prinzipiell auf jeder der Entwicklungsschritte verblieben werden kann. Das bedeutet, dass in einer Domäne noch viele Individuen Kompetenzen erwerben können, jedoch nur wenige Eminenz erreichen werden, genauso wie Eminenz aufgrund hoher Investitionszeit von einer Person in der Regel nur in einem sehr eng definierten Feld erlangt werden kann. Dieses Zusammenspiel der Entwicklungsschritte mit Profilschärfung, Fähigkeitszunahme und Einflussfaktoren ist in Abbildung 1 dargestellt. Einige der Elemente, wie konkrete personale und Umwelteinflüsse, die in den herangezogenen Modellen an prominenter Stelle aufgeführt sind, sind dabei nicht in ihrer Ausführlichkeit mit in die Systematisierung aufgenommen, sondern lediglich angedeutet.

Für die vorliegende Arbeit bedeutet dies Folgendes: Die Schüler:innen der TIMS-Studie befinden sich zum Zeitpunkt der Erhebung am Ende der vierten Klasse, laut Übersicht lassen sie sich somit in den Phasen der Fähigkeitsdifferenzierung und Profilbildung verorten. Sie sind dabei, Kompetenzen in einem noch eher breiten Feld mit aber ersten inhaltlichen Schwerpunkten zu entwickeln und können entsprechende Leistungen zeigen. Zu diesem Zeitpunkt können Schüler:innen i. d. R. noch keine Expertise oder Eminenz in einer Domäne erreicht haben, dennoch sind sie im Rahmen der Möglichkeiten in der Lage, sehr hohe Leistungen zu erbringen.

Diese hohen, in schulischen oder schulähnlichen Anforderungssituationen gezeigten, das heißt nachgewiesenen, Leistungen werden in dieser Arbeit in den Blick genommen. Im Modell vorgelagerte oder nicht-kognitive Konstrukte wie Begabung oder Talent sind hingegen nicht Teil der Untersuchung, dies gilt insbesondere auch für das Konstrukt Hochbegabung, das häufig mit hohen Leistungen gleichgesetzt wird. Dabei muss jedoch beachtet werden, dass Hochbegabung zwar mit hohen Leistungen einhergehen kann, wie dargestellt, jedoch einige grundsätzliche Abweichungen in der Konzeptualisierung (Hochbegabung als Möglichkeit, latentes Merkmal etc.) aufweist.

In groß angelegten Schulvergleichsstudien wird wie in dieser Arbeit der Fokus auf die Erfassung und Untersuchung von Leistungen und dahinterliegenden Kompetenzen

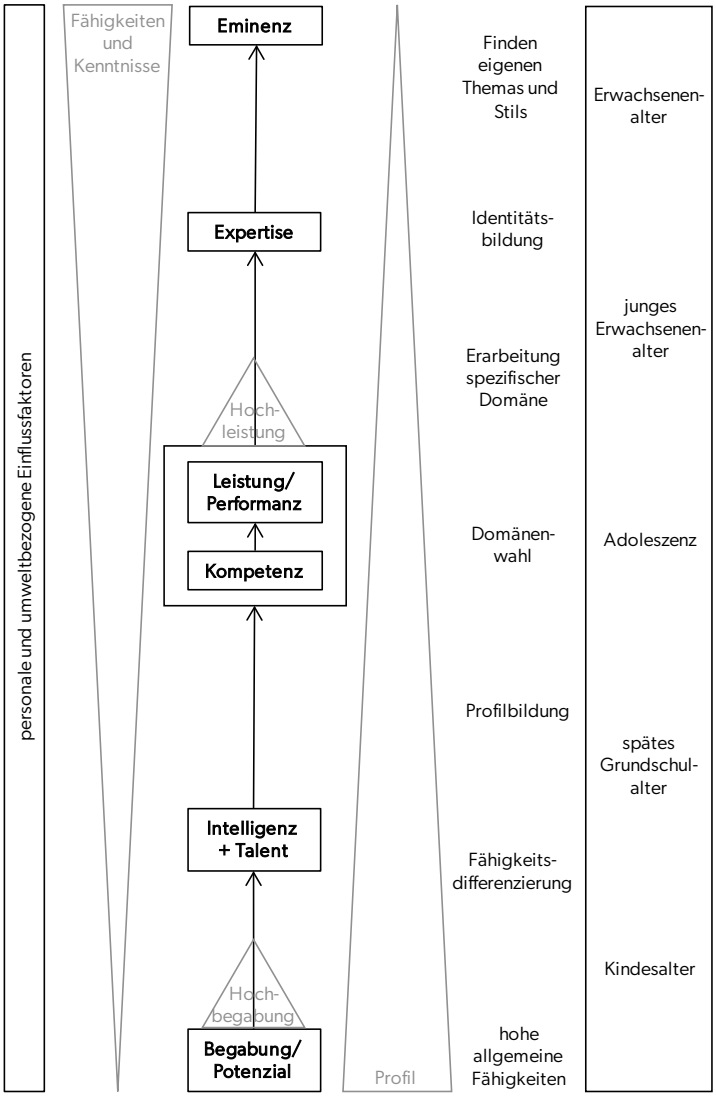


Abb.: 1 Zusammenfassende Systematisierung anhand des TAD-Frameworks nach Preckel et al. (2020), des Überblicksschemas nach Deiglmayr et al. (2017), des Talent-Development-Megamodells nach Subotnik et al. (2011) und des Differenzierten sowie Erweiterten Begabungs- und Talentmodell nach Gagné (2013); (eigene Darstellung)

gelegt. Dazu werden die Leistungen von Schüler:innen über Tests[7] erhoben, die dann auf sogenannten vorab definierten Kompetenzstufen verortet werden können. Dieses Vorgehen ermöglicht es, Aussagen darüber zu treffen, welche Kompetenzen – als nicht direkt messbares Konstrukt – Schüler:innen wahrscheinlich aufweisen. Im Rahmen von TIMSS werden insgesamt fünf Kompetenzstufen definiert, wobei Schüler:innen auf Kompetenzstufe I die niedrigsten Leistungen im TIMSS-Test gezeigt haben und Schüler:innen auf Kompetenzstufe V die höchsten (für nähere Informationen zu den Kompetenzstufen in TIMSS siehe Kapitel 6.4.1). Diese Einordnung ist grundlegend für die Definition der leistungsstarken Schüler:innen in dieser Arbeit, da die Analysen wie eingangs erläutert auf den Befunden von TIMSS 2019 aufbauen und mit Daten der TIMS-Studie durchgeführt werden: Als ‚leistungsstärkste Schüler:innen' werden im Folgenden Schüler:innen auf Kompetenzstufe V bezeichnet, in Abgrenzung dazu Schüler:innen, die Leistungen auf Kompetenzstufe IV erbringen, als ‚potenziell leistungsstarke Schüler:innen'. Mit dem Ausdruck ‚leistungsstarke Schüler:innen' werden beide Gruppen gemeinsam adressiert.

Wird in der vorliegenden Arbeit also über die leistungsstarken Schüler:innen gesprochen, bedeutet das, dass ein Merkmal im Fokus steht, das – anders als (Hoch) Begabung, Intelligenz oder Talent – in hohem Maße veränderlich und somit förderbar ist. Dies betonte schon Weinert (2001), der die Bedeutung von Lerngelegenheiten gegenüber Begabungen hervorhebt:

> „Unabhängig von den unterschiedlichen Fähigkeiten und Talenten der Schüler muss alles gelernt werden, was später gewusst und gekonnt wird. Lernen ist der mächtigste Mechanismus der kognitiven Entwicklung. Das gilt uneingeschränkt sowohl für hochbegabte Kinder als auch für schwächer begabte Schüler (Weinert, 2001, S. 85)."

Dies zeigt, welche Bedeutung der Lernumwelt und damit insbesondere der Schule eines Kindes darin zukommt, Lerngelegenheiten zu schaffen, in denen Schüler:innen

7 Die schulischen Leistungen von Schüler:innen können alternativ auch über deren Noten in den einzelnen Fächern erhoben werden. Da jedoch in verschiedenen Untersuchungen gezeigt wurde, dass Noten durchaus auch andere Facetten als die Leistung bewerten und sozialen Bezugsnormen unterliegen (Maaz et al., 2008; Valtin, 2012), werden in dieser Arbeit Testleistungen als sachliche bzw. kriteriale Leistungsmaße zur Definition der leistungsstarken Schüler:innen herangezogen (Rheinberg, 2001).

bei entsprechender Nutzung ihre Kompetenzen und Interessen entwickeln können. Deshalb wird das Zusammenspiel von personalen Faktoren, Umwelteinflüssen und Schulleistungen im folgenden Kapitel näher beleuchtet.

3

Erklärungsmodelle zur Entstehung von Schulleistungen

In den vergangenen Jahrzehnten gab es in der Bildungsforschung verschiedene Ansätze zur Erklärung von Schulleistungen. Viele davon resultierten in Modellen, die verschiedene Einflussfaktoren auf die Schulleistung abbilden, in Beziehung zueinander setzen und Wirkrichtungen postulieren, wobei die verschiedenen Modelle unterschiedlichen Ziel- und Schwerpunktsetzungen unterliegen (z. B. das Angebots-Nutzungsmodell nach Fend [1980] und Helmke und Weinert [2017], das Prozess-Produkt-Paradigma [Gräsel, 2020] oder das Diagnose-Prognose-Modell nach Heller [1991]). Im Rahmen der IGLU-Studie 2001 wurde dabei ein viel rezipiertes Modell weiterentwickelt, das auf einem Modell von Helmke und Weinert (1997) und u. a. Vorarbeiten von Wang et al. (1993) basiert und seitdem auch in der TIMS-Studie Verwendung findet. Insbesondere in *Large-Scale*-Studien kann dieses aufgrund des Einbezugs vielfältiger Perspektiven umfassend und adäquat untersucht werden (Lankes et al., 2002). Im Rahmenmodell zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen wird der Fokus auf die verschiedenen Bereiche des Bildungssystems und deren Zusammenhang mit den Leistungsergebnissen der Schüler:innen gelegt. Dabei werden vor allem übergeordnete Themenblöcke in den Mittelpunkt gestellt, die verschiedene Merkmale in sich vereinen (siehe Abbildung 2).

Endpunkt des Modells bilden die Leistungsergebnisse der Schüler:innen, also beispielsweise Testleistungen, Noten oder erreichte Abschlüsse. Diese werden von individuellen Schülermerkmalen wie Geschlecht, kognitiven Fähigkeiten, Vorwissen oder motivationalen Faktoren beeinflusst, die wiederum mit dem elterlichen Erziehungs- und Unterstützungsverhalten, also beispielsweise der Kooperation mit der Schule oder dem Umgang mit Hausaufgaben, sowie den schulischen Bedingungen zusammenhängen. Zu den schulischen Bedingungen werden dabei schulinterne Faktoren (z. B. Schulleitung, Ressourcen oder Schulkonzept), Merkmale der Lehrkraft (z. B. professionelle Kompetenzen oder Persönlichkeit), Merkmale der Klassen (z. B. Klassengröße, sozioökonomische Merkmale der Klasse, Klassenklima) sowie der Unterricht (z. B. Unterrichtsqualität oder Einsatz digitaler Medien) gezählt. Das elterliche Erziehungs- und Unterstützungsverhalten nimmt zudem Einfluss auf den schulischen und den außerschulischen Bereich (z. B. Vereinsmitgliedschaft oder Nachmittagsbetreuung), wobei es selbst – wie auch die schulischen Bedingungen und der außerschulische Kontext – durch das soziale und kulturelle Kapital der Eltern bestimmt wird. Dazu zählen beispielsweise das Einkommen, die Wohnsituation oder der Migrationshintergrund der Eltern. Die schulischen Bedingungen werden wiederum beeinflusst von bildungspolitischen und äußeren schulischen Rahmenbedingungen (z. B.

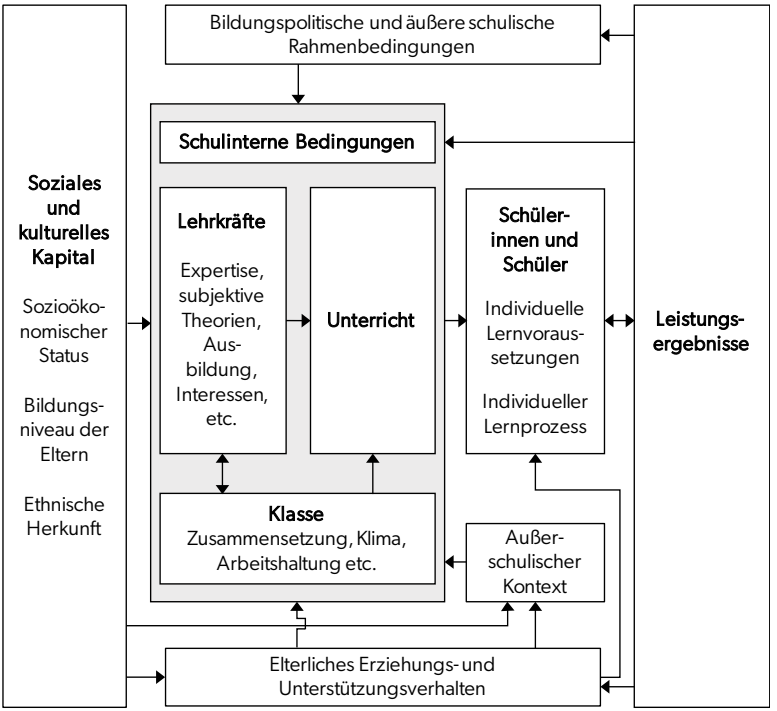


Abb.: 2 Rahmenmodell zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen (Lankes et al., 2002)

curriculare Vorgaben, regionale Schulverwaltungen oder übergreifende Initiativen), durch den außerschulischen Kontext und das elterliche Erziehungs- und Unterstützungsverhalten. Neben direkten und indirekten Effekten der einzelnen Bereiche auf die Leistung werden im Modell auch direkte Effekte der Leistungsergebnisse auf die Bereiche Bildungspolitische und äußere schulische Rahmenbedingungen, Schulinterne Bedingungen, Schülerinnen und Schüler und Elterliches Erziehungs- und Unterstützungsverhalten angenommen. Damit verdeutlicht das Modell übereinstimmend mit zahlreichen anderen Modellen und Veröffentlichungen die komplexen Zusammenhänge zwischen Leistungen und ihren Bedingungsfaktoren (Brühwiler, Helmke & Schrader, 2017; Hosenfeld, Helmke & Schrader, 2002b; Hußmann, Wendt, Kasper, Bos & Goy, 2017; Pekrun, 2002).

Die einzelnen Bereiche, die das Modell enthält, lassen sich klassifizieren in sogenannte proximale und distale Faktoren. Dabei meint der Begriff ‚proximal' Merkmale, die eine hohe Nähe zu den Schulleistungen aufweisen und direkt mit ihnen zusammenhängen, und der Begriff ‚distal' weiter entfernte Merkmale. Betrachtet man dabei die Anordnung der einzelnen Bereiche im Rahmenmodell zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen, zeigt sich, dass die Merkmale jedoch eher im Sinne einer Wirkungskette präsentiert werden. Mit der Bezeichnung proximal und distal werden in diesem Kontext also keine zwei klar voneinander abgrenzbaren Pole definiert, vielmehr sollen sie eine Einordnung der unterschiedlichen Bereiche auf einem Kontinuum ermöglichen. Es gibt somit Merkmale, die sehr weit entfernt vom Zielkriterium Leistung liegen, aber vermittelt über weniger distale sowie proximale Faktoren trotzdem noch Einfluss nehmen, sowie Merkmale, die direkt die Leistung beeinflussen (König, Felske & Kaiser, 2023; Schrader, Helmke & 2008). Zu den eher proximalen Bedingungsfaktoren der Schulleistung zählen im Modell die individuellen Schülermerkmale, der Unterricht, die Klasse, die Lehrkraft sowie das elterliche Erziehungs- und Unterstützungsverhalten. Zu den eher distalen Faktoren, die entsprechend weiter entfernt von der Leistung eingeordnet werden können, gehören das soziale und kulturelle Kapital der Eltern, das Bildungssystem und äußere schulische Rahmenbedingungen sowie schulinterne als auch außerschulische Bedingungen (Brühwiler et al., 2017; Elsäßer, 2017; Maag Merki, 2021).

Zur Bedeutung der einzelnen im Modell angenommenen Bereiche schulischer Bedingungsfaktoren gibt es bereits eine Vielzahl an empirischen Untersuchungen aus den letzten Jahrzehnten. Die meisten Studien fokussierten sich dabei darauf, konkrete ausgewählte Merkmale aus einem oder mehreren Bereichen auf ihren Einfluss auf die schulischen Leistungen von Schüler:innen hin zu untersuchen (Helmke & Schrader, 2010). Es gibt jedoch auch Übersichtsarbeiten sowie Metaanalysen, die die Bedeutung mehrerer Bereiche über verschiedene Schulformen und Fächer hinweg im Gesamten betrachten. Eine häufig zitierte Übersichtsarbeit im deutschsprachigen Raum stammt von Lipowsky (2006), der zu dem Schluss kommt, dass der einflussreichste Bereich für schulische Leistungsergebnisse der der individuellen Merkmale der Schüler:innen ist. Diese Einschätzung stützt er dabei auf Studien von Gruehn (2000), Hosenfeld, Helmke, Ridder und Schrader (2001) und Marzano (2001), die zu dem Ergebnis kamen, dass durch individuelle Merkmale von Schüler:innen zwischen 50 % und 80 % der Varianz in den Leistungen aufgeklärt werden können. Gefolgt wird der Bereich der individuellen Schülermerkmale von Lehrkräfte- und Klassenmerkmalen sowie

den Schulmerkmalen, wobei er begründet durch längsschnittliche Befunde zu der Vermutung kommt, dass die Bereiche Lehrkraft, Unterricht und Klasse relevanter für die Leistungen sind als bisher angenommen. Zudem betont Lipowsky (2006), jedoch ohne diesen Bereich konkret einzuordnen, die vielfach belegte Bedeutung des sozioökonomischen Hintergrundes der Eltern.

Zu einem ähnlichen Ergebnis kommen Köller und Baumert (2018), die ebenfalls die individuellen Merkmale als einflussreichste Faktoren identifizieren. Auf diesen Bereich folgen die Merkmale der Lehrkraft und des Unterrichts, wobei zusätzlich darauf hingewiesen wird, dass Schulmerkmale im Vergleich deutlich weniger bedeutsam für die schulischen Leistungen sind. Ergänzend kommen Kulik und Kulik (1989) in einer umfangreichen Metaanalyse zu verschiedenen Bereichen schulischer Einflussfaktoren zu dem Schluss, dass die Leistungen von Schüler:innen stärker durch die Lehrkraft beeinflusst werden als beispielsweise durch demografische Merkmale von Schüler:innen oder Merkmale der Klassenzusammensetzung. Ähnlich resümieren Kyriakides, Christoforou und Charalambous (2013) in ihrer Metaanalyse von *single* und *multilevel analysis*, dass sowohl Merkmale der Lehrkraft als auch Merkmale des Unterrichts bedeutsam für die Leistungen von Schüler:innen sind. Hingegen keine großen Unterschiede in der Bedeutsamkeit der Bereiche individuelle Merkmale (*aptitude*), Unterricht (*instruction*) und Umwelt (*environment*) können Fraser, Walberg, Welch und Hattie (1987) in ihrer Synthese von 2575 weltweit durchgeführten Studien und Metaanalysen unterschiedlichster methodischer Ansätze zu Einflussfaktoren von Leistung (*educational productivity*) feststellen. Alle drei Bereiche schätzen sie in ihrer Gesamtheit auf Basis von Korrelationen mit Leistung bzw. Effektstärken im niedrigen, teilweise auch negativen Bereich bis hin zu einem sehr hohen Bereich (für *aptitude*: r = .18-.71; für *instruction*: r = -.12-1.17; für *environment*: r = -.05-.79) als besonders bedeutsam für die Leistungsergebnisse von Schüler:innen ein. Unterschiede zeigen sich erst, wenn die einzelnen Konstrukte, die zu den Überthemen zusammengefasst wurden, separat mit ihren Effektstärken betrachtet werden.

Während die bisher vorgestellten Arbeiten primär auf proximale Faktoren abzielen, gibt es auch einige Untersuchungen, die sowohl proximale als auch distale Faktoren gleichermaßen berücksichtigen. So konnten beispielsweise Wang et al. (1993) in ihrer Metaanalyse zeigen, dass „distal variables, like state, district, and school level policy and demographics, have little influence on school learning" (S. 276). Proximale Faktoren wie die individuellen, psychologischen Schülermerkmale, Merkmale des Unterrichts und Merkmale der häuslichen Lernumgebung können hingegen als starke Einfluss-

faktoren von Leistung eingestuft werden (*T*-Werte der proximalen Faktoren liegen mit 51.4 bis 54.7 über dem Normwert von 50, *T*-Werte der distalen Faktoren liegen mit 35.0 bis 47.3 durchgängig darunter). Ähnliches lassen auch die Ergebnisse von Kyriakides und Tsangaridou (2008) vermuten, die verschiedene *educational-effectiveness-studies* hinsichtlich ihrer Befunde zu verschiedenen Einflussfaktoren gegenüberstellen. Die meisten signifikanten Einflüsse auf die Leistung lassen sich dabei für die individuellen Schülermerkmale finden, gefolgt von Merkmalen des Unterrichts. Distale Faktoren wie Schulmerkmale und Schulpolitik stellen sich hingegen als deutlich weniger bedeutsam heraus. Eine geringere Bedeutsamkeit schulischer Merkmale (z. B. Kooperation im Kollegium mit $z = .019$ oder pädagogische Leitung mit $z = .046$; Korrelationen z-transformiert) lässt sich auch in der Zusammenschau der Ergebnisse einer Metaanalyse von Scheerens, Witziers und Steen (2013) feststellen. Faktoren des Unterrichts, insbesondere quantitative Merkmale (z. B. *effective learning time* mit $z = .147$; *opportunity to learn* mit $z = .145$), werden hingegen als besonders einflussreich für die Leistung von Schüler:innen eingestuft. Zwar ebenfalls bedeutsam, aber in deutlich geringerem Maße fallen Merkmale des elterlichen Unterstützungsverhalten ($z = .093$) aus.

Die gewiss bekannteste umfassende Arbeit zu schulischen Einflussfaktoren stammt von Hattie (2009), der in seiner Arbeit eine Metaanalyse mit über 800 Metaanalysen durchführte. Die zentrale Erkenntnis dieser Studie ist, dass neben Merkmalen der Schüler:innen ($d = .40$) die Lehrkräfte ($d = .49$) mit ihrem Unterricht ($d = .42$) den größten Einfluss auf die Leistungen von Schüler:innen haben und damit deutlich einflussreicher sind als schulische Merkmale ($d = .23$). Zudem konnte festgestellt werden, dass unter den elternhausbezogenen Einflussfaktoren ($d = .31$) Merkmale wie das elterliche Erziehungs- und Unterstützungsverhaltens ($d = .57$) oder der soziale und kulturelle Hintergrund ($d = .57$) deutlich bedeutsamer sind als Merkmale wie beispielsweise die Familienstruktur ($d = .17$) oder das Fernsehverhalten ($d = -.18$). Außerschulische Merkmale spielen wie in den anderen gefundenen Übersichtsarbeiten allerdings kaum eine Rolle. Die Befunde zu den wenigen im Rahmen der Studie untersuchten Merkmalen (z. B. außerschulische Erfahrungen mit $d = .09$) verweisen jedoch auf eine geringe Bedeutsamkeit des außerschulischen Bereiches für Leistungsergebnisse.

Relativ übereinstimmend wird somit in der aktuellen Literatur der Bereich der individuellen Merkmale der Schüler:innen als einflussreichster Faktor identifiziert. Ebenfalls recht eindeutig, wenn auch mit kleinen Abweichungen zwischen den verschiedenen Übersichten, werden die Bereiche Unterricht und Lehrkraft als zweitwichtigste Bereiche eingestuft. Darauf folgen die Bereiche des elterlichen Erziehungs- und

Unterstützungsverhaltens, der Klassenmerkmale und der schulischen Bedingungen, wobei hier den Klassenmerkmalen nochmal eine höhere Bedeutung zugeschrieben wird als den Schulmerkmalen. Zu den eher wenig relevanten Prädiktoren individueller Leistungen werden hingegen das Bildungssystem und der außerschulische Kontext gezählt. Zur Einordnung des sozialen und kulturellen Kapitals sowie des außerschulischen Bereiches im Verhältnis zu den Klassen- und Schulmerkmalen konnten in der bisherigen Forschung keine eindeutigen Aussagen gefunden werden. Dies lässt sich damit erklären, dass die meisten Überblicksforschungen und Metaanalysen im Bereich der Erziehungswissenschaft bzw. Pädagogischen Psychologie einen starken Fokus auf die unmittelbar schulkontextbezogenen Bereiche legen, beginnend bei bildungspolitischen und äußeren schulischen Rahmenbedingungen über Schul-, Klassen-, Unterrichts- und Lehrkräftemerkmale sowie Schülermerkmale bis hin zu Leistungsergebnissen. Umfassendere Betrachtungen, die auch, aber nicht ausschließlich Merkmale des Elternhauses oder den außerschulischen Bereich untersuchen, sind hingegen selten, da sie thematisch weit von den schulischen Merkmalen entfernt sind und deshalb häufig in separaten, enger geführten Untersuchungen behandelt werden. Ebenfalls relevant dürften bei der Auswahl der zu untersuchenden Bereiche insbesondere auch zwei übergeordnete Überlegungen sein: Erstens sind die unterschiedlichen Bereiche von Bedingungsfaktoren in unterschiedlichem Ausmaß beeinflussbar. Während beispielsweise Merkmale des sozialen und kulturellen Kapitals – wenn überhaupt – nur schwer direkt verändert werden können, lassen sich Merkmale des Unterrichts oder der Klasse einfacher beeinflussen. Übereinstimmend wird der Unterricht als der Bereich identifiziert, der am leichtesten im gesamten Rahmenmodell veränderbar ist (Hosenfeld et al., 2002b; Lipowsky, 2006).

Zweitens wird davon ausgegangen, dass Merkmale der Schule oder der Klassen nicht direkt auf die Leistungen wirken, sondern über den Unterricht und die individuellen Schülermerkmale vermittelt werden. Sollen also weniger proximale Merkmale betrachtet werden, ist es angezeigt, auch Merkmale der vermittelnden Bereiche einzubeziehen und eine ‚Wirkungskette' vom entferntesten Merkmal zum Leistungskriterium nachzuzeichnen. Dabei wird davon ausgegangen, dass entferntere Faktoren zwar einen Einfluss auf die Leistung haben, der Zusammenhang zwischen beiden Merkmalen jedoch geringer ist als für sehr nahe Merkmale (Lipowsky, 2006):

> „Lernrelevante Aktivitäten können durch Unterricht ebenso wie durch den Kontext und außerunterrichtliche Faktoren (Familie, Medien) zwar angeregt

> und unterstützt, gefördert und kontrolliert werden, liegen aber letztlich in der Kontrolle des Lernenden und hängen von dessen individuellen Merkmalen ab (Brühwiler et al., 2017, S. 294)."

In der Gesamtschau bestätigen die bisherigen empirischen Befunde die größere Bedeutung proximaler und veränderbarer Faktoren (insbesondere der individuellen Schülermerkmale, des Unterrichts und der Lehrkraft) gegenüber den weniger einflussreichen distaleren Faktoren. Dabei scheinen die umfassendsten Publikationen zu verschiedenen Teilbereichen als Einflussfaktoren auf schulische Leistungen aus den 1980er- bis 2000er-Jahren zu stammen. Aktuellere (Übersichts)-Quellen befassen sich hingegen häufiger mit vertiefenden Untersuchungen zu einzelnen, enger begrenzten Themenbereichen und liefern somit eher eine Einschätzung zu einzelnen Merkmalen innerhalb der Bereiche als Erkenntnisse für eine globale Beurteilung der Bereiche. Befunde zu einzelnen Merkmalen ausgewählter Bereiche des Rahmenmodells zur Erklärung von Leistungsergebnissen werden im Folgenden näher beleuchtet. Da das Ziel der vorliegenden Arbeit ist, bedeutsame Prädiktoren für besondere Leistungsstärke und mögliche Ansatzpunkte für eine Förderung von (potenziell) leistungsstarken Schüler:innen herauszuarbeiten, wird dabei der Fokus auf die Bereiche gelegt, die in diesem Kapitel als am relevantesten identifiziert wurden, nämlich Lehrkräfte, Unterricht und individuelle Schülermerkmale.

4

Prädiktoren von Schulleistungen

In diesem Kapitel werden die abgeleiteten Bereiche individuelle Schülermerkmale, Unterricht und Lehrkräfte vom Proximalen (individuelle Schülermerkmale) zum Distalen (Lehrkräfte) vorgehend näher betrachtet. Da in der bisherigen Forschung explizite Untersuchungen zu Prädiktoren besonderer Leistungsstärke nur sehr vereinzelt durchgeführt wurden, werden in den einzelnen Unterkapiteln primär Befunde genereller Forschungen zu den Bedingungsfaktoren von Schulleistungen dargestellt und anschließend die differenziellen Befunde in die allgemeine Forschungslage eingeordnet. Aus diesen Darstellungen werden jeweils Merkmale abgeleitet, die in der vorliegenden Arbeit untersucht werden. Anschließend werden die ausgewählten Merkmale noch einmal näher beleuchtet. Da eines der Ziele dieser Arbeit ist, Ansatzpunkte für eine Förderung potenziell leistungsstarker Schüler:innen zu identifizieren, werden dabei schwerpunktmäßig veränderliche, d. h. direkt förderbare, Merkmale in den verschiedenen, ausgewählten Bereichen betrachtet.

4.1 Merkmale der Schüler:innen

Die Untersuchung von individuellen Schülermerkmalen hat in der erziehungswissenschaftlichen Forschung bereits eine lange Tradition. Dabei fallen unter diesen Überbegriff zahlreiche Merkmale aus dem kognitiven sowie nicht-kognitiven Bereich.[8] Die am häufigsten untersuchten Merkmale des kognitiven Bereiches sind Intelligenz (und verwandte Konzepte), Vorwissen und Leistungsergebnisse, beispielsweise gemessen über Schulnoten, Leistungstests oder Abschlüsse. Unter den kognitiven sowie nicht-kognitiven Prädiktoren konnte in verschiedensten Untersuchungen die Intelligenz als einflussreichster Prädiktor für Leistung ausgemacht werden (Kuncel, Hezlett & Ones, 2004; Rost & Sparfeldt, 2008). Dabei wird davon ausgegangen, dass die Korrelation von Leistung und Intelligenz bei etwa $r = .50$ liegt und die Intelligenz somit ca. ein Viertel der Varianz ($R^2 = .25$) in den Schulleistungen erklärt (Gustafsson & Undheim, 1996; Roth et al., 2015; Stern & Guthke, 2001). Je nach Schulform, Fach oder Messung der Schulleistung können die Werte jedoch variieren. So konnten für standardisierte Leistungstests höhere Zusammenhänge mit der Intelligenz gefunden werden als für Noten, bei denen Referenzgruppeneffekte sowie weitere Merkmale der Schüler:innen

8 Das Verständnis des Begriffes ‚nicht-kognitiv' orientiert sich auch im Kontext von Einflussfaktoren auf die Leistung an den Erläuterungen in Kapitel 2.

bei der Bewertung eine Rolle spielen können (beispielsweise $r = .81$ bei Deary, Strand, Smith und Fernandes [2007], bis zu $r = .69$ bei Colom und Flores-Mendoza [2007] und $r = .48$-$.86$ bei Frey und Dettermann [2004]). Zudem geht Jensen (1980) davon aus, dass der Zusammenhang zwischen Intelligenz und Schulleistung in Grundschulen höher als in anderen Schulformen ist und zwischen ca. $r = .60$ und $r = .70$ liegt. Für weiterführende Schulen bzw. Bildungsinstitutionen scheint der Zusammenhang hingegen zunehmend nachzulassen. Die höchste Varianzaufklärung durch Intelligenz zeigt sich i. d. R. für Mathematik und Naturwissenschaften sowie z. T. für Sprachen. Für sozialwissenschaftliche sowie künstlerisch-praktische Fächer ist die Intelligenz hingegen weniger relevant (Deary et al., 2007; Roth et al., 2015; Vock, Gronostaj & Preckel, 2017).

Mit dem bereichsspezifischen Vorwissen gibt es jedoch einen weiteren kognitiven Prädiktor, der besonders vorhersagekräftig ist und sich in einigen Studien sogar als bedeutsamer herausstellte als die Intelligenz (Siegler et al., 2012; Simonsmeier, Flaig, Deiglmayr, Schalk & Schneider, 2021; Weinert, Helmke & Schneider, 1989). Diese Befunde sind jedoch nicht weiter verwunderlich, berücksichtigt man die Tatsache, dass mit dem bereichsspezifischen Vorwissen ein mit den schulischen Leistungen hochgradig verwandtes Konzept untersucht wird, das die Grundlage für (aktuelle) Leistungen darstellt. Zur Untersuchung besonderer Leistungsstärke scheint das Vorwissen somit ein besonders geeigneter Prädiktor zu sein. Jedoch wäre der Erkenntnisgewinn für das Anliegen dieser Untersuchung eher gering und würde vor allem Aussagen über die Stabilität von Leistungsstärke ermöglichen. Die eigentliche Frage nach Prädiktoren von Leistungsstärke würde damit nur verschoben werden, nämlich zum Zeitpunkt der Erhebung des Vorwissens.

Während die Zahl an kognitiven Merkmalen noch relativ überschaubar ist, sind die nicht-kognitiven Merkmale so vielzählig, dass es einer Strukturierung bedarf, um einen Überblick für diesen Bereich geben zu können. Farrington et al. (2012) erarbeiteten basierend auf einem umfangreichen systematischen Review zu lernrelevanten, nicht-kognitiven Merkmalen ein Modell über fünf Kategorien nicht-kognitiver Merkmale (siehe Abbildung 3). Zentrales Element des Modells sind die fünf Kategorien nicht-kognitiver Merkmale. Diese lassen sich wie folgt charakterisieren:

1. Zu der Kategorie *Academic Behaviors* werden Merkmale gezählt, die das Verhalten von Schüler:innen rund um die Schule und das Lernen widerspiegeln. Sie zeichnen sich dadurch aus, dass sie beobachtbar und somit direkt messbar sind. Beispiele für diese Kategorie sind die regelmäßige Anwesenheit und aktive Teilnahme im Unterricht, das Erledigen von Hausaufgaben, das Lernen

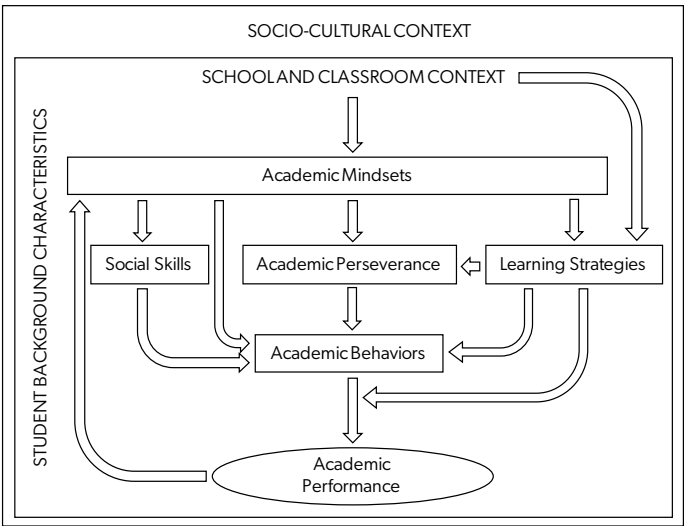


Abb.: 3 Modell der fünf Kategorien nicht-kognitiver Merkmale nach Farrington et al. (2012)

für Tests oder die Organisation von Schulmaterialien. Kurz gefasst, können *Academic Behaviors* beschrieben werden als „those behaviors commonly associated with being a 'good student' " (Farrington et al., 2012, S. 8).

2. Die Kategorie *Academic Perseverance* umfasst vor allem psychologische Merkmale, die das Durchhaltevermögen der Schüler:innen betreffen, z. B. Ausdauer, Belohnungsaufschub, Disziplin, Selbstkontrolle und die Fähigkeit sich zu fokussieren. Bei dieser Kategorie geht es also nicht um Merkmale einer kurzfristigen Problembearbeitung, sondern um eine intensive und langanhaltende Beschäftigung mit dem Lerngegenstand, die trotz auftretender Hindernisse beibehalten wird.
3. Unter die Kategorie *Academic Mindsets* fallen psycho-soziale Vor- und Einstellungen in Bezug auf die eigene Person. Bekannte Theorien, die sich mit Merkmalen dieser Kategorie auseinandersetzen sind beispielsweise die Erwartung x Wert-Theorie (Atkinson, 1957; Eccles & Wigfield, 2002) oder die Zielsetzungstheorie (Locke & Latham, 1990). Insgesamt unterscheiden Farrington et al. (2012) vier verschiedene Mindsets, denen unterschiedliche Merkmale zugeordnet werden können. Das erste Mindset ‚I belong in this academic

community‘ umfasst Merkmale wie z. B. das Zugehörigkeitsgefühl zur Schule oder Klasse, das zweite Mindset ‚My ability and competence grow with my effort‘, beispielsweise das Konstrukt (Miss-)Erfolgsattribuierung. Zum dritten Mindset ‚I can succeed at this‘ werden Merkmale wie die Selbstwirksamkeit gezählt, zum vierten Mindset ‚This work has value for me‘ beispielsweise die intrinsische und extrinsische Motivation.

4. Die vierte Kategorie nicht-kognitiver Faktoren, *Learning Strategies*, umfasst Taktiken und Prozesse, also Strategien, um eigene kognitive Ziele zu erreichen und somit den Lernprozess zu strukturieren und das Lernen zu maximieren. Strategien, die dieser Kategorie zugeordnet werden können, sind beispielsweise Memorierstrategien (z. B. um Fakten abrufen zu können), Verständnisstrategien (z. B. gezielt Informationen beim Lesen von Texten herausfiltern) oder Korrekturstrategien (z. B. bei Fehlkonzepten oder Unklarheiten), aber auch Merkmale wie Zeitmanagement, Selbstregulation oder Metakognitionen, sodass diese Kategorie einen hohen Bezug zur Kognitionswissenschaft aufweist.
5. Unter der Kategorie *Social Skills* werden diejenigen Faktoren zusammengefasst, die sich als zwischenmenschliche Kompetenzen beschreiben lassen. Dazu zählen beispielsweise Faktoren wie Empathie, Kooperation, Durchsetzungsvermögen oder Verantwortung. Damit sind Faktoren dieser Kategorie relevant für die soziale Interaktion im Schulkontext, z. B. zwischen Peers oder Schüler:in und Lehrkraft sowie in Gruppenarbeitssituationen oder Klassendiskussionen.

Diese fünf Kategorien nicht-kognitiver Merkmale werden gerahmt von Hintergrundmerkmalen der Schüler:innen (in Abgrenzung zu nicht-kognitiven Merkmalen beispielsweise sozioökonomischer Hintergrund und Geschlecht), dem Schul- und Klassenkontext sowie dem soziokulturellen Kontext. Besonders bedeutsam scheinen dabei die Kategorien *Academic Mindsets* und *Academic Behaviors* zu sein, da sie entweder alle anderen Kategorien sowie Leistung (*Academic Performance*) indirekt oder direkt beeinflussen (*Academic Mindsets*) oder zumindest mit allen anderen vier Kategorien sowie Leistung direkt zusammenhängt (*Academic Behaviors*). Farrington et al. (2012) kommen in ihrem systematischen Review zu dem Schluss, dass die relevanteste Kategorie für schulische Leistungen *Academic Behaviors* ist: "[…] it is hard to imagine how noncognitive factors could improve student performance without working through the classroom behaviors that directly shape academic performance." (Farrington et al., 2012, S. 8–9)

Merkmale der Kategorie *Academic Mindsets* sind hingegen neben *Learning Strategies* am geeignetsten, um über *Academic Behaviors* die Leistungen von Schüler:innen zu beeinflussen. Dabei muss jedoch berücksichtigt werden, dass Farrington et al. (2012) in ihrem systematischen Review zwar eine große Anzahl an Faktoren berücksichtigen, jedoch keine vollumfängliche Systematisierung nicht-kognitiver Merkmale vornehmen. So fällt zum Beispiel auf, dass neben vielen Einzelfaktoren, die sich jedoch zum Teil den Bereichen nachträglich zuordnen lassen, größere Bereiche wie Emotionen oder Wohlbefinden der Schüler:innen nicht in die Untersuchung Eingang gefunden haben. Zudem liegt der Fokus der Arbeit auf Schüler:innen von der Mittelstufe bis hin zur postsekundären Bildung, sodass die Grundschulzeit nicht miteinbezogen wurde und somit die Passung des Modells bzw. die Relevanz der einzelnen Bereiche nicht unhinterfragt angenommen werden kann. So zeigen beispielsweise Schmalfuß (2022) und Wolke (2008), dass Absentismus, der nach Farrington et al. (2012) einflussreichste Faktor der Kategorie *Academic Behaviors*, in der Grundschule ein deutlich geringeres Problem darstellt als in weiterführenden Schulen und somit im Grundschulbereich deutlich weniger bedeutsam sein dürfte als in diesem Modell angenommen.

Trotz der Kritikpunkte konnte in der Literatur keine passendere oder auch ähnlich umfangreiche Arbeit zu nicht-kognitiven Faktoren gefunden werden. Einzig ein Literaturreview von Gutman und Schoon (2013) soll ergänzend vorgestellt werden, da in diesem ähnliche Merkmale wie bei Farrington et al. (2012) betrachtet werden, dabei jedoch die gesamte Spanne der Schulzeit untersucht wird. Es bietet damit die Möglichkeit, das vorgestellte Modell zumindest zum Teil um eine Gültigkeit für die Grundschule zu ergänzen. Gutman und Schoon (2013) konzentrieren sich in ihrer Untersuchung auf die veränderlichen Faktoren Selbstwahrnehmungen (Fähigkeitsselbstkonzept und Selbstwirksamkeit), Motivation (Zielsetzung, intrinsische und extrinsische Motivation, Erwartungs- und Wertkomponenten), Ausdauer (Engagement und Grit), Selbstkontrolle, Metakognitive Strategien, Soziale Kompetenzen (Führungskompetenzen und soziale Fähigkeiten), Resilienz und Coping sowie Kreativität. Sie kommen zu dem Schluss, dass das Fähigkeitsselbstkonzept, die Erfolgserwartungen und die Wertzuschreibungen als Komponenten der Motivation sowie das Durchhaltevermögen für die Leistungen am bedeutsamsten sind. Damit stellen sie übereinstimmend mit Farrington et al. (2012) die Bedeutung von Merkmalen, die sich nach deren Klassifikation den *Academic Mindsets* zuordnen lassen, heraus. Dabei wird in beiden Publikationen explizit auf die Veränderbarkeit und damit die Förderbarkeit

der Merkmale des *Academic Mindsets* hingewiesen. Sie treffen somit den Anspruch der vorliegenden Arbeit, veränderliche Merkmale in den Fokus zu stellen, um Ansatzpunkte für eine Förderung besonderer Leistungsstärke zu identifizieren.

In der Zusammenschau der Untersuchungen stellten sich somit insbesondere Merkmale der Mindsets bzw. Bereiche ‚My ability and competence grow with my effort' bzw. Fähigkeitsselbstkonzept, ‚I can succeed at this' bzw. Erfolgserwartungen und ‚This work has value for me' bzw. Wertzuschreibungen als bedeutsame Prädiktoren schulischer Leistungen heraus. Ordnet man die Merkmale dieser Mindsets in den Gesamtkontext der pädagogisch-psychologischen Forschung ein, zeigt sich, dass damit insbesondere Merkmale, die im Zusammenhang mit den Erwartung x Wert-Theorien[9] stehen, als besonders bedeutsam für schulische Leistungen herausgearbeitet wurden. In verschiedenen Studien, die im Rahmen der Erwartung- x Wert-Theorien entstanden, konnte die Bedeutung von Erwartungskomponenten wie dem Selbstkonzept und der Selbstwirksamkeit sowie von Wert-Komponenten wie intrinsischer Motivation, Einstellung oder Lernfreude übereinstimmend mit Farrington et al. (2012) und Gutman und Schoon (2013) nachgewiesen werden (Afari, 2015; Kriegbaum et al., 2018; Lipnevich, Preckel & Krumm, 2016; Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005).

Das Verhältnis von Wertkomponenten zu Erwartungskomponenten hinsichtlich ihrer Bedeutsamkeit für schulische Leistungen ist in der aktuellen Forschung dabei noch nicht eindeutig geklärt. Verschiedene Studien kommen zu dem Schluss, dass Erwartungskomponenten zwar etwas stärker mit der Leistung von Schüler:innen zusammenhängen als Wertkomponenten (Kriegbaum et al., 2018; Ozel, Caglak & Erdogan, 2013; Steinmayr & Spinath, 2009). Da sich die untersuchten Konstrukte jedoch stark zwischen den Studien unterscheiden und einerseits somit kein vollständiges Bild gezeichnet werden kann und andererseits der Theorie nach beide Komponenten gemeinsam Einfluss nehmen, scheint es sinnvoll, Merkmale aus beiden Komponenten heranzuziehen. Einer der in der Forschung prominentesten Faktoren aus dem

9 Den Erwartung x Wert-Theorien zufolge hängen Leistung maßgeblich von der Motivation zur Durchführung ab. Die Motivation wiederum ergibt sich unter der Prämisse limitierter Ressourcen aus dem Abwägen von Erwartungen an eine bestimmte Handlung (Erfolgswahrscheinlichkeit) und dem Wert, der dieser zugesprochen wird. Das Resultat beider Abwägungen beeinflusst, wie sehr in das Erreichen dieses Ziels investiert wird (Schuster, 2017). Seit den Anfängen dieser Theorien, die häufig bei Atkinson (1957, 1964) verortet wird, wurden zahlreiche Theorie-Varianten und Weiterentwicklungen publiziert, die aber in Kern diesen Grundgedanken verfolgen (z. B. Eccles & Wigfield [2002]; Feather [1992]; Heckhausen [1991]).

Bereich Erwartungskomponente ist das akademische Selbstkonzept von Schüler:innen. Für dieses konnte bereits in zahlreichen Studien ein bedeutsamer Zusammenhang mit schulischen Leistungen belegt werden, vereinzelt wiesen Ergebnisse darauf hin, dass das Selbstkonzept ähnlich erklärungskräftig sein kann wie das Vorwissen oder die Intelligenz (Hattie, 2009; Seaton, Parker, Marsh, Craven & Yeung, 2013; Steinmayr & Spinath, 2009). Bei den Wertkomponenten ist hingegen ein viel beachteter Faktor das Interesse. Krapp (2010) resümierte jedoch, dass eindeutige Zusammenhänge zwischen Interesse und Leistung häufig nicht nachgewiesen werden können. Er begründet dies damit, dass die schulische Realität mit ihren festgelegten Inhalten und Überprüfungen der Entwicklung inhaltlicher Interessen entgegensteht. Ein vielversprechendes, schulnaheres Konstrukt könnte stattdessen die Einstellung von Schüler:innen zu den verschiedenen Unterrichtsfächern sein. Dieses vereint unterschiedliche Facetten der Wertkomponente wie intrinsische Werte oder Freude in sich und zeigte in bisherigen Studien bereits bedeutsame Zusammenhänge mit der Leistung (Hemmings & Kay, 2010; Lipnevich et al., 2016; Teig & Nilsen, 2022).

Dafür sprechen auch die wenigen differenziellen Befunde, die zu besonderer Leistungsstärke vorliegen. Untersuchungen in diesem Bereich liegen insbesondere für die Merkmale Selbstkonzept, intrinsische und extrinsische Motivation, Freude und Einstellung vor und zeigen in hoher Übereinstimmung, dass leistungsstarke Schüler:innen sich durch eine ausgeprägtere intrinsische und extrinsische Motivation, eine größere Freude am Thema und eine positivere Einstellung zum Fach auszeichnen (Abu-Hamour & Al-Hmouz, 2013; Bos et al., 2012; Kartal & Kutlu, 2017; Kibrislioglu, 2015; Neuendorf, Kuhl & Jansen, 2017). Eine der wenigen Untersuchungen, die mehrere Konstrukte parallel betrachtet, stammt aus Irland. Sie kommt anhand von Analysen von *Large-Scale*-Daten zu dem Schluss, dass sowohl Einstellungen als auch Selbstwahrnehmungen bedeutsam für besondere Leistungsstärke sind (Pitsia, 2022). Die hohe Bedeutsamkeit des Selbstkonzeptes für besondere Leistungsstärke konnte ebenfalls in bereits mehreren Studien gezeigt werden (Bos et al., 2012; Neuendorf et al., 2017; Schmidtner, 2017; Tuorón, López-González, Hernandéz, García San Pedro & Navarro Ascencio, 2018). Dings und Spinath (2021) fanden zudem, dass dem Selbstkonzept nochmal eine größere Bedeutung zukommt als der Selbstwirksamkeit.

Das Selbstkonzept und die Einstellung zum Fach scheinen somit nach den bisherigen Befunden vielversprechende Prädiktoren von Schulleistungen und besonderer Leistungsstärke zu sein. Beide Konzepte werden in den zwei Folgekapiteln sowohl theorie- als auch empiriebezogen genauer vorgestellt.

Bei der Sichtung der Literatur zu differenziellen Befunden fallen jedoch auch die Hintergrundmerkmale als häufig untersuchte Prädiktoren ins Auge. Dies kann nach Farrington et al. (2012) damit begründet werden, dass Hintergrundmerkmale alle anderen nicht-kognitiven Persönlichkeitsmerkmale und das gesamte Lernen rahmen und beeinflussen. Bei den untersuchten Merkmalen handelt es sich i. d. R. um die Hintergrundmerkmale, die sich bereits in Studien zu generellen Prädiktoren von Schulleistungen als bedeutsam herausgestellt haben. So konnte gezeigt werden, dass die kognitiven Grundfähigkeiten von leistungsstarken Schüler:innen mindestens durchschnittlich und im Vergleich zu weniger leistungsstarken Schüler:innen signifikant höher sind (Schmidtner, 2017). Ähnliches gilt auch für den sozioökonomischen Hintergrund. In verschiedenen Untersuchungen bestätigte sich, dass sich die Gruppe der leistungsstarken Schüler:innen häufiger durch einen hohen sozioökonomischen Hintergrund oder ein höheres Bildungsniveau der Eltern auszeichnet (Neuendorf et al., 2017; Pitsia, 2022; Schmidtner, 2017; Wendt, Willems, Tarelli, Euen & Bos, 2013). Seltener untersucht ist hingegen die Bedeutung des Migrationshintergrundes und des häuslichen Sprachgebrauchs. In einer Untersuchung von Wendt et al. (2013) zeigte sich jedoch, dass die Subgruppe der leistungsstarken Schüler:innen seltener einen Migrationshintergrund aufweist als die Schülerschaft in ihrer Gesamtheit. Der häusliche Sprachgebrauch wird in einer Untersuchung von Gilleece, Cosgrove und Sofroniou (2010) neben anderen Hintergrundmerkmalen untersucht. Es zeigt sich, dass der Sprachgebrauch zwar bedeutsam für besondere Leistungsstärke ist, alle untersuchten Hintergrundmerkmale jedoch relevanter für die Gruppe der leistungsschwachen Schüler:innen sind. Neben dem sozioökonomischen Hintergrund scheint das Geschlecht zu den am besten beforschten Hintergrundmerkmalen in diesem Feld zu gehören. Die *greater male variability hypothesis* (Ellis, 1894) legte den Grundstein für die Annahme, dass Jungen in den Extremgruppen häufiger vorkommen als Mädchen. Dies konnte inzwischen in verschiedenen Studien für verschiedene Fächer, aber insbesondere für das Fach Mathematik mit Befunden von bis zu 70 % Jungen unter den leistungsstärksten Schüler:innen, belegt werden (Forgasz & Hill, 2013; Lee & Burkam, 1996; Wendt et al., 2013).

Bei den bisherigen Befunden zu Einflussfaktoren besonderer Leistungsstärke müssen jedoch verschiedene Aspekte berücksichtigt werden:

1. Die meisten der rezipierten Studien basieren auf *Large-Scale*-Daten und sind somit querschnittlicher Natur. Aussagen über kausale Zusammenhänge lassen sich deshalb mit diesen Untersuchungen nicht treffen, sodass nur Ver-

mutungen über die Wirkrichtungen gemacht werden können. Dafür verfügen diese Daten i. d. R. über sehr große, repräsentative Stichproben und besondere Ziehungsverfahren, die generalisierte Aussagen über die Schülerschaft in den jeweils untersuchten Staaten zulassen.

2. Die meisten Untersuchungen mit differenziellem Ansatz stammen aus den letzten zehn Jahren. Dabei sind komplexere Analysen, die über Deskriptiva und Gruppenvergleiche hinausgehen, erst zunehmend unter den neueren Publikationen zu finden. Die Datenlage für belastbare, multivariate Informationen ist dementsprechend noch eher gering.
3. Die meisten der Untersuchungen basieren auf Daten von Schüler:innen der Sekundarstufe. Befunde für die Grundschule liegen hingegen bisher nur in geringer Zahl vor.
4. Die Gruppe der leistungsstarken Schüler:innen wird in den einzelnen Studien teils sehr unterschiedlich definiert, je nachdem wie die Leistungen der Schüler:innen erhoben wurden. So gibt es beispielsweise Studien, in denen Prozentkriterien gewählt wurden oder Studien, in denen vorab definierte Leistungsgruppen genutzt wurden. Darüber hinaus unterscheidet sich auch die Definition der Referenzgruppe in den verschiedenen Untersuchungen. Teilweise ist dies die Gesamtheit der Schüler:innen, die nicht als leistungsstark definiert wurden, teilweise eine weitere anhand eines Kriteriums definierte, ausgewählte Gruppe.

All diese Aspekte nehmen Einfluss auf die konkreten Ergebnisse der einzelnen Studien. Eine Beurteilung dieses Forschungsfeldes ist somit zum jetzigen Zeitpunkt aufgrund der geringen Anzahl an bisherigen Studien sowie der fehlenden Vergleichbarkeit der Ergebnisse nur sehr eingeschränkt möglich. Dies muss auch bei den differenziellen Befunden berücksichtigt werden, die in den nachfolgenden Kapiteln aufgeführt werden. Es kann nach den bisherigen Befunden jedoch vermutet werden, dass das Selbstkonzept und die Einstellung zum Fach, aber auch die kognitiven Fähigkeiten, der sozioökonomische Hintergrund, der Migrationshintergrund sowie das Geschlecht geeignete Prädiktoren von besonderer Leistungsstärke sind.

4.1.1 Fachbezogenes Selbstkonzept

Das Selbstkonzept gehört zu einem der bekanntesten und bestuntersuchtesten Merkmalen von Schüler:innen (Buch, Sparfeldt & Rost, 2019; Hellmich & Günther, 2011; Köller, Trautwein, Lüdtke & Baumert, 2006; Marsh, Byrne & Shavelson, 1988; Shavelson, Hubner & Stanton, 1976). Darunter werden die subjektiven Vorstellungen, Gedanken und Bewertungen der eigenen Person verstanden. Dabei lässt sich das generelle Selbstkonzept in verschiedene Facetten unterteilen, die sich auf unterschiedliche Bereiche des Lebens beziehen, z. B. auf das Äußere, auf kognitive Fähigkeiten oder auf die Empathiefähigkeit. Diese bereichsspezifischen Selbstkonzepte lassen sich wiederum in weitere Unterselbstkonzepte differenzieren, sodass davon ausgegangen wird, dass das generelle Selbstkonzept multidimensional und hierarchisch aufgebaut ist. Eine viel rezipierte Konzeptualisierung des Selbstkonzeptes wird als Shavelson-Modell bezeichnet (Möller & Trautwein, 2020; Shavelson et al., 1976; siehe Abbildung 4).

Für die vorliegende Arbeit ist das Fähigkeitsselbstkonzept (auch leistungsbezogenes oder akademisches Selbstkonzept) relevant. Für dieses findet sich eine Unterteilung in das verbale und das mathematische Selbstkonzept, denen wiederum orientiert an den schulischen Unterrichtsfächern verschiedene fachspezifische Selbstkonzepte zugeordnet werden (Marsh et al., 1988; Marsh & Craven, 2006; Shavelson et al., 1976). Dabei wird basierend auf verschiedenen Untersuchungen davon ausgegangen, dass sich die verschiedenen Selbstkonzepte, insbesondere das verbale und das mathematische Selbstkonzept, schon im frühen Schulalter ausbilden und differenzieren lassen (Arens, Trautwein & Hasselhorn, 2011; Eccles, Wigfield, Harold & Blumenfeld, 1993; Hellmich & Günther, 2011). Dies impliziert gleichzeitig, dass es sich bei dem Selbstkonzept um ein veränderbares Konstrukt handelt.

Das (akademische) Selbstkonzept formt sich vor allem über Vergleichsprozesse. Als Erklärungsansatz hat sich dafür früh das *Internal/External-Frame-of-Reference-Model* durchgesetzt (Marsh, 1986). Nach diesem werden Selbstkonzepte sowohl durch interne als auch externe Vergleiche beeinflusst. Erstere können durch den Vergleich eigener Leistungen in verschiedenen Bereichen oder zu verschiedenen Zeitpunkten verstanden werden, beispielsweise indem bei der Herausbildung des verbalen Selbstkonzeptes die wahrgenommene Leistungsfähigkeit im mathematischen Bereich als Referenz herangezogen wird. Mit externen Vergleichen hingegen werden beispielsweise Vergleiche der eigenen Leistungen mit denen anderer Personen, wie beispielsweise den Mitschüler:innen, umschrieben. Zusätzlich zu diesen beiden Quellen kön-

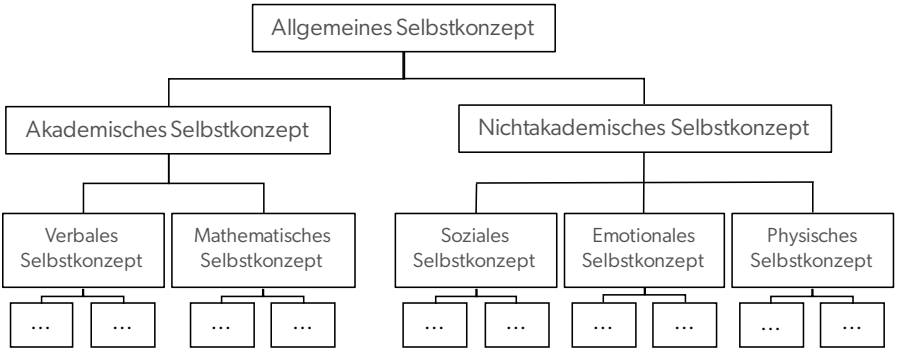


Abb.: 4 Struktur des Selbstkonzeptes nach Marsh et al. (1988)

nen aber auch Rückmeldungen von Bezugspersonen wie Eltern oder Lehrkräften, Kausalattribuierungen, bisherige Erfahrungen und das Selbstvertrauen Einfluss auf das Selbstkonzept haben (Gebauer, 2013; Hellmich & Günther, 2011; Marsh, 1986; Skaalvik, 1997). Gleichzeitig bedeutet das, dass das akademische Selbstkonzept nicht nur von den eigenen Leistungen abhängt, sondern auch von den Bezugsgruppen, die für Vergleiche herangezogen werden. Dieser Umstand wird auch als Referenzgruppeneffekt bezeichnet. Zu den bekanntesten Referenzgruppeneffekten gehören der *Big-Fish-Little-Pond-Effekt* (BFLPE) und der *Basking-in-Reflected-Glory-Effekt* (BIRGE). Beim BFLPE wird davon ausgegangen, dass die durchschnittliche Leistung der Referenzgruppe, i. d. R. der Klasse, negativ mit dem eigenen Selbstkonzept zusammenhängt. Das heißt, dass das Selbstkonzept einer Person in einer leistungsstarken Klasse niedriger ausfällt, als es unter gleichen Voraussetzungen in einer leistungsschwachen Gruppe der Fall wäre, weil der Vergleichsrahmen, der an der Formung des Selbstkonzeptes beteiligt ist, andere Informationen liefert (Marsh, 1987; Marsh & Parker, 1984). Der BIRGE hingegen geht davon aus, dass das Selbstkonzept in einer leistungsstarken Klasse sogar ansteigen kann, da sich die Person einer besonderen Gruppe zugehörig fühlt und somit ihre eigenen Fähigkeiten höher bewertet (Cialdini et al., 1976).

Bisherige Forschungen gehen davon aus, dass bei der Betrachtung der gesamten Schülerschaft der BFLPE gegenüber dem BIRGE überwiegt, wobei sich dies auch in der deutlich größeren Anzahl an Forschungen zum BFLPE widerspiegelt (Becker & Birkelbach, 2017; Köller, 2004; Marsh, Kong & Hau, 2000). Für beide Effekte konnten allerdings kaum differenzielle Untersuchungen zu besonderer Leistungsstärke aus-

gemacht werden. Dennoch zeigte sich in den wenigen, existierenden Studien bereits für mehrere Kontexte eine geringere Bedeutung des BFLPE für besonders leistungsstarke Schüler:innen als für die Gesamtheit aller Schüler:innen. Gleichzeitig wurde in diesen Studien – sowohl für Schüler:innen in Spezialklassen für leistungsstarke oder hochbegabte Schüler:innen als auch für Schüler:innen in Regelklassen – häufig eine gleichwertige oder eher höhere Bedeutung des BIRGE festgestellt (Cunningham & Rinn, 2007; Köller, 2004; Marsh et al., 2000; Preckel & Brüll, 2010). Ausnahme bildet eine Studie von Zeidner und Schleyer (1999), die für israelische Schüler:innen in Spezialklassen für leistungsstarke bzw. begabte Schüler:innen ein geringeres Selbstkonzept fanden als für die Schüler:innen in gemischten Klassen. Unabhängig von der fokussierten Personengruppe wird auch davor gewarnt, dem sehr populären BFLPE als nur einer von vielen Informationsquellen zu viel Bedeutung in der Entwicklung von Selbstkonzepten beizumessen.

> "Our main argument is that the BFLPE, while having added to our understanding of the origins of self-concepts, disproportionately emphasizes one aspect of social comparison to the exclusion of many other intervening factors (Dai & Rinn, 2008, S. 283)."

Neben der Entstehung des akademischen Selbstkonzeptes steht aber vor allem dessen Zusammenhang mit der Leistung im Zentrum der Forschung. Dabei finden sich in den meisten Fällen signifikante positive Zusammenhänge mit schwachen bis mittleren Werten, je nachdem wie die Konstrukte erfasst und ob mehrere Zeitpunkte betrachtet wurden (beispielsweise Hansford & Hattie, 1982 mit r = .21 - .42; Helmke & van Aken, 1995 mit bis zu r = .27; Huang, 2011 mit r = .19 - .27; Marsh et al., 2018 mit r = .30 - .62; Marsh et al., 2005 mit r = .30 - .40), bei Hattie (2009) stellte sich das akademische Selbstkonzept sogar als einer der wichtigsten Einflussfaktoren auf Leistung heraus. Übereinstimmend konnten dabei höhere Zusammenhänge für Noten als für Testleistungen gefunden werden (siehe dafür auch Jansen, Schroeders & Lüdtke, 2014; Schicke & Fagan, 1994). Vereinzelte Befunde zum Zusammenhang von akademischem Selbstkonzept und Leistung für leistungsstarke Schüler:innen legen nahe, dass diese einerseits für Mathematik häufiger über ein hohes Selbstkonzept berichten als ihre Mitschüler:innen, andererseits dass der Zusammenhang zwischen beiden Merkmalen mit zunehmendem Fähigkeitsniveau stärker ausfällt (Drüke-Noe, Gniewosz & Paasch, 2022; Hansford & Hattie, 1982).

Wie beide Merkmale kausal miteinander zusammenhängen, ist ein vieldiskutierter Aspekt in der Selbstkonzeptforschung. Dabei sind vor allem zwei Ansätze vorherrschend: der *Skill-Development*-Ansatz und der *Self-Enhancement*-Ansatz. Während nach dem *Self-Enhancement*-Ansatz davon ausgegangen wird, dass das akademische Selbstkonzept Einfluss auf die Leistung nimmt, postuliert der *Skill-Development*-Ansatz die gegenläufige Wirkrichtung, also von der Leistung auf das akademische Selbstkonzept (Möller & Trautwein, 2020). Insgesamt deutet die bisherige Befundlage eher auf das Vorherrschen des *Skill-Development*-Ansatzes hin (Helmke & van Aken, 1995; Retelsdorf, Köller & Möller, 2014; Skaalvik & Valås, 1999). So konnte in verschiedenen Studien für verschieden Fächer und Altersstufen gezeigt werden, dass die Leistung vornehmlich Einfluss auf das akademische Selbstkonzept hat. Allerdings gibt es eine fast ähnlich starke Befundlage, die darauf hindeutet, dass die Annahme des *Self-Enhancement*-Ansatzes zutreffend ist (Köller et al., 2006; Steinmayr & Spinath, 2009). Die meisten Untersuchungen finden Hinweise auf beide Wirkrichtungen, weshalb sich in der jüngeren Vergangenheit die Annahme eines reziproken, also wechselseitigen Zusammenhangs zwischen beiden Merkmalen durchgesetzt hat (Arens et al., 2017; Huang, 2011; Marsh & Martin, 2011; Marsh & O'Mara-Eves, 2008). Ebenfalls als relativ gesichert gilt, dass das akademische Selbstkonzept zu Beginn der Schulzeit im Mittel sehr hoch und in der Regel überhöht ausgeprägt ist. Im Laufe der Grundschulzeit wird es dann realistischer und sinkt ab, wobei es im Durchschnitt aber trotzdem im positiven Bereich verbleibt (Donat, Radant & Dalbert, 2017; Hellmich & Günther, 2011; Helmke, 1998). Dabei wird davon ausgegangen, dass das akademische Selbstkonzept und die Leistung mit zunehmendem Alter immer stärker zusammenhängen. Das bedeutet gleichzeitig, dass das Grundschulalter noch gut für die Förderung des akademischen Selbstkonzeptes geeignet ist, während in der Sekundarstufe beide Konstrukte bereits so stark zusammenhängen, dass eine Förderung eher schwer möglich ist (Helmke, 1992; Praetorius et al., 2016; Spinath & Spinath, 2005).

Bei getrennter Betrachtung der reziproken Zusammenhänge für verschiedene Altersstufen zeigt sich ein etwas differenzierteres Bild darüber, welche Wirkrichtung in welchen Phasen überwiegt. Für den Grundschulbereich konnte häufig die Dominanz des *Skill-Development*-Ansatzes nachgewiesen werden, wobei die Annahme reziproker Zusammenhänge jedoch prinzipiell bestehen bleibt. Diese Dominanz kann mit dem überhöhten akademischen Selbstkonzept der Schüler:innen zu Beginn der Grundschulzeit erklärt werden, das wie beschrieben durch erbrachte Leistungen im Laufe der ersten Schuljahre korrigiert und realistischer wird (Helmke, 1998; Helmke

& van Aken, 1995; Praetorius et al., 2016; Skaalvik & Valås, 1999). Mit zunehmendem Alter scheint sich dieser Zusammenhang jedoch zu verändern, sodass im späteren Schulverlauf häufig der *Self-Enhancement*-Ansatz vorherrscht. Grund dafür könnten Selbstbestätigungsmechanismen sein, nach denen durch Leistung versucht wird, das bestehende akademische Selbstkonzept zu bestätigen (Dalbert & Stöber, 2008; Marsh, 1990; Retelsdorf et al., 2014). Wann genau der Wechsel in der Dominanz stattfindet, ist jedoch nicht eindeutig. Während viele Studien den Wechsel zeitlich mit dem Übergang von der Grundschule in die weiterführende Schule verorten, gibt es auch Studien, die zu dem Schluss kommen, dass der *Skill-Development*-Ansatz bereits im Laufe der Grundschulzeit nicht mehr vorherrschend ist (Martschinke & Kammermeyer, 2006; Weidinger, Steinmayr & Spinath, 2017).

Wie bereits zuvor in diesem Kapitel erwähnt, hängt der Zusammenhang von Selbstkonzept und Leistung auch von der Art der Konzeptualisierung ab. So konnte vielfach gezeigt werden, dass der Zusammenhang zwischen beiden Konstrukten umso höher ausfällt, desto spezifischer diese erfasst werden (Buch et al., 2019; Marsh & Craven, 2006). Damit schließt sich der Kreis zum Shavelson-Modell:

> "[…] the three-level Shavelson model of self-concept is important for any understanding of relationships between self-concept and achievement. Without distinguishing levels of self-concept we could not interpret the relationships. The level most related to achievement is the most specific (that is, discipline-specific) (Hoge, Smit & Crist, 1995, S. 312)."

Eine Besonderheit stellt dabei das naturwissenschaftliche Selbstkonzept, das für die vorliegende Arbeit relevant ist, dar. Da die Naturwissenschaften aus mehreren Disziplinen zusammengesetzt sind, finden sich unterschiedliche Konzeptualisierungen für das domänenspezifische Selbstkonzept: einerseits als eindimensionales Konstrukt, andererseits als multidimensionales Konstrukt mit Subfacetten für die Domänen Biologie, Physik und Chemie. Dabei zeigen sich überwiegend Hinweise auf eine höhere Passung des multidimensionalen Konstruktes gegenüber einem allgemeinen naturwissenschaftlichen Selbstkonzept (für eine umfangreiche Aufarbeitung für die Sekundarstufe siehe Jansen et al., 2014).

Da in Deutschland Naturwissenschaften in der Grundschule anders als in der Sekundarstufe als integratives Fach (je nach Bundesland beispielsweise als Sachunterricht, Sachkunde oder Heimatkunde bezeichnet) unterrichtet werden, ist allerdings

fraglich, inwiefern diese Ergebnisse auf den Grundschulbereich übertragbar sind, in dem die Schüler:innen die einzelnen Fächer i. d. R. nicht klar voneinander getrennt kennenlernen. Stattdessen scheint es wahrscheinlicher, dass in der Grundschule die Erfassung eines unterrichtsnahen Selbstkonzeptes vielversprechender ist. Dies zeigt auch die Konzeptualisierung des domänenspezifischen Selbstkonzeptes in verschiedenen Studien in der Grundschule, die zumeist das globale Sachunterrichtsselbstkonzept oder themenspezifische Selbstkonzepte, die sich auf konkrete Unterrichtseinheiten beziehen, untersuchen (Blumberg, 2008; Lossen, Tillmann, Holtappels, Rollett & Hannemann, 2016; Valtin, Wagner & Schwippert, 2005).

4.1.2 Fachbezogene Einstellung zum Sachunterricht

Wie im Kapitel zu den generellen Einflussfaktoren auf Ebene der Schüler:innen bereits dargestellt, ist die Einstellung ein Konstrukt, das verschiedene Facetten der Wertkomponente in sich vereint. Im Vergleich zum Selbstkonzept ist für Einstellung jedoch sowohl die theoretische als auch die empirische Datenlage im erziehungswissenschaftlichen Bereich deutlich weniger umfangreich und konsistent. Dies zeigt sich bereits bei dem Versuch, eine einheitliche Definition für Einstellung auszumachen. Es scheinen in der bisherigen Forschung vor allem zwei Definitionsansätze vorherrschend zu sein: beim ersten Ansatz wird Einstellung als (Bereitschaft für) konsistente Reaktionen und Handlungen auf eine Situation verstanden (z. B. Campbell, 1963; Seiffge-Krenke, 1974), der zweite Ansatz hingegen ist weitergefasst und bezieht Einstellungen nicht nur auf verhaltensbezogene, sondern auch auf affektive und kognitive Komponenten (z. B. Maio & Haddock, 2010). Diese zweite Gruppe spiegelt die Annahmen des Drei-Komponenten-Modells der Einstellung nach Rosenberg und Hovland (1960) wider (siehe Abbildung 5).

Nach diesem Modell entstehen Einstellungen als Reaktionen auf Reize (‚Einstellungsgegenstände'), wobei sich die Einstellung in die drei Komponenten Affekt, Kognition und Verhalten unterteilen lässt. Die affektive Komponente der Einstellung zeigt sich beispielsweise über das autonome Nervensystem oder verbale Äußerungen über Bewunderung oder Angst und verkörpert vor allem die Gefühle und Emotionen, die mit dem Einstellungsgegenstand verbunden sind. Annahmen und Einschätzungen bezüglich des Einstellungsgegenstandes, wie Wahrnehmungsurteile und (verbal geäußerte) Überzeugungen (*beliefs*), sind hingegen Teil der kognitiven Komponente. Der

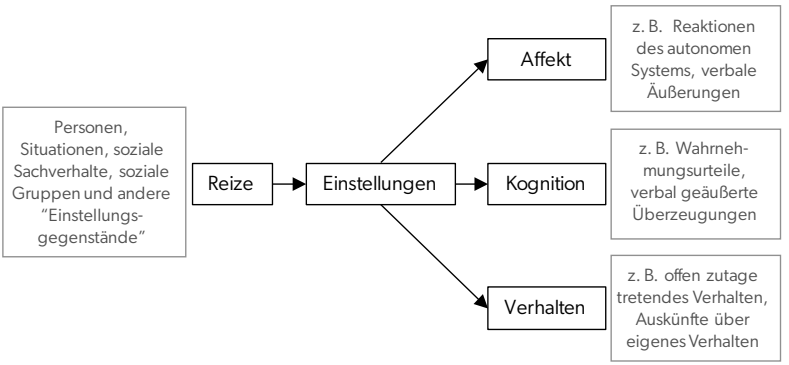


Abb.: 5 Drei-Komponenten-Modell der Einstellung nach Rosenberg und Hovland (1960)

Verhaltensanteil der Einstellung drückt sich wiederum durch (potenzielles) beobachtbares Verhalten oder Auskünfte darüber aus. Einstellungen können dabei bezüglich Stärke, Richtung bzw. Valenz (positive, neutrale oder negative Einstellung) und Konsistenz (beispielsweise zwischen den einzelnen Komponenten) variieren (Ajzen, 2005; Haddock & Maio, 2014; Landwehr, 2017; Rosenberg & Hovland, 1960). Die Bedeutung von Einstellungen ergibt sich aus ihrer Funktion zur Orientierung. Sie sorgen dafür, dass die große Anzahl an Umweltreizen entlang der bestehenden Einstellungen selektiert wird und haben somit eine „wahrnehmungsformierende Wirkung" (Fischer & Wiswede, 2009, S. 322; siehe auch Grigutsch, Raatz & Törner, 1998; Schröder, 2001). Dabei werden Einstellungen zwar als relativ stabile Eigenschaften angesehen, jedoch gelten sie als beeinflussbar, insbesondere durch bestimmte Ereignisse oder wenn beispielsweise neue Informationen zur Verfügung stehen (Ajzen, 2005).

Wie Fischer und Wiswede (2009) herausarbeiten, unterstützen bisherige Befunde die dreikomponentige Struktur jedoch nur teilweise. So finden sich häufig deutlich höhere Korrelationen zwischen der affektiven und kognitiven Komponente als angenommen, während die Verhaltenskomponente in besonderem Maße situationsabhängig zu sein scheint. Deshalb werden inzwischen eher Definitionen, in denen die Einstellung nicht mehr in Komponenten unterteilt bzw. diese getrennt behandelt werden oder zweikomponentige Definitionen, die nur Affekte und Kognitionen umfassen, vorgezogen. Dabei stimmen jedoch alle Definitionen darin überein, dass Einstellungen Bewertungen von Einstellungsgegenständen sind. Welche Annahmen darüber getrof-

fen werden, wie diese zusammengesetzt sind, hängt dann stark von dem jeweiligen Kontext ab, in dem Einstellungen betrachtet werden (z. B. politische Einstellungen oder Einstellungen zur Schule).

Eine inzwischen häufig genutzte Definition im Schul- und Unterrichtskontext stammt von Helmke (1993), der die Einstellung als „die affektiv (negativ oder positiv) getönte Orientierung gegenüber dem Lernen […] oder gegenüber bestimmten Fächern" beschreibt (zitiert nach Brühwiler & Helmke, 2018, S. 80). Demnach spielt also für die fachbezogene Einstellung insbesondere die affektive Komponente der Einstellung mit Gefühlen bzw. Emotionen eine Rolle. Gleichzeitig bedeutet dies aber nicht, dass sich die Einstellung nicht auch durch Kognitionen, die ohnehin stark mit Affekten zusammenhängen, oder Verhalten ausdrücken kann – sie sind nur nicht (mehr) zwingend notwendig, um fachbezogene Einstellung näher zu bestimmen. In der empirischen, insbesondere naturwissenschaftlichen, Forschung wird darüber hinaus eine hohe Nähe von Einstellung und intrinsischer Motivation angenommen. Operationalisiert wird Einstellung deshalb häufig über Komponenten der Lernfreude und des Interesses (Teig & Nilsen, 2022).

Die meisten bisherigen Untersuchungen kommen zu dem Schluss, dass schwache bis moderate positive Zusammenhänge zwischen Einstellung und Leistung vorliegen (Ai, 2002; Oliver & Simpson, 1988; Tuner & Yilmaz, 2020; Vachon, 1986). Vereinzelt finden sich jedoch auch schwache negative (Ozel et al., 2013) oder deutlich höhere Zusammenhänge. Chen et al. (2018) beispielsweise finden für Mathematik hohe, signifikante Effekte der fachbezogenen Einstellung auf die Leistung, wobei diese auch unter Kontrolle von Alter, IQ, Arbeitsgedächtnis und Mathematikangst signifikant bleiben. Sie resümieren, dass in Mathematik die fachbezogene Einstellung somit einer der bedeutsamsten Prädiktoren von Leistung ist. Auch Weinburgh (1995) kommt in ihrer Metaanalyse zu vergleichsweise hohen Werten von $r = .50$ bzw. $r = .55$ je nach Geschlecht für den Zusammenhang zwischen beiden Konstrukten in den Naturwissenschaften. Die meisten anderen Metaanalysen, die sich auf Naturwissenschaften fokussieren, finden hingegen niedrigere Zusammenhänge, beispielsweise $r = .16$ bei Willson (1983) oder $r = .25$ bei Mao, Cai, He, Chen und Fan (2021). Zusammenhänge zwischen fachbezogener Einstellung und Leistung konnten auch in *Large-Scale*-Studien bereits wiederholt für die Naturwissenschaften nachgewiesen werden, wobei sich in Analysen mit TIMSS-Daten zusätzlich gezeigt hat, dass die fachbezogenen Einstellungen höher ausfallen, wenn Naturwissenschaften als integriertes Fach und nicht als separate Fächer unterrichtet werden (Martin et al., 2000; OECD, 2016; Steffensky et al., 2020).

Zudem hat sich in verschiedenen Untersuchungen herausgestellt, dass der Zusammenhang zwischen der fachbezogenen Einstellung und der Leistung in der Grundschule zwar noch relativ schwach ist, bis zur 10. Klasse jedoch ansteigt. In der Abiturstufe nimmt der Zusammenhang wieder ab (Ma & Kishor, 1997; Mao et al., 2021; Willson, 1983). Das bedeutet – ähnlich wie bereits für das Selbstkonzept gezeigt –, dass in der Grundschulzeit der Grundstein für die späteren Einstellungen der Schüler:innen gelegt wird, diese allerdings noch veränderbar und gut förderbar sind (Upmeier zu Belzen, 2007). Deutlich weniger untersucht ist hingegen die Wirkrichtung im Zusammenhang zwischen fachbezogener Einstellung und Leistung. Die meisten Studien gehen von Einflüssen der Einstellung auf die Leistung oder reziproken Effekten aus (Chen et al., 2018; Grootenboer & Hemmings, 2007; Hattie, 2009). Vereinzelt finden sich jedoch auch Hinweise auf die umgekehrte Wirkrichtung (Willson, 1983). In vielen Studien konnten zudem neben signifikanten Zusammenhängen mit der Leistung auch signifikante moderate Korrelationen von fachspezifischem Selbstkonzept und fachspezifischer Einstellung festgestellt werden (Ozel et al., 2013; Vachon, 1986).

Befunde, die die fachbezogene Einstellung gezielt für leistungsstarke Schüler:innen untersuchen sind hingegen äußerst selten. Eine Studie von Grootenboer und Hemmings (2007) konnte für Mathematik zeigen, dass die fachbezogene, positive Einstellung nicht nur bedeutsam mit der Leistung zusammenhängt (r = .36) und damit relevanter als andere Merkmale wie *utilitarian belief* (r = .20) und Selbstkonzept (r = .23) ist, sondern auch dafür geeignet ist, die Zugehörigkeit von Schüler:innen zur Gruppe der leistungsstarken Schüler:innen besser vorherzusagen, als die Zugehörigkeit zur Gruppe der leistungsschwachen Schüler:innen. Auch zwischen Einstellung und Selbstkonzept findet sich für leistungsstarke Schüler:innen eine signifikante moderate Korrelation.

Die teilweise eher schwachen Zusammenhänge zwischen der fachbezogenen Einstellung und der Leistung, die sich auch in Metaanalysen zeigen, erklären Ma und Kishor (1997) damit, dass die Einstellung in den bisherigen Studien teils sehr unterschiedlich und mit unzureichenden Instrumenten erhoben wird. Dazu passt auch die Diskussion um die Trennbarkeit von fachbezogenen Einstellungen und der Einstellung zur unterrichtenden Lehrkraft (Kibrislioglu, 2015). Hier ist nicht immer eindeutig, ob von Schüler:innen beides klar voneinander getrennt werden kann, weshalb die Operationalisierung der fachbezogenen Einstellung bei Untersuchungen besonders berücksichtigt werden muss.

4.2 Merkmale des Unterrichts

Wie in Kapitel 3 herausgearbeitet, zählen Unterrichtsmerkmale neben Schülermerkmalen und Lehrkräftemerkmalen zu den bedeutsamsten Einflussfaktoren schulischer Leistungen. In der bisherigen Forschung fanden zahlreiche Unterrichtsmerkmale Eingang in verschiedene Untersuchungen, so beispielsweise Häufigkeit des Unterrichts, Lernzeit, Klima, Lehrer-Schüler-Beziehung, Unterrichtsqualität, Organisation des Unterrichts, Sozialformen oder Differenzierungsformen (Fend, 1977; Hattie, 2009; Klieme & Rakoczy, 2008; Kulik & Kulik, 1992; Moser & Angelone, 2011; Seidel & Shavelson, 2007). Ziel der Untersuchung solcher Merkmale ist zumeist die Identifikation von Faktoren, die guten Unterricht ausmachen, und herauszufinden, wie diese mit den verschiedenen Zielsetzungen des Unterrichts zusammenhängen. Nicht wenige Studien erstellen als Resultat ihrer Forschung Listen mit Eigenschaften, die ein Unterricht aufweisen muss, um als gut eingestuft zu werden (beispielsweise Brophy, 2000; Helmke & Weinert, 2017; Meyer, 2003).

Grundlegend ist jedoch zunächst die Frage, wie guter Unterricht überhaupt definiert werden kann. Eine inzwischen sehr etablierte Konzeptualisierung lieferte Berliner (1987, 2005). Nach dieser muss ‚qualitätsvoller' Unterricht, wie er ihn bezeichnet, zwei Bedingungen erfüllen: Zum einen muss Unterricht effektiv sein, d. h. er muss dazu führen, dass anvisierte Ziele wie beispielsweise Lernerfolg oder eine positive motivationale Entwicklung der Schüler:innen erreicht werden. Zum anderen muss Unterricht vor dem Hintergrund aktueller, regionaler Werte und Standards, aus einer normativen, gesellschaftlichen Perspektive als gut eingestuft werden. Erfüllt ein Unterricht beide Kriterien, kann er als qualitätsvoll gelten (Kunter & Ewald, 2016).

Die unterschiedlichen Merkmale zur Beschreibung von Unterricht können in verschiedene Perspektiven eingeordnet werden (Kunter & Trautwein, 2013; Willems, 2016). Auf der einen Seite gibt es Faktoren, die sich der Quantität von Unterricht zuordnen lassen (z. B. Anzahl der Unterrichtsstunden oder aktive Lernzeit), auf der anderen Seite Merkmale, die die Qualität des Unterrichts beschreiben (z. B. Strukturiertheit, Lehrer-Schüler-Beziehung). Andererseits lässt sich Unterricht jedoch auch durch Prozess- (Interaktionen im Unterricht, Motivierungsqualität) oder Produktmerkmale (z. B. Lernerfolg) charakterisieren oder in Sichtstrukturen (Organisations- oder Sozialformen) und Tiefenstrukturen (kognitive Aktivierung, Verstehensprozesse) von Unterricht einteilen. Dabei widersprechen sich die einzelnen Perspektiven nicht und verfolgen auch keinen Anspruch darauf, einen Rahmen für alle potenziellen Unterrichtsmerk-

male bieten zu können. Merkmale können also beispielsweise gleichzeitig die Qualität des Unterrichts beschreiben und zu den Prozessmerkmalen gezählt werden.

In bisherigen Untersuchungen fanden besonders Sicht- und Tiefenstrukturen Beachtung, wobei die Bedeutung der Tiefenstrukturen für qualitätsvollen Unterricht und die Leistungen von Schüler:innen vielfach herausgestellt wurde (Clausen, Reusser & Klieme, 2003; Seidel & Shavelson, 2007; Wang et al., 1993), weshalb diese auch in erwähnten Listen und Merkmalskatalogen für qualitätsvollen Unterricht am stärksten vertreten sind. Während unter Sichtstrukturen des Unterrichts all jene Merkmale gezählt werden, die methodisch-organisatorischer Natur sind, die das Unterrichtsgeschehen strukturieren und somit ‚sichtbar' sind, werden unter Tiefenstrukturen die Merkmale verstanden, die die tieferliegenden Komponenten des Unterrichts wie die tatsächlichen Interaktionen oder die Verstehens- und Lernprozesse beschreiben (Bohl, 2017). Tiefenstrukturen werden dabei häufig auch als Merkmale der Unterrichtsqualität bezeichnet. Die Unterrichtsqualität stellt somit eine zentrale Bedingung für gelingenden Unterricht und die schulischen Leistungen von Schüler:innen dar und scheint ein vielversprechendes Konstrukt für besondere Leistungsstärke zu sein. Bisherige Forschungen zur Unterrichtsqualität werden im nachfolgenden Kapitel ausführlicher behandelt.

Ein Merkmal, das bislang nicht eindeutig den Sicht- oder Tiefenstrukturen zugeordnet werden konnte, ist Differenzierung. Diese wird in verschiedenen Publikationen mal als Sicht, mal als Tiefenstruktur beschrieben (Hess & Lipowsky, 2016; Inckemann, 2014; Kunter & Ewald, 2016; Praetorius et al., 2018). Da Differenzierung jedoch die Förderung von anhand bestimmter Kriterien (i. d. R. Leistung) definierter Gruppen von Schüler:innen meint, birgt Differenzierung per definitionem ein hohes Potenzial für die Entwicklung besonderer Leistungsstärke. Deshalb wird zusätzlich zur Unterrichtsqualität auch Differenzierung im Folgenden genauer betrachtet.

4.2.1 Unterrichtsqualität

Unterrichtsqualität kann definiert werden als „stable pattern of instructional behavior which (as a whole or using single components) allow for substantive prediction and/or explanations of the achievement of educational goals by students" (Weinert, Schrader & Helmke, 1989, S. 899). Diese Definition legt den Schwerpunkt auf die Komponente ‚effektiver Unterricht', wobei Weinert, Schrader et al. (1989) bereits selbst auf die Schwierigkeiten bei der Definition von Unterrichtsqualität aufgrund der Komplexität

des Konstruktes sowie der Vielzahl an assoziierten Merkmalen hinweisen. Ähnlich wie Berliner (2005) kommen Klieme und Tippelt (2008) zwar zu dem Schluss, dass Unterrichtsqualität auch eine normative Komponente beinhaltet, allerdings scheint der Fokus in der Forschung zur Unterrichtsqualität vor allem auf dem Prozesscharakter von Unterrichtsqualität und dessen Wirksamkeit zu liegen. Gleichzeitig resümieren sie ebenfalls, dass die Definition von Unterrichtsqualität Schwierigkeiten mit sich zu bringen scheint und in vielen Publikationen deshalb eher auf Umschreibungen mithilfe anderer Konstrukte gesetzt wird.

Zur näheren Beschreibung von Unterrichtsqualität haben sich aufgrund der skizzierten Komplexität statt Definitionen vor allem Modelle und Systematisierungen (je nach Publikation beispielsweise als Faktoren, Prinzipien oder Dimensionen beschrieben) etabliert. Diese verfolgen zumeist einen hierarchischen Ansatz, indem sie verschiedene Kriterien unter verschiedenen Überbegriffen bündeln und diese teils noch zueinander in Beziehung setzen (Heinitz & Nehring, 2020). Etablierte Systematisierungen von Unterrichtsqualität sind beispielsweise das QuAIT-Modell nach Slavin (2000) und das CLASS-Modell nach Pianta und Hamre (2009). Beim QuAIT-Modell wird davon ausgegangen, dass es vier Faktoren der Unterrichtsqualität gibt: *quality of instruction* (Präsentation der Inhalte, sodass sie gut erlernt werden können), *appropriateness* (Angemessenheit des Schwierigkeitslevels), *incentives* (Fähigkeit die Schüler:innen zu motivieren) und *time* (Möglichkeiten effektiv zu lernen). Damit werden ausschließlich Merkmale in das Modell integriert, die von der Lehrkraft beeinflusst und bestimmt werden können, sodass die Qualität von Unterricht also vor allem von der unterrichtenden Lehrkraft abhängt. Eine ähnliche Perspektive nimmt auch das CLASS-Modell ein, wobei dies zusätzlich die Interaktionen zwischen Lehrkräften und Schüler:innen miteinbezieht. Es fokussiert sich dabei auf die drei übergeordneten Domänen Klassenführung, emotionale Unterstützung und Lernunterstützung. Diesen werden wiederum verschiedene Dimensionen zugeordnet, die durch konkrete Indikatoren verkörpert werden. Damit ist das CLASS-Modell gleichzeitig ein Instrument zur Beobachtung von Unterricht, das zur Bestimmung der Qualität von Unterricht sowie Rückmeldung dazu eingesetzt werden kann (Pianta & Hamre, 2009).

Eine andere Perspektive wird hingegen mit dem gestuften Modell der Unterrichtsqualität vertreten (Pietsch, 2010). Dieses ordnet die einzelnen Komponenten der Unterrichtsqualität nicht nebeneinander an, sondern geht davon aus, dass diese aufeinander aufbauende Niveaustufen darstellen. Unterricht, der auf der ersten Stufe angesiedelt ist, zeichnet sich dadurch aus, dass er die grundsätzlichen Voraussetzun-

gen für gelingenden Unterricht (Herstellung eines lernförderlichen Klimas und Einhaltung grundlegender Strukturen und Regeln) erfüllt. Auf der nachfolgenden Stufe wird zusätzlich zu Regeln eine effiziente Klassenführung mit Grundzügen von Differenzierung wie beispielsweise der Anpassung des Tempos realisiert. Auf der dritten Stufe ist Unterricht verortet, der sich zusätzlich durch eine hohe Motivierungsqualität, Mitbestimmungsmöglichkeiten für die Schüler:innen sowie transparente Diagnostik von Lernständen und Weiterentwicklungsoptionen auszeichnet. Auf der höchsten Stufe wird darüber hinaus in hohem Maße differenziert bzw. individualisiert und den Schüler:innen Freiraum für selbstgewählte Lerninhalte und Reflexion gegeben (Pietsch, 2010).

Als besonders bedeutsam hat sich insbesondere im deutschsprachigen Raum, aber auch in der internationalen Forschung das Modell der drei Basisdimensionen von Unterrichtsqualität durchgesetzt (Klieme et al., 2001). Dieses wurde im Rahmen der TIMSS-Videostudie 1995 entwickelt, bei der der Mathematikunterricht von u. a. 100 Klassen in Deutschland videografiert wurde. In den meisten dieser Klassen bearbeiteten die Schüler:innen zusätzlich einen Mathematiktest, einen Fragebogen zu Interesse und fachbezogener Einstellung und genau wie die Lehrkräfte Fragen zum Unterricht, sodass umfangreiche Analysen zum Zusammenhang von Unterricht, Leistungsergebnissen und motivationalen Merkmalen der Schüler:innen möglich waren. Dabei ergaben explorative Faktorenanalysen drei übergeordnete Basisdimensionen von Unterrichtsqualität: Klassenführung, kognitive Aktivierung und konstruktive Unterstützung (Klieme et al., 2001; Praetorius et al., 2018). Breite Akzeptanz findet dieses Modell, weil es, wie Praetorius et al. (2018) herausarbeiten, bereits in zahlreichen Untersuchungen für verschiedene Fächer und Altersstufen genutzt und bestätigt werden konnte. Dass die drei Basisdimensionen, obwohl anhand des Mathematikunterrichtes entwickelt, auf andere Fächer übertragen werden können, liegt darin begründet, dass die drei Dimensionen generisch, also fachunspezifisch, angelegt sind.

Inwiefern Unterrichtsqualität allerdings überhaupt fachunspezifisch operationalisiert werden kann oder ob Unterrichtsqualität immer fachspezifisch betrachtet werden sollte, ist ein eigener Diskurs im Rahmen der Forschung zur Unterrichtsqualität. Ausgangspunkt dieses Diskurses bildet die Annahme, dass es Merkmale gibt, die in allen Unterrichtsfächern gleichermaßen von Bedeutung sind, während andere in hohem Maße fachspezifisch sind. Klassifiziert werden können diese über das Fachwissen der Lehrkräfte. Während für die Umsetzung generischer Unterrichtsqualitätsmerkmale kein spezifisches Fachwissen von Nöten ist, zeichnen sich fachspezifische Kriterien

dadurch aus, dass sie nur mit entsprechenden Fachkenntnissen umgesetzt werden können (Wüsten, 2010). Dabei scheint sich zunehmend die Einschätzung durchzusetzen, dass die Generik bzw. Fachspezifität eines Merkmals zu großen Teilen auf der Ebene der Operationalisierung bestimmt wird (Heinitz & Nehring, 2020; Seidel, Renkl & Rieß, 2021). In einem Syntheseframework systematisieren Praetorius und Charalambous (2018) verschiedene Konzeptualisierungen der Unterrichtsqualität und identifizieren drei Kategorien von Merkmalen: Merkmale der ersten Kategorie (G) sind vor allem generisch, Merkmale der zweiten Kategorie (S) vor allem fachspezifisch. Darüber hinaus gibt es jedoch auch eine ganze Reihe an Merkmalen, die sich nicht primär einer der beiden Kategorien zuordnen lassen. Diese werden in einer dritten Kategorie (G*S) als integrativ für beide Perspektiven angesehen (Praetorius & Charalambous, 2018). Heinitz und Nehring (2020) ergänzen diese Aufstellung noch um eine vierte Kategorie (G+S), die Merkmale enthält, die hauptsächlich generisch sind, sich jedoch auch durch fachspezifische Komponenten auszeichnen. Damit schränken sie die Kategorie (G*S) gleichzeitig auf die Merkmale ein, die beide Perspektiven zu ähnlichen Anteilen enthalten.

Wie fachspezifische Anteile auf der Ebene der Operationalisierung genau aussehen können, verdeutlichen Heinitz und Nehring (2020) über drei verschiedene Typen der Spezifikation. Die Fokussierungs-Spezifik erfolgt durch die Ausrichtung generischer Merkmale auf einen konkreten, fachspezifischen Unterrichtsgegenstand. Theoretisierungs-Spezifik wird hingegen dadurch erreicht, dass generische Merkmale in eine fachspezifische Theorie bzw. Terminologie überführt werden. Der dritte Typ ist die genuine Fachspezifik, die Merkmale umfasst, die nicht Teil generischer Überlegungen sind, sondern ausschließlich in fachspezifischen Kontexten vorkommen (Heinitz & Nehring, 2020). Helmke und Schrader (2008) resümieren zum Verhältnis von Generik und Fachspezifität: „eine angemessene Beurteilung der Unterrichtsqualität erfordert in der Tat die Berücksichtigung von beidem: fachübergreifender und fachspezifischer Merkmale“ (S. 32), womit sie zu einer ähnlichen Einschätzung kommen wie Seidel et al. (2021), die zusammenfassen, dass es bei der Beurteilung von Unterrichtsqualität „immer beides [ist]“ (S. 296). Wie Unterrichtsqualität angemessen beurteilt werden kann, berührt jedoch darüber hinaus auch die Frage, wie Informationen über den Unterricht generiert werden.

Die häufigsten Formen zur Erfassung der Unterrichtsqualität sind Befragungen von Schüler:innen oder Lehrkräften oder Beobachtungen durch Außenstehende. Als besonders verlässliche Quellen gelten Beobachtungen des Unterrichts, die durch

externe und speziell geschulte Personen durchgeführt werden (Clausen, 2002; Göllner et al., 2016; Praetorius, 2013). Allerdings ist dieses Verfahren im Vergleich zu Befragungen sehr zeitaufwändig und teuer, weshalb in vielen Studien stattdessen Informationen von Lehrkräften und Schüler:innen herangezogen werden. Lehrkräfte als Informationsquelle bieten sich bei der Erfassung von Unterrichtsqualität wiederum besonders an, weil sie den Unterricht aktiv gestalten und durch ihre professionellen Kompetenzen in der Lage sind, unterschiedliche Facetten des Unterrichts getrennt voneinander zu betrachten. Allerdings bedeutet dies auch, dass Lehrkräfte stark in den Unterricht involviert sind, über den sie Auskunft geben sollen, weshalb fraglich ist, wie objektiv sie diesen beurteilen können (Göllner et al., 2016; Henke, Bosse & Spörer, 2018). Alternativ können auch Aussagen von Schüler:innen zur Einschätzung der Unterrichtsqualität genutzt werden. Diese können für die Beurteilung von Unterricht einen längeren Zeitraum heranziehen und auch insbesondere Merkmale wie kognitive Aktivierung, die für externe Beobachter:innen und Lehrkräfte nicht direkt ersichtlich sind, einschätzen. Zudem ist die Befragung sehr ökonomisch, insbesondere, wenn alle Schüler:innen einer Klasse befragt werden. Inwiefern Schüler:innen hingegen verschiedene Aspekte des Unterrichts voneinander abgrenzen und einschätzen können, ist nicht eindeutig belegt (Helmke, 2007; Henke et al., 2018; Stahns, Rieser & Hußmann, 2020; Wagner, Göllner, Helmke, Trautwein & Lüdtke, 2013). In verschiedenen Studien konnte allerdings wiederholt gezeigt werden, dass Schüler:innen als Expert:innen für den Unterricht angesehen werden können (Rutter, Maughan & Mortimore, 1980) und sehr wohl in der Lage sind, Qualitätsmerkmale separat einzuschätzen (Fauth, Decristan, Rieser, Klieme & Büttner, 2014b; Kloss, 2014). Da es explizit Befunde dafür gibt, dass dies auch schon für Kinder im Grundschulalter gilt, werden Schülerbefragungen sehr häufig zur Erfassung der Unterrichtsqualität in allen Phasen der Schulzeit eingesetzt. Ein Vorteil, den Schüleraussagen zusätzlich bieten, ist, dass sie nicht nur auf individueller Ebene betrachtet werden können, sondern auch in aggregierter Form für eine objektivere Erfassung des Unterrichts auf Klassenebene genutzt werden können (Helmke, 2007; Lüdtke, Robitzsch, Trautwein & Kunter, 2009). Allerdings zeigen Untersuchungen auch, dass die Einschätzungen von Beobachter:innen, Lehrkräften und Schüler:innen zum Teil stark voneinander abweichen (de Jong & Westerhof, 2001; Reibnegger & Nausner, 2018). Die Auswahl der Informationsquellen kann somit einen bedeutsamen Einfluss auf die Ergebnisse von Untersuchungen haben, sodass die Auswahl mit Bedacht und vor dem Hintergrund der konkreten Forschungsfragen getroffen werden muss.

Während beim CLASS-Modell die Herkunft der Informationen zur Einschätzung der Unterrichtsqualität bereits in der Systematisierung festgelegt ist (Pianta & Hamre, 2009), ist dies bei den meisten anderen Modellen und so auch bei dem Modell der drei Basisdimensionen offen. Deshalb wurden im Rahmen der TIMSS-Videostudie 1995 die Einschätzungen von externen Beobachter:innen, von Lehrkräften und von Schüler:innen parallel erhoben. In einer umfangreichen Untersuchung dieser Daten kam Clausen (2002) zu dem Schluss, dass sich die Einschätzungen der verschiedenen Informationsquellen auch bei den drei Basisdimensionen voneinander unterscheiden. Zusätzlich fand er unterschiedliche Facetten bei der Einschätzung des Unterrichts durch Schüler:innen, anders als in den meisten anderen Untersuchungen zeigten sich jedoch ebenfalls Hinweise auf globale, wenig differenzierte Beurteilungen.

Da es sich bei den drei Basisdimensionen der Unterrichtsqualität wie zuvor beschrieben um eine der etabliertesten und breit erforschtesten Konzeptualisierungen handelt, wird diese auch in dieser Arbeit als grundlegend angesehen. Die Basisdimensionen Klassenführung, kognitive Aktivierung und konstruktive Unterstützung werden im Folgenden näher erläutert.

Klassenführung

Die erste Basisdimension, die Klieme et al. (2001) identifizieren konnten, wird als Klassenführung bezeichnet. Darunter fallen „alle Handlungen und Strategien, die dazu dienen, Ordnungsstrukturen im Klassenzimmer herzustellen und aufrechtzuerhalten" (Kunter & Trautwein, 2013, S. 78). Alternative Begrifflichkeiten aus der deutsch- und englischsprachigen Forschung sind beispielsweise Klassenmanagement oder *classroom management*, wobei diese keine konzeptuellen Unterschiede zur Klassenführung aufweisen (Reibnegger & Nausner, 2018). Ursprünglich wurde der Begriff *classroom management* von Kounin (1976) etabliert, der Unterricht in zahlreichen Studien auf dieses Merkmal hin untersuchte. Seitdem hat sich die Forschung zur Klassenführung stark verbreitet. Der Dimension Klassenführung werden zwei Hauptziele zugeschrieben: zum einen erwünschtes Verhalten bei Schüler:innen zu fördern, zum anderen unerwünschtes Verhalten zu minimieren (Hochweber, Hosenfeld & Klieme, 2014). Dazu gibt es in der aktuellen Literatur eine ganze Zahl an Indikatoren, die Zeichen für gute Klassenführung darstellen. Zu diesen zählen u. a. der Umgang mit und die Prävention von Störungen und Disziplinproblemen, klare Regeln, Disziplin im Klassenzimmer, Kontrolle bzw. Überwachung des Unterrichtsgeschehens,

Aufbau von Routinen, Strukturierung des Unterrichts, flüssige Übergänge zwischen verschiedenen Phasen des Unterrichts, angemessenes Zeitmanagement, gute Vorbereitung der Lernmaterialien, klare Aufgabenstellungen und Schritte während des Unterrichts, plausible Gliederung der Inhalte, eindeutige Rückmeldestrukturen, effektive Nutzung der Lernzeit, Gestaltung von Lernsituationen, Ermöglichen von konzentrierten Lernaktivitäten, Flexibilität/Adaptivität, Umgang mit parallelen Ereignissen im Klassenzimmer, Fokussierung auf die ganze Gruppe, Eingehen auf individuelle Bedürfnisse der Schüler:innen, verständliche, eindeutige (Körper-)Sprache und klare Rollendefinition (Brühwiler et al., 2017; Decristan et al., 2016; Helmke, 2007; Klieme et al., 2001; König et al., 2023; Künstning, Neuber & Lipowsky, 2016; Meyer, 2003; Milles & Jansen, 2021; Neuenschwander, 2006; Praetorius et al., 2018; Schmal, 2020).

Damit zählt Klassenführung zu den eher generischen Merkmalen der Unterrichtsqualität, da sie Verhalten und Strukturen umfassen, die in jedem Unterricht unabhängig von den tatsächlichen Inhalten benötigt werden (Heinitz et al., 2022; Reusser & Pauli, 2021). Wird eine Klasse effektiv geführt, ermöglicht dies eine effiziente Beschäftigung mit den Fachinhalten und eine größtmögliche Nutzung der zur Verfügung stehenden Unterrichtszeit. Deshalb wird Klassenführung häufig auch als Grundvoraussetzung für gelingenden Unterricht und die Umsetzung von verschiedenen Methoden und Maßnahmen beschrieben (Helmke & Schrader, 2008; Milles & Jansen, 2021; Stahns et al., 2020). Wenig überraschend konnte deshalb schon in verschiedensten Studien ein Zusammenhang zwischen Klassenführung bzw. Facetten der Klassenführung und den Leistungen von Schüler:innen nachgewiesen werden (Dubberke, Kunter, McElvany, Brunner & Baumert, 2008; Fauth, Decristan, Rieser, Klieme & Büttner, 2014b; Seidel & Shavelson, 2007).

Im Rahmen des Projektes *Cognitive Activation in the Classroom: The Orchestration of Learning Opportunities for the Enhancement of Insightful Learning in Mathematics (COACTIV)* wurden Zusammenhänge zwischen den Basisdimensionen der Unterrichtsqualität und kognitiven sowie motivational-affektiven Zielkriterien im Mathematikunterricht untersucht. Für die Klassenführung zeigte sich auf Basis aggregierter Schüleraussagen ein signifikanter, positiver Zusammenhang mit der Leistung in Höhe von $r = .26$ (Kunter & Voss, 2011). Ähnliche Ergebnisse finden sich bei Fauth, Decristan, Rieser, Klieme und Büttner (2014a), in deren Untersuchung sich signifikante, positive Zusammenhänge zwischen der Leistung im Sachunterricht und der Klassenführung von $r = .30$ bis $r = .46$ zeigten, je nachdem wessen Perspektive zur Einschätzung der Klassenführung herangezogen wird. In einer Zusammenschau

verschiedener Studien zum Zusammenhang zwischen Klassenführung und Leistung, in die auch Ergebnisse von COACTIV und der Studie von Fauth, Decristan, Rieser, Klieme und Büttner (2014a) eingeflossen sind, wurden überwiegend signifikante positive Zusammenhänge berichtet, die größtenteils als schwach bis moderat eingestuft werden können (Praetorius et al., 2018). Eine ganze Reihe Studien konnte jedoch auch deutlich höhere Zusammenhänge nachweisen. Ein mit $r = .62$ vergleichsweise hoher, positiver Zusammenhang zwischen Leistung und effizienter Klassenführung wurde beispielsweise in einer Untersuchung zur Leseleistung am Ende der Grundschulzeit gefunden (Helmke et al., 2010). Dies passt auch zu den Ergebnissen der Metaanalyse von Wang et al. (1993), die eine überdurchschnittlich hohe Bedeutung der Klassenführung herausstellten. Es finden sich jedoch auch einige Studien, die keinen Zusammenhang zwischen Klassenführung und Leistung nachweisen konnten (Henschel, Rjosk, Holtmann & Stanat, 2019; Roßbach, 2002). Roßbach (2002) vermutet, dass ein Grund für fehlende Zusammenhänge das Alter in der Stichprobe sein könnte. Möglicherweise ist die Klassenführung während der Anpassungsphase in der Grundschule noch nicht so bedeutsam, sondern wird erst ab der 3. Klassenstufe relevant für das schulische Lernen, weshalb sich erst mit zunehmendem Alter Zusammenhänge zeigen.

Im Sinne multikriterialer Zielsetzungen von Unterricht werden in viele Untersuchungen von Unterrichtsqualität neben der Leistung auch motivational-affektive Merkmale einbezogen. Praetorius et al. (2018) fanden in ihrer Untersuchung verschiedener Studien uneinheitliche Ergebnisse für den Zusammenhang von Klassenführung und motivationalen Merkmalen. So gab es zwar signifikante positive, aber auch negative Effekte. Signifikante, negative Effekte von moderater Stärke fanden sich dabei vor allem zwischen der Klassenführung und fachbezogenen Selbstkonzepten, nicht signifikante bis schwache signifikante, positive Ergebnisse fanden sich hingegen vor allem für Interesse und Lernfreude (Praetorius et al., 2018).

Für Freude und Interesse konnten diese Ergebnisse auch bei der Untersuchung von *Large-Scale*-Daten gefunden werden. Bei der Analyse von PISA-Daten von 2012 zeigten sich beispielsweise signifikante, positive Zusammenhänge beider Merkmale mit der Klassenführung (Schiepe-Tiska et al., 2016), ein Befund, den auch Klieme et al. (2001) bereits 15 Jahre zuvor bei der Entwicklung des Modells der drei Basisdimensionen von Unterrichtsqualität zeigen konnten. Mit Daten des IQB-Bildungstrends 2018 konnte hingegen ein negativer Zusammenhang mit Interesse ebenso wie mit dem Selbstkonzept festgestellt werden. Dabei zeigte sich, dass Unterrichtsqualitätsmerkmale auf Klassenebene vor allem für Leistung bedeutsam sind, für motivatio-

nal-affektive Merkmale Unterrichtsqualitätsmerkmale auf Individualebene (Henschel et al., 2019; siehe auch Fauth, Decristan, Rieser, Klieme & Büttner, 2014b). Anders als in den zuvor beschriebenen Untersuchungen konnten in einigen Studien auch positive Zusammenhänge zwischen Klassenführung und Selbstkonzept nachgewiesen werden (Milles & Jansen, 2021; Schiepe-Tiska et al., 2016; Seidel & Shavelson, 2007). Milles und Jansen (2021) vermuten, dass eine effektive Klassenführung mehr Zeit für Lernprozesse ermöglicht, die selbstkonzeptförderliche Merkmale aufweisen und deshalb auch positive Zusammenhänge zwischen beiden Merkmalen vorkommen. Während für die Leistungen also über verschiedene Fächer und Altersklassen hinweg der Zusammenhang mit der Klassenführung relativ gut belegt ist, ist die Befundlage für motivationale Merkmale deutlich uneinheitlicher. Es scheinen jedoch Befunde zu überwiegen, die einen negativen Zusammenhang der Klassenführung mit dem Selbstkonzept und positive Zusammenhänge mit Lernfreude und Interesse annehmen.

Im Bereich leistungsdifferenzieller Befunde für Klassenführung wurden bisher hauptsächlich leistungsschwache Schüler:innen in den Blick genommen. In diesen Studien wird meist die Annahme vertreten, dass schwächere Schüler:innen stärkere Strukturierung und Vorgaben benötigen und weniger in der Lage sind, eine mangelnde Klassenführung zu kompensieren als ihre leistungsstarken Mitschüler:innen (Lipowsky, 2006; Rakoczy, Klieme, Lipowsky & Drollinger-Vetter, 2010). Dadurch entstand gleichzeitig das Bild, dass Klassenführung ein Merkmal sei, dass insbesondere für diese Gruppe bedeutsam ist. Differenzielle Befunde, die diese Annahme für die Gegenseite, also für die geringere Bedeutsamkeit der Klassenführung für besondere Leistungsstärke, belegen, gibt es bisher jedoch kaum.

Eine der wenigen Studien stammt von Lüftenegger et al. (2015), die Unterschiede zwischen leistungsstarken und nicht-leistungsstarken Schüler:innen, allerdings im Kontext hoher mathematischer Begabung, untersuchen. Damit sind die Ergebnisse nur begrenzt auf die vorliegende Arbeit übertragbar, da die Gruppeneinteilung für eine selektierte Stichprobe erfolgte. Dennoch können sie Hinweise auf mögliche Zusammenhänge liefern. Lüftenegger et al. (2015) kamen zu dem Schluss, dass die leistungsstarken Begabten sowohl höhere motivationale Merkmale wie Selbstkonzept und Interesse aufwiesen als auch ihren Unterricht anders einschätzten. Dabei berichteten sie insbesondere von einer höher wahrgenommenen Passung von Aufgaben und Fähigkeitsniveau und einer höheren Eingebundenheit in unterrichtliche Entscheidungsprozesse – Merkmale, die dem Bereich der Klassenführung zugeordnet werden könnten. Eine weitere Studie untersucht am Beispiel von Äthiopien und Englisch als

erster Fremdsprache, inwiefern die Basisdimensionen der Unterrichtsqualität für verschiedene Leistungsgruppen bedeutsam sind. Dabei finden sie für Klassenführung keinen Einfluss auf die Leistungen und das für alle Fähigkeitsgruppen gleichermaßen (Sanfo & Malgoubri, 2021).

Kognitive Aktivierung

Die zweite Basisdimension der Unterrichtsqualität ist die kognitive Aktivierung. Diese kann definiert werden als „Prozess der Induktion eines Problems beim Lernenden" (Minnameier, Hermkes & Mach, 2015, S. 842), d. h., dass bei den Schüler:innen die Bereitschaft hervorgerufen werden soll, sich selbstständig, aktiv und tiefgreifend mit den Unterrichtsgegenständen auseinander zu setzen (Kunter & Trautwein, 2013). Die kognitive Aktivierung umfasst drei Hauptmerkmale: das Formulieren von herausfordernden Aufgaben, das Anknüpfen an das Vorwissen der Schüler:innen und das Führen fachbezogener Unterrichtsgespräche sowie Argumentationen (Lipowsky et al., 2009).

Neben diesen übergeordneten Merkmalen gibt es in der Forschung zahlreiche Indikatoren, welche Merkmale kognitiv aktivierender Unterricht aufweisen kann. Kognitiv aktivierender Unterricht zeichnet sich beispielsweise aus über eine genetisch-sokratische Herangehensweise, anspruchsvolle anstatt repetitive Übungen, Motivierung der Schüler:innen, komplexe Aufgabenstellungen mit multiplen Lösungswegen, Auslösen kognitiver Konflikte, Differenzierung von Lernangeboten, herausfordernde Aufgaben stellen, die nicht durch einfaches Abrufen bekannter Informationen lösbar sind (*higher order thinking)*, Schüler:innen Konzepte und Ideen entdecken lassen, Aufgaben stellen, deren Bearbeitung sich aus Sicht der Schüler:innen lohnt, Möglichkeiten für die Entwicklung eigener Lösungswege bieten, Aufgaben, die das Vorwissen und die Lebensrealität der Lernenden einbeziehen, Lernprozesse auf hohem, aber zu den Schüler:innen passendem Niveau, Adaptivität, um unterschiedlichen Lernausgangslagen gerecht zu werden, Interpretieren von Problemstellungen und Suchen relevanter Informationen zur Lösung einer Aufgabe, auf Unterschiede in Ideen, Konzepten, Positionen, Interpretationen und Lösungen hinweisen, Fragenstellen von Schüler:innen fördern, Schüler:innen die Reflexion des eigenen Lernprozesses ermöglichen und Schüler:innen ihre Antworten erklären lassen (Fauth & Leuders, 2018; Klieme & Rakoczy, 2008; Klieme et al., 2001; Künstning et al., 2016; Kunter & Voss, 2011; Lipowsky et al., 2009; Lotz, 2016; Schmal, 2020; Seidel et al., 2021).

Damit wird auch deutlich, dass kognitive Aktivierung nicht mit hoher allgemeiner Aktivität, z. B. beim Ausführen praktischer Tätigkeiten wie dem Durchführen eines Experiments, gleichzusetzen ist (Fauth & Leuders, 2018). Stattdessen steht die vertiefte Beschäftigung mit den Inhalten des Unterrichtsfaches im Vordergrund. Dementsprechend wird kognitive Aktivierung auch als deutlich fachspezifischer als die Klassenführung angesehen, auch wenn sie nach Klieme et al. (2001) als generische Dimension definiert worden ist. Wie Praetorius und Charalambous (2018) und Heinitz und Nehring (2020) mit ihren Systematisierungen gezeigt haben, ist dies jedoch kein Widerspruch per se. Merkmale der Unterrichtsqualität können beispielsweise generisch angelegt und für alle Fächer generell bedeutsam sein, aber dann fachspezifisch ausgestaltet werden.

Ein entscheidender Aspekt der Basisdimension kognitive Aktivierung ist, dass sie – anders als die Klassenführung – ausschließlich über die Eigeninitiative der Schüler:innen realisiert werden kann. Zwar können Lehrkräfte ihren Unterricht in allen Aspekten möglichst kognitiv anregend gestalten und so die Chancen auf tiefgehende Denkprozesse erhöhen. Ob dieses Angebot (im Sinne des Angebots-Nutzungs-Modells) an die Schüler:innen jedoch von diesen angenommen wird, liegt nicht im Einflussbereich der Lehrkräfte. Deshalb wird in der Literatur anstelle der Bezeichnung ‚kognitive Aktivierung' häufig auch die Formulierung ‚Potenzial zur kognitiven Aktivierung' genutzt, um dieses Verhältnis deutlich werden zu lassen (Fauth & Leuders, 2018).

Dadurch, dass ein kognitiv aktivierender Unterricht zu einer höheren kognitiven Auseinandersetzung mit dem Lerngegenstand und vertieften Denkprozessen bei den Schüler:innen beitragen soll, wird vermutet, dass so ein besseres Verständnis der Thematik und bessere schulische Leistungen erreicht werden können (Lipowsky et al., 2009). Diese Annahme konnte auch bereits in verschiedenen Studien für verschiedene Fächer und Altersklassen belegt werden (Dubberke et al., 2008; Fraser et al., 1987; Hattie, 2009; Kunter et al., 2013; Li, Liu, Zhang & Liu, 2020; Lipowsky et al., 2009; Schiepe-Tiska et al., 2016; Stahns et al., 2020; Teig & Nilsen, 2022), so auch in verschiedenen *Large-Scale*-Studien wie TIMSS, PISA oder dem IQB-Bildungstrend: Im Rahmen der Analysen zu PISA 2000, während der das Modell der Basisdimensionen entwickelt wurde, konnten Klieme et al. (2001) einen Zusammenhang von $r = .22$ zwischen kognitiver Aktivierung und Leistung nachweisen. Im IQB-Bildungstrend 2022 fanden sich hingegen signifikante negative Zusammenhänge von kognitiver Aktivierung mit den Leistungen im Fach Deutsch sowie dem fachbezogenen Selbstkonzept, wobei diese nur unter Betrachtung der individuellen Perspektive auftraten.

Für die geteilte Wahrnehmung der kognitiven Aktivierung zeigten sich keine bedeutsamen Zusammenhänge. Für Interesse konnten darüber hinaus weder auf Basis von Individual- noch auf Klassenebene signifikante Zusammenhänge gefunden werden (Henschel, Rjosk & Heinschel, 2023). In TIMSS 2019 wurde ebenfalls im Rahmen von Mehrebenenmodellen der Zusammenhang zwischen kognitiver Aktivierung und Unterrichtszielen untersucht. Es zeigte sich, dass ebenfalls nur Zusammenhänge auf der Individualperspektive signifikant wurden. Sowohl für Mathematik als auch Naturwissenschaften zeigten sich dabei negative Zusammenhänge, die sich unter Kontrolle verschiedener Hintergrundmerkmale nicht mehr zeigten (Stang, Lepper, Steffensky & McElvany, 2020).

Ebenfalls Hinweise auf einen Zusammenhang zwischen naturwissenschaftlichen Leistungen und kognitiver Aktivierung fanden beispielsweise Förtsch, Werner, Dorfner, von Kotzebue und Neuhaus (2016). In den untersuchten Modellen lag der Zusammenhang bei rund $r = .20$. Allerdings wurden diese Zusammenhänge knapp nicht signifikant, was die Autor:innen jedoch bei einer größeren Stichprobe vermuten würden. Auf Basis der umfangreicheren Daten der Untersuchung *Professionswissen in den Naturwissenschaften* (ProwiN) zeigten sich dann auch deutliche Zusammenhänge zwischen Leistungen im Fach Biologie und kognitiver Aktivierung mit $r = .39$, was einer Varianzaufklärung von ca. 15 % entspricht (Förtsch, Werner, von Kotzebue & Neuhaus, 2016). Zusammenhänge in dieser Stärke konnten neben Fauth, Decristan, Rieser, Klieme und Büttner (2014b) für Sachunterricht auch Praetorius et al. (2018) in ihrer Übersichtsarbeit für verschiedene Fächer und Altersstufen finden. Die Befunde ergaben durchgängig signifikante schwache bis moderate positive Zusammenhänge für kognitive Aktivierung und Leistung. Ähnlich, wenn auch mit deutlich geringerer Befundlage, fielen die Ergebnisse für Interesse aus (Praetorius et al., 2018). Für Freude fand sich in dieser Untersuchung hingegen nur ein Befund, der einen nicht signifikanten negativen Zusammenhang mit der kognitiven Aktivierung ergab.

Demgegenüber stehen Studien, die durchaus auch signifikante positive Zusammenhänge zwischen Freude und kognitiver Aktivierung zeigen konnten (Lazarides & Buchholz, 2019; Schiepe-Tiska et al., 2016). Für das Selbstkonzept konnten ebenfalls in verschiedenen Untersuchungen und Kontexten positive Zusammenhänge mit der kognitiven Aktivierung nachgewiesen werden (Drexl & Streb, 2016; Scherer, Nilsen & Jansen, 2016; Seidel & Shavelson, 2007). In vereinzelten Studien zeigen sich jedoch auch negative oder nicht signifikante Zusammenhänge (Henschel et al., 2023; Henschel et al., 2019; Milles & Jansen, 2021). Insgesamt zeigt sich für den Zusammenhang

zwischen kognitiver Aktivierung und Leistung also eine recht einheitliche Befundlage, die von signifikanten, positiven Zusammenhängen ausgeht. Für motivationale Faktoren sind die bisherigen Befunde uneinheitlicher und verweisen sowohl auf nicht signifikante als auch auf signifikante positive und negative Zusammenhänge. Es scheinen für Freude und Selbstkonzept jedoch positive Zusammenhänge vorzuherrschen.

Differenzielle Befunde, die kognitive Aktivierung im Zusammenhang mit besonderer Leistungsstärke untersuchen, sind bisher äußerst selten. Eine der wenigen Studien ist die bereits im Abschnitt zur Klassenführung herangezogene Untersuchung mit äthiopischen Schüler:innen. Für kognitive Aktivierung ergaben sich dabei für alle Leistungsgruppen positive Effekte auf die Leistung, wobei sich keine bedeutsamen Unterschiede zwischen den Gruppen zeigten (Sanfo & Malgoubri, 2021).

Konstruktive Unterstützung

Die dritte Basisdimension der Unterrichtsqualität stellt die konstruktive Unterstützung dar. Diese kann definiert werden als „Anleitung und Begleitung des […] in Gang gesetzten Problemlöseprozesses" durch die unterrichtende Lehrkraft (Minnameier et al., 2015, S. 842). Dabei finden sich in verschiedenen Publikationen unterschiedliche Bezeichnungen wie Schülerorientierung, Unterrichtsklima oder *student support* – im Kern meinen diese Bezeichnungen jedoch vergleichbare Konstrukte (Steffensky & Neuhaus, 2018). In vielen Studien, wenn auch nicht unbedingt explizit benannt, umfasst konstruktive Unterstützung zwei große Subfacetten. Die erste kann als inhaltliche Strukturierung benannt werden und umfasst eine konkret auf den Unterrichtsgegenstand bzw. Lern- und Verstehensprozess bezogene Strukturierung und Hilfestellung. Die zweite Subfacette hingegen wird als emotionale Unterstützung bezeichnet und bezieht sich vor allem auf motivationale Unterstützung und eine positive Beziehung zwischen Lehrkraft und Schüler:in (Kunter & Voss, 2011; Steffensky & Neuhaus, 2018; Teig & Nilsen, 2022). Indikatoren für eine erfolgreiche konstruktive Unterstützung können die Dekomposition komplexer Inhalte in wesentliche Schritte, der Umgang mit Verständnisschwierigkeiten und Fehlern bzw. Fehlvorstellungen, eine positive Fehlerkultur sowie ein konstruktiver Umgang mit Fehlern, individuelle Hilfestellungen, gemeinschaftliche Aktivitäten im Klassenverband, adaptive Unterrichtsgestaltung, individualisierter Unterricht, klare Kommunikation von Erwartungen, konstruktives und effektives Feedback, Aufbau positiver und fürsorglicher Beziehungen, ein wertschätzendes Lernklima, gegenseitiger Respekt,

positive Arbeitshaltung, verantwortungsvoller Umgang mit Personen und Gegenständen, Gerechtigkeit, zufriedene und fröhliche Grundstimmung, Höflichkeit, Geduld, Ansprechbarkeit, Feingefühl für individuelle Bedürfnisse und deren Berücksichtigung oder den Schüler:innen zuhören sein (Bohl, Budde & Rieger-Ladich, 2017; Brophy, 2000; Decristan et al., 2016; Fauth, Decristan, Rieser, Klieme & Büttner, 2014b; Hess & Lipowsky, 2016; Klieme et al., 2001; König et al., 2023; Kunter & Trautwein, 2013; Kunter & Voss, 2011; Meyer, 2003; Minnameier et al., 2015; Praetorius et al., 2018; Salner-Gridling, 2020; Schiepe-Tiska et al., 2016; Sliwka, Klopsch & Dumont, 2019; Teig & Nilsen, 2022)

Konstruktive Unterstützung enthält dementsprechend fachspezifische und generische Anteile. Während die Subfacette emotionale Unterstützung eher generisch angelegt ist, erfordert die inhaltliche Strukturierung hingegen in deutlich höherem Maße fachspezifisches Wissen. Es wird davon ausgegangen, dass eine hohe konstruktive Unterstützung dazu führt, dass Schüler:innen sich im Unterricht wohl fühlen und so günstige motivationale Ausprägungen aufweisen. Gleichzeitig können insbesondere durch die Subfacette inhaltliche Strukturierung auch Lernprozesse gefördert werden. Theoretisch wird somit vermutet, dass konstruktive Unterstützung vor allem einen Einfluss auf motivationale Faktoren, aber auch auf Leistung haben kann, wobei dies stark von der Operationalisierung des Konstruktes abhängt (Kunter & Trautwein, 2013).

Der angenommene Einfluss konstruktiver Unterstützung im Unterricht auf motivationale Merkmale konnte bereits in zahlreichen Studien belegt werden. So konnten beispielsweise Zusammenhänge mit dem Merkmal Interesse wiederholt nachgewiesen werden, wobei diese häufig eher schwach ausfallen (Klieme et al., 2001; Kunter & Voss, 2011; Wentzel, 1998). In einer Studie von Fauth, Decristan, Rieser, Klieme und Büttner (2014b) zeigte sich darüber hinaus, dass konstruktive Unterstützung das Interesse sowohl auf Schüler- als auch auf Klassenebene vorhersagen konnte. Für intrinsische Motivation konnten ebenfalls schwache Zusammenhänge festgestellt werden, wobei sich diese ausschließlich auf Individualebene zeigten (Ohle-Peters et al., 2021). Schwache bis moderate Zusammenhänge konnten für die Merkmale Selbstkonzept und Lernfreude nachgewiesen werden (Lazarides & Buchholz, 2019; Sakiz, Pape & Woolfolk Hoy, 2012; Weißeno & Landwehr, 2015). Im Rahmen der IQB-Bildungstrends 2018 sowie 2022 fanden sich dabei für Selbstkonzept (wie auch Interesse) höhere Zusammenhänge für die Individualebene als für die Klassenebene (Henschel et al., 2023; Henschel et al., 2019). Ausnahme bildet die Studie von Allen und Fraser (2007), in der weder für Freude noch für Einstellung signifikante Zusammenhänge

mit der konstruktiven Unterstützung nachgewiesen werden. In eine ähnliche Richtung deuten auch die Befunde von Praetorius et al. (2018), die signifikante positive Effekte nur bei zwei von sechs untersuchten Zusammenhängen feststellen konnten. Die signifikanten, moderaten Zusammenhänge zeigten sich dabei für die Lernfreude und das Selbstkonzept.

Für die Leistungen finden die meisten Studien wie erwartet keinen Zusammenhang mit der konstruktiven Unterstützung (Kunter & Voss, 2011; Praetorius et al., 2018; Roßbach, 2002; Weißeno & Landwehr, 2015). Dieser Befund ergab sich auch für naturwissenschaftliche Kompetenzen und dabei sowohl für Grund- als auch für Sekundarschulen (Allen & Fraser, 2007; Fauth, Decristan, Rieser, Klieme & Büttner, 2014b). Andererseits konnte Cornelius-White (2007) in seiner Metaanalyse auch moderate positive Effekte von Merkmalen konstruktiver Unterstützung auf kognitive abhängige Variablen finden. Dass dieser Befund so stark von anderen abweicht, könnte in der Definition der kognitiven Merkmale begründet sein. In der Metaanalyse werden unterschiedliche Formen wie IQ, Testleistungen oder Noten in verschiedenen Domänen zusammengefasst, während in den meisten anderen Untersuchungen vor allem Noten oder Testleistungen eines Unterrichtsfaches herangezogen werden. Unter diesen enger gefassten Leistungsvariablen finden sich nur vereinzelt positive Zusammenhänge zwischen Leistung und konstruktiver Unterstützung (Fauth, Decristan, Rieser, Klieme & Büttner, 2014b; Hughes, Luo, Kwok & Loyd, 2008; Kleickmann, Steffensky & Praetorius, 2020).

In *Large-Scale*-Studien fällt das Bild ähnlich aus, wobei sich etwas häufiger positive Zusammenhänge zu zeigen scheinen. Auf Klassenebene finden sich dabei meist keine Zusammenhänge zwischen konstruktiver Unterstützung und Leistung, Zusammenhänge zeigen sich dabei größtenteils auf Schülerebene (Henschel et al., 2023; Henschel et al., 2019; Klieme et al., 2001; Stang et al., 2020). Ausnahme bildet der IQB-Bildungstrend 2018, in dem für Mathematik auf Klassenebene ein signifikanter, negativer Zusammenhang von Schülerorientierung und Leistung gefunden wurde. Auf Schülerebene fällt der Zusammenhang hingegen positiv, allerdings deutlich niedriger aus als auf Klassenebene. In TIMSS 2019 konnte für Naturwissenschaften ebenfalls ein signifikanter positiver Zusammenhang gefunden werden, der jedoch unter Kontrolle von Hintergrundmerkmalen der Schüler:innen nicht weiter bestehen blieb (Stang et al., 2020).

Keinen Unterschied zwischen der Einschätzung des Klassenklimas konnten Lüftenegger et al. (2015) in ihrer Untersuchung begabter und leistungsstarker Schüler:innen

ausmachen. Sanfo und Malgoubri (2021) hingegen zeigten mit Mehrebenenanalysen, dass die konstruktive Unterstützung zum einen einen signifikanten, positiven Einfluss auf die Leistung hat, zum anderen, dass dieser Zusammenhang für die leistungsstarken Schüler:innen nochmal höher ausfällt. Dieses Ergebnis stützen auch Affandi, Husniati und Saputra (2020), die das Wohlbefinden leistungsstarker Schüler:innen untersuchten und dabei insbesondere zwei Faktoren als relevant ausmachen konnten. Einen der beiden Faktoren beschreiben sie als Unterstützung bei der individuellen Weiterentwicklung und identifizieren somit einen Aspekt der konstruktiven Unterstützung als besonders bedeutsam.

Gemeinsame Betrachtung der drei Basisdimensionen

Wie bereits in den Darstellungen zu den einzelnen Basisdimensionen angeklungen, untersuchen viele Studien, vor allem große Projekte wie COACTIV (Kunter et al., 2013; Kunter & Voss, 2011) oder Pythagoras (Lipowsky et al., 2009) sowie *Large-Scale*-Studien wie der IQB-Bildungstrend (Henschel et al., 2023; Henschel et al., 2019), TIMSS (Klieme et al., 2001; Stang et al., 2020) und PISA (Schiepe-Tiska et al., 2016), alle drei Basisdimensionen gemeinsam. Mit diesen Studien kann das Verhältnis der Basisdimensionen untereinander sowie deren Bedeutsamkeit für Unterrichtsziele bei gleichzeitiger Betrachtung beschrieben werden. In der Gesamtheit lässt sich resümieren, dass es neben einigen nicht signifikanten Befunden für den Zusammenhang der Basisdimensionen der Unterrichtsqualität mit Leistung und motivationalen Merkmalen viele schwach bis moderat ausgeprägte Zusammenhänge gibt, die sich darüber hinaus auch zwischen den drei Basisdimensionen zeigen (Schiepe-Tiska et al., 2016). Da die meisten Untersuchungen dabei Mehrebenenanalysen beinhalten, werden die Ergebnisse im Folgenden getrennt für die Schüler- und Klasseneben zusammengefasst. Auf Klassenebene finden sich die höchsten und häufigsten signifikanten Zusammenhänge zwischen kognitiver Aktivierung und Leistungswerten sowie zwischen konstruktiver Unterstützung und motivationalen Faktoren. Dabei fallen diese zumeist positiv und schwach bis moderat aus. In einigen Studien finden sich zudem schwache positive Zusammenhänge für konstruktive Unterstützung und Leistung, Ausnahme bildet ein Befund des IQB-Bildungstrends, der bedeutsame negative Befunde nachweisen konnte. Etwas seltener finden sich in diesen Studien Zusammenhänge für Klassenführung. Die Befunde sind schwächer als die für konstruktive Unterstützung mit motivationalen Faktoren bzw. kognitiver Aktivierung mit Leistung, allerdings

zeigen sie sich sowohl für Leistungen als auch motivationale Faktoren. Insbesondere Zusammenhänge zwischen kognitiver Aktivierung und motivationalen Faktoren konnten bisher kaum ausgemacht werden, einzig im Rahmen des IQB-Bildungstrends 2022 konnte ein schwacher negativer Zusammenhang von kognitiver Aktivierung und Selbstkonzept festgestellt werden. Auf Klassenebene können also für Leistung vor allem Zusammenhänge mit der kognitiven Aktivierung und Klassenführung, für motivationale Faktoren mit der konstruktiven Unterstützung und ebenfalls der Klassenführung ausgemacht werden.

Auf Schülerebene finden sich insgesamt etwas weniger Befunde, da in einigen Studien die Basisdimensionen der Unterrichtsqualität ausschließlich auf der Klassenebene modelliert wurden. Auf Schülerebene zeigen sich die meisten bedeutsamen Zusammenhänge für konstruktive Unterstützung. Anders als auf Klassenebene finden sich dabei Zusammenhänge mit Leistung und motivationalen Faktoren zu etwa gleichen Teilen. Für kognitive Aktivierung zeigen sich auch auf Schülerebene vor allem signifikante Zusammenhänge mit der Leistung, allerdings sind diese etwas seltener signifikant. Vereinzelt und in deutlich geringerer Höhe konnten auch Zusammenhänge zwischen Klassenführung und motivationalen Faktoren ausgemacht werden, wobei diese uneinheitlich mal positiv und mal negativ ausfallen. Auf Schülerebene ist somit für Leistung vor allem ein positiver Zusammenhang mit konstruktiver Unterstützung und kognitiver Aktivierung, für motivationale Faktoren ein positiver Zusammenhang mit konstruktiver Unterstützung anzunehmen. Es zeigt sich, dass alle drei Basisdimensionen bedeutsam für multikriteriale Bildungsziele des Unterrichts sind und dementsprechend auch gemeinsam betrachtet werden sollten (Bohl, 2017).

Bei den differenziellen Befunden finden sich kaum Untersuchungen, die alle drei Basisdimensionen gemeinsam betrachten. Ausnahme bildet die Untersuchung von Sanfo und Malgoubri (2021), die zeigen konnten, dass konstruktive Unterstützung besonders bedeutsam für die Leistung leistungsstarker Schüler:innen ist. Kognitive Aktivierung stellte sich hingegen für die Leistung aller Schüler:innen unabhängig vom Leistungsniveau als relevant heraus, während sich für Klassenführung keine signifikanten Zusammenhänge zeigten. Lüftenegger et al. (2015) konnten in ihrer Untersuchung zudem Unterschiede für die Wahrnehmung der Passung von Aufgaben zum Leistungsniveau und der Eingebundenheit in Entscheidungsprozesse zwischen unterschiedlichen Leistungsgruppen ausmachen. Für Merkmale der wahrgenommenen Klassenführung und das Klassenklima zeigten sich keine Unterschiede. Somit geben bisherige differenzielle Befunde Hinweise darauf, dass die konstruktive Unter-

stützung, möglicherweise insbesondere die inhaltliche Strukturierung, für leistungsstarke Schüler:innen besonders bedeutsam für die Leistung ist. Es bleibt allerdings festzuhalten, dass die leistungsdifferenzielle Befundlage im Bereich der Unterrichtsqualitätsdimensionen äußerst unzureichend ist. Dies ist umso mehr erstaunlich, als dass Unterrichtsqualität in der Forschung zu den prominentesten Merkmalen von Unterricht zählt und in zahlreichen Studien untersucht wurde. Die wenigen Quellen, die entsprechende Untersuchungen behandeln, stammen vor allem aus den letzten Jahren. Grund dafür könnte das Wiedererstarken des *Aptitude-Treatment*-Ansatzes (Snow, 1991), bei dem davon ausgegangen wird, dass die Wirksamkeit von Instruktionen von den individuellen Voraussetzungen der Schüler:innen abhängt, vor dem Hintergrund einer immer heterogener werdenden Schülerschaft sein. Gleichzeitig wird in diesem Zuge wiederholt der Bedarf an Studien mit differenziellem Ansatz formuliert (Kunter & Ewald, 2016; Neuhaus, 2007).

Ein Aspekt soll zum Abschluss des Kapitels noch einmal aufgegriffen werden: Merkmale wie Adaptivität oder Berücksichtigung individueller Bedürfnisse und Interessen werden in der Literatur jeder der drei Basisdimensionen zugeordnet. Wie zuvor beschrieben, zeigt sich somit für den Bereich der Differenzierung eine Uneinheitlichkeit in der Systematisierung und Einordnung in den Diskurs. Im Folgenden wird Differenzierung deshalb in einem eigenen Kapitel genauer in den Blick genommen.

4.2.2 Differenzierung

Unter Differenzierung versteht man, „die individuellen Lernvoraussetzungen von Lernenden so zu berücksichtigen, dass alle Lernenden gemäß ihren Fähigkeiten, Bereitschaften und Einstellungen gefordert werden“ (Prediger & von Aufschnaiter, 2017, S. 291). Dabei wird Differenzierung als Sammelbegriff verstanden, unter dem verschiedene organisatorische, methodische und didaktische Ansätze wie Individualisierung bzw. individuelle Förderung, innere Differenzierung, adaptives Unterrichten oder offener Unterricht zusammengefasst werden. Übereinstimmend reagieren diese Ansätze auf die Heterogenität der Schüler:innen, indem sie eine größere Passung von Unterrichtsgestaltung und Voraussetzung ermöglichen und so bessere Leistungen möglich machen sollen (Häcker, 2017; Sächsisches Bildungsinstitut, 2017).

Differenzierung lässt sich allgemein unterteilen in Formen der äußeren Differenzierung und Formen der inneren Differenzierung (auch Binnendifferenzierung

genannt). Die äußere Differenzierung beschreibt Strukturierungsmaßnahmen zur Erzeugung kriterienbezogener Gruppen auf der organisatorischen Meso- und Makroebene des Schulsystems, beispielsweise durch unterschiedliche Schulformen, Profilschulen, Jahrgänge oder Klassen. Unter innerer Differenzierung werden Maßnahmen verstanden, die innerhalb einer Lerngruppe und vor allem initiiert durch die Lehrkraft umgesetzt werden (Inckemann, 2014; Klafki, 2007; Letzel, 2021). Dabei basiert innere Differenzierung auf einer weiteren Gruppenbildung innerhalb der Lerngruppe, z. B. anhand zuvor gezeigter Leistungen oder Interessen, sodass die Passung zwischen Inhalten und Voraussetzungen unterschiedlicher Schülergruppen während der einzelnen Unterrichtseinheiten und für konkrete Inhalte erhöht wird (Häcker, 2017). Nach Klafki (2007) gibt es vor allem zwei grundlegende Wege innere Differenzierung zu realisieren. Zum einen können zur Erreichung der Lernziele für Schüler:innen unterschiedliche Methoden oder Medien eingesetzt, zum anderen können die Inhalte und die individuellen Lernziele der Schüler:innen angepasst werden, wobei unter Lernziele sowohl kognitive als auch motivationale Merkmale fallen. Für die Umsetzung des zweiten Weges wäre ein zweiteiliger Lehrplan von Nöten, der sich aus Basisinhalten (Fundamentum) und Aufbauinhalten (Additum) zusammensetzt, sodass Schüler:innen Inhalte entsprechend ihrer Voraussetzungen in unterschiedlicher Tiefe und Ausführlichkeit behandeln können (Klafki, 2007). Zur Realisierung des ersten Weges, der Differenzierung von Methoden und Medien, bieten sich in erster Linie unterschiedliche Sozialformen an. In heterogenen Lerngruppen kommt dabei der Gruppenarbeit, neben Formen der Partner- und Einzelarbeit, eine besondere Bedeutung zu. Darüber hinaus kann die Lehrkraft mit den Schüler:innen direkt zusammenarbeiten und dies zur Umsetzung von Differenzierung mit den ersten drei Sozialformen für ausgewählte Schülergruppen kombinieren (Klafki, 2007).

Als Spezialfall von innerer Differenzierung kann die individuelle Förderung bzw. die Individualisierung bezeichnet werden. Beide Begriffe werden dabei in der Literatur überwiegend synonym verwendet. Anders als bei innerer Differenzierung richtet individuelle Förderung bzw. Individualisierung den Blick nicht auf unterschiedliche Gruppen von Schüler:innen, sondern betrachtet alle Schüler:innen einzeln. Deshalb wird individuelle Förderung bzw. Individualisierung häufig auch als das gegenüberliegende Extrem zu innerer Differenzierung beschrieben (Häcker, 2017; Hess & Lipowsky, 2016). Darüber hinaus können innere Differenzierung und individuelle Förderung bzw. Individualisierung hinsichtlich ihrer Ausgansperspektive unterschieden werden. Während Differenzierung eher von der Lehrkraft und der Organisation der Schüler:in-

nen in ‚händelbaren' Gruppen aus gedacht wird, setzt individuelle Förderung bzw. Individualisierung an den einzelnen Schüler:innen an und denkt den Unterricht von deren Seite her (Altrichter, Trautmann, Wischer, Sommerauer & Doppler, 2009).

Eine weitere verbreitete Form innerer Differenzierung ist die natürliche Differenzierung, die auch offene Differenzierung oder Selbstdifferenzierung genannt wird. Natürliche Differenzierung äußert sich über Arbeitsaufträge und Aufgaben, die verschiedene Umsetzungs- und Wahlmöglichkeiten mit unterschiedlichen Anforderungsniveaus ermöglichen. Die Schüler:innen können so eigenverantwortlich und ihren eigenen Kompetenzen und Interessen entsprechend die Inhalte bearbeiten und den Unterricht so an die eigenen Bedürfnisse anpassen (Prediger & von Aufschnaiter, 2017; Scherer & Moser Opitz, 2010).

Da wie eingangs herausgearbeitet in dieser Arbeit Merkmale des Unterrichts und nicht des Jahrgangs oder der Schule betrachtet werden sollen, liegt der Fokus in dieser Arbeit auf Formen der inneren Differenzierung. Diese haben gemein, dass sie innerhalb von bestehenden Lerngruppen die Voraussetzungen der Schüler:innen berücksichtigen und darauf abzielen, alle Schüler:innen entsprechend ihrer Voraussetzungen individuell in ihrem Lernen zu unterstützen.

Die verschiedenen Ansätze, Unterricht an die unterschiedlichen Voraussetzungen der Schüler:innen anzupassen, entstanden vor allem im Kontext einer zunehmenden Heterogenität in der Schülerschaft und der Vermutung, dass es die sogenannten ‚mittleren' Schüler:innen (Gebauer, 2019), an denen Unterricht sich klassischerweise orientiert, nicht gibt. Insbesondere in der Grundschule wird Maßnahmen der inneren Differenzierung eine hohe Bedeutung beigemessen, da im Bereich dieser Schulform seltener äußere Differenzierungsformen vorkommen. In den meisten Fällen setzt diese erst verstärkt mit dem Übergang in die Sekundarstufe und dem Eintritt in ein mehrgliederiges Schulsystem ein. Die Heterogenität in der Schülerschaft ist somit höher als in den anderen Bereichen des Schulsystems, sodass Unterricht stärker auf die unterschiedlichen Voraussetzungen ausgerichtet sein muss, um die Chancengleichheit für Lernerfolge beizubehalten (Hadeler, 2015; Warwas, Hertel & Labuhn, 2011).

Diesen sehr nachvollziehbaren Begründungen für die Relevanz differenzierender Maßnahmen gegenüber werden zumeist fünf Argumente gegen Differenzierung bzw. als Schwierigkeit dieses Ansatzes angeführt (Altrichter et al., 2009; Bruder, Linneweber-Lammerskitten & Reibold, 2015; Gehrer & Nusser, 2020; Jennek, Gronostaj & Vock, 2019; Roeder, 1997). Der wohl meist benannte Nachteil von innerer Differenzierung ist der hohe Aufwand, der für Lehrkräfte in der Unterrichtsvorbereitung besteht.

Da die Materialien beispielsweise in unterschiedlichen Schwierigkeitsstufen oder für unterschiedliche Schülerinteressen mit verschiedenen Inhalten vorbereitet werden müssen, erhöht sich der Vorbereitungsaufwand teilweise enorm. Ein zweites Argument geht in eine ähnliche Richtung: Nicht nur die Vorbereitung von differenzierenden Unterrichtsangeboten ist aufwändig, auch die Organisation und Implementation dieser Methoden und Materialien im Unterricht kann mehr Zeit in Anspruch nehmen, sodass die Zeit, die Schüler:innen aktiv mit den Inhalten verbringen, geringer ausfällt. Ebenfalls häufig angeführt werden außerdem Bedenken hinsichtlich der Wirksamkeit differenzierender Maßnahmen. Das vierte und fünfte Argument betreffen die personellen Voraussetzungen bei der Realisierung von Differenzierung. Einerseits erfordern differenzierende Maßnahmen hohe diagnostische Fähigkeiten der Lehrkräfte, da sie in der Lage sein müssen, die Voraussetzungen der Schüler:innen adäquat einzuschätzen, um geeignete Angebote machen zu können. Andererseits müssen die Schüler:innen jedoch auch gewisse Fähigkeiten wie Selbstständigkeit, Disziplin, Teamfähigkeit und Selbstregulation mitbringen, damit differenzierende Maßnahmen im Unterricht erfolgreich umgesetzt werden können, was insbesondere in frühen Klassenstufen nicht immer der Fall ist (Altrichter et al., 2009; Bruder et al., 2015; Gehrer & Nusser, 2020; Jennek et al., 2019; Roeder, 1997).

Diese Bedenken und Herausforderungen zeigen sich auch in den Befunden zum Einsatz differenzierender Maßnahmen in Deutschland. Insbesondere *Large-Scale*-Studien bieten zur Identifikation von Einsatzhäufigkeiten großes Potenzial, da sie in der Regel sehr große und repräsentative Stichproben umfassen. Eine umfassende Sichtung dieser Studien hinsichtlich Differenzierungsmaßnahmen findet sich bei Letzel und Otto (2019) sowie Letzel (2021). Sie kommen zu dem Schluss, dass bisherige Studien ein geringes Maß an Implementation von innerer Differenzierung im Unterricht herausstellen konnten, wobei die Vielfalt erfragter Methoden in einigen dieser Studien stark eingeschränkt ist. Zudem konnte gezeigt werden, dass sich die Methoden, die hauptsächlich eingesetzt wurden, vor allem durch eine geringe Vorbereitungszeit und Geschlossenheit, d. h. Entscheidung des Einsatzes durch die Lehrkraft, auszeichnen, wie z. B. individuelle Unterstützungsmaßnahmen durch die Lehrkraft oder angepasste Bearbeitungszeiten (Hadeler, 2015; Hertel, 2014; Jennek et al., 2019).

Zur Wirksamkeit binnendifferenzierender Maßnahmen gibt es eine große Anzahl von Untersuchungen, die die Zusammenhänge für unterschiedliche Differenzierungsmethoden, Fächer und Altersgruppen untersuchen. Die zahlreichen Ergebnisse wurden bereits in mehreren Übersichtsarbeiten und Metaanalysen gebündelt, sodass für

einen Überblick vorrangig diese Befunde vorgestellt werden. In seiner Metaanalyse konnte Hattie (2009) einen schwachen positiven Zusammenhang für Leistung mit binnendifferenzierenden Maßnahmen von $d = .16$, mit Individualisierung von $d = .23$ und mit Enrichment-Angeboten (zusätzliche Aufgaben für leistungsstarke Schüler:innen) von $d = .39$ feststellen. Diese positiven Zusammenhänge finden sich auch in weiteren Metaanalysen für verschiedene Fächer und Altersstufen. So konnten beispielsweise Lou et al. (1996) für Mathematik, Naturwissenschaften, Sprachen, Kunst und weitere Fächer zeigen, dass *within-class grouping* positive Auswirkungen auf die Leistung der Schüler:innen hat, wobei der Effekt für Mathematik und Naturwissenschaften am höchsten ausfällt. Darüber hinaus ergab sich, dass positive Auswirkungen für alle Altersstufen angenommen werden können. Die höchsten Zusammenhänge zeigen sich dabei für das späte Grundschulalter (Lou et al., 1996). Moderate Zusammenhänge mit der Leistung in Mathematik im späten Grundschulalter konnte auch Slavin (1987) bestätigen. Zusammenhänge von $g = .19$ bis $g = .30$ konnten Steenbergen-Hu, Makel und Olszewski-Kublilius (2016) für diese Maßnahme finden. Kulik und Kulik (1992) hingegen finden für *within-class grouping* lediglich kleine oder gar keine Zusammenhänge mit der Leistung. Sie kommen zu dem Schluss, dass differenzierende Maßnahmen umso bedeutsamer für schulische Leistungen sind, je stärker sie für eine Anpassung des Curriculums sorgen. Am einflussreichsten stellten sich in ihrer Untersuchung Maßnahmen wie Enrichment und Akzeleration (schnelleres Durchlaufen der Inhalte und Schullaufbahn) heraus (Kulik & Kulik, 1992). Horak (1981) konnte hingegen für individualisierten Mathematikunterricht sogar kleine negative Zusammenhänge feststellen. In einer umfangreichen Bestandsaufname zu binnendifferenzierendem und individualisierendem Unterricht, in die u. a. auch die Studien von Horak (1981) und Hattie (2009) einfließen, kommen Lipowsky und Lotz (2015) zu dem Schluss, dass innere Differenzierung und Individualisierung zwar Einfluss auf die Leistung haben, dieser jedoch meist weniger stark ausfällt als nach theoretischen Überlegungen erhofft. Dies führen sie darauf zurück, dass differenzierende Maßnahmen zumeist aufwändig in der Vorbereitung sind und deshalb möglicherweise aus Zeitgründen zu wenig Aufmerksamkeit auf eine den Tiefenstrukturen entsprechende, qualitativ hochwertige Umsetzung gelegt wird.

Anders als für die Basisdimensionen der Unterrichtsqualität gibt es für Differenzierung häufiger auch leistungsdifferenzielle Untersuchungen. Dies dürfte in der Anlage des Konzeptes Differenzierung begründet liegen, da häufig nach Leistung unterschiedene Gruppen im Fokus der Unterrichtsgestaltung stehen. In einem syste-

matischen Review zur Bedeutung differenzierender Maßnahmen für leistungsstarke Schüler:innen kommen Ziernwald, Hillmayr und Holzberger (2022) zu drei wesentlichen Erkenntnissen:

1. Differenzierungsmaßnahmen haben für leistungsstarke Schüler:innen überwiegend positive Auswirkungen, wobei die Befunde stark zwischen den untersuchten Studien variieren.
2. Sowohl Lehrkräfte als auch leistungsstarke Schüler:innen schätzen Differenzierungsmaßnahmen als wünschenswerte bzw. geeignete Methoden zur Förderung leistungsstarker Schüler:innen ein.
3. Trotzdem erfahren leistungsstarke Schüler:innen Differenzierung in Schule und Unterricht nicht auf regelmäßiger, üblicher Basis (Ziernwald et al., 2022).

Für die Frage, ob Differenzierung dabei eher für leistungsschwache oder leistungsstarke Schüler:innen relevant ist, ist die bisherige Befundlage uneinheitlich. Während einige Studien zu dem Schluss kommen, dass vor allem leistungsschwache Schüler:innen differenzierende Maßnahmen erfahren und davon profitieren (Gehrer & Nusser, 2020; Lou et al., 1996; Weinert & Helmke, 1996), können andere Studien zeigen, dass die Einsatzhäufigkeit und der Nutzen für leistungsstarke Schüler:innen höher ist (Inckemann, 2014; Kulik & Kulik, 1992).

Neben dem Zusammenhang von differenzierenden Maßnahmen mit Leistungen wurde auch der Zusammenhang mit motivationalen Merkmalen der Schüler:innen untersucht, wobei die Befundlage hier deutlich weniger umfangreich ist. Viele Untersuchungen kommen zu dem Schluss, dass sich differenzierende Maßnahmen positiv auswirken (Altrichter et al., 2009; Böhnel & Svik, 1993; Letzel, 2021; Ohle-Peters et al., 2021). In ihrer Metaanalyse betrachten Lou et al. (1996) neben der Leistung auch das Selbstkonzept und die Einstellung der Schüler:innen. Für beide Merkmale können sie kleine positive Effekte von *within-class grouping* nachweisen. Ähnlich kommen Beck et al. (2008) zu dem Ergebnis, dass ein Unterricht, der sich den Voraussetzungen der Schüler:innen anpasst, positive Auswirkungen auf das Selbstkonzept und das Interesse hat. Beloshitskii und Dushkin (2005), Guay, Roy und Valois (2017) und Ohle-Peters et al. (2021) konnten zudem zeigen, dass innere Differenzierung positive Effekte auf die (intrinsische) Motivation der Schüler:innen hat.

Differenzielle Befunde für motivationale Merkmale leistungsstarker Schüler:innen konnten darüber hinaus kaum ausgemacht werden. Letzel (2021) vermutet jedoch, dass sich wie für Leistungen auch für motivationale Merkmale leistungsstarker Schüler:in-

nen Unterschiede in der Wirksamkeit verschiedener Maßnahmen zeigen. In einer Übersichtsarbeit zu Differenzierungsmaßnahmen kommen Ziernwald, Holzberger, Hillmayr und Reiss (2020) zu dem Schluss, dass bisherige Befunde darauf hindeuten, dass bei leistungsstarken Schüler:innen Differenzierung nur über homogene Lerngruppen (also Maßnahmen, die eher der äußeren Differenzierung zugeordnet werden können) positive Auswirkungen auf motivationale Merkmale hat. Für innere Differenzierung finden sie ebenfalls keine Befunde. In ihrem systematischen Review zur Wirkung von Differenzierungsmaßnahmen für leistungsstarke Schüler:innen finden Ziernwald et al. (2022) zwei Jahre später erneut Hinweise auf eine positive Auswirkung vornehmlich äußerer Differenzierungsmaßnahmen auf motivationale Merkmale leistungsstarker Schüler:innen. Aufgrund fehlender Untersuchungen, die sich explizit mit Methoden innerer Differenzierung befassen, können jedoch keine genaueren Aussagen über einen möglichen differenziellen Zusammenhang zwischen motivationalen Merkmalen und innerer Differenzierung bei besonderer Leistungsstärke getroffen werden.

4.3 Merkmale der Lehrkraft

Wie in Kapitel 3 herausgearbeitet, gehören neben Schülermerkmalen und Unterricht auch Lehrkräfte zu den einflussreichsten Bedingungsfaktoren schulischer Leistungen von Schüler:innen. Auf Grund ihrer bedeutsamen Funktion im Bildungssystem liegt der Fokus in den meisten Untersuchungen deshalb auf berufsbezogenen Merkmalen von Lehrkräften. Zu häufig untersuchten Merkmalen zählen beispielsweise Aus- und Fortbildung, diagnostische, didaktische oder fachliche Kompetenzen, Selbstregulation, Selbstwirksamkeit, Persönlichkeitsfaktoren, Enthusiasmus, Wohlbefinden, Achtsamkeit, Belastungserleben, Ressourcen sowie Berufszufriedenheit (Baumert et al., 2010; Bermejo-Toro, Prieto-Ursúa & Hernández, 2016; Darling-Hammond, 2000; Fauth et al., 2019; Hattie, 2009; Hofmann & Gottein, 2011; Hosenfeld, Helmke & Schrader, 2002a; Jackson-Crossland, 2000; Jennings & Greenberg, 2009; Jerusalem & Schwarzer, 1999; Kunter et al., 2011; Shulman, 1987; Skaalvik & Skaalvik, 2014).

Klarer Fokus liegt dabei in bisherigen Untersuchungen auf den professionellen Kompetenzen der Lehrkräfte, die je nach Konzeptualisierung einen Großteil der genannten Merkmale beinhalten können. In den letzten Jahren hat sich ein Modell durchgesetzt, das die professionellen Kompetenzen von Lehrkräften in vier Teilaspekte unterteilt (siehe Abbildung 6).

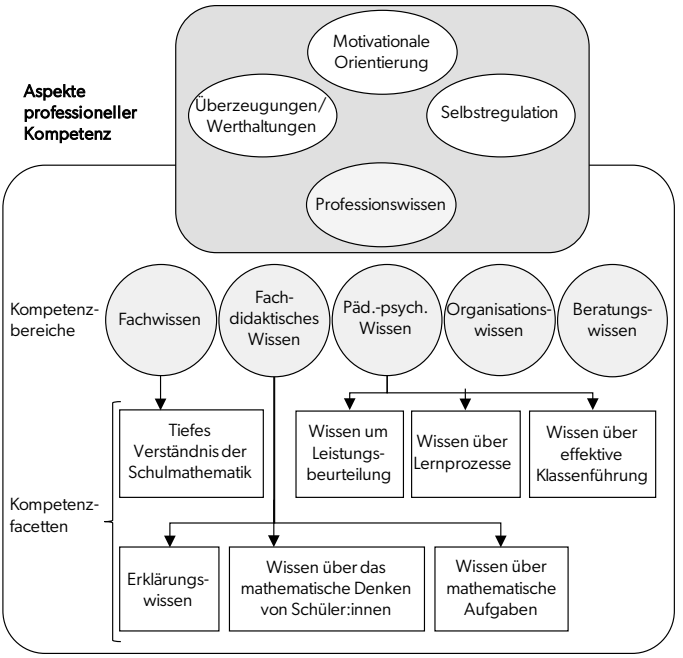


Abb.: 6 Komponenten des Professionswissens nach Baumert und Kunter (2011) sowie Shulman (1987)

Der umfangreichste Teilaspekt ist das Professionswissen. Dieses lässt sich wiederum in fünf Kompetenzbereiche unterteilen, die zum Teil weitere Facetten aufweisen. Der erste Bereich ist das Fachwissen, das beispielsweise ein vertieftes Verständnis der Inhalte sowie Faktenwissen umfasst. Der zweite Kompetenzbereich nennt sich fachdidaktisches Wissen und beinhaltet Facetten wie Erklärungswissen, Wissen über die Denkweisen von Schüler:innen oder Wissen über mögliche Aufgaben. Der dritte Kompetenzbereich, das pädagogisch-psychologische Wissen, umfasst die Facetten Wissen über Leistungsbeurteilung, Lernprozesse und effektive Klasseführung. Ergänzt werden diese drei zentralen Kompetenzen um das Organisationswissen und das Beratungswissen. Neben dem großen Bereich des Professionswissens umfasst die professionelle Kompetenz von Lehrkräften außerdem deren Überzeugungen und Werthaltungen. Dazu werden sogenannte *beliefs*, also epistemologische Überzeugungen und subjektive Theorien zum Lehren und Lernen, gezählt. Motivationale Orientierungen

als dritter Aspekt umfassen im Kontext professioneller Kompetenz vor allem deren intrinsische Motivation sowie selbstbezogene Kognitionen wie Selbstwirksamkeit oder Kontrollüberzeugungen. Der letzte Aspekt, die Selbstregulation der Lehrkräfte, bezieht sich auf deren Fähigkeit, mit Ressourcen und Beanspruchungen im Beruf umzugehen (Baumert & Kunter, 2006, 2011). Die beiden letztgenannten Aspekte weisen dabei eine gewisse Nähe auf:

> „Motivationale Orientierungen und selbstregulative Fähigkeiten sind für die psychische Dynamik des Handelns, die Aufrechterhaltung der Intention und die Überwachung und Regulation des beruflichen Handelns über einen langen Zeitraum verantwortlich. Beide Aspekte sind somit zentrale Merkmale der psychologischen Funktionsfähigkeit von handelnden Personen (Baumert & Kunter, 2011, S. 42)."

Die Bedeutung dieser eher distalen Merkmale für die Leistungen von Schüler:innen lässt sich mithilfe des Kaskadenmodells nach Krauss et al. (2020) verdeutlichen (siehe Abbildung 7). Es stellt eine Weiterentwicklung des der COACTIV-Studie zugrunde gelegten Wirkmodells mit etablierten Kompetenzmodellen dar, wobei das Wirkmodell zunächst nur die Säulen 1, 3 und 5 umfasste. Nach dem Kaskadenmodell von Krauss et al. (2020) bilden die professionellen Kompetenzen der Lehrkräfte als erste Säule den Ausgangspunkt. Die professionellen Kompetenzen haben einen Einfluss auf situationsspezifische Kompetenzen, die die Wahrnehmung, Interpretation und Entscheidungsfindung im Unterricht umfassen. Diese wiederum wirken sich darauf aus, wie sich die Lehrkräfte im Unterricht verhalten und Lerngelegenheiten für die Schüler:innen bereitstellen. Über Aktivitäten der Schüler:innen, wie beispielsweise die Beteiligung im Unterricht oder die Nutzung von Lerngelegenheiten, wirken diese wiederum indirekt auf die unterrichtlichen Zielkriterien. Dieser Schritt wird im Modell als Mediation der Schüler:innen bezeichnet. Unter unterrichtlichen Zielkriterien, die den Endpunkt des Kaskadenmodells darstellen, werden im Kaskadenmodell sowohl kognitive als auch motivationale Merkmale verstanden. Das Modell versteht die einzelnen Säulen dabei als Elemente einer kausalen Wirkungskette von professionellen Kompetenzen bis hin zu unterrichtlichen Zielkriterien.

Die Bedeutung der professionellen Kompetenz wurde bereits in zahlreichen Untersuchungen belegt. Dabei gibt es eine Vielzahl an Studien, die entweder das Modell der professionellen Kompetenz sowie dessen einzelne Aspekte überprüfen oder

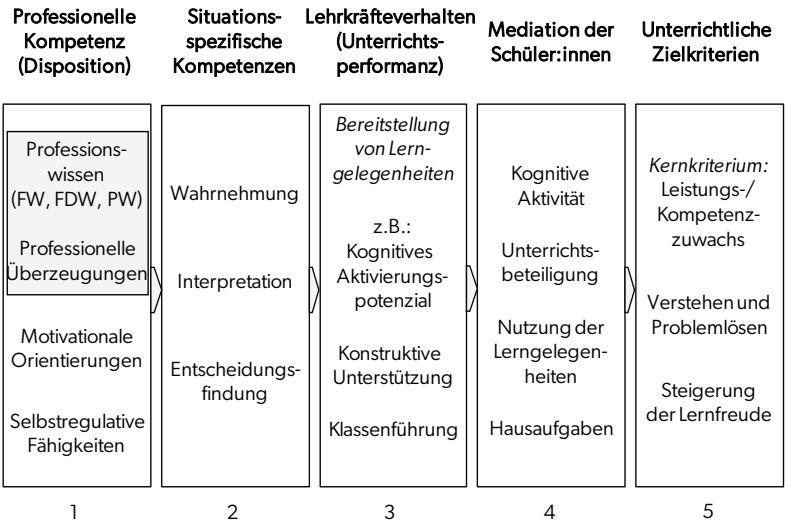


Abb.: 7 Kaskadenmodell des Professionswissens nach Krauss et al. (2020)

deren Zusammenhang mit weiteren Merkmalen untersuchen (Blömeke et al., 2014; Dicke et al., 2019; Kaiser et al., 2016; Kleickmann, Vehmeyer & Möller, 2010; König & Pflanzl, 2015; Lay & Chandrasegaran, 2018; Voss, Kunter, Seiz, Hoehne & Baumert, 2014). Zunehmend finden sich jedoch auch Studien, die das Kaskadenmodell oder Ausschnitte davon in den Blick nehmen und die indirekte wie direkte Bedeutung der professionellen Kompetenz für schulische Leistungen herausarbeiten (Arens & Morin, 2016; Baumert et al., 2010; Fauth et al., 2019; Förtsch, Werner, von Kotzebue, et al., 2016; Guo, McDonald Connor, Yang, Roehrig & Morrison, 2012; Kunter et al., 2013). In diesen Untersuchungen werden verschiedene Merkmale, die hauptsächlich der professionellen Kompetenz von Lehrkräften zugeordnet werden können, mit den Basisdimensionen der Unterrichtsqualität und Leistung in Mathematik, Naturwissenschaften und Lesen sowie zum Teil auch mit motivationalen Merkmalen in Zusammenhang gebracht, wobei dabei insbesondere Aspekte des Professionswissens im Vordergrund stehen. Die gefundenen Zusammenhänge sind in diesen Studien schwach bis moderat.

In den Untersuchungen zeigen sich vor allem Zusammenhänge des pädagogisch-psychologischen Wissens mit der konstruktiven Unterstützung und der kognitiven Aktivierung. Zusätzlich konnten für die kognitive Aktivierung auch Effekte auf die Leistung nachgewiesen werden (Baumert et al., 2010; Förtsch, Werner, von Kotzebue, et al., 2016; Kunter et al., 2013). Direkte Effekte auf Leistung zeigten sich für das pädagogisch-psychologische Wissen dabei nicht, jedoch konnte in einer Untersuchung ein direkter Effekt des Fachwissens auf die Leistung ohne Mediation durch Merkmale der Unterrichtsqualität festgestellt werden (Förtsch, Werner, von Kotzebue, et al., 2016). Einige Studien beziehen zusätzlich Merkmale aus den Bereichen motivationale Orientierungen und Selbstregulation mit ein, vereinzelt stehen diese allein im Fokus. Direkte Effekte auf die Leistungen von Schüler:innen konnten in den betrachteten Untersuchungen für Berufszufriedenheit, Wohlbefinden und Erschöpfung der Lehrkräfte gefunden werden (Arens & Morin, 2016; Banerjee, Stearns, Moller & Mickelson, 2017; Collie & Martin, 2017). Für weitere Aspekte der Selbstregulation (Berufszufriedenheit, berufsbezogenes Wohlbefinden, konstruktivistische Überzeugungen und Selbstregulationsstile) zeigten sich darüber hinaus Effekte auf konstruktive Unterstützung und Klassenführung (Dicke et al., 2019; Grams, 2014; Guo et al., 2012; Kunter et al., 2013). Demgegenüber konnten für die Merkmale Selbstwirksamkeit und Enthusiasmus als Teil der motivationalen Orientierung, ähnlich wie beim pädagogisch-psychologischen Wissen, vor allem indirekte Zusammenhänge mit Leistungen über die konstruktive Unterstützung und die kognitive Aktivierung nachgewiesen werden (Fauth et al., 2019; Guo et al., 2012; Kunter et al., 2013). Darüber hinaus ist Selbstwirksamkeit neben konstruktivistischer Überzeugung das einzige Merkmal, für das bedeutsame Zusammenhänge mit der Klassenführung gezeigt werden konnten. Hervorzuheben ist somit, dass Selbstwirksamkeit das einzige Merkmale ist, für das Zusammenhänge mit allen drei Basisdimensionen festgestellt werden konnten, wobei in der Untersuchung von Fauth et al. (2019) zusätzlich zu diesem Befund auch alle drei Basisdimensionen signifikante Auswirkungen auf kognitive wie nicht-kognitive Merkmale der Schüler:innen haben. Der Selbstwirksamkeit von Lehrkräften scheint somit eine besondere Bedeutung zuzukommen.

Im Bereich der professionellen Lehrkräftekompetenz konnte nur ein einziger differenzieller Befund von 1997 ausgemacht werden. Insgesamt muss die Befundlage in diesem Feld damit als äußerst unzureichend hinsichtlich Aktualität und Menge beschrieben werden, was daran liegen könnte, dass Untersuchungen, die professionelle Kompetenzen von Lehrkräften mit Schulleistungen zusammenbringen erst in

jüngerer Zeit verstärkt durchgeführt werden. Rowen et al. (1997) konnten in einer Untersuchung zum Professionswissen der Lehrkräfte in Mathematik schwache Zusammenhänge mit der Leistung der Schüler:innen zeigen, wobei sich herausstellte, dass das Professionswissen für leistungsschwächere Schüler:innen bedeutsamer zu sein scheint als für leistungsstarke Schüler:innen. Möglicherweise sind also andere Aspekte professioneller Kompetenz für leistungsstärkere Schüler:innen bedeutsamer. Diese Annahme erscheint plausibel, da generelle Befunde sowohl die Bedeutung der motivationalen Orientierungen (insbesondere die Facette Selbstwirksamkeit) und der selbstregulativen Fähigkeiten von Lehrkräften (vorrangig positiv angelegte Konstrukte wie berufsbezogenes Wohlbefinden oder Berufszufriedenheit) bereits nachweisen konnten. In den folgenden Kapiteln werden deshalb die Merkmale Berufszufriedenheit und Selbstwirksamkeit genauer betrachtet.

4.3.1 Berufszufriedenheit

Der Berufszufriedenheit von Lehrkräften wird in der Forschung eine hohe Bedeutung beigemessen. Sie ist Teil des beruflichen Wohlbefindens und steht in engem Zusammenhang mit der Selbstwirksamkeit von Lehrkräften (Bach, 2022). Darüber hinaus wird angenommen, dass die Berufszufriedenheit im Zusammenhang mit der eigenen Leistung steht und somit ausschlaggebend für die Qualität der eigenen Arbeit ist (Buer, Squarra, Ebermann-Richter & Kirchner, 1995). Für Berufszufriedenheit existieren in der Forschung verschiedene Definitionen. Ein vielzitierter Ansatz stammt von Merz (1979), der Berufs- bzw. Arbeitszufriedenheit definiert als „innerseelischer Zustand, der aus der emotional affektiven und rationalen Beurteilung des Arbeitsverhältnisses resultiert und mit dem Verhalten in einem gewissen Zusammenhang steht“ (S. 29). Dieterich und Rietz (1996) hingegen schlagen eine weniger restriktive Definition vor, nach der Zufriedenheit als Akzeptanz, welche sich jeder Mensch im Kontext des gesamten Lebens, eines Teilbereiches oder eines bestimmten Ereignisses selbst zuschreibt, bezeichnet werden kann. Einen anderen Aspekt berücksichtigen wiederum Homans und Prokop (1968), die in ihrer Definition auf das Verhältnis von erlangter und ausstehender Belohnung im Beruf als zentralen Punkt hinweisen. Somit meint Berufszufriedenheit im Kern eine Einschätzung der eigenen Berufs- bzw. Arbeitssituation. Wie auch in der Definition von Merz (1979) wird dabei deutlich, dass nicht durchgängig eine klare Trennung zwischen Berufszufriedenheit und Arbeitszu-

friedenheit vorgenommen wird. Eine in der Forschung viel genutzte Unterscheidung von berufsbezogenen Formen der Zufriedenheit, in der auch diese beiden Konzepte aufgegriffen werden, dreht sich um die insgesamt vier Formen *job satisfaction*, *vocational satisfaction*, *job attitude* und *industrial morale* (Bruggemann, Groskurth & Ulich, 1975). Unter *job satisfaction* versteht man dabei die Zufriedenheit mit der aktuellen Anstellung, also die Berücksichtigung eines konkreten Arbeitsverhältnisses mit seinen Arbeitsumständen. Für diese Form wird im Deutschen die Bezeichnung Arbeitszufriedenheit genutzt. Demgegenüber steht *voactional satisfaction*, zu Deutsch Berufszufriedenheit. Anders als die Arbeitszufriedenheit bezieht sich diese nicht auf eine konkrete Stelle, sondern auf die Berufswahl im Ganzen. Die dritte Form nennt sich *job attitude* und wird häufig mit Einstellung zur Arbeit übersetzt. In Abgrenzung zur Arbeitszufriedenheit bezieht sich diese aber auf konkrete Aspekte der Anstellung wie beispielsweise das Gehalt oder das Kollegium. Die vierte Form in diesem Modell wird als *industrial morale* bezeichnet, worunter das Zugehörigkeitsgefühl zur eigenen Berufsgruppe verstanden wird (Bruggemann et al., 1975). Die Berufszufriedenheit wird in diesem Sinne z. T. auch verkürzt als durchschnittliche Arbeitszufriedenheit über mehrere Anstellungen hinweg beschrieben (Crites, 1969). Der Fokus in der Forschung liegt zumeist auf diesen beiden ersten Formen der berufsbezogenen Zufriedenheit, häufiger werden dabei auch die Aspekte *job attitude* und *industrial morale* unter die anderen beiden Begriffe subsumiert (Ammann, 2004). Eine besonders häufige synonyme Verwendung der Begriffe Arbeits- und Berufszufriedenheit findet sich im Kontext des Lehrberufes, was auf die Besonderheit zurückzuführen ist, dass Lehrkräfte selten die Arbeitsstelle und den Arbeitgeber Staat wechseln. Deshalb wird für Lehrkräfte von einer hohen Überschneidung von Berufs- und Arbeitszufriedenheit ausgegangen (Ammann, 2004). Insgesamt bestätigen Studien immer wieder, dass Lehrkräfte im Durchschnitt eine hohe Berufszufriedenheit aufweisen und diese nicht vorrangig mit Hintergrundmerkmalen der Lehrkräfte wie Geschlecht, Alter oder Dauer der Berufsausübung zusammenhängen. Gezeigt werden konnte jedoch, dass Lehrkräfte an Grundschulen zufriedener sind als Lehrkräfte in anderen Schulformen (Bruggemann et al., 1975; Krampen, 1978; Merz, 1979; Shen, Leslie, Spybrook & Ma, 2012).

Diese und weitere Einflussfaktoren von Berufszufriedenheit standen auch bereits in zahlreichen weiteren Untersuchungen im Fokus. Weinert (2015) führt dazu aus, dass viele dieser Studien über Faktorenanalysen Bedingungsfaktoren von Arbeits- bzw. Berufszufriedenheit bestimmen. Als zentrale Dimensionen, die immer wieder nachgewiesen werden können, können die Arbeit selbst, Führungsstil und Supervision, Organisation und Organisationsleitung, Beförderungsmöglichkeiten, Mitarbeiter:in-

nen, Arbeitsbedingungen, finanzielle und nicht-finanzielle Be- und Entlohnungen sowie Anerkennung ausgemacht werden (Weinert, 2015). Studien, die konkret die Arbeits- und Berufszufriedenheit von Lehrkräften untersuchen, machen als Hauptursachen von Berufszufriedenheit zumeist pädagogische Aspekte, unterrichtliche Aktivitäten sowie die emotionale Qualität der Arbeit und erlebte Wirksamkeit im Unterricht, also Facetten der ersten Dimension der Arbeit an sich, aus (Buer et al., 1995; Caprara, Barbaranelli, Borgogni & Steca, 2003; Krampen, 1981; Morbitzer, 2009). Ebenfalls relevant sind für die Arbeits- und Berufszufriedenheit von Lehrkräften außerdem die Aspekte Sicherheit des Arbeitsplatzes und Bezahlung und somit Facetten der Dimensionen Arbeitsbedingungen und finanzielle Entlohnung. Hingegen eher selten werden Aufstiegschancen und die schulische Verwaltung als Ursachen für Berufszufriedenheit identifiziert (Buer et al., 1995; Krampen, 1981). Komplementär werden häufiger Interaktionsschwierigkeiten, z. B. mit Eltern, Erziehungsprobleme bei Schüler:innen und unterrichtliche Misserfolge als Ursachen für Unzufriedenheit benannt (Morbitzer, 2009).

Neben den Determinanten von Berufszufriedenheit stehen aber vor allem auch die Auswirkungen der Berufszufriedenheit von Lehrkräften auf schulische Aspekte im Fokus von Untersuchungen. Aus der Perspektive der positiven Psychologie, nach der Wohlbefinden den Berufserfolg bestimmt, wird für den Lehrberuf angenommen, dass die Berufszufriedenheit Auswirkungen auf die Qualität der eigenen Arbeit hat und Einfluss auf die Entwicklung und Förderung der Schüler:innen (Buer et al., 1995; Grams, 2014; Harrison, King & Wang, 2023; Hilgenfeld, 2012). Hilgenfeld (2012) fasst diese Sichtweise folgendermaßen zusammen: „Zufriedene Lehrer gehen mit mehr Freude, Motivation, Gefühl und Ausdauer täglich an ihren Arbeitsplatz. Diese Herangehensweise wirkt sich auf die Kinder aus“ (S. 124–125).

Insgesamt zeigt sich, dass Berufszufriedenheit in der Mehrheit der bisherigen Untersuchungen als abhängige Variable und zumeist im Kontext von Wohlbefindensforschung (also im Zusammenhang mit Faktoren wie Stress, Belastung, Kündigung oder Selbst- und Kontrollüberzeugungen) betrachtet wurde. Studien, die Berufszufriedenheit als erklärende Variable einbeziehen, sind hingegen deutlicher seltener, wobei dies insbesondere für den Kontext der Unterrichtsqualität bzw. -gestaltung und die kognitiven wie nicht-kognitiven Merkmale von Schüler:innen zutrifft. Dementsprechend sind auch Untersuchungen, die die in den Schulleistungsmodellen postulierte Wirkungskette von Lehrkräftemerkmalen über Unterrichtsmerkmale auf Merkmale der Schüler:innen untersuchen, rar. Entsprechend stellt sich die Befundlage für differenzielle Untersuchungen zur Berufszufriedenheit und besonderer Leistungsstärke

dar. Außer der im vorangegangenen Kapitel vorgestellten Untersuchung konnten keine weiteren Studien, die diesen Zusammenhang in den Blick nehmen, ausgemacht werden.

Entsprechend sind die Befunde für den Zusammenhang mit Merkmalen der Unterrichtsgestaltung und -qualität weniger passgenau als für die in den vorangegangenen Kapiteln beschriebenen Merkmale. In verschiedenen Studien konnte beispielsweise ein Zusammenhang zwischen der Berufszufriedenheit von Lehrkräften und positiven Ausprägungen motivational-affektiver Merkmale wie beispielsweise Motivation, Selbstwirksamkeit, Engagement und Enthusiasmus festgestellt werden (Judge, Weiss, Kammeyer-Mueller & Hulin, 2017; Meyer, Stanley, Herscovitch & Topolnytsky, 2002; Rupprecht, 2015). Ebenso zeigte sich wiederholt, dass eine höhere Arbeits- bzw. Berufszufriedenheit mit Merkmalen, die sich der konstruktiven Unterstützung bzw. dem Verhältnis zwischen Lehrkräften und Schüler:innen zuordnen lassen, einhergeht. So konnten Opdenakker und Van Damme (2006) zeigen, dass eine höhere Arbeitszufriedenheit mit einer höheren Unterstützung der Schüler:innen im Unterricht einhergeht. Speziell für Binnendifferenzierung konnte ein schwacher Zusammenhang mit der Berufszufriedenheit festgestellt werden (Kullmann, 2012). Ebenso konnten Lopes und Oliveira (2020) zeigen, dass zwischenmenschliche Interaktionen, insbesondere die im Unterricht, besonders mit Berufszufriedenheit zusammenhängen. Virtanen, Sørensen Vaaland und Ertesvåg (2019) kommen ebenfalls zu dem Schluss, dass beide Merkmale eng miteinander zusammenhängen und gehen davon aus, dass reziproke Effekte vorliegen, sodass sich beide Merkmale gegenseitig steigern können. Konsistent mit diesen Befunden zeigen sich für Merkmale von Burnout (z. B. emotionale Erschöpfung) Zusammenhänge mit einer geringen emotionalen Unterstützung der Schüler:innen (Klusmann, Aldrup, Roloff, Lüdtke & Hamre, 2022).

Ähnliche Befunde finden sich auch für die Merkmale Klassenführung und kognitive Aktivierung, wobei sich zeigt, dass konstruktive Unterstützung bzw. verwandte Merkmale deutlich häufiger als die anderen beiden Basisdimensionen der Unterrichtsqualität betrachtet werden. Dabei konnte für Klassenführung vereinzelt gezeigt werden, dass diese signifikant mit der Berufszufriedenheit und entgegengesetzt mit der emotionalen Erschöpfung der Lehrkräfte zusammenhängt (Klusmann et al., 2022; Opdenakker & Van Damme, 2006). Für kognitive Aktivierung hingegen können keine direkten Zusammenhänge mit der Berufszufriedenheit festgestellt werden. Allerdings zeigt eine Untersuchung von Harrison et al. (2023) einen indirekten Zusammenhang zwischen beiden Merkmalen, der über das von Lehrkräften wahrgenommene Verhältnis von Lehrkräften und Schüler:innen mediiert wird. Zusammenhänge mit der

Klassenführung zeigen sich dabei nur zum Teil. In einer Studie zum *inquiry-based teaching*, das Merkmale eines hohen Potenzials zur kognitiven Aktivierung zeigt, ergeben sich ebenfalls keine signifikanten direkten Zusammenhänge zwischen beiden Merkmalen (Perera et al., 2022). Eine der wenigen Untersuchungen, die zusätzlich zu Merkmalen des Unterrichts auch Merkmale der Schüler:innen als abhängige Variablen einbezieht, stammt von Pakarinen et al. (2010), die für Kinder im *kindergarten* (ähnlich eines vorbereitenden Jahres in einer Vorschule) Determinanten sprachlicher Fähigkeiten untersuchten. Dabei zeigte sich, dass sowohl das Stressempfinden der Lehrkräfte als auch die Unterrichtsqualität (basierend auf dem CLASS-Instrument) vermittelt über die Motivation der Schüler:innen auf deren Fähigkeiten wirkt. Die Zusammenhänge fallen dabei durchgängig moderat aus, der Zusammenhang zwischen Stress und Motivation ist im Gegensatz zu den anderen Effekten negativ und noch einmal höher als der Effekt von Unterrichtsqualität. Dabei werden die Unterrichtsqualität und das Stressempfinden der Lehrkräfte allerdings als separate Prädiktoren betrachtet, sodass keine Aussagen über eventuelle Mediationen zwischen diesen beiden Faktoren möglich sind. Der Zusammenhang zwischen Leistungen von Schüler:innen und der Berufszufriedenheit von Lehrkräften konnte auch in einer australischen Studie belegt werden. Dabei zeigte sich ein direkter Effekt der Arbeitszufriedenheit der Lehrkräfte auf die Leistung der Schüler:innen sowie die von Schüler:innen wahrgenommene Klassenführung, mediierende Effekte wurden auch in dieser Studie nicht näher untersucht. Ein direkter Zusammenhang mit Leistung konnte dabei für das Merkmal Wohlbefinden belegt werden (Collie & Martin, 2017). In einer längsschnittlich angelegten Studie können Caprara, Barbaranelli, Steca und Malone (2006) hingegen keinen Effekt der Berufszufriedenheit auf die Leistung der Schüler:innen ausmachen. Zusätzlich zeigt sich ein moderater Zusammenhang zwischen der Berufszufriedenheit und der Selbstwirksamkeit. In Untersuchungen mit *Large-Scale*-Daten können für emotionale Erschöpfung direkte Zusammenhänge mit den Leistungen der Schüler:innen und der wahrgenommenen Unterstützung durch die Lehrkraft festgestellt werden (Arens & Morin, 2016; Klusmann, Richter & Lüdtke, 2016). Anders als Pakarinen et al. (2010) zeigten sich dabei aber zum Teil keine Zusammenhänge mit motivationalen Merkmalen der Schüler:innen.

Die zahlreichen Einzelbefunde, die für verschiedene Facetten des Wohlbefindens von Lehrkräften wie Berufs- und Arbeitszufriedenheit, emotionale Erschöpfung und Stress vorliegen, zeigen neben einigen nicht signifikanten Befunden vor allem bedeutsame positive Zusammenhänge mit den Leistungen der Schüler:innen, vereinzelt mit

ihren motivationalen Merkmalen sowie mit Aspekten der Unterrichtsqualität. Indirekte Effekte der Arbeits- und Berufszufriedenheit über Merkmale des Unterrichts auf Merkmale der Schüler:innen wurden bisher kaum untersucht. Die bisherigen Befunde legen jedoch nahe, dass neben direkten Effekten auch indirekte Effekte, vermittelt über den Unterricht, existieren könnten.

4.3.2 Lehrkräfteselbstwirksamkeit

Die Selbstwirksamkeit gilt als eines der wichtigsten Merkmale von Lehrkräften im Zusammenhang mit ihrem Verhalten in Schule und Unterricht (Kunter & Trautwein, 2013; Tella, 2008). Nach Bandura (1977) kann Selbstwirksamkeit definiert werden als "conviction that one can successfully execute the behavior required to" (S. 193). Diese grenzt er von der Ergebniserwartung ab, die die Einschätzung beschreibt, dass ein bestimmtes Verhalten zu einem erwarteten Ergebnis führt. Im Lehrkräftekontext wird hier häufig die Erwartung benannt, dass die Bemühungen von Lehrkräften einen Einfluss auf die Leistungen der Schüler:innen haben können. Die Selbstwirksamkeit weist somit einen stärkeren Bezug zur eigenen Person auf und fokussiert darauf, ob eine Person sich selbst zutraut, bestimmte Herausforderungen erfolgreich zu bewältigen (Bandura, 1977; Ross, 1994). Eine hohe konzeptuelle Nähe weist die Selbstwirksamkeit somit zum Konstrukt des Selbstkonzeptes auf. Die wesentliche Unterscheidung liegt dabei in dem Einschätzungsgegenstand. Während bei der Selbstwirksamkeit in der Regel die eigenen Kompetenzen in Bezug auf die Bewältigung enggefasster, konkreter zukünftiger Ereignisse, Herausforderungen oder Aufgaben eingeschätzt werden, bezieht sich das Selbstkonzept auf eine globalere, relationale Einschätzung (z. B. eine fachliche Domäne) und schließt dabei sowohl kognitive, affektive als auch evaluative Dimensionen mit ein. Dabei generiert sich das Selbstkonzept über verschiedene interne und externe Vergleichsmechanismen (siehe dafür auch Kapitel 4.1.1) und steht somit in engem Zusammenhang mit der Umwelt, während für die Selbstwirksamkeit angenommen wird, dass vor allem selbstbezogene Informationen wie vorherige Erfahrungen, aber auch stellvertretende Erfahrungen, verbale Überzeugungen und die psychologische sowie affektive Verfassung relevant sind (Bandura, 1977; Köller & Möller, 2010; Kriegbaum et al., 2018). Vorherigen Erfahrungen wird deshalb ein besonders hoher Stellenwert beigemessen, weil es sich um persönliche Erfolgs- oder auch Misserfolgserlebnisse handelt und diese somit sehr eng mit der eigenen Kompetenzein-

schätzung zusammenhängen. Dementsprechend werden stellvertretende Erfahrungen als weniger bedeutsam eingeschätzt. Durch das Beobachten anderer Personen, die beispielweise eine herausfordernde Situation bewältigen, kann jedoch das eigene Vermögen, in ähnlicher Weise erfolgreich zu sein, positiver beurteilt werden. Ebenso können verbale Überzeugungen, i. d. R. die positive Bestärkung durch Dritte, einen gewissen Einfluss auf die Selbstwirksamkeit haben. Neben Kompetenzerfahrungen und anderen Personen können zudem auch psychologische und affektive Reaktionen des Körpers die eigene Selbstwirksamkeit beeinflussen. So kann das Ausmaß dieser Reaktionen in herausfordernden Situationen, z. B. Angst oder freudige Erregtheit, als Anzeichen bzw. Information dafür dienen, über welches Kompetenzvermögen zur Bewältigung man selbst verfügt (Bach, 2022; Bandura, 1977). Neben den Quellen für das jeweilige Konstrukt unterscheiden sich Selbstkonzept und Selbstwirksamkeit zudem hinsichtlich ihrer angenommenen Stabilität. Während davon ausgegangen wird, dass es sich beim Selbstkonzept um ein eher stabiles, überdauerndes Konstrukt handelt, ist die Selbstwirksamkeit dynamischer und anlassbezogener (Breker, 2016). Damit geht ähnlich wie beim Selbstkonzept eine gewisse Bereichsspezifität der Selbstwirksamkeit einher (Gebauer, 2013). Dementsprechend wird in der Forschung häufig auch von einer konkreten Lehrkräfteselbstwirksamkeit gesprochen, die als Überzeugung verstanden wird, berufsbezogene Anforderungen sowie schwierige Situationen im Schulalltag allein bewältigen zu können (Bach, 2022; Florin, Stebler, Pauli & Reusser, 2022; Gebauer, 2013). Diese vereint wiederum mehrere bereichsspezifische Lehrkräfteselbstwirksamkeiten. So konnten beispielsweise Skaalvik und Skaalvik (2007) sechs unterschiedliche Bereiche (*instruction, adapting education to individual students' needs, motivating students, keeping discipline, cooperating with colleagues and parents* und *coping with changes and challenge*), Ma, Luo, Cavanagh, Dong und Sun (2022) hingegen zwei globalere Faktoren (*ethos* und *teaching*) ausmachen. In der bekanntesten Skala der deutschsprachigen Forschung legen Schwarzer und Jerusalem (1999) fünf Facetten der Lehrkräfteselbstwirksamkeit zugrunde: berufliche Leistung, berufliche Weiterentwicklung, soziale Interaktionen mit Schülern, Eltern und Kollegen sowie Umgang mit Berufsstress.

Die Spezifität der Lehrkräfteselbstwirksamkeit wurde außer in Zusammenhang mit der Validierung von neuentwickelten Testinstrumenten in der bisherigen Forschung eher selten untersucht (Ausnahme besipielsweise Thommen, Grob, Lauermann, Klassen & Praetorius, 2022). Ähnlich gering ausgeprägt ist die Befundlage zur Stabilität der Lehrkräfteselbstwirksamkeit (Klassen, Tze, Betts & Gordon, 2011), wobei

bisherige Befunde keine eindeutigen Hinweise geben. So gibt es sowohl Studien, die auf eine hohe Stabilität der Selbstwirksamkeit hindeuten (Künstning et al., 2016), als auch Studien, deren Befunde eher eine höhere Dynamik vermuten lassen (Holzberger, Philip & Kunter, 2013; Woolfolk Hoy & Burke Spero, 2005).

Die Unterscheidung zwischen der globalen und der bereichsspezifischen Lehrkräfteselbstwirksamkeit hat auch Eingang in das heuristische Modell der Selbstwirksamkeit für den Schulkontext nach Zee und Koomen (2016) gefunden. Dieses wurde mithilfe eines umfangreichen Reviews auf Basis der in der bisherigen Forschung angenommenen, reziproken Zusammenhänge von Lehrkräfteselbstwirksamkeit mit verschiedenen schulrelevanten Bereichen entwickelt (siehe Abbildung 8). Durchgezogene Pfeile repräsentieren dabei die Zusammenhänge, die im Rahmen des Reviews explizit untersucht wurden, gestrichelte Pfeile jene, die plausibel sind, deren Überprüfung allerdings nicht Teil des Reviews war.

Nach dem Modell wird davon ausgegangen, dass die Lehrkräfteselbstwirksamkeit direkten Einfluss auf die Qualität des Unterrichts, die kognitiven sowie motivationalen Outcomes der Schüler:innen und das Wohlbefinden der Lehrkräfte hat, wobei alle Wirkrichtungen auch andersherum gedacht werden können. Darüber hinaus hängt die Unterrichtsqualität sowohl mit den Schüleroutcomes als auch dem Wohlbefinden der Lehrkräfte zusammen. Zwischen diesen beiden Konstrukten wird zudem ebenfalls ein Zusammenhang angenommen. Der Zusammenhang zwischen der Lehrkräfteselbstwirksamkeit und der Unterrichtsqualität ist nach Zee und Koomen (2016) dabei am besten untersucht, wobei die Befundlage für die Subfacette emotionale Unterstützung trotzdem gering ausfällt. Für die Leistungen und motivationalen Merkmale der Schüler:innen stellen die Autorinnen hingegen fest, dass die meisten Publikationen die Zusammenhänge mit der Lehrkräfteselbstwirksamkeit nur theoretisch betrachten und empirische Befunde deutlich seltener sind. Diese deuten jedoch darauf hin, dass sowohl direkte als auch indirekte Effekte vorstellbar sind. Für den Bereich Wohlbefinden finden sich vorrangig Untersuchungen, die Lehrkräfteselbstwirksamkeit im Zusammenhang mit negativen Aspekten des Wohlbefindens betrachten. Bisherige Befunde zeigen allerdings, dass die Selbstwirksamkeit stärker mit den positiven Facetten zusammenhängt (Zee & Koomen, 2016).

So kommen bisherige Untersuchungen zu dem Schluss, dass eine hohe Lehrkräfteselbstwirksamkeit mit einem höheren Enthusiasmus und Engagement, einer höheren Bindung an den Lehrberuf und einer höheren Berufszufriedenheit einhergeht (Skaalvik & Skaalvik, 2017; Trentham, Silvern & Brogdon, 1985; Tschannen-Moran,

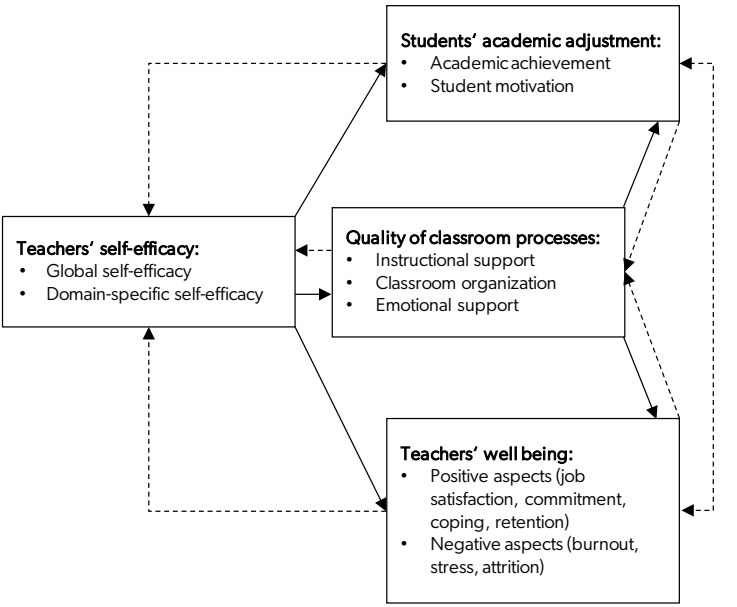


Abb.: 8 Heuristisches Modell der Selbstwirksamkeit nach Zee und Koomen (2016)

Woolfolk Hoy & Hoy, 1998; Zee & Koomen, 2016). Dabei ist die Mehrheit der bisherigen Untersuchungen allerdings querschnittlich angelegt, sodass keine empirischen Aussagen über die Wirkrichtung der Zusammenhänge möglich sind. In einer umfangreichen Aufarbeitung der bisherigen Forschung zur Lehrkräfteselbstwirksamkeit kommt Bach (2022) ebenfalls zu diesem Schluss und resümiert, dass, wenn vorhanden, vor allem indirekte und direkte Effekte von Lehrkräfteselbstwirksamkeit auf Berufszufriedenheit nachgewiesen werden konnten. Dabei verweist er gleichzeitig auf fehlende Untersuchungen, die die gegensätzliche Richtung überprüfen, sowie auf neuere Publikationen, in denen von einem reziproken Zusammenhang ausgegangen werden kann (Bach, 2022). In zunehmendem Maße werden darüber hinaus auch Zusammenhänge der Lehrkräfteselbstwirksamkeit mit Unterrichtsmerkmalen oder mit Merkmalen der Schüler:innen betrachtet. Grundlegende Annahme dieser Untersuchungen ist, dass selbstwirksame Lehrkräfte in der Lage sind, eine lernförderlichere Umgebung sowie einen qualitativ hochwertigeren Unterricht zu gestalten und so Einfluss auf die Schüler:innen nehmen (Perera et al., 2022). Studien, die diesen

Zusammenhang in Gänze und insbesondere mit Fokus auf die motivationalen Orientierungen der Lehrkräfte abbilden, sind bisher rar, da in den meisten Fällen nur Teile dieser angenommenen Zusammenhänge untersucht werden. Entsprechend konnten auch keinerlei Studien, außer der in Kapitel 4.3 aufgeführten, ausgemacht werden, die einen differenziellen Ansatz verfolgen. In diesem Abschnitt können deshalb ausschließlich generelle Befunde herangezogen werden.

In einer Untersuchung mit Daten von PIRLS 2006 bilden Arens und Morin (2016) die Zusammenhänge über das Lehrkräftemerkmal Erschöpfung, das Unterrichtsmerkmal wahrgenommene Unterstützung, das Schülermerkmal Selbstkonzept und die Leistung (sowohl über Noten als auch die Ergebnisse im Kompetenztest) ab. Dabei zeigten sich kleine signifikante negative Effekte der Erschöpfung auf die wahrgenommene Unterstützung, die wiederum sowohl auf die Noten als auch auf die Testleistungen kleine signifikante positive Effekte hatte. Zudem zeigten sich kleine bis mittlere signifikante positive Effekte des Selbstkonzeptes auf die Leistungen sowie direkte kleine signifikante negative Effekte der Erschöpfung auf die Leistung. Sie konnten in ihrer Untersuchung somit zeigen, dass Erschöpfung vermittelt über die wahrgenommene Unterstützung die Leistungen beeinflusst, wobei sich zusätzlich auch ein direkter Effekt der Erschöpfung auf die Leistung zeigte (Arens & Morin, 2016). In einer anderen Studie mit TIMSS-2015-Daten untersuchten Perera et al. (2022), inwiefern die Lehrkräfteselbstwirksamkeit die wahrgenommene Unterstützung und den Einsatz von *inquiry teaching* beeinflusst und so wiederum Einfluss auf die naturwissenschaftlichen Leistungen nimmt. Dabei konnten moderate signifikante positive Effekte der Lehrkräfteselbstwirksamkeit auf *inquiry teaching*, jedoch nicht auf wahrgenommene Unterstützung gefunden werden. Beide Unterrichtsmerkmale hängen darüber hinaus schwach signifikant mit der Leistung zusammen, wobei dieser Zusammenhang für *inquiry teaching* positiv, für wahrgenommene Unterstützung negativ ausfällt. Zudem zeigte sich ein moderater signifikanter positiver Einfluss der Lehrkräfteselbstwirksamkeit auf die Berufszufriedenheit (Perera et al., 2022). In einer weiteren Studie zum Sachunterricht untersuchen Fauth et al. (2019) für Drittklässler:innen, wie die professionelle Kompetenz von Lehrkräften mit den Basisdimensionen der Unterrichtsqualität und dem Interesse und der Leistung zusammenhängt. Dabei zeigen sich vor allem schwache bis moderate signifikante positive Effekte der Lehrkräfteselbstwirksamkeit auf alle drei Basisdimensionen, die, wie in Kapitel 4.2.1 ausführlicher dargestellt, zum Teil wiederum moderate signifikante positive Effekte auf das Interesse und die Leistung aufweisen (Fauth et al., 2019).

Studien, die zwar keine Leistungen, stattdessen aber nicht-kognitive Outcomes untersuchen, kommen zu gemischten Ergebnissen. So konnte beispielsweise gezeigt werden, dass die Lehrkräfteselbstwirksamkeit auf unterschiedliche Unterrichtsmerkmale (prozessorientiertes Unterrichten, Lebensweltbezug, kooperatives Lernen und Differenzierung) wirkt, die aber wiederum entweder keinen oder nur einen kleinen signifikanten Einfluss auf die intrinsische Motivation und das Wohlbefinden der Schüler:innen haben (Thoonen, Sleegers, Peetsma & Oort, 2011). Für Unterricht, der ein tiefes Verständnis fördert und die Schüler:innen kognitiv aktiviert, fanden Schiefele und Schaffner (2015) einen signifikanten Zusammenhang mit der Lehrkräfteselbstwirksamkeit, wenn dieser durch die Lehrkräfte eingeschätzt wurde. Effekte der Unterrichtsmerkmale auf die Motivation der Schüler:innen fanden sich hingegen nur, wenn diese von Schüler:innen eingeschätzten wurden. Für die Basisdimensionen der Unterrichtsqualität konnten Burić und Kim (2020) wiederum nachweisen, dass die Lehrkräfteselbstwirksamkeit signifikant mit allen drei Merkmalen zusammenhängt und diese ebenso alle Einfluss auf die intrinsische Motivation der Schüler:innen haben. Die gefundenen Effekte sind dabei schwach bis moderat ausgeprägt und somit konsistenter und größer als bei Thoonen et al. (2011).

Befunde, die sich schwerpunktmäßig mit dem Zusammenhang zwischen Lehrkräfteselbstwirksamkeit und Unterrichtsqualität auseinandersetzen, bestätigen im Wesentlichen diesen Befund. So konnte bereits in mehreren Metaanalysen ein bedeutsamer Zusammenhang nachgewiesen werden (Hidayah, Wangid, Wuryandani & Salimi, 2023; Klassen & Tze, 2014; Kuusinen, 2016). In einer detaillierten Untersuchung der Lehrkräfteselbstwirksamkeit und der drei Basisdimensionen der Unterrichtsqualität, bei der pro Dimension zwei Subfacetten erfasst und zusätzlich das übergeordnetes Konstrukt *Instructional Practice* gebildet wurde, zeigen sich wie in der Untersuchung von Burić und Kim (2020) durchgängig signifikante positive Effekte. So findet sich ein moderater Zusammenhang der Lehrkräfteselbstwirksamkeit mit der *Instructional Practice*, bei detaillierterer Betrachtung zeigen sich schwache bis moderate Zusammenhänge mit den einzelnen Subfacetten der Basidimensionen, die höchsten für die beiden Subfacetten der kognitiven Aktivierung (Depaepe & König, 2018). Keine signifikanten Zusammenhänge zwischen der Lehrkräfteselbstwirksamkeit und Merkmalen der Unterrichtsqualität konnten hingegen Vieluf, Kunter und van de Vijver (2013) finden.

Eine der wenigen längsschnittlichen Studien in diesem Bereich kommt zu dem Ergebnis, dass es keine oder schwache Effekte der Selbstwirksamkeit auf die Unterrichtsqualität gibt. Stattdessen zeigen sich eher Hinweise auf reziproke Zusammen-

hänge zwischen zwei Schuljahren (Holzberger et al., 2013). Für größere Zeiträume und mehr Messzeitpunkte konnten Künstning et al. (2016) wiederum nachweisen, dass die Lehrkräfteselbstwirksamkeit ein stabiler Langzeitprädiktor der Unterrichtsqualität ist, wobei sich die höchsten Zusammenhänge für das Klassenklima fanden.

Während der Zusammenhang der Lehrkräfteselbstwirksamkeit mit der Unterrichtsqualität bereits gut untersucht ist, stand der Zusammenhang mit innerer Differenzierung seltener im Fokus. Werden Formen der inneren Differenzierung betrachtet, geschieht dies häufig im Kontext des inklusiven Unterrichtens. Es wurde bereits vielfach nachgewiesen, dass eine hohe Lehrkräfteselbstwirksamkeit generell mit einer höheren Experimentierfreude und Offenheit bei unterrichtlichen Tätigkeiten, mit einer stärkeren Vorbereitung und Planung des Unterrichts und mit einer höheren Innovationsbereitschaft einhergeht (Allinder, 1994; Döring-Seipel & Dauber, 2010; Smylie, 1988; Tschannen-Moran & McMaster, 2009). So konnte beispielsweise gezeigt werden, dass die Lehrkräfteselbstwirksamkeit im Bereich Unterrichtsmethoden in moderater Höhe mit dem Ausmaß an innerer Differenzierung im Unterricht zusammenhängt (Dixon, Yssel, McConnell & Hardin, 2014). Schwache bis moderate signifikante positive Zusammenhänge der Lehrkräfteselbstwirksamkeit mit innerer Differenzierung fanden beispielsweise Schiefele, Streblow und Retelsdorf (2013) sowie Suprayogi, Valcke und Godwin (2017). Kein signifikanter Zusammenhang zwischen Lehrkräfteselbstwirksamkeit und der Individualisierungspraxis im Unterricht ergab sich bei Schüle, Schrieck, Besa und Arnold (2016). Im Kontext inklusiven Unterrichtens zeigte sich darüber hinaus, dass die Lehrkräfteselbstwirksamkeit der stärkste Prädiktor für den Einsatz inklusiver Unterrichtsmethoden (operationalisiert über Differenzierung und Individualisierung) und damit bedeutsamer ist als die Einstellung, die Schulform oder die Lehrerfahrung (Schwab & Alnahdi, 2020).

Anders als indirekte Effekte der Lehrkräfteselbstwirksamkeit wurden direkte Zusammenhänge der Lehrkräfteselbstwirksamkeit auf die Leistungen nur selten überprüft. Direkte Effekte finden sich in der bisherigen Forschung stattdessen vor allem für aggregierte Leistungen (Caprara et al., 2006; Perera & John, 2020). In einer der wenigen Studien mit Individualleistungen konnte Tella (2008) zeigen, dass es einen kleinen direkten Effekt der Lehrkräfteselbstwirksamkeit auf die Mathematikleistungen von Grundschüler:innen gibt. Gleichzeitig zeigte sich auch ein direkter Effekt auf das Interesse der Schüler:innen. Darüber hinaus muss auch die Befundlage für direkte Zusammenhänge von Lehrkräfteselbstwirksamkeit und nicht-kognitiven Merkmalen von Schüler:innen als äußerst unzureichend bezeichnet werden.

5

Forschungsfrage und Ableiten der Hypothesen

Die Darstellungen im vorangegangenen Kapitel haben gezeigt, dass die Bereiche Lehrkräfte, Unterricht und Schüler:innen viele bedeutsame Bedingungsfaktoren von schulischen Leistungen umfassen. Im Kontext besonderer Leistungsstärke scheinen dabei auf Basis vorliegender leistungsdifferenzieller Befunde insbesondere die Merkmale fachspezifisches Selbstkonzept der Schüler:innen, die Einstellung der Schüler:innen zum Fach, die kognitiven Fähigkeiten, das Geschlecht, der sozioökonomische Hintergrund, der Migrationshintergrund, die Klassenführung, die kognitive Aktivierung, die konstruktive Unterstützung, der Einsatz binnendifferenzierender Maßnahmen im Unterricht, die Berufszufriedenheit der Lehrkraft sowie die Lehrkräfteselbstwirksamkeit zu sein. In der bisherigen Literatur konnten für diese zwölf Merkmalen direkte und indirekte Effekte auf die Leistung nachgewiesen werden. Die Grundannahme des in Kapitel 3 vorgestellten Rahmenmodells zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen, dass Lehrkräfte im Sinne einer Wirkungskette über den Unterricht und die individuellen Merkmale der Schüler:innen auf Leistungen Einfluss haben können, kann somit nach den bisherigen Forschungen auch für die Gruppe der leistungsstarken Schüler:innen als plausibel angesehen werden. Einige Befunde deuten jedoch darauf hin, dass Merkmale der Lehrkräfte und des Unterrichts darüber hinaus auch direkte Effekte auf die Leistungen der Schüler:innen haben können. Insgesamt zeigen die Darstellungen in den vorherigen Kapiteln, dass die theoretischen Grundlagen zur Definition leistungsstarker Schüler:innen sowie die Forschungslage hinsichtlich differenzieller Befunde gering ausfallen und deshalb wenig darüber bekannt ist, wie die Zusammenhänge von Merkmalen von Schüler:innen, Unterricht und Lehrkräften mit schulischen Leistungen für diese Schülergruppe ausgeprägt sind. Aus diesem Grund werden in der vorliegenden Arbeit mögliche direkte und indirekte Prädiktoren besonderer Leistungsstärke daraufhin untersucht, inwiefern sie dazu beitragen, dass Schüler:innen nicht nur gute Leistungen auf Kompetenzstufe (KS) IV, sondern auch Leistungen, die auf KS V verortet werden können, erreichen und somit ihr Potenzial besser ausschöpfen. Die Hypothesen, die dabei den Analysen zugrunde liegen, werden im Folgenden getrennt für die Merkmale der Schüler:innen, des Unterrichts und der Lehrkräfte dargestellt.

5.1 Hypothesen zu den Merkmalen der Schüler:innen

Das fachbezogene Selbstkonzept und die fachbezogene Einstellung konnten in der bisherigen Forschung als bedeutsame Prädiktoren schulischer Leistungen identifiziert werden. Für die Gruppe der leistungsstarken Schüler:innen zeigen sich zudem höhere Ausprägungen für beide Merkmale als für leistungsschwächere Schüler:innen (Ai, 2002; Drüke-Noe et al., 2022; Marsh et al., 2005; Vachon, 1986). Darüber hinaus konnte bereits auch in differenziellen Untersuchungen die Bedeutung dieser beiden Merkmale belegt werden (Neuendorf et al., 2017; Pitsia, 2022; Tuorón et al., 2018). So konnte beispielsweise gezeigt werden, dass der Zusammenhang zwischen fachbezogenem Selbstkonzept und Leistung umso stärker ausfällt, desto leistungsfähiger Personen sind (Hansford & Hattie, 1982) und dass die fachbezogene Einstellung besser für die Prädiktion von Leistungsstärke als von Leistungsschwäche geeignet ist (Grootenboer & Hemmings, 2007). Beide Merkmale sind somit vielversprechende Merkmale zur Erklärung besonderer Leistungsstärke. Dementsprechend lauten die ersten beiden Hypothesen:

(H1) Die fachbezogene Einstellung der Schüler:innen zum Fach beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.

(H2) Das fachbezogene Selbstkonzept der Schüler:innen beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.

Neben dem fachbezogenen Selbstkonzept und der fachbezogenen Einstellung stellten sich auch die Hintergrundmerkmale kognitive Fähigkeiten, Geschlecht, sozioökonomischer Hintergrund und Migrationshintergrund als mögliche Prädiktoren von besonderer Leistungsstärke heraus (Forgasz & Hill, 2013; Gilleece et al., 2010; Schmidtner, 2017; Wendt et al., 2013). Da diese jedoch nur eingeschränkt beeinflussbar und somit förderbar sind, werden sie nicht als Prädiktoren, sondern als Kontrollvariablen ohne gesonderte Hypothesen hinzugezogen. Die Hypothesen zu den individuellen Schülermerkmalen sind in Abbildung 9 veranschaulicht.

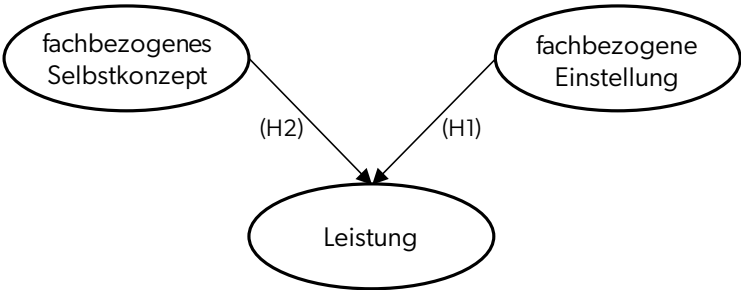


Abb.: 9 Hypothesen zu individuellen Merkmalen der Schüler:innen

5.2 Hypothesen zu den Merkmalen des Unterrichts

Für die Leistungen der Schüler:innen gehört die Unterrichtsqualität zu den bedeutsamsten Einflussfaktoren aus dem Bereich des Unterrichts. In Untersuchungen zu den drei Basisdimensionen der Unterrichtsqualität nach Klieme et al. (2001) konnten bedeutsame positive Zusammenhänge mit der Leistung vor allem für die Klassenführung und die kognitive Aktivierung nachgewiesen werden (Fauth, Decristan, Rieser, Klieme & Büttner, 2014a; Kunter et al., 2013; Schiepe-Tiska et al., 2016). In differenziellen Untersuchungen finden sich jedoch Hinweise darauf, dass insbesondere die konstruktive Unterstützung für das Erreichen besonderer Leistungsstärke bedeutsam sein könnte (Affandi et al., 2020; Sanfo & Malgoubri, 2021), weshalb für den vorliegenden Kontext zudem ein Zusammenhang zwischen konstruktiver Unterstützung und Leistung vermutet werden kann. Für nicht-kognitive Merkmale von Schüler:innen zeigen sich vorwiegend positive Zusammenhänge (beispielsweise für Einstellung und Interesse; Fauth et al., 2019; Klieme et al., 2001; Scherer et al., 2016; Schiepe-Tiska et al., 2016), z. T. aber auch negative Zusammenhänge (z. B. mit dem Selbstkonzept; Henschel et al., 2019; Praetorius et al., 2018).

Zusätzlich zur Unterrichtsqualität verweisen sowohl theoretische Überlegungen als auch empirische Befunde auf eine mögliche Bedeutsamkeit der inneren Differenzierung zur Prädiktion besonderer Leistungsstärke. Generelle Befunde legen schwache positive Zusammenhänge zwischen innerer Differenzierung und Leistung sowie Selbstkonzept und Einstellung nahe (Böhnel & Svik, 1993; Guay et al., 2017; Hattie, 2009; Lipowsky & Lotz, 2015; Lou et al., 1996). Für Schüler:innen am Ende der

Grundschulzeit zeigen sich dabei die höchsten Zusammenhänge mit der Leistung (Lou et al., 1996), für besonders leistungsstarke Schüler:innen konnten vorwiegend positive Auswirkungen binnendifferenzierender Maßnahmen nachgewiesen werden (Kulik & Kulik, 1992; Ziernwald et al., 2022). Für motivational-affektive Merkmale liegen bisher keine differenziellen Befunde zum Zusammenhang mit innerer Differenzierung vor, Befunde zu äußerer Differenzierung lassen jedoch vermuten, dass positive Effekte vorliegen könnten (Letzel, 2021; Ziernwald et al., 2020).

Da sich somit in der bisherigen Forschung sowohl direkte als auch indirekte Zusammenhänge der Unterrichtsmerkmale mit der Leistung zeigen, werden für beide Zusammenhangsformen Hypothesen formuliert (siehe Abbildung 10 und 11).

(H3) Die konstruktive Unterstützung beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H4) Die kognitive Aktivierung beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H5) Die Klassenführung beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H6) Die individuelle Förderung beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.

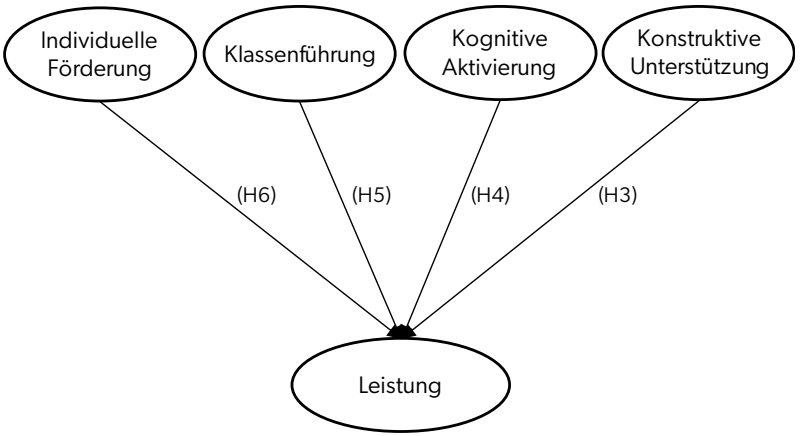


Abb.: 10 Hypothesen zu Merkmalen des Unterrichts I

(H7) Die konstruktive Unterstützung hängt positiv mit der fachbezogenen Einstellung zusammen.
(H8) Die konstruktive Unterstützung hängt positiv mit dem fachbezogenen Selbstkonzept zusammen.
(H9) Die kognitive Aktivierung hängt positiv mit der fachbezogenen Einstellung zusammen.
(H10) Die kognitive Aktivierung hängt positiv mit dem fachbezogenen Selbstkonzept zusammen.
(H11) Die Klassenführung hängt positiv mit der fachbezogenen Einstellung zusammen.
(H12) Die Klassenführung hängt negativ mit dem fachbezogenen Selbstkonzept zusammen.
(H13) Die individuelle Förderung hängt positiv mit der fachbezogenen Einstellung zusammen.
(H14) Die individuelle Förderung hängt positiv mit dem fachbezogenen Selbstkonzept zusammen.

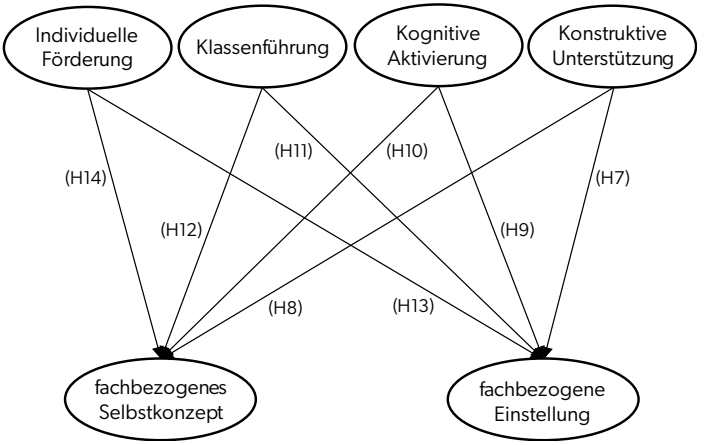


Abb.: 11 Hypothesen zu Merkmalen des Unterrichts II

5.3 Hypothesen zu den Merkmalen der Lehrkräfte

Aus dem Bereich der Lehrkräfte konnten Merkmale der professionellen Kompetenzen als mögliche Prädiktoren besonderer Leistungsstärke identifiziert werden. Dabei scheinen auf Basis der differenziellen Untersuchung von Rowen et al. (1997) insbesondere die Bereiche Selbstregulation und Motivationale Orientierungen vielversprechend zu sein. In bisherigen Untersuchungen konnte gezeigt werden, dass die Merkmale Selbstwirksamkeit (Fauth et al., 2019; Tella, 2008) und Berufszufriedenheit (Arens & Morin, 2016; Banerjee et al., 2017; Collie & Martin, 2017) bedeutsam und positiv mit der Leistung der Schüler:innen zusammenhängen. Die meisten Studien untersuchen jedoch weniger direkte Effekte der beiden Merkmale auf die Leistung als indirekte Effekte über Merkmale der Unterrichtsqualität und weitere individuelle Schülermerkmale. Es konnten bisher insbesondere bedeutsame positive Zusammenhänge beider Merkmale mit den Unterrichtsmerkmalen konstruktive Unterstützung, Klassenführung und Formen der inneren Differenzierung ausgemacht werden (Dixon et al., 2014; Fauth et al., 2019; Hidayah et al., 2023; Kullmann, 2012; Opdenakker & Van Damme, 2006). Einzig für den Zusammenhang zwischen Berufszufriedenheit und kognitiver Aktivierung konnten in bisherigen Untersuchungen wiederholt keine signifikanten Effekte nachgewiesen werden (z. B. Perera et al., 2022). Da somit sowohl Hinweise auf direkte als auch indirekte Effekte der Berufszufriedenheit und Lehrkräfteselbstwirksamkeit auf die Leistungen von Schüler:innen vorliegen, werden für beide Zusammenhangsformen Hypothesen formuliert und in Abbildung 12 und 13 visualisiert:

(H15) Die Lehrkräfteselbstwirksamkeit beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H16) Die Berufszufriedenheit beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.

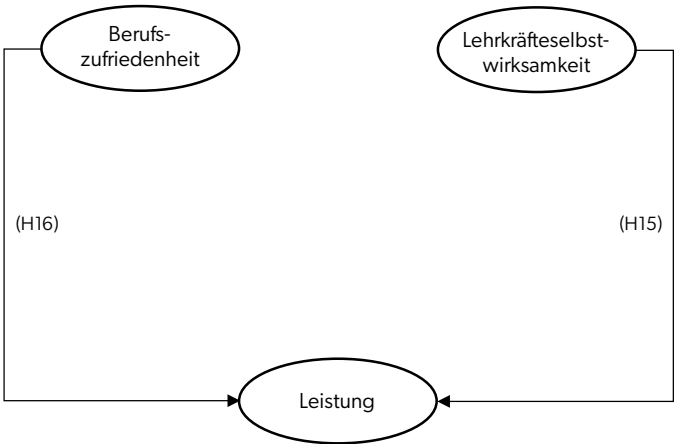


Abb.: 12 Hypothesen zu den Merkmalen der Lehrkraft I

(H17) Die Lehrkräfteselbstwirksamkeit hängt positiv mit der konstruktiven Unterstützung zusammen.
(H18) Die Lehrkräfteselbstwirksamkeit hängt positiv mit der kognitiven Aktivierung zusammen.
(H19) Die Lehrkräfteselbstwirksamkeit hängt positiv mit der Klassenführung zusammen.
(H20) Die Lehrkräfteselbstwirksamkeit hängt positiv mit der individuellen Förderung zusammen.
(H21) Die Berufszufriedenheit hängt positiv mit der konstruktiven Unterstützung zusammen.
(H22) Die Berufszufriedenheit hängt nicht mit der kognitiven Aktivierung zusammen.
(H23) Die Berufszufriedenheit hängt positiv mit der Klassenführung zusammen.
(H24) Die Berufszufriedenheit hängt positiv mit der individuellen Förderung zusammen.

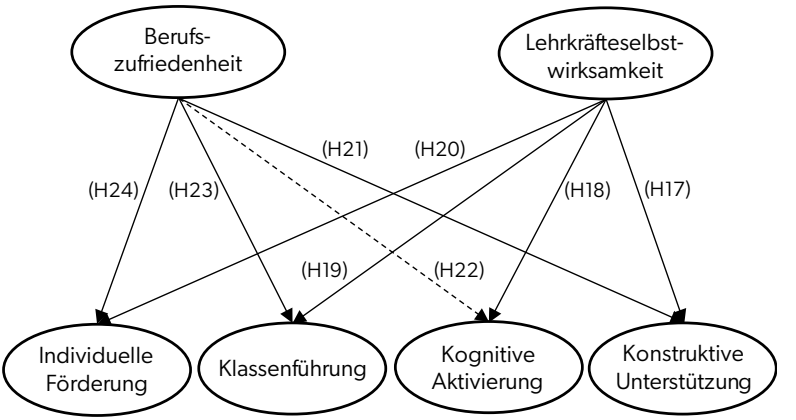


Abb.: 13 Hypothesen zu den Merkmalen der Lehrkraft II

5.4 Zusammenfassung

In Abbildung 14 sind die Hypothesen aus den Bereichen Schüler:innen, Unterricht und Lehrkräfte gemeinsam dargestellt. Durchgezogene Pfeile stehen dabei für angenommene positive Zusammenhänge zwischen zwei Merkmalen. Ausnahme bildet der Zusammenhang zwischen kognitiver Aktivierung und fachbezogenem Selbstkonzept, für den eine negative Ausprägung erwartet wird, und der deshalb mit einem Minus-Zeichen gekennzeichnet ist. Der gestrichelte Pfeil zwischen Berufszufriedenheit und kognitiver Aktivierung wiederum steht dafür, dass kein signifikanter Zusammenhang zwischen den beiden Merkmalen erwartet wird. Wie beschrieben, werden Zusammenhänge mit Kontrollvariablen nicht als eigene Hypothesen formuliert und sind deshalb nicht in Abbildung 14 enthalten.

Neben den angenommenen Zusammenhängen kann auf Basis der bisherigen Untersuchungen auch vermutet werden, dass die Merkmale der einzelnen Bereiche miteinander korrelieren. Da diese Beziehungen jedoch nicht im Fokus der vorliegenden Untersuchung stehen, werden dazu ebenfalls keine Hypothesen formuliert. Bei den Berechnungen werden Korrelationen innerhalb der einzelnen Bereiche zugelassen und berichtet, jedoch nicht umfassend ausgewertet.

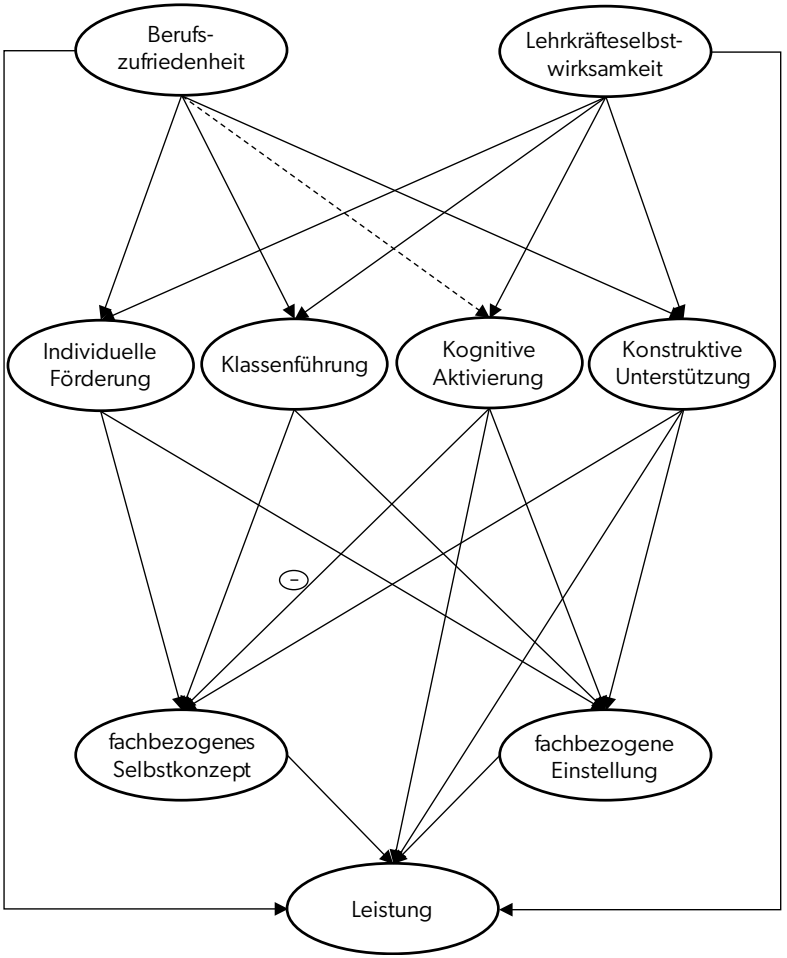


Abb.: 14 Grafische Übersicht über die Hypothesen

II Methodischer Teil

6

Datengrundlage: TIMSS 2019

Da sich im Rahmen von TIMSS 2019 zum inzwischen vierten Mal gezeigt hat, dass in Deutschland nur wenige Schüler:innen die höchste Kompetenzstufen erreichen und der Anteil trotz wiederholter Forderungen zur stärkeren Berücksichtigung und Förderung leistungsstarker Schüler:innen seit 2007 noch gesunken ist, werden in der vorliegenden Arbeit die Daten von TIMSS 2019 genutzt, um die leistungsstarken Schüler:innen näher zu untersuchen und schwerpunktmäßig in den Blick zu nehmen. Dazu werden in diesem Kapitel zunächst die TIMS-Studie, die Stichprobe von TIMSS 2019 und ausgewählte für diese Arbeit relevante Test- und Befragungsinstrumente vorgestellt.

6.1 Hintergründe und Ziele der TIMS-Studie

Die Daten, die den Analysen in dieser Arbeit zu Grunde liegen, entstammen der *Trends in International Mathematics and Science Study (TIMSS)* 2019, die in Deutschland von Prof. Dr. Knut Schwippert als wissenschaftlichem Leiter an der Universität Hamburg verantwortet wurde. Initiiert wurde die TIMS-Studie von der *International Association for the Evaluation of Educational Achievement* (IEA), die auch für die Durchführung anderer großangelegter Schulvergleichsuntersuchungen wie PISA oder IGLU zuständig ist. Die wissenschaftliche Leitung auf internationaler Ebene liegt beim *TIMSS & PIRLS International Study Center* (ISC) am Boston College (Schwippert, Scholz, et al., 2020).

Deutschland beteiligt sich auf gemeinsamen Beschluss der *Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland* (KMK) und des *Ministeriums für Bildung und Forschung* (BMBF) an der international durchgeführten Schulvergleichsuntersuchung. Diese findet regelmäßig in einem vierjährigen Rhythmus statt und legt den Schwerpunkt auf die Betrachtung von Trends in den mathematischen und naturwissenschaftlichen Kompetenzen von Viert- und Achtklässler:innen und deren individuellen und schulischen Lernumfeldern in den teilnehmenden Staaten und Regionen. Deutschland beteiligt sich seit 2007 mit der vierten Klasse an TIMSS (Schwippert, Kasper, et al., 2020b).

Mit der TIMS-Studie werden im Wesentlichen drei Ziele verfolgt:

1. Bestandsaufnahme der allgemeinen Situation in den Bildungssystemen und Betrachtung von Entwicklungen über einen längeren Zeitraum hinweg: International findet TIMSS seit 1995 statt. Deutschland nimmt mit der Grundschule seit 2007 teil, sodass mit TIMSS 2019 für Deutschland ein Trend über zwölf Jahre

abbildbar ist. Eines der Hauptanliegen ist dabei, den bestehenden Trend weiterführen zu können. Dieser ermöglicht es, langfristig die Entwicklung in den Bildungssystemen zu verfolgen und beispielsweise die Umsetzung von Reformen oder Veränderungen infolge größerer politischer oder gesellschaftlicher Ereignisse zu untersuchen. Gleichzeitig kann durch das wiederholte Querschnittsdesign betrachtet werden, inwiefern sich die Zusammensetzung der Schülerschaft in den teilnehmenden Staaten verändert und welche Merkmale die Lernumgebungen der jeweiligen Kohorten aufweisen. Eine umfangreiche Erfassung von Merkmalen der Lernumgebung ermöglicht es zudem, die Zusammenhänge zwischen diesen potenziellen Einflussfaktoren und den Kompetenzen der Schüler:innen in Mathematik und den Naturwissenschaften zu untersuchen. Durch das wiederholte Querschnittsdesign können mit TIMSS-Daten jedoch keine kausalen Aussagen getroffen werden (Schwippert, Scholz, et al., 2020).

2. Vergleiche auf internationaler Ebene zwischen den teilnehmenden Staaten und Regionen: An TIMSS 2019 nahmen insgesamt 64 Staaten und Regionen teil, wodurch Vergleiche mit einer großen Zahl an Bildungssystemen möglich sind. Durch die Bildung von Vergleichsgruppen können zudem Gruppen von an TIMSS teilnehmenden Staaten, die einen ähnlichen ökonomischen, kulturellen oder geschichtlichen Hintergrund aufweisen, herangezogen werden. Für Deutschland sind dies die Vergleichsgruppe der OECD-Staaten (in TIMSS 2019: 29 Staaten) und die Gruppe der EU-Staaten (in TIMSS 2019: 24 Staaten). Umfangreiche Informationen zu den teilnehmenden Bildungssystemen, die über die reinen Leistungswerte hinaus für die inhaltliche Einordnung der Ergebnisse herangezogen werden können, werden über zwei Wege generiert: Zum einen stellen alle Staaten Informationen zur allgemeinen Struktur des Bildungssystems, zu den Schulformen und zur Ausbildung der Lehrkräfte bereit, die begleitend in einer Enzyklopädie veröffentlicht werden, zum anderen werden den Schulleitungen, Lehrkräften, Eltern und Schüler:innen Hintergrundfragebögen vorgelegt, in denen deren individuelle Situation bzw. Einschätzung erfasst wird, um ein umfassendes Bild der Lernumgebung der teilnehmenden Schüler:innen zu erzeugen. Die Fragebögen enthalten dabei einen Kern an Fragen, der für alle Staaten gleich ist, um Vergleiche von Hintergrundinformationen zu ermöglichen (Schwippert, Scholz, et al., 2020).
3. Untersuchung staatenspezifischer, ggf. aktueller Herausforderungen: Um aktuelle Entwicklungen oder Herausforderungen im eigenen Bildungssystem

gezielt untersuchen zu können, gibt es die Möglichkeit, die international vorgegebenen Fragebögen, die in allen Staaten administriert werden, um nationale Fragen zu erweitern. In Deutschland gibt es neben den generellen nationalen Erweiterungen zu Spezifika der Lehr- und Lernumgebungen (z. B. Fragen zu den Themen Ganztagsschule, Hausaufgaben, Schulorganisation oder häuslichem Unterstützungsverhalten) zumeist auch einen inhaltlichen Schwerpunkt. In TIMSS 2019 war dies die tiefergehende Untersuchung der Unterrichtsqualität des Mathematik- und Sachunterrichts (Schwippert, Scholz, et al., 2020).

Hauptanliegen der TIMS-Studie ist es, die Bildungssysteme der teilnehmenden Staaten über einen längeren Zeitraum hinweg zu betrachten, zu vergleichen sowie Veränderungen zu dokumentieren und zu interpretieren, um deren Qualität zu sichern und voranzubringen. Dazu werden nicht nur die Kompetenzen der Schüler:innen in Mathematik und in den Naturwissenschaften erhoben, sondern auch umfangreiche Informationen zur häuslichen und schulischen Lernumgebung generiert. So kann ein umfangreiches Bild der unterschiedlichen teilnehmenden Bildungssysteme gezeichnet werden. Der Fokus der TIMS-Studie liegt auf den teilnehmenden Staaten mit ihren Bildungssystem und dient nicht der Betrachtung einzelner Schüler:innen im Sinne einer Einzelfalldiagnostik (Schwippert, Scholz, et al., 2020).

6.2 Organisation der TIMS-Studie und Schlussfolgerungen

Wie im vorangegangenen Kapitel erwähnt, folgt die TIMS-Studie einem vierjährigen Zyklus. Die Projektlaufzeit von TIMSS 2019 begann in Deutschland am 1.4.2017 und endete am 31.12.2020, wobei die Haupterhebung, an der sich die Benennung der einzelnen Zyklen orientiert, im Frühjahr 2019 stattfand. Der zeitliche Ablauf der Studie über diese vier Jahre, die Gestaltung der Testungen sowie die Stichprobenziehung werden im Folgenden dargestellt. Dabei wird zusätzlich darauf eingegangen, welche Relevanz die jeweiligen Aspekte für die vorliegende Arbeit haben.

6.2.1 Zeitlicher Ablauf der Studienzyklen

Im Rahmen von TIMSS 2019 wurden, anders als in den meisten Zyklen, die nur zwei Erhebungen umfassen, insgesamt drei Erhebungen realisiert: eine Moduseffektstudie, ein Feldtest und eine Haupterhebung. Die zusätzliche Moduseffektstudie ist in der Umstellung von papierbasierter auf computerbasierte Testung begründet. Um den Trend zwischen den verschiedenen Zyklen aufrecht zu erhalten, wurde die Funktionsweise und Schwierigkeit der Testaufgaben in den verschiedenen Testmodi überprüft. Unter Berücksichtigung der Ergebnisse der Moduseffektstudie können die Ergebnisse von TIMSS 2019 als anschlussfähig an den Trend der vorherigen Zyklen betrachtet werden (Schwippert, Scholz, et al., 2020). Diese erste Teilstudie fand im Frühjahr 2017 mit knapp 850 Schüler:innen statt. 2018 folgte dann ein Feldtest mit ca. 1.330 Schüler:innen, in dem neuentwickelte Testaufgaben und Hintergrundfragebögen sowie Testabläufe und Materialien getestet und anhand der Befunde überarbeitet sowie für die Haupterhebung finalisiert wurden. Diese fand in Deutschland ein Jahr später im Mai und Juni 2019 in allen 16 Ländern der Bundesrepublik Deutschland statt. Gleichzeitig wurde im Rahmen der Haupterhebung eine Brückenstudie durchgeführt, die die Ergebnisse aus der Modusstudie absichern bzw. Veränderungen seitdem feststellen sollte. Dafür bearbeitete ein Teil der Schüler:innen die Testaufgaben wie in vorherigen TIMSS-Zyklen papierbasiert, der andere Teil bearbeitete die Aufgaben an Laptops (Schwippert, Kasper, et al., 2020a; Schwippert, Scholz, et al., 2020).

Nach der Haupterhebung wurden die Daten aller Staaten von der internationalen Studienleitung aufbereitet und skaliert, sodass die einheitlichen, finalen Datensätze ab Sommer 2020 den teilnehmenden Staaten zur Verfügung standen. Im selben Jahr wurden außerdem die internationalen Ergebnisse veröffentlicht und parallel in Deutschland ein nationaler Bericht mit umfangreichen internationalen wie nationalen Ergebnissen verfasst (Schwippert, Scholz, et al., 2020).

6.2.2 Ablauf der Testung

Der Ablauf des Testtages sowie Manuale für die Testleitungen und begleitende Materialien sind von der internationalen Studienleitung vorgegeben, um die Vergleichbarkeit zwischen den teilnehmenden Staaten zu sichern. Zur Qualitätssicherung dieser Prozeduren werden zudem externe Beobachter:innen eingesetzt, die stichprobenartig

Testungen besuchen und die vorgabenkonforme Umsetzung überprüfen (Johansone & Flicop, 2020). Während die Fachlehrkräfte, Schulleitungen und Eltern ihre jeweiligen Fragebögen vor bzw. nach dem Testtag ausfüllten, bearbeiteten die Schüler:innen die Kompetenztests und die Fragebögen an einem Testtag, dessen Ablauf standardisiert war, in ihrer Schule.

Dieser beginnt damit, dass die Testleitung den Raum, der von der Schule für die Testung zur Verfügung gestellt wird, vorbereitet. Nach einer Begrüßung und Einweisung bearbeiten die Schüler:innen zunächst ein erstes Testheft mit Mathematik- oder Naturwissenschaftsaufgaben. Im Anschluss gibt es für die Schüler:innen die erste von mehreren Pausen, die den Testtag in kindgerechte Arbeitsphasen aufteilen, woraufhin die Schüler:innen ein zweites Testheft bearbeiten. In einer größeren Pause nach der Kompetenztestung bereiten die Testleitungen alles für die Befragung der Schüler:innen vor (Einsammeln der Testmaterialien und Austeilen der Befragungsunterlagen). Im ersten Schritt bearbeiten die Kinder dann eine Kurzversion des *Kognitiven Fähigkeitstests* (KFT) sowie der *Grundintelligenztest Skala 2 – Revision* (CFT 20-R), bevor sie mit der Bearbeitung des Fragebogens beginnen. Da dieser für Viertklässler:innen sehr umfangreich ist, ist die Bearbeitung des Fragebogens nochmal durch eine Pause unterbrochen. Nachdem die Testleitung alle Materialien abschließend wieder eingesammelt hat, ist der Testtag für die Kinder beendet. Im Anschluss an die Testung werden die erhobenen Informationen digitalisiert und zunächst der IEA zur Überprüfung und anschließend der internationalen Studienleitung für die Skalierung übergeben (Schwippert, Scholz, et al., 2020).

6.2.3 Stichprobenziehung und Besonderheiten des Datensatzes[10]

Bei der Durchführung solcher groß angelegten Studien müssen einige Besonderheiten berücksichtigt werden. Dies betrifft zum einen die Stichprobenziehung und Repräsentativität der Ergebnisse, zum anderen die Notwendigkeit, die Informationen aus den verschiedenen Quellen einander zuordnen zu können. Diese Besonderheiten werden im Folgenden erläutert.

[10] Die Datensätze zu TIMSS 2019 können über das *Forschungsdatenzentrum* (FDZ) des *Instituts zur Qualitätsentwicklung im Bildungswesen* (IQB) beantragt werden. Für Antragsinformationen und Erläuterungen zur Aufbereitung des Datensatzes, der in dieser Arbeit genutzt wurde, siehe Kapitel 7.1.1.

Um den Zielen der Studie gerecht werden zu können, müssen die Ergebnisse der Befragung bzw. Testung in den teilnehmenden Staaten generalisierbar sein. Dazu werden in jedem Zyklus repräsentative Stichproben gezogen, die spezifische Besonderheiten der Bildungssysteme (für Deutschland: Ziehung von 16 repräsentativen Stichproben zur Berücksichtigung der föderalen Struktur) widerspiegeln (Schwippert, Scholz, et al., 2020). Um eine repräsentative Stichprobe zu erzeugen, wird in TIMSS ein mehrstufiges Ziehungsverfahren eingesetzt. Unter Berücksichtigung der verschiedenen Schulgrößen werden dabei zunächst für jedes Bundesland Schulen gezogen und anschließend aus diesen wiederum per Zufall eine vierte Klasse ausgewählt. Gibt es nur eine vierte Klasse in einer Schule, ist diese automatisch Teil der Stichprobe. Da es aufgrund dieses Designs zu unterschiedlichen Ziehungswahrscheinlichkeiten zwischen den einzelnen Schüler:innen kommt, werden zusätzlich Gewichte berechnet, die diesen Umstand bei Analysen berücksichtigen und die Repräsentativität der Stichprobe sicherstellen (Schwippert, Scholz, et al., 2020). Die Stichprobe wird dabei repräsentativ für die Schüler:innen der teilnehmenden Staaten gezogen. Das bedeutet gleichzeitig, dass die Stichprobe für die Lehrkräfte, Schulleitungen und Erziehungsberechtigten nicht als repräsentativ angesehen werden kann, da diese über die gezogenen Schüler:innen und nicht unabhängig ausgewählt werden. Sollen beispielsweise Aussagen über die Lehrkräfte getroffen werden, müssen diese auf die Schüler:innen bezogen werden.

Damit die Aussagen der Lehrkräfte und die der Schüler:innen aus den verschiedenen Datenquellen einander zugeordnet werden können, wird für die Analysen neben den Gewichten eine weitere Funktionsvariable, eine sogenannte *Identification Variable*, benötigt. Am häufigsten wird in Analysen für die Zuordnung von Lehrkräfte- und Schüleraussagen die Variable IDCLASS, mit der sich jede Klasse eindeutig identifizieren und der jeweiligen Lehrkraft zuordnen lässt, verwendet. Da jedoch im naturwissenschaftlichen Unterricht in Deutschland einige Klassen von zwei Lehrkräften unterrichtet werden, würden mit der Verwendung von IDCLASS einigen Klassen die Aussagen zweier Lehrkräfte zugeordnet werden. So würde allerdings eine der Voraussetzungen der Mehrebenenanalyse (siehe Kapitel 7.3.2) verletzt werden, die besagt, dass keine Varianz auf der Klassenebene vorliegen darf. Dementsprechend muss für die geplanten Analysen eine ID-Variable genutzt werden, die eine ein-eindeutige Zuordnung von Lehrkräfte- und Schüleraussagen ermöglicht. Dies kann mit der Variable IDTEALIN realisiert werden, die jeder Lehrkraft die eigene Klasse zuordnet. Daraus folgt, dass bei Klassen mit zwei Lehrkräften die Schüleraussagen zweimal im Datensatz (einmal pro Lehrkraft) vorkommen. Damit die Aussagen dieser Schüler:innen nicht doppelt in die Analysen eingehen, gewichtet eine Gewichtungsvariable diese Klassen entsprechend

herunter (für eine ausführliche Darstellung der Gewichtungs- und Identifikationsvariablen siehe beispielsweise Fishbein, Foy & Yin, 2021; Schulz-Heidorf, 2016).

6.3 Beschreibung der Stichprobe

Um ausreichend große Stichproben zur Sicherstellung von Repräsentativität und Vergleichbarkeit zwischen den Staaten zu generieren, wurden in jedem der teilnehmenden Staaten mindestens 150 Schulen und somit ca. 4.000 Schüler:innen für die Haupterhebung ausgewählt. Zusätzlich wurde in Staaten, die 2019 die Umstellung zur digitalen Testung realisierten, eine weitere Stichprobe von ca. 1.500 Schüler:innen für die Brückenstudie gezogen, die den Leistungstest papierbasiert bearbeitete (Schwippert, Scholz, et al., 2020).

Da für die Analysen in der vorliegenden Arbeit nur die Daten der eTIMSS-Haupterhebung verwendet werden, beziehen sich die folgenden Angaben auf diese Stichprobe und nicht zusätzlich auf die Stichprobe der Brückenstudie. Zudem werden die naturwissenschaftlichen Kompetenzen untersucht, weshalb sich in der Darstellung im Folgenden auf diese Domäne konzentriert wird.

In Deutschland umfasst die Stichprobe in TIMSS 2019 203 Schulen bzw. 211 Klassen mit insgesamt 3.437 Schüler:innen. Von diesen Schüler:innen nahmen ca. 97 % an den Leistungstests und 87 % an der Befragung teil. Befragt wurden zudem 216 Mathematiklehrkräfte und 218 Sachunterrichtslehrkräfte, wobei die Beteiligungsquote bei ca. 90 % lag. Dass mehr Lehrkräfte als Klassen befragt wurden, lässt sich damit erklären, dass einige wenige Klassen von zwei Lehrkräften unterrichtet wurden (Schwippert, Scholz, et al., 2020).

Die Schüler:innen erreichten in TIMSS 2019 in den Naturwissenschaften im Mittel einen Punktwert von 518. Damit befindet sich Deutschland im Vergleich mit allen anderen Teilnehmern in TIMSS im Mittelfeld und deutlich oberhalb des internationalen Mittelwertes von TIMSS 2019, der bei 491 Punkten liegt. Allerdings ist sowohl der Mittelwert der Vergleichsgruppe der EU-Staaten (522 Punkte) als auch der der OECD-Staaten (526 Punkte) etwas höher als das Ergebnis der Schüler:innen in Deutschland (Steffensky et al., 2020). Für die Verteilung auf die Kompetenzstufen (s. auch Kapitel 6.4.1) zeigt sich folgendes Bild: Jeweils ca. 7 % der Schüler:innen liegen mit ihren Leistungen auf der niedrigsten bzw. höchsten Kompetenzstufe. Ungefähr 20 % der Schüler:innen erreichten die Kompetenzstufe II und knapp 36 % die Kompetenzstufe III. Die verbleibenden 30 % erreichten die Kompetenzstufe IV (Steffensky

et al., 2020). Somit befinden sich in der zu untersuchenden Stichprobe 30 % potenziell leistungsstarke Schüler:innen und 7 % leistungsstärkste Schüler:innen. Dies entspricht gewichtet ca. 1.000 bzw. 240 Schüler:innen.

6.4 Beschreibung ausgewählter Instrumente

Im Rahmen der TIMSS-Untersuchung werden neben Kompetenztests in Mathematik und den Naturwissenschaften sowie einer Kurzversion des KFTs und des CFTs auch Hintergrundfragebögen für die Schüler:innen, ihre Fachlehrkräfte, ihre Erziehungsberechtigten sowie die Schulleitungen erhoben (Schwippert, Scholz, et al., 2020). Mit diesen werden umfangreiche Informationen zu den Lernbedingungen der befragten Schüler:innen in der Schule, im Fachunterricht und in der Familie sowie zu weiteren, teils deutschlandspezifischen Einflussfaktoren generiert. Da für diese Arbeit die Ergebnisse des Kompetenztests in Naturwissenschaften sowie die Antworten aus den Schüler- und Lehrkräftefragebögen relevant sind, werden diese Instrumente im Folgenden ausführlicher vorgestellt.

6.4.1 Kompetenztests

Um die Leistungen der Viertklässler:innen in Mathematik und den Naturwissenschaften zu erfassen, werden speziell für die TIMS-Studie entwickelte Kompetenztests eingesetzt. Insgesamt bestand dieser Test in TIMSS 2019 aus 169 Naturwissenschafts- und 171 Mathematikaufgaben, die in Form eines rotierten Testheftdesigns von den Schüler:innen bearbeitet wurden (für ausführlichere Informationen zum rotierten Testheftdesign siehe Martin, Mullis & Foy, 2017; Schwippert, Scholz, et al., 2020).

Der Test zur Messung der naturwissenschaftlichen Kompetenzen umfasst Aufgaben aus den Inhaltsbereichen Geografie, Biologie und Physik/Chemie. Damit entspricht die Konzeptualisierung dieser Domäne dem Fach *science* bzw. Naturwissenschaften. Der in Grundschulen in Deutschland übliche Sachunterricht[11] entspricht jedoch nicht dieser Konzeptualisierung und enthält darüber hinaus auch Elemente aus dem technischen,

[11] Je nach Bundesland gibt es alternative Bezeichnungen für dieses Fach (z. B. Heimatkunde oder Sachkunde). In dieser Arbeit werden die Begriffe Sachunterricht und Naturwissenschaften verwendet.

gesellschaftlich-sozialwissenschaftlichen und historischen Bereich (Steffensky et al., 2020). Um die Vergleichbarkeit der Ergebnisse zwischen den Staaten sicherzustellen, werden die Materialien in Deutschland, insbesondere die Fragebögen, mit einem entsprechenden Zusatz versehen. Die Testaufgaben lassen sich außer in Inhaltsbereiche auch in kognitive Anforderungsbereiche einteilen, die in TIMSS 2019 folgendermaßen besetzt waren: Reproduzieren (41 %), Anwenden (38 %) sowie Problemlösen (21 %). 43 % der Aufgaben entstammten dem Bereich Biologie, 36 % dem Inhaltsbereich Physik/Chemie und 21 % dem Bereich Geografie. Diese Verteilung lässt sich damit begründen, dass geografische Inhalte deutlich weniger im naturwissenschaftlichen Unterricht in den teilnehmenden Staaten vertreten sind als biologische oder physikalisch/chemische Themen (Steffensky et al., 2020).

Der Einsatz des rotierten Testheftdesigns hat zur Folge, dass alle befragten Schüler:innen jeweils nur einen Teil der Mathematik- und Naturwissenschaftsaufgaben bearbeiten. Das bedeutet, dass je nachdem welche Testhefte die jeweiligen Schüler:innen erhalten haben, nur Leistungsinformationen für diesen Teil aller Aufgaben vorliegen. Die fehlenden Werte werden deshalb mithilfe des Verfahrens der multiplen Imputation geschätzt. In TIMSS enthalten die Datensätze somit fünf mögliche Kompetenzwerte pro Domäne, genannt *plausible values* (Foy & LaRoche, 2020; Schwippert, Scholz, et al., 2020, S. 44).

Die Ergebnisse der Schüler:innen in den Kompetenztests wurden in TIMSS 1995 auf einer Punkteskala mit einem festgelegten Mittelwert von 500 Punkten und einer Standardabweichung von 100 Punkten verortet (Gonzalez, 1998). Über sogenannte Benchmarks können die Punktwerte der Schüler:innen zusätzlich auf Kompetenzstufen eingeordnet werden. Insgesamt gibt es fünf Kompetenzstufen, die durch vier Benchmarks abgegrenzt werden (Wendt et al., 2016). Schüler:innen mit einem maximalen Punktwert von 399 können auf der Kompetenzstufe I verortet werden, Schüler:innen mit Leistungen zwischen 400 und 474 Punkten auf Kompetenzstufe II und Schüler:innen, die einen Punktwert von mindestens 475 und maximal 549 erreichen, werden der dritten Kompetenzstufe zugeordnet. Ab einem Punktwert von 550 erreichen Schüler:innen die Kompetenzstufe IV und ab 625 Punkten gehören Schüler:innen der höchsten Kompetenzstufe, Kompetenzstufe V, an (Steffensky et al., 2020).

Die unterschiedlichen Kompetenzstufen lassen sich dabei folgendermaßen kategorisieren: Schüler:innen auf Kompetenzstufe I verfügen nur über rudimentäres Anfangswissen in den Naturwissenschaften und haben selbst bei einfachen Aufgaben Schwierigkeiten, diese zu lösen. Schüler:innen auf Kompetenzstufe II verfügen

bereits über ein Basiswissen in den Inhaltsbereichen Biologie und Physik/Chemie, auf Kompetenzstufe III wird darüber hinaus auch ein Basiswissen in Geografie erkennbar und die Schüler:innen können ihr Wissen bei einfachen Aufgaben auch anwenden. Auf Kompetenzstufe IV zeigen Schüler:innen zudem die Fähigkeit, ihr vorhandenes Wissen auch in anspruchsvolleren Kontexten anzuwenden und Schlussfolgerungen zu ziehen. Sie verfügen über grundlegende Kenntnisse der naturwissenschaftlichen Denk- und Arbeitsweisen. Schüler:innen auf Kompetenzstufe V verfügen ebenfalls über elementare Fähigkeiten in den naturwissenschaftlichen Denk- und Arbeitsweisen und Grundwissen in allen drei Inhaltsbereichen, dies jedoch auf einem fortgeschrittenen Niveau. So sind sie beispielsweise in der Lage, einfache Experimente zu planen (Steffensky et al., 2020).

Diese Einteilung auf die Kompetenzstufen ist grundlegend für die vorliegende Arbeit. Wie eingangs beschrieben, erreichen 6.9 % aller Schüler:innen in Deutschland in den Naturwissenschaften die Kompetenzstufe V. Diese werden in den folgenden Kapiteln als leistungsstärkste Schüler:innen bezeichnet. Unter potenziell leistungsstarken Schüler:innen werden die Schülerinnen verstanden, die Leistungen auf Kompetenzstufe IV zeigen.

6.4.2 Schülerfragebogen

Um lern- und leistungsbezogene Hintergrundmerkmale der Schüler:innen zu erfassen, wird in TIMSS ein aus internationalen und nationalen Fragen bestehender Fragebogen eingesetzt. In TIMSS 2019 umfasste der Fragebogen 45 Fragen, von denen knapp die Hälfte international vorgegeben war. Der internationale Teil des Fragebogens umfasst altersgemäße Fragen zu allgemeinen Informationen über die eigene Person (z. B. Geschlecht oder Alter), zum sozioökonomischen und Migrationshintergrund, zur Anwesenheit in der Schule, zu den Erfahrungen, die die Schüler:innen in der Schule machen sowie Fragen zu fachbezogenen Selbsteinschätzungen und zur Einschätzung von lehrkrafts- und unterrichtsbezogenen Aspekten des Mathematik- und Sachunterrichts. Im nationalen Teil wird der Block zu den Einschätzungen um weitere Fragen ergänzt. Insgesamt umfasst dieser Bereich Fragen zu selbstbezogenen Kognitionen sowie zur Unterrichtsqualität und -gestaltung. Daran anschließend werden Fragen zum elterlichen (Unterstützungs)Verhalten und zu digitalen Medien innerhalb und außerhalb der Schule gestellt. Darüber hinaus enthält der Fragebogen auch Fragen, bei

denen die Schüler:innen angeben sollen, wie sie andere Schülergruppen hinsichtlich verschiedener Aspekte einschätzen. Der letzte Block des Fragebogens umfasst Fragen zur Teilnahme an außerunterrichtlichen Angeboten der Schule.

Im theoretischen Teil dieser Arbeit wurden bedeutsame Einflussfaktoren für naturwissenschaftliche Leistungen, mit besonderem Fokus auf Prädiktoren besonderer Leistungsstärke, identifiziert. Dabei konnte gezeigt werden, dass auf individueller Schülerebene insbesondere die Merkmale fachbezogene Einstellung und fachbezogenes Selbstkonzept für besonders hohe Leistungen von Schüler:innen relevant sind. Bedeutsame Hintergrundmerkmale sind darüber hinaus die kognitiven Fähigkeiten, das Geschlecht, der sozioökonomische Hintergrund und der Migrationshintergrund. Auf Ebene des Unterrichts stellten sich vor allem Merkmale der Unterrichtsqualität und der inneren Differenzierung als wichtig heraus. Als bedeutsame Merkmale der Lehrkräfte können die Berufszufriedenheit und Lehrkräfteselbstwirksamkeit angenommen werden.

In den in TIMSS 2019 eingesetzten Instrumenten können für die individuellen Schülermerkmale die Skalen ‚Fachbezogenes Selbstkonzept Sachunterricht I und intrinsische Werte Sachunterricht' und ‚Fachbezogenes Selbstkonzept Sachunterricht II' zur Erfassung der fachbezogenen Einstellung und des fachbezogenen Selbstkonzeptes genutzt werden. Die kognitiven Fähigkeiten werden in TIMSS 2019 über Subskalen des *Kognitiven Fähigkeitstests* (KFT) und der *Grundintelligenztest Skala 2 – Revision* (CFT 20-R) erhoben. Dabei erfassen die Subskalen des KFTs eher Facetten der verbalen Fähigkeiten, die des CFT 20-R eher Facetten der mathematischen Fähigkeiten. Aufgrund der größeren konzeptionellen Nähe zum naturwissenschaftlichen Sachunterricht, der in dieser Arbeit im Fokus steht, wird deshalb der CFT 20-R zur Analyse der kognitiven Fähigkeiten genutzt. Für den sozioökonomischen Hintergrund gibt es ebenfalls mehrere mögliche Skalen im Schülerfragebogen (z. B. zum Besitz von Wohlstandsgütern oder zum kulturellen Kapital). Skalen, die umfangreiche Informationen zur Einschätzung des sozioökonomischen Status liefern (z. B. Informationen über Schul- und Berufsabschlüsse der Eltern, finanzielle Mittel oder die Verfügbarkeit von Ressourcen) finden sich wiederum im Elternfragebogen. Da der sozioökonomische Hintergrund jedoch nicht im Fokus der vorliegenden Untersuchung steht, sondern lediglich als Kontrollvariable hinzugezogen wird, soll ein möglichst einfacher Prädiktor gewählt werden. Mit der Frage nach der Anzahl der Bücher zu Hause bietet der Schülerfragebogen einen einfacheren, aber adäquaten Indikator für den sozioökonomischen Hintergrund (Eriksson, Lindvall, Helenius & Ryve, 2021; Schwippert, 2019; Stubbe, Krieg, Beese & Jusufi, 2020). Darüber hinaus finden sich mit ‚Geschlecht'

und ‚Migrationshintergrund' im Schülerfragebogen Informationen, um die beiden weiteren Kontrollvariablen abbilden zu können. Die Information ‚Migrationshintergrund' stellt dabei eine rekodierte Version der Skalen ‚Deutschland als Herkunftsland der Mutter' und ‚Deutschland als Herkunftsland des Vaters' dar (siehe Kapitel 6.4.4).

Auch die drei Basisdimensionen der Unterrichtsqualität lassen sich in TIMSS 2019 über den Schülerfragebogen erfassen. Dieser beinhaltet die Skalen ‚Wahrgenommene Klassenführung (Sachunterricht)' und ‚Wahrgenommene Unterstützung und kognitive Aktivierung durch die Lehrperson (Sachunterricht)'. Skalen zu Erfassung von Formen der inneren Differenzierung finden sich sowohl im Schülerfragebogen als auch im Lehrkräftefragebogen. Um die Merkmale des Unterrichts im Rahmen der Analysen vergleichen zu können, wird mit der Skala ‚Wahrnehmung individueller Förderung im Sachunterricht' wie für die Unterrichtsqualitätsmerkmale eine Skala aus dem Schülerfragebogen genutzt.

Die Lehrkräftemerkmale lassen sich mit den Informationen aus dem Schülerfragebogen nicht abdecken. Deshalb wird im Folgenden zunächst der Lehrkräftefragebogen vorgestellt, bevor eine detaillierte Darstellung der einzelnen ausgewählten Skalen in Kapitel 6.4.4 erfolgt.

6.4.3 Lehrkräftefragebogen

Um Merkmale der Lehrkraft und des Unterrichts zu erfassen, wird in TIMSS ein Lehrkräftefragebogen eingesetzt, der sowohl generelle Fragen enthält, die auf alle Lehrkräfte zutreffen, als auch fachspezifische Fragen, die sich konkret auf den Mathematik- oder naturwissenschaftlichen Sachunterricht beziehen und somit je nach Fachkonstellation nur für einen Teil der Lehrkräfte zutreffen.

Der internationale Teil der Fragebögen in TIMSS 2019 beginnt mit einigen allgemeinen Fragen zur ausfüllenden Lehrkraft. Anschließend folgen Fragen zu den erworbenen Abschlüssen und zur Ausbildung sowie zur Einschätzung von schul- und unterrichtsbezogenen Aspekten. Zudem werden Angaben zur getesteten TIMSS-Klasse sowie zum Unterricht und zum Einsatz digitaler Medien in dieser Klasse erfragt. Darauf folgen Fragen zu den unterrichteten Inhalten sowie zum Umgang mit Hausaufgaben und Bewertungen. Abgeschlossen werden die internationalen fachspezifischen Fragebogenteile, zunächst für Mathematik, anschließend für den Sachunterricht, mit Fragen zu Fortbildungen.

Der nationale Teil des Lehrkräftefragebogens beginnt mit Fragen zu Erfahrungen und Einstellungen im und zum Beruf sowie einem Block mit Fragen zum Umgang mit digitalen Medien. Anschließend folgen Fragen zu schulbezogenen Tätigkeiten außerhalb des Unterrichts und zur Einschätzung verschiedener schulbezogener Aspekte. Darüber hinaus werden mathematikspezifische Fragen gestellt, die die Kenntnis bzw. den Einsatz von Materialien aus aktuellen Projekten abdecken. Darauf folgen zunächst für Mathematik und anschließend für den Sachunterricht Fragen zur Unterrichtsgestaltung mit und ohne digitale Medien.

Die Ergebnisse der theoretischen Darlegungen haben gezeigt, dass auf Lehrkräfteebene die Berufszufriedenheit und die Lehrkräfteselbstwirksamkeit besonders relevant sind. Diese Merkmale lassen sich über die beiden Skalen im Lehrkräftefragebogen ‚Zufriedenheit mit dem Lehrberuf' und ‚Sicherer Umgang beim Unterrichten im Sachunterricht' abbilden.

Auf Unterrichtsebene wurden wie im vorherigen Kapitel beschrieben Unterrichtsqualitätsmerkmale sowie Formen der inneren Differenzierung als potenzielle Prädiktoren besonderer Leistungsstärke identifiziert. Über Skalen im Schülerfragebogen lassen sich diese vollständig abbilden. Im Lehrkräftefragebogen finden sich hingegen nur zum Teil Skalen, die diese Merkmale widerspiegeln. In den Analysen werden deshalb die Skalen aus dem Schülerfragebogen eingesetzt.

6.4.4 Darstellung der relevanten Skalen

Wie im vorherigen Kapitel beschrieben, werden in der folgenden Untersuchung die Skalen ‚Fachbezogenes Selbstkonzept Sachunterricht I und intrinsische Werte Sachunterricht', ‚Fachbezogenes Selbstkonzept Sachunterricht II', ‚Wahrnehmung individueller Förderung im Sachunterricht', ‚Wahrgenommene Klassenführung (Sachunterricht)', ‚Wahrgenommene Unterstützung und kognitive Aktivierung durch die Lehrperson (Sachunterricht)', ‚Zufriedenheit mit dem Lehrberuf' und ‚Sicherer Umgang beim Unterrichten im Sachunterricht' zur Untersuchung der Hypothesen eingesetzt. Zusätzlich werden die Informationen ‚*Grundintelligenztest Skala 2 – Revision* (CFT 20-R)', ‚Geschlecht', ‚Anzahl der Bücher zu Hause' und ‚Migrationshintergrund' als Kontrollvariablen hinzugezogen. In diesem Kapitel erfolgt eine detailliertere Darstellung dieser Skalen. Deskriptive Statistiken zu den acht Skalen finden sich im Anhang.

Fachbezogenes Selbstkonzept Sachunterricht I und intrinsische Werte Sachunterricht

Die Skala ‚Fachbezogenes Selbstkonzept Sachunterricht I und intrinsische Werte Sachunterricht' wird in Deutschland seit der ersten Teilnahme 2007 eingesetzt. Die Skala umfasste zunächst fünf Items, seit TIMSS 2015 wird sie mit neun Items erhoben. Sie ist Teil des internationalen Fragebogens und heißt im Englischen ‚*Students Like Learning Science*'. In Anlehnung an den englischen Skalentitel wird in der nationalen Berichterstattung auch der Kurztitel der Skala ‚Fachbezogene Einstellung (Sachunterricht)' verwendet. Diese Kurzform wird auch in den folgenden Ausführungen genutzt.[12]

Die Items der Skala ‚Fachbezogene Einstellung' umfassen Facetten der affektiven Komponente von Einstellung, dabei auch explizit Aspekte des Interesses bzw. intrinsischer Werte, sowie Facetten der kognitiven Komponente von Einstellung. Nicht vertreten ist hingegen die Verhaltenskomponente, sodass die in TIMSS eingesetzte Skala eher ein zweikomponentiges Verständnis von Einstellung widerspiegelt (siehe Kapitel 4.1.2).

Skala: Fachbezogenes Selbstkonzept Sachunterricht I und intrinsische Werte Sachunterricht

Wie sehr stimmst du diesen Aussagen über das Lernen im naturwissenschaftlichen Sachunterricht zu?
(1 = stimme völlig zu; 2 = stimme eher zu; 3 = stimme eher nicht zu; 4 = stimme überhaupt nicht zu)

ASBS07A	Ich lerne gerne im Sachunterricht.
ASBS07B	Ich wünschte, ich hätte keinen Sachunterricht.
ASBS07C	Sachunterricht ist langweilig.
ASBS07D	Ich lerne viele interessante Dinge im Sachunterricht.
ASBS07E	Ich mag Sachunterricht.
ASBS07F	Ich freue mich auf den Sachunterricht in der Schule.
ASBS07G	Im Sachunterricht lerne ich, wie die Dinge in der Welt funktionieren.
ASBS07H	Ich mache gern Versuche im Sachunterricht.
ASBS07I	Sachunterricht gehört zu meinen Lieblingsfächern.

12 Da in dieser Arbeit mit Ausnahme der ‚Zufriedenheit mit dem Lehrberuf' ausschließlich Skalen für das Fach Sachunterricht betrachtet werden, wird in den folgenden Kapiteln auf den Klammerausdruck ‚(Sachunterricht)' bei allen fachspezifisch erhobenen Skalen verzichtet.

Fachbezogenes Selbstkonzept Sachunterricht II

Mit der Skala ‚Fachbezogenes Selbstkonzept Sachunterricht II' lässt sich das fachbezogene Selbstkonzept der Schüler:innen erfassen. Genau wie die Skala ‚Fachbezogene Einstellung' ist sie Teil der internationalen Vorgaben und wird seit 2007 in Deutschland erhoben. Dabei wurde die Skala in TIMSS 2011 um die Items ASBS09E-ASBS09G ergänzt und wird seitdem unverändert eingesetzt.

Wie in den theoretischen Grundlagen dargestellt, entsteht das Selbstkonzept vor allem über Vergleiche. Dies spiegelt sich auch in den Items der TIMSS-Skala wider, die sowohl internale Vergleichsprozesse – beispielsweise über den Einbezug anderer Fächer als Referenzkategorie –, als auch externale Vergleichsprozesse – über den Vergleich mit Mitschüler:innen – abbilden. Anders als bei der Skala ‚Fachbezogene Einstellung', die im Kern fachspezifische Facetten des Sachunterrichtes beinhaltet, entsteht die Fachspezifität der Skala zum fachbezogenen Selbstkonzept, indem neutrale Formulierungen, die für alle Fächer genutzt werden könnten, explizit auf den Sachunterricht bezogen werden.

Skala: Fachbezogenes Selbstkonzept Sachunterricht II

Wie sehr stimmst du diesen Aussagen über den naturwissenschaftlichen Sachunterricht zu? (1 = stimme völlig zu; 2 = stimme eher zu; 3 = stimme eher nicht zu; 4 = stimme überhaupt nicht zu)	
ASBS09A	Normalerweise bin ich gut im Sachunterricht.
ASBS09B	Sachunterricht fällt mir schwerer als vielen meiner Mitschüler.
ASBS09C	Ich bin einfach nicht gut im Sachunterricht.
ASBS09D	Ich lerne schnell im Sachunterricht.
ASBS09E	Meine Lehrerin/mein Lehrer sagt, dass ich gut im Sachunterricht bin.
ASBS09F	Ich finde Sachunterricht schwerer als alle anderen Fächer.
ASBS09G	Sachunterricht bringt mich durcheinander.

Wahrnehmung individueller Förderung im Sachunterricht

Die Skala ‚Wahrnehmung individueller Förderung im Sachunterricht' ist für TIMSS 2019 neu entwickelt und somit das erste Mal erhoben worden. Sie ist Teil der nationalen Ergänzung im Fragebogen und basiert auf der Skala ‚Förderung nach individuellen

Lernvoraussetzungen‘ aus PISA 2006 (Frey et al., 2009). Die Skala wird im Folgenden verkürzt als ‚Individuelle Förderung‘ bezeichnet.[13]

Die Items der Skala beinhalten entsprechend dem in den theoretischen Grundlagen erarbeiteten Verständnis von individueller Förderung vor allem einen Bezug zu Förderformen, die auf individuelle Schüler:innen abzielen, einige wenige Items beziehen sich aber auch auf Gruppierungsformen. Dabei ist das Differenzierungskriterium in den meisten Items die Leistung der Schüler:innen, in einigen werden aber auch interessenbezogene Differenzierungsmaßnahmen abgebildet.

Skala: Wahrnehmung individueller Förderung im Sachunterricht

Wie oft passieren diese Dinge in eurem naturwissenschaftlichen Sachunterricht? (1 = immer oder fast immer; 2 = häufig; 3 = manchmal; 4 = selten; 5 = nie)	
ASXS22A	Im Unterricht gibt meine Lehrerin/mein Lehrer Schülern Zusatzaufgaben, wenn sie etwas noch nicht verstanden haben.
ASXS22B	Meine Lehrerin/mein Lehrer gibt den guten Schülern zusätzlich schwierige Aufgaben, bei denen sie sich wirklich anstrengen müssen.
ASXS22C	Bei der Stillarbeit gibt meine Lehrerin/ mein Lehrer den Schülern unterschiedlich schwierige Aufgabenarten, die zu jedem Schüler passen.
ASXS22D	Bei Gruppenarbeiten verteilt meine Lehrerin/mein Lehrer unterschiedlich schwierige Aufgaben an die verschiedenen Gruppen.
ASXS22E	Wenn wir alleine oder in Gruppen arbeiten, bekommen wir
	unterschiedlich schwierige Aufgaben.
ASXS22F	Im Unterricht erhalten die Schüler unterschiedliche Aufgaben, die gut zu ihren Interessen passen.
ASXS22G	Im Unterricht erhalten die Schüler unterschiedliche Aufgaben, je nachdem, wofür sich die Schüler interessieren.

Wahrgenommene Klassenführung (Sachunterricht)

Die Skala ‚Wahrgenommene Klassenführung (Sachunterricht)‘, im Weiteren als ‚Klassenführung‘ bezeichnet, wird in Deutschland seit 2015 als Teil der nationalen Erwei-

[13] Da alle ausgewählten Skalen zu Unterrichtsmerkmalen die individuelle Wahrnehmung der Schüler:innen erfassen und dieser Zusatz somit nicht zur Differenzierung verschiedener Perspektiven benötigt wird, wird darauf einheitlich bei allen vier Skalen verzichtet.

terung eingesetzt. Sie basiert auf der Skala ‚Klassenführung' und dabei insbesondere auf der Subskala ‚Unterrichtsstörungen und Zeitverschwendung', die 2003/2004 in der PISA-2003-Ergänzungsstudie COACTIV eingesetzt wurde (Baumert et al., 2009). Entsprechend liegt der Fokus der Skala auf störendem Verhalten der Schüler:innen sowie auf daraus resultierenden Wartezeiten während des Unterrichts. Damit werden jedoch weitere Aspekte der Klassenführung, wie beispielsweise eine gute Vorbereitung und klare Strukturierung des Unterrichts oder das Implementieren von Regeln und Routinen, nicht direkt in der Skala adressiert.

Skala: Wahrgenommene Klassenführung (Sachunterricht)

Wie oft passieren diese Dinge in eurem naturwissenschaftlichen Sachunterricht? (1 = in jeder Stunde; 2 = in den meisten Stunden; 3 = in manchen Stunden; 4 = nie oder fast nie)	
ASXS27A	Die Schüler hören der Lehrerin/dem Lehrer nicht zu.
ASXS27B	Es ist laut und unruhig.
ASXS27C	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer muss lange warten, bis alle Schüler leise sind.
ASXS27D	Die Schüler können nicht gut arbeiten.
ASXS27E	Die Schüler fangen erst an zu arbeiten, wenn der Unterricht schon lange begonnen hat.

Wahrgenommene Unterstützung und kognitive Aktivierung durch die Lehrperson (Sachunterricht)

Die Skala ‚Wahrgenommene Unterstützung und kognitive Aktivierung durch die Lehrperson (Sachunterricht)' wird seit 2015 in TIMSS erhoben. Sie ist Teil des nationalen Fragebogens und wurde in Anlehnung an Fauth, Decristan, Rieser, Klieme und Büttner (2014a) entwickelt. Während die Items ASXS28A bis ASXS28E das Konstrukt ‚konstruktive Unterstützung' abbilden, erfassen die Items ASXS28F bis ASXS28L die ‚kognitive Aktivierung'. Für die folgenden Analysen werden entsprechend dieser Konzeption die beiden Konstrukte wie zwei getrennte Skalen behandelt und benannt. Die Items der Skala ‚konstruktive Unterstützung' bilden vor allem die Subfacette emotionale Unterstützung und dabei insbesondere eine positive Fehlerkultur und Beziehung ab. Merkmale der inhaltlichen Strukturierung sind hingegen nur angedeutet. Ähnlich liegt auch bei der Skala ‚kognitive Aktivierung' der Fokus nicht auf den drei Subfacetten (1. Stellen von herausfordernden Aufgaben, 2. Anknüpfen an Vorwissen und 3. Führen fachbezogener Gespräche und Argumentationen) gleichermaßen. Zwar zeigen

sich alle drei Subfacetten in der Skala, es finden sich dabei jedoch vor allem Items, die den ersten Aspekt abbilden. Nur einige wenige Items lassen sich darüber hinaus den beiden anderen Subfacetten zuordnen.

Skala: Wahrgenommene Unterstützung und kognitive Aktivierung durch die Lehrperson (Sachunterricht)	
Wie sehr stimmst du den folgenden Aussagen zum naturwissenschaftlichen Sachunterricht zu? (1 = stimme völlig zu; 2 = stimme eher zu; 3 = stimme eher nicht zu; 4 = stimme überhaupt nicht zu)	
ASXS28A	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer ist auch dann nett zu mir, wenn ich einen Fehler mache.
ASXS28B	Ich bin unserer Sachunterrichtslehrerin/unserem Sachunterrichtslehrer wichtig.
ASXS28C	Wenn ich einen Fehler mache, sagt unsere Sachunterrichtslehrerin/ unser Sachunterrichtslehrer mir, wie ich es besser machen kann.
ASXS28D	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer mag mich.
ASXS28E	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer glaubt, dass ich schwierige Aufgaben lösen kann.
ASXS28F	Wir bearbeiten im Sachunterricht Aufgaben, über die ich sehr gründlich nachdenken muss.
ASXS28G	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer fragt mich, was ich verstanden habe und was nicht.
ASXS28H	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer stellt Fragen, über die ich sehr gründlich nachdenken muss.
ASXS28I	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer stellt uns Aufgaben, die auf den ersten Blick schwierig wirken.
ASXS28J	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer fragt uns bei einem neuen Thema, was wir schon wissen.
ASXS28K	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer stellt uns Aufgaben, über die ich gerne nachdenke.
ASXS28L	Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer möchte, dass ich meine Antworten erkläre.

Zufriedenheit mit dem Lehrberuf

Die Skala ‚Zufriedenheit mit dem Lehrberuf' ist Teil des internationalen Fragbogens in TIMSS 2019. Sie wird seit TIMSS 2011 in dieser Form erhoben und wurde in Anlehnung an IGLU 2006 entwickelt (Bos et al., 2010). Im Englischen wird diese

Skala mit ‚Job Satisfaction' überschrieben und würde entsprechend der in den theoretischen Darlegungen vorgestellten Differenzierung mit Arbeitszufriedenheit übersetzt werden (siehe Kapitel 4.3.1). Die Items zeigen jedoch eine klare Fokussierung auf das allgemeine Berufsbild und nicht auf die aktuelle Arbeitsstelle, sodass im Englischen die Bezeichnung *vocational satisfaction* (auf Deutsch: Berufszufriedenheit) passender wäre. Für eine eindeutigere Benennung und klare Verortung entsprechend des aktuellen Forschungsstandes wird für die Skala in dieser Arbeit der Titel ‚Berufszufriedenheit' verwendet.

Skala: Zufriedenheit mit dem Lehrberuf

Wie häufig empfinden Sie folgendermaßen in Bezug auf Ihre Tätigkeit als Lehrerin/Lehrer? (1 = sehr häufig; 2 = häufig; 3 = manchmal; 4 = nie)	
ATBG08A	Ich bin mit meinem Beruf als Lehrerin/Lehrer zufrieden.
ATBG08B	Ich empfinde meine Arbeit als sehr sinnvoll und wichtig.
ATBG08C	Mein Beruf begeistert mich.
ATBG08D	Meine Arbeit inspiriert mich.
ATBG08E	Ich bin stolz auf meine Arbeit.

Sicherer Umgang beim Unterrichten im Sachunterricht

Die Skala ‚Sicherer Umgang beim Unterrichten im Sachunterricht' wird in Deutschland seit TIMSS 2011 eingesetzt. Während die Skala in den Zyklen 2011 und 2015 noch international vorgegeben war, ist sie in TIMSS 2019 Teil des nationalen Fragebogens. Im Englischen trägt die Skala den Titel ‚Confidence in Teaching Science', für TIMSS 2015 wurde sie zudem mit der *Ohio State Teacher Efficacy Scale* überarbeitet (Hooper, 2016). Das Konstrukt Selbstwirksamkeit zeichnet sich (vor allem im Vergleich zum ähnlichen Konstrukt Selbstkonzept) hinsichtlich der Selbsteinschätzung durch einen starken Eigenbezug und Fokus auf konkrete Anforderungssituationen aus. Dies zeigt sich auch in den Items der Skala ‚Sicherer Umgang beim Unterrichten im Sachunterricht'. Sie deckt somit das Konstrukt Selbstwirksamkeit von Lehrkräften ab und wird im Folgenden deshalb als ‚Lehrkräfteselbstwirksamkeit' bezeichnet. Dabei beziehen sich die meisten Items auf die Leistungen und motivationalen Merkmale der Schüler:innen, einige wenige fokussieren auch auf die Selbstwirksamkeit im Umgang

mit bzw. beim Einsatz bestimmter aktueller und konkreter Unterrichtsmethoden (z. B. forschend-entdeckendes Lernen).

Skala: Sicherer Umgang beim Unterrichten im Sachunterricht	
Wie sicher fühlen Sie sich im naturwissenschaftlichen Sachunterricht in dieser Klasse bei den folgenden Tätigkeiten? (1 = sehr sicher; 2 = sicher; 3 = einigermaßen sicher; 4 = unsicher)	
ATXS17A	Schüler für das Lernen in Naturwissenschaften zu begeistern
ATXS17B	naturwissenschaftliche Konzepte oder Prinzipien durch naturwissenschaftliche Experimente erklären
ATXS17C	den leistungsfähigsten Schülern herausfordernde Aufgaben stellen
ATXS17D	meinen Unterricht so gestalten, dass das Interesse der Schüler geweckt wird
ATXS17E	den Schülern vermitteln, wie wichtig es ist, Naturwissenschaften zu lernen
ATXS17F	das naturwissenschaftliche Verständnis der Schüler bewerten
ATXS17G	Schüler, die Schwierigkeiten haben, darin unterstützen, den Lernstoff zu verstehen
ATXS17H	die Bedeutsamkeit und Relevanz von Naturwissenschaften verdeutlichen
ATXS17I	Förderung der Denkfähigkeiten höherer Ordnung der Schüler
ATXS17J	Naturwissenschaft mithilfe der Methode des forschend-entdeckenden Lernens unterrichten

Kontrollvariablen

Die Kontrollvariablen werden anders als die im Fokus stehenden Konstrukte nicht über Skalen in den Hintergrundfragebögen, sondern über einfache Fragen mit mehreren Antwortoptionen bzw. einen separaten Subtest erhoben. Für die Kontrollvariablen werden in den weiteren Ausführungen die folgenden (Kurz)Bezeichnungen verwendet: ‚CFT', ‚Geschlecht', ‚Bücher' und ‚Migrationshintergrund' bzw. in Abbildungen ‚MGH'.

Bevor in Kapitel 8 die Ergebnisse der Analysen präsentiert und in diesem Zuge auch die aufgeführten Skalen einer empirischen Überprüfung unterzogen werden, werden zunächst in Kapitel 7 das methodische Vorgehen sowie die Grundlagen für die Analysen beschrieben.

Test: Grundintelligenztest Skala 2 – Revision (CFT 20-R)
Subskalen ‚Reihenfortsetzen' und ‚Matrizen'
CFT 20-R-T-Werte

Frage: Geschlecht (ASBG01)
Bist du ein Mädchen oder ein Junge?
1 = Mädchen
2 = Junge

Frage: Anzahl der Bücher zu Hause (ASBG04)
Wie viele Bücher gibt es bei dir zu Hause ungefähr? (Zähle nicht mit: Zeitschriften, Zeitungen und deine Schulbücher.)
1 = keine oder nur sehr wenige (0 – 10 Bücher)
2 = genug, um ein Regalbrett zu füllen (11 – 25 Bücher)
3 = genug, um ein Regal zu füllen (26 – 100 Bücher)
4 = genug, um zwei Regale zu füllen (101 – 200 Bücher)
5 = genug, um drei oder mehr Regale zu füllen (über 200 Bücher)

Frage: Migrationshintergrund (migrat)
Rekodierung der Fragen:
Ist deine Mutter (oder Stiefmutter oder weibliche Erziehungsberechtigte) in Deutschland geboren? (ASBG06A)
Ist dein Vater (oder Stiefvater oder männlicher Erziehungsberechtigter) in Deutschland geboren? (ASBG06B)
(1 = ja; 2 = nein; 3 = weiß ich nicht; 4 = nichts trifft zu)
0 = beide nicht im Ausland geboren
1 = ein Elternteil im Ausland geboren
2 = beide Elternteile im Ausland geboren

7

Methodisches Vorgehen

Aus den Hypothesen und den Spezifika der Datengrundlage ergibt sich die Notwendigkeit, zur Beantwortung der Forschungsfrage ein Mehrebenenstrukturgleichungsmodell zu rechnen. Diese Methode kombiniert die Vorteile der Strukturgleichungsmodellierung mit denen der Mehrebenenanalyse. Die Analysen zur Beantwortung der Forschungsfrage umfassen dabei insgesamt vier übergeordnete Schritte: Im ersten Schritt werden verschiedene Voranalysen zur Einschätzung der Datengrundlage durchgeführt und im zweiten Schritt die zu analysierenden latenten Konstrukte geprüft. Unter Berücksichtigung der Erkenntnisse aus den Voranalysen werden dann verschiedene Vormodelle berechnet, mit denen die direkten Effekte der einzelnen Bereiche untersucht werden. Auf der Basis der Ergebnisse dieser Vormodelle wird anschließend das Gesamtmodell zur Beantwortung der Forschungsfrage aufgestellt und im letzten Schritt schrittweise berechnet. In diesem Kapitel werden zunächst die Datenaufbereitung und die Voranalysen sowie anschließend die Grundlagen der konfirmatorischen Faktorenanalyse und der Mehrebenenstrukturgleichungsmodellierung dargestellt. Die Analysen in dieser Arbeit werden mit SPSS 27 und Mplus 8 durchgeführt.[14]

7.1 Datenaufbereitung und Voranalysen

Für die Analysen bildet ein Datensatz, der sowohl die Informationen der Lehrkräfte als auch die der Schüler:innen enthält, die Grundlage.[15] Dieser wurde für die zuvor beschriebenen Analysen noch einmal spezifisch aufbereitet. Die dabei vorgenommenen Schritte werden im Folgenden erläutert.

---

14 Aufgrund unterschiedlicher Methoden der beiden Programme mit fehlenden Werten umzugehen, weichen die Datengrundlagen bei den jeweils durchgeführten Analysen voneinander ab. In SPSS werden jedoch nur Voranalysen zur Überprüfung der Daten durchgeführt. Die Modelle zur Beantwortung der Hypothesen werden ausschließlich in Mplus durchgeführt, sodass deren Ergebnisse vergleichbar sind.

15 Die Daten wurden vom FDZ am IQB zur Verfügung gestellt (https://doi.org/10.17616/R3M30D; Datennutzungsvertrag mit der Kennnummer 2310-24a; ausführliche Quellenangabe zum Datensatz: Kasper, D., Beese, C., Scholz, L. A. & Schwippert, K. (2023). *Trends in International Mathematics and Science Study 2019 (TIMSS 2019)* (Version 2_1) [Datensatz]. Berlin: IQB – Institut zur Qualitätsentwicklung im Bildungswesen. https://doi.org/10.5159/IQB_TIMSS_2019_v2).

7.1.1 Vorbereitung des Datensatzes

Da die Items der einzelnen in Kapitel 6.4.4 vorgestellten Skalen unterschiedlich gepolt formuliert sind, wurde im ersten Schritt eine Invertierung ausgewählter Items durchgeführt, sodass alle Items eine hohe Ausprägung des jeweiligen Konstruktes widerspiegeln. Nach der Invertierung wurde für alle Items außer für die Leistungsschätzer über eine Verschiebung des gültigen Wertebereiches von 1 bis 4 zu 0 bis 3 ein natürlicher Nullpunkt zur besseren Interpretation der späteren Ergebnisse erzeugt (Pedhazur, 1997). Zudem wurden die über den Fragebogen für Schüler:innen erhobenen Informationen zum naturwissenschaftlichen Sachunterricht aggregiert, um sie als Variablen auf Klassenebene einbeziehen zu können. Darüber hinaus erfolgte eine Umkodierung der fünf *plausible values* zur Kompetenzstufenzugehörigkeit in kategoriale Variablen mit drei Ausprägungen. Dabei wurden die Kompetenzstufen unter KS IV zusammengefasst und die Kompetenzstufen IV und V unverändert beibehalten. Durch diese Form der Kodierung ergibt sich die Möglichkeit, multinomial logistische Regressionen zu berechnen, die die Übergangschancen von Kompetenzstufe IV auf eine höhere Kompetenzstufe bzw. niedrigere Kompetenzstufen widerspiegeln[16]. Da in Mplus automatisch die am höchsten kodierte Option als Referenzkategorie herangezogen wird, erfolgte die Umkodierung (zunächst scheinbar kontraintuitiv) mit KS I–III = 0; KS V = 1 und KS IV = 2 (Muthén & Muthén, 2017).

Bevor die eigentlichen Analysen zur Beantwortung der Forschungsfrage durchgeführt werden können, werden verschiedene Grundvoraussetzungen der Daten geprüft. Dazu gehört neben der Klassifizierung fehlender Werte und der Prüfung der Daten auf Normalverteilung auch die Bestimmung des Intraklassenkorrelationskoeffizienten. Mit diesem wird geprüft, welche Varianzanteile in der zu erklärenden Variable auf Variablen der Klassenebene zurückgehen und ob dieser Varianzanteil die Berücksichtigung der Mehrebenenstruktur anzeigt.

---

16 Da die Kompetenzstufenzuordnung auf den *plausible values* basiert, enthalten die fünf Datensätze z. T. unterschiedliche Anzahlen an Schüler:innen auf Kompetenzstufe IV und V, je nach zu Grunde liegendem *plausible value*. Da bei der Verwendung von type = imputation in Mplus alle Datensätze die gleiche Anzahl gültiger Fälle für die *plausible-value*-Variable aufweisen müssen, ist die Berechnung einer binär logistischen Regression mit einer dichotomen Variable, die nur einen Teil der Schüler:innen umfasst, nicht möglich. Unter Nutzung einer multinomialen Variable, bei der die ursprüngliche, identische Anzahl an Fällen pro Datensatz erhalten bleibt, können ebenfalls die Übergangschancen von KS IV auf KS V bestimmt werden, es kommt lediglich eine zusätzliche Information (Übergangschancen von KS IV auf KS I bis III) hinzu.

7.1.2 Fehlende Werte

Wie Kim und Bentler (2002) beschreiben, ist das Vorhandensein fehlender Werte (*missing values*) in Datensätzen eines der am weitverbreitetsten Probleme bei der Analyse von Daten. Sie entstehen, wenn Fragen von den befragten Personen nicht beantwortet werden, entweder unabsichtlich – durch Überblättern, Geben einer ungültigen Antwort oder durch Probleme mit dem Testinstrument – oder absichtlich, indem einzelne Fragen übersprungen werden oder die Befragung vorzeitig abgebrochen wird. Problematisch sind fehlende Werte unter anderem, da Datensätze mit fehlenden Werten weniger valide sind und Ergebnisse komplexer Analysen verzerrt sein können (Backhaus, Erichson, Gensler, Weiber & Weiber, 2021; Lüdtke, Robitzsch, Trautwein & Köller, 2007). Die Art und Weise, wie fehlende Werte zu Stande kommen, wird als *missing*-Mechanismus beschrieben. In der aktuellen Forschung werden dabei vor allem drei Formen unterschieden: *missing completely at random* (MCAR), *missing at random* (MAR) und *missing not at random* (MNAR; Rubin, 1976).

Kim und Bentler (2002) fassen die drei Hauptprobleme fehlender Daten wie folgt zusammen:

> "(a) if the missingness is neither MCAR or MAR, then any analysis that ignores the nonrandomness may be biased,
> (b) the existence of incomplete data usually implies a loss of information, so that estimates will be less efficient than planned, and
> (c) standard statistical methods are designed for complete data (S. 609)."

Da Datensätze ohne *missings* jedoch die Ausnahme darstellen, wurden verschiedene Möglichkeiten zum Umgang mit fehlenden Werten entwickelt. Die in den letzten Jahrzehnten gängigsten Methoden sind der paarweise oder listenweise Ausschluss der fehlenden Werte (*pairwise* oder *listwise deletion*), *Single-Imputation*-Ansätze wie das Ersetzen von fehlenden Werten durch Mittelwerte (*mean imputation*) oder durch mittels Regression vorhergesagte Werte (*conditional mean imputation*), der Einsatz eines *Full-Information-Maximum-Likelihood*-Schätzers (FIML) oder die Multiple Imputation der fehlenden Werte (Enders, 2010; Lin & Bentler, 2012). Während die erstgenannten Methoden dabei inzwischen eher veraltet bzw. mit großen Nachteilen behaftet sind, werden der FIML und die Multiple Imputation aktuell als geeignetste Methoden zum Umgang mit fehlenden Werten angesehen (Enders, 2010). Bei der Multiplen Imputation werden in einem ersten Schritt mehrere Datensätze erstellt, die

je einen geschätzten Wert pro *missing* enthalten, wobei sich die geschätzten Werte in den unterschiedlichen Datensätzen voneinander unterscheiden. Während der Analysen werden dann alle Berechnungen so oft durchgeführt, wie es imputierte Datensätze gibt, wobei die unterschiedlichen geschätzten Werte (*plausible values*) für ein eindeutiges Ergebnis zusammengeführt werden (Graham, Taylor, Olchowski & Cumsille, 2006; Rubin, 1978).[17] Der FIML hingegen ergänzt keine fehlenden Werte, sondern schätzt die wahrscheinlichsten Werte automatisch während der Analyse, indem alle zur Verfügung stehenden Informationen genutzt werden. Enders (2010) beschreibt den Ansatz der Maximum-Likelihood-Methode folgendermaßen:

> "The goal of maximum likelihood estimation is to identify the population parameter values that have the highest probability of producing a particular sample of data. Identifying the most likely parameter values requires a summary fit measure for the entire sample, not just a single score (S. 59)."

Die fehlenden Werte und Modellparameter werden somit bei der FIML-Methode, anders als bei der Multiplen Imputation, im selben Schritt und nicht nacheinander geschätzt (Graham et al., 2006). Hingegen gemein haben beide Methoden, dass die zu analysierenden Daten normalverteilt und die *missing values* dem *missing*-Mechanismus MAR oder MCAR zuordbar sein müssen (Böwing-Schmalenbrock & Jurczok, 2011; Enders, 2010).

Fehlen Werte in einem Datensatz nicht systematisch, das heißt sie hängen also nicht von anderen Variablen ab und sind somit komplett zufällig, werden sie der Kategorie MCAR zugeordnet. Zusätzliche Bedingung ist dabei, dass die fehlenden Werte auch nicht auf Grund ihres tatsächlichen Wertes (den eine Person angegeben hätte, wenn sie die Frage beantwortet hätte) fehlen dürfen. Dies ist häufiger bei Fragen zu sensiblen Themen der Fall, wenn eben jene Befragten die Fragen auslassen, die einer bestimmten Antwortkategorie zuzuordnen sind (beispielsweise bei Fragen zu Gesundheit oder Einkommen). Gleichzeitig bedeutet dies, dass auch keine Muster fehlender Werte (*missing patterns*) über alle Fälle im Datensatz hinweg erkennbar sein dürfen. Sind hingegen jedoch systematische *missing values* und damit auch Muster fehlender Werte im Datensatz erkennbar, handelt es sich bei diesen entweder um den Mechanismus MAR oder MNAR. Bei MAR hängt das Fehlen von Werten mit

[17] Diese Technik wird bei der Erstellung der TIMSS-Datensätze für die Leistungsvariablen genutzt.

anderen Variablen zusammen, es lässt sich jedoch durch diese anderen Variablen vollständig erklären. Eine Abhängigkeit von der tatsächlichen Ausprägung darf auch bei dieser Form nicht vorliegen. Im Gegensatz dazu fallen fehlende Werte unter den MNAR-Mechanismus, wenn ihr Fehlen nur durch die tatsächliche Ausprägung und nicht durch andere Variablen erklärt werden kann (Enders, 2010; Little & Rubin, 2020; Rubin, 1976).

Für die meisten statistischen Verfahren gilt, dass sie ‚ignorierbare' fehlende Werte, das heißt Werte, die dem MCAR- oder MAR-Mechanismus zuzuordnen sind, voraussetzen (Li, 2013; Little & Rubin, 2020; Lüdtke et al., 2007). Dementsprechend wichtig ist es, den *missing*-Mechanismus in den Daten zu bestimmen. Mit dem Test nach Little liegt in SPSS eine Möglichkeit vor, zu testen, ob fehlende Werte MCAR sind oder nicht (Little, 1988; Lüdtke et al., 2007). Dazu werden ähnlich wie beim t-Test Mittelwertsunterschiede zwischen Subgruppen mit gleichen *missing patterns* verglichen (Enders, 2010). Die Nullhypothese, die dabei getestet wird, ist, dass die Daten völlig zufällig im Datensatz fehlen. Wird der Test nach Little also nicht signifikant, kann die Annahme, dass die Werte MCAR sind, angenommen werden. Wird der Test nach Little hingegen auf einem Signifikanzniveau von $p \leq .05$ signifikant, deutet dies darauf hin, dass die Werte entweder MAR oder MNAR sind. Ist dies der Fall, ist eine Unterscheidung zwischen MAR und MNAR jedoch weder mit diesem noch mit anderen Tests möglich, da dafür die tatsächlichen Werte der fehlenden Daten bekannt sein müssten. Die Unterscheidung zwischen MAR und MNAR muss demnach entweder über eine Analyse der Daten – beispielsweise über die Betrachtung der *missing patterns* – oder auf inhaltlicher Ebene – über die Prüfung der inhaltlichen Plausibilität der fehlenden Werte oder das Ausmaß an erfragten sensiblen Themen – erfolgen.

Im vorliegenden Fall wird der Test nach Little signifikant ($\chi^2 = 22159.656$; $df = 12515$; $p \leq .001$), weshalb die Nullhypothese abgelehnt werden muss. Da die fehlenden Werte im zu analysierenden Datensatz also entweder MAR oder MNAR sind, wird in weiteren Schritten untersucht, welcher der beiden Mechanismen am wahrscheinlichsten im Datensatz vorliegt. Dazu wird zunächst eine Prüfung der *missing patterns*, die ebenfalls in SPSS analysiert werden können, vorgenommen. Dabei zeigt sich, dass die relevanten Muster im Schülerfragebogen vorliegen. Bei genauerer Betrachtung lassen sich vor allem zwei übergeordnete Formen erkennen: Ein großer Teil der fehlende Werte geht auf Fälle zurück, in denen für die Schüler:innen keinerlei Informationen vorliegen, d. h. die Schüler:innen haben den Fragebogen nicht bearbeitet. Für die übrigen *missing patterns* kann angenommen werden, dass sie verschiedene

Bearbeitungsstände widerspiegeln. In allen Fällen liegen für die Variablen, die eher zu Beginn des Fragebogens erfragt wurden, die Informationen nahezu vollständig vor, während im hinteren Teil des Fragebogens die Anzahl der fehlenden Werte in den unterschiedlichen Mustern zunimmt. Somit handelt es sich bei den vorliegenden *missing patterns* um Muster, die sich höchstwahrscheinlich zum einen durch den Teilnahmestatus (beispielsweise über bundeslandspezifische Verpflichtungsgrade, Einwilligung der Erziehungsberechtigten oder Abwesenheit am Testtag), zum anderen durch Erhebungsumstände wie Abbruch des Fragebogens oder Ende der Bearbeitungszeit erklären lassen. Zudem kann davon ausgegangen werden, dass auch keine inhaltlichen Gründe vorliegen, die eine MNAR-Annahme bestätigen würden, da die zu analysierenden Variablen keine hochsensiblen Fragen enthalten und auf Grund des Studiendesigns auch keine Mechanismen wie ein *selection bias* oder die Nichterreichbarkeit bestimmter Gruppen angenommen werden müssen. Damit wird für die vorliegende Arbeit die Annahme getroffen, dass die fehlenden Werte als *missing at random* angesehen werden können. Damit ist die erste Voraussetzung für den Einsatz fortgeschrittenerer Methoden zum Umgang mit fehlenden Werten wie dem FIML oder der Multiplen Imputation erfüllt.

Als nächstes wird geprüft, ob die Daten normalverteilt sind. In SPSS gibt es mit dem Kolmogorov-Smirnov-Test und dem Shapiro-Wilks-Test zwei Möglichkeiten, Daten auf Normalverteilung zu testen. Dabei ist der Shapiro-Wilks-Test dem Kolmogorov-Smirnov-Test aufgrund höherer Power vorzuziehen (Keskin, 2006; Mendes & Pala, 2003; Razali & Yap, 2011). Da jedoch eine Gewichtung der Daten erfolgt und die Fallzahlen dadurch die maximal erlaubte Fallzahl für die Anwendung des Shapiro-Wilks-Tests überschreiten (Razali & Yap, 2011), werden die Daten stattdessen mit dem ebenfalls häufig eingesetzten Kolmogorov-Smirnov-Test auf Normalverteilung geprüft. Der Kolmogorov-Smirnov-Test prüft die Nullhypothese, dass die Daten normalverteilt sind (Mendes & Pala, 2003). Die Nullhypothese muss dementsprechend abgelehnt werden, wenn der Kolmogorov-Smirnov-Test signifikant wird. Im vorliegenden Fall zeigt sich, dass der Kolmogorov-Smirnov-Test für alle Items auf dem Niveau $p < .01$ signifikant wird. Das bedeutet, dass die Daten für keines der Items normalverteilt sind. Somit ist die Voraussetzung der Normalverteilung für den Einsatz des FIMLs oder der multiplen Imputation verletzt. Mit dem *Maximum-Likelihood-Robust*-Schätzer (MLR) ist in Mplus jedoch ein FIML-Schätzer implementiert, der robust gegenüber einer Verletzung dieser Voraussetzung ist und die Nutzung des FIML trotzdem ermöglicht (Muthén & Muthén, 2017).

Bei der Auswahl des Schätzers müssen zusätzlich das Skalenniveau der abhängigen Variable sowie die geplante Analyse berücksichtigt werden. Je nach Kombination werden unterschiedliche Schätzer empfohlen. In den geplanten Analysen werden die Mplus-Befehle TYPE = TWOLEVEL sowie TYPE = TWOLEVEL RANDOM verwendet. Für beide wird bei kontinuierlichen oder kategorialen abhängigen Variablen die Nutzung des MLR-Schätzers mit numerischer Integration vorgegeben (Muthén & Muthén, 2017). Da Berechnungen unter diesen Voraussetzungen sehr komplex und zeitaufwändig werden können, kann eine Reduktion der Integrationspunkte vorgenommen werden. Standardmäßig werden in Mplus bei numerischer Integration 15 Integrationspunkte pro Dimension verwendet. Muthén und Muthén (2017) empfehlen, dass die Integrationspunkte auf zehn oder sieben für ausreichend präzise Schätzungen reduziert werden können, nach Rabe-Hesketh und Skrondal (2012) ist unter Nutzung numerischer Integration aber auch eine Reduktion bis auf einen Integrationspunkt möglich. Da in den geplanten Modellen bis zu sieben Dimensionen der Integration vorkommen können, wobei von Muthén und Muthén (2017) Modelle mit fünf oder mehr Dimensionen als „very heavy" eingeschätzt werden (S. 528), ist eine Berechnung mit sieben Integrationspunkten teilweise mit den gegebenen technischen Möglichkeiten nicht durchführbar. Deshalb werden für die Berechnung der komplexen Hauptmodelle sechs Integrationspunkte verwendet.

Die Prüfung der Voraussetzungen zum Umgang mit fehlenden Werten hat gezeigt, dass die fehlenden Werte als MAR angesehen werden können und somit die erste Voraussetzung für multiple Imputation oder den FIML erfüllt ist. Die zweite Voraussetzung hingegen, dass die Daten normalverteilt sind, ist verletzt. Deshalb wird für diese Arbeit der Einsatz des FIML-Ansatzes mit dem MLR-Schätzer gewählt, der robust gegenüber dieser Verletzung ist. Darüber hinaus wird der MLR-Schätzer auf Grund der kategorialen abhängigen Variable mit der numerischen Integration kombiniert und zur Reduktion der Komplexität und Zeitintensität der Berechnungen die Zahl der Integrationspunkte verringert.

7.1.3 Intraklassenkorrelation

Wie eingangs erläutert, ergibt sich aus den Hypothesen und der Datengrundlage die Notwendigkeit, eine Mehrebenenanalyse zu berechnen. Dafür muss in einem ersten Schritt geprüft werden, ob die Datengrundlage die Durchführung einer Mehrebenen-

analyse stützt. Dazu kann der sogenannte Intraklassenkorrelationskoeffizient (engl.: *Intraclasscorrelationcoefficient*; ICC) herangezogen werden. Dieser gibt an, wie viel Prozent der Varianz in der zu erklärenden Variable auf Varianz auf der Klassenebene[18] und auf Varianz auf der Individualebene zurückgeht (Bickel, 2007; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Der ICC kann mithilfe des sparsamsten Modells der Mehrebenenanalyse, dem Nullmodell, berechnet werden. Beim Nullmodell werden keine erklärenden Variablen in das Modell aufgenommen, sondern nur die zu erklärende, intervallskalierte Variable sowie eine Funktionsvariable zur Bestimmung der Gruppenzugehörigkeit der einzelnen Beobachtungen, sodass die Varianz in der Zielvariable auf die beiden Ebenen aufgeteilt wird (Langer, 2010). Der ICC zeigt somit, ob die Varianz auf Klassenebene so hoch ist, dass anstatt einfacher OLS-Regressionen Mehrebenenanalysen durchzuführen sind (Goldstein, 1999). Der ICC kann dabei Werte zwischen 0 und 1 annehmen, wobei ein ICC von 0 bedeutet, dass keine Varianz auf der Klassenebene vorliegt und ein Wert von 1, dass 100 % der Varianz in der Zielvariable ausschließlich durch Merkmale auf der Klassenebene erklärt werden können. Im erziehungswissenschaftlichen Kontext wird häufig ein ICC von 0.1 als Grenzwert angesehen, ab dem eine Mehrebenenanalyse durchgeführt werden sollte (Urban, 2022; siehe auch Eid, Gollwitzer & Schmitt, 2010; Lee, 2000), einige Autor:innen vertreten jedoch die Meinung, dass auch bereits ICCs ab einem Wert von 0.05 die Berechnung eines Mehrebenenmodells rechtfertigen (Cohen, Cohen, West & Aiken, 2003; Hox, 2010).

Da bei den geplanten Analysen eine kategoriale Variable untersucht werden soll und somit die Voraussetzung der Intervallskalierung nicht gegeben ist, lässt sich der ICC nicht direkt berechnen. Um dennoch einen Richtwert zu erhalten, ob eine Mehrebenenanalyse gerechnet werden sollte, wird das Nullmodell mit der ursprünglichen intervallskalierten Leistungsvariable berechnet. Zusätzlich erfolgt die Bestimmung der ICCs auch für die Kompetenzstufenvariable, um einen möglichst nahen Wert zu erhalten. Die Kompetenzstufenvariable, die die Grundlage für die kategoriale Zielvariable bildet, ist zwar nicht intervallskaliert, kann nach Bentler und Chou (1987) jedoch als solche behandelt werden, wenn wie in diesem Fall mindestens vier Ausprägungen vorliegen. Da die Kompetenzstufenvariable eine gruppierte Version der intervallskalierten Leistungsvariable ist, kann angenommen werden, dass die ICCs

[18] Da in der vorliegenden Arbeit ein 2-Ebenen-Modell mit Schüler:innen auf Ebene 1 und Klassen auf Ebene 2 gerechnet wird, werden die methodischen Grundlagen exemplarisch für diese Form des Modells erklärt. Weitere Informationen zur Mehrebenenanalyse finden sich in Kapitel 7.3.2.

aufgrund geringerer Varianz in der Variable niedriger ausfallen als für die intervallskalierte Leistungsvariable und stärker den tatsächlichen ICCs entsprechen.

Da die Leistungswerte über fünf *plausible values* geschätzt werden, wird pro Nullmodell nicht nur ein ICC berechnet, sondern fünf ICCs. Insgesamt liegen die Werte für die ungruppierten Leistungsvariablen zwischen 11.2 % und 12.9 % und für die Kompetenzstufen-Variable wie erwartet etwas niedriger zwischen 10.0 % und 11.9 %. Damit liegen alle Werte bei oder über den empfohlenen Grenzwerten von 5 % bzw. 10 %. Die hilfsweise herangezogenen ICCs deuten somit darauf hin, dass die Mehrebenenstruktur bei den Analysen berücksichtigt werden sollte.

7.1.4 Zusammenfassung und Schlussfolgerungen

Wie der Test nach Little und die Betrachtung der *missing patterns* gezeigt haben, können die fehlenden Werte im vorliegenden Datensatz als *missing at random* (MAR) angesehen werden. Damit ist die erste Voraussetzung für den Einsatz des FIML oder der Multiplen Imputation erfüllt. Die zweite Voraussetzung, dass die Daten normalverteilt sein müssen, ist hingegen verletzt. Deshalb werden die Analysen mithilfe des MLR-Schätzers durchgeführt. Aufgrund der kategorialen abhängigen Variable wird zudem die numerische Integration unter Nutzung von sechs Integrationspunkten bei den komplexen Modellen eingesetzt. Die Analysen zum Interklassenkorrelationskoeffizienten haben zudem ergeben, dass ein Mehrebenenmodell zur Beantwortung der Hypothesen gerechnet werden sollte.

7.2 Konfirmatorische Faktorenanalysen

Bevor die in Kapitel 6.4.4 beschriebenen Merkmale im Rahmen der Modelle untersucht werden, werden die Skalen mittels konfirmatorischer Faktorenanalysen (*confirmatory factor analysis*; CFA) geprüft. Eine zusätzliche Prüfung im Rahmen einer *Explorativen Faktorenanalyse* (EFA), wie sie beispielsweise bei neu entwickelten Skalen zur Überprüfung der Qualität und Faktorstruktur zum Einsatz kommt (Brandt, 2020), ist bei den zu analysierenden Skalen nicht notwendig, da in TIMSS etablierte und teils seit mehreren Zyklen eingesetzte sowie empirisch überprüfte Skalen mit klaren theoretischen Vorannahmen vorliegen.

7.2.1 Grundlagen der konfirmatorischen Faktorenanalyse

Die konfirmatorische Faktorenanalyse beruht auf der Annahme, dass latente, also nicht direkt beobachtbare Konstrukte über direkt messbare Indikatoren (manifeste Variablen) erfasst werden können (Backhaus, Erichson & Weiber, 2015). Dabei werden vorab theoriegeleitet Hypothesen aufgestellt, wie die manifesten Variablen mit den latenten Konstrukten zusammenhängen bzw. wie sie diese abbilden. Dazu wird ein sogenanntes Messmodell entwickelt, das diese angenommenen Beziehungen widerspiegelt und anschließend untersucht, wie gut diese Messmodelle zur Erfassung latenter Konstrukte zu den Daten passen. Die CFA lässt sich somit den strukturprüfenden Verfahren zuordnen (Gäde, Schermelleh-Engel & Brandt, 2020). Die Passung der Variablen zum latenten Konstrukt wird über sogenannte Faktorladungen bestimmt, die die Korrelationen der einzelnen Items mit dem angenommenen Konstrukt angeben (Backhaus et al., 2021). Diese sind auf Werte zwischen -1 und 1 normiert und sollten eine Höhe von mindestens .40 (Backhaus et al., 2015; Homburg & Baumgartner, 1995), besser aber von .70 oder größer haben (Huber, Herrmann, Meyer, Vogel & Vollhardt, 2007). Insbesondere bei großen Stichproben konnte jedoch gezeigt werden, dass auch niedrigere Faktorladungen von .20 oder .30 als ausreichend eingestuft werden können (Gagné & Hancock, 2006; Swisher, Beckstead & Bebeau, 2004). Zudem sollte die Auswahl der Items nicht nur datengestützt erfolgen, sondern auch inhaltlich plausibel sein. Dementsprechend ist es auch möglich, Items beizubehalten, die diese Grenzwerte verfehlen, wenn es aus inhaltlichen Gründen als sinnvoll erachtet wird (Berning, 2018). Für die vorliegende Arbeit wird in Anbetracht der unterschiedlichen Grenzwerte folgende Klassifikation der Ergebnisse genutzt:

Tab.: 1 Klassifikation der Ergebnisse konfirmatorischer Faktorenanalysen

ausreichend	akzeptabel	gut
> .30	> .40	> .70

7.2.2 Überprüfung der Datengrundlage

Voraussetzung für die Durchführung von CFAs ist die Überprüfung der Datengrundlage auf deren Eignung für diese Analyse. Dies lässt sich mithilfe des *Kaiser-Meyer-Olkin-Kriteriums* (KMO) oder des Bartlett-Tests auf Sphärizität prüfen. Da der Bartlett-Test normalverteilte Daten voraussetzt (Cleff, 2015) und diese in diesem Fall nicht vorliegen, wird das *Kaiser-Meyer-Olkin*-Kriterium zur Einschätzung der Daten herangezogen. Dieses prüft über partielle Korrelationen, ob die gemeinsame Streuung von Items auf einen Faktor zurückgeht (Fromm, 2008). Das KMO-Kriterium kann Werte zwischen 0 und 1 annehmen, wobei höhere Werte eine bessere Eignung der Daten anzeigen. Kaiser (1974) schlägt dabei folgende Kategorisierung vor:

< .50	.50 – .59	.60 – .69	.70 – .79	.80 – .89	> .90
unacceptable	miserable	mediocre	middling	meritorious	marvelous

Abb.: 15 Klassifikation des KMO-Kriteriums nach Kaiser (1974)

7.2.3 Kriterien zur Beurteilung der Ergebnisse

Für die Überprüfung von Ergebnissen konfirmatorischer Faktorenanalysen gibt es eine Vielzahl verschiedener Kriterien, die für die Einschätzung von Modellen kombiniert werden sollten (Hu & Bentler, 1999; Kline, 2016). Da diese Kriterien gleichermaßen zur Beurteilung von Strukturgleichungsmodellen und Mehrebenenanalysen geeignet sind, werden sie auch zur Einschätzung der berechneten Teil- und Gesamtmodelle herangezogen.[19]

Eines der bekanntesten klassischen Kriterien zur Einschätzung der Modellgüte ist das Chi-Quadrat (χ^2) bzw. der Chi-Quadrat-Differenztest, die zum einen Aussa-

[19] Da Maße der Varianzaufklärung wie beispielsweise Cohen´s *d* oder Nagelkerkes R^2, die zur Anwendung kommen, um die Bedeutung von Ergebnissen einzuschätzen, unter Nutzung der beschriebenen Modellspezifikationen nicht in Mplus ausgegeben werden und auch nicht aus den zur Verfügung stehenden Kennwerten sinnvoll berechnet werden können, wird in diesem Kapitel nicht näher auf diese Maße eingegangen.

gen über die Passung des Modells zu den Daten, zum anderen Vergleiche genesteter Modelle ermöglichen. Ergänzt wird dieser Wert i. d. R. mit Informationskriterien und *Model-Fit*-Indizes. Mit der Devianz, dem *Akaike Information Criterion* (AIC) und dem *Bayesian Information Criterion* (BIC) werden in dieser Arbeit drei Kriterien verwendet, die ebenfalls einen Vergleich von Modellen möglich machen, ohne dabei die Schachtelung der Modelle vorauszusetzen. Gleichzeitig sind diese drei Kriterien nicht nur für Modelle mit metrischen Daten, sondern auch für Modelle mit kategorialen und nominalen Variablen geeignet. Damit die Beurteilung der Modelle aber auch ohne Referenzmodelle möglich ist, werden zudem *Model-Fit*-Indizes ausgewählt, die dies ermöglichen. Die gängigsten *Model-Fit*-Indizes, die häufig in verschiedenen Kombinationen zur Modellbeurteilung herangezogen werden, sind für absolute Fit-Indizes der *Root Mean Square Error of Approximation* (RMSEA) und der *Standardized Root Mean Square Residual* (SRMR) sowie für relative Fit-Indizes der *Comparative Fit Index* (CFI) und der *Tucker-Lewis-Index* (TLI). Bezogen auf die relativen Fit-Indizes argumentiert Kline (2016), dass der CFI und der TLI sehr ähnliche Werte produzieren, sodass einer der beiden Werte zur Einschätzung eines Modelles ausreichend ist. Da der TLI zudem komplexe Modelle – wie sie in der vorliegenden Arbeit berechnet werden – eher bestraft, wird in dieser Arbeit der Empfehlung von Kline gefolgt, neben dem χ^2 noch den RMSEA, CFI und SRMR zur Beurteilung von Modellen zu verwenden. Die ausgewählten Kriterien werden im Folgenden kurz skizziert.

Mithilfe des χ^2-Wertes lässt sich die Güte eines Modells darüber bestimmen, wie gut ein Modell zu den Daten passt. Wird das χ^2 signifikant, muss die Nullhypothese, dass ein Modell exakt mit der Datengrundlage übereinstimmt, verworfen werden (Geiser, 2010). Dabei ist jedoch zu berücksichtigen, dass bei „sample sizes between 75 and 200, the chi-square test is typically an appropriate indicator of model fit. However, when the sample size is over 400, most models are rejected" (Stone, 2021, S. 2). Für die Analysen in dieser Arbeit ist dieser Wert also aufgrund der großen Stichprobe nur eingeschränkt aussagekräftig. Darüber hinaus bietet das χ^2 mit dem χ^2-Differenztest aber auch die Möglichkeit zwei Modelle miteinander zu vergleichen. In Mplus wird das χ^2 standardmäßig ausgegeben, das allerdings nur unter Nutzung des ML-Schätzers direkt verwendet werden kann. Wird stattdessen wie in der vorliegenden Arbeit der MLR-Schätzer eingesetzt, muss stattdessen der Satorra-Bentler-χ^2-Differenztest

mit einem Scaling-Korrekturfaktor manuell berechnet werden (Kleinke, Schlüter & Christ, 2017; Twele, 2022).[20] Der Korrekturfaktor *cd* berechnet sich dabei wie folgt:

$$cd = \frac{d0 * c0 - d1 * c1}{d0 - d1},$$

wobei *d0* die Freiheitsgrade des Nullmodells (bzw. des stärker restriktiven Modells), *d1* die Freiheitsgrade des Alternativmodells, *c0* der Korrekturfaktor für das Nullmodell (bzw. des stärker restriktiven Modells) und *c1* der Korrekturfaktor für das Alternativmodell sind. Anschließend kann der Satorra-Bentler-Differenztest *T* berechnet werden:

$$TRd = \frac{(T0 * c0 - T1 * c1)}{cd},$$

wobei *T0* für das nach Satorra-Bentler korrigierte χ^2 des Nullmodells (bzw. des stärker restriktiven Modells), *T1* für das korrigierte χ^2 des Alternativmodells und *c0* bzw. *c1* erneut für die Korrekturfaktoren stehen (Kleinke et al., 2017). Unter Nutzung der Tabellen zur χ^2-Verteilung und der Differenz aus den Freiheitsgraden kann anschließend geprüft werden, ob das Alternativmodell signifikant besser als das Referenzmodell auf die Daten passt.

Mit der Devianz, dem AIC und dem BIC stehen drei weitere Informationskriterien zur Verfügung, mit denen Modelle im Vergleich beurteilt werden können. Der Devianzwert beruht auf dem der Mehrebenenanalyse zu Grunde liegenden *Maximum-Likelihood*-Ansatz. Da die Devianz in Mplus nicht automatisch ausgegeben wird, muss dieser Wert manuell berechnet werden. Die Formel zur Berechnung lautet:

$$D = -2LL,$$

wobei *LL* für den *Loglikelihood*-Wert steht. Das Modell mit dem kleineren Devianzwert ist dabei dem mit dem größeren Wert vorzuziehen (Jäckle & Schärdel, 2017; Tausendpfund, 2020). Das *Akaike Information Criterion* (AIC) hingegen ermöglicht Aussagen darüber, wie unterschiedlich zwei Modelle sind (Akaike, 1973). Das Modell mit dem niedrigsten AIC ist „‚closest' to the unknown reality that generated the data, from among the candidate models considered" (Burnham & Anderson, 2002, S. 62).

---

20 Da der χ^2-Wert bei Modellen mit der Leistungsvariable auf Grund der numerischen Integration nicht verfügbar ist, kommt der χ^2-Differenztest nur bei konfirmatorischen Faktorenanalysen zur Anwendung.

Da der AIC ebenfalls keine absoluten, sondern relative Werte produziert, müssen zur Beurteilung mindestens zwei Modelle berechnet werden, die mit dem AIC in eine Rangfolge gebracht werden können. Das Modell mit dem niedrigsten AIC weist dabei den besten Fit auf. Mit zunehmender Stichprobengröße belohnt der AIC jedoch komplexe Modelle (Burnham & Anderson, 2004; Cavanaugh & Neath, 2019). Das *Bayesian Information Criterion* (BIC), seltener auch Schwarz-Kriterium genannt, wurde 1978 als Alternative zum AIC entwickelt (Schwarz, 1978). So verfolgt der BIC einen bayesianischen Ansatz und verwendet zudem eine abgewandelte Form des Strafterms. Die Veränderung im Strafterm im Vergleich zum AIC führt dazu, dass der BIC bereits ab einer Stichprobengröße von $n = 8$ komplexere Modelle bestraft und einfachere Modelle vorzieht. Auch für den BIC gilt jedoch, dass der niedrigste Wert das Modell mit dem besten Fit anzeigt (Burnham & Anderson, 2002; Cavanaugh & Neath, 2019).

Neben den beschriebenen Kriterien ist es angezeigt, auch Fit-Indizes heranzuziehen, die ohne Vergleichsmodell aussagekräftig sind. Zu den absoluten Fit-Indizes gehört unter anderem der *Root Mean Square Error of Approximation* (RMSEA). Der RMSEA kann Werte zwischen 0 und 1 annehmen, wobei ein Wert von 0 den perfekten *Model-Fit* anzeigt (Kleinke et al., 2017; Kline, 2016). Curran, Bollen, Chen, Paxton und Kirby (2003) fassen die Daumenregeln zur Einschätzung der Werte folgendermaßen zusammen: „values of less than .05 to denote good model fit, values less than .08 to denote adequate model fit, and values exceeding .10 to denote poor model fit" (S. 213). Dabei neigt der RMSEA dazu, einfache Modelle und große Stichproben zu belohnen (Kline, 2016). Wie der RMSEA gehört auch der *Standardized Root Mean Square* (SRMR) zu den absoluten Fit-Indizes und *badness-of-fit-statistics*. Das bedeutet, dass wie beim RMSEA kleinere Werte für den SRMR größeren vorzuziehen sind, wobei Werte über .10 als schlechter Fit gelten (Kline, 2016). Nach Hu und Bentler (1999) gilt zudem für ein akzeptables Modell ein „cutoff value close to .08" (S. 27), während Schermelleh-Engel und Moosbrugger (2003) raten, für ein gutes Modell Werte von unter .05 anzustreben. Da der SRMR einen ähnlichen Berechnungsansatz wie das χ^2 verfolgt, empfehlen Asparouhov und Muthén (2018) den SRMR insbesondere dann anzugeben, wenn große Stichproben analysiert werden, da das χ^2 in diesem Fall dazu neigt, signifikant zu werden und somit nur eingeschränkt interpretierbar ist.

Mit dem *Comparative Fit Index* (CFI) wird außerdem ein relativer Fit-Index angegeben. Diese werden berechnet, indem das Modell automatisch mit einem Referenzmodell, häufig dem Nullmodell, verglichen wird (Kline, 2016). Der CFI kann Werte zwischen 0 und 1 annehmen, wobei für diesen Index hohe Werte einen besseren Fit

bedeuten, da der CFI zu den *goodness-of-fit-statistics* gehört. Kline (2016) führt zur Interpretation des CFIs aus, dass ein "CFI of 0.9 says that the fit of the researchers model is about 90% better than that of the baseline model" (S. 276). Von einem guten *Model-Fit* kann gesprochen werden, wenn der CFI bei .95 liegt (Hu & Bentler, 1999), von einem akzeptablen Fit ab .90 (Nagy, 2007).

Zusätzlich zu den beschriebenen Gütekriterien wird speziell für die Faktorenanalysen Cronbach´s Alpha als Maß der internen Konsistenz der einzelnen Skalen berechnet. Cronbach´s Alpha lässt sich über die Betrachtung der Korrelationen einzelner Items einer Skala untereinander berechnen, mit dem Ziel zu bestimmen, ob sie in ähnlicher Weise, also konsistent, beantwortet wurden (Moosbrugger & Kelava, 2020). Entscheidend bei der Berechnung von Cronbach´s Alpha ist es deshalb, gegensätzlich formulierte Items vorab zu invertieren. Trotzdem ist es möglich, dass invertierte Items eine eigene Systematik aufweisen, also sich von den nicht-invertierten Items unterscheiden und einen separaten Faktor bilden (Moosbrugger & Kelava, 2020). Dies muss bei der Interpretation von Cronbach´s Alpha, aber auch den Faktorladungen einer CFA berücksichtigt werden. Taber (2018) zeigte, dass in den meisten rezenten Publikationen ein unterer Grenzwert von .70 für Cronbach´s Alpha für eine akzeptable interne Konsistenz angenommen wird. Als gute interne Konsistenz wird darüber hinaus ein Wert von mindestens .80, als sehr gute interne Konsistenz ein Wert von .90 oder höher angesehen (Döring & Bortz, 2016; Tachtsoglou & König, 2017).

Bei der Interpretation der dargestellten *Model-Fit*-Werte ist zu berücksichtigen, dass diese häufig als strikte Grenzwerte missverstanden werden. Stattdessen sind sie als Daumenregeln entwickelt worden, die zudem nur unter bestimmten Bedingungen wie der Nutzung des ML-Schätzers bei kontinuierlichen Variablen gültig sind. Wie in Kapitel 7.1.2 gezeigt, liegt jedoch für die analysierten Variablen keine Normalverteilung vor, weshalb stattdessen der MLR-Schätzer zum Einsatz kommt. Wie die einzelnen *Model-Fit*-Indizes unter alternativen Bedingungen funktionieren, wurde bisher jedoch eher selten untersucht. Für den RMSEA konnten Gao, Shi und Maydeu-Olivares (2020) beispielsweise zeigen, dass dieser unter Nutzung des MLR höhere Werte produziert als beim ML-Schätzer. Brosseau-Liard und Savalei (2014) kamen zu dem Schluss, dass sich ebenso für den CFI bei der Verletzung der Normalverteilungsvoraussetzung höhere Werte ergeben. Auf Basis dieser Befunde werden in der vorliegenden Arbeit auch die als ‚ausreichend' einzustufenden Werte für den RMSEA und den SRMR (jeweils .10), die in Publikationen mit Analysen auf Basis des ML-Schätzers meist als zu hoch erachtet werden, noch akzeptiert. Für den CFI konnte

Kriterium		*Model-Fit*	
χ^2-Differenztest		signifikant besseres Modell zu bevorzugen	
Devianz*		Modell mit kleinerer Devianz zu bevorzugen	
AIC*		Modell mit kleinerem AIC zu bevorzugen	
BIC*		Modell mit kleinerem BIC zu bevorzugen	
	Ausreichender *Model-Fit*	Akzeptabler *Model-Fit*	Guter *Model-Fit*
RMSEA	≤.0	<.08	<.05
SRMR	≤.10	<.08	<.05
CFI	-	≥.90	>.95

* = verfügbare Modell-Güte-Kriterien bei Modellberechnungen mit Leistungsvariable

Abb.: 16 Übersicht über die in dieser Arbeit genutzten *Model-Fit*-Werte

keine entsprechende Abstufung in der Literatur gefunden werden, weshalb hier nur zwei Qualitäten (‚akzeptabel' und ‚gut') unterschieden werden. Eine Übersicht über die für diese Arbeit gültigen Richtwerte, die sich aus den vorangegangenen Darstellungen ergibt, findet sich in Abbildung 16.

Auch die Bedingung der kontinuierlichen abhängigen Variable ist in dieser Arbeit nicht erfüllt. Wie in Kapitel 7.1.2 beschrieben, hat dies unter anderem zur Folge, dass der MLR-Schätzer in Kombination mit der numerischen Integration bei den Analysen eingesetzt wird. Unter diesen Voraussetzungen können das χ^2, der RMSEA, der SRMR und der CFI nicht berechnet werden, sodass bei Modellen, die als abhängige Variable die Leistungsvariable beinhalten, nur die Devianz, der AIC und der BIC zur Einschätzung der Modelle zur Verfügung stehen (Kelava & Brandt, 2014; Muthén, 2014; Schermelleh-Engel, Kerwer & Klein, 2014). In den Ausführungen zu den einzelnen *Model-Fit*-Indizes konnte zudem gezeigt werden, dass auch Bedingungen wie die Stichprobengröße oder die Komplexität des Modells die *Model-Fit*-Werte beeinflussen können. Insbesondere aufgrund der großen Stichprobe, die bei einigen der ausgewählten Kriterien zu Verzerrungen führen kann, wird darauf geachtet, dass χ^2 immer in Kombination mit dem SRMR sowie AIC und BIC nur gemeinsam interpretiert werden. Für den RMSEA zeigen sich wiederum niedrigere Werte, wenn einfache Modelle und

große Stichproben analysiert werden. Da in den Analysen in dieser Arbeit eher komplexe Modelle, die vom RMSEA bestraft werden, mit großen Stichproben untersucht werden, wird angenommen, dass sich beide Effekte gegenseitig aufheben und somit keine besondere Berücksichtigung bei der Interpretation vorgenommen werden muss.

7.3 Mehrebenenstrukturgleichungsmodelle

Nachdem in den vorangegangenen Kapiteln die Voraussetzungen für die Analysen überprüft und erste methodische Grundlagen geschaffen wurden, werden im Folgenden die Methoden zur Modellberechnung erläutert. Aus den Hypothesen und der Datengrundlage ergibt sich die Notwendigkeit, zur Abbildung der komplexen postulierten Beziehungen ein Strukturgleichungsmodell zu berechnen. Dieses muss zusätzlich die Mehrebenenstruktur in den Daten berücksichtigen. Deshalb werden zunächst Strukturgleichungsmodelle und anschließend Mehrebenenanalysen genauer vorgestellt und anschließend zusammengeführt.

7.3.1 Strukturgleichungsmodelle

Strukturgleichungsmodelle (engl. *strucutral equation model;* SEM) können als Erweiterung zu einfachen oder multiplen Regressionen mit metrischen, kategorialen oder nominalen Variablen eingesetzt werden, wenn mehrere abhängige Variablen mit weiteren Beziehungen zu einander untersucht werden sollen (Backhaus, Erichson, Plinke & Rolf, 2018; Wolf & Best, 2010). Diese Form von Regressionsmodellen nennt sich Pfadmodell. Strukturgleichungsmodelle bieten die Möglichkeit, solche Pfadmodelle zu testen und simultan mit den in Kapitel 7.2.1 beschriebenen Messmodellen, also Faktorenanalysen zur Spezifikation latenter Variablen, zu kombinieren. Dieses Vorgehen bietet den Vorteil, dass durch das Messmodell Messfehler und durch das Pfadmodell komplexe Beziehungen zwischen den Konstrukten gleichzeitig berücksichtigt werden können (Kleinke et al., 2017).

Zwei besondere Formen von Zusammenhängen, die mit Strukturgleichungsmodellen untersucht werden können, sind Mediator- und Moderatoranalysen. Bei Moderatoranalysen, auch Interaktionen oder konditionale Beziehungen genannt, ‚moderiert', also beeinflusst, eine Variable den Zusammenhang zwischen zwei anderen Variablen.

Das bedeutet, dass die Stärke dieses Zusammenhanges abhängig von der Ausprägung der dritten Variable ist (Berning, 2018). Bei Mediatoranalysen hingegen wird der Zusammenhang zwischen zwei Variablen nicht moderiiert, sondern mediiert. Das bedeutet, dass ein Effekt von einer unabhängigen auf eine abhängige Variable teilweise oder vollständig von einer dritten Variable vermittelt wird. Von einer vollständigen Mediation spricht man, wenn sich ein signifikanter Effekt einer Variablen auf eine andere auflöst, wenn eine Mediatorvariable ergänzt wird. Verringert sich lediglich die Stärke des direkten Zusammenhangs und es kommt ein signifikanter, indirekter Effekt hinzu, spricht man von einer teilweisen Mediation. Ist der indirekte Zusammenhang nicht signifikant und der ursprüngliche direkte Effekt bleibt (nahezu) unverändert erhalten, liegt kein indirekter Effekt vor. Neben ‚einfachen' Mediationen sind auch serielle Mediationen, bei denen mehrere Mediatoren aufeinander folgen, möglich (Baron & Kenny, 1986; Berning, 2018).

Aufgrund der komplexen Pfad-Messmodelle, bei denen Variablen gleichzeitig erklärend als auch abhängig sein können, haben sich im Kontext von Strukturgleichungsmodellen alternative Begrifflichkeiten und feste Regeln zur grafischen Darstellung der Zusammenhänge etabliert. Bei den Begrifflichkeiten ist die Unterscheidung zwischen exogenen Variablen (ausschließlich erklärende bzw. unabhängige Variablen) und endogenen Variablen, die erklärt werden sollen oder die gleichzeitig abhängig und unabhängig sind, üblich (Berning, 2018). Für die Darstellung gelten folgende Konventionen, die in Abbildung 17 schematisch visualisiert sind: manifeste Variablen werden in Rechtecken, latente Variablen in Ellipsen und Fehlerterme in Kreisen (manchmal auch ohne Umrandung) dargestellt. Ein einfacher Pfeil zwischen den Elementen kennzeichnet einen Regressionspfad, ein Doppelpfeil Korrelationen. Häufig werden zudem die Messmodelle der exogenen und endogenen Variablen und das Pfadmodell gekennzeichnet (Berning, 2018).

Auch wenn die einfachen Pfeile in Strukturgleichungsmodellen gerichtete Zusammenhänge darstellen, muss beachtet werden, dass mit Strukturgleichungsmodellen keine echten Kausalaussagen getroffen werden können, solange keine Längsschnittdaten verwendet werden. Bei der Nutzung von querschnittlichen Daten wie in der vorliegenden Arbeit müssen die Zusammenhänge und deren Richtungen vorab theoretisch abgeleitet werden. Mit Strukturgleichungsmodellen können dann Aussagen darüber getroffen werden, ob die abgeleiteten Zusammenhänge zwischen zwei Variablen tatsächlich existieren. Eine Überprüfung, ob zudem die Annahme über Ursache und Wirkung zutrifft, ist mit einem Strukturgleichungsmodell mit Querschnittsdaten

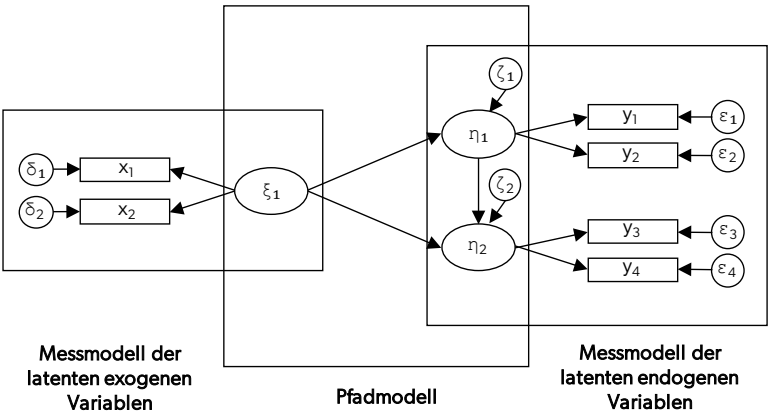


Abb.: 17 Strukturgleichungsmodell in Anlehnung an Backhaus et al. (2015)

nicht möglich. Prinzipiell könnte es somit sein, dass ein alternatives Modell mit entgegengesetzten Effekten genauso auf die Daten passt, weshalb kausale Interpretationen nicht getroffen werden sollten (Wolf & Best, 2010).

7.3.2 Mehrebenenmodelle

Die Mehrebenenanalyse (MEA; engl. *multilevel analysis*) ist ein Verfahren, das auf der Regressionsanalyse basiert. Während die Regressionsanalyse voraussetzt, dass die einzelnen Individuen voneinander unabhängig sind, werden häufig Daten genutzt, die eine geschachtelte Struktur aufweisen. Von einer geschachtelten Datenstruktur spricht man, wenn die einzelnen Elemente der Stichprobe unterschiedlichen Gruppen angehören. Dies ist häufig der Fall, wenn die Befragten in einem mehrstufigen Ziehungsverfahren und/oder aus Gründen der Effizienz oder Erreichbarkeit über ihre Gruppenzugehörigkeit ausgewählt wurden (z. B. Befragung aller Kinder ausgewählter Klassen oder Mitarbeiter:innen ausgewählter Firmen). Dabei kann angenommen werden, dass sich die Elemente innerhalb von Gruppen ähnlicher sind als zwischen Gruppen und somit die Annahme der Unabhängigkeit der Fälle verletzt ist (Eid, Gollwitzer & Schmitt, 2015). Für die Analyse solcher Daten wurde die MEA ent-

wickelt, die es im Gegensatz zur OLS-Regressionsanalyse ermöglicht, die Abhängigkeit der Beobachtungen zu berücksichtigen. Während bei der OLS-Regressionsanalyse die Standardfehler unterschätzt würden, wenn die hierarchische Struktur der Daten unberücksichtigt bleibt, können die Standardfehler mit dem MEA korrekt geschätzt werden (Pötschke, 2014).

Voraussetzung für eine MEA ist, dass in den Daten Informationen darüber enthalten sind, welcher Gruppe die einzelnen Elemente angehören. Dabei ist für die klassische MEA eine eindeutige Zuordnung notwendig, d. h. dass jedes Element nur einer Gruppe zugeordnet werden kann (Jäckle & Schärdel, 2017). In vielen Fällen befinden sich auf der untersten Ebene (auch Individualebene oder Level 1 genannt) Individuen wie beispielsweise Schüler:innen, es ist aber auch möglich, dass hier Institutionen, Städte oder Beobachtungen über die Zeit hinweg untersucht werden. Auf der nächsthöheren Ebene sind entsprechend z. B. Klassen, Träger, Bundesländer oder Personen, die im Längsschnitt betrachtet werden, angesiedelt. Nach dieser Logik können beliebig viele weitere Ebenen ergänzt werden, wenn die entsprechende Software zur Berechnung vorhanden ist. Da in dieser Arbeit Schüler:innen in Klassen betrachtet werden, wird in den folgenden Erläuterungen der Fokus auf das 2-Ebenen-Modell gelegt. Diese Herangehensweise verdeutlicht, dass die MEA auf der Annahme basiert, dass Merkmale auf Level 1 nicht nur von den individuellen Merkmalen einer Person abhängen, sondern auch von den äußeren Einflüssen, dem Kontext, beeinflusst werden. Sie ermöglicht es, Merkmale beider Ebenen gleichzeitig zu betrachten und deren Einfluss auf eine abhängige Variable zu schätzen (Langer, 2010; Pötschke, 2014).

Dabei sind verschiedene Formen von Zusammenhängen möglich (siehe Abbildung 18). Im linken Kasten ist ein einfacher Zusammenhang zwischen einer erklärenden Variable auf Level 2 und einer zu erklärenden Variable auf Level 1 dargestellt. Im Kasten in der Mitte wird zusätzlich zu dem Effekt einer Level-2-Variable auf die Level-1-Variable auch noch ein Effekt einer weiteren Level-1-Variable angenommen. Im rechten Kasten schließlich ist die komplexeste Variante dargestellt. Bei dieser sind zusätzlich zu direkten Effekten auf den beiden Ebenen sogenannte Interaktionseffekte dargestellt, bei denen Variablen nicht nur Einfluss auf die zu erklärende Variable haben, sondern auch auf die Zusammenhänge zwischen zwei Variablen (Jäckle & Schärdel, 2017; Langer, 2010).

Damit auch die komplexen Effekte noch korrekt geschätzt werden können, setzt die MEA möglichst umfangreiche Stichproben voraus. Etabliert hat sich die Faustregel, dass mindestens 30 Gruppen mit jeweils mindestens 30 Individuen bzw. ein Produkt

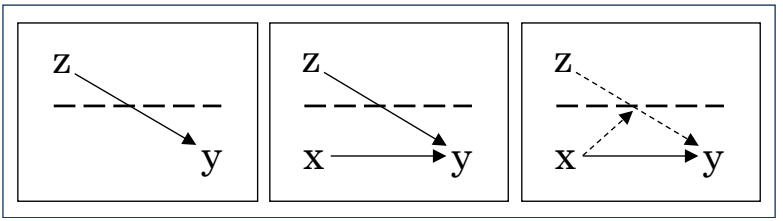


Abb.: 18 Zusammenhänge in MEAs in Anlehnung an Jäckle und Schärdel (2017)

von $n = 900$ vorhanden sein sollten. Dies konnte auch für logistische abhängige Variablen bestätigt werden (Bryan & Jenkins, 2015; Tausendpfund, 2020). Hox, Moerbeek und van de Schoot (2018) betonen jedoch, dass „for accuracy and high power a large number of groups appears more important than a large number of individuals per group" (S. 214). Mit 211 Klassen mit durchschnittlich 24 Schüler:innen ($SD = 11.16$) ist die Datengrundlage vor diesem Hintergrund für einer MEA geeignet.

Bevor die Mehrebenanalyse durchgeführt werden kann, muss außerdem entschieden werden, ob die unabhängigen Variablen zentriert werden sollten. Zentrierungen können eingesetzt werden, um eine sinnvolle Interpretation der Variablen zu ermöglichen, da die Mehrebenenanalyse wie die Regressionsanalyse zur Interpretation des Achsenabschnitts einen inhaltlich plausiblen Nullpunkt voraussetzt. Es werden dabei zwei Varianten unterschieden, zum einen die Zentrierung am Gesamtmittelwert (*grand-mean-centering*; GMC), zum anderen die Zentrierung am Gruppenmittelwert (*cluster-mean-centering*; CMC; Abramowski, 2020). Bei der Zentrierung am Gesamtmittelwert wird von allen Individualwerten der (identische) Mittelwert der gesamten Stichprobe abgezogen, wodurch eine Verschiebung der Werte auf der x-Achse erreicht wird. Die so zentrierten Individualwerte können als jeweilige Abweichung vom Gesamtmittelwert interpretiert werden (Tausendpfund, 2020). Bei der Zentrierung am Clustermittelwert wird hingegen der jeweilige Mittelwert der Gruppe, dem ein Individuum angehört, abgezogen. Die so entstehenden Werte können interpretiert werden als „um die Clusterzugehörigkeit bereinigt" (Janke, 2006, S. 113).

Die CMC wird von vielen früheren wie heutigen Autor:innen jedoch kritisch gesehen, da ein Modell mit am Gruppenmittelwert zentrierten Variablen nicht mehr mit Modellen mit Rohdaten vergleichbar ist (siehe bspw. Hox et al., 2018; Jäckle & Schärdel, 2017; Kreft, de Leeuw & Aiken, 1995; Paccagnella, 2006). Kelley, Evans,

Lowman und Lykes (2017) schlussfolgern in einer Untersuchung mit fünf großen, qualitativ hochwertigen Datensätzen (u. a. PISA) zum CMC folgendes:

> "We have presented five diverse examples covering a broad range of social science topics and in all of them, consistent with much of the technical literature, group-mean centering the data substantially distorts the results. Our results demonstrate that the distortions are sometimes neither arcane nor small. Instead, […] the distortions induced by group-mean centering in multi-level models can be sufficiently substantial to lead to seriously incorrect interpretations of social processes" (S. 280).

Als Alternative zur Zentrierung kann auch eine Rekodierung der Antwortoptionen zur Erzeugung eines natürlichen Nullpunktes vorgenommen werden (Mayerl & Urban, 2019; siehe auch Kapitel 7.1.1). Diese Möglichkeit bietet sich insbesondere bei Likert-Skalen an, da diese häufig nur aus Konventionsgründen eine Kodierung von 1 und höher und deshalb keinen Nullpunkt haben. Über eine einfach Rekodierung (z. B. bei vier Skalenpunkten von 1 bis 4 zu 0 bis 3) kann ein natürlicher Nullpunkt erzeugt werden.

Da den zu untersuchenden Konstrukten in dieser Arbeit Likert-Skalen zu Grunde liegen und eine inhaltlich sinnvolle Interpretation auch ohne Zentrierung möglich ist, wird eine Rekodierung für die Herstellung eines inhaltlich sinnvollen Nullpunktes vorgenommen und auf eine Zentrierung mit GMC oder CMC verzichtet.

Am besten lässt sich eine MEA schrittweise über vier Modellstufen aufbauen:

1. Nullmodell
2. *Random-Intercept*-Modell
3. *Random-Slope*-Modell
4. *Cross-Level-Interaction*-Modell (Tausendpfund, 2020).

Das Nullmodell, auch leeres Modell genannt, ist das einfachste Modell in der Mehrebenenanalyse und erfüllt vor allem zwei Funktionen. Zum einen dient es dazu, die erklärbare Varianz auf die verschiedenen zu untersuchenden Ebenen aufzuteilen und damit zu bestimmen, ob die Varianz, die über Level-2-Merkmale erklärt werden kann, eine Mehrebenenanalyse anzeigt. Dazu wird zunächst nur die abhängige Variable modelliert, ohne dass erklärende Variablen in das Modell aufgenommen werden. Auf dieser Grundlage kann wie bereits in Kapitel 7.1.3 beschrieben der ICC berechnet

werden. Liegt dieser über einem gewissen Grenzwert (je nach Quelle über .05 oder .10), sollte eine Mehrebenanalyse anstatt einer OLS-Regression gerechnet werden. Zum anderen kann das Nullmodell für die weiteren Schritte als Referenzmodell herangezogen werden (Pötschke, 2020). Da im vorliegenden Fall eine kategoriale abhängige Variable untersucht wird, für die keine ICCs berechnet werden können, wurde der ICC in den Voranalysen mit zwei metrischen Leistungsschätzern, auf denen die kategoriale abhängige Variable basiert, überprüft. Im Rahmen des schrittweisen Aufbaus der Mehrebenenanalyse, die nicht mehr zur Überprüfung der Voraussetzungen, sondern zu Beantwortung der Hypothesen dient, wird für eine Vergleichbarkeit über verschiedene Modellschritte hinweg wieder die kategoriale Variable verwendet. Im Anschluss an das Nullmodell wird in einer MEA das sogenannte *Random-Intercept*-Modell berechnet. Bei diesem werden erklärende Variablen hinzugenommen unter der Annahme, dass sich die Achsenabschnitte (*Intercepte*) zwischen den Klassen unterscheiden, die Effekte (*Slopes*) auf die abhängige Variable zwischen den Klassen jedoch gleich sind. Dabei werden zunächst nur erklärende Variablen der Individualebene in das *Random-Intercept*-Modell aufgenommen und erst in einem zweiten Schritt Merkmale der Kontextebene ergänzt. Die Interpretation der Ergebnisse erfolgt dabei analog zur Interpretation bei der Regressionsanalyse (Pötschke, 2014; Tausendpfund, 2020).

Der dritte Modellschritt erweitert das *Random-Intercept*-Modell um die Annahme, dass nicht nur der Gruppenmittelwert, sondern auch der Effekt der erklärenden Variablen zwischen den Klassen variieren darf. Es wird deshalb als *Random-Slope*-Modell bezeichnet. Im letzten Schritt der Mehrebenenanalyse, dem *Cross-Level-Interaction*-Modell, können zusätzlich noch Wechselwirkungen zwischen den Merkmalen beider Ebenen modelliert werden. Dies geschieht besonders häufig in Form von Moderatoreffekten, bei denen eine Level-2-Variable, den Effekt einer Level-1-Variable auf die abhängige Variable beeinflusst (Pötschke, 2014). Die Mehrebenenanalyse bietet somit die Möglichkeit komplexe Beziehungen zwischen Merkmalen verschiedener Ebenen abzubilden. Jäckle und Schärdel (2017) warnen jedoch davor, immer alle Schritte einer MEA durchzuführen, wenn nicht ausreichend inhaltliche oder statistische Argumente dafür vorliegen:

> „Auch wenn die Flexibilität, welche Random-Slope-Modelle und Cross-Level-Interaktionen bieten, verführerisch sein kann, sollten Forscher stets bedenken, dass jede Erweiterung des Modells durch zusätzliche Parameter immer auch auf Kosten der Modellschätzung geht. Insbesondere bei MEA gilt deshalb, dass

> die Modelle zwar so genau wie nötig, aber eben auch so sparsam wie möglich gehalten sein sollten“ (S. 153).

Wie Abbildung 18 zu entnehmen ist, wird bei Mehrebenenmodellen üblicherweise nur eine sehr begrenzte Anzahl von Variablen bzw. latenten Konstrukten untersucht. Sollen mehrere Merkmale und komplexe Beziehungen zwischen diesen untersucht werden, muss der Mehrebenenansatz mit der Strukturgleichungsmodellierung kombiniert werden.

7.3.3 Zusammenfassung und Schlussfolgerungen für die Analysen

Aufgrund der geclusterten Datenstruktur mit Merkmalen auf Individual- und Klassenebene, der ausreichend hohen Intraklassenkorrelation und der hohen Anzahl latenter Konstrukte auf Individual- und Klassenebene werden für die Überprüfung der Hypothesen Strukturgleichungsmodelle unter Berücksichtigung der Mehrebenenstruktur durchgeführt. Durch diese Verknüpfung beider Ansätze können die Vorteile der beiden Methoden kombiniert und die komplexen Zusammenhänge inhaltlich wie statistisch adäquat abgebildet werden.

Aufgrund der Komplexität des Mehrebenenstrukturgleichungsmodells ist eine Berechnung des vollständigen *Random-Intercept*-Modells mit den vorliegenden technischen Möglichkeiten nur unter Reduktion der Integrationspunkte möglich, was bei der Interpretation der Ergebnisse berücksichtigt werden muss. Da für die Berechnung des *Random-Slopes*-Modells eine weitere starke Reduktion der Integrationspunkte nötig wäre[21], wird bei den Berechnungen das *Random-Intercept*-Modell als finales Modell gewählt. Dies ist auch inhaltlich vertretbar, da so zum einen dem Prinzip der Sparsamkeit nach Jäckle und Schärdel (2017) gefolgt wird. Zum anderen zielt die Forschungsfrage nicht primär auf die Untersuchung der Unterschiede zwischen den

[21] Proberechnungen mit einer variierenden Anzahl an Integrationspunkten zeigten verschiedene Resultate: entweder war eine Berechnung mit den vorliegenden technischen Möglichkeiten nicht machbar oder die Berechnungen wiesen deutlich zu hohe Rechenzeiten auf, um im Rahmen einer Dissertation sinnvoll durchgeführt werden zu können. Eine stärkere Reduktion der Integrationspunkte als alternative Herangehensweise führte in Kombination mit dem Ansatz der Multiplen Imputation zu nicht-sinnvollen Ergebnissen.

befragten Klassen ab, sondern legt den Fokus auf die Erklärung der Zielvariable, was mit dem *Random-Intercept*-Modell ebenfalls geleistet werden kann. Der Fokus in den Analysen liegt somit darauf, ein Strukturgleichungsmodell unter Berücksichtigung der geschachtelten Datenstruktur zu rechnen.

8

Ergebnisse

Im Folgenden werden die Ergebnisse der Analysen zur Beantwortung der Forschungsfrage und der Hypothesen vorgestellt. Dazu werden zunächst in einem ersten Schritt die Skalen bzw. latenten Konstrukte aus TIMSS 2019 wie in Kapitel 7.2 beschrieben überprüft und die Messmodelle für die komplexeren Analysen abgeleitet. Anschließend werden das Nullmodell sowie die Ergebnisse der Vormodelle präsentiert, mit denen getestet wird, für welche Merkmale neben indirekten Pfaden auf die naturwissenschaftliche Leistung auch direkte Pfade im Gesamtmodell modelliert werden. Anhand dieser Vormodelle wird das finale Gesamtmodell zur Beantwortung der Forschungsfrage aufgestellt und dieses anschließend schrittweise unter Hinzunahme der ausgewählten Merkmale aufgebaut und überprüft.

8.1 Prüfung der latenten Konstrukte und Ableitung der Messmodelle

Da die ausgewählten Skalen von TIMSS 2019 wie in Kapitel 6.4.4 beschrieben zum Teil über die Zyklen weiterentwickelt und ergänzt wurden, wird zunächst eine empirische Überprüfung der Merkmale vorgenommen, die als latente Konstrukte in das Modell einfließen. Da es sich bei den Kontrollvariablen um manifeste Variablen handelt, werden diese im Folgenden nicht überprüft.

Da in dieser Arbeit sowohl Schüler- als auch Lehrkräfteaussagen fachspezifisch für den naturwissenschaftlichen Unterricht untersucht werden, wird für eine adäquate Gewichtung bei den Analysen für die vorliegende Arbeit primär das *weight for science teacher data* (SCIWGT) genutzt. Ausnahme bildet die Überprüfung der einzelnen Skalen aus dem Schülerfragebogen im Rahmen der konfirmatorischen Faktorenanalysen, da in diesem Schritt noch keine Verbindung zwischen den einzelnen Konstrukten hergestellt wird. Alternativ wird deshalb das *student house weight* (HOUWGT) genutzt. Beide Gewichte ermöglichen repräsentative Aussagen auf Schülerebene (Beese, Scholz, Jentsch, Jusufi & Schwippert, 2020).

Mit dem Bartlett-Test auf Sphärizität und dem *Kaiser-Meyer-Olkin*-Kriterium (KMO) sind in SPSS zwei Methoden implementiert, mit denen überprüft werden kann, ob die zugrundeliegenden Daten für konfirmatorische Faktorenanalysen (*confirmatory factor analysis*; CFA) geeignet sind. Da der Bartlett-Test wie in Kapitel 7.2.2 beschrieben normalverteilte Daten voraussetzt, die in den zu untersuchenden Konstrukten nicht vorliegen, wurden die Daten mithilfe des *Kaiser-Meyer-Olkin*-Kriteriums

getestet. Bei der Prüfung der acht latenten Konstrukte, die im Mehrebenenstrukturgleichungsmodell aufgenommen werden, ergeben sich Werte für das *Kaiser-Meyer-Olkin*-Kriterium (KMO) zwischen .754 und .924. Damit liegen alle KMO-Werte deutlich über dem kritischen Wert von .50 und können nach Backhaus et al. (2021) und Kaiser und Rice (1974) als mindestens ‚ziemlich gut' (*middling*) bezeichnet werden. Die Daten sind somit für eine Faktorenanalyse geeignet.

Die Ergebnisse für die CFAs der Messmodelle für individuelle Förderung, Klassenführung, konstruktive Unterstützung und Berufszufriedenheit sind in den Tabellen 2 bis 5 dargestellt.[22] Zur Gewichtung der Daten wurde wie beschrieben für die Skalen des Schülerfragebogens das HOUWGT genutzt, für die Skalen des Lehrkräftefragebogens das SCIWGT.

Die vier Messmodelle weisen Werte für Cronbach´s Alpha von mindestens $\alpha = .782$ auf und liegen somit über dem in Kapitel 7.2.1 abgeleiteten Grenzwert von $\alpha = .700$. Darüber hinaus verweisen die *Model-Fit*-Werte RMSEA, CFI und SRMR für das Messmodell zur individuellen Förderung auf einen akzeptablen bis guten *Model-Fit* (siehe Tabelle 2; für die in dieser Arbeit gültigen Grenzwerte siehe Kapitel 7.2.3). Die *Model-Fit*-Werte für die Messmodelle zur Klassenführung (Tabelle 3), zur konstruktiven Unterstützung (Tabelle 4) und zur Berufszufriedenheit (Tabelle 5) zeigen durchgängig einen guten *Model-Fit* an. Cronbach´s Alpha und den *Model-Fit*-Werten zufolge können die Messmodelle der vier latenten Konstrukte in dieser Form verwendet werden.

22 In der Spalte Variablenname kennzeichnet * Variablen, die für eine positive Ausrichtung invertiert wurden. Variablen, die in aggregierter Form aufgenommen wurden, sind mit ** gekennzeichnet. Variablen mit *** sind sowohl invertiert als auch aggregiert.

Tab.: 2 Messmodell individuelle Förderung

Individuelle Förderung (α = .873)

Variablenname	Faktorladung
ASXS22A***	.543
ASXS22B***	.542
ASXS22C***	.745
ASXS22D***	.751
ASXS22E***	.761
ASXS22F***	.809
ASXS22G***	.807

(χ^2=68.634; df=14; RMSEA=.028; CFI=.910; SRMR=.049)

Tab.: 3 Messmodell Klassenführung

Klassenführung (α = .875)

Variablenname	Faktorladung
ASXS27A**	.672
ASXS27B**	.825
ASXS27C**	.828
ASXS27D**	.607
ASXS27E**	.687

(χ^2=22.243; df=5; RMSEA=.026; CFI=.955; SRMR=.039)

Tab.: 4 Messmodell konstruktive Unterstützung

Konstruktive Unterstützung (α = .782)

Variablenname	Faktorladung
ASXS28A***	.462
ASXS28B***	.585
ASXS28C***	.664
ASXS28D***	.725
ASXS28E***	.410

(χ^2=8.574; df=5; RMSEA=.012; CFI=.977; SRMR=.029)

Tab.: 5 Messmodell Berufszufriedenheit

Berufszufriedenheit (α = .875)

Variablenname	Faktorladung
ATBG08A**	.699
ATBG08B**	.668
ATBG08C**	.827
ATBG08D**	.805
ATBG08E**	.821

(χ^2=6.921; *df*=5; RMSEA=.009; CFI=.993; SRMR=.024)

Einige der Faktorladungen in den vier CFAs in den Tabellen 2 bis 5 liegen hingegen deutlich unter dem Wert von .700, ab dem Ladungen als gut eingeschätzt werden. Wie in Kapitel 7.2.1 beschrieben, können Ladungen ab .400 als akzeptabel, mit inhaltlicher Begründung ab .300 als ausreichend angesehen werden. Die Ladungen in den vier Messmodellen weisen somit mindestens eine akzeptable Höhe auf. Da zudem keine inhaltlichen Gründe für den Ausschluss einzelner Items vorliegen, werden die vier Konstrukte individuelle Förderung, Klassenführung, konstruktiven Unterstützung und Berufszufriedenheit in unveränderter Form als Messmodell in den folgenden Analysen verwendet. Damit liegen für diese Messmodelle keine alternativen Modellversionen vor, weshalb das χ^2 und die Freiheitsgrade, die nur im Vergleich mehrerer Modelle sinnvoll interpretiert werden können, nicht herangezogen werden.

Für die vollständige Skala der fachbezogenen Einstellung (Messmodell I, Tabelle 6) zeigt sich hingegen, dass zwar Cronbach´s Alpha eine sehr gute interne Konsistenz aufweist, die *Model-Fit*-Werte für den RMSEA, den CFI und den SRMR jedoch größtenteils einen lediglich akzeptablen Fit des Messmodells anzeigen (siehe Tabelle 7). Da für dieses Messmodell zudem einige Items eher niedrige Ladungen aufweisen und die Skala mit neun Item insgesamt recht umfangreich ist, wird mit Messmodell II zusätzlich eine gekürzte Version berechnet, bei dem die Items mit den niedrigsten Ladungen ausgeschlossen wurden (siehe Tabelle 6). In Messmodell II können bis auf eine Ladung alle Faktorladungen als gut eingestuft werden, die Ladung von ASBS7B verfehlt den Grenzwert von .700 nur knapp. Cronbach´s Alpha ist im Vergleich zu Messmodell I nahezu unverändert, die *Model-Fit*-Werte für den RMSEA, den CFI und den SRMR fallen hingegen deutlich besser aus und deuten durchgängig auf einen

guten *Model-Fit* hin. Mithilfe des nach Satorra-Bentler korrigierten χ^2-Differenztests lässt sich darüber hinaus überprüfen, ob eines der beiden Messmodelle signifikant besser auf die Daten passt. Der korrigierte χ^2-Differenztest für Messmodell I und II zur fachbezogenen Einstellung wird signifikant ($\Delta\chi^2 = 667.0055$; $df = 16$; $p \leq .001$) und deutet somit darauf hin, dass Messmodell II signifikant besser auf die Daten passt als Messmodell I. In den folgenden Analysen wird deshalb Messmodell II genutzt.

Tab.: 6 Messmodell I und II für fachbezogene Einstellung

Fachbezogene Einstellung Modell I ($\alpha = .919$)

Variablenname	**Faktorladung**
ASBS07A*	.832
ASBS07B	.700
ASBS07C	.690
ASBS07D*	.728
ASBS07E*	.904
ASBS07F*	.884
ASBS07G*	.563
ASBS07H*	.549
ASBS07I*	.822

Fachbezogene Einstellung Modell II ($\alpha = .918$)

Variablenname	**Faktorladung**
ASBS07A*	.844
ASBS07B	.680
ASBS07D*	.713
ASBS07E*	.909
ASBS07F*	.889
ASBS07I *	.820

Tab.: 7 Model-Fit-Werte für Messmodell I und II für fachbezogene Einstellung

	χ^2	df	**RMSEA**	**CFI**	**SRMR**
Modell I	755.733	27	.079	.925	.048
Modell II	53.636	9	.034	.993	.011

Die Ergebnisse für die vollständige Skala des fachbezogenen Selbstkonzepts sind in Tabelle 8 dargestellt. Drei der sieben Faktorladungen verfehlen den Grenzwert von .700. Auffällig ist dabei, dass sich die unterschiedlich hohen Ladungen entlang der ursprünglichen Formulierungsrichtung der Items zeigen. Dies lässt vermuten, dass die positive bzw. negative Formulierung der Items Einfluss auf die Beantwortung

der Fragen hat (siehe Kapitel 7.2.3). Deshalb wird alternativ ein zweites Messmodell geprüft, bei dem das fachbezogene Selbstkonzept als zwei separate Facetten modelliert wird. In den geplanten Analysen können diese über einen Faktor zweiter Ordnung, der das fachbezogene Selbstkonzept in der Gesamtheit repräsentiert, in die Modelle aufgenommen werden. Tabelle 8 zeigt, dass die Faktorladungen von Messmodell II durchweg höher ausfallen als in Messmodell I, die *Model-Fit*-Werte sprechen ebenfalls für Messmodell II (siehe Tabelle 9). Während die *Model-Fit*-Werte für Messmodell II durchgängig auf einen guten Fit hindeuten, zeigt für Messmodell I nur der SRMR einen ausreichenden Fit an. Der RMSEA und der CFI hingegen verfehlen die Grenzwerte für einen ausreichenden Fit deutlich. Für Messmodell II spricht auch das Ergebnis des nach Satorra-Bentler korrigierten χ^2-Differenztests ($\Delta\chi^2 = 1184.436$; $df = 1$; $p \leq .001$), weshalb dieses in den folgenden Analysen verwendet wird.

Tab.: 8 Messmodell I und II für fachbezogenes Selbstkonzept

Fachbezogenes Selbstkonzept Modell I ($\alpha = .848$)

Variablenname	**Faktorladung**
ASBS09A*	.600
ASBS09B	.735
ASBS09C	.738
ASBS09D*	.574
ASBS09E*	.453
ASBS09F	.771
ASBS09G	.729

Fachbezogenes Selbstkonzept Modell II

Variablenname	**Faktorladung**
Fachbezogenes Selbstkonzept + ($\alpha = .789$)	
ASBS09A*	.793
ASBS09D*	.774
ASBS09E*	.676
Fachbezogenes Selbstkonzept - ($\alpha = .848$)	
ASBS09B	.750
ASBS09C	.745
ASBS09F	.802
ASBS09G	.751

Tab.: 9 Model-Fit-Werte für Messmodell I und II für fachbezogenes Selbstkonzept

	χ^2	df	**RMSEA**	**CFI**	**SRMR**
Modell I	1012.641	14	.130	.786	.089
Modell II	94.982	13	.039	.984	.025

Die Ergebnisse für die kognitive Aktivierung sind in Tabelle 10 aufgeführt. Für die vollständige Skala liegt Cronbach´s Alpha über dem Grenzwert von .700 ist somit als gut einzustufen. Es zeigen sich allerdings durchgängig niedrige Faktorladungen unter .700, wobei zwei Ladungen mit unter .200 besonders gering ausfallen. Eine weitere Ladung verfehlt den Grenzwert von .400 knapp. Darüber hinaus fallen die *Model-Fit*-Werte für den RMSEA, den CFI und den SRMR sehr unterschiedlich aus (siehe Tabelle 11). Während der RMSEA einen guten *Model-Fit* anzeigt, deutet der SRMR auf einen akzeptablen Fit hin. Der CFI verfehlte hingegen den niedrigsten Grenzwert von .900 deutlich. Da sich für die vollständige Skala zur kognitiven Aktivierung somit uneinheitliche und zum Teil unzureichende Ergebnisse zeigen, wird mit Messmodell II ein Alternativmodell gerechnet, bei dem die Items mit den niedrigsten Ladungen entfernt wurden.

Tab.: 10 Messmodell I und II für kognitive Aktivierung

Kognitive Aktivierung Modell I (α = .754)

Variablenname	**Faktorladung**
ASXS28F***	.500
ASXS28G***	.652
ASXS28H***	.538
ASXS28I***	.579
ASXS28J***	.104
ASXS28K***	.399
ASXS28L***	.193

Kognitive Aktivierung Modell II (α = .708)

Variablenname	**Faktorladung**
ASXS28F***	.524
ASXS28G***	.623
ASXS28H***	.556
ASXS28I***	.594
ASXS28K***	.373

Tab.: 11 Model-Fit-Werte für Messmodell I und II für kognitive Aktivierung

	χ^2	*df*	**RMSEA**	**CFI**	**SRMR**
Modell I	32.424	14	.016	.814	.064
Modell II	9.211	5	.013	.944	.037

Cronbach´s Alpha kann für Messmodell II als gut eingestuft werden. Die Faktorladungen sind wie in Messmodell I eher gering, aber bis auf eine Ausnahme akzeptabel. Das Item mit der niedrigsten Faktorladung (ASXS28K) kann als ausreichend eingestuft werden. Da ASXS28K eine Facette der kognitiven Aktivierung abbildet, die über die anderen Items in Messmodell II nicht abgedeckt wird, stützen inhaltliche Gründe dessen Beibehaltung. Darüber hinaus zeigt Messmodell II für den RMSEA, den CFI und den SRMR deutlich bessere und homogenere Werte als Messmodell I, die als akzeptabel bis gut eingestuft werden können. Auch der nach Satorra-Bentler korrigierte χ^2-Differenztest zeigt an, dass das Messmodell II signifikant besser auf die Daten passt als Messmodell I ($\Delta\chi^2 = 23.391$; $df = 9$; $p \leq .010$). Für kognitive Aktivierung wird somit Messmodell II in den folgenden Analysen genutzt.

Tabelle 12 zeigt die Ergebnisse für die CFA mit der vollständigen Skala zur Lehrkräfteselbstwirksamkeit. Es zeigt sich, dass Cronbach´s Alpha als gut eingestuft werden kann, die Faktorladungen fallen jedoch eher niedrig aus, wobei es keine Ausreißer gibt, die unter .400 oder .300 liegen. Die *Model-Fit*-Werte in Tabelle 13 deuten auf einen akzeptablen Fit hin, der RMSEA zeigt sogar einen guten *Model-Fit* an. Bei der inhaltlichen Betrachtung der einzelnen Items zur Lehrkräfteselbstwirksamkeit fällt jedoch auf, dass sich die Items ATXS17B und ATXS17J auf sehr konkrete Unterrichtsmethoden und deren Einsatz beziehen. Die übrigen Items der Skala hingegen umfassen eher allgemeine unterrichtsbezogene Aspekte, die nicht davon abhängen, ob die Lehrkräfte bestimmte Methoden generell nutzen oder nicht. Es wird deshalb trotz der akzeptablen Ergebnisse für Messmodell I ein zweites Messmodell ohne die beiden benannten Items berechnet und mit Messmodell I verglichen.

Messmodell II weist einen ähnlich hohen Wert für Cronbach´s Alpha auf wie Messmodell I, die Faktorladungen liegen etwas höher, sind in der Gesamtheit aber ähnlich wie in Messmodell I (siehe Tabelle 12). Ähnliches gilt auch für die *Model-Fit*-Werte. Der RMSEA ist nahezu unverändert, der SRMR fällt nominell besser aus und der CFI ist höher als in Messmodell I. Der nach Satorra-Bentler korrigierte χ^2-Differenztest zeigt, dass Messmodell II signifikant besser auf die Daten passt als Messmodell I ($\Delta\chi^2 = 36.475$; $df = 15$; $p \leq .010$). Da sich in den Faktorladungen und den Werten für Cronbach´s Alpha kaum Unterschiede zwischen beiden Messmodellen zeigen, die *Model-Fit*-Werte und der χ^2-Differenztest jedoch eine höhere Passung von Messmodell II anzeigen und die inhaltliche Ausrichtung der Skala in diesem Modell angemessener ist, wird in den folgenden Analysen Messmodell II genutzt.

Tab.: 12 Messmodell I und II für Lehrkräfteselbstwirksamkeit

Lehrkräfteselbstwirksamkeit Modell I (α = .873)

Variablenname	Faktorladung
ATXS17A**	.644
ATXS17B**	.600
ATXS17C**	.747
ATXS17D**	.626
ATXS17E**	.691
ATXS17F**	.641
ATXS17G**	.576
ATXS17H**	.652
ATXS17I**	.668
ATXS17J**	.622

Lehrkräfteselbstwirksamkeit Modell II (α = .859)

Variablenname	Faktorladung
ATXS17A**	.623
ATXS17C**	.715
ATXS17D**	.647
ATXS17E**	.714
ATXS17F**	.654
ATXS17G**	.595
ATXS17H**	.670
ATXS17I**	.654

Tab.: 13 Model-Fit-Werte für Messmodell I und II für Lehrkräfteselbstwirksamkeit

	χ^2	df	RMSEA	CFI	SRMR
Modell I	87.277	35	.021	.900	.055
Modell II	50.762	20	.022	.923	.049

Insgesamt zeigt die Überprüfung der Messmodelle mit konfirmatorischen Faktorenanalysen für die Skalen konstruktive Unterstützung, Klassenführung, individuelle Förderung sowie Berufszufriedenheit gute Faktorladungen und Werte für Cronbach´s Alpha sowie die *Model-Fit*-Kriterien, sodass sie in unveränderter Form genutzt werden. Die Skalen fachbezogene Einstellung, kognitiver Aktivierung und Lehrkräfteselbstwirksamkeit hingegen mussten aufgrund inhaltlicher Aspekte sowie unzureichender Faktorladungen und Werte für Cronbach´s Alpha bzw. *Model-Fit*-Werte leicht modifiziert werden, sodass gekürzte Versionen der Skalen in den Datensätzen für die anschließenden Analysen genutzt werden. Stärkere Anpassungen waren für die Skala fachbezogenes Selbstkonzept nötig. Diese wird über einen Faktor zweiter Ordnung, auf den die Subfacetten positives fachbezogenes Selbstkonzept und negatives fachbezogenes Selbstkonzept laden, in die Modelle aufgenommen. Nach den vor-

genommenen Anpassungen weisen alle Items der acht latenten Faktoren mindestens ausreichende, zumeist jedoch akzeptable bis gute Faktorladungen, gute Werte für Cronbach´s Alpha sowie akzeptable bis gute *Model-Fit*-Werte auf.

In den Modellen, die im Folgenden berechnet werden, werden diese Messmodelle unverändert genutzt, für eine bessere Übersichtlichkeit jedoch ohne die zugrundeliegenden Einzelitems und Faktorladungen aufgeführt. Für das fachbezogene Selbstkonzept wird ausschließliche der Faktor zweiter Ordnung in den Modellen dargestellt. Ausführliche Modellinformationen zu den jeweiligen Teil- und Gesamtmodellen sowie zu den Ladungen der zwei Subfacetten des fachbezogenen Selbstkonzepts finden sich im Anhang. Bevor die Vor- und Hauptmodelle zur Beantwortung der Forschungsfrage berechnet werden, wird im nächsten Schritt zunächst ein zweites Nullmodell ohne Prädiktoren als Referenzmodell für die nachfolgenden Analysen gerechnet.

8.2 Berechnung des Nullmodells als Referenzmodell

Wie in Kapitel 7.3.2 erläutert, können mit dem Nullmodell vor allem zwei Informationen gewonnen werden: zum einen zeigen die Intraklassenkorrelationen, ob anstatt einer einfachen OLS-Regression Mehrebenenanalysen durchgeführt werden sollten, zum anderen bietet es Referenzwerte zur Beurteilung alternativer Modellversionen. Dabei konnte mit dem ersten Nullmodell in Kapitel 7.1.3 gezeigt werden, dass die Intraklassenkorrelationen für die vorliegenden Analysen eine Mehrebenenuntersuchung rechtfertigen. Dabei mussten zur Bestimmung der Intraklassenkorrelationen aus methodischen Gründen im Nullmodell zwei alternative Formen der abhängigen Variable (erreichte Punkte im Kompetenztest und Kompetenzstufenzugehörigkeit) genutzt werden. Dies bringt mit sich, dass das in Kapitel 7.1.3 berechnete Nullmodell nicht als Referenz für die folgenden Analysen dienen kann. In diesem Kapitel wird deshalb ein zweites Nullmodell als Referenzmodell mit der rekodierten, multinomialen abhängigen Variable berechnet, mit der in den nachfolgenden Modellen die angestrebten Übergangswahrscheinlichkeiten generiert werden können.

Da die multinomial-logistische abhängige Variable in das Modell eingebunden ist, können wie in Kapitel 7.2.3 erläutert ausschließlich Informationskriterien wie die Devianz, der AIC und der BIC berechnet werden. Dadurch ist es nicht möglich, unabhängige Aussagen über die Güte der Modelle zu treffen, sondern ausschließlich relative Beurteilungen mithilfe von Referenzmodellen vorzunehmen. Somit lassen sich

mit den in Tabelle 14 dargestellten Informationskriterien noch keine Aussagen über die Güte des zweiten Nullmodells (NM) treffen. Stattdessen stellen sie die Referenzwerte für die nachfolgenden Modelle dar. Die Devianz dieses Referenzmodells beträgt 116491.382, der AIC 116635.382 und der BIC 117078.939.

Die Vor- und Hauptmodelle in den folgenden Kapiteln werden als gut eingestuft, wenn die Werte des Alternativmodells für alle drei *Model-Fit*-Kriterien niedriger ausfallen als die des Nullmodells und somit auf einen besseren *Model-Fit* das Alternativmodells hindeuten.

Tab.: 14 Model-Fit-Werte Nullmodell II

Modell	Devianz	AIC	BIC
NM	116491.382	116635.382	117078.939

8.3 Berechnung der Vormodelle

Da sich in der bisherigen Forschung neben vor allem indirekten auch vermehrt direkte Effekte der verschiedenen Prädiktoren auf Schulleistungen gezeigt haben, werden zunächst die möglichen direkten Effekte der Lehrkräfte- und Unterrichtsmerkmalen auf die Übergangswahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen, überprüft. Dazu werden zwei Modelle gerechnet, die die Merkmale der Bereiche Unterricht und Lehrkraft getrennt voneinander betrachten. Für Merkmale, bei denen sich in den Vormodellen direkte Effekte auf die Leistung zeigen, werden diese zusätzlich zu den indirekten Effekten im Gesamtmodell modelliert.

Eine Besonderheit muss bei der Interpretation der direkten Pfade berücksichtigt werden: Üblicherweise werden im Rahmen (multinomial) logistischer Regressionen *odds ratios* zur Quantifizierung des Einflusses von Prädiktoren berichtet. Bei der Verwendung von multipler Imputation werden in Mplus standardmäßig lediglich die Regressionskoeffizienten (β) ausgegeben. Die *odds ratios* müssen deshalb nachträglich mit folgender Formel berechnet werden (Best & Wolf, 2012):

$$odds\,ratios = e^{\beta}$$

Interpretiert werden können die *odds ratios*, die Werte zwischen 0 und +∞ annehmen können, folgendermaßen: *odds ratios*, die zwischen 0 und 1 liegen, verringern die Wahrscheinlichkeit, dass ein anvisiertes Ereignis eintritt, *odds ratios* über 1 erhöhen

die Wahrscheinlichkeit. Je dichter die *odds ratios* dabei am Wert 1 liegen, desto weniger bedeutsam sind die betrachteten Prädiktoren.

Der aus methodischen Gründen notwendige Einsatz einer multinomial logistischen Zielvariable (siehe Kapitel 7.1.1) hat darüber hinaus zur Folge, dass nicht nur die Regressionskoeffizienten für den Übergang von Kompetenzstufe IV auf Kompetenzstufe V ausgegeben werden, sondern auch die für den Übergang von Kompetenzstufe IV auf die zusammengefassten Kompetenzstufen I bis III. Da diese jedoch nicht zur Beantwortung der Forschungsfrage beitragen, werden keine entsprechenden *odds ratios* berechnet und berichtet.

Im Folgenden werden die Vormodelle zur Überprüfung der direkten Pfade und zur Ableitung des Gesamtmodells vorgestellt. Begonnen wird dabei mit den Merkmalen der Lehrkräfte.

8.3.1 Vormodell I: Lehrkräftemerkmale

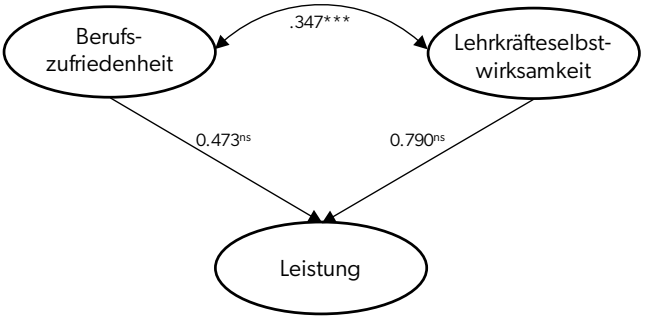
Mit dem ersten Vormodell wird untersucht, ob die Lehrkräfteselbstwirksamkeit und die Berufszufriedenheit direkt die Wahrscheinlichkeit beeinflussen, dass Schüler:innen anstatt der Kompetenzstufe IV die Kompetenzstufe V erreichen (Hypothese 15 und 16). Das Modell weist eine Devianz von 10594.210, einen AIC von 10686.209 und einen BIC von 10969.593 auf (siehe Tabelle 15). Da alle drei Werte deutlich kleiner ausfallen als die *Model-Fit*-Werte des Nullmodells, ist das Vormodell I besser zur Beschreibung der Daten geeignet und kann somit als Grundlage für die Überprüfung der direkten Zusammenhänge genutzt werden.

Tab.: 15 Model-Fit-Werte Vormodell I

Modell	**Devianz**	**AIC**	**BIC**
Nullmodell (NM)	116491.382	116635.382	117078.939
Vormodell I (VM I)	10594.210	10686.209	10969.593

In Abbildung 19 sind die Ergebnisse des Vormodells I (VM I) dargestellt. Die *odds ratios* für das Merkmal Lehrkräfteselbstwirksamkeit liegen bei 0.790. Die *odds ratios* sind damit kleiner als Eins und deuten entsprechend auf einen negativen Effekt der Lehrkräfteselbstwirksamkeit auf die Übergangswahrscheinlichkeit hin. Für die Berufs-

zufriedenheit zeigen sich *odds ratios* von 0.473, womit sie ebenfalls unter Eins liegen. Da sie weniger dicht an Eins liegen als die *odds ratios* der Lehrkräfteselbstwirksamkeit, ist die Berufszufriedenheit zwar der bedeutsamere Prädiktor, allerdings ist auch dieser Effekt nicht signifikant. Die Korrelation beider Prädiktoren ist signifikant und liegt bei $r = .347$, was nach Cohen (1988) einem mittleren Zusammenhang entspricht. Da beide Prädiktoren keinen direkten signifikanten Effekt auf die Übergangswahrscheinlichkeit haben, müssen die Hypothesen 15 und 16 verworfen werden. Entsprechend werden im Gesamtmodell ausschließlich indirekte Zusammenhänge zwischen Lehrkräftemerkmalen und Leistung modelliert.



*** = $p \leq .001$; ** = $p \leq .01$; * = $p \leq .05$; ns = nicht signifikant
Anmerkung: für logistische Regressionen wird ausschließlich der Wert für den Übergang von KS IV auf KS V angegeben

Abb.: 19 Ergebnisse Vormodell I

8.3.2 Vormodell II: Unterrichtsmerkmale

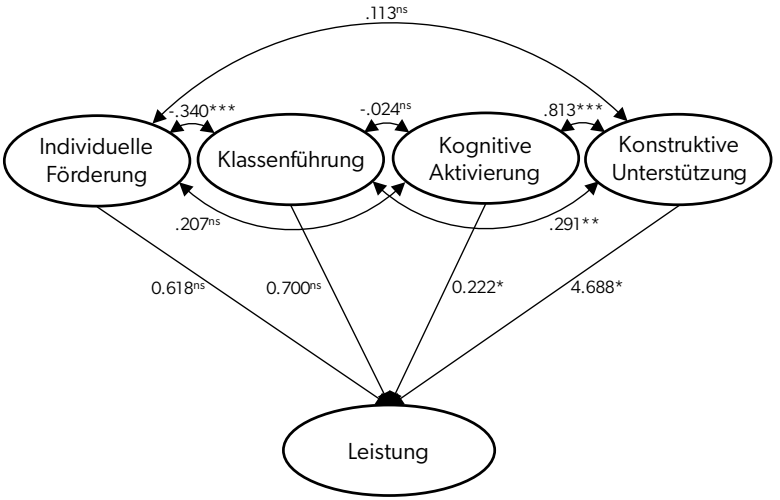
Im zweiten Vormodell werden die direkten Zusammenhänge zwischen den Merkmalen des Unterrichts und der naturwissenschaftlichen Leistung überprüft (Hypothese 3 bis 6). In Tabelle 16 sind die Modell-Güte-Kriterien für Vormodell II (VM II) dargestellt. Mit einer Devianz von 13810.708, einem AIC von 13974.709 und einem BIC von 14479.871 liegen auch für das zweite Vormodell alle drei Gütekriterien deutlich unter den Werten des Nullmodells, sodass dieses dem Nullmodell vorzuziehen ist.

Tab.: 16 Model-Fit-Werte Vormodell II

Modell	Devianz	AIC	BIC
Nullmodell (NM)	116491.382	116635.382	117078.939
Vormodell II (VM II)	13810.708	13974.709	14479.871

Die Ergebnisse des zweiten Vormodells können der Abbildung 20 entnommen werden. Für die konstruktive Unterstützung zeigen sich signifikante *odds ratios* von 4.688. Da sie größer als Eins sind, deuten sie auf einen positiven Effekt der konstruktiven Unterstützung auf die Leistung hin. Ebenfalls signifikante *odds ratios* ergeben sich für die kognitive Aktivierung. Diese liegen bei 0.222 und zeigen somit einen negativen Effekt auf die Übergangswahrscheinlichkeit an. Für die Merkmale Klassenführung (*odds ratios* = 0.700) und individuelle Förderung (*odds ratios* = 0.618) zeigen sich hingegen keine signifikanten Effekte. Hohe signifikante Korrelationen ergeben sich zwischen den Merkmalen konstruktive Unterstützung und kognitive Aktivierung (r = .813), schwache bis mittlere signifikante Korrelationen zwischen konstruktive Unterstützung und Klassenführung (r = .291) sowie Klassenführung und individueller Förderung (r = -.340).

Da sich in Vormodell II für die konstruktive Unterstützung und die kognitive Aktivierung signifikante direkte Effekte auf die Übergangswahrscheinlichkeit zeigen, werden für die konstruktive Unterstützung und die kognitive Aktivierung im Gesamtmodell zusätzlich zu indirekten Effekten auch direkte Effekte modelliert. Allerdings kann nur Hypothese 3 auf Basis dieses Vormodells vorläufig angenommen werden. Hypothese 4 muss hingegen vorläufig verworfen werden, da sich in Vormodell II ein negativer und somit nicht der erwartete, sondern der entgegengesetzte Effekt der kognitiven Aktivierung auf die Leistung zeigt. Eine abschließende Prüfung dieser beiden Hypothesen erfolgt in Hauptmodell IV. Für die Klassenführung und die individuelle Förderung zeigen sich hingegen keine direkten Effekte, sodass Hypothese 5 und 6 verworfen werden müssen. Im Gesamtmodell werden deshalb ausschließlich indirekte Effekte dieser beiden Merkmale auf die Leistung angenommen.



*** = $p \leq .001$; ** = $p \leq .01$; * = $p \leq .05$; ns = nicht signifikant
Anmerkung: für logistische Regressionen wird ausschließlich der Wert für den Übergang von KS IV auf KS V angegeben

Abb.: 20 Ergebnisse Vormodell II

8.4 Ableitung und Berechnung des Gesamtmodells

Mit den Ergebnissen aus der Überprüfung der latenten Konstrukte, der Berechnung des Nullmodells und den Befunden aus den Vormodellen lässt sich das Gesamtmodell zur Bearbeitung der Forschungsfrage dieser Arbeit aufstellen (siehe Abbildung 21). Die latenten Konstrukte werden in der Form, wie sie in Kapitel 8.1 abgeleitet wurden, in die Modelle aufgenommen. Das zweite Nullmodell aus Kapitel 8.2 bildet das Referenzmodell für die folgenden Modelle. Diese werden – anders als im vorherigen Kapitel – auch miteinander verglichen, um das geeignetste Modell zur Beschreibung der Daten zu identifizieren. Aus den Vormodellen ergibt sich darüber hinaus, dass im Gesamtmodell neben indirekten Effekten von Lehrkräfte- und Unterrichtsmerkmalen zusätzlich direkte Effekte von der konstruktiven Unterstützung und der kognitiven Aktivierung auf die Leistung modelliert werden. Die direkten Effekte auf Leistung stellen dabei multinomial logistische Regressionen dar, Effekte, die nicht direkt auf Leistung bezo-

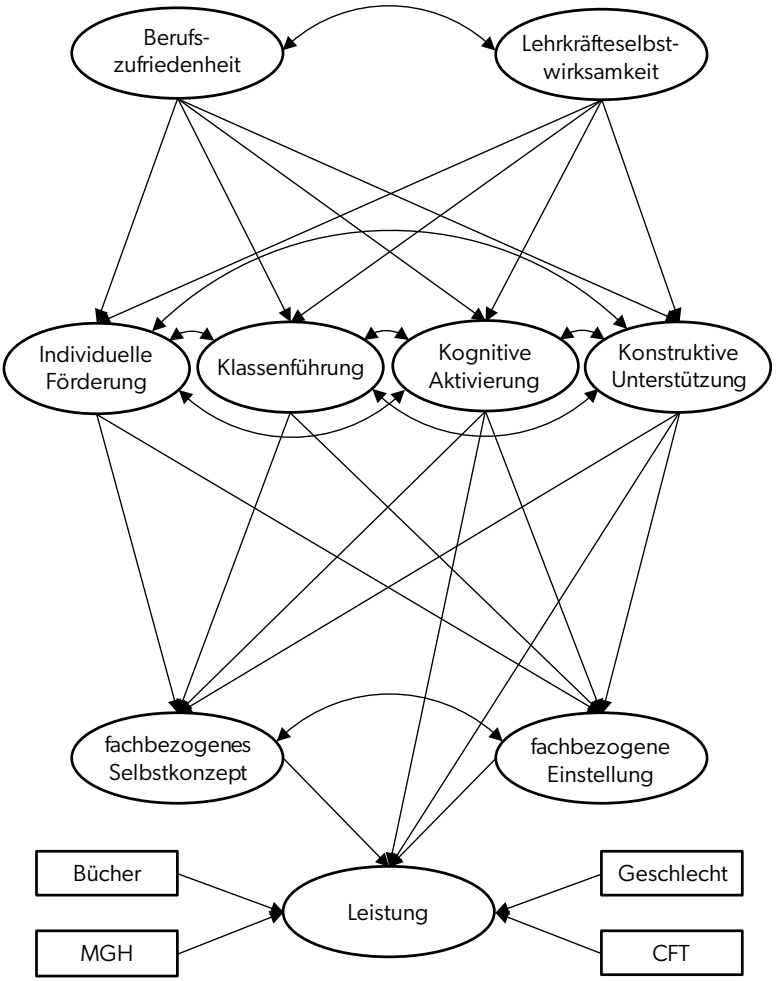


Abb.: 21 Abgeleitetes Gesamtmodell

gen sind, sind lineare Regressionen. Zusätzlich werden wie in Kapitel 5 beschrieben Korrelationen zwischen den Merkmalen der einzelnen Bereiche zugelassen.

Das Gesamtmodell aus Abbildung 21 wird im Folgenden schrittweise aufgebaut. Im ersten Schritt (Hauptmodell I) werden zunächst die Leistung und die Schülermerkmale analysiert. In einem zweiten Schritt (Hauptmodell II) werden dann zusätzlich die Unterrichtsmerkmale mit allen Zusammenhängen in das Modell aufgenommen. Im dritten Schritt (Hauptmodell III) werden die Lehrkräftemerkmale ergänzt. Im letzten Schritt (Hauptmodell IV) wird erneut Hauptmodell III, aber unter Kontrolle relevanter Hintergrundvariablen, berechnet. Hauptmodell IV stellt damit das vollständige Gesamtmodell dar. Dabei wird die zugrundeliegende Mehrebenenstruktur in den einzelnen berechneten Modellen berücksichtigt, auf die Berechnung von *Random-Slope-* oder *Crosslevel-Interaction*-Modelle wird jedoch verzichtet (siehe Kapitel 7.3.3). Eine ausführliche Einordnung der Ergebnisse erfolgt anschließend in Kapitel 9.

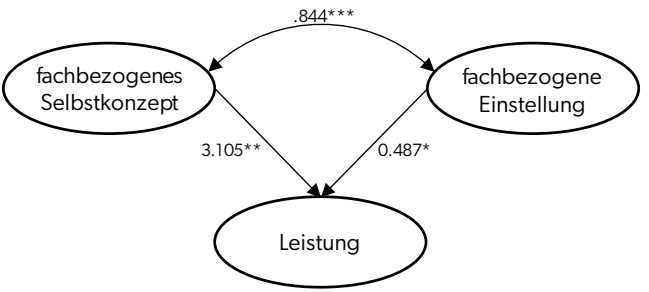
8.4.1 Hauptmodell I: Schülermerkmale

Mit dem ersten Schritt zur Berechnung des Gesamtmodells, dem Hauptmodell I (HM I), wird der Zusammenhang des fachbezogenen Selbstkonzeptes und der fachbezogenen Einstellung mit der Leistung untersucht. Die verfügbaren Modell-Güte-Kriterien für die Hauptmodelle sind wie in den Vormodellen die Devianz, der AIC und der BIC. Für Hauptmodell I ergibt sich eine Devianz von 83859.790, ein AIC von 83955.789 und ein BIC von 84251.494. Im Vergleich zum Nullmodell fallen alle drei Kriterien niedriger aus und verweisen somit auf einen besseren Fit von Hauptmodell I (siehe Tabelle 17).

Tab.: 17 Model-Fit-Werte Hauptmodell I im Vergleich zu vorherigen Modellschritten

Modell	**Devianz**	**AIC**	**BIC**
Nullmodell (NM)	116491.382	116635.382	117078.939
Hauptmodell I (HM I)	83859.790	83955.789	84251.494

Wie Abbildung 22 entnommen werden kann, zeigen sich für das fachbezogene Selbstkonzept signifikante *odds ratios* von 3.105. Steigt das fachbezogene Selbstkonzept um eine Einheit an, erhöht sich somit die Wahrscheinlichkeit, Leistungen auf KS V



*** = $p \leq .001$; ** = $p \leq .01$; * = $p \leq .05$; ns = nicht signifikant
Anmerkung: für logistische Regressionen wird ausschließlich der Wert für den Übergang von KS IV auf KS V angegeben

Abb.: 22 Ergebnisse Hauptmodell I

anstatt von KS IV zu erbringen, um ca. das 3-Fache. Für die fachbezogene Einstellung zeigen sich hingegen signifikante *odds ratios* von 0.487, die auf einen negativen Effekt hindeuten. Das fachbezogene Selbstkonzept ist somit ein bedeutsamerer individueller Prädiktor als die fachbezogene Einstellung. Das fachbezogene Selbstkonzept und die fachbezogene Einstellung korrelieren signifikant in Höhe von $r = .844$ miteinander, was einem starken Zusammenhang entspricht. Nach den Ergebnissen von Hauptmodell I kann die Hypothese 1, dass das fachbezogene Selbstkonzept die Übergangswahrscheinlichkeit positiv beeinflusst, vorläufig angenommen werden. Hypothese 2 muss hingegen vorläufig verworfen werden, da sich in Vormodell I ein negativer Effekt der fachbezogenen Einstellung zeigt. Eine abschließende Überprüfung dieser beiden Hypothesen erfolgt mit der Berechnung von Hauptmodell IV.

8.4.2 Hauptmodell II: Schüler- und Unterrichtsmerkmale

Im zweiten Schritt zur Berechnung des Gesamtmodells werden zusätzlich zu den individuellen Schülermerkmalen die Unterrichtsmerkmale konstruktive Unterstützung, kognitive Aktivierung, Klassenführung und individuelle Förderung in das Modell aufgenommen. Die Modell-Güte-Kriterien in Tabelle 18 verweisen für das Hauptmodell II (HM II) zum einen auf einen besseren Fit gegenüber dem Nullmodell, zum

anderen durchgängig auf einen etwas schlechteren Fit gegenüber dem Hauptmodell I (Devianz = 91638.446; AIC = 91912.446; BIC = 92756.437).

Tab.: 18 Model-Fit-Werte Hauptmodell II im Vergleich zu vorherigen Modellschritten

Modell	**Devianz**	**AIC**	**BIC**
Nullmodell (NM)	116491.382	116635.382	117078.939
Hauptmodell I (HM I)	83859.790	83955.789	84251.494
Hauptmodell II (HM II)	91638.446	91912.446	92756.437

Die Ergebnisse für Hauptmodell II sind in Abbildung 23 dargestellt. Wie bereits in Hauptmodell I zeigen sich für die fachbezogene Einstellung und das fachbezogene Selbstkonzept signifikante *odds ratios*. Mit *odds ratios* von 0.455 für die fachbezogene Einstellung und 3.629 für das fachbezogene Selbstkonzept ergibt sich unter Hinzunahme von Unterrichtsmerkmalen für die fachbezogene Einstellung eine minimale Abnahme der *odds ratios*, für das fachbezogene Selbstkonzept eine leichte Zunahme der *odds ratios*. Die Korrelation beider Merkmale bleibt nahezu unverändert. Hauptmodell II spricht somit ebenfalls für eine vorläufige Annahme der Hypothese 1 und eine vorläufige Ablehnung der Hypothese 2.

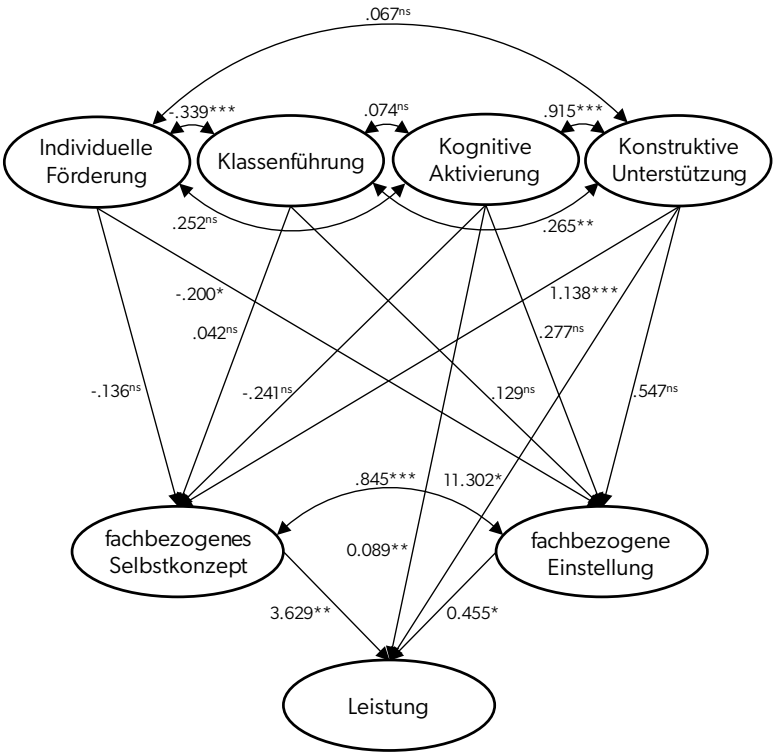
Für die modellierten direkten Effekte der Unterrichtsmerkmale konstruktive Unterstützung und kognitive Aktivierung ergeben sich ebenfalls signifikante *odds ratios*. Für konstruktive Unterstützung betragen diese 11.302. Nehmen Schüler:innen also eine um eine Einheit höhere konstruktive Unterstützung im naturwissenschaftlichen Sachunterricht wahr, steigt ihre Wahrscheinlichkeit Leistungen auf KS V anstatt KS IV zu erbringen um ca. das 11-Fache. Die *odds ratios* für kognitive Aktivierung hingegen sind mit 0.089 deutlich kleiner als Eins, was auf einen bedeutsamen negativen Effekt hindeutet.

Die gleichzeitig modellierten indirekten Effekte von kognitiver Aktivierung über individuelle Schülermerkmale auf Leistung zeigen keine durchgängigen signifikanten Pfade. Zwar sind sowohl die fachbezogene Einstellung als auch das fachbezogene Selbstkonzept signifikante Prädiktoren von Leistung, für beide Merkmale zeigen sich jedoch in den linearen Regressionen von kognitiver Aktivierung keine signifikanten Regressionskoeffizienten (fachbezogene Einstellung: $\beta = .277$; fachbezogenes Selbstkonzept: $\beta = -.241$). Für kognitive Aktivierung ergibt sich somit in diesem Modell aus-

schließlich ein direkter Effekt. Konstruktive Unterstützung hingegen weist zusätzlich zum direkten Effekt auch einen indirekten Effekt auf die Leistung vermittelt über das fachbezogene Selbstkonzept auf (β = 1.138).[23] Der Zusammenhang zwischen konstruktiver Unterstützung und fachbezogener Einstellung ist hingegen nicht signifikant (β = .547). Für die konstruktive Unterstützung zeigt sich somit sowohl ein direkter Effekt als auch ein vermittelter Effekt über das fachbezogene Selbstkonzept auf die Leistung. Hauptmodell II spricht somit wie Hauptmodell I für eine vorläufige Annahme von Hypothese 3, dass es einen direkten positiven Effekt der konstruktiven Unterstützung auf die Leistung gibt und die Ablehnung von Hypothese 4 (positiver direkter Effekt der kognitiven Aktivierung auf die Leistung). Darüber hinaus ergibt sich für die Hypothesen zu den Zusammenhängen zwischen konstruktiver Unterstützung bzw. kognitiver Aktivierung und dem fachbezogenen Selbstkonzept bzw. der fachbezogenen Einstellung, dass die Hypothese 8 zum Zusammenhang von konstruktiver Unterstützung und fachbezogenem Selbstkonzept vorläufig angenommen werden kann, die Hypothesen 7, 9 und 10 jedoch vorläufig abgelehnt werden müssen. Auch hier erfolgt die endgültige Überprüfung im Rahmen von Hauptmodell IV.

Für die Merkmale Klassenführung und individuelle Förderung wurden auf Basis der Ergebnisse der Vormodelle ausschließlich indirekte Effekte modelliert. Für Klassenführung wird dabei weder der Zusammenhang mit der fachbezogenen Einstellung (β = .129) noch der mit dem fachbezogenen Selbstkonzept (β = .042) signifikant. Für individuelle Förderung hingegen ergibt sich ein signifikanter linearer Zusammenhang mit der fachbezogenen Einstellung von β = -.200, während der Zusammenhang mit dem fachbezogenen Selbstkonzept nicht signifikant wird (β = -.136). Die Ergebnisse verweisen somit auf einen indirekten Effekt der individuellen Förderung auf die Leistung, der über die fachbezogene Einstellung vermittelt wird. Für die Unterrichtsmerkmale zeigen sich hohe signifikante Korrelationen zwischen der konstruktiven Unterstützung und der kognitiven Aktivierung (r = .915), schwache bis mittlere signifikante Korrelationen zwischen der konstruktiven Unterstützung und der Klassenführung (r = .265) sowie zwischen der Klassenführung und der individuellen Förderung

[23] Üblicherweise nehmen lineare Regressionskoeffizienten Werte zwischen -1 und 1 an. Werte, die außerhalb dieses Bereiches liegen, sind jedoch rechnerisch ebenfalls möglich, können dann aber ein Hinweis auf Multikollinearitätsprobleme sein (Jöreskog, 1999). Da die Daten vorab auf Multikollinearität geprüft wurden, kann dies für die vorliegenden Analysen ausgeschossen werden. Nach Jöreskog (1999) können ungewöhnlich hohe oder niedrige Regressionskoeffizienten in diesem Fall als unproblematisch angesehen werden.



*** = $p \leq .001$; ** = $p \leq .01$; * = $p \leq .05$; ns = nicht signifikant
Anmerkung: für logistische Regressionen wird ausschließlich der Wert für den Übergang von KS IV auf KS V angegeben

Abb.: 23 Ergebnisse Hauptmodell II

(r = -.339). Die übrigen Korrelationen werden nicht signifikant. Für die Hypothesen 11 bis 14 bedeutet dies folgendes: Die Hypothesen 11 und 12, die Zusammenhänge der Klassenführung mit den individuellen Schülermerkmalen formulieren, sowie Hypothese 14 (individuelle Förderung und fachbezogenes Selbstkonzept) müssen vorläufig verworfen werden. Die Hypothese 13 zum Zusammenhang zwischen individueller Förderung und der fachbezogenen Einstellung kann vorläufig angenommen werden.

8.4.3 Hauptmodell III: Schüler-, Unterrichts- und Lehrkräftemerkmale

Mit dem dritten Schritt zur Berechnung des Gesamtmodells werden neben den individuellen Merkmalen und den Merkmalen des Unterrichts auch Lehrkräftemerkmale in das Hauptmodell III mitaufgenommen. In Tabelle 19 sind die Modell-Güte-Kriterien für Hauptmodell III im Vergleich zum Nullmodel und den Hauptmodellen I und II dargestellt. Es zeigt sich, dass alle drei Kriterien für das Hauptmodell III (Devianz = 96450.264; AIC = 96820.265; BIC = 97959.961) gegenüber dem Nullmodell eine bessere Passung zu den Daten anzeigen. Im Vergleich zu den beiden anderen Hauptmodellen HM I und HM II fällt der *Model-Fit* von Hauptmodell III jedoch schlechter aus.

Tab.: 19 Model-Fit-Werte Hauptmodell III im Vergleich zu vorherigen Modellschritten

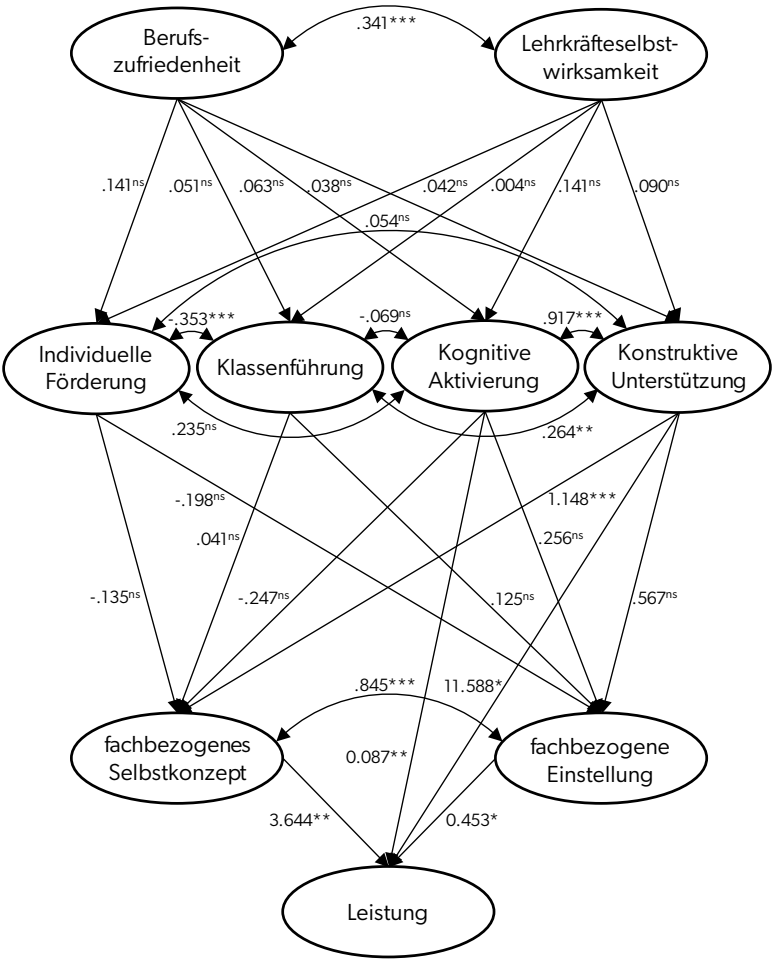
Modell	**Devianz**	**AIC**	**BIC**
Nullmodell (NM)	116491.382	116635.382	117078.939
Hauptmodell I (HM I)	83859.790	83955.789	84251.494
Hauptmodell II (HM II)	91638.446	91912.446	92756.437
Hauptmodell III (HM III)	96450.264	96820.265	97959.961

Abbildung 24 zeigt die Ergebnisse von Hauptmodell III. Insgesamt ergeben sich trotz der neu aufgenommenen Merkmale nur wenig Veränderungen im Vergleich zu Hauptmodell II. So werden auch in diesem dritten Schritt die *odds ratios* für die individuellen Schülermerkmale signifikant. Nahezu unverändert zu Hauptmodell II liegen die *odds ratios* für die fachbezogene Einstellung bei 0.453, für das fachbezogene Selbstkonzept bei 3.644. Die Korrelation liegt erneut bei $r = .845$ und kann somit als hoch eingestuft werden.

Ebenso haben sich auch die *odds ratios* der konstruktiven Unterstützung und der kognitiven Aktivierung im Vergleich zum Hauptmodell II kaum verändert. Die *odds ratios* für konstruktive Unterstützung sind unter Hinzunahme der Lehrkräftemerkmale leicht auf 11.588 angestiegen. Die *odds ratios* der kognitiven Aktivierung sind mit 0.087 hingegen nahezu unverändert. Gleiches gilt für die linearen Zusammenhänge von Klassenführung und individueller Förderung mit fachbezogener Einstellung

und fachbezogenem Selbstkonzept. Für beide Merkmale ergeben sich dabei minimal niedrigere Zusammenhänge, wobei diese wie in Hauptmodell II für Klassenführung nicht signifikant (fachbezogene Einstellung: β = .125; fachbezogenes Selbstkonzept: β = .041) und für individuelle Förderung nur für die fachbezogene Einstellung signifikant ausfallen (fachbezogene Einstellung: β = -.198; fachbezogenes Selbstkonzept: β = -.135). Ein ähnliches Bild ergibt sich auch für die Korrelationen zwischen den Unterrichtsmerkmalen. Die Korrelationen zwischen konstruktiver Unterstützung und kognitiver Aktivierung (r = .917) sowie zwischen konstruktiver Unterstützung und Klassenführung (r = .264) bleiben nahezu unverändert, eine leicht höhere negative Korrelation zeigt sich zwischen den Merkmalen Klassenführung und individuelle Förderung (r = -.353). Die Bedeutung dieser Ergebnisse für die Hypothesen bleibt im Vergleich zu Hauptmodell II somit ebenfalls unverändert.

Für die Lehrkräfteselbstwirksamkeit und Berufszufriedenheit werden basierend auf den Ergebnissen der Vormodelle im Hauptmodell III ausschließlich indirekte Effekte über Unterrichtsmerkmale und individuelle Schülermerkmale auf Leistung modelliert. Die höchsten linearen Zusammenhänge zeigen sich zwischen der Lehrkräfteselbstwirksamkeit und der kognitiven Aktivierung (β = .141) sowie zwischen der Berufszufriedenheit und der individuellen Förderung (β = .141). Beide Zusammenhänge werden jedoch nicht signifikant, wobei der Zusammenhang zwischen Berufszufriedenheit und individueller Förderung nur knapp das Signifikanzniveau von $p \leq .05$ verfehlt. Alle weiteren möglichen Zusammenhänge zwischen den Lehrkräfte- und den Unterrichtsmerkmalen werden ebenfalls nicht signifikant und weisen durchgängig Regressionskoeffizienten von unter .100 auf. Einzig die Korrelation zwischen der Lehrkräfteselbstwirksamkeit und der Berufszufriedenheit mit r = .341 fällt in diesem Teil des Modells signifikant aus. Für beide untersuchten Lehrkräftemerkmale zeigen sich in Hauptmodell III somit keine indirekten Effekte auf die Leistung. Die Hypothesen zum Zusammenhang von Lehrkräfteselbstwirksamkeit und Unterrichtsmerkmalen (Hypothese 17 bis 20) sowie von Berufszufriedenheit und Unterrichtsmerkmalen (Hypothese 21, 23 und 24) müssen basierend auf den Ergebnissen von Hauptmodell III vorläufig verworfen werden. Einzig Hypothese 22, nach der kein Zusammenhang zwischen Berufszufriedenheit und kognitiver Aktivierung erwartet wurde, kann vorläufig angenommen werden. Eine abschließende Prüfung dieser Hypothesen erfolgt im nächsten Schritt bei der Berechnung des Gesamtmodells.



*** = $p \leq .001$; ** = $p \leq .01$; * = $p \leq .05$; ns = nicht signifikant

Anmerkung: für logistische Regressionen wird ausschließlich der Wert für den Übergang von KS IV auf KS V angegeben

Abb.: 24 Ergebnisse Hauptmodell III

8.4.4 Hauptmodell IV: Gesamtmodell mit Kontrollvariablen

Im letzten Schritt, dem Hauptmodell IV, wird das vollständige Gesamtmodell unter Hinzunahme von relevanten Kontrollvariablen berechnet. Zusätzlich zu den Merkmalen, die bereits in Hauptmodell III enthalten sind, werden mit dem CFT, dem Geschlecht, dem Migrationshintergrund und dem heimischen Bücherbesitz vier individuelle Hintergrundmerkmale der Schüler:innen aufgenommen. Da in Mplus Schätzprobleme auftreten können, wenn einzelne Variablen einen stark abweichenden Wertebereich im Vergleich zu den anderen Variablen in einem Modell aufweisen (Schulz-Heidorf, 2016), wird der CFT dabei z-standardisiert aufgenommen. Die rekodierte, nominale Variable zum Migrationshintergrund wird über zwei Dummy-Variablen im Modell abgebildet.

In Tabelle 20 sind die Modell-Güte-Kriterien für das Nullmodell und alle Hauptmodelle dargestellt. Das Hauptmodell IV weist eine Devianz von 83830.600, einen AIC von 84220.599 und einen BIC von 85367.651 auf. Damit liegen die Werte für alle drei Kriterien deutlich unter den Werten des Nullmodells, sodass das Hauptmodell diesem vorgezogen werden sollte. Darüber hinaus weist das Hauptmodell IV auch im Vergleich zu Hauptmodell II und III geringere Werte auf und ist somit auch besser zur Beschreibung der Daten geeignet als die Modelle ohne Kontrollvariablen bzw. Kontroll- und Lehrkräftevariablen. Im Vergleich zum Hauptmodell I weist das Hauptmodell IV hingegen sehr ähnliche *Model-Fit*-Werte auf. Während für die Devianz der Wert des Hauptmodells IV (83830.600) knapp unter dem des Hauptmodells I (83859.790) liegt und somit für einen besseren Fit das Hauptmodells IV spricht, liegen die Werte für den AIC (HM IV: 84220.599; HM I: 83955.789) und den BIC (HM IV: 85367.651; HM I: 84251.494) etwas darüber, wobei sich die größte Differenz zwischen beiden Modellen für den BIC zeigt. Wie in Kapitel 7.2.3 beschrieben neigen jedoch sowohl der AIC als auch der BIC dazu, bei großen Stichproben verzerrte Ergebnisse zu produzieren, wobei insbesondere der BIC komplexe Modelle gegenüber einfachen Modellen bestraft. Unter Berücksichtigung dieses Umstandes können sowohl das Hauptmodell I als auch das Hauptmodell IV als beste Modelle zur Beschreibung das Datengrundlage ausgemacht werden. Da das Hauptmodell IV im Gegensatz zum sehr einfachen Hauptmodell I einen deutlich detaillierteren Einblick in das Zusammenspiel möglicher Prädiktoren besonderer Leistungsstärke ermöglicht

und zudem die Zusammenhänge das Hauptmodells I ebenfalls abdeckt, wird dieses Modell für die Beantwortung der Forschungsfrage dieser Arbeit präferiert.

Tab.: 20 Model-Fit-Werte Hauptmodell VI im Vergleich zu vorherigen Modellschritten

Modell	**Devianz**	**AIC**	**BIC**
Nullmodell (NM)	116491.382	116635.382	117078.939
Hauptmodell I (HM I)	83859.790	83955.789	84251.494
Hauptmodell II (HM II)	91638.446	91912.446	92756.437
Hauptmodell III (HM III)	96450.264	96820.265	97959.961
Hauptmodell IV (HM IV)	83830.600	84220.599	85367.651

Die Ergebnisse zu Hauptmodell IV sind in Abbildung 25 dargestellt. Die Kontrollvariablen CFT, Geschlecht, Migrationshintergrund und heimischer Bücherbesitz werden zusätzlich zu den bisherigen Merkmalen direkt auf Leistung modelliert. Es zeigt sich, dass der CFT und der heimische Bücherbesitz signifikante Hintergrundmerkmale der Schüler:innen sind. Die höchsten *odds ratios* ergeben sich mit 1.840 für den CFT. Die Wahrscheinlichkeit, Leistungen auf KS V anstatt auf KS IV zu erbringen, erhöht sich damit um das 1.8-Fache, wenn der CFT um eine Einheit ansteigt. Etwas niedriger fallen die *odds ratios* (1.452) für den heimischen Bücherbesitz aus. Die Merkmale Geschlecht (*odds ratios* = 1.141) und Migrationshintergrund – sowohl für die Dummyvariable ‚ein Elternteil mit Migrationshintergrund' (*odds ratios* = 0.905), als auch für die Dummyvariable ‚beide Elternteile mit Migrationshintergrund' (*odds ratios* = 0.799) – stellen sich nicht als bedeutsame Prädiktoren besonderer Leistungsstärke heraus.

Für die individuellen Schülermerkmale fachbezogene Einstellung und fachbezogenes Selbstkonzept zeigen sich unter Kontrolle der Hintergrundvariablen Veränderungen in den *odds ratios* beider Merkmale. Die *odds ratios,* Leistungen auf KS V anstatt auf KS IV zu erbringen, betragen in Hauptmodell IV für die fachbezogene Einstellung 0.640 und zeigen damit eine geringere Bedeutung als in den vorherigen Modellschritten an. Gleichzeitig ist der Effekt der fachbezogenen Einstellung auf die Leistung nicht mehr signifikant. Auch die *odds ratios* für das fachbezogene Selbstkonzept fallen mit 2.261 niedriger aus als in den anderen Hauptmodellen, der Effekt des Selbstkonzeptes ist jedoch weiterhin signifikant. Die Korrelation zwischen fachbezo-

gener Einstellung und fachbezogenem Selbstkonzept liegt mit $r = .809$ leicht niedriger als in den Hauptmodellen I bis III, ist aber nach wie vor hoch. Mit den Ergebnissen aus dem Hauptmodell IV kann somit die Hypothese 1 angenommen werden. Hypothese 2 muss unter Kontrolle relevanter Hintergrundmerkmale der Schüler:innen jedoch verworfen werden. Eine Übersicht über die einzelnen angenommenen und verworfenen Hypothesen findet sich in Kapitel 9.1.

Ähnlich zeigen sich auch Veränderungen in den Zusammenhängen von Unterrichtsmerkmalen mit individuellen Schülermerkmalen und Leistung. Die *odds ratios* der kognitiven Aktivierung betragen im Hauptmodell IV 0.250. Im Vergleich mit den vorherigen Modellschritten fällt die Bedeutsamkeit der kognitiven Aktivierung somit geringer aus, ist jedoch nach wie vor signifikant. Nicht mehr signifikant ist hingegen der direkte Zusammenhang zwischen der konstruktiven Unterstützung und der Leistung. Mit *odds ratios* von 3.404 fallen beide Werte dabei deutlich niedriger aus als in den Hauptmodellen I bis III. Die Hypothesen 3 und 4 müssen somit verworfen werden. Zwar wurde mit Hypothese 4 ein direkter Zusammenhang zwischen kognitiver Aktivierung und Leistung erwartet, anders als die Ergebnisse von Hauptmodell IV zeigen, wurde jedoch ein positiver Einfluss angenommen.

Weiterhin signifikant ist der indirekte Effekt der konstruktiven Unterstützung auf die Leistung vermittelt über das fachbezogene Selbstkonzept. Der lineare Zusammenhang zwischen konstruktiver Unterstützung und dem fachbezogenen Selbstkonzept ($\beta = .987$) ist ebenso wie der bereits dargestellte Zusammenhang zwischen fachbezogenem Selbstkonzept und Leistung nach wie vor signifikant. Insgesamt verliert jedoch auch dieser indirekte Effekt unter Kontrolle der Hintergrundvariablen an Bedeutung. Anders als in den vorherigen Modellschritten zeigt sich nun zudem ein direkter linearer Zusammenhang zwischen konstruktiver Unterstützung und fachbezogener Einstellung. Mit einem Regressionskoeffizienten von $\beta = .632$ fällt dieser Zusammenhang positiv und signifikant aus. Da der Pfad von fachbezogener Einstellung auf die Leistung jedoch in diesem Modell nicht signifikant ist, zeigt sich kein indirekter Effekt der konstruktiven Unterstützung über die fachbezogene Einstellung auf die Leistung. Mit dem Hauptmodell IV können die Hypothesen 7 und 8 somit angenommen werden.

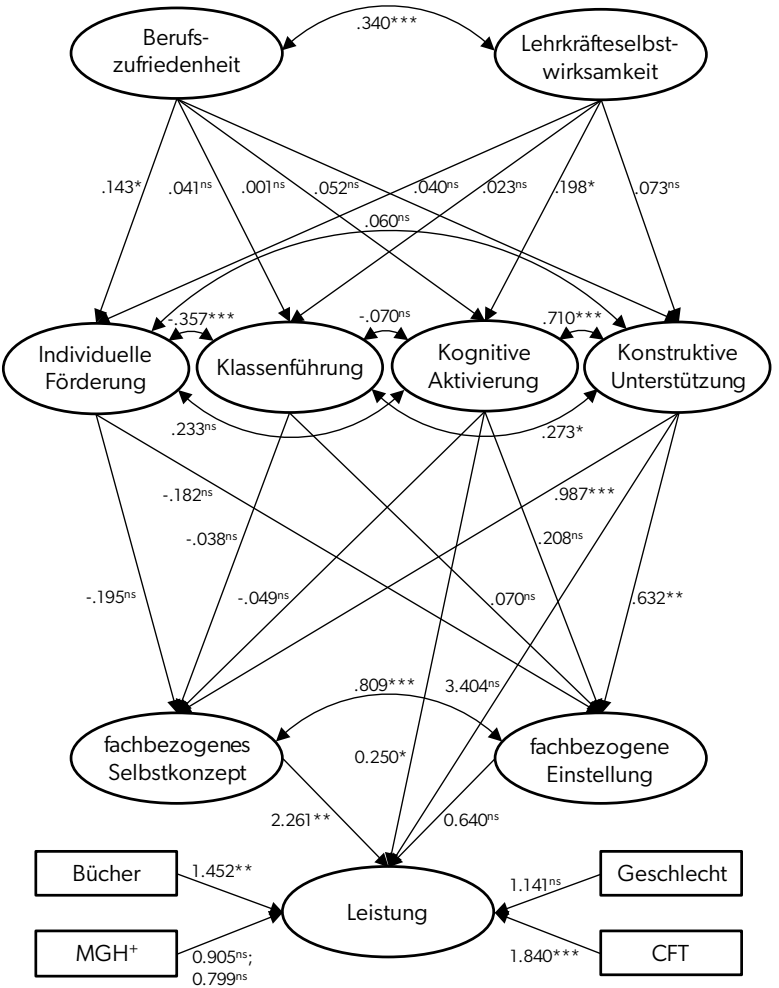
Für kognitive Aktivierung zeigt sich übereinstimmend mit den vorherigen Modellschritten, dass es keine indirekten Zusammenhänge über das fachbezogene Selbstkonzept oder die fachbezogene Einstellung mit der naturwissenschaftlichen Leistung gibt. Zwar gibt es einen signifikanten Zusammenhang zwischen fachbezogenem Selbstkonzept und Leistung, allerdings wird sowohl der Pfad von kognitiver Aktivierung

auf das fachbezogene Selbstkonzept (β = -.049) als auch der Pfad zur fachbezogenen Einstellung (β = .208) nicht signifikant. Die Hypothesen 9 und 10 können dementsprechend nicht angenommen werden.

Für den indirekten Zusammenhang zwischen individueller Förderung und Leistung vermittelt über die fachbezogene Einstellung zeigt sich, dass dieser nicht mehr signifikant wird. Dabei betrifft dies nicht nur den Zusammenhang zwischen fachbezogener Einstellung und Leistung, sondern auch den Pfad von individueller Förderung auf fachbezogene Einstellung (β = -.182). Letzterer verfehlt das Signifikanzniveau von $p \leq .05$ nur knapp. Die Hypothese 13 muss somit verworfen werden. Ebenfalls nicht signifikant fallen die Zusammenhänge zwischen der individuellen Förderung und dem fachbezogenen Selbstkonzept (β = -.195) sowie zwischen Klassenführung und fachbezogenem Selbstkonzept (β = -.038) bzw. fachbezogener Einstellung (β = .070) aus. Diese Pfade wurden jedoch in keinem der berechneten Hauptmodelle signifikant. Die Hypothesen 11, 12 und 14 müssen somit auch auf Basis der Ergebnisse von Hauptmodell IV verworfen werden.

Wenig Veränderungen gibt es auch bei den Korrelationen zwischen den Unterrichtsmerkmalen. Signifikante Korrelationen zeigen sich nach wie vor zwischen den Merkmalen konstruktive Unterstützung und Klassenführung (β = .273), Klassenführung und individuelle Förderung (β = -.357) sowie konstruktive Unterstützung und kognitive Aktivierung (β = .710), wobei letztere um ca. .200 niedriger ausfällt als in Hauptmodell II und III. Die Korrelationen von kognitiver Aktivierung mit Klassenführung (β = -.070) bzw. individueller Förderung (β = .233) sowie von konstruktiver Unterstützung und individueller Förderung (β = .060) werden nicht signifikant.

Anders als in den vorherigen Modellschritten zeigen sich unter Kontrolle der Hintergrundmerkmale der Schüler:innen nun zum Teil auch signifikante lineare Zusammenhänge zwischen den Merkmalen der Lehrkräfte und des Unterrichts. Für die Lehrkräfteselbstwirksamkeit zeigt sich in Hauptmodell IV ein signifikanter linearer Zusammenhang mit der kognitiven Aktivierung (β = .198). Damit ergibt sich gleichzeitig ein indirekter Zusammenhang von Lehrkräfteselbstwirksamkeit über die kognitive Aktivierung auf die Leistung. Für die konstruktive Unterstützung (β = .073), die Klassenführung (β = .001) und die individuelle Förderung (β = .040) zeigen sich keine signifikanten Effekte der Lehrkräfteselbstwirksamkeit. Die Korrelation zwischen den beiden Lehrkräftemerkmalen beträgt weiterhin nahezu unverändert .340. Die Hypothese 18 kann somit mit den Ergebnissen von Hauptmodell IV bestätigt werden, die Hypothesen 17, 19 und 20 werden abgelehnt.



*** = $p \leq .001$; ** = $p \leq .01$; * = $p \leq .05$; ns = nicht signifikant;
+ = ein Elternteil in Ausland geboren; beide Elternteile im Ausland geboren

Anmerkung: für logistische Regressionen wird ausschließlich der Wert für den Übergang von KS IV auf KS V angegeben

Abb.: 25 Ergebnisse Hauptmodell IV

Der Zusammenhang zwischen der Berufszufriedenheit und der individuellen Förderung, der in Hauptmodell III noch knapp das Signifikanzniveau von $p \leq .05$ verfehlt, fällt in Hauptmodell IV signifikant aus ($\beta = .143$). Die Hypothese 24 kann somit angenommen werden. Die Zusammenhänge von Berufszufriedenheit mit Klassenführung ($\beta = .041$), kognitiver Aktivierung ($\beta = .023$) und konstruktiver Unterstützung ($\beta = .52$) sind weiterhin nicht signifikant. Die Hypothese 22, die davon ausgeht, dass es keinen Zusammenhang zwischen Berufszufriedenheit und kognitiver Aktivierung gibt, kann entsprechend dieser Befunde angenommen werden. Die Hypothesen 21 und 23 müssen hingegen verworfen werden.

9

Zusammenfassung und Diskussion der Ergebnisse

In diesem Kapitel werden die Ergebnisse der vorangegangenen Analysen zusammengefasst und diskutiert. In einem ersten Schritt werden dazu die Befunde der durchgeführten Untersuchung und deren Bedeutung für die in Kapitel 5 formulierten Hypothesen zusammengefasst. Anschließend werden die Ergebnisse vor dem Hintergrund der in Kapitel 4 dargestellten theoretischen Grundlagen diskutiert und in das Bild bisheriger empirischer Befunde eingeordnet sowie interpretiert. Abschließend erfolgt eine kritische Betrachtung der vorliegenden Arbeit. In Kapitel 10 wird schließlich die Bedeutung der Befunde für die Praxis dargelegt sowie ein Ausblick auf weitere mögliche Forschungen gegeben.

9.1 Ergebnisse

In den vorangegangenen Kapiteln wurden die Analysen zur Bearbeitung der Hypothesen vorgestellt. Dabei ergaben Nullmodelle ICCs von näherungsweise 10 bis 13 % für die abhängige Variable (naturwissenschaftliche Leistung), sodass Mehrebenenstrukturgleichungsmodelle gerechnet wurden. Diese kombinieren sowohl multinominal logistische Regressionen als auch lineare Regressionen, je nachdem, ob Regressionspfade auf die multinomial logistische Zielvariable oder auf eines der acht metrischen latenten Konstrukte modelliert wurden.

Da sich die Analysen unter den gegebenen Voraussetzungen als zeitlich und technisch sehr anspruchsvoll herausstellten, wurden die Integrationspunkte so weit verringert, dass eine Berechnung unter den gegebenen Voraussetzungen möglich wurde. In Vormodellen wurde zunächst überprüft, welche Lehrkräfte- und Unterrichtsmerkmale einen direkten signifikanten Effekt auf die Übergangswahrscheinlichkeit, dass Schüler:innen Leistungen auf KS V anstatt auf KS IV erbringen, haben. Aus forschungspragmatischen Gründen wurden in den weiteren Modellschritten nur die direkten Pfade zusätzlich in das Gesamtmodell mit indirekten Pfaden aufgenommen, die sich in Vormodellen als signifikant herausstellten.

In den Vormodellen zeigte sich, dass die konstruktive Unterstützung und die kognitive Aktivierung direkt signifikant mit der naturwissenschaftlichen Leistung zusammenhängen. Für die Unterrichtsmerkmale Klassenführung und individuelle Förderung sowie die Lehrkräftemerkmale Berufszufriedenheit und Lehrkräfteselbstwirksamkeit ergaben sich keine direkten Effekte. Entsprechend wurden im Gesamtmodell zusätzlich zu indirekten Pfaden von Lehrkräftemerkmalen über Unterrichts- und

individuelle Schülermerkmale auf Leistung direkte Pfade der kognitiven Aktivierung und der konstruktiven Unterstützung auf die Leistung modelliert. Das Gesamtmodell wurde anschließend in vier Schritten aufgebaut. Im letzten Schritt wurde das vollständige Modell unter Berücksichtigung relevanter Kontrollvariablen berechnet.

Im vollständigen Gesamtmodell zeigt sich, dass von den individuellen Schülermerkmalen das fachbezogene Selbstkonzept direkt signifikant positiv mit der Übergangswahrscheinlichkeit, Leistungen auf KS V anstatt auf KS IV zu erbringen, zusammenhängt. Ein direkter signifikanter Effekt auf die Leistung zeigt sich auch für die kognitive Aktivierung. Im Gegensatz zum Effekt des fachbezogenen Selbstkonzeptes fällt dieser jedoch negativ aus. Die fachbezogene Einstellung und die konstruktive Unterstützung sind keine bedeutsamen direkten Prädiktoren der Übergangswahrscheinlichkeit. Für die konstruktive Unterstützung zeigt sich allerdings ein signifikanter indirekter positiver Effekt auf die Leistung vermittelt über das fachbezogene Selbstkonzept sowie ein positiver Effekt auf die fachbezogene Einstellung. Für Klassenführung und individuelle Förderung ergeben sich keine indirekten Effekte auf die naturwissenschaftliche Leistung. Für die Lehrkräfteselbstwirksamkeit zeigt sich ein signifikanter Zusammenhang mit der kognitiven Aktivierung und für Berufszufriedenheit mit der individuellen Förderung. Gleichzeitig ergibt sich somit ein indirekter Effekt der Lehrkräfteselbstwirksamkeit über die kognitive Aktivierung auf die Übergangswahrscheinlichkeit, Leistungen auf KS V anstatt auf KS IV zu erbringen.

Auffällig bei den Ergebnissen im letzten Schritt der Berechnung des Gesamtmodells ist, dass sich für einige Pfade Ergebnisse zeigen, die in den vorherigen Modellschritten nicht in der Art auftraten. So war die fachbezogene Einstellung in allen vorherigen Modellschritten ein signifikanter negativer Prädiktor der Übergangswahrscheinlichkeit, ebenso die konstruktive Unterstützung, für die sich zusätzlich zum indirekten auch ein direkter signifikanter positiver Effekt ergab. Ein indirekter Effekt zeigte sich darüber hinaus auch für die individuelle Förderung auf die Leistung vermittelt über die fachbezogene Einstellung. Zwischen Lehrkräfte- und Unterrichtsmerkmalen ergaben sich hingegen zuvor gar keine direkten Zusammenhänge.

Die Überprüfung der Hypothesen erfolgt auf Basis des vollständigen Gesamtmodells. Die Ergebnisse aus den vorherigen Modellschritten werden jedoch bei den Darstellungen in diesem und dem nachfolgenden Kapitel mitberücksichtigt.

Tab.: 21 Übersicht über die bestätigten Hypothesen

Bestätigte Hypothesen	
(H2)	Das fachbezogene Selbstkonzept der Schüler:innen beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H7)	Die konstruktive Unterstützung hängt positiv mit der fachbezogenen Einstellung zusammen.
(H8)	Die konstruktive Unterstützung hängt positiv mit dem fachbezogenen Selbstkonzept zusammen.
(H18)	Die Lehrkräfteselbstwirksamkeit hängt positiv mit der kognitiven Aktivierung zusammen.
(H22)	Die Berufszufriedenheit hängt nicht mit der kognitiven Aktivierung zusammen.
(H24)	Die Berufszufriedenheit hängt positiv mit der individuellen Förderung zusammen.

In den Tabellen 21 und 22 sind die Hypothesen aus Kapitel 5 aufgeführt. Tabelle 21 beinhaltet dabei die Hypothesen, die auf Basis des vollständigen Gesamtmodells angenommen werden können. Tabelle 22 können die Hypothesen entnommen werden, die auf Basis der Ergebnisse abgelehnt werden müssen.

Es zeigt sich, dass die Hypothesen H2, H7, H8, H18, H22 und H24 auf Basis der Befunde angenommen werden können. Auf zwei Hypothesen soll an dieser Stelle nochmal genauer eingegangen werden, da deren Zuordnung unter Umständen nicht intuitiv nachvollziehbar ist: Die Hypothese 4 postuliert zwar übereinstimmend mit den Ergebnissen einen direkten signifikanten Zusammenhang der kognitiven Aktivierung mit der Übergangswahrscheinlichkeit, Leistungen auf KS V anstatt auf KS IV zu erbringen, allerdings wurde auf Grundlage bisheriger Forschungen ein positiver Effekt erwartet. Stattdessen zeigte sich ein negativer Effekt, sodass die Hypothese 4 abgelehnt werden muss. Hypothese 22 hingegen wird in Tabelle 21 aufgeführt. Zwar zeigt sich im Gesamtmodell kein signifikanter Zusammenhang zwischen Berufszufriedenheit und kognitiver Aktivierung, da diese Annahme – abweichend zur üblichen Formulierung in den anderen Hypothesen – allerdings so in der Hypothese abgebildet wurde, wird Hypothese 22 bestätigt. Alle anderen Hypothesen weisen eine einheitliche Formulierung auf, sodass deren Zuordnung zu Tabelle 21 und 22 eindeutiger nachvollziehbar ist. Sie werden deshalb nicht nochmal separat erläutert. Die Bedeutung der in diesem Kapitel zusammengefassten Befunde wird im Folgenden diskutiert.

Tab.: 22 Übersicht über die abgelehnten Hypothesen

Abgelehnte Hypothesen	
(H1)	Die fachbezogene Einstellung der Schüler:innen zum Fach beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H3)	Die konstruktive Unterstützung beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H4)	Die kognitive Aktivierung beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H5)	Die Klassenführung beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H6)	Die individuelle Förderung beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H9)	Die kognitive Aktivierung hängt positiv mit der fachbezogenen Einstellung zusammen.
(H10)	Die kognitive Aktivierung hängt positiv mit dem fachbezogenen Selbstkonzept zusammen.
(H11)	Die Klassenführung hängt positiv mit der fachbezogenen Einstellung zusammen.
(H12)	Die Klassenführung hängt negativ mit dem fachbezogenen Selbstkonzept zusammen.
(H13)	Die individuelle Förderung hängt positiv mit der fachbezogenen Einstellung zusammen.
(H14)	Die individuelle Förderung hängt positiv mit dem fachbezogenen Selbstkonzept zusammen.
(H15)	Die Lehrkräfteselbstwirksamkeit beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H16)	Die Berufszufriedenheit beeinflusst die Übergangswahrscheinlichkeit Kompetenzen auf KS V anstatt auf KS IV zu zeigen positiv.
(H17)	Die Lehrkräfteselbstwirksamkeit hängt positiv mit der konstruktiven Unterstützung zusammen.
(H19)	Die Lehrkräfteselbstwirksamkeit hängt positiv mit der Klassenführung zusammen.
(H20)	Die Lehrkräfteselbstwirksamkeit hängt positiv mit der individuellen Förderung zusammen.
(H21)	Die Berufszufriedenheit hängt positiv mit der konstruktiven Unterstützung zusammen.
(H23)	Die Berufszufriedenheit hängt positiv mit der Klassenführung zusammen.

9.2 Einordnung und Interpretation

Mit der vorliegenden Arbeit soll die Frage nach bedeutsamen Prädiktoren besonderer Leistungsstärke in den Naturwissenschaften in der Grundschule beantwortet werden. Dazu wurde das Rahmenmodell zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen als Ausgangpunkt für die Überlegungen zu möglichen Einflussfaktoren herangezogen. Die Diskussion der Ergebnisse erfolgt entlang der drei Bereiche aus dem Rahmenmodell, die in dieser Arbeit in den Blick genommen wurden, beginnend mit den individuellen Schülermerkmalen. Abschließend wird die Bedeutung der Ergebnisse im Hinblick auf das gesamte Rahmenmodell erörtert. Ein wichtiger Aspekt, der bei der Diskussion und Interpretation der Ergebnisse berücksichtigt werden muss, ist, dass der Untersuchung Querschnittsdaten zugrunde liegen. Das bedeutet, dass auf Basis der Daten keine Aussagen über kausale Zusammenhänge zwischen Merkmalen, sondern lediglich theoriegeleitete Annahmen zur Einordnung der gefundenen Zusammenhänge getroffen werden können. Ausführlichere Darstellungen zu diesem Aspekt finden sich in Kapitel 9.3.

9.2.1 Merkmale der Schüler:innen

Die Ergebnisse im Bereich der individuellen Schülermerkmale zeigen, dass insbesondere das fachbezogene Selbstkonzept der Schüler:innen ein bedeutsamer Prädiktor besonderer Leistungsstärke in einem Fach ist. Über alle Modelle hinweg erhöht das fachbezogene Selbstkonzept bei einem Anstieg um eine Einheit die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen, zwischen 226.14 und 264.40 %, wobei sich der geringste Wert unter Kontrolle relevanter Hintergrundmerkmale ergibt. Insgesamt kann der Einfluss des fachbezogenen Selbstkonzeptes, insbesondere im Vergleich mit den anderen untersuchten Prädiktoren, als hoch beschrieben werden. Damit stimmen die Ergebnisse der vorliegenden Arbeit nicht nur mit denen vorheriger genereller Forschungen überein (Hansford & Hattie, 1982; Marsh et al., 2018), sondern auch mit den Ergebnissen differenzieller Untersuchungen zu Einflussfaktoren besonderer Leistungsstärke, die übereinstimmend eine hohe Bedeutsamkeit des fachbezogenen Selbstkonzeptes belegen (Neuendorf et al., 2017; Tuorón et al., 2018). Dabei scheint das fachbezogene Selbstkonzept im Kontext differenzieller Untersuchungen aktuell der am besten untersuchte sowie

am häufigsten bestätigte Prädiktor besonderer Leistungsstärke zu sein. Die vorliegende Untersuchung kommt ebenfalls zu diesem Ergebnis und konnte darüber hinaus zeigen, dass das fachbezogene Selbstkonzept sogar ein bedeutsamerer Prädiktor besonderer Leistungsstärke ist als die kognitiven Fähigkeiten. In generellen Untersuchungen zeigt sich dieser Befund bisher eher selten (z. B. Steinmayr & Spinath, 2009), was die differenzielle Bedeutsamkeit dieses Merkmals untermauert. Gleichzeitig steht diese Einschätzungen im Einklang mit dem Ergebnis einer Untersuchung von Helmke (1992), der nachweisen konnte, dass der Zusammenhang zwischen dem fachbezogenen Selbstkonzept und den schulischen Leistungen mit zunehmender Leistungsstärke höher ausfällt.

Für den Zusammenhang zwischen fachbezogener Einstellung und naturwissenschaftlichen Leistungen zeigen sich entgegen den Erwartungen z. T. deutlich niedrigere, aber vor allem negative Effekte. Damit erhöht ein Anstieg um eine Einheit die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen, nicht, sondern reduziert sie. Die *odds ratios* liegen dabei je nach Modell zwischen 0.453 und 0.640, wobei der Effekt der fachbezogenen Einstellung unter Kontrolle relevanter Hintergrundmerkmale nicht mehr signifikant ausfällt. Erwartungswidrig ist dieser Befund vor dem Hintergrund, dass bisherige generelle Studien vor allem nominelle oder signifikante positive Effekte der fachbezogenen Einstellung feststellen konnten (Ai, 2002; Chen et al., 2018), sich nur ganz vereinzelt negative Zusammenhänge zeigten (Ozel et al., 2013) und auch differenzielle Befunde eher auf einen positiven Zusammenhang hindeuten (Grootenboer & Hemmings, 2007).

Eine mögliche Erklärung für die in dieser Arbeit gefundenen negativen Zusammenhänge könnte in der angenommenen Wirkrichtung liegen. Auf Basis theoretischer Annahmen, die im Kontext genereller Untersuchungen entstanden (z. B. Erwartung x Wert-Modell), wurde vermutet, dass vor allem die fachbezogene Einstellung die naturwissenschaftlichen Leistungen beeinflusst (Atkinson, 1957; Eccles & Wigfield, 2002). Es könnte jedoch sein, dass diese Annahme bei besonders leistungsstarken Schüler:innen nicht zutrifft. Ein möglicher Erklärungsansatz wäre, dass diese Schüler:innen eine weniger positive affektive Einstellung zum Sachunterricht haben, da ihnen die naturwissenschaftlichen Inhalte eher leichtfallen und sie dementsprechend weniger in das Fach investieren müssen, um sehr gute Leistungen zu erzielen. Gleichzeitig kann vermutet werden, dass Schüler:innen deshalb im schulischen Alltag häufiger Langeweile im Sachunterricht empfinden und seltener für sie neue und interessante Unterrichtsinhalte behandelt werden, was vermutlich zu einer höheren positiven affektiven

Einstellung führen würde. Die fachbezogene affektive Einstellung wäre somit eher als Resultat der eigenen Erfahrungen und Leistungen zu verstehen. Ein alternativer Erklärungsansatz könnte auch im Alter der in TIMSS befragten Schüler:innen liegen. Für die fachbezogene Einstellung konnte vereinzelt gezeigt werden, dass ähnlich wie beim Selbstkonzept beide Wirkrichtungen vorkommen können. Dabei scheint im Kindesalter der Effekt der Leistung auf die Einstellung, ab dem Jugendalter der Effekt der Einstellung auf die Leistung zu überwiegen (Willson, 1983). Da wie bereits erläutert in dieser Arbeit jedoch Querschnittsdaten genutzt wurde, können hierzu ausschließlich Hypothesen formuliert werden. Eine längsschnittliche Untersuchung zur Identifikation von kausalen Zusammenhängen ist mit TIMSS-Daten nicht möglich.

Dass die fachbezogene Einstellung insgesamt weniger bedeutsam ist als das fachbezogene Selbstkonzept, steht wiederum in Übereinstimmung mit bisherigen Befunden zum Verhältnis von Erwartungs- und Wertkomponenten. Bei vorherigen Untersuchungen stellten sich Erwartungskomponenten zumeist als relevanter für die schulischen Leistungen heraus als Wertkomponenten (Kriegbaum et al., 2018; Ozel et al., 2013). Dies kann mit den Ergebnissen dieser Arbeit auch für den Kontext besonderer Leistungsstärke bestätigt werden.

Abschließend werden für diesen Bereich die Ergebnisse zu den Kontrollvariablen betrachtet. Von den vier Kontrollvariablen stellten sich zwei als bedeutsame Prädiktoren von besonderer Leistungsstärke heraus (CFT und Büchervariable). Sowohl das Geschlecht der Schüler:innen als auch der Migrationshintergrund waren hingegen keine signifikanten Prädiktoren. Für den Migrationshintergrund, der über zwei dummykodierte Variablen im Modell abgebildet wurde, zeigte sich dabei, dass das Vorhandensein eines Elternteils mit Migrationshintergrund mit *odds ratios* von knapp unter Eins nahezu keinen Effekt auf die schulischen Leistungen hat. Wenn beide Elternteile einen Migrationshintergrund aufweisen, fällt der Effekt zwar nominell stärker negativ aus, ist jedoch ebenfalls nicht signifikant. Dass damit eine Diskrepanz zu Ergebnissen genereller Untersuchungen vorliegt, bei denen der Migrationshintergrund häufig ein bedeutsamer (negativer) Prädiktor von schulischen Leistungen und Bildungserfolg ist (Bellin, 2009; Schöber, Retelsdorf & Köller, 2015; Wendt, Schwippert, Stubbe & Jusufi, 2020), könnte möglicherweise dadurch erklärt werden, dass der Migrationshintergrund von Schüler:innen weniger im Kontext leistungsstarker als schwacher und durchschnittlich leistungsstarker Schüler:innen relevant sein könnte.

Für das Geschlecht zeigt sich hingegen ein nomineller, positiver Effekt auf die Wahrscheinlichkeit Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV

zu erbringen. Der positive Effekt dieser binären Variable kann dabei folgendermaßen interpretiert werden: Da die Mädchen im Datensatz mit einem niedrigeren Wert kodiert sind als Jungen, bedeutet der Anstieg um eine Einheit in diesem Fall, dass ein Kind anstatt des weiblichen das männliche Geschlecht angegeben hat. Da die *odds ratios* nominell positiv ausfallen steigt somit die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstelle von Kompetenzstufe IV zu zeigen, bei Angabe des männlichen anstatt des weiblichen Geschlechts. Der Effekt fällt jedoch nicht signifikant aus. Damit zeigt sich in dieser Arbeit eine deutlich niedrigere Bedeutsamkeit des Geschlechts, als es nach der *greater male hypothesis* und bisherigen Untersuchungen zu erwarten war (Ellis, 1894; Forgasz & Hill, 2013; Wendt et al., 2013). Eine mögliche Erklärung könnte in der Definition der Spitzengruppe liegen, die teils stark zwischen Untersuchungen variiert. Während Forgasz und Hill (2013) lediglich die obersten 2 % der Schülerschaft als leistungsstark definierten und dabei größere Unterschiede zugunsten der Jungen fanden, basiert die Untersuchung von Wendt et al. (2013) auf Daten aus TIMSS/IGLU 2011, sodass, wie in dieser Arbeit, die Zugehörigkeit zur höchsten Kompetenzstufe bzw. das 95. Perzentil als Kriterium verwendet wird. Die Ergebnisse von Wendt et al. (2013) zeigen dabei zwar ebenfalls einen höheren Anteil von Jungen in der Spitzengruppe, der Unterschied fällt jedoch nicht so stark aus wie bei Forgasz und Hill (2013). Da die Definition der leistungsstarken Schüler:innen von Wendt et al. (2013) mit der Definition der vorliegenden Arbeit vergleichbar ist, stehen die eher geringen Geschlechterunterschiede in der vorliegenden Untersuchung somit in Übereinstimmung mit den Befunden dieser Studie. Möglicherweise hat also die Definition der Spitzengruppe Auswirkungen darauf, wie stark sich Geschlechterunterschiede zugunsten der Jungen zeigen.

Ein bedeutsamer Einfluss hat sich hingegen für den über die Büchervariable erhobenen sozioökonomischen Hintergrund ergeben. Mit einer um eine Einheit höheren Angabe zum Bücherbesitz steigt die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen, um 145.10 %. Ähnliches zeigte sich auch für den CFT als Maß der kognitiven Fähigkeiten. Die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V zu zeigen, erhöht sich bei einem Anstieg um eine Einheit um 183.90 %. Damit sind sowohl der sozioökonomische Hintergrund als auch die kognitiven Fähigkeiten der Schüler:innen übereinstimmend mit bisherigen differenziellen Untersuchungen (z. B. Pitsia, 2022; Schmidtner, 2017) relevante Prädiktoren besonderer Leistungsstärke. Unter allen individuellen Schülermerkmalen ist allerdings weiterhin das fachbezogene Selbstkonzept das einflussreichste Merkmal.

Erfreulich ist dabei, dass mit den kognitiven Fähigkeiten ein Hintergrundmerkmal am bedeutsamsten ist, das zumindest in gewissem Maße auch direkt gefördert werden kann. Wie in Kapitel 2 beschrieben, sind die kognitiven Fähigkeiten insbesondere im jungen Kindesalter noch entwickelbar, sodass im Kontext der Grundschule eine systematische Förderung möglich ist. Dagegen sind die Merkmale Geschlecht, Migrationshintergrund und sozioökonomischer Hintergrund von Schüler:innen grundsätzlich feststehend und kaum durch äußere Einflüsse veränderbar. Langfristig kann die Bedeutsamkeit dieser Merkmale für schulische Leistung beispielsweise auf Ebene des Bildungssystems zwar prinzipiell verändert werden, die unmittelbaren Möglichkeiten, die Lehrkräfte und Schulen im Hinblick auf diese Merkmale haben, sind jedoch stark begrenzt. Die Ergebnisse dieser Arbeit machen somit Hoffnung, dass besondere Leistungsstärke am Ende der Grundschulzeit bei einer adäquaten Förderung der kognitiven Fähigkeiten unabhängig von individuellen bzw. familiären Hintergrundmerkmalen für mehr Schüler:innen als bisher möglich ist.

9.2.2 Merkmale des Unterrichts

Die Ergebnisse zu den Merkmalen des Unterrichts fallen über die verschiedenen Modelle hinweg unterschiedlich aus. Konsistent zeigen die Ergebnisse jedoch, dass die Klassenführung kein bedeutsamer Prädiktor besonderer Leistungsstärke ist, weder direkt noch indirekt. Damit kommt die vorliegende Untersuchung zu einem anderen Ergebnis als die meisten im ersten Teil dieser Arbeit vorgestellten generellen Studien (Praetorius et al., 2018; Seidel & Shavelson, 2007; Wang et al., 1993), da in den berichteten Untersuchungen nur vereinzelt kein Zusammenhang zwischen der Klassenführung und der Leistung gefunden wurde (Henschel et al., 2019; Roßbach, 2002). Roßbach (2002) vermutet, dass das Ausbleiben signifikanter Ergebnisse in seiner Untersuchung darauf zurückzuführen sein könnte, dass sich die untersuchten Kinder noch eher am Beginn der Grundschulzeit befanden und Klassenführung möglicherweise erst nach der Anpassungsphase, also ab ca. der dritten Klassenstufe, relevanter wird.

Die Schüler:innen in dieser Untersuchung sind jedoch bereits am Ende der Grundschulzeit, sodass diese Vermutung zumindest für leistungsstarke Schüler:innen nicht zuzutreffen scheint. Diese Einschätzung lässt sich zusätzlich dadurch untermauern, dass die Ergebnisse dieser Arbeit auch mit den Befunden einer der wenigen diffe-

renziellen Untersuchungen von Sanfo und Malgoubri (2021) übereinstimmt, die im Schnitt noch etwas ältere Kinder untersuchten (M = 14.36 Jahre). Diese fanden für leistungsstarke Schüler:innen ebenfalls keinen signifikanten Zusammenhang zwischen Klassenführung und Leistung, sodass vermutet werden kann, dass die Klassenführung für leistungsstarke Schüler:innen generell weniger bedeutsam ist. Dazu passt, dass sich die Klassenführung in dieser Arbeit nicht nur nicht für die Leistungen, sondern auch nicht für das fachbezogene Selbstkonzept und die fachbezogene Einstellung als relevant herausgestellt hat. Auch dies widerspricht den meisten generellen Befunden, die zu dem Schluss kommen, dass Klassenführung zwar schwach, aber in ähnlichem Maße bedeutsam für motivationale Merkmale ist wie für Leistung (Klieme et al., 2001; Schiepe-Tiska et al., 2016). Werden diese Befunde im Gesamten betrachtet, liefert die vorliegende Arbeit auch empirische Hinweise darauf, dass die häufig im Kontext leistungsschwacher Schüler:innen geäußerte Vermutung, die Klassenführung sei stärker für leistungsschwache als leistungsstarke Schüler:innen relevant, da letztere besser in der Lage sind eine mangelnde Klassenführung zu kompensieren, nicht nur für die Gruppe leistungsschwacher Schüler:innen, sondern auch für die Gruppe der leistungsstarken Schüler:innen zutreffend sein könnte.

Eindeutig sind ebenfalls die Ergebnisse zur kognitiven Aktivierung. Über alle Modelle hinweg zeigt sich ein signifikanter direkter Effekt der kognitiven Aktivierung auf die Wahrscheinlichkeit, naturwissenschaftliche Leistungen auf Kompetenzstufe V anstatt Kompetenzstufe IV zu erbringen. Allerdings fallen die Befunde trotzdem erwartungswidrig aus, da sie nicht wie auf Basis vorheriger genereller (z. B. Dubberke et al., 2008; Klieme et al., 2001) sowie differenzieller Befunde (Sanfo & Malgoubri, 2021) vermutet positiv sind, sondern negativ. Damit reduziert eine um eine Einheit höhere kognitive Aktivierung die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen und zwar je nach Modell zwischen -75.00 und -91.30 %. Da die einzige Untersuchung, die ebenfalls einen negativen Zusammenhang zwischen kognitiver Aktivierung und Leistung fand (Stang et al., 2020), auf den gleichen Daten basiert wie diese Arbeit, scheint es angezeigt, die Operationalisierung der kognitiven Aktivierung genauer zu überprüfen.

Die TIMSS-Skala zur kognitiven Aktivierung beinhaltet sieben Items, von denen aufgrund unzureichender Faktorladungen nur fünf für die vorliegende Analyse genutzt werden konnten. Bereits bei der Überprüfung der latenten Konstrukte in Kapitel 8.1 zeigte sich, dass die drei Hauptfacetten der kognitiven Aktivierung (herausfordernde Aufgaben stellen, an das Vorwissen der Schüler:innen anknüpfen und

anspruchsvolle Unterrichtsgespräche bzw. -argumentationen führen) nicht in gleicher Gewichtung in der Skala abgebildet werden. Durch die Streichung zweier Items, die aus methodischen Gründen notwendig war, verstärkte sich diese Ungleichgewichtung noch. Die finale Skala enthielt somit hauptsächlich Items, die auf herausfordernde Aufgaben bzw. Fragen ausgerichtet sind. Insbesondere die Facette des Vorwissens ist jedoch nicht mehr in der Skala abgebildet. Darüber hinaus sind die Items stark auf Aktivitäten der Lehrkraft und weniger auf die der Schüler:innen bezogen sowie teils sehr ähnlich formuliert, sodass die inhaltlichen Unterschiede zwischen den Items für Schüler:innen möglicherweise nicht auf den ersten Blick erkennbar sind.

Dass sich in der vorliegenden Arbeit ein negativer Zusammenhang zwischen der kognitiven Aktivierung und der Leistung zeigt, könnte damit erklärt werden, dass in den Items hauptsächlich der herausfordernde Charakter von Unterrichtsaktivitäten adressiert wird. Es kann vermutet werden, dass besonders leistungsstarke Schüler:innen wie sie in dieser Arbeit untersucht wurden, seltener herausfordernde Aufgaben und Aktivitäten im Unterricht erleben und deshalb ein negativer Zusammenhang zwischen den beiden Merkmalen zustande kommt. Der zugrundliegende Mechanismus wäre damit ein anderer als angenommen: Anstatt dass kognitiv herausfordernde Unterrichtsaktivitäten dafür sorgen, dass (potenziell) leistungsstarke Schüler:innen ihre schulischen Leistungen verbessern, führen höhere Leistungen dazu, dass die Unterrichtsaktivitäten seltener als herausfordernd erlebt werden.

Dabei muss jedoch berücksichtigt werden, dass die kognitive Aktivierung in dieser Arbeit auf Klassenebene aggregiert wurde. Für diese Form wurden bisher nur sehr selten negative Zusammenhänge mit der Leistung gefunden, weshalb bisherige Erklärungsansätze für diese erwartungswidrigen Befunde rar sind. Jaeuthe, Lenkeit, Bosse und Spörer (2023) vermuten jedoch, dass eine Ursache für den negativen Zusammenhang, den sie im Kontext inklusiv arbeitender Schulen nachweisen konnten, ein an den leistungsschwachen Schüler:innen ausgerichteter Unterricht sein könnte. Hinweise, ob diese Vermutung auch für diese Arbeit zutrifft, könnten anhand der Untersuchung der Klassenzusammensetzung ausgemacht werden. Darüber hinaus muss jedoch berücksichtigt werden, dass die Skala zur kognitiven Aktivierung in den Modellen teils unzureichende Faktorladungen aufweist und die Ergebnisse deshalb mit einer gewissen Vorsicht interpretiert werden müssen. Nichtsdestotrotz ergeben die einzelnen negativen Zusammenhänge und die geringe Anzahl an Erklärungsansätzen in der bisherigen Forschung Hinweise darauf, dass Mechanismen vorzuliegen scheinen, die bisher nicht ausreichend untersucht sind.

Indirekte Effekte der kognitiven Aktivierung über individuelle Schülermerkmale auf die Leistung zeigten sich entgegen den Erwartungen nicht. Nach den Befunden von Lazarides und Buchholz (2019) oder Drexl und Streb (2016) wären durchaus Zusammenhänge zwischen kognitiver Aktivierung und motivationalen Faktoren erwartbar gewesen. Stattdessen stehen die Befunde zur kognitiven Aktivierung eher im Einklang mit dem Bild, das im Rahmen der COACTIV-Studie gezeichnet wurde und nach dem kognitive Aktivierung insbesondere für Leistungen und nicht für motivationale Merkmale bedeutsam ist (Kunter & Voss, 2011).

Etwas weniger eindeutig fallen die Ergebnisse für die Merkmale konstruktive Unterstützung und individuelle Förderung aus. Für die konstruktive Unterstützung zeigen sich in den Hauptmodellen gleichzeitig direkte und indirekte Effekte, wobei der direkte Effekt im Gesamtmodell unter Kontrolle der Hintergrundmerkmale nicht mehr signifikant wird. In den beiden vorherigen Modellschritten ergeben sich dabei außerordentlich hohe Werte für den Zusammenhang zwischen konstruktiver Unterstützung und Leistung. Die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V statt auf Kompetenzstufe IV zu erbringen, erhöht sich bei einem Anstieg der konstruktiven Unterstützung um eine Einheit um 1030.20 bis 1058.70 %. Damit zeigen sich für die konstruktive Unterstützung die mit Abstand höchsten Werte im gesamten Modell. Dabei muss jedoch berücksichtigt werden, dass unter Kontrolle von Hintergrundmerkmalen der Schüler:innen die konstruktive Unterstützung trotz weiterhin sehr hoher Werte kein direkter signifikanter Prädiktor von besonderer Leistungsstärke mehr ist. Stattdessen zeigt sich im Zusammenspiel aller Variablen, dass der indirekte Effekt der konstruktiven Unterstützung über das individuelle Schülermerkmal fachbezogenes Selbstkonzept bedeutsamer als der direkte Effekt ist. Dieser zeigt sich bereits in den vorherigen Modellschritten und kann dabei ebenfalls durchgängig als hoch eingeschätzt werden. Die Bedeutsamkeit der konstruktiven Unterstützung für individuelle Schülermerkmale ergibt sich zudem dadurch, dass im vollständigen Modell auch der direkte Pfad von konstruktiver Unterstützung auf die fachbezogene Einstellung erstmals signifikant wird.

Damit stehen die Ergebnisse zur konstruktiven Unterstützung im Einklang mit den meisten bisherigen generellen Befunden zur konstruktiven Unterstützung, nach denen sich vor allem Zusammenhänge mit motivationalen Faktoren zeigen (Kunter & Voss, 2011; Sakiz et al., 2012). Im Gegensatz dazu zeichnet sich für differenzielle Untersuchungen bisher kein so eindeutiges Bild ab. Die bisherige Befundlage spiegelt gemischte Ergebnisse wider, die zum Teil eine besondere Bedeutsamkeit konstruk-

tiver Unterstützung für leistungsstarke Schüler:innen (Affandi et al., 2020; Sanfo & Malgoubri, 2021), zum Teil keine differenzielle Relevanz herausstellten (Lüftenegger et al., 2015). Mit Vorlage dieser Arbeit zeigt sich jedoch verstärkt die Tendenz, dass konstruktive Unterstützung zumindest indirekt als relevanter Prädiktor besonderer Leistungsstärke angesehen werden kann. Ob in Untersuchungen eher direkte Effekte auf Leistungen oder auf motivationale Merkmale gefunden wurden, kann mit der Operationalisierung der konstruktiven Unterstützung erklärt werden (Kunter & Trautwein, 2013). In TIMSS 2019 spiegelt die Skala zur konstruktiven Unterstützung insbesondere die Subfacette emotionale Unterstützung wider und weniger die inhaltliche Strukturierung. Damit ist der vorliegende Befund, dass die konstruktive Unterstützung insbesondere für die individuellen Schülermerkmale bedeutsam ist, plausibel, da der emotionalen Unterstützung vor allem eine Bedeutung für motivationale Merkmale zugesprochen wird, während die inhaltliche Strukturierung eher für die schulischen Leistungen relevant ist. Gleichzeitig sprechen die Ergebnisse dafür, dass – unter Berücksichtigung der Komplexität der schulischen Lehr- und Lernumwelten – die Annahme eines indirekten Zusammenhangs im Fall der konstruktiven Unterstützung im Kontext leistungsstarker Schüler:innen angemessener ist als die Annahme eines direkten Effektes.

Die Ergebnisse zur individuellen Förderung zeigen hingegen keinen direkten Zusammenhang mit der Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen. Stattdessen ergeben sich in den Hauptmodellen II und III indirekte Effekte über die fachbezogene Einstellung. Im vollständigen Gesamtmodell zeigen sich jedoch keine bedeutsamen Zusammenhänge mehr, da unter Kontrolle der Hintergrundvariablen sowohl der Pfad von individueller Förderung auf fachbezogene Einstellung als auch der von fachbezogener Einstellung auf Leistung nicht signifikant werden. Damit passen die Ergebnisse zur bisherigen generellen Forschung zu Formen der inneren Differenzierung, die als heterogen eingestuft werden kann und zwar viele signifikante positive (Hattie, 2009; Lou et al., 1996; Steenbergen-Hu et al., 2016), aber auch keine (Kulik & Kulik, 1992) oder signifikante negative Zusammenhänge (Horak, 1981) zwischen Differenzierung und Leistung bzw. individuellen Schülermerkmalen aufweist. Ähnlich sieht auch die Befundlage zu differenziellen Untersuchungen aus. In verschiedenen Studien wird eine besondere Bedeutung für die Leistungen sowohl von leistungsstarken (Inckemann, 2014; Kulik & Kulik, 1992) als auch leistungsschwachen Schüler:innen angenommen (Gehrer & Nusser, 2020; Lou et al., 1996). Die Ergebnisse dieser Untersuchung deuten darauf hin,

dass die individuelle Förderung als Form der inneren Differenzierung kein Prädiktor besonderer Leistungsstärke ist.

Die indirekten Zusammenhänge, die sich in Hauptmodell II und III zeigten, weisen zudem entgegen den Erwartungen negative Vorzeichen auf und deuten somit darauf hin, dass ein höheres Maß an individueller Förderung zu einer niedrigeren, also negativeren, fachbezogenen Einstellung (sowie zu einem niedrigeren fachbezogenen Selbstkonzept) führt. Da jedoch beide Teilpfade des indirekten Zusammenhangs auf Leistung über die fachbezogene Einstellung negativ ausfallen, würde ein höheres Maß an individueller Förderung in der Gesamtschau zu besseren Leistungen führen. Vor dem Hintergrund, dass die fachbezogene Einstellung möglicherweise im Kontext besonderer Leistungsstärke nicht die Ursache, sondern die Folge von Leistung ist, ist dieser indirekte Effekt jedoch mit Vorsicht zu interpretieren. Nichtsdestotrotz sind die Zusammenhänge für individuelle Förderung überraschend. Weder in der Operationalisierung von individueller Förderung und fachbezogener Einstellung noch in bisherigen Untersuchungen lassen sich eindeutige Erklärungsmöglichkeiten für diese Zusammenhänge ausmachen. Es lässt sich lediglich vermuten, dass die Einschätzung der individuellen Förderung durch die Schüler:innen möglicherweise nicht mit dem tatsächlichen Ausmaß an individueller Förderung im Unterricht zusammenhängt, da sie eventuell gar nicht alle Maßnahmen wahrnehmen oder als individuelle Förderung einstufen (Lenske, 2016; Schwantner, 2015).

Berichten Schüler:innen trotzdem von einem hohen Ausmaß individueller Förderung, könnte dies ein Anzeichen für eine besonders hohe Heterogenität in der Klasse sein, die dieses Vorgehen notwendig macht. Eine hohe Heterogenität in der Klassenzusammensetzung könnte gleichzeitig zu ungünstigeren motivationalen Ausprägungen bei Schüler:innen führen (für einen Einblick in den Diskurs zu Auswirkungen von Heterogenität und Kompositionsmerkmalen auf Schülermerkmale siehe beispielsweise Trautmann & Wischer, 2011 und Vock & Gronostaj, 2017). Da dies jedoch lediglich eine Vermutung für einen möglichen Erklärungsansatz darstellt, sollte dieser Befund entsprechend in weiteren Untersuchungen mit zusätzlichen potenziellen erklärenden Merkmalen näher in den Blick genommen werden, um fundiertere Aussagen treffen zu können.

Die Befunde zu den Merkmalen des Unterrichts verweisen also insbesondere auf die Bedeutsamkeit der kognitiven Aktivierung und der konstruktiven Unterstützung. Anders als erwartet, reduziert ein höheres Maß an kognitiver Aktivierung im Sachunterricht jedoch die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt

Kompetenzstufe IV zu erbringen. Auf Basis der Ergebnisse wird deshalb angenommen, dass in dieser Arbeit eher die Leistungen die Wahrnehmung der kognitiven Aktivierung bestimmt, als dass die kognitive Aktivierung als Prädiktor von Leistung fungiert. Deutlich höhere Werte ergeben sich für die konstruktive Unterstützung, die sich als sehr bedeutsamer indirekter Prädiktor besonderer Leistungsstärke herausstellte. Klassenführung ist hingegen durchgängig kein Prädiktor besonderer Leistungsstärke, für individuelle Förderung ergeben sich Hinweise auf einen möglichen Zusammenhang, die jedoch im Gesamtmodell nicht mehr signifikant ausfallen.

9.2.3 Merkmale der Lehrkraft

Die Ergebnisse zu den Merkmalen der Lehrkraft zeigen, dass die in dieser Arbeit ausgewählten Faktoren nur eingeschränkt bedeutsam sind für die Wahrscheinlichkeit, dass Schüler:innen Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV erbringen. Für die Berufszufriedenheit zeigt sich entgegen den Erwartungen, dass sie keinen direkten Effekt auf die naturwissenschaftliche Leistung der Schüler:innen hat. Damit kommt die vorliegende Untersuchung zu anderen Ergebnissen als beispielsweise Banerjee et al. (2017), Collie und Martin (2017) oder Arens und Morin (2016), die in ihren Arbeiten direkte Effekte von Lehrkräftemerkmalen wie Arbeitszufriedenheit, Wohlbefinden oder emotionaler Erschöpfung auf die Leistungen der Schüler:innen nachweisen konnten. Andererseits passen die Befunde dieser Arbeit zur längsschnittlichen Untersuchung von Caprara et al. (2006), die keinen direkten Effekt der Berufszufriedenheit auf die Leistungen nachweisen konnten. Möglicherweise ist hier die klare Abgrenzung der einzelnen verwandten Konstrukte Berufszufriedenheit, Arbeitszufriedenheit, Wohlbefinden und emotionale Erschöpfung ausschlaggebend dafür, ob sich in Untersuchungen signifikante direkte Zusammenhänge mit den Leistungen der Schüler:innen zeigen. Während unter Berufszufriedenheit eher eine übergeordnete, dauerhaftere Einstellung gegenüber der Profession verstanden werden kann (Bruggemann et al., 1975), können Arbeitszufriedenheit, Wohlbefinden sowie emotionale Erschöpfung als stärker veränderbare und anlassbezogenere Merkmale angesehen werden. Dies könnte ein Hinweis darauf sein, dass weniger die grundlegenden Haltungen bzw. Einstellungen von Lehrkräften zu ihrem Beruf ausschlaggebend für die Leistungen der Schüler:innen sind, sondern vielmehr die aktuelle Verfassung der Lehrkräfte. Gleichzeitig würde dies darauf hindeuten, dass

Berufs- und Arbeitszufriedenheit bei Lehrkräften, anders als von Ammann (2004) vermutet, möglicherweise doch nicht gleichgesetzt werden können.

Unter Kontrolle der Hintergrundmerkmale ergibt sich ein schwacher signifikanter Zusammenhang der Berufszufriedenheit mit der individuellen Förderung, der dafür spricht, dass zufriedenere Lehrkräfte ein höheres Maß an individueller Förderung in ihrem Unterricht realisieren können. Damit passen die Ergebnisse dieser Arbeit zu den Befunden von Kullmann (2012), der ebenfalls einen schwachen Zusammenhang zwischen beiden Merkmalen nachweisen konnte. Die meisten Befunde im Zusammenhang mit Unterrichtsmerkmalen konnten bisher zur Berufszufriedenheit und konstruktiven Unterstützung ausgemacht werden, wobei diese wiederholt einen positiven Zusammenhang zwischen beiden Merkmalen fanden (Lopes & Oliveira, 2020; Virtanen et al., 2019). In der vorliegenden Untersuchung zeigten sich erwartungswidrig in keinem der Modelle signifikante Zusammenhänge. Ein ähnliches Bild ergibt sich auch für die Klassenführung. Zwar ist hier die Befundlage deutlich geringer als für konstruktive Unterstützung, trotzdem wäre nach den Ergebnissen bisheriger Untersuchungen (z. B. Opdenakker & Van Damme, 2006) ein bedeutsamer positiver Zusammenhang plausibel gewesen. Anders sieht es für den Zusammenhang zwischen Berufszufriedenheit und kognitiver Aktivierung aus. Für diesen konnte in bisherigen Studien kein oder lediglich ein indirekter Zusammenhang über das Lehrer-Schüler-Verhältnis ausgemacht werden (Harrison et al., 2023; Perera et al., 2022), sodass auch für diese Untersuchung kein signifikanter direkter Effekt der Berufszufriedenheit auf die kognitive Aktivierung angenommen wurde. Die Zusammenhänge in den Modellen fallen erwartungsgemäß nicht signifikant aus.

Die Berufszufriedenheit kann in dieser Arbeit somit weder als relevanter Prädiktor besonderer Leistungsstärke noch als bedeutsamer Einflussfaktor von Unterrichtsmerkmalen angesehen werden. Dabei sollte der gefundene Effekt der Berufszufriedenheit auf die individuelle Förderung und insbesondere die unterschiedliche Signifikanz zwischen Hauptmodell III und IV nicht überinterpretiert werden. Dieser erreicht im vollständigen Gesamtmodell nur knapp das Signifikanzniveau von $p \leq .05$, während er in Hauptmodell III dieses Signifikanzniveau nur knapp verfehlt. Somit zeigen sich deutlich weniger große Unterschiede hinsichtlich der Bedeutsamkeit zwischen beiden Effekten, als aufgrund der etablierten Grenzwerte für Signifikanzniveaus angenommen werden könnte.

Für die Lehrkräfteselbstwirksamkeit ergibt sich ein ähnliches Bild wie für die Berufszufriedenheit. In dieser Untersuchung kann kein direkter Effekt der Lehrkräf-

teselbstwirksamkeit auf die Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstelle von Kompetenzstufe IV zu erbringen, nachgewiesen werden. Damit widersprechen die Ergebnisse den wenigen Untersuchungen, die direkte Zusammenhänge zwischen den beiden Merkmalen in den Blick genommen haben (Perera & John, 2020; Tella, 2008). Generelle Studien kamen außerdem zu dem Schluss, dass es bedeutsame Effekte auf die Unterrichtsmerkmale sowie indirekt auf motivationale und kognitive Merkmale der Schüler:innen gibt (z. B. Burić & Kim, 2020; Dixon et al., 2014; Fauth et al., 2019; Schiefele et al., 2013). In der vorliegenden Arbeit zeigen sich jedoch eher in Übereinstimmung mit Vieluf et al. (2013), Holzberger et al. (2013), Schüle et al. (2016) sowie mit Depaepe und König (2018) im vollständigen Gesamtmodell kaum signifikanten Zusammenhänge. Einzig für die kognitive Aktivierung konnte ein signifikanter Effekt der Lehrkräfteselbstwirksamkeit festgestellt werden.

Mit dem Effekt auf kognitive Aktivierung ergibt sich im vollständigen Gesamtmodell ein indirekter Effekt der Lehrkräfteselbstwirksamkeit auf die Leistung, vermittelt über die kognitive Aktivierung. Selbstwirksamere Lehrkräfte realisieren somit in ihrem Unterricht ein höheres Maß an wahrgenommener kognitiver Aktivierung, was wiederum zu einer reduzierten Wahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen, führt. Insgesamt kann dieser indirekte Zusammenhang als schwach eingestuft werden. Wie der negative Effekt der kognitiven Aktivierung interpretieret werden könnte und einzuordnen ist, wurde bereits in Kapitel 9.2.2 erläutert. Ob die Interpretation, dass eine höhere Lehrkräfteselbstwirksamkeit indirekt zu einer geringeren Wahrscheinlichkeit beiträgt, zulässig ist, ist somit fraglich und Bedarf näherer Untersuchung.

Bei den bisherigen Erläuterungen zu den beiden Lehrkräftemerkmalen Berufszufriedenheit und Lehrkräfteselbstwirksamkeit muss berücksichtigt werden, dass alle herangezogenen Befunde aus generellen Untersuchungen stammen. Im Bereich der Lehrkräfte konnte bisher nur eine einzige differenzielle Untersuchung ausgemacht werden, die auch lediglich Hinweise darauf enthält, dass motivationale Orientierungen und selbstregulative Fähigkeiten für besondere Leistungsstärke relevant sein könnten (Rowen et al., 1997). Die Befunde dieser Arbeit bestätigen diese Einschätzung jedoch nicht.

9.2.4 Zusammenführung vor dem Hintergrund des Rahmenmodells

Im letzten Schritt werden die Ergebnisse dieser Arbeit in Bezug auf das zugrundeliegende Rahmenmodell zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen aus Kapitel 3 eingeordnet. Verschiedene Untersuchungen kamen zu dem Schluss, dass der bedeutsamste Bereich dieses Modells für schulische Leistungen die individuellen Schülermerkmale ist, gefolgt von Merkmalen des Unterrichts und der Lehrkräfte (Hattie, 2009; Köller & Baumert, 2018; Lipowsky, 2006; Wang et al., 1993). Die Ergebnisse der vorliegenden Arbeit, in der sich auf diese drei Bereiche fokussiert wurde, bestätigen diese Rangfolge. Über die verschiedenen berechneten Modelle hinweg kommt den individuellen Merkmalen der Schüler:innen sowie ihren Hintergrundmerkmalen die größte Bedeutung zu. Für den Unterricht zeigen sich ebenfalls bedeutsame Zusammenhänge mit der Leistung, wobei die Bedeutung der einzelnen Merkmale zumeist geringer ausfällt und stärker über die Modelle hinweg variiert. Die untersuchten Lehrkräftemerkmale stellen sich wiederum als deutlich weniger bedeutsam heraus. In den meisten Modellen zeigen sich keine signifikanten Zusammenhänge zwischen Merkmalen der Lehrkraft und den anderen betrachteten Bereichen, erst unter Kontrolle von Hintergrundmerkmalen der Schüler:innen ergeben sich zum Teil bedeutsame Zusammenhänge mit den Unterrichtsmerkmalen. Dabei muss jedoch berücksichtigt werden, dass im Bereich Lehrkräfte nur zwei Merkmale betrachtet wurden, die als eher subjektive, personelle Merkmale eingestuft werden können (siehe Kapitel 9.3 für weitere Überlegungen zur Auswahl der Lehrkräftemerkmale). Somit konnte mit der vorliegenden Arbeit nur ein kleiner Ausschnitt des Bereiches Lehrkräfte abgedeckt werden.

Nach dem Rahmenmodell hängen die drei Bereiche sowohl direkt als auch indirekt zusammen. Konkret wird ein indirekter Zusammenhang zwischen Lehrkräftemerkmalen und Schülerleistungen angenommen, der über den Unterricht und individuelle Schülermerkmale vermittelt wird. Die Ergebnisse dieser Untersuchung bestätigen, dass sich auch für Prädiktoren besonderer Leistungsstärke übereinstimmend mit dem Modell indirekte Zusammenhänge zwischen Lehrkräfte- bzw. Unterrichtsmerkmalen und schulischen Leistungen zeigen. Anders als im Modell postuliert, konnten allerdings auch direkte Zusammenhänge zwischen Unterrichtsmerkmalen und Leistungen nachgewiesen werden. Bisher werden jedoch ausschließlich entgegengesetzte Effekte der schulischen Leistungen auf schulinterne Bedingungen, zu denen auch der Unter-

richt gehört, im Rahmenmodell abgebildet. Eine Überprüfung der Zusammenhänge, die im Rahmenmodell für die gesamte Schülerschaft angenommen werden, erscheint entsprechend für einzelne Schülergruppen wie den leistungsstärksten Schüler:innen angezeigt, insbesondere hinsichtlich ihrer generellen Bedeutung sowie möglicher direkter sowie entgegengesetzt gerichteter Zusammenhänge (wie beispielsweise im Fall der fachbezogenen Einstellung und der Leistung), die sich in der vorliegenden Arbeit gezeigt hat.

9.3 Kritische Reflexion der vorliegenden Arbeit

In diesem Kapitel wird wie eingangs beschrieben die vorliegende Arbeit insbesondere hinsichtlich methodischer Aspekte und Limitationen aber auch Potenziale kritisch betrachtet. Einige mögliche Kritikpunkte, wie beispielsweise die Auswahl der Lehrkräftemerkmale, wurden bereits in den vorherigen Kapiteln angeschnitten. In diesem Kapitel werden die verschiedenen Aspekte im Gesamten betrachtet und eingeordnet.

Die größte Limitation, aber auch gleichzeitig der größte Vorteil, dieser Arbeit liegt in der Durchführung von Sekundäranalysen unter Nutzung bereits bestehenden Daten. Die Nutzung von Sekundärdaten ermöglicht es, neue Erkenntnisse zu generieren, ohne dass Schulen mit zusätzlichen Erhebungen belastet werden. Stattdessen werden bestehende Daten genutzt, sodass Forschungen ökonomisch und mit geringerem zeitlichem Aufwand durchgeführt sowie bereits erhobene Datensätze besser ausgeschöpft werden können. Für diese Arbeit wurden die Daten der TIMSS-2019-Untersuchung genutzt. Diese bieten den großen Vorteil, dass mit ihnen repräsentative Aussagen über die Grundgesamtheit aller Viertklässler:innen in den teilnehmenden Staaten und somit auch für Deutschland getroffen werden können. Gleichzeitig lassen umfangreiche Fragebögen für Lehrkräfte, Schulleitungen und Erziehungsberechtigte, die zusätzlich zum Test zur Erfassung der mathematischen und naturwissenschaftlichen Kompetenzen erhoben werden, die Untersuchung bzw. Berücksichtigung zahlreicher Hintergrundmerkmale in Analysen zu, sodass die angestrebten komplexen Modelle mit den TIMSS-Datensätzen durchführbar waren. Die Durchführung von Kompetenztests zur Erfassung der schulischen Leistungen bietet zudem einen weiteren großen Vorteil für die vorliegende Untersuchung. Diese Vorgehensweise ermöglichte es, den Analysen ein objektives Kompetenzmaß zugrunde legen zu können – anders

als bei der Nutzung von Noten, die weitere, z. T. subjektive Komponenten enthalten (Maaz et al., 2008; Valtin, 2012).

Durch die Einordnung der erzielten Kompetenzwerte der Schüler:innen auf sogenannten Kompetenzstufen bietet die TIMS-Studie darüber hinaus den Vorteil, eine über verschiedene Zyklen und Teilnehmerstaaten hinweg einheitliche Unterteilung der Schüler:innen in unterschiedlich leistungsstarke Gruppen zur Verfügung zu stellen. Dadurch bieten die Daten für die Fragestellung dieser Arbeit bereits Gruppenzuordnungen und klare Definitionen der einzelnen Kompetenzstufen, die auch für zukünftige Untersuchungen zu (potenziell) leistungsstarken Schüler:innen genutzt werden können. Durch diese einheitliche Definition können Befunde verschiedener Untersuchungen zu leistungsstarken Schüler:innen besser miteinander in Verbindung gebracht werden, als es bei individuell gewählten Einstufungen, wie sie zur Zeit noch häufig vorkommen (Neuendorf, 2022), der Fall ist. Ein weiterer Vorteil der TIMSS-Daten liegt in ihrem Umfang und ihrer Struktur. Durch die Ziehung ganzer Klassen in einer Schule und der zusätzlichen Befragung der dazugehörigen Lehrkräfte und Schulleitungen sowie der Erziehungsberechtigten entsteht neben einem Schülerdatensatz auch je ein Datensatz für Lehrkräfte, Schulleitungen und Erziehungsberechtigte mit vielfältigen Informationen. Gleichzeitig ist es möglich, die Angaben der verschiedenen Informationsquellen einander zuzuordnen, sodass umfangreiche Datensätze entstehen, die eine Clusterstruktur und Informationen zur Gruppenzugehörigkeit beinhalten und somit Analysen mit Merkmalen verschiedener Ebenen zulassen. Dabei muss jedoch berücksichtigt werden, dass die Lehrkräftestichprobe aufgrund des Ziehungsverfahrens nicht repräsentativ ist. Für generalisierbare Aussagen müssen die Informationen über die Lehrkräfte somit auf die Schüler:innen bezogen werden.

Die Nutzung von Sekundärdaten bringt jedoch auch einige Nachteile mit sich. So ist der verfügbare Datensatz in der Regel weniger aktuell als ein Primärdatensatz und die Passgenauigkeit der enthaltenen Informationen für die zu untersuchende Forschungsfrage nur selten hundertprozentig gegeben. Der in dieser Arbeit genutzte Datensatz von TIMSS 2019 ist zum Zeitpunkt der Analysen noch recht aktuell, aufgrund der Zeit, die zwischen Erhebung und Veröffentlichung der Arbeit liegt, muss jedoch bei der Einordnung der Ergebnisse insbesondere ein Aspekt berücksichtigt werden. In Folge der COVID-19-Pandemie kam es zu wiederholten Lockdowns und damit Schulschließungen, wodurch sich der Schulalltag der Schüler:innen massiv veränderte. Die Daten von TIMSS 2019 wurden jedoch im Frühjahr 2019 und somit vor der COVID-19-Pandemie erhoben. Die Daten sind somit nicht durch die Pandemie

beeinflusst, allerdings muss in Betracht gezogen werden, dass die Erkenntnisse die (aktuellen) Verhältnisse nach der Pandemie möglicherweise nicht vollständig widerspiegeln.

Damit zeigt sich ein weiterer Nachteil der TIMSS-Daten: Auch wenn die Studie alle vier Jahre durchgeführt wird, stellt sie keine Längsschnittuntersuchung dar, da nicht dieselben Schüler:innen wiederholt befragt werden, sondern in jedem Zyklus eine neue Kohorte von Viertklässler:innen im Fokus steht. Damit liegen der Untersuchung in dieser Arbeit Querschnittsdaten zugrunde, auf Basis derer kausale Annahmen nicht empirisch überprüft werden können. Mit Querschnittsdaten ist es lediglich möglich zu testen, ob die aufgestellten Hypothesen verworfen werden müssen oder nicht (Reinecke, 2014). Die Ergebnisse können somit Hinweise darauf geben, ob die theoretisch hergeleiteten Zusammenhänge anhand der Daten als plausibel angesehen werden können, allerdings kann die Wirkrichtung eines Zusammenhangs nicht bestimmt werden. Es muss also in Betracht gezogen werden, dass zwei Merkmale nicht wie erwartet, sondern entgegengesetzt zusammenhängen. Definitive Aussagen darüber können lediglich mit Längsschnittuntersuchungen getroffen werden. Für diese Arbeit bedeutet das, dass die gefundenen Zusammenhänge, deren Richtung auf Basis theoretischer Überlegungen und bisheriger Befunde angenommen wurde, prinzipiell auch entgegengesetzt möglich sind. Insbesondere beim Zusammenhang von Leistung und fachbezogener Einstellung könnte dies eine Erklärung für die gefundenen Zusammenhänge sein. Darüber hinaus ist aber auch vorstellbar, dass die Lehrkräftemerkmale Berufszufriedenheit und Lehrkräftemerkmale nicht (nur) als Prädiktoren von Unterrichtsmerkmalen fungieren. So kann beispielsweise auch angenommen werden, dass die Berufszufriedenheit nicht die Voraussetzung für individuelle Förderung im Unterricht ist, sondern, dass die Möglichkeit, ein hohes Maß an individueller Förderung im Unterricht umzusetzen und positive Auswirkungen auf die Kinder zu erleben, die Ursache für eine höhere Berufszufriedenheit der Lehrkräfte darstellt.

Wie zuvor beschrieben, können Sekundärdaten darüber hinaus den Nachteil haben, dass sie nur bedingt auf die zu untersuchende Fragestellung passen. Dies trifft zum Teil auch auf die in dieser Arbeit genutzten Daten von TIMSS 2019 zu. Bereits erwähnt wurde, dass die TIMSS-Daten mit der Einteilung der Schüler:innen auf Kompetenzstufen bereits eine gute Grundlage für leistungsgruppenspezifische Analysen enthalten. Darüber hinaus beinhaltet der Datensatz zahlreiche Hintergrundvariablen, die zu vertieften Analysen der Leistungen herangezogen werden können. Es gibt konkret für die vorliegende Arbeit jedoch insbesondere zwei nachteilige Aspekte, die

berücksichtigt werden sollten: zum einen gibt es ein paar Merkmale, deren Untersuchung im Rahmen dieser Forschungsfrage hätte sinnvoll sein können, die jedoch nicht im Datensatz enthalten waren. Zum anderen gibt es einige Skalen, die zwar vorhanden sind, aber hinsichtlich ihrer Konzeptualisierung Nachteile aufweisen.

Zum ersten Aspekt gehört insbesondere eine Kategorie von Informationen, die nicht (in geeigneter Weise) mit den TIMSS-Daten abgedeckt wird, aber für die Untersuchung leistungsstärkster Schüler:innen von Interesse wäre – auch um noch anschlussfähiger an bestehende Ansätze zur Förderung leistungsstarker bzw. hochbegabter Schüler:innen zu sein. Dabei handelt es sich um Informationen zur Teilnahme von leistungsstarken Schüler:innen an etablierten Förderansätzen wie Enrichment (Anreicherung bestehender Angebote) oder Akzeleration (beschleunigtes Durchlaufen von Angeboten). Da diese Strategien Fördermöglichkeiten darstellen, die auch innerhalb des Unterrichts bzw. innerhalb der Schule umgesetzt werden können, hätte eine Betrachtung dieser Merkmale entweder als zusätzlicher Prädiktor oder als Kontrollvariable wertvolle Hinweise darauf liefern können, inwiefern eine Teilnahme an solchen Angeboten besondere Leistungsstärke bedingt und in welchem Verhältnis die Stärke eines möglichen Zusammenhangs im Verhältnis zu der anderer Merkmale steht. Damit hätte zudem eine Erkenntnis hinsichtlich der Bedeutsamkeit dieser eher aus der Hochbegabtenförderung stammenden Strategien für die Förderung von Leistungen und damit ein zusätzliches Argument für oder gegen die Übertragbarkeit dieser Methoden gewonnen werden können.

Zum zweiten Aspekt finden sich Kritikpunkte insbesondere hinsichtlich der Unterrichtsskalen. Die größte Problematik zeigt sich bei der Skala zur kognitiven Aktivierung. Wie bereits in Kapitel 8.1 ausführlicher erläutert, decken die Einzelitems die verschiedenen Facetten der kognitiven Aktivierung nicht in der gleichen Gewichtung ab. Dies resultierte in teils unzureichenden Faktorladungen, sodass Items für die Skalenbildung entfernt werden mussten, was zwar die Ungleichgewichtung weiter verstärkte, aber zu ausreichenden statistischen Kennwerten führte. In den verschiedenen Modellschritten veränderten sich die Ladungen der Items der Skala zur kognitiven Aktivierung (andere Skalen waren nicht betroffen) allerdings zum Teil so stark, dass einzelne Items erneut unzureichende Ladungen aufwiesen (siehe Anhang). Die Ergebnisse zur kognitiven Aktivierung sollten somit nur mit Vorsicht interpretiert werden, da die Skala sowohl inhaltliche als auch methodische Defizite aufweist. Eine Revision der Skala für eine höhere Ausgewogenheit sollte in Betracht gezogen werden. Um den Trend seit TIMSS 2015 nicht zu gefährden, sollte dies jedoch vorrangig über

die Ergänzung von Items aus den unterrepräsentierten Subfacetten realisiert werden, sodass die bisherige Operationalisierung für Trendanalysen weiterhin möglich ist.

Ein Ungleichgewicht hinsichtlich der beiden theoretisch angenommenen Subfacetten (emotionale Unterstützung und inhaltliche Strukturierung) zeigt sich auch bei der Skala zur konstruktiven Unterstützung. Diese spiegelt ausschließlich Aspekte der emotionalen Unterstützung wider, sodass eine eindeutigere Benennung der Skala angebracht wäre, um eine klare theoretische Verortung zu erreichen. Auch bei dieser Skala sollte darüber hinaus die Ergänzung von Einzelitems zur inhaltlichen Strukturierung für eine höhere konzeptuelle Passung zur Theorie und für die Trenderhaltung in Erwägung gezogen werden. Für die vorliegende Arbeit bedeutet das, dass, anders als in den Hypothesen formuliert, die konstruktive Unterstützung nicht vollumfänglich untersucht wurde. Stattdessen sind die Ergebnisse nur für die Subfacette emotionale Unterstützung gültig und können nicht ohne weitere Überprüfung auf die inhaltliche Strukturierung und damit die konstruktive Unterstützung im Gesamten übertragen werden. So ist gleichzeitig der Aspekt der inhaltlichen, fachlichen Unterstützung durch die Lehrkraft, bei der eher ein Zusammenhang mit besonderer Leistungsstärke angenommen werden könnte als mit der emotionalen Unterstützung, nicht untersucht worden, sodass keine Aussagen über die Bedeutung dieser Facette getroffen werden können. Eine Betrachtung der inhaltlichen Strukturierung scheint jedoch weiterhin vielversprechend.

Ein letzter Kritikpunkt lässt sich im weiteren Sinne beiden Aspekten zuordnen. Zwar werden die relevanten Unterrichtsmerkmale im Rahmen des Schülerfragebogens recht umfassend erhoben, allerdings finden sich auch in anderen Fragebögen in TIMSS Informationen zum Unterricht. Im Fragebogen der Lehrkräfte werden allerdings nicht alle Faktoren des Unterrichtens, die in den Blick genommen werden sollten, erfragt, sodass eine eingeschränkte oder gemischte Form hinsichtlich der Informationsquellen hätte gewählt werden müssen. Um die Ergebnisse zu den Unterrichtsmerkmalen jedoch möglichst gut vergleichen zu können, wurden in dieser Arbeit ausschließlich die Informationen aus nur einem Fragebogen, dem Schülerfragebogen, genutzt. Dieses Vorgehen führte allerdings gleichzeitig dazu, dass eine möglicherweise vielversprechende Skala des Lehrkräftefragebogens aus dem Bereich der inneren Differenzierung nicht in den Analysen berücksichtigt wurde. In dieser wird von den Lehrkräften erfragt, inwiefern leistungsstärkere Schüler:innen in ihrem Unterricht angepasste Aufgaben und Materialien erhalten. Möglicherweise könnte das selbstberichtete Vorgehen der Lehrkräfte im Gegensatz zur wahrgenommenen individuellen Förderung ein ange-

messenerer Prädiktor für besondere Leistungsstärke sein. Mit dieser Überlegung geht zudem einher, dass bei Differenzierung im Unterricht insbesondere die Perspektive der Lehrkraft von Interesse sein könnte, da Schüler:innen möglicherweise gar nicht in vollem Umfang mitbekommen, wie häufig und auch welche Weise differenzierende Maßnahmen im Unterricht eingesetzt werden, sodass Lehrkräfte für diese Information die zuverlässigere Quelle darstellen (Lenske, 2016; Schwantner, 2015). Der Umstand, dass die erfragte Wahrnehmung der Schüler:innen möglicherweise nicht mit der objektiven Realität übereinstimmt und deshalb weniger verlässlich ist, muss bei der Interpretation der Befunde zur individuellen Förderung berücksichtigt werden.

Die umfangreiche Datengrundlage in TIMSS, die die Untersuchung der komplexen Modelle in dieser Arbeit erst möglich machte, geht jedoch mit einem potenziellen weiteren Nachteil einher. Um diese große Menge an Informationen zu generieren, müssen die Fragebögen eine hohe Zahl an Fragen beinhalten. Dadurch dauert die Bearbeitung der Fragebögen beispielsweise für die Schüler:innen ca. 45 Minuten, was dazu führen kann, dass die Aufmerksamkeit oder aber auch die Bereitschaft zur sorgfältigen Bearbeitung mit fortschreitendem Fragebogen nachlassen. Dies kann dazu führen, dass die Aussagen zu Fragen, die weiter hinten im Fragebogen positioniert sind, weniger verlässlich sind bzw. die Anzahl an *missings* deutlich höher ausfällt als bei Fragen zu Beginn der Befragung (Positionseffekt; siehe beispielsweise Sedlmeier & Renkewitz, 2018). Die *missing patterns*, die in Kapitel 7.1.2 untersucht wurden, deuten darauf hin, dass dies bei der vorliegenden Untersuchung für den Schülerfragebogen zutrifft: Je später die Fragen im Fragebogen auftauchen, desto größer ist die Anzahl der *missings*. Das bedeutet, dass prinzipiell die Informationen zu den individuellen Schülermerkmalen sowie den Kontrollvariablen als verlässlicher eingestuft werden können, als die Informationen zum Unterricht. Diese sind Teil der nationalen Ergänzung und entsprechend in der zweiten Hälfte des Fragebogens aufgeführt. Da jedoch im Vergleich zu der Anzahl gültiger Werte zu Beginn des Fragebogens noch immer ca. 90 % aller Personen auch die Fragen zu Unterrichtsmerkmalen beantwortet haben und die Analysen unter Einsatz des FIMLs erfolgten, können die Informationen trotzdem als zuverlässig eingestuft werden.

Ebenfalls bedeutsame Limitationen ergeben sich auch durch die Anlage dieser Untersuchung und der dabei durchgeführten Analysen. Dies betrifft insbesondere die Maßnahmen zum Umgang mit der Komplexität der Berechnungen und die Auswahl von und den Umgang mit den verschiedenen Skalen.

Ersteres bezieht sich vor allem auf die Durchführung multinomial logistischer Regressionen unter Nutzung numerischer Integration im Rahmen von Mehrebenenstrukturgleichungsmodellen mit multipler Imputation. Diese Kombination verschiedener Methoden führte dazu, dass nur eine sehr eingeschränkte Anzahl an Kennwerten, vor allem zur objektiven Bewertung der Modellgüte und der Bedeutsamkeit der Analysen, zur Verfügung steht. So konnten für die Vor- und Hauptmodelle beispielsweise weder absolute noch relative Fit-Indizes, wie der RMSEA, SRMR und CFI, noch Maße der Varianzaufklärung, wie R^2 oder Nagelkerkes R^2, berechnet werden. Damit können kaum absolute Aussagen über die Ergebnisse der vorliegenden Untersuchung und ihre Bedeutung getroffen werden, was eine Einordnung lediglich relational möglich macht.

Des Weiteren bezieht sich der erste Überpunkt auch auf die Reduktion der Integrationspunkte (siehe Kapitel 7.1.2). Diese wurde vorgenommen, um die geplanten Berechnungen trotz ihrer Komplexität realisieren zu können. Eine solche Reduktion ist zwar zulässig, gleichzeitig wird dadurch jedoch in Kauf genommen, dass die Ergebnisse der Analysen im Verhältnis zu Berechnungen mit den empfohlenen Integrationspunkten weniger genau ausfallen können (Muthén & Muthén, 2017; Rabe-Hesketh & Skrondal, 2012). Dementsprechend ist dieser Umstand insbesondere für die Ergebnisse relevant, die bestimmte Richtwerte (wie z. B. Signifikanzniveaus) nur knapp erreichen oder verfehlen, da möglich ist, dass diese bei der Berechnung mit mehr Integrationspunkten anders ausfallen. Somit sollten speziell diese Ergebnisse (z. B. zum Zusammenhang zwischen Berufszufriedenheit und individueller Förderung) zwar mit Vorsicht interpretiert werden, trotzdem bieten sie Hinweise auf mögliche Zusammenhänge, die in fokussierten Analysen unter Nutzung der empfohlenen Anzahl an Integrationspunkten überprüft werden sollten.

Die Komplexität der Berechnungen führte zusätzlich dazu, dass keine Mehrebenenanalysen mit *Random-Slopes* oder Interaktionseffekten durchgeführt wurden. Damit konnte das Potenzial dieser Methode im Rahmen dieser Arbeit zwar nicht voll ausgeschöpft werden, allerdings waren diese Modellschritte zur Beantwortung der Forschungsfrage auch aus inhaltlichen Gründen nicht nötig (siehe Kapitel 7.3.3). Aus forschungspragmatischen Gründen wurde zudem versucht, das Hauptmodell in seiner Komplexität zu reduzieren sowie die Zahl der zu berechnenden Modelle möglichst gering zu halten. Dazu wurden zum einen in zwei einfacheren Vormodellen die direkten Effekte von Lehrkräfte- und Unterrichtsmerkmalen auf die Leistung modelliert, um anschließend nur die signifikanten Pfade in das Gesamtmodell zu übernehmen. Zum anderen wurde beim schrittweisen Aufbau des Gesamtmodells

darauf verzichtet, auch die Zwischenschritte unter Kontrolle der Hintergrundmerkmale zu rechnen. Diese beiden Maßnahmen können allerdings zu einem möglichen Informationsverlust führen. So wurden im Gesamtmodell keine direkten Effekte von Lehrkräftemerkmalen sowie zwei Unterrichtsmerkmalen modelliert, sodass mögliche Änderungen in der Signifikanz im Zusammenspiel mit den anderen betrachteten Merkmalen nicht festgestellt werden konnten. Ebenso sind Informationen über die Bedeutsamkeit der verschiedenen Hintergrundmerkmale nur für das vollständige Gesamtmerkmal vorhanden, sodass auch hier keine Veränderungen zwischen den verschiedenen Modellschritten beobachtbar sind. Es sollte in Betracht gezogen werden, die verschieden Aspekte in weniger komplexen Modellen erneut genauer in den Blick zu nehmen, um sicherzustellen, dass keine bedeutsamen Erkenntnisse verloren gehen.

Zum zweiten Aspekt, der Auswahl von und dem Umgang mit Skalen, lassen sich wiederum insbesondere drei konkrete Punkte ausmachen. Der erste Punkt, der diesem Bereich zugeordnet werden kann, bezieht sich auf die Auswahl bestimmter Bereiche im ersten Schritt dieser Arbeit (Kapitel 3). Auf Basis von Metaanalysen und Übersichtsarbeiten wurden dabei die bedeutsamsten Bereiche des Rahmenmodells zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen für die schulischen Leistungen identifiziert und der Fokus dieser Arbeit entsprechend ausgewählt. Möglicherweise bieten für diese spezielle Zielgruppe andere Bereiche des Modells weitere vielversprechende oder sogar bedeutsamere Prädiktoren. Insbesondere die Bereiche, die im Zusammenhang mit dem familiären Hintergrund der Kinder stehen (soziales und kulturelles Kapitel sowie elterliches Erziehungs- und Unterstützungsverhalten), könnten für die Leistungen bedeutsam sein. So konnten beispielsweise Hoover-Dempsey et al. (2005) oder Lehrl, Evangelou und Sammons (2020) zeigen, dass das Unterstützungsverhalten der Eltern, aber auch ihre Ambitionen positiv mit den Leistungen der Schüler:innen zusammenhängen. Darüber hinaus fällt dieser Zusammenhang umso höher aus, je jünger die Kinder sind. Da in dieser Arbeit Kinder im Grundschulalter im Fokus stehen, könnten also möglicherweise die Eltern ein ausschlaggebender Prädiktor für besondere Leistungsstärke sein.

Der zweite Punkt steht ebenfalls im Zusammenhang mit den Bereichen, allerdings konkret mit dem der Lehrkräfte und den ausgewählten Merkmalen. Wie im vorangegangenen Kapitel beschrieben, könnte sich die geringe Bedeutung des Bereiches der Lehrkräfte in dieser Arbeit durch die (wenigen) ausgewählten Merkmale erklären lassen. Die Untersuchung bzw. Hinzunahme weiterer Lehrkräftemerkmale könnten möglicherweise zu anderen Ergebnissen und somit zu einer veränderten Einschätzung

der Bedeutsamkeit der Lehrkräftemerkmale führen. Da sich in der bisherigen Literatur jedoch keine Hinweise auf eine Bedeutsamkeit anderer Lehrkräftemerkmale (wie z. B. Professionswissen oder Berufserfahrung) für besondere Leistungsstärke zeigten, wurden diese Merkmale nicht für die Untersuchung herangezogen. Der Grund für die fehlenden Hinweise ist nach der bisherigen Forschungslage allerdings eher in der geringen Anzahl an differenziellen Untersuchungen zu sehen, sodass ein breiterer Blick auch auf generelle Prädiktoren schulischer Leistung zum jetzigen Stand der Forschung angebracht ist. So konnte beispielsweise die Bedeutung von Merkmalen wie Professionswissen, Ausbildung oder Berufserfahrung für schulische Leistungen allgemein bereits mehrfach nachgewiesen werden (beispielsweise bei Baumert et al., 2010; Darling-Hammond, 2000), sodass auch eine Untersuchung im Kontext besonderer Leistungsstärke vielversprechend sein kann, um bedeutsame Merkmale der Lehrkräfte zu identifizieren. Dabei muss jedoch berücksichtigt werden, dass die Interklassenkorrelation zwar für einen bedeutsamen Anteil der Varianz auf Level 2 spricht (siehe Kapitel 7.1.3), allerdings umfasst dieser nur etwas mehr als 10 %, sodass der Großteil der Varianz durch Prädiktoren auf Schülerebene erklärt werden kann. Somit ist die Bedeutsamkeit der Lehrkräftemerkmale insgesamt als geringer für das Erreichen besonderer Leistungsstärke einzustufen als die von Schülermerkmalen.

Der dritte Punkt in diesem Abschnitt betrifft die Erfassung der Unterrichtsmerkmale über Aussagen der Schüler:innen sowie deren anschließende Aggregation auf Level 2. Die Gründe über die Auswahl der Unterrichtsskalen wurden bereits zu Beginn des Kapitels erläutert. Die Aggregation wurde vorgenommen, da dieses konkrete Forschungsfeld noch am Anfang steht und bisher wenig über den Einfluss von Unterrichtsmerkmalen bekannt ist. Deshalb erschien es sinnvoll, diese als Merkmal der Klasse zu betrachten, um ein möglichst objektives Maß der Unterrichtsqualität und -gestaltung zu erhalten. Gleichzeitig sind die Erkenntnisse dieser Arbeit so besser mit Aussagen anderer Studien vergleichbar, die auf Informationen externer Beobachter:innen und Lehrkräfte basieren und die ebenfalls den Unterricht an sich und nicht dessen Wahrnehmung erfassen. In einem nächsten Schritt wäre jedoch die Betrachtung der subjektiv wahrgenommenen Unterrichtsmerkmale interessant (für weitere Überlegungen dazu siehe Kapitel 10.1).

Insbesondere im Vergleich verschiedener Bereiche und in der umfangreichen Berücksichtigung möglicher Einflussfaktoren, die die Komplexität schulischer Lehr- und Lernumwelten abbilden, liegt wiederum eine zentrale Stärke dieser Arbeit. Dieses Vorgehen ermöglicht es, in diesem bisher wenig beforschten Feld einen ersten grö-

ßeren Überblick und damit Ansätze für vertiefende Forschungen zu liefern. Darüber hinaus können mit dem differenziellen Ansatz adäquatere Forschungsergebnisse zu Prädiktoren besonderer Leistungsstärke als bei generellen Untersuchungen präsentiert werden. Dabei muss jedoch berücksichtigt werden, dass in dieser Arbeit nur die direkten auf die Leistung modellierten Effekte tatsächlich differenzielle Befunde darstellen. Direkte Effekte zwischen den anderen untersuchten Bereichen sind hingegen als generelle Befunde einzustufen, da die Differenzierung erst durch den Bezug auf multinomial logistische Zielvariable zustande kommt. Indirekte Effekte sind somit eine Mischform, da sie sich aus linearen Regressionen ohne konkrete Differenzierung und multinomial logistischen Regressionen mit Differenzierung zusammensetzen. Eine zwar etwas anders gelagerte, aber dennoch vielversprechende Herangehensweise, mit der durchgängig differenzielle Befunde generiert werden, könnten multiple logistische Regressionen oder aber ein Modellvergleich sein, bei dem das gleiche Modell für unterschiedliche Fähigkeitsgruppen berechnet wird (siehe Kapitel 10.1).

Insgesamt gehört diese Arbeit mit allen erläuterten Vor- und Nachteilen zu den umfangreichsten und komplexesten Untersuchungen in diesem Feld in der Erziehungswissenschaft, die darüber hinaus auf einer repräsentativen Datengrundlage basiert. Dabei bietet sie sowohl die Überprüfung theoretischer Annahmen (insbesondere denen des Rahmenmodells) als auch Generierung von Befunden zu einem Forschungsgegenstand, der nicht nur national, sondern auch international bisher sehr wenig beforscht wurde. Weiterführende Ansätze zur Erforschung besonderer Leistungsstärke werden im anschließenden Kapitel erläutert.

10 Fazit

In dieser Arbeit wurde der Frage nachgegangen, was die leistungsstärksten Schüler:innen von potenziell leistungsstarken Schüler:innen unterscheidet und welche Prädiktoren aus den Bereichen individuelle Schülermerkmale, Unterricht und Lehrkräfte dazu beitragen, dass Schüler:innen nicht nur Kompetenzen auf Kompetenzstufe IV, sondern auf Kompetenzstufe V erbringen. Dazu wurden Mehrebenenstrukturgleichungsmodelle mit multinomial logistischen und linearen Regressionen gerechnet, sodass verschiedene Merkmale aus unterschiedlichen Bereichen adäquat auf ihre Bedeutsamkeit hin untersucht werden konnten. In diesem Kapitel wird nun abschließend beleuchtet, welche Bedeutung die Ergebnisse dieser Untersuchung haben. Dazu werden zunächst weitere Forschungsansätze erläutert, die sich aus dieser Arbeit und den Ergebnissen der Analysen ergeben (Kapitel 10.1). Abschließend werden Überlegungen präsentiert, welche Relevanz die Ergebnisse insbesondere für die pädagogische Praxis haben (Kapitel 10.2).

10.1 Ausblick für weitere Forschungen

Im Rahmen der vorliegenden Arbeit konnten verschiedene inhaltliche wie methodische Ansatzpunkte identifiziert werden, deren weiterführende Untersuchung angezeigt scheint. Im Folgenden werden zunächst inhaltliche Ansatzpunkte für weitere Forschungen erläutert, bevor mögliche alternative bzw. weiterführende methodische Herangehensweisen zur Bearbeitung der Forschungsfrage sowie einzelner auffälliger Befunde dieser Arbeit dargestellt werden. Abschließend werden einige Ansatzpunkte für weiterführende Untersuchungen im Bereich der Methodenforschung benannt.

Während der inhaltlichen Vorarbeiten zu dieser Arbeit zeigte sich insbesondere ein Defizit in der theoretischen Systematisierung und Nomenklatur nicht-kognitiver Merkmale. So ist die Bezeichnung ‚nicht-kognitiv' zur Beschreibung aller individuellen Merkmale, die nicht explizit den kognitiven Faktoren zugeordnet werden können, zwar relativ etabliert, sie impliziert jedoch einerseits fälschlicherweise die Abwesenheit kognitiver Facetten und lässt andererseits darauf schließen, dass das gemeinsame Element dieser Merkmale bisher nicht eindeutig bestimmt wurde. Entsprechend basiert deren Klassifikation nach wie vor auf der Abgrenzung von einer anderen klar definierten und als komplementär angenommenen Merkmalsgruppe anstatt auf immanenten Gemeinsamkeiten. Für eine klarere Definition der Gruppe der nicht-kognitiven Merkmale und eine angemessenere theoretische Basis sollte in zukünftigen Arbeiten

verstärkt Fokus darauf gelegt werden, dieses Theoriedefizit zu schließen. Dies würde gleichzeitig eine vollständigere und eindeutigere Systematisierung der zugehörigen Merkmale ermöglichen, als es in bisherigen Arbeiten (Farrington et al., 2012; Gutman & Schoon, 2013) erfolgen konnte.

Unter anderem basierend auf diesen beiden herangezogenen Systematisierungen wurden für diese Arbeit zwei Merkmale als potenzielle individuelle Prädiktoren ausgewählt. Es lassen sich allerdings auch andere Merkmale bzw. Merkmalsgruppen als mögliche bedeutsame Prädiktoren bzw. Unterscheidungsmerkmale identifizieren. Insbesondere vor dem Hintergrund, dass der Großteil der Varianz in der Zielvariable in dieser Arbeit durch Merkmale auf Schülerebene erklärt werden kann, sollte in zukünftigen Untersuchungen der Fokus vor allem auf diese Ebene gelegt werden. Im Folgenden werden zunächst potenzielle Prädiktoren aus dem Bereich der individuellen Schülermerkmale benannt. Anschließend erfolgen einerseits Erläuterungen zu den Bereichen Unterricht und Lehrkraft, andererseits auch zu weiteren möglicherweise relevanten Bereichen und Merkmalen des Rahmenmodells zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen, die in dieser Arbeit nicht in den Blick genommen wurden. Aufgrund der geringen Anzahl bisheriger Forschungen in diesem Bereich lässt sich dabei eine Vielzahl an Merkmalen identifizieren, deren Untersuchung vielversprechend erscheint.

Konkret scheint es lohnenswert, im Bereich der individuellen Schülermerkmale weitere verwandte Merkmale aus dem Bereich der Erwartung x Wert-Theorie zu betrachten. So konnte beispielweise insbesondere die Bedeutung der Selbstwirksamkeitserwartung oder der Motivation für die schulischen Leistungen generell vielfach belegt werden (Arens, Frenzel & Goetz, 2020; Kriegbaum et al., 2018; Steinmayr & Spinath, 2009) und auch in differenziellen Untersuchungen zeigen sich erste Hinweise auf eine besondere Bedeutung für leistungsstarke Schüler:innen (Kartal & Kutlu, 2017; Pitsia, 2022). Darüber hinaus wäre auch denkbar, dass nicht – wie in dieser Arbeit untersucht – die Einstellung zu einem bestimmten Fach für besondere Leistungsstärke in diesem Bereich ausschlaggebend ist, sondern die generelle Einstellung zu Schule und Lernen. So konnte Schmidtner (2017) beispielsweise zeigen, dass die leistungsstärksten Schüler:innen häufig auch in anderen Fächern leistungsstark sind, was als Indiz für eine hohe Bedeutung fachübergreifender Merkmale (wie der allgemeinen kognitiven Fähigkeiten, der Einstellung zu Schule und Lernen oder die Unterstützung durch die Eltern), sprechen könnte. Auf Basis der Befunde von Farrington et al. (2012) kann vermutet werden, dass dabei auch insbesondere fachunabhängige Merkmale des

Bereiches *Academic Perseverance* für besondere Leistungsstärke relevant sein könnten. Diese Kategorie nicht-kognitiver Merkmale umfasst Faktoren wie *grit* oder Ausdauer – Eigenschaften, die insbesondere für langfristige Zielerreichungen und damit für das Erlangen und Aufrechterhalten hoher Kompetenzen bedeutsam sind. Somit wäre eine Überprüfung weiterer individueller Faktoren angezeigt, um eine höhere Zahl an potenziellen Prädiktoren besonderer Leistungsstärke zu überprüfen, als es in der bisherigen differenziellen Forschung geschehen ist.

Dazu könnte ganz ähnlich wie in dieser Arbeit der Ansatz der multiplen multinomial logistischen Regression gewählt werden. Diese Herangehensweise würde es zudem ermöglichen, eine deutlich höhere Anzahl individueller Merkmale (schrittweise) in nur einem Regressionsmodell zu überprüfen, da bei einem reinen Verbleib auf Individualebene die Berechnungen zeitlich und technisch deutlich weniger voraussetzungsreich sind. Mit einer solchen Untersuchung könnte somit ökonomisch ein hoher Zugewinn an Erkenntnissen im Bereich der differenziellen Untersuchungen besonderer Leistungsstärke generiert werden.

Über die individuellen Merkmale hinaus gibt es weitere relevante Faktoren aus anderen Bereichen des Rahmenmodells zur Betrachtung des Zusammenhangs von Schülerleistungen und deren Bedingungen, die dazu beitragen könnten, dass Schüler:innen Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV erbringen und in dieser Arbeit nicht betrachtet wurden. Ein Bereich, der in vielen Untersuchungen Betrachtung findet, sich aber in Metaanalysen und Übersichtsarbeiten zumeist als weniger bedeutsam herausstellte als der der individuellen Schülermerkmale, des Unterrichts oder der Lehrkräfte ist der der Eltern und ihrem Erziehungs- und Unterstützungsverhalten (Hattie, 2009; Scheerens et al., 2013). Es konnte dabei jedoch auch gezeigt werden, dass sich die Bedeutung der einzelnen Faktoren dieses Bereiches stark unterscheidet. So können vor allem Merkmale, die eng mit dem Lernprozess und der Lernumgebung der Schüler:innen zusammenhängen, als relevante Prädiktoren in generellen Untersuchungen ausgemacht werden. Bei Hattie (2009) stellten sich beispielsweise die häusliche (Lern)Umgebung, das elterliche Unterstützungsverhalten und der sozioökonomische Hintergrund als bedeutsamste Faktoren heraus, während Merkmale wie die Familienzusammensetzung oder der Fernsehkonsum wenig relevant für die Leistungen von Schüler:innen sind (siehe beispielsweise auch Dearing, Kreider, Simpkins & Weiss, 2006; Hillmayr, Täschner, Brockmann & Holzberger, 2021).

In dieser Arbeit wurde von diesen Merkmalen der sozioökonomische Hintergrund betrachtet. Da dieses Merkmal jedoch nicht im Fokus stand, sondern lediglich als Kontrollvariable hinzugezogen wurde, wurde als einfacher Indikator des sozioökonomischen Hintergrundes die Büchervariable genutzt. Die Ergebnisse verweisen dabei auf einen bedeutsamen Effekt der Büchervariable auf die Übergangswahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen. Da die Büchervariable das Konstrukt des sozioökonomischen Hintergrundes jedoch nicht vollumfänglich abbildet, sollte auch im Kontext besonderer Leistungsstärke eine genauere Betrachtung des sozioökonomischen Hintergrundes mit komplexeren Operationalisierungen vorgenommen werden.

Schmidtner (2017) konnte in der Sekundarstufe bereits die Bedeutsamkeit des sozioökonomischen Hintergrundes, operationalisiert über den HISEI und die EGP-Klassen, für leistungsstarke Schüler:innen nachweisen. Dies stützen auch die Befunde von Gilleece et al. (2010), die ebenfalls Sekundarschüler:innen in den Blick nehmen. In Anlehnung an die Operationalisierung des sozioökonomischen Hintergrundes in OECD-Studien (*index of economic, social, and cultural status*; ESCS), in denen ökonomische, soziale und kulturelle Faktoren gemeinsam in einem übergeordneten Konstrukt betrachtet werden, nutzen sie jedoch eine noch umfangreichere Operationalisierung als Schmidtner (2017). Anders als in OECD-Studien gehandhabt, setzen sie dabei mit Blick auf mögliche praktische Schlussfolgerungen separate Konstrukte für die verschiedenen Bereiche des ESCS ein. Eine vertiefende Untersuchung der Bedeutung des sozioökonomischen Hintergrundes sollte in Anlehnung an die Operationalisierungen in den Studien von Gilleece et al. (2010) und Schmidtner (2017) im Kontext besonderer Leistungsstärke auch für die Grundschule vorgenommen werden, um nähere Einblicke hinsichtlich der Bedeutung verschiedener Facetten des sozioökonomischen Hintergrundes in frühen Jahren der Schulzeit zu gewinnen.

Auch in den Bereichen Unterricht, Lehrkraft und Klasse gibt es potenzielle Prädiktoren besonderer Leistungsstärke, die in dieser Arbeit nicht betrachtet wurden. Im Zusammenhang mit diesen Bereichen, die alle Level 2 zugeordnet werden können, ist jedoch eine Berücksichtigung des ICCs angezeigt. Dieser liegt für die multinomiale Zielvariable in dieser Arbeit bei etwas über 10 % und spricht somit zwar für eine gewisse Bedeutung von Level-2-Merkmalen. Gleichzeitig impliziert dieser Befund allerdings, dass deren Relevanz, insbesondere im Vergleich mit Merkmalen der Individualebene, nicht überschätzt werden sollte. Eine grundlegende Frage, die bei der Untersuchung besonderer Leistungsstärke im Kontext von Level-2-Merkmalen zentral

ist, betrifft die Verteilung der leistungsstarken Schüler:innen auf die Klassen sowie die Klassenkomposition. Für die Sekundarstufe konnte beispielsweise gezeigt werden, dass leistungsstarke Schüler:innen am häufigsten das Gymnasium besuchen und sich dabei relativ gleichmäßig über die untersuchten Gymnasialklassen verteilen. Hinsichtlich weiterer Klassenmerkmale konnte zudem nachgewiesen werden, dass beispielsweise der Anteil leistungsstarker Schüler:innen in einer Klasse, eine hohe Leistungsstreuung innerhalb einer Klasse, die Unterrichtszeit pro Woche und zusätzliche Unterrichtsstunden bedeutsame Merkmale im Kontext besonderer Leistungsstärke darstellen (Schmidtner, 2017; Tuorón et al., 2018).

Dabei muss jedoch berücksichtigt werden, dass in Deutschland das mehrgliedrige Schulsystem für eine gewisse Selektion beim Übergang von der Grundschule in die Sekundarstufe sorgt, die in der Grundschule noch nicht wirksam wird. Entsprechend sind die bisherigen Ergebnisse für die Sekundarstufe nur bedingt für die Grundschule gültig. Da der gesamte Bereich der Klasse und damit die Klassenzusammensetzung in der vorliegenden Arbeit nicht im Fokus stand, wurde die Verteilung leistungsstarker Schüler:innen auf die Klassen im Kontext der Grundschule bisher in Deutschland nicht genauer untersucht. Eine systematische Überprüfung der Klassenzusammensetzung ist jedoch als wichtige Grundlagenarbeit anzusehen und sollte in zukünftigen Forschungen angestrebt werden, da sie vertiefende Einblicke gewähren und damit eine Basis für weitere Forschungen und Schwerpunktsetzungen liefern kann.

Ziel dieser Untersuchung war es, neben der Untersuchung von möglichen Prädiktoren besonderer Leistungsstärke auch zu überprüfen, inwiefern die Merkmale der Lehrkräfte, des Unterrichts und der Schüler:innen zusammenwirken und direkte sowie indirekte Effekte auf die Leistung haben. Dabei hat sich gezeigt, dass Lehrkräftemerkmale die größten Effekte auf die Unterrichtsgestaltung haben, insgesamt aber eher wenig bedeutsam für die Leistung sind. Während für die meisten Unterrichtsmerkmale über die verschiedenen Modellvarianten hinweg eindeutige Zusammenhänge mit den Lehrkräftemerkmalen und den Merkmalen der Schüler:innen ausgemacht werden konnten, zeigte sich insbesondere bei Pfaden, die im Zusammenhang mit der individuellen Förderung stehen, weniger eindeutige Ergebnisse. Je nach konkreten Modellspezifikationen lagen die Zusammenhänge mal knapp unter, mal knapp über dem Signifikanzniveau von $p \leq .05$. Die Ergebnisse dieser Untersuchung liefern somit Hinweise auf mögliche Zusammenhänge, diese können jedoch nicht zweifelsfrei festgestellt werden. Um die Zusammenhänge von individueller Förderung sowohl mit Merkmalen der Lehrkraft, aber auch denen der Schüler:innen und ihren Leistun-

gen eindeutiger bestimmen zu können, sollte eine fokussierte Untersuchung dieses Merkmals, auch in Kombination mit weiteren Ansätzen zur Binnendifferenzierung, erfolgen.

Ein Ansatz für weitere Untersuchungen, die an der Schnittstelle zwischen inhaltlichen und methodischen Überlegungen anzusiedeln ist, bezieht sich auf die Erfassung der Unterrichtsmerkmale. In dieser Arbeit wurde entschieden, die Merkmale des Unterrichts über Schüleraussagen zu erfassen und die Informationen dann auf Level 2 zu aggregieren, um ein Klassenmerkmal und damit objektiveres Maß anstelle der subjektiven Einschätzungen zu erhalten. Welches Vorgehen in diesem Zusammenhang gewählt werden sollte, ist ein vieldiskutiertes Thema in der Forschung und für jedes Forschungsinteresse individuell zu entscheiden. Für die Frage, welche Faktoren besondere Leistungsstärke begünstigen können, könnte statt der Betrachtung von Klassenmerkmalen beispielsweise auch die individuelle Wahrnehmung von Unterrichtsfacetten durch die Schüler:innen im Sinne des Angebots-Nutzungs-Modells relevant sein.

Insbesondere für das Merkmal kognitive Aktivierung bietet diese Perspektive einen hohen Mehrwert, da so nicht das generelle Potenzial zur kognitiven Aktivierung in der Klasse, sondern die individuell wahrgenommene kognitive Aktivierung der einzelnen Schüler:innen mit ihren individuellen Leistungsvoraussetzungen betrachtet werden kann (signifikante Zusammenhänge für die Individualebene zeigte sich beispielsweise bei Henschel et al., 2023 oder Stang et al., 2020). Bei der Untersuchung besonderer Leistungsstärke kann vermutet werden, dass das mittlere Maß der kognitiven Aktivierung in der Klasse nicht (vollständig) mit der Einschätzung der leistungsstärksten Schüler:innen übereinstimmt, da diese mit höheren Vorkenntnissen ausgestattet sind und entsprechend bei gleichem Impuls möglicherweise ein niedrigeres Maß an kognitiver Aktivierung wahrnehmen. Somit würde es eine Betrachtung auf Individualebene ermöglichen, zu überprüfen, ob die individuelle wahrgenommene kognitive Aktivierung ein passenderer Prädiktor besonderer Leistungsstärke ist als die kognitive Aktivierung als Klassenmerkmal. Da prinzipiell ähnliche Zusammenhänge auch für die anderen Unterrichtsmerkmale vorstellbar sind, könnte eine Untersuchung angebracht sein, in der verschiedene Merkmale des Unterrichts sowohl auf Klassenebene aggregiert als auch auf Individualebene modelliert werden und dann auf ihre Effekte auf die Übergangswahrscheinlichkeit, Leistungen auf Kompetenzstufe V anstatt auf Kompetenzstufe IV zu erbringen, überprüft werden.

Ein alternativer Ansatz zur näheren Untersuchung von Unterrichtsmerkmalen und Leistung ergibt sich aus dem erwartungswidrigen Ergebnis für den Zusammenhang zwischen kognitiver Aktivierung und Leistung in dieser Arbeit. In Kapitel 9.2.2 wurde erläutert, dass dieser möglicherweise durch eine entgegengesetzte Wirkrichtung zwischen beiden Merkmalen erklärt werden könnte. Denkbar wäre jedoch auch, dass der Effekt von Unterrichtsmerkmalen auf motivationale und kognitive Schülermerkmale in Abhängigkeit von der individuellen Leistung bzw. von verwandten Merkmalen, wie beispielsweise dem Vorwissen, unterschiedlich stark ausfällt. Um die Zusammenhänge besser zu verstehen, sollte somit überprüft werden, ob das Leistungsniveau von Schüler:innen als Moderator der Zusammenhänge zwischen Unterrichts- und individuellen Schülermerkmalen fungiert.

Neben multiplen multinomial logistischen Regressionen und Moderatorenanalysen lassen sich noch weitere methodische Ansätze ausmachen, mit denen sich der Frage nach Prädiktoren besonderer Leistungsstärke genähert werden kann. Eine Erkenntnis dieser Arbeit ist das Wiedererstarken des Aptitude-Treatment-Ansatzes in den letzten Jahren. In Anbetracht zunehmender Heterogenität in der Schülerschaft und der Ergebnisse in dieser sowie vorheriger Arbeiten, die zeigen, dass verschiedene Einflussfaktoren für unterschiedlich leistungsstarke Schüler:innen unterschiedlich bedeutsam sind, sollte dieser Ansatz zukünftig verstärkt verfolgt werden. Neben eher einfachen t-Tests, die eingesetzt werden könnten, um Merkmale zu identifizieren, die sich bedeutsam zwischen den potenziell leistungsstarken und den leistungsstärksten Schüler:innen unterscheiden, gibt es auch die Möglichkeit multiple Gruppenvergleiche durchzuführen. So könnten nicht nur einzelne Merkmale betrachtet werden, sondern komplexere Modelle für verschiedene Leistungsgruppen berechnet und anschließend auf Gemeinsamkeiten und Unterschiede hin analysiert werden, um differenzielle Effekte und Prädiktoren zu identifizieren. Gleichzeitig würde dieser Ansatz den Vorteil bieten, für eine Differenzierung zwischen Gruppen nicht auf multinomial logistische Regressionen angewiesen zu sein. Stattdessen können lineare Regressionen genutzt werden, sodass auch bei der Formulierung komplexerer (Mehrebenen)Modelle aussagekräftigere Parameter zur Einschätzung der Güte und Bedeutung von Modellen (z. B. zur Varianzaufklärung) zur Verfügung stehen als in den Analysen dieser Arbeit.

Während diesem Kritikpunkt aus Kapitel 9.3 in weiteren Untersuchungen also mithilfe von multiplen Gruppenvergleichen begegnet werden könnte, liefern die bisher aufgeführten Ansätze noch keine Lösung für die bedeutsamste methodische Einschränkung dieser Arbeit: Aufgrund der Nutzung von Querschnittsdaten können mit

den vorliegenden Analysen keine Aussagen über kausale Zusammenhänge zwischen verschiedenen Merkmalen getroffen werden. Annahmen über die Richtung dieser Zusammenhänge können lediglich auf Basis bisheriger theoretischer Überlegungen und Forschungen erfolgen, jedoch nicht anhand statistischer Befunde. Vor diesem Hintergrund ist es besonders bedeutsam, weitere Untersuchungen zu Merkmalen leistungsstarker Schüler:innen und Prädiktoren dieser Leistungsstärke auf Basis längsschnittlicher Daten durchzuführen. Dadurch wäre es möglich, auch statistische Belege für kausale Zusammenhänge zu liefern und belastbarere Ergebnisse zu generieren. Insbesondere zur besseren Einordnung der Zusammenhänge zwischen kognitiver Aktivierung und Leistung sowie fachbezogene Einstellung und Leistung wären solche Analysen nötig, da mit der vorliegenden Arbeit lediglich Vermutungen für mögliche Erklärungsansätze für die erwartungswidrigen Befunde geliefert werden können. Mit längsschnittlichen Untersuchungen ließen sich diese genauer untersuchen, sodass weitere tiefgehende Erkenntnisse zur differenziellen Wirkung verschiedener Merkmale auf die Leistung unter Berücksichtigung des zeitlichen Verlaufes zur Verfügung gestellt werden können.

Nichtsdestotrotz fallen auch solche Untersuchungen, genau wie Sekundäranalysen auf Basis von *Large-Scale*-Daten, in den Bereich der nicht-experimentellen Studien. Diese haben im Gegensatz zu experimentellen oder quasi-experimentellen Untersuchungen den Nachteil, dass die untersuchten Personen nicht zufällig den einzelnen Untersuchungsbedingungen zugeordnet und im Sinne eines Experiments mögliche Störvariablen kontrolliert werden können. Stattdessen werden natürliche Gruppen in der Stichprobe zur Untersuchung von Zusammenhängen unter stark eingeschränkten Kontrollmöglichkeiten genutzt, weshalb von einer geringeren internen Validität solcher Studien ausgegangen werden muss. Dies betrifft nicht nur die vorliegende Arbeit, sondern auch zahlreiche andere komplexere Untersuchungen im Kontext besonderer Leistungsstärke. Viele der Studien, die Prädiktoren besonderer Leistungsstärke betrachten, basieren wie auch diese Arbeit auf Daten von *Large-Scale-Assessment*, (quasi)experimentelle Studien finden sich bisher nur vereinzelt (siehe auch Neuendorf, 2022). Die meisten Studien in diesem Bereich können zwar je nach konkretem Design somit begründete Hinweise auf mögliche (kausale) Zusammenhänge liefern, um allerdings eindeutige Kausalaussagen zu Einflussfaktoren besonderer Leistungsstärke treffen zu können, wäre eine Untersuchung mit (quasi)experimentellem Design notwendig. In Anlehnung an das Vorgehen von Sontag und Stoeger (2015) könnten beispielsweise zur Untersuchung der Bedeutung individueller Merkmale Trainings zu

Förderung einzelner Faktoren in der Schule durchgeführt werden und deren Einfluss für unterschiedlich leistungsstarke Schüler:innen in den Kontroll- und Interventionsklassen überprüft werden (für genauere Informationen zum Untersuchungsdesign siehe Sontag & Stoeger, 2015). Die Bedeutsamkeit von Unterrichtsmerkmalen könnte wiederum über eine systematische Anpassung einer Unterrichtsmethode und deren Auswirkungen auf verschiedene Schülergruppen untersucht werden, wie es beispielsweise Midouser und Betzer (2007) für *problem-based-learning* über einen Zeitraum von zwei Jahren überprüft haben.

Neben inhaltlichen und alternativen methodischen Ansätzen zur Bearbeitung der übergeordneten Forschungsfrage haben sich zudem zwei Anhaltspunkte für weitere Untersuchungen im Bereich der Methodenforschung ergeben. Der erste Aspekt betrifft den Umstand, dass sich die allermeisten Untersuchungen – wobei in dieser Arbeit insbesondere solche, die sich mit *Model-Fit*-Werten auseinandersetzen, von Bedeutung waren – auf ‚ideale', d. h. vollständige und vor allem normalverteilte Daten, beziehen. Solche ‚idealen' Daten sind in der Forschung jedoch eher selten, häufig liegt in zumindest einem Teil der Daten mindestens eine Verletzung dieser Voraussetzungen vor. Die Ausrichtung bisheriger Methodenforschungen hat dabei zur Folge, dass es nur sehr wenige Anhaltspunkte dafür gibt, wie beispielsweise mit *Model-Fit*-Werten bei ‚nicht-idealen' Daten umzugehen ist. In dieser Arbeit konnten einige wenige Untersuchungen ausgemacht werden, die darauf hindeuten, dass die üblichen Faustregeln zur Interpretation bei Verletzung von Voraussetzungen nicht unbedingt adäquat sind (Brosseau-Liard & Savalei, 2014; Gao et al., 2020). Diesen zufolge scheinen auch weniger strenge Richtwerte geeignet zur Einschätzung der Daten zu sein. Fehlendes Wissen über einen angemessenen Umgang mit ‚nicht-idealen' Daten könnte entsprechend dazu führen, dass fälschlicherweise Modelle aufgrund vermeintlich unzureichender *Model-Fit*-Werte abgelehnt werden und damit wertvolle Erkenntnisse verloren gehen.

Der zweite Ansatzpunkt für weitere Methodenforschungen bezieht sich auf die geringe Anzahl (absoluter) Kennwerte, die bei der Berechnung logistischer Regressionen unter Nutzung numerischer Integration und multipler Imputation in Mplus zur Verfügung stehen. Dies führt dazu, dass insgesamt nur wenig robuste Aussagen über die Passung der Modelle zu den Daten sowie die Relevanz der Modelle im Gesamten getroffen werden können und somit die Frage nach der Bedeutung dieser Arbeit im Forschungsfeld bis zu einem gewissen Grad offen bleiben muss. Um Arbeiten, die diese oder ähnliche Ansätze verfolgen, mehr Gewicht zu verleihen und belastbarere Aussagen zu ermöglichen, sollten Maße (weiter)entwickelt bzw. implementiert wer-

den, die die Grundlage für Aussagen über die Passung derartiger Modelle (insbesondere absolute *Model-Fit*-Werte) sowie die Varianzaufklärung im Gesamten bilden können. Zwar gibt es insbesondere im Bereich der Varianzaufklärung inzwischen eine ganze Reihe von Maßzahlen, die für verschiedenste Anwendungsfälle genutzt werden können, für die konkreten Spezifikationen in dieser Arbeit wurden jedoch weder geeignete Kennwerte durch Mplus bereitgestellt noch ließen sich aus zur Verfügung stehenden Werten sinnvoll Maße der Varianzaufklärung berechnen.

Somit ergeben sich aus der vorliegenden Arbeit sowohl weiterführende Forschungsansätze, die im Bereich der Theoriebildung und inhaltlichen Fortführung der mit diesen Analysen begonnenen Untersuchung verortet werden können, als auch Ansätze im Bereich der Methoden mit möglichen alternativen methodischen Herangehensweisen sowie der Notwendigkeit weitere Erkenntnisse zum Einsatz verschiedener Kennzahlen zu generieren.

10.2 Bedeutung für die pädagogische Praxis

Nachdem im vorangegangenen Kapitel Ansätze für weiterführende Forschungen abgeleitet wurden, die sich aus den Vorarbeiten, den Analysen und den Ergebnissen in dieser Arbeit ergeben haben, wird in diesem Kapitel abschließend die Bedeutung der Befunde für die pädagogische Praxis dargelegt. Die zentrale Frage dieser Arbeit war, herauszufinden, wodurch sich die leistungsstärksten Schüler:innen im Bereich der Naturwissenschaften am Ende der Grundschulzeit, die im Rahmen der TIMSS-Testung die höchste Kompetenzstufe erreichen, von den Schüler:innen unterscheiden, die Leistungen auf Kompetenzstufe IV zeigen und damit als potenziell leistungsstark angesehen werden können. Dazu wurden Analysen mit einem multinomial logistischen Ansatz durchgeführt, um Merkmale zu identifizieren, die dazu beitragen, dass Schüler:innen anstelle von Kompetenzstufe IV Kompetenzstufe V erreichen. So können mögliche Prädiktoren besonderer Leistungsstärke und damit Ansatzpunkte für eine Förderung potenziell leistungsstarker Schüler:innen ausgemacht werden, um es diesen zu ermöglichen, ihre Potenziale besser auszuschöpfen und ebenfalls zu den leistungsstärksten Schüler:innen am Ende der Grundschulzeit zu gehören. Durch die Untersuchung dieser Fragestellung im Kontext der Grundschule bieten die Erkenntnisse die Möglichkeit, frühzeitig und passgenau in der schulischen Laufbahn von Schüler:innen anzusetzen und so dafür zu sorgen, dass nicht schon in frühen

Jahren entscheidende Rückstände entwickelt werden, die in späteren Jahren nicht mehr kompensiert werden können.

Die wohl wichtigste, übergeordnete Erkenntnis, die sich bei der Untersuchung dieser Fragestellung zeigt, ist, dass sich Schüler:innen der beiden betrachteten Kompetenzstufen unterscheiden und wie vermutet verschiedene Merkmale unterschiedlich relevant für die Übergangswahrscheinlichkeit sind, Leistungen auf Kompetenzstufe V anstelle von Kompetenzstufe IV zu erbringen. Das bedeutet, dass – bevor anhand konkreter Prädiktoren über mögliche Wege zur Förderung besonderer Leistungsstärke nachgedacht werden kann – zunächst einmal das Bewusstsein in Schule und Unterricht dafür geschaffen werden muss, dass für unterschiedlich leistungsstarke Schüler:innen unterschiedliche Merkmale von Bedeutung sind. Die vorliegende Arbeit konnte dabei insbesondere zeigen, dass sich dieser Umstand nicht nur auf sehr unterschiedlich leistungsstarke Schülergruppen (wie z. B. zwischen den leistungsschwächsten und den leistungsstärksten Schüler:innen) bezieht, sondern auch auf Schüler:innen, die auf den ersten Blick als nicht besonders unterschiedlich leistungsstark wahrgenommen werden könnten.

Konkret zeigen die Analysen in dieser Arbeit, dass sich vor allem drei relevante Prädiktoren besonderer Leistungsstärke aus den betrachteten Bereichen individuelle Schülermerkmale, Unterricht und Lehrkraft ausmachen lassen, weshalb diese in den folgenden Ausführungen im Fokus stehen. Die identifizierten Merkmale stammen dabei aus dem Bereich der individuellen Schülermerkmale und des Unterrichts.

Als bedeutsamster Prädiktor besonderer Leistungsstärke konnte in dieser Arbeit das fachbezogene Selbstkonzept ausgemacht werden. Die Ergebnisse zeigen, dass das fachbezogene Selbstkonzept maßgeblich und von allen untersuchten Prädiktoren am stärksten dazu beiträgt, dass Schüler:innen zu den leistungsstärksten Schüler:innen am Ende der Grundschulzeit in den Naturwissenschaften gehören. Die generelle Bedeutung dieses Merkmals für Leistung gilt in der Forschung bereits als etablierter Befund. Die vorliegende Arbeit zeigt, dass diese auch in diesem konkreten Kontext gültig ist. Insbesondere vor dem Hintergrund, dass Helmke (1992) zeigen konnte, dass der Zusammenhang zwischen dem Selbstkonzept und der Leistung enger wird, je höher die erbrachten Leistungen sind und dass vielfach belegt werden konnte, dass das Selbstkonzept in reziprokem Zusammenhang mit der Leistung steht, kann dieses Merkmal als plausibler Prädiktor besonderer Leistungsstärke angenommen werden. Da in Kapitel 4.1.1 zudem gezeigt werden konnte, dass das Selbstkonzept im Grundschulalter noch nicht so stark von der Leistung abhängt wie in späteren

Phasen der Schulzeit, wird davon ausgegangen, dass eine Förderung des Selbstkonzeptes vor allem im Grundschulalter möglich und angebracht ist, damit sich positive Entwicklungen des Selbstkonzepts besser auf die Leistung auswirken können (Praetorius et al., 2016; Weinert, Schrader, et al., 1989). Entsprechend ist eine Förderung des Selbstkonzepts in den ersten Schuljahren von besonderer Bedeutung nicht nur für die generelle schulische Leistungsentwicklung, sondern auch für die Entwicklung besonderer Leistungsstärke.

Da das Selbstkonzept bereits seit mehreren Jahrzehnten im Fokus der Forschung steht, gibt es bereits einige Strategien zur Förderung des Selbstkonzeptes bei Kindern. Eine Vielzahl dieser Programme fokussiert sich auf bestimmte Alters- oder Zielgruppen, z. B. das Projekt „Vivo – Bildung von Grund auf!" für Grundschulkinder in sozial benachteiligten Nachbarschaften (Hauenstein, Pfeil & Wittmer-Gerber, 2022), die Intervention „Die Entdeckerreise in die neue Schule", die speziell am Übergang in die Sekundarstufe I ansetzt (Kienle & Kopp, 2014), oder das „Ich bin Ich-Konzept", das alltagsintegriert das Selbstkonzept bereits ab dem Kindergartenalter fördern soll (Merget, Engel & Aich, 2021). Dabei sollte jedoch verstärkter Fokus darauf gelegt werden, dass sich Maßnahmen nicht nur an Kinder mit ungünstigen Voraussetzungen oder niedrigem Selbstkonzept richten, sondern auch Kinder mit durchschnittlichem oder hohem Selbstkonzept in den Blick genommen werden, um auch für diese Kinder positive Auswirkungen auf die Leistungen zu ermöglichen und somit das Erreichen besonderer Leistungsstärke zu fördern. Bereits bestehende Interventionen bieten dabei den Vorteil, dass sie bereits über einen längeren Zeitraum hinweg erprobt wurden und mit wenig Aufwand an andere Zielgruppen angepasst werden könnten.

Das zweite Merkmal, das sich in dieser Arbeit als bedeutsamer Prädiktor besonderer Leistungsstärke herausstellte, ist die konstruktive und dabei insbesondere die emotionale Unterstützung im Unterricht. Die Besonderheit dieses Merkmals ist, dass es sowohl direkt als auch indirekt zur Erreichung besonderer Leistungsstärke beiträgt, wobei sich in dieser Arbeit gezeigt hat, dass insbesondere der indirekte Effekt vorzuherrschen scheint. Indirekt hat die konstruktive Unterstützung zunächst einen Effekt auf das Selbstkonzept der Schüler:innen, das dann wiederum positiv auf die Leistungen wirken kann. Somit kann neben den externen Ansätzen, die zuvor erläutert wurden, auch die Unterstützung durch die Lehrkraft selbst als Möglichkeit zur Förderung des Selbstkonzeptes von Grundschüler:innen angesehen werden. Wie in Kapitel 4.1.1 dargestellt, formt sich das Selbstkonzept von Schüler:innen über verschiedene Informationsquellen, unter anderem über Rückmeldungen von wichti-

gen Bezugspersonen oder aber durch Vergleiche z. B. mit Mitschüler:innen, anderen Fächern oder früheren Leistungen. Dieses Wissen können sich Lehrkräfte zunutze machen, um aktiv das Selbstkonzept ihrer Schüler:innen zu stärken, z. B. indem sie positive Kompetenzeinschätzungen formulieren oder selbst verstärkt auf individuelle Verbesserungen der Schüler:innen hinweisen. Da sich in dieser Arbeit in einigen Modellen zudem bedeutsame direkte Effekte der konstruktiven Unterstützung zeigen, können Lehrkräfte auch durch emotional unterstützendes Verhalten direkt zum Erreichen besonderer Leistungsstärke beitragen. Dieses zeichnet sich insbesondere durch positive Beziehungen zwischen Lehrkräften und Schüler:innen, das Herstellen einer positiven und wertschätzenden Grundstimmung, durch das Erkennen und Berücksichtigen individueller Bedürfnisse der Schüler:innen oder auch durch die Ansprechbarkeit der Lehrkräfte aus (siehe Kapitel 4.2.1). Werden diese Aspekte im Unterricht umgesetzt, kann also eine leistungsfördernde Wirkung vermutet werden, die für Schüler:innen bedeutsam ist, um besonders hohe Leistungen am Ende der Grundschulzeit zu erbringen.

Als dritter bedeutsamer Prädiktor besonderer Leistungsstärke konnten in dieser Arbeit zwei Subskalen des CFT 20-R als Facetten der kognitiven Fähigkeiten der Schüler:innen identifiziert werden. Aufgrund der Konzeptualisierung des CFT 20-R gilt dieser Befund vor allem für mathematische Facetten kognitiver Fähigkeiten. Dabei muss jedoch berücksichtigt werden, dass die Erhebung dieser zwei Subskalen nicht mit etablierten, psychologischen Verfahren zur Diagnostik der kognitiven Fähigkeiten gleichgesetzt werden können und die Ergebnisse somit nur eingeschränkt gültig sind. Nichtsdestotrotz können sie wertvolle Hinweise auf die Bedeutung kognitiver Fähigkeiten liefern. Wie eingangs im Kapitel zur Begriffsbestimmung 2 dargestellt wurde, können die kognitiven Fähigkeiten, auch als Intelligenz bezeichnet, bis zu einem gewissen Grad entwickelt werden. Dies ist insbesondere im Kindesalter möglich, da sich kognitive Fähigkeiten mit zunehmendem Alter zu stabilen Persönlichkeitsmerkmalen entwickeln. Kognitive Fähigkeiten bilden gleichzeitig die Grundlage für die weitere Kompetenzentwicklung, die durch wiederholte Lerngelegenheiten angeregt wird. In der Zeit, in der sich die kognitiven Fähigkeiten fördern lassen, finden zudem erste Fähigkeitsdifferenzierungen und damit einhergehend eine Profilbildung statt (siehe Abbildung 1 in Kapitel 2). Somit stellt das Kindesalter eine bedeutsame Phase für die weitere schulische und berufliche Entwicklung dar, da sich sowohl kognitive Fähigkeiten als auch inhaltliche Interessen in gewissem Maße in dieser Zeit formen. Mit der vorliegenden Arbeit konnte gezeigt werden, dass dies nicht nur für generelle

Tendenzen in schulischen (naturwissenschaftlichen) Leistungen gilt, sondern dass kognitive Fähigkeiten auch für eine Unterscheidung zwischen potenziell leistungsstarken und den leistungsstärksten Schüler:innen relevant sind. Das bedeutet, dass Lehrkräfte, aber auch weitere Bezugspersonen, in dieser frühen Phase der schulischen Laufbahn die Möglichkeit haben, die kognitiven Fähigkeiten der Schüler:innen und damit besondere Leistungsstärke zu fördern.

Mittlerweile existiert eine ganze Reihe etablierter Programme zur Förderung kognitiver Fähigkeiten, die in Vorschule und Schule eingesetzt werden können. Schneider (2019) resümiert jedoch, dass von dieser Möglichkeit bisher eher selten Gebrauch gemacht wird, da bisherige Erkenntnisse aus Forschung und erprobten Interventionen nicht ausreichend in die Praxis transferiert werden konnten. Entsprechend zeigt sich hier – neben der Notwendigkeit, Programme zur Förderung kognitiver Fähigkeiten stärker in den Schulalltag zu integrieren – auch, dass von Forschung, Praxis, aber auch Politik eine stärkere Nutzung von Befunden außerhalb der Forschung angestrebt werden sollte. Während die vorliegende Arbeit eher im Bereich der Grundlagenforschung eingeordnet werden kann und somit vor allem als Basis für weitere Forschungen und Überlegungen sowie praxisorientiertere Projekte dient, trifft diese Forderung vor allem Projekte, die sich durch eine gewisse konzeptuelle Praxisnähe auszeichnen. Im Bereich der Begabungsförderung wurden mit LemaS und LemaS-Transfer als Reaktion auf Befunde international vergleichender Studien wie TIMSS oder PISA solche Projekte ins Leben gerufen (KMK, 2015, 2016). Im Rahmen dieser Initiativen werden im ersten Schritt gemeinsam mit Schulen ko-konstruktiv unterschiedliche Produkte für den Einsatz in Schule und Unterricht entwickelt und diese im zweiten Schritt in weitere Schulen transferiert (Weigand, 2020). Durch eine Kombination von weiterführender Grundlagenforschung, wie sie in dieser Arbeit erfolgt ist, und der Durchführung praxisnaher Initiativen wie LemaS kann zukünftig besondere Leistungsstärke am Ende der Grundschulzeit besser gefördert werden, sodass im Rahmen internationaler Schulvergleichsuntersuchungen hoffentlich bald höhere Anteile von Schüler:innen auf Kompetenzstufe V in Deutschland berichtet werden können.

Literaturverzeichnis

Abramowski, R. (2020). *Das bisschen Haushalt. Zur Kontinuität traditioneller Arbeitsteilung in Paarbeziehungen – ein europäischer Vergleich*. Opladen: Verlag Barbara Budrich. https://doi.org/10.3224/96665008

Abu-Hamour, B. & Al-Hmouz, H. (2013). A Study of Gifted High, Moderate and Low Achievers in their Personal Characteristics and Attitudes toward School and Teachers. *International Journal of Special Education*, *28*(3), 5–15.

Afari, E. (2015). Relationship of students' attitudes towards science and academic achievement. In M. S. Khine (Hrsg.), *Attitude Measurements in Science Education: Classic and Contemporary Approaches* (S. 245–262). Charlotte: Information Age Publishing.

Affandi, L. H., Husniati, H. & Saputra, H. H. (2020). Exploring the Source of Well-Being for High Achiever Students. *Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran*, *11*(1), 104–119. https://doi.org/10.25273/pe.v11i1.8767

Ai, X. (2002). Gender Differences in Growth in Mathematics Achievement: Three-Level Longitudinal and Multilevel Analyses of Individual, Home, and School Influences. *Mathematical Thinking and Learning*, *4*(1), 1–22. https://doi.org/10.1207/S15327833MTL0401_1

Ajzen, I. (2005). *Attitudes, Personality and Behavior*. Maidenhead: Open University Press.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *2nd International Symposium on Information Theory* (S. 267–281). Budapest: Akadémiai Kiadó.

Allen, D. & Fraser, B. J. (2007). Parent and Student Perceptions of Classroom Learning Environment and its Association with Student Outcomes. *Learning Environments Research*, *10*, 67–82. https://doi.org/10.1007/s10984-007-9018-z

Allinder, R. M. (1994). The Relationship Between Efficacy and the Instructional Practices of Special Education Teachers and Consultants. *Teacher Education and Special Education*, *17*(2), 86–95. https://doi.org/10.1177/088840649401700203

Altrichter, H., Trautmann, M., Wischer, B., Sommerauer, S. & Doppler, B. (2009). Unterrichten in heterogenen Gruppen: Das Qualitätspotenzial von Individualisierung, Differenzierung und Klassenschülerzahl. In W. Specht (Hrsg.), *Nationaler Bildungsbericht Österreich 2009, Band 2. Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 344–360). Graz: Leykam.

Ammann, T. (2004). *Zur Berufszufriedenheit von Lehrerinnen. Erfahrungsbilanzen in der mittleren Berufsphase*. Bad Heilbrunn: Klinkhardt.

Arens, A. K., Frenzel, A. C. & Goetz, T. (2020). Self-Concept and Self-Efficacy in Math: Longitudinal Interrelations and Reciprocal Linkages with Achievement. *The Journal of Experimental Education*, *90*(3), 615–633. https://doi.org/10.1080/00220973.2020.1786347

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Muryama, K. & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores. Long-term reciprocal effects across five waves and three achievement tracks. *The Journal of Educational Psychology*, *109*(5), 621–634. https://doi.org/10.1037/edu0000163

Arens, A. K. & Morin, A. J. S. (2016). Relations Between Teachers' Emotional Exhaustion and Students' Educational Outcomes. *Journal of Educational Psychology*, *108*(6), 800–813. https://doi.org/10.1037/edu0000105

Arens, A. K., Trautwein, U. & Hasselhorn, M. (2011). Erfassung des Selbstkonzepts im mittleren Kindesalter: Validierung einer deutschen Version des SDQ I. *Zeitschrift für Pädagogische Psychologie*, *25*(2), 131–144. https://doi.org/10.1024/1010-0652/a000030

Arnold, R., Kleß, E. & Prescher, T. (2014). Systemik des Talents und der Begabung aus Sicht der Pädagogischen Psychologie sowie der Personal- und Sozialpsychologie. In M. Stamm (Hrsg.), *Handbuch Talententwicklung. Theorien, Methoden und Praxis in Psychologie und Pädagogik* (S. 45–62). Bern: Verlag Hans Huber.

Asparouhov, T. & Muthén, B. (2018). SRMR in Mplus. Verfügbar unter: https://www.statmodel.com/download/SRMR2.pdf

Atkinson, J. W. (1957). Motivational Determinants of RIsk-Taking Behavior. *Psychological Review*, *64*, 359–372.

Atkinson, J. W. (1964). *An Introduction to Motivation*. New York: Van Nostrand Reinhold Company. https://doi.org/10.1037/14156-004

Bach, A. (2022). *Selbstwirksamkeit im Lehrberuf. Entstehung und Veränderung sowie Effekte auf Gesundheit und Unterricht*. Münster: Waxmann. https://doi.org/10.31244/9783830995166

Backhaus, K., Erichson, B., Gensler, S., Weiber, R. & Weiber, T. (2021). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (16. Aufl.). Wiesbaden: Springer. https://doi.org/10.1007/978-3-662-56655-8

Backhaus, K., Erichson, B., Plinke, W. & Rolf, W. (2018). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (15. Aufl.). Berlin: Springer. https://doi.org/10.1007/978-3-662-56655-8

Backhaus, K., Erichson, B. & Weiber, R. (2015). *Fortgeschrittene Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (3. Aufl.). Berlin: Springer Gabler. https://doi.org/10.1007/978-3-662-46087-0

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037//0033-295x.84.2.191

Banerjee, N., Stearns, E., Moller, S. & Mickelson, R. A. (2017). Teacher Job Satisfaction and Student Achievement: The Roles of Teacher Professional Community and Teacher Collaboration in Schools. *American Journal of Education*, *123*(2), 203–241. https://doi.org/10.1086/689932

Baron, R. M. & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M. & Tsai, Y.-M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV). Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, *9*(4), 469–520. https://doi.org/10.1007/s11618-006-0165-2

Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann. https://doi.org/10.31244/9783830974338

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Beck, E., Baer, M., Guldimann, T., Bischoff, S., Brühwiler, C., Müller, P., Niedermann, R., Rogalla, M. & Vogt, F. (2008). *Adaptive Lehrkompetenz. Analyse und Struktur, Veranderbarkeit und Wirkung handlungssteuernden Lehrerwissens*. Münster: Waxmann.

Becker, D. & Birkelbach, K. (2017). Bildungsungleichheit durch Schul- und Schulklasseneffekte. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (3. Aufl., S. 179–210). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-15272-7_6

Beese, C., Scholz, L. A., Jentsch, A., Jusufi, D. & Schwippert, K. (2020). *TIMSS 2019. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente und Arbeit mit den Datensätzen*. Münster: Waxmann. https://doi.org/10.31244/9783830995814

Bellin, N. (2009). *Klassenkomposition, Migrationshintergrund und Leistung. Mehrebenenanalysen zum Sprach- und Leseverständnis von Grundschülern*. Wiesbaden: VS Research. https://doi.org/10.1007/978-3-531-91359-9

Beloshitskii, A. V. & Dushkin, A. V. (2005). An Experiment in Differentiated Instruction in a Higher Technical Educational Institution. *Russian Education and Society*, *47*(9), 54–61.

Bentler, P. M. & Chou, C.-P. (1987). Practical Issues in Structural Modeling. *Sociological Methods & Research*, *16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Berliner, D. C. (1987). Simple Views of Effective Teaching and a Simple Theory of Classroom Instruction. In D. C. Berliner & B. Rosenshine (Hrsg.), *Talks to Teachers: A Festschrift for N.L. Gage* (S. 93–110). New York: Random House.

Berliner, D. C. (2005). The Near Impossibility of Testing for Teacher Quality. *Journal of Teacher Education*, *56*(3), 205–213. https://doi.org/10.1177/0022487105275904

Bermejo-Toro, L., Prieto-Ursúa, M. & Hernández, V. (2016). Towards a Model of Teacher Well-Being: Personal and Job Resources Involved in Teacher Burnout and Engagement. *Educational Psychology*, *36*(3), 481–501. https://doi.org/10.1080/01443410.2015.1005006

Berning, C. C. (2018). Strukturgleichungsmodelle. In C. Wagemann, A. Goerres & M. Siewert (Hrsg.), *Handbuch Methoden der Politikwissenschaft* (S. 1–18). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-16937-4_30-2

Best, H. & Wolf, C. (2012). Modellvergleich und Ergebnisinterpretation in Logit- und Probit-Regressionen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *64*, 377–395.

Bickel, R. (2007). *Multilevel Analysis for Applied Research. It´s Just Regression!* New York: The Guilford Press.

Binet, A. & Simon, T. (1916). *The Intelligence of the Feeble-Minded*. übersetzt von Elizabeth S. Kite, Baltimore: Williams & Wilkins. https://doi.org/10.1037/11070-000

Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J. & Döhrmann, M. (2014). Von der Lehrerausbildung in den Beruf – Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft*, *17*, 509–542. https://doi.org/10.1007/s11618-014-0564-8

Blumberg, E. (2008). *Multikriteriale Zielerreichung im naturwissenschaftsbezogenen Sachunterricht der Grundschule – Eine Studie zum Einfluss von Strukturierung in schülerorientierten Lehr-Lernumgebungen auf das Erreichen kognitiver, motivationaler und selbstbezogener Zielsetzungen*. Westfälische Wilhelms-Universität Münster. univ Diss.

Bohl, T. (2017). Umgang mit Heterogenität im Unterricht: Forschungsbefunde und didaktische Implikationen. In T. Bohl, J. Budde & M. Rieger-Ladich (Hrsg.), *Umgang mit Heterogenität in Schule und Unterricht. Grundlagentheoretische Beiträge, empirische Befunde und didaktische Reflexionen* (S. 257–274). Bad Heilbrunn: Julius Klinkhardt. https://doi.org/10.36198/9783838547558

Bohl, T., Budde, J. & Rieger-Ladich, M. (Hrsg.). (2017). *Umgang mit Heterogenität in Schule und Unterricht. Grundlagentheoretische Beiträge, empirische Befunde und didaktische Reflexionen*. Bad Heilbrunn: Julius Klinkhardt. https://doi.org/10.36198/9783838547558

Böhnel, E. & Svik, G. (1993). Modellbeschreibung und Evaluation des Schulversuchs „Innere Differenzierung und Individualisierung im Mathematikunterricht". *Unterrichtswissenschaft*, *21*(1), 66–89.

Boring, E. G. (1923). Intelligence as the Test Test It. *The New Republic*, *34*, 35–36.

Bos, W., Strietholt, R., Goy, M., Stubbe, T. C., Tarelli, I. & Hornberg, S. (2010). *IGLU 2006. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

Bos, W., Wendt, H., Ünlü, A., Valtin, R., Euen, B., Kasper, D. & Tarelli, I. (2012). Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland. In W. Bos, I. Tarelli, A. Bremerich-Voß & K. Schwippert (Hrsg.), *Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 227–259). Münster: Waxmann.

Böwing-Schmalenbrock, M. & Jurczok, A. (2011). Multiple Imputation in der Praxis: ein sozialwissenschaftliches Anwendungsbeispiel. https://nbn-resolving.org/urn:nbn:de:kobv:517-opus-58111

Brandt, H. (2020). Exploratorische Faktorenanalyse (EFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 575–614). Wiesbaden: Springer. https://doi.org/10.1007/978-3-662-61532-4

Breker, T. A. (2016). *Fähigkeitsselbstkonzept, Selbstwirksamkeit & Mindset – Wie können Lehrkräfte Erkenntnisse aus der Sozial-Kognitiven-Psychologie nutzen, um die Potenzialentfaltung von Schülerinnen und Schülern zu fördern?* Europa-Universität Viadrina: univ. Diss.

Brophy, J. (2000). Teaching. *Educational Practices Series*, *1*, 3–36.

Brosseau-Liard, P. E. & Savalei, V. (2014). Adjusting Incremental Fit Indices for Nonnormality. *Multivariate Behavioral Research*, *49*, 460–470. https://doi.org/10.1080/00273171.2014.933697

Bruder, R., Linneweber-Lammerskitten, H. & Reibold, J. (2015). Individualisieren und differenzieren. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 513–534). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-642-35119-8_19

Bruggemann, A., Groskurth, P. & Ulich, E. (1975). *Arbeitszufriedenheit*. Bern: Huber.

Brühwiler, C. & Helmke, A. (2018). Determinanten der Schulleistung. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 78–91). Weinheim: Basel.

Brühwiler, C., Helmke, A. & Schrader, F.-W. (2017). Determinanten der Schulleistung. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (S. 291–314). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-15083-9_13

Bryan, M. L. & Jenkins, S. P. (2015). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*, *32*(1), 1–20. https://doi.org/10.1093/esr/jcv059

Buch, S. R., Sparfeldt, J. R. & Rost, D. H. (2019). Diagnostik schulfachspezifischer Selbstkonzepte: Differentielles Schulisches-Selbstkonzept-Gitter (DISK-Gitter). In H. Gaspard, U. Trautwein & M. Hasselhorn (Hrsg.), *Diagnostik und Förderung von Motivation und Volition* (S. 69–86). Göttingen: Hogrefe. https://doi.org/10.1026/03001-000

Buer, J. v., Squarra, D., Ebermann-Richter, P. & Kirchner, C. (1995). Pädagogische Freiräume, berufliche Zufriedenheit und berufliche Belastung. Analysen zum unterrichtlichen Alltagshandeln von Lehrern an Wirtschaftsschulen in den neuen Bundesländern. *Zeitschrift für Pädagogik*, *41*(4), 555–577.

Burić, I. & Kim, L. E. (2020). Teacher Self-Efficacy, Instructional Quality, and Student Motivational Beliefs: An Analysis Using Multilevel Structural Equation Modeling. *Learning and Instruction*, *66*, 101302. https://doi.org/10.1016/j.learninstruc.2019.101302

Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach* (2. Aufl.). New York: Springer. https://doi.org/10.1007/b97636

Burnham, K. P. & Anderson, D. R. (2004). Multimodel Inference. Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*, 261–304. https://doi.org/10.1177/0049124104268644

Campbell, D. T. (1963). Social Attitudes and Other Acquired Behavioral Dispositions. In S. Koch (Hrsg.), *Psychology: A Study of a Science Study II. Empirical Substructure and Relations with other Sciences. Investigations of Man as Socius: Their Place in Psychology and the Social Sciences* (S. 94–172). New York: McGraw-Hill. https://doi.org/10.1037/10590-003

Caprara, G. V., Barbaranelli, C., Borgogni, L. & Steca, P. (2003). Efficacy Beliefs as Determinants of Teachers' Job Satisfaction. *Journal of Educational Psychology*, *95*(4), 821–832. https://doi.org/10.1037/0022-0663.95.4.821

Caprara, G. V., Barbaranelli, C., Steca, P. & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, *44*, 473–490. https://doi.org/10.1016/j.jsp.2006.09.001

Cavanaugh, J. E. & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*, *11*, 1–11. https://doi.org/10.1002/wics.1460

Chen, L., Bae, S. R., Battista, C., Qin, S., Chen, T., Evans, T. M. & Menon, V. (2018). Positive Attitude Toward Math Supports Early Academic Success: Behavioral Evidence and Neurocognitive Mechanisms. *Psychological Science*, *29*(3), 390–402. https://doi.org/10.1177/0956797617735528

Chomsky, N. (1980). Rules and Representations. *The Behavioral and Brain Sciences*, *3*, 1–61. https://doi.org/10.1017/s0140525x00001515

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S. & Sloan, L. R. (1976). Basking in Reflected Glory: Three (Football) Field Studies. *Journal of Personality and Social Psychology*, *34*(3), 366–375. https://doi.org/10.1037//0022-3514.34.3.366

Clausen, M. (2002). *Unterrichtsqualität: eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.

Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen. Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, *31*(2), 122–141.

Cleff, T. (2015). *Deskriptive Statistik und Explorative Datenanalyse. Eine computergestützte Einführung mit Excel, SPSS und STATA* (3. Aufl.). Wiesbaden: Springer. https://doi.org/10.1007/978-3-8349-4748-2

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates.

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3. Aufl.). London: Lawrence Erlbaum Associates.

Collie, R. J. & Martin, A. J. (2017). Teachers' Sense of Adaptability: Examining Links with Perceived Autonomy Support, Teachers' Psychological Functioning, and Students' Numeracy Achievement. *Learning and Individual Differences*, *55*, 29–39. https://doi.org/10.1016/j.lindif.2017.03.003

Colom, R. & Flores-Mendoza, C. E. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence*, *35*, 243–251. https://doi.org/10.1016/j.intell.2006.07.008

Cornelius-White, J. (2007). Learner-Centered Teacher-Student Relationships Are Effective: A Meta-Analysis. *Review of Educational Research*, *77*(1), 113–143. https://doi.org/10.3102/003465430298563

Crites, J. O. (1969). *Vocational Psychology: The Study of Vocational Behavior and Development*. New York: McGraw-Hill.

Cunningham, L. G. & Rinn, A. N. (2007). The Role of Gender and Previous Participation in a Summer Program on Gifted Adolescents' Self-Concepts Over Time. *Journal for the Education of the Gifted*, *30*(3), 326–352. https://doi.org/10.1177/016235320703000303

Curran, P. J., Bollen, K. A., Chen, F., Paxton, P. & Kirby, J. B. (2003). Finite Sampling Properties of the Point Estimates and Confidence Intervals of the RMSEA. *Sociological Methods & Research*, *32*(2), 208–252. https://doi.org/10.1177/0049124103256130

Dai, D. Y. & Rinn, A. N. (2008). The Big-Fish-Little-Pond Effect: What Do We Know and Where Do We Go from Here? *Educational Psychology Review*, *20*, 283–317. https://doi.org/10.1007/s10648-008-9071-x

Dalbert, C. & Stöber, J. (2008). Forschung zur Schülerpersönlichkeit. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (2. Aufl., S. 905–925). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91095-6_37

Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A Review of State Policy Evidence. *Education Policy Analysis Archives*, *8*(1). https://doi.org/10.14507/epaa.v8n1.2000

de Groot, A. D. (1965). *Thought and Choice in Chess*. Den Haag: Mouton.

de Jong, R. & Westerhof, K. J. (2001). The Quality of Student Ratings of Teacher Behaviour. *Learning Environments Research*, *4*, 51–85.

Dearing, E., Kreider, H., Simpkins, S. & Weiss, H. B. (2006). Family Involvement in School and Low-Income Children's Literacy: Longitudinal Associations Between and Within Families. *Journal of Educational Psychology*, *98*(4), 653–664. https://doi.org/10.1037/0022-0663.98.4.653

Deary, I. J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and Educational Achievement. *Intelligence*, *35*, 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. & Fox, H. C. (2004). The Impact of Childhood Intelligence on Later Life: Following Up the Scottisch Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130–147. https://doi.org/10.1037/0022-3514.86.1.130

Decristan, J., Kunter, M., Fauth, B., Büttner, G., Hardy, I. & Hertel, S. (2016). What Role Does Instructional Quality Play for Elementary School Children's Science Competence? A Focus on Students at Risk. *Journal for educational research online*, *8*(1), 66–89.

Deiglmayr, A., Schalk, L. & Stern, E. (2017). Begabung, Intelligenz, Talent, Wissen, Kompetenz und Expertise: eine Begriffsklärung. In U. Trautwein & M. Hasselhorn (Hrsg.), *Begabungen und Talente* (S. 1–16). Göttingen: Hogrefe. https://doi.org/10.1026/02846-000

Depaepe, F. & König, J. (2018). General Pedagogical Knowledge, Self-Efficacy and Instructional Practice: Disentangling Their Relationship in Pre-Service Teacher Education. *Teaching and Teacher Education*, *69*, 177–190. https://doi.org/10.1016/j.tate.2017.10.003

Dicke, T., Marsh, H. W., Parker, P. D., Guo, J., Riley, P. & Waldeyer, J. (2019). Job Satisfaction of Teachers and Their Principals in Relation to Climate and Student Achievement. *Journal of Educational Psychology*, *112*(5), 1061–1073. https://doi.org/10.1037/edu0000409

Dieterich, R. & Rietz, I. (1996). *Psychologisches Grundwissen für Schule und Beruf: ein Wörterbuch*. Donauwörth: Auer.

Dings, A. & Spinath, F. M. (2021). Motivational and Personality Variables Distinguish Academic Underachievers from High Achievers, Low Achievers, and Overachievers. *Social Psychology of Education*, *24*, 1461–1485. https://doi.org/10.1007/s11218-021-09659-2

Dixon, F. A., Yssel, N., McConnell, J. M. & Hardin, T. (2014). Differentiated Instruction, Professional Development, and Teacher Efficacy. *Journal for the Education of the Gifted*, *37*(2), 111–127. https://doi.org/10.1177/0162353214529042

Donat, M., Radant, M. & Dalbert, C. (2017). Psychologie der Schülerpersönlichkeit. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion* (3. Aufl., S. 167–190). Wiesbaden: Springer VS.

Döring-Seipel, E. & Dauber, H. (2010). Was hält Lehrer und Lehrerinnen gesund – die Bedeutung von Ressourcen, subjektiver Bewertung und Verarbeitung von Belastung für die Gesundheit von Lehrern und Lehrerinnen. *Schulpädagogik heute*, *1*(2), 1–16.

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer.

Drexl, D. & Streb, J. (2016). Der Einfluss von Unterrichtsqualität auf die soziale und emotionale Schulerfahrung von Erstklässlern. In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler & R. Schmidt (Hrsg.), *Facetten grundschulpädagogischer und -didaktischer Forschung* (S. 27–32). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-11944-7_4

Drüke-Noe, C., Gniewosz, B. & Paasch, D. (2022). Mathematisches Selbstkonzept und Schülerleistungen – Zusammenhänge für den Unterricht nutzbar machen. In A. C. George, S. Götz, M. Illetschko & E. Süss-Stepancik (Hrsg.), *Empirische Befunde zu Kompetenzen im Mathematikunterricht der Sekundarstufe 1 und Folgerungen für die Praxis. Ergänzende Analysen zu den Bildungsstandardüberprüfungen* (S. 211–232). Münster: Waxmann. https://doi.org/10.31244/9783830995586

Dubberke, T., Kunter, M., McElvany, N., Brunner, M. & Baumert, J. (2008). Lerntheoretische Überzeugungen von Mathematiklehrkräften. Einflüsse auf die Unterrichtsgestaltung und den Lernerfolg von Schülerinnen und Schülern. *Zeitschrift für Pädagogische Psychologie*, *22*(3–4), 193–206. https://doi.org/10.1024/1010-0652.22.34.193

Eccles, J. S. & Wigfield, A. (2002). Motivational Beliefs, Values and Goals. *Annual Review of Psychology*, *53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J. S., Wigfield, A., Harold, R. D. & Blumenfeld, P. (1993). Age and Gender Differences in Children's Self- and Task Perceptions during Elementary School. *Child Development*, *64*(3), 830–847. https://doi.org/10.2307/1131221

Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz.

Eid, M., Gollwitzer, M. & Schmitt, M. (2015). *Statistik und Forschungsmethoden* (4. Aufl.). Weinheim: Beltz.

Ellis, H. (1894). *Man and Woman: A Study of Human Secondary Sexual Characters*. London: Scott.

Elsäßer, S. (2017). *Komponenten von schulischen Leistungen. Eine Analyse zu Einflussfaktoren auf die Notengebung in der Grundschule*. Ludwig-Maximilians-Universität: univ. Diss.

Enders, C. K. (2010). *Applied Missing Data Analysis*. New York: The Gilford Press.

Eriksson, K., Lindvall, J., Helenius, O. & Ryve, A. (2021). Socioeconomic Status as a Multidimensional Predictor of Student Achievement in 77 Societies. *Frontiers in Psychology*, *6*, 1–10. https://doi.org/10.3389/feduc.2021.731634

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W. & Beechum, N. O. (2012). *Teaching Adolescents to Become Learners. The Role of Noncognitive Factors in Shaping School Performance: A Critical Literature Review*. Chicago: University of Chicago Consortium on Chicago School Research.

Fauth, B., Decristan, J., Decker, A.-T., Büttner, G., Hardy, I., Klieme, E. & Kunter, M. (2019). The Effects of Teacher Competence on Student Outcomes in Elementary Science Education: The Mediating Role of Teaching Quality. *Teaching and Teacher Education*, *86*, 1–14. https://doi.org/10.1016/j.tate.2019.102882

Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014a). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive. Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, *28*(3), 127–137. https://doi.org/10.1024/1010-0652/a000129

Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014b). Student Ratings of Teaching Quality in Primary School. Dimensions and Prediction of Student Outcomes. *Learning and Instruction*, *29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fauth, B. & Leuders, T. (2018). *Kognitive Aktivierung im Unterricht*. Stuttgart: Landesinstitut für Schulentwicklung.

Feather, N. T. (1992). Values, Valences, Expectations, and Actions. *Journal of Social Issues*, *48*(2), 109–124. https://doi.org/10.1111/j.1540-4560.1992.tb00887.x

Fend, H. (1977). *Schulklima: soziale Einflussprozesse in der Schule*. Weinheim: Beltz.

Fend, H. (1980). *Theorie der Schule*. München: Urban und Schwarzenberg.

Fischer, C. (2015). Potenzialorientierter Umgang mit Vielfalt. Individuelle Förderung im Kontext Inklusiver Bildung. In C. Fischer (Hrsg.), *(Keine) Angst vor Inklusion: Herausforderungen und Chancen gemeinsamen Lernens in der Schule* (S. 21–35). Münster: Waxmann.

Fischer, L. & Wiswede, G. (2009). *Grundlagen der Sozialpsychologie*. München: Oldenbourg Verlag. https://doi.org/10.1524/9783486847826

Fishbein, B., Foy, P. & Yin, L. (2021). *TIMSS 2019. User Guide for the International Database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Florin, M., Stebler, R., Pauli, C. & Reusser, K. (2022). Selbstwirksamkeit von Lehrpersonen bei der Entwicklung einer personalisierten Lehr-Lernkultur. *Zeitschrift für Bildungsforschung*, *12*, 515–534. https://doi.org/10.1007/s35834-022-00366-2

Forgasz, H. J. & Hill, J. C. (2013). Factors Implicated in High Mathematics Achievement. *International Journal of Science and Mathematics Education*, *11*, 481–499. https://doi.org/10.1007/s10763-012-9348-x

Förtsch, C., Werner, S., Dorfner, T., von Kotzebue, L. & Neuhaus, B. J. (2016). Effects of Cognitive Activation in Biology Lessons on Students' Situational Interest and Achievement. *Research in Science Education*, *47*, 559–578. https://doi.org/10.1007/s11165-016-9517-y

Förtsch, C., Werner, S., von Kotzebue, L. & Neuhaus, B. J. (2016). Effects of Biology Teachers' Professional Knowledge and Cognitive Activation on Students' Achievement. *International Journal of Science Education*, *38*(17), 2642–2666. https://doi.org/10.1080/09500693.2016.1257170

Foy, P. & LaRoche, S. (2020). Estimating Standard Errors in the TIMSS 2019 Results. In M. O. Martin, M. von Davier & I. V. S. Mullis (Hrsg.), *Methods and Procedures: TIMSS 2019 Technical Report* (S. 14.11–14.60). Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Fraser, B. J., Walberg, H. J., Welch, W. W. & Hattie, J. (1987). Syntheses od Educational Productivity Research. *International Journal of Educational Research*, *11*(2), 147–252.

Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R. (2009). *PISA 2006 Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

Frey, M. C. & Dettermann, D. K. (2004). Scholastic Assessment or g? The Relationship Between the Scholastic Assessment Test and General Cognitive Ability. *Psychological Science*, *15*(6), 373–378. https://doi.org/10.1111/j.0956-7976.2004.00687.x

Fromm, S. (2008). Faktorenanalyse. In N. Baur & S. Fromm (Hrsg.), *Datenanalyse mit SPSS für Fortgeschrittene* (2. Aufl., S. 314v344). Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-91034-5_15

Gäde, J. C., Schermelleh-Engel, K. & Brandt, H. (2020). Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 615–660). Wiesbaden: Springer. https://doi.org/10.1007/978-3-662-61532-4_24

Gagné, F. (1985). Giftedness and Talent: Reexamining a Reexamination of the Definitions. *Gifted Child Quarterly*, *29*(3), 103–112. https://doi.org/10.1177/001698628502900302

Gagné, F. (2013). The DMGT: Changes Within, Beneath, and Beyond. *Talent Development & Excellence*, *5*(1), 5–19.

Gagné, P. & Hancock, G. R. (2006). Measurement Model Quality, Sample Size, and Solution Propriety in Confirmatory Factor Models. *Multivariate Behavioral Research*, *41*(1), 65–83. https://doi.org/10.1207/s15327906mbr4101_5

Gao, C., Shi, D. & Maydeu-Olivares, A. (2020). Estimating the Maximum Likelihood Root MeanSquare Error of Approximation (RMSEA) with Non-normal Data: A Monte-Carlo Study. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*, 192–201. https://doi.org/10.1080/10705511.2019.1637741

Gebauer, M. M. (2013). *Determinanten der Selbstwirksamkeitsüberzeugung von Lehrenden. Schulischer Berufsalltag an Gymnasien und Hauptsschulen*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-00613-6

Gebauer, S. (2019). *Förderung von Lehrerkompetenzen zur adaptiven Unterrichtsgestaltung. Zum Potenzial situierter Lernumgebungen in der Lehrerfortbildung*. Bad Heilbrunn: Julius Klinkhardt.

Gehrer, K. & Nusser, L. (2020). Binnendifferenzierender Deutschunterricht und dessen Einfluss auf die Lesekompetenzentwicklung in der Sekundarstufe I. *Journal for Educational Research Online*, *12*(2), 166–189. https://doi.org/10.25656/01:20976

Geiser, C. (2010). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-92042-9

Gilleece, L., Cosgrove, J. & Sofroniou, N. (2010). Equity in Mathematics ans Science Outcomes: Characteristics Associated with High and Low Achievement on PISA 2006 in Ireland. *International Journal of Science and Mathematics Education*, *8*, 475–496. https://doi.org/10.1007/s10763-010-9199-2

Glaser, R. & Chi, M. T. H. (2009). Overwiev. In M. T. H. Chi, R. Glaser & M. J. Farr (Hrsg.), *The Nature of Expertise* (S. xv–xxiix). New York: Psychology Press.

Goldstein, H. (1999). *Multilevel Statistical Models*. London: Institute of Education.

Göllner, R., Wagner, W., Klieme, E., Lüdtke, O., Nagengast, B. & Trautwein, U. (2016). Erfassung der Unterrichtsqualität mithilfe von Schülerurteilen: Chancen, Grenzen und Forschungsperspektiven. In BMBF (Hrsg.), *Forschungsvorhaben in Ankopplung an Large-Scale-Assessments* (S. 63–82). Berlin: Bundesministerium für Bildung und Forschung. https://doi.org/10.25656/01:12674

Gonzalez, E. J. (1998). Reporting Student Achievement in Mathematics and Science Literacy, Advanced Mathematics, and Physics. In M. O. Martin & D. L. Kelly (Hrsg.), *Third International Mathematics and Science Study. Technical Report Volume III: Implementation and Analysis. Final Year of Secondary School (Population 3)* (S. 121–140). Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Gottfredson, L. S. (1997). Mainstream Science on Intelligence: An Editorial With 52 Signatories, History, and Bibliography. *Intelligence*, *24*(1), 13–23.

Grabner, R. H. & Meier, M. A. (2021). Die Entwicklung von Expertise. In V. Müller-Oppliger & G. Weigand (Hrsg.), *Handbuch Begabung* (S. 149–167). Weinheim: Beltz.

Graham, J. W., Taylor, B. J., Olchowski, A. E. & Cumsille, P. E. (2006). Planned Missing Data Designs in Psychological Research. *Psychological Methods*, *11*(4), 323–343. https://doi.org/10.1037/1082-989X.11.4.323

Grams, S. L. (2014). *Zufriedene Lehrer - zufriedene Lerner? Der Zusammenhang zwischen berufsgezogenem Lehrerwohlbefinden und der Lehrer-Schüler-Beziehung im Unterrichtsalltag*, Universität Hildesheim & Aalborg University: univ. Diss.

Gräsel, C. (2020). Prozess-Produkt-Paradigma in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch der Lehrerinnen- und Lehrerbildung* (S. 148–153). Bad Heilbrunn: Verlag Julius Klinkhardt. https://doi.org/10.35468/hblb2020-017

Grigutsch, S., Raatz, U. & Törner, G. (1998). Einstellungen gegenüber Mathematik bei Mathematiklehrern. *Journal für Mathematik-Didaktik*, *19*(1), 3–45. https://doi.org/10.1007/BF03338859

Grootenboer, P. & Hemmings, B. (2007). Mathematics Performance and the Role Played by Affective and Background Factors. *Mathematics Education Research Journal*, *19*(3), 3–20. https://doi.org/10.1007/BF03217459

Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung.* Münster: Waxmann.

Guay, F., Roy, A. & Valois, P. (2017). Teacher structure as a predictor of students' perceived competence and autonomous motivation: The moderating role of differentiated instruction. *British Journal of Educational Psychology*, *87*, 224–240. https://doi.org/10.1111/bjep.12146

Guo, Y., McDonald Connor, C., Yang, Y., Roehrig, A. D. & Morrison, F. J. (2012). The Effects of Teacher Qualification, Teacher Self-Efficacy, and Classroom Practices on Fifth Graders' Literacy Outcomes. *The Elementary School Journal*, *113*(1), 3–24. https://doi.org/10.1086/665816

Gustafsson, J.-E. & Undheim, J. O. (1996). Individual Differences in Cognitive Functions. *Handbook of Educational Psychology*, 186–242.

Gutman, L. M. & Schoon, I. (2013). *The Impact of Non-Cognitive Skills on Outcomes for Young People*. Education Endowment Foundation: London.

Häcker, T. (2017). Individualisierter Unterricht. In T. Bohl, J. Budde & M. Rieger-Ladich (Hrsg.), *Umgang mit Heterogenität in Schule und Unterricht* (S. 275–290). Bad Heilbrunn: Klinkhardt.

Haddock, G. & Maio, G. R. (2014). Einstellungen. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 197–230). Wiesbaden: Springer. https://doi.org/10.1007/978-3-642-41091-8_6

Hadeler, S. (2015). *Fordern und Fördern. Leistungsanforderungen und Differenzierung in der Lern- und Förderumwelt privater Grundschulen*. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-09876-6

Hansford, B. C. & Hattie, J. A. (1982). The Relationship Between Self and Achievement/Performance Measures. *Review of Educational Research*, *52*(1), 123–142. https://doi.org/10.3102/00346543052001123

Harrison, M. G., King, R. B. & Wang, H. (2023). Satisfied Teachers Are Good Teachers: The Association Between Teacher Job Satisfaction and Instructional Quality. *British Educational Research Journal*, *49*(3), 476–498. https://doi.org/10.1002/berj.3851

Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer. https://doi.org/10.1007/3-540-33020-8_9

Hattie, J. A. C. (2009). *Visible Learning. A Synthesis of Over 800 Metaanalyses Relating to Achievement*. New York: Routledge.

Hauenstein, H., Pfeil, S. & Wittmer-Gerber, S. (2022). *Vivo – Bausteine für die 4. Klasse zur Begleitung des Übergangs. Zukunftschancen für Kinder am Übergang in die Sekundarstufe I*. Berlin: sdw.

Heckhausen, H. (1991). *Motivation and Action*. Berlin: Springer. https://doi.org/10.1007/978-3-642-75961-1

Heinitz, B. & Nehring, A. (2020). Kriterien naturwissenschaftsdidaktischer Unterrichtsqualität – ein systematisches Review videobasierter Unterrichtsforschung. *Unterrichtswissenschaft*, *48*, 319–360. https://doi.org/10.1007/s42010-020-00074-8

Heinitz, B., Szogs, M., Förtsch, C., Korneck, F., Neuhaus, B. J. & Nehring, A. (2022). Unterrichtsqualität in den Naturwissenschaften. Eine vergleichende Gegenüberstellung von Ansätzen zwischen Fachspezifik und Generik. *Zeitschrift für Didaktik der Naturwissenschaften*, *28*, 169–186. https://doi.org/10.1007/s42010-020-00074-8

Heller, K. (1976). *Intelligenz und Begabung*. München: E. Reinhardt.

Heller, K. (1991). *Begabungsdiagnostik in der Schul- und Erziehungsberatung*. Bern: Huber.

Heller, K. & Perleth, C. (2007). Talentförderung und Hochbegabtenberatung in Deutschland. In K. Heller & A. Ziegler (Hrsg.), *Begabt sein in Deutschland* (S. 139–170). Berlin: Lit.

Hellmich, F. & Günther, F. (2011). Entwicklung von Selbstkonzepten bei Kindern im Grundschulalter – ein Überblick. In F. Hellmich (Hrsg.), *Selbstkonzepte im Grundschulalter* (S. 19–46). Stuttgart: Kohlhammer. https://doi.org/10.17433/978-3-17-022863-4

Helmke, A. (1992). *Selbstvertrauen und schulische Leistungen*. Göttingen: Hogrefe.

Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5. Klassenstufe. *Zeitschrift für Pädagogische Psychologie*, *7*(2–3), 77–86.

Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzeptes vom Kindergarten bis zur 6. Klassenstufe. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 119–132). Weinheim: Beltz.

Helmke, A. (2007). *Unterrichtsqualität erfassen, bewerten, verbessern* (5. Aufl.). Seelze: Klett Kallmeyer.

Helmke, A., Helmke, T., Heyne, N., Hosenfeld, A., Schrader, F.-W. & Wagner, W. (2010). Effiziente Klassenführung als Schlüsselmerkmal der Unterrichtsqualität – ein Untersuchungsbeispiel aus der Grundschule. In C. Spiel, R. Reimann, B. Schober & P. Wagner (Hrsg.), *Bildungspsychologie* (S. 101–105). Göttingen: Hogrefe.

Helmke, A. & Schrader, F.-W. (2008). Merkmale der Unterrichtsqualität: Potenzial, Reichweite und Grenzen. *Seminar - Lehrerbildung und Schule*, *3*, 17–47.

Helmke, A. & Schrader, F.-W. (2010). Merkmale der Unterrichtsqualität. Potenzial, Reichweite und Grenzen. In B. Schaal & F. Huber (Hrsg.), *Qualitätssicherung im Bildungswesen. Auftrag und Anspruch der bayerischen Qualitätsagentur* (S. 69–108). Münster: Waxmann.

Helmke, A. & van Aken, M. A. G. (1995). The Causal Ordering of Academic Achievement and Self-Concept of Ability During Elementary School: A Longitudinal Study. *Journal of Educational Psychology*, *87*(4), 624–637. https://doi.org/10.1037/0022-0663.87.4.624

Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 71–176). Göttingen: Hogrefe.

Helmke, A. & Weinert, F. E. (2017). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (7. Aufl.). Seelze-Velber: Klett-Kallmeyer.

Hemmings, B. & Kay, R. (2010). Prior Achievement, Effort, And Mathematics Attitude as Predictors of Current Achievement. *The Australian Educational Researcher*, *37*(2), 41–58. https://doi.org/10.1007/BF03216921

Henke, T., Bosse, S. & Spörer, N. (2018). Binnendifferenzierung im inklusiven Unterricht. Ein Vergleich der Schüler- und Beobachterperspektive. In F. Hellmich, G. Görel & M. F. Löper (Hrsg.), *Inklusive Schul- und Unterrichtsentwicklung. Vom Anspruch zur erfolgreichen Umsetzung* (S. 143–159). Stuttgart: Kohlhammer. https://doi.org/10.17433/978-3-17-032720-7

Henschel, S., Rjosk, C. & Heinschel, A. (2023). Merkmale der Unterrichtsqualität im Fach Deutsch. In P. Stanat, S. Schipolowski, R. Schneider, S. Weirich, S. Henschel & K. A. Sachse (Hrsg.), *IQB-Bildungstrend 2022. Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im dritten Ländervergleich* (S. 359–388). Münster: Waxmann. https://doi.org/10.31244/9783830997771

Henschel, S., Rjosk, C., Holtmann, M. & Stanat, P. (2019). Merkmale der Unterrichtsqualität im Fach Mathematik. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 355–384). Münster: Waxmann.

Hertel, S. (2014). Adaptive Lerngelegenheiten in der Grundschule. Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. In B. Kopp, S. Martschinke, M. Munser-Kiefer, M. Haider, E.-M. Kirschhock, G. Ranger & G. Renner (Hrsg.), *Individuelle Förderung und Lernen in der Gemeinschaft* (S. 19–34). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-04479-4_2

Hess, M. & Lipowsky, F. (2016). Lernen individualisieren und Unterrichtsquailtät verbessern. In F. Heinzel & K. Koch (Hrsg.), *Individualisierung im Grundschulunterricht. Anspruch, Realisierung und Risiken* (S. 23–31). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-15565-0_3

Hidayah, R., Wangid, M. N., Wuryandani, W. & Salimi, M. (2023). The Influence of Teacher Efficacy on Education Quality: A Meta-Analysis. *International Journal of Educational Methodology*, *9*(2), 435–450. https://doi.org/10.12973/ijem.9.2.435

Hilgenfeld, F. (2012). Die Berufszufriedenheit von Lehrern, ihre Einflussfaktoren und Auswirkungen auf Lebens- und Schulqualität. In I. Mortag & I. Nowosad (Hrsg.), *Qualität des Lebens und Qualität der Schule. Wohlfühlen in der Schule aus der Sicht der Beteiligten* (S. 119–128): Uniwersytet Zielonogórski.

Hillmayr, D., Täschner, J., Brockmann, L. & Holzberger, D. (2021). *Elternbeteiligung im schulischen Kontext – Potenzial zur Förderung des schulischen Erfolgs von Schülerinnen und Schülern*. Münster: Waxmann. https://doi.org/10.31244/9783830993667

Hochweber, J., Hosenfeld, I. & Klieme, E. (2014). Classroom Composition, Classroom Management, and the Relationship Between Student Attributes and Grades. *Journal of Educational Psychology*, *106*(1), 289–300. https://doi.org/10.1037/a0033829

Hofmann, F. & Gottein, H.-P. (2011). Zusammenhänge zwischen Persönlichkeitsmerkmalen und der Bereitschaft zur Individualisierung im Unterricht. *Zeitschrift für Bildungsforschung*, *1*, 55–67. https://doi.org/10.1007/s35834-011-0005-1

Hoge, D. R., Smit, E. K. & Crist, J. T. (1995). Reciprocal Effects of Self-Concept and Academic Achievement in Sixth and Seventh Grade. *Journal of Youth and Adolescence*, *24*(3), 295–314. https://doi.org/10.1007/BF01537598

Holzberger, D., Philip, A. & Kunter, M. (2013). How Teachers' Self-Efficacy Is Related to Instructional Quality: A Longitudinal Analysis. *Journal of Educational Psychology*, *105*(3), 774–786. https://doi.org/10.1037/a0032198

Homans, G. C. & Prokop, D. (1968). *Elementarformen sozialen Verhaltens*. Köln: Westdeutscher Verlag. https://doi.org/10.1007/978-3-663-02391-3

Homburg, C. & Baumgartner, H. (1995). Beurteilung von Kausalmodellen. Bestandsaufnahme und Anwendungsempfehlungen. *Marketing: ZFP - Journal of Research and Management*, *3*, 162–176. https://doi.org/10.15358/0344-1369-1995-3-162

Hooper, M. (2016). Developing the TIMSS 2015 Context Questionnaires. In M. O. Martin, I. V. S. Mullis & M. Hooper (Hrsg.), *Methods and Procedures in TIMSS 2015*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Hoover-Dempsey, K. V., Walker, J. M. T., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S. & Closson, K. (2005). Why Do Parents Become Involved? Research Findings and Implications. *The Elementary School Journal*, *106*(2), 105–130. https://doi.org/10.1086/499194

Horak, V. M. (1981). A Meta-Analysis of Research Findings on Individualized Instruction in Mathematics. *The Journal of Educational Research*, *74*(4), 249–253. https://doi.org/10.1080/00220671.1981.10885318

Horn, J. L. & Cattell, R. B. (1966). Refinement and Test of the Theory of Fluid and Crystallized General Intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. https://doi.org/10.1037/h0023816

Hosenfeld, I., Helmke, A., Ridder, A. & Schrader, F.-W. (2001). Eine mehrebenenanalytische Betrachtung von Schul- und Klasseneffekten. *Empirische Pädagogik*, *15*, 513–534.

Hosenfeld, I., Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz. Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 65–82). Weinheim: Beltz.

Hox, J. J. (2010). *Multilevel Analysis. Techniques and Applications* (2. Aufl.). New York: Routledge. https://doi.org/10.4324/9780203852279

Hox, J. J., Moerbeek, M. & van de Schoot, R. (2018). *Multilevel Analysis. Techniques and Applications* (3. Aufl.). New York: Routledge. https://doi.org/10.4324/9781315650982

Hu, L.-t. & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, C. (2011). Self-concept and Academic Achievement: A Meta-Analysis of Longitudinal Relations. *Journal of School Psychology*, *49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Huber, F., Herrmann, A., Meyer, F., Vogel, J. & Vollhardt, K. (2007). *Kausalmodellierung mit Partial Least Squares. Eine anwendungsorientierte Einführung.* Wiesbaden: Gabler.

Hughes, J. N., Luo, W., Kwok, O.-M. & Loyd, L. K. (2008). Teacher–Student Support, Effortful Engagement, and Achievement: A 3-Year Longitudinal Study. *Journal of Educational Psychology*, *100*(1), 1–14. https://doi.org/10.1037/0022-0663.100.1.1

Hußmann, A., Wendt, H., Kasper, D., Bos, W. & Goy, M. (2017). Ziele, Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU 2016). In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Voß, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 29–78). Münster: Waxmann.

Hüther, G. & Hauser, U. (2012). *Jedes Kind ist hoch begabt. Die angeborenen Talente unserer Kinder und was wir aus ihnen machen.* München: Knaus.

Inckemann, E. (2014). Binnendifferenzierung - Individualisierung - adaptiver Unterricht. In W. Einsiedler, M. Götz, A. Hartinger & F. Heinzel (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (4. Aufl.). Bad Heilbrunn: Klinkhardt.

iPEGE. (2009). *Professionelle Begabtenförderung. Empfehlungen zur Qualifizierung von Fachkräften in der Begabtenförderung.* Österreichisches Zentrum für Begabtenförderung und Begabungsforschung (özbf).

Jäckle, S. & Schärdel, J. (2017). Mehrebenenanalyse. In S. Jäckle (Hrsg.), *Neue Trends in den Sozialwissenschaften. Innovative Techniken für qualitative und quantitative Forschung* (S. 147–176). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-17189-6_6

Jackson-Crossland, B. A. (2000). *The Relationships Between Teacher Empowerment, Teachers' Sense of Responsibility for Student Outcomes, and Student Achievement.* University of Missouri: univ. Diss.

Jaeuthe, J., Lenkeit, J., Bosse, S. & Spörer, N. (2023). Zur Bedeutsamkeit der Unterrichtsqualität für die Entwicklung von Rechtschreibkompetenzen in der Grundschule. *Zeitschrift für Erziehungswissenschaft*, *26*, 1571–1596. https://doi.org/10.1007/s11618-023-01188-2

Janke, N. (2006). *Soziales Klima an Schulen aus der Lehrer-, Schulleiter- und Schülerperspektive. Eine Sekundäranalyse der Studie >Kompetenzen und Einstellungen von Schülerinnen und Schülern - Jahrgangsstufe 4 (KESS 4)<*. Münster: Waxmann.

Jansen, M., Schroeders, U. & Lüdtke, O. (2014). Academic Self-Concept in Science: Multidimensionality, Relations to Achievement Measures, and Gender Differences. *Learning and Individual Differences*, *30*, 11–21. https://doi.org/10.1016/j.lindif.2013.12.003

Jennek, J., Gronostaj, A. & Vock, M. (2019). Wie Lehrkräfte im Englischunterricht differenzieren. Eine Re-Analyse der DESI-Videos. *Unterrichtswissenschaft*, *47*, 99–116. https://doi.org/10.1007/s42010-018-0027-7

Jennings, P. A. & Greenberg, M. T. (2009). The Prosocial Classroom: Teacher Social and Emotional Competence in Relation to Student and Classroom Outcomes. *Review of Educational Research*, *79*(1), 491–525. https://doi.org/10.3102/0034654308325693

Jensen, A. R. (1980). *Bias in Mental Testing*. New York: The Free Press.

Jerusalem, M. & Schwarzer, R. (Hrsg.). (1999). *Förderung von Selbstwirksamkeit bei Schülern und Lehrern. Abschlußbericht der wissenschaftlichen Begleitung des Modellversuchs Verbund Selbstwirksame Schulen*. Berlin.

Johansone, I. & Flicop, S. (2020). International Quality Assurance Program for TIMSS 2019. In M. O. Martin, M. von Davier & I. V. S. Mullis (Hrsg.), *Methods and Procedures: TIMSS 2019 Technical Report* (S. 7.1–7.25). Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Jöreskog, K. G. (1999). How Large Can a Standardized Coefficient be? https://www.statmodel.com/download/Joreskog.pdf

Judge, T. A., Weiss, H. M., Kammeyer-Mueller, J. D. & Hulin, C. L. (2017). Job Attitudes, Job Satisfaction, and Job Affect: A Century of Continuity and of Change. *Journal of Applied Psychology*, *102*(3), 356–374. https://doi.org/10.1037/apl0000181

Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M. & Hoth, J. (2016). Professional competencies of (prospective) mathematics teachers – cognitive versus situated approaches. *Educational Studies in Mathematics*, *94*, 161–182. https://doi.org/10.1007/s10649-016-9713-8

Kaiser, H. F. (1974). An Index of Factorial Simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kaiser, H. F. & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kartal, S. K. & Kutlu, Ö. (2017). Identifying the Relationships between Motivational Features of High and Low Performing Students and Science Literacy Achievement in PISA 2015 Turkey. *Journal of Education and Training Studies*, *5*(12), 146–154. https://doi.org/10.11114/jets.v5i12.2816

Kelava, A. & Brandt, H. (2014). A general non-linear multilevel structural equation mixture model. *Frontiers in Psychology*, *5*, 1–16. https://doi.org/10.3389/fpsyg.2014.00748

Kelley, J., Evans, M. D. R., Lowman, J. & Lykes, V. (2017). Group-mean-centering independent variables in multi-level models is dangerous. *Quality & Quantity*, *51*, 261–283. https://doi.org/10.1007/s11135-015-0304-z

Keskin, S. (2006). Comparison of Several Univariate Normality Tests Regarding Type I Error Rate and Power of the Test in Simulation based Small Samples. *Journal of Applied Science Research*, *2*(5), 296–300.

Kibrislioglu, N. (2015). An Investigation About 6th Grade Students' Attitudes Towards Mathematics. *Procedia – Social and Behavioral Sciences*, *186*, 64–69. https://doi.org/10.1016/j.sbspro.2015.04.024

Kienle, S. & Kopp, B. (2014). „Die Entdeckerreise in die neue Schule" – Eine Intervention zur Persönlichkeitsförderung im Übergang von der Grundschule in die Sekundarstufe. In D. Blömer, M. Lichtblau, A.-K. Jüttner, K. Koch, M. Krüger & R. Werning (Hrsg.), *Perspektiven auf inklusive Bildung. Gemeinsam anders lehren und lernen* (S. 328–334). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-06955-1_50

Kim, K. H. & Bentler, P. M. (2002). Tests of Homogenity of Means and Covariance Matrices for Multivariate Incomplete Data. *Psychometrica*, *67*(4), 609–624. https://doi.org/10.1007/BF02295134

Klafki, W. (2007). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik*. Weinheim: Beltz.

Klassen, R. M. & Tze, V. M. C. (2014). Teachers' Self-Efficacy, Personality, and Teaching Effectiveness: A Meta-Analysis. *Educational Research Review*, *12*, 59–76. https://doi.org/10.1016/j.edurev.2014.06.001

Klassen, R. M., Tze, V. M. C., Betts, S. M. & Gordon, K. A. (2011). Teacher Efficacy Research 1998–2009: Signs of Progress or Unfulfilled Promise? *Educational Psychology Review*, *23*, 21–43. https://doi.org/10.1007/s10648-010-9141-8

Kleickmann, T., Steffensky, M. & Praetorius, A.-K. (2020). Quality of Teaching in Science Education. More Than Three Basic Dimensions? In A.-K. Praetorius, J. Grünkorn & E. Klieme (Hrsg.), *Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen* (S. 37–55). Weinheim: Beltz.

Kleickmann, T., Vehmeyer, J. & Möller, K. (2010). Zusammenhänge zwischen Lehrervorstellungen und kognitiven Strukturierungen im Unterricht am Beispiel von Scaffolding-Maßnahmen. *Unterrichtswissenschaft*, *38*(3), 210–228.

Kleinke, K., Schlüter, E. & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus* (2. Aufl.). Oldenbourg: De Gruyter. https://doi.org/10.1515/9783486989458

Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, *6*, 10–13.

Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, *54*(2), 222–237.

Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur" und Unterrichtsgestaltung im internationalen Vergleich. In E. Klieme & J. Baumert (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43–57). Bonn: BMBF.

Klieme, E. & Tippelt, R. (2008). Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz. In E. Klieme & R. Tippelt (Hrsg.), *Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz* (S. 7–13). Weinheim: Beltz.

Kline, R. B. (2016). *Principals and Practice of Structural Equation Modeling*. New York: The Guilford Press.

Kloss, J. (2014). *Grundschüler als Experten für Unterricht. Empirische Überprüfung der Validität von Unterrichtsbeurteilungen durch Schüler der dritten und vierten Jahrgangsstufe*. Frankfurt a. M.: Peter Lang. https://doi.org/10.3726/978-3-653-03897-2

Klusmann, U., Aldrup, K., Roloff, J., Lüdtke, O. & Hamre, B. K. (2022). Does Instructional Quality Mediate the Link Between Teachers' Emotional Exhaustion and Student Outcomes? A Large-Scale Study Using Teacher and Student Reports. *Journal of Educational Psychology*, *114*(6), 1442–1460. https://doi.org/10.1037/edu0000703

Klusmann, U., Richter, D. & Lüdtke, O. (2016). Teachers' Emotional Exhaustion Is Negatively Related to Students' Achievement: Evidence From a Large-Scale Assessment Study. *Journal of Educational Psychology*, *108*(8), 1193–1203. https://doi.org/10.1037/edu0000125

KMK. (2009). *Grundsatzposition der Länder zur begabungsgerechten Förderung*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2009/2009_12_12-Begabungsgerechte-Foerderung.pdf

KMK. (2015). Förderstrategie für leistungsstarke Schülerinnen und Schüler. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Foerderstrategie-leistungsstarke-Schueler.pdf

KMK. (2016). Gemeinsame Initiative von Bund und Ländern zur Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler. https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2018/Beschluss_zur_gemeinsamen_Bund-Laender-Initiative.pdf

Knoblauch, H. (2009). Von der Kompetenz zur Performanz. Wissenschaftssoziologische Aspekte der Kompetenz. In T. Kurtz & M. Pfadenhauer (Hrsg.), *Soziologie der Kompetenz* (S. 237–255). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91951-5_13

Köller, O. (2004). *Konsequenzen von Leistungsgruppierungen*. Münster: Waxmann.

Köller, O. & Baumert, J. (2018). Schulische Leistungen und ihre Messung. In W. Schneider & U. Lindenberger (Hrsg.), *Entwicklungspsychologie* (S. 681–695). Weinheim: Beltz.

Köller, O. & Möller, J. (2010). Selbstwirksamkeit. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 767–773). Weinheim: Beltz.

Köller, O., Trautwein, U., Lüdtke, O. & Baumert, J. (2006). Zum Zusammenspiel von schulischer Leistung, Selbstkonzept und Interesse in der gymnasialen Oberstufe. *Zeitschrift für Pädagogische Psychologie*, *20*(1/2), 27–39. https://doi.org/10.1024/1010-0652.20.12.27

König, J., Felske, C. & Kaiser, G. (2023). Professionelle Kompetenz von Mathematiklehrkräften aus einer pädagogischen Perspektive. In S. Krauss & A. Lindl (Hrsg.), *Professionsiwssen von Mathematiklehrkräften. Implikationen aus der Forschung für die Praxis* (S. 293–331). Wiesbaden: Springer. https://doi.org/10.1007/978-3-662-64381-5_9

König, J. & Pflanzl, B. (2015). Is Teacher Knowledge Associated With Performance? On The Relationship Between Teachers' General Pedagogical Knowledge and Instructional Quality. *European Journal of Teacher Education*, *39*(4), 419–436. https://doi.org/10.1080/02619768.2016.1214128

Köster, H., Mehrtens, T. & Müller, F. (2022). Merkmale naturwissenschaftlicher Begabung bei Kindern im Kita- und Grundschulalter. In G. Weigand, C. Fischer, F. Käpnick, C. Perleth, F. Preckel, M. Vock & H.-W. Wollersheim (Hrsg.), *Dimensionen der Begabungs- und Begabtenförderung in der Schule*. Bielefeld: wbv Publikation.

Kounin, J. S. (1976). *Techniken der Klassenführung*. Stuttgart: Klett-Kallmeyer.

Krampen, G. (1978). Über die Beziehung von Berufszufriedenheit und beruflichen Wertorientierungen bei Lehrern. *Psychologie und Praxis*, *22*(2), 50–57.

Krampen, G. (1981). Berufszufriedenheit und Zielorientierungen von Lehrern. Interdependenzanalyse zweier motivationaler Variablenkomplexe. *Psychologie und Praxis*, *25*(2), 45–55.

Krapp, A. (2010). Interesse. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 311–323). Weinheim: Beltz.

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N. & Blum, W. (2020). Competence As a Continuum in The COACTIV Study: The "Cascade Model". *ZDM - Mathematics Education*, *52*(3), 311–327. https://doi.org/10.1007/s11858-020-01151-z

Kreft, I. G. G., de Leeuw, J. & Aiken, L. S. (1995). The Effect of Different Forms of Centering in Hierarchical Linear Models. *Multivariate Behavioral Research*, *30*(1), 1–21. https://doi.org/10.1207/s15327906mbr3001_1

Kriegbaum, K., Becker, N. & Spinath, B. (2018). The Relative Importance of Intelligence and Motivation as Predictors of School Achievement: A Meta-Analysis. *Methodological Reviews*, *25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Kuhl, J. (2021). Begabung, Bildung und Beziehung aus persönlichkeitspsychologischer Sicht. In V. Müller-Oppliger & G. Weigand (Hrsg.), *Handbuch Begabung* (S. 185–203). Weinheim: Beltz.

Kulik, J. A. & Kulik, C.-L. C. (1989). The Concept of Meta-Analysis. *International Journal of Educational Research*, *13*(3), 227–340.

Kulik, J. A. & Kulik, C.-L. C. (1992). Meta-Analytic Findings on Grouping Programs. *Gifted Child Quarterly*, *36*(2), 73–77. https://doi.org/10.1177/001698629203600204

Kullmann, H. (2012). Erwünschte Charakteristika von Partner/-innen für Lehrerkooperation. Eine empirische Analyse anhand der Selbstbestimmungstheorie der Motivation. In E. Baum, T.-S. Idel & H. Ullrich (Hrsg.), *Kollegialität und Kooperation in der Schule. Theoretische Konzepte und empirische Befunde* (S. 77–89). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-94284-1_5

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. https://doi.org/10.1037/0022-3514.86.1.148

Künstning, J., Neuber, V. & Lipowsky, F. (2016). Teacher Self-Efficacy as a Long-Term Predictor of Instructional Quality in The Classroom. *European Journal of Psychology of Education*, *31*, 299–322. https://doi.org/10.1007/s10212-015-0272-7

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.

Kunter, M. & Ewald, S. (2016). Bedingungen und Effekte von Unterricht: Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 9–31). Münster: Waxmann.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional Competence of Teachers: Effects on Instructional Quality and Student Development. *Journal of Educational Psychology*, *105*(3), 805–820. https://doi.org/10.1037/a0032583

Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Paderborn: utb.

Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85–114). Münster: Waxmann.

Kuusinen, C. M. (2016). *The Meaning and Measure of Teacher Self-Efficacy for Effective Classroom Teaching Practices*. University of Michigan: univ. Diss.

Kyed, O. (2018). *Hochbegabt. Der Ratgeber für Eltern*. Freiburg: Herder.

Kyriakides, L., Christoforou, C. & Charalambous, C. Y. (2013). What Matters for Student Learning Outcomes: A Meta-Analysis of Studies Exploring Factors of Effective Teaching. *Teaching and Teacher Education*, *36*, 143–152. https://doi.org/10.1016/j.tate.2013.07.010

Kyriakides, L. & Tsangaridou, N. (2008). Towards the Development of Generic and Differentiated Models of Educational Effectiveness: A Study on School and Teacher Effectiveness in Physical Education. *British Educational Research Journal*, *34*(6), 807–838. https://doi.org/10.1080/01411920802041467

Landwehr, B. (2017). *Partizipation, Wissen und Motivation im Politikunterricht. Eine Interventionsstudie*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-16507-9

Langer, W. (2010). Mehrebenenanalyse mit Querschnittsdaten. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 741–774). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92038-2_28

Lankes, E.-M., Bos, W., Mohr, I., Plaßmeier, N., Schwippert, K., Sibberns, H. & Voss, A. (Hrsg.). (2002). *Anlage und Durchführung der Internationalen Grundschule-Lese-Untersuchung (IGLU) und ihrer Erweiterung um Mathematik und Naturwissenschaften (IGLU-E)*. Münster: Waxmann.

Lay, Y. F. & Chandrasegaran, A. L. (2018). The Contribution of Teacher Preparation on TIMSS Grade 8 Students' Science Achievement: A Comparative Study between Malaysia and Singapore. *Journal of Baltic Science Education*, *17*(4), 576–589. https://doi.org/10.33225/jbse/18.17.576

Lazarides, R. & Buchholz, J. (2019). Student-perceived teaching quality: How is it related to different achievement emotions in mathematics classrooms? *Learning and Instruction*, *61*, 45–59. https://doi.org/10.1016/j.learninstruc.2019.01.001

Lee, V. E. (2000). Using Hierarchical Linear Modeling to Study Social Contexts: The Case of School Effects. *Educational Psychologist*, *35*(2), 125–141. https://doi.org/10.1207/S15326985EP3502_6

Lee, V. E. & Burkam, D. T. (1996). Gender Differences in Middle Grade Science Achievement: Subject Domain, Ability Level, and Course Emphasis. *Science Education*, *80*(6), 613–650.

Lehrl, S., Evangelou, M. & Sammons, P. (2020). *The Home Learning Environment and its Role in Shaping Children's Educational Development*: School Effectiveness and School Improvement, 31(1). https://doi.org/10.1080/09243453.2020.1693487

Lenske, G. (2016). *Schülerfeedback in der Grundschule. Untersuchungen zur Validität*. Münster: Waxmann.

Letzel, V. (2021). *Binnendifferenzierung in der Schulpraxis. Eine quantitative Studie zur Einsatzhäufigkeit und zu Kontextfaktoren*. Universität Trier.

Letzel, V. & Otto, J. (2019). Binnendifferenzierung und deren konkrete Umsetzung in der Schulpraxis. Eine qualitative Studie. *Zeitschrift für Bildungsforschung*, *9*, 375-393. https://doi.org/10.1007/s35834-019-00256-0

Li, C. (2013). Little's Test of Missing Completely at Random. *Tha Stata Journal*, *13*(4), 795–809. https://doi.org/10.1177/1536867X1301300407

Li, H., Liu, J., Zhang, D. & Liu, H. (2020). Examining The Relationships Between Cognitive Activation, Self-Efficacy, Socioeconomic Status, and Achievement in Mathematics: A Multi-Level Analysis. *British Journal of Educational Psychology*, *91*(1), 101–126. https://doi.org/10.1111/bjep.12351

Lin, J. & Bentler, P. M. (2012). *A Probability Based Test for Missing Completely at Random Data Patterns*. Vortrag gehalten auf National Council on Measurement in Education, Vancouver.

Lipnevich, A. A., Preckel, F. & Krumm, S. (2016). Mathematics Attitudes and Their Unique Contribution To Achievement: Going Over and Above Cognitive Ability and Personality. *Learning and Individual Differences*, *47*, 70–79. https://doi.org/10.1016/j.lindif.2015.12.027

Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern.* (S. 47–70). (Zeitschrift für Pädagogik; Beiheft 51). Weinheim: Beltz. https://doi.org/10.25656/01:7370

Lipowsky, F. & Lotz, M. (2015). Ist Individualisierung der Königsweg zum erfolgreichen Lernen? Eine Auseinandersetzung mit Theorien, Konzepten und empirischen Befunden. In G. Mehlhorn, K. Schöppe & F. Schulz (Hrsg.), *Begabungen entwickeln & Kreativität fördern* (S. 155–219). München: kopaed.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E. & Reusser, K. (2009). Quality of Geometry Instruction and Its Short-Term Impact on Students' Understanding of The Pythagorean Theorem. *Learning and Instruction*, *19*, 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Little, R. J. A. & Rubin, D. B. (2020). *Statistical Analysis wirh Missing Data* (3. Aufl.). New York: Wiley.

Locke, E. & Latham, G. P. (1990). *A Theory of Goal Setting & Task Performance*. Englewood Cliffs: Prentice-Hall.

Lopes, J. & Oliveira, C. (2020). Teacher and School Determinants of Teacher Job Satisfaction: A Multilevel Analysis. *School Effectiveness and School Improvement*, *31*(4), 641–659. https://doi.org/10.1080/09243453.2020.1764593

Lossen, K., Tillmann, K., Holtappels, H. G., Rollett, W. & Hannemann, J. (2016). Entwicklung der naturwissenschaftlichen Kompetenzen und des sachunterrichtsbezogenen Selbstkonzepts bei Schüler/-innen in Ganztagsgrundschulen. Ergebnisse der Längsschnittstudie StEG-P zu Effekten der Schülerteilnahme und der Angebotsqualität. *Zeitschrift für Pädagogik*, *62*(6), 760–779. https://doi.org/10.25656/01:16888

Lotz, M. (2016). *Kognitive Aktivierung im Leseunterricht der Grundschule. Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-10436-8

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B. & d'Appolonia, S. (1996). Within-Class Grouping: A Meta-Analysis. *Review of Educational Research*, *66*(4), 423–458. https://doi.org/10.2307/1170650

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. https://doi.org/10.1026/0033-3042.58.2.103

Lüdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the Impact of Learning Environments: How to Use Student Ratings of Classroom or School Characteristics in Multilevel Modeling. *Contemporary Educational Psychology*, *34*, 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Lüftenegger, M., Kollmayer, M., Bergsmann, E., Jöstl, G., Spiel, C. & Schober, B. (2015). Mathematically Gifted Students and Highachievement: The Role of Motivation and Classroom Structure. *High Ability Studies*, *26*(2), 227–243. https://doi.org/10.1080/13598139.2015.1095075

Ma, K., Luo, J., Cavanagh, M., Dong, J. & Sun, M. (2022). Measuring Teacher Self-Efficacy: Validating a New Comprehensive Scale Among Chinese Pre-Service Teachers. *Frontiers in Psychology*, *13*, 1063830. https://doi.org/10.3389/fpsyg.2022.1063830

Ma, X. & Kishor, N. (1997). Assessing the Relationship between Attitude toward Mathematics and Achievement in Mathematics: A Meta-Analysis. *Journal for Research in Mathematics Education*, *28*(1), 26–47. https://doi.org/10.2307/749662

Maag Merki, K. (2021). Schulqualitätsforschung. In T. Hascher (Hrsg.), *Handbuch Schulforschung* (3. Aufl., S. 1–22). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-24729-4_22

Maaz, K., Neumann, M., Trautwein, U., Wendt, W., Lehmann, R. & Baumert, J. (2008). Der Übergang von der Grundschule in die weiterführende Schule. Die Rolle von Schüler- und Klassenmerkmalen beim Einschätzen der individuellen Lernkompetenz durch die Lehrkräfte. *Schweizerische Zeitschrift für Bildungswissenschaften*, *30*(3), 519–548. https://doi.org/10.24452/sjer.30.3.4801

Maio, G. R. & Haddock, G. (2010). *The Psychology of Attitudes and Attitude Change*. London: Sage. https://doi.org/10.4135/9781446214299

Mandl, H. & Krause, U.-M. (2001). *Lehrkompetenz für die Wissensgesellschaft*. München: Ludwig-Maximilians-Universität.

Mao, P., Cai, Z., He, J., Chen, X. & Fan, X. (2021). The Relationship Between Attitude Toward Science and Academic Achievement in Science: A Three-Level Meta-Analysis. *Frontiers in Psychology*, *12*, 1–12. https://doi.org/10.3389/fpsyg.2021.784068

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.2307/1163048

Marsh, H. W. (1987). The Big-Fish-Little-Pond Effect on Academic Self-Concept. *Journal of Educational Psychology*, *79*(3), 280–295. https://doi.org/10.1037//0022-0663.79.3.280

Marsh, H. W. (1990). Causal Ordering of Academic Self-concept and Academic Achievement: A Multiwave, Longitudinal Panel Analysis. *Journal of Educational Psychology*, *82*(4), 646–656. https://doi.org/10.1037//0022-0663.82.4.646

Marsh, H. W., Byrne, B. M. & Shavelson, R. J. (1988). A Multifaceted Academic Self-Concept: Its Hierarchical Structure and Its Relation to Academic Achievement. *Journal of Educational Psychology*, *80*(3), 366–380. https://doi.org/10.1037//0022-0663.80.3.366

Marsh, H. W. & Craven, R. G. (2006). Reciprocal Effects of Self-Concept and Performance From a Multidimensional Perspective Beyond Seductive Pleasure and Unidimensional Perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Kong, C.-K. & Hau, K.-T. (2000). Longitudinal Multilevel Models of the Big Fish Little Pond Effect on Academic Self-concept: Counterbalancing Contrast and Reflected Glory Effects in Hong Kong Schools. *Journal of Personality and Social Psychology*, *78*(2), 337–349. https://doi.org/10.1037//0022-3514.78.2.337

Marsh, H. W. & Martin, A. J. (2011). Academic Self-Concept and Academic Achievement: Relations and Causal Ordering. *British Journal of Educational Psychology*, *81*, 59–77. https://doi.org/10.1348/000709910x503501

Marsh, H. W. & O'Mara-Eves, A. (2008). Reciprocal Effects Between Academic Self-Concept, Self-Esteem, Achievement, and Attainment Over Seven Adolescent Years: Unidimensional and Multidimensional Perspectives of Self-Concept. *Personality and Social Psychology*, *34*(4), 542–552. https://doi.org/10.1177/0146167207312313

Marsh, H. W. & Parker, J. W. (1984). Determinants of Student Self-Concept: Is It Better to Be a Relatively Large Fish in a Small Pond Even if You Don't Learn to Swim as Well? *Journal of Personality and Social Psychology*, *47*(1), 213–231. https://doi.org/10.1037//0022-3514.47.1.213

Marsh, H. W., Pekrun, R., Muryama, K., Arens, K. A., Parker, P. D., Guo, J. & Dicke, T. (2018). An Integrated Model of Academic Self-Concept Development. Academic Self-Concept, Grades, Test Scores, and Tracking Over Six Years. *Developmental Psychology*, *54*(2), 263–280. https://doi.org/10.1037/dev0000393

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Martin, M. O., Mullis, I. V. S. & Foy, P. (2017). TIMSS 2019 Assessment Design. In I. V. S. Mullis & M. O. Martin (Hrsg.), *TIMSS 2019 Assessment Frameworks* (S. 79–91). Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Martin, M. O., Mullis, I. V. S., Gonzalez, E., Gregory, K. D., Smith, T. A., Chrostowski, S. J., Garden, R. A. & O'Connor, K. M. (2000). *TIMSS 1999. International Science Report.* Chestnutt Hill, MA: TIMSS & PIRLS International Study Center.

Martschinke, S. & Kammermeyer, G. (2006). Selbstkonzept, Lernfreude und Leistungsangst und ihr Zusammenspiel im Anfangsunterricht. In A. Schründer-Lenzen (Hrsg.), *Risikofaktoren kindlicher Entwicklung. Migration, Leistungsangst und Schulübergang* (S. 125–139). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90075-9_6

Marzano, R. J. (2001). *A New Era of School Reform: Going Where The Research Takes Us*. Aurora: Midcontinent Research for Education and Learning (McREL).

Mayerl, J. & Urban, D. (2019). Vorsicht (!) bei Regressionsanalysen mit Interaktionsvariablen. Die Skalierung der unabhängigen Variablen kann die Höhe, Richtung und Signifikanz der Regressionskoeffizienten bestimmen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *71*, 135–156. https://doi.org/10.1007/s11577-019-00590-1

Mendes, M. & Pala, A. (2003). Type I Error Rate and Power of Three Normality Tests. *Pakistan Journal of Information and Technology*, *2*(2), 135–139. https://doi.org/10.3923/itj.2003.135.139

Merget, E., Engel, E.-M. & Aich, G. (2021). *Ich bin Ich. Die gezielte Erfassung und Stärkung des kindlichen Selbstkonzeptes in Kindertageseinrichtungen*. Weinheim: Beltz.

Merz, J. (1979). *Berufszufriedenheit von Lehrern: eine empirische Untersuchung*. Weinheim: Basel.

Meyer, H. (2003). Zehn Merkmale guten Unterrichts. Empirische Befunde und didaktische Ratschläge. *Zeitschrift für Pädagogik*, *55*(10), 36–43.

Meyer, J. P., Stanley, D. J., Herscovitch, L. & Topolnytsky, L. (2002). Affective, Continuance, and Normative Commitment to the Organization: A Meta-analysis of Antecedents, Correlates, and Consequences. *Journal of Vocational Behavior*, *61*(1), 20–52. https://doi.org/10.1006/jvbe.2001.1842

Midouser, D. & Betzer, N. (2007). The Contribution of Project-Based-Learning To High-Achievers' Acquisition of Technological Knowledge and Skills. *International Journal of Technology and Design Education*, *18*, 55–77.

Milles, F. & Jansen, M. (2021). Die Bedeutung von Unterrichtsmerkmalen für das mathematische Selbstkonzept und für die Moderation des Big-Fish-Little-Pond Effekts. In R. Lazarides & D. Raufelder (Hrsg.), *Motivation in unterrichtlichen fachbezogenen Lehr-Lernkontexten Perspektiven aus Pädagogik, Psychologie und Fachdidaktiken* (S. 299–330). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-31064-6_11

Minnameier, G., Hermkes, R. & Mach, H. (2015). Kognitive Aktivierung und Konstruktive Unterstützung als Prozessqualitäten des Lehrens und Lernens. *Zeitschrift für Pädagogik*, *61*(6), 837–856. https://doi.org/10.25656/01:15429

Möller, J. & Trautwein, U. (2020). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (3. Aufl., S. 187–210). Berlin: Springer. https://doi.org/10.1007/978-3-662-61403-7_8

Mönks, F. J. (1985). Hochbegabte: eine Situationsbeschreibung. In F. J. Mönks & P. Span (Hrsg.), *Hochbegabte in der Gesellschaft*. Nijmegen: Dekker & van de Vegt.

Moosbrugger, H. & Kelava, A. (Hrsg.). (2020). *Testtheorie und Fragebogenkonstruktion* (3 Aufl.). Berlin: Springer Nature. https://doi.org/10.1007/978-3-662-61532-4

Morbitzer, D. (2009). *Die Berufszufriedenheit von Lehrerinnen und Lehrern an den steirischen Berufsschulen*. Karl-Franzens-Universität Graz: univ. Diss.

Moser, U. & Angelone, D. (2011). *PISA 2009: Porträt des Kantons Züric*. Zürich: Institut für Bildungsevaluation.

Muthén, L. K. (2014). *Mplus Discussion. Goodness of fit tests*. https://www.statmodel.com/discussion/messages/11/3031.html?1528419519

Muthén, L. K. & Muthén, B. O. (2017). *Mplus User´s Guide. Version 8*. Los Angeles: Muthén & Muthén.

Nagy, G. (2007). *Berufliche Interessen, kognitive und fachgebundene Kompetenzen: Ihre Bedeutung für die Studienfachwahl und die Bewährung im Studium*. Freie Universität Berlin: univ. Diss.

Neuendorf, C. (2022). *Leistungsstarke Schülerinnen und Schüler in Deutschland: Definition, Entwicklung, soziale Integration*. univ. Diss.: Universität Potsdam.

Neuendorf, C., Kuhl, P. & Jansen, M. (2017). Leistungsstarke Schülerinnen und Schüler in Deutschland. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (S. 317–334). Münster: Waxmann.

Neuenschwander, M. P. (2006). Überprüfung einer Typologie der Klassenführung. *Schweizerische Zeitschrift für Bildungswissenschaften*, *28*(2), 243–258. https://doi.org/10.24452/sjer.28.2.4727

Neuhaus, B. (2007). Unterrichtsqualität als Forschungsfeld für empirische biologiedidaktische Studien. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung* (S. 243–254). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-540-68166-3_22

Ohle-Peters, A., Igler, J., Schlitter, T., Teerling, A., Köller, O. & McElvany, N. (2021). Unterrichtsqualität und intrinsische Lesemotivation im Kontext der Bund-Länder-Initiative „Bildung durch Sprache und Schrift" (BiSS). *Zeitschrift für Erziehungswissenschaft*, *24*, 861–882. https://doi.org/10.1007/s11618-021-01022-7

Oliver, J. S. & Simpson, R. D. (1988). Influences of Attitude Toward Science, Achievement Motivation, and Science Self Concept on Achievement in Science: A Longitudinal Study. *Science Education*, *72*(2), 143–155. https://doi.org/10.1002/sce.3730720204

Opdenakker, M.-C. & Van Damme, J. (2006). Teacher Characteristics and Teaching Styles as Effectiveness Enhancing Factors of Classroom Practice. *Teaching and Teacher Education*, *22*(1), 1–21. https://doi.org/10.1016/j.tate.2005.07.008

OECD (Organisation for Economic Co-operation and Development). (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing. https://doi.org/10.1787/9789264266490-en

Ozel, M., Caglak, S. & Erdogan, M. (2013). Are Affective Factors a Good Predictor of Science Achievement? Examining The Role of Affective Factors Based on PISA 2006. *Learning and Individual Differences*, *24*, 73–82. https://doi.org/10.1016/j.lindif.2012.09.006

Paccagnella, O. (2006). Centering or Not Centering in Multilevel Models? The Role of The Group Mean and The Assessment of Group Effects. *Evaluation Review*, *30*(1), 66–85. https://doi.org/10.1177/0193841x05275649

Pakarinen, E., Kiuru, N., Lerkkanen, M.-K., Poikkeus, A.-M., Siekkinen, M. & Nurmi, J.-E. (2010). Classroom Organization and Teacher Stress Predictlearning Motivation in Kindergarten Children. *European Journal of Psychology of Education*, *25*, 281–300. https://doi.org/10.1007/s10212-010-0025-6

Pedhazur, E. J. (1997). *Mulitple Regression in Behavioral Research. Explanation and Prediction* (3. Aufl.). New York: Wadsworth.

Pekrun, R. (2002). Vergleichende Evaluationsstudien zu Schülerleistungen: Konsequenzen für die Bildungsforschung. *Zeitschrift für Pädagogik*, *48*(1), 111–128.

Perera, H. N. & John, J. E. (2020). Teachers' Self-Efficacy Beliefs For Teaching Math: Relations With Teacher and Student Outcomes. *Contemporary Educational Psychology*, *61*, 101842. https://doi.org/10.1016/j.cedpsych.2020.101842

Perera, H. N., Maghsoudlou, A., Miller, C. J., McIlveen, P., Barber, D., Part, R. & Reyes, A. L. (2022). Relations of Science Teaching Self-Efficacy With Instructional Practices, Student Achievement and Support, and Teacher Job Satisfaction. *Contemporary Educational Psychology*, *69*, 1–15. https://doi.org/10.1016/j.cedpsych.2021.102041

Peter, T. (2018). Talent. Zu einem Schlüsselbegriff zeitgenössischer Bildung. In A. Böker & K. Horvath (Hrsg.), *Begabung und Gesellschaft. Sozialwissenschaftliche Perspektiven auf Begabung und Begabtenförderung* (S. 71–93). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-21761-7_4

Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher*, *38*(2), 109–119. https://doi.org/10.3102/0013189x09332374

Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft*, *13*, 121–148. https://doi.org/10.1007/s11618-010-0113-z

Pitsia, V. (2022). Examining High Achievement in Mathematics and Science Among Post-Primary Students in Ireland: A Multilevel Binary Logistic Regression Analysis of PISA Data. *Large-scale Assessments in Education*, *10*(14), 1–30. https://doi.org/10.1186/s40536-022-00131-x

Plomin, R. & von Stumm, S. (2018). The New Genetics of Intelligence. *Nature Reviews Genetics*, *19*, 148–159. https://doi.org/10.1038/nrg.2017.104

Pötschke, M. (2014). Mehrebenenanalyse. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 1101–1115). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-18939-0_87

Pötschke, M. (2020). Mehrebenenmodelle. In C. Wagemann, A. Goerres & M. B. Siewert (Hrsg.), *Handbuch Methoden der Politikwissenschaft* (S. 495–524). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-16936-7

Praetorius, A.-K. (2013). Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter. Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis. *Beiträge zur Lehrerinnen- Und Lehrerbildung*, *31*(2), 174–185. https://doi.org/10.36950/bzl.31.2.2013.9650

Praetorius, A.-K. & Charalambous, C. Y. (2018). Classroom Observation Frameworks For Studying Instructional Quality: Looking Back and Looking Forward. *ZDM - Mathematics Education*, *50*, 535–553. https://doi.org/10.1007/s11858-018-0946-0

Praetorius, A.-K., Kastens, C., Hartig, J. & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? Zusammenhänge zwischen optimistischen, realistischen und pessimistischen Selbstkonzepten und der Leistungsentwicklung von Grundschulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *48*(1), 14–26. https://doi.org/10.1026/0049-8637/a000140

Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic Dimensions of Teaching Quality: The German Framework of Three Basic Dimensions. *ZDM - Mathematics Education*, *50*, 407–426. https://doi.org/10.1007/s11858-018-0918-4

Preckel, F. & Brüll, M. (2010). The Benefit of Being a Big Fish in A Big Pond: Contrast and Assimilation Effects on Academic Self-Concept. *Learning and Individual Differences*, *20*, 522–531. https://doi.org/10.1177/1745691619895030

Preckel, F., Golle, J., Grabner, R., Jarvin, L., Kozbelt, A., Müllensiefen, D., Olszewski-Kublilius, P., Schneider, W., Subotnik, R., Vock, M. & Worrell, F. C. (2020). Talent Development in Achievement Domains: A Psychological Framework for Within- and Cross-Domain Research. *Perspectives on Psychological Science*, *15*(3), 691–722.

Preckel, F., Stumpf, E. & Schneider, W. (2018). Hochbegabung, Expertise und außergewöhnliche Leistung. In W. Schneider & U. Lindenberger (Hrsg.), *Entwicklungspsychologie* (8. Aufl., S. 681–695). Weinheim: Beltz.

Prediger, S. & von Aufschnaiter, C. (2017). Umgang mit heterogenen Lernvoraussetzungen aus fachdidaktischer Perspektive: Fachspezifische Anforderungs- und Lernstufungen berücksichtigen. In T. Bohl, J. Budde & M. Rieger-Ladich (Hrsg.), *Umgang mit Heterogenität in Schule und Unterrich* (S. 291–307). Bad Heilbrunn: Klinkhardt. https://doi.org/10.36198/9783838547558

Rabe-Hesketh, S. & Skrondal, A. (2012). *Categorical Responses, Counts, and Survival* (3. Aufl.). College Station: Stata Press.

Rakoczy, K., Klieme, E., Lipowsky, F. & Drollinger-Vetter, B. (2010). Strukturierung, kognitive Aktivität und Leistungsentwicklung im Mathematikunterricht. *Unterrichtswissenschaft*, *38*(3), 229–246.

Raudenbush, S. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2. Aufl.). Thousand Oaks: Sage.

Razali, N. M. & Yap, B. W. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Reibnegger, H. & Nausner, E. (2018). Klassenführungshandeln aus der Perspektive von Schüler/-innen und Lehrer/-innen. *Pädagogische Horizonte*, *2*(1), 63–82.

Reinecke, J. (2014). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: De Gruyter.

Retelsdorf, J., Köller, O. & Möller, J. (2014). Reading Achievement and Reading Self-Concept – Testing the Reciprocal Effects Model. *Learning and Instruction*, *29*, 21–30.

Reusser, K. & Pauli, C. (2021). Unterrichtsqualität ist immer generisch und fachspezifisch. Ein Kommentar aus kognitions- und lehr-lerntheoretischer Sicht. *Unterrichtswissenschaft*, *49*, 189–202.

Rheinberg, F. (2001). Bezugsnormen und schulische Leistungsbeurteilung. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 59–71). Weinheim: Beltz.

Roeder, P. M. (1997). Binnendifferenzierung im Urteil von Gesamtschullehrern. *Zeitschrift für Pädagogik*, *43*(2), 241–59. https://doi.org/10.25656/01:6981

Röhr-Sendlmeier, U. & Käser, U. (2017). Kompetenz. In L. Kühnhardt & T. Mayer (Hrsg.), *Bonner Enzyklopädie der Globalität* (S. 235–248). Wiesbaden: Springer.

Rosenberg, M. J. & Hovland, C. I. (1960). Cognitive, Affective, and Behavioral Components of Attitudes. In M. J. Rosenberg (Hrsg.), *Attitude organization and change* (S. 1–14). New Haven: Yale University Press.

Ross, J. A. (1994). *Beliefs That Make a Difference: The Origins and Impacts of Teacher Efficacy*. Calgary: Annual Meeting of the Canadian Association for Curriculum Studies.

Roßbach, H.-G. (2002). Unterrichtsqualität im 2. Schuljahr – Ergebnisse einer empirischen Untersuchung. *Unterrichtswissenschaft*, *30*(3), 230–245.

Rost, D. H. (2004). Über «Hochbegabung» und «hochbegabte» Jugendliche und Mythen, Fakten, Forschungsstandards. In J. Abel, R. Möller & C. Palentien (Hrsg.), *Jugend im Fokus empirischer Forschung* (S. 39–86). Münster: Waxmann.

Rost, D. H. & Buch, S. R. (2018). Hochbegabung. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5. Aufl., S. 226–241). Weinheim: Beltz.

Rost, D. H. & Sparfeldt, J. R. (2008). Intelligenz und Hochbegabung. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion* (S. 303–325). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91104-5_12

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.

Rowen, B., Chiang, F.-S. & Miller, R. J. (1997). Using Research on Employees› Performance to Study the Effects of Teachers on Students› Achievement. *Sociology of Education*, *70*(4), 256–284.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, *3*, 581–592.

Rubin, D. B. (1978). Multiple Imputations in Sample Surveys. A Phenomenological Bayesian Approach to Nonresponse. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 20–28.

Rupprecht, S. (2015). *Achtsamkeit macht Schule?! Fördert ein Achtsamkeitstraining das Lehrerwohlbefinden und die Unterrichtsqualität?*, Leuphana Universität Lüneburg: univ. Diss.

Rutter, M., Maughan, B. & Mortimore, P. (1980). *Fünfzehntausend Stunden. Schulen und ihre Wirkung auf die Kinder*. Weinheim: Beltz.

Sächsisches Bildungsinstitut (Hrsg.). (2017). *Binnendifferenzierung und lernzieldifferenter Unterricht*.

Sakiz, G., Pape, S. J. & Woolfolk Hoy, A. (2012). Does Perceived Teacher Affective Support Matter For Middle School Students in Mathematics Classrooms? *Journal of School Psychology*, *50*(2), 235–255.

Salner-Gridling. (2020). *Querfeldein: individuell lernen – differenziert lehren*: BMBF.

Sanfo, J.-B. M. B. & Malgoubri, I. (2021). Teaching Quality and Students› EFL Achievements in Ethiopia: Analysis From the Perspective of the Basic Dimensions of Teaching Quality. *Theory and Practice in Language Studies*, *11*(10), 1131–1145.

Scheerens, J., Witziers, B. & Steen, R. (2013). A Meta-analysis of School Effectiveness Studies. *Revista de educación*, *361*, 619–645.

Scherer, P. & Moser Opitz, E. (2010). *Fördern im Mathematikunterricht der Primarstufe*. Heidelberg: Spektrum Akademischer Verlag.

Scherer, R., Nilsen, T. & Jansen, M. (2016). Evaluating Individual Students' Perceptions of Instructional Quality: An Investigation of their Factor Structure, Measurement Invariance, and Relations to Educational Outcomes. *Frontiers in Psychology*, *7*(110), 1–16.

Schermelleh-Engel, K., Kerwer, M. & Klein, A. G. (2014). Evaluation of Model Fit in Nonlinear Multilevel Structural Equation Modeling. *Frontiers in Psychology*, *5*, 1–11.

Schermelleh-Engel, K. & Moosbrugger, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schicke, M. C. & Fagan, T. K. (1994). Contributions of Self-concept and Intelligence to the Prediction of Academic Achievement Among Grade 4, 6, and 8 Students. *Canadian Journal of School Psychology*, *10*(1), 62–69.

Schiefele, U. & Schaffner, E. (2015). Teacher Interests, Mastery Goals, and Self-Efficacy as Predictors of Instructional Practices and Student Motivation. *Contemporary Educational Psychology*, *42*, 159–171.

Schiefele, U., Streblow, L. & Retelsdorf, J. (2013). Dimensions of Teacher Interest and Their Relations to Occupational Well-Being and Instructional Practices. *Journal for educational research online*, *5*(1), 7–37.

Schiepe-Tiska, A., Heine, J.-H., Lüdtke, O., Seidel, T. & Prenzel, M. (2016). Mehrdimensionale Bildungsziele im Mathematikunterricht und ihr Zusammenhang mit den Basisdimensionen der Unterrichtsqualität. *Unterrichtswissenschaft*, *44*(3), 211–225.

Schmal, J. (2020). Unterrichtsqualität: Gut, besser, am besten? *Pflege Zeitschrift*, *73*, 39–41.

Schmalfuß, M. (2022). *Schulabsentismus vor dem Hintergrund regionaler Disparitäten*. Wiesbaden: Springer.

Schmidtner, S. S. (2017). *Leistungsstarke Jugendliche in Naturwissenschaften. Vertiefende Analysen zu PISA 2006 und PISA 2012*. TUM School of Education, univ. Diss.

Schneider, W. (2019). Programme zur Förderung kognitiver Fähigkeiten in Vorschule und Schule: Wie effektiv sind sie, und wie gut sind die Verfahren praktisch implementiert? *Zeitschrift für Pädagogische Psychologie*, *33*(1), 5–16.

Schöber, C., Retelsdorf, J. & Köller, O. (2015). Verbales schulisches Selbstkonzept und sprachliche Leistungen in Gruppen mit und ohne Migrationshintergrund. *Psychologie in Erziehungs und Unterricht*, *62*(2), 89–105.

Schrader, F.-W., Helmke, A. & (2008). Determinanten der Schulleistung. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion* (S. 285–302). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91104-5_11

Schröder, H. (2001). *Didkatisches Wörterbuch* (3. Aufl.). München: Oldenbourg Verlag.

Schüle, C., Schrieck, J., Besa, K.-S. & Arnold, K.-H. (2016). Der Zusammenhang der Theorie des geplanten Verhaltens mit der selbstberichteten Individualisierungspraxis von Lehrpersonen. *Empirische Sonderpädagogik*, *8*(2), 140–152.

Schulz-Heidorf, K. (2016). Individuelle Förderung im Unterricht: Eine Möglichkeit, soziale Herkunft und Schulerfolg zu entkoppeln? Eine Re-Analyse aus IGLU-E 2011. Universität Hamburg: univ. Diss.

Schuster, B. (2017). *Pädagogische Psychologie. Lernen, Motivation und Umgang mit Auffälligkeiten*. Wiesbaden: Springer.

Schwab, S. & Alnahdi, G. H. (2020). Do They Practise What They Preach? Factors Associated With Teachers' Use of Inclusive Teaching Practices Among In-Service Teachers. *Journal of Research in Special Educational Needs*, *20*(4), 321–330.

Schwantner, U. (2015). Innere Differenzierung im Mathematikunterricht. In B. Suchań, C. Wallner-Paschon & C. Schreiner (Hrsg.), *PIRLS & TIMSS 2011. Die Kompetenzen in Lesen, Mathematik und Naturwissenschaft am Ende der Volksschule. Österreichischer Expertenbericht* (S. 127–138). Graz: Leykam.

Schwarz, G. (1978). Estimating The Dimension of A Model. *The Annals of Statistics*, *6*(2), 461–464.

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.

Schwippert, K. (2019). Was wird aus den Büchern? Sozialer Hintergrund von Lernenden und Bildungsungleichheit aus Sicht der international vergleichenden Erziehungswissenschaft. *JERO*, *11*(1), 92–117.

Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. & Wendt, H. (Hrsg.). (2020a). *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.

Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. & Wendt, H. (2020b). TIMSS 2019: Wichtige Ergebnisse im Überblick. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 13–24). Münster: Waxmann.

Schwippert, K., Scholz, L. A., Beese, C., Kasper, D., Schulz-Heidorf, K. & Girelli, A.-L. (2020). Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2019). In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 25–55). Münster: Waxmann.

Seaton, M., Parker, P., Marsh, H. W., Craven, R. G. & Yeung, A. S. (2013). The reciprocal relations between self-concept, motivation and achievement: juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology*, 1–24. https://doi.org/10.1080/01443410.2013.825232

Sedlmeier, P. & Renkewitz, F. (2018). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler* (3. Aufl.). Hallbergmoos: Pearson.

Seidel, T., Renkl, A. & Rieß, W. (2021). Basisdimensionen für Unterrichtsqualität im Fachkontext konkretisieren: Die Rolle von Unterrichtsartefakten und Bestimmung von Standardsituationen. *Unterrichtswissenschaft*, *49*, 293–301. https://doi.org/10.1007/s42010-021-00108-9

Seidel, T. & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research*, *77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Seiffge-Krenke, I. (1974). *Probleme und Ergebnisse der Kreativitätsforschung*. Bern: Huber.

Shavelson, R. J., Hubner, J., J. & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Shen, J., Leslie, J. M., Spybrook, J. K. & Ma, X. (2012). Are Principal Background and School Processes Related to Teacher Job Satisfaction? A Multilevel Study Using Schools and Staffing Survey 2003-04. *American Educational Research Journal*, *49*(2), 200–230. https://doi.org/10.3102/0002831211419949

Shulman, L. S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, *57*, 1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperrreguy, M. I. & Chen, M. (2012). Early Predictors of High School Mathematics Achievement. *Psychological Science*, *23*(7), 691––697. https://doi.org/10.1177/0956797612440101

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L. & Schneider, M. (2021). Domain-Specific Prior Knowledge and Learning: A Meta-Analysis. *Educational Psychologist*, *57*, 31–54. https://doi.org/10.1080/00461520.2021.1939700

Skaalvik, E. M. (1997). Self-Enhancing and Self-Defeating Ego Orientation: Relations With Task and Avoidance Orientation, Achievement, Self-Perceptions, and Anxiety. *Journal of Educational Psychology*, *89*(1), 71––81. https://doi.org/10.1037/0022-0663.89.1.71

Skaalvik, E. M. & Skaalvik, S. (2007). Dimensions of Teacher Self-Efficacy and Relations With Strain Factors, Perceived Collective Teacher Efficacy, and Teacher Burnout. *Journal of Educational Psychology*, *99*(3), 611–625. https://doi.org/10.1037/0022-0663.99.3.611

Skaalvik, E. M. & Skaalvik, S. (2014). Teacher Self-Efficiacy and Perceived Autonomy: Relations With Teacher Engagement, Job Satisfaction, and emotional Exhaustion. *Psychological Reports: Employment Psychology & Marketing*, *114*(1), 68–77. https://doi.org/10.2466/14.02.PR0.114k14w0

Skaalvik, E. M. & Skaalvik, S. (2017). Motivated For Teaching? Associations With School Goal Structure, Teacher Self-Efficacy, Job Satisfaction and Emotional Exhaustion. *Teaching and Teacher Education*, *67*, 152–160. https://doi.org/10.1016/j.tate.2017.06.006

Skaalvik, E. M. & Valås, H. (1999). Relations Among Achievement, Self-concept, and Motivation in Mathematics and Language Arts: A Longitudinal Study. *The Journal of Experimental Education*, *67*, 135–149. https://doi.org/10.1080/00220979909598349

Slavin, R. E. (1987). Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis. *Review of Educational Research*, *57*(3), 293–336.

Slavin, R. E. (2000). *Educational Psychology: Theory and Practice* (6. Aufl.). Boston: Allyn and Bacon.

Sliwka, A., Klopsch, B. & Dumont, H. (2019). *Konstruktive Unterstützung im Unterricht*. Stuttgart: Institut für Bildungsanalysen Baden-Württemberg.

Smylie, M. A. (1988). The Enhancement Function of Staff Development: Organizational and Psychological Antecedents to Individual Teacher Change *American Educational Research Journal*, *25*(1), 1–30. https://doi.org/10.3102/00028312025001001

Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. Los Angeles: Sage.

Snow, R. E. (1991). Aptitude-Treatment Interaction as a Framework for Research on Individual Differences in Psychotherapy. *Journal of Consulting and Clinical Psychology*, *59*(2), 205–216.

Sontag, C. & Stoeger, H. (2015). Can Highly Intelligent and High-Achieving Students Benefit From Training in Self-Regulated Learning in A Regular Classroom Context? *Learning and Individual Differences*, *41*, 43–53.

Spinath, B. & Spinath, F. M. (2005). Development of Self-Perceived ability in Elementary School: The Role of Parents' Perceptions, Teacher Evaluations, and Intelligence. *Cognitive Development*, *20*(2), 190–204. https://doi.org/10.1016/j.cogdev.2005.01.001

Stahns, R., Rieser, S. & Hußmann, A. (2020). Können Viertklässlerinnen und Viertklässer Unterrichtsqualität valide einschätzen? Ergebnisse zum Fach Deutsch. *Unterrichtswissenschaft*, *48*, 663–682. https://doi.org/10.1007/s42010-020-00084-6

Stang, J., Lepper, C., Steffensky, M. & McElvany, N. (2020). Einblicke in die Gestaltung des Mathematik- und naturwissenschaftsbezogenen Sachunterrichts an Grundschulen in Deutschland. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 187–208). Münster: Waxmann. https://doi.org/10.31244/9783830993193

Steenbergen-Hu, S., Makel, M. C. & Olszewski-Kublilius, P. (2016). What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement: Findings of Two Second-Order Metaanalyses. *Review of Educational Research*, *86*(4), 849–899. https://doi.org/10.3102/0034654316675417

Steffensky, M. & Neuhaus, B. J. (2018). Unterrichtsqualität im naturwissenschaftlichen Unterricht. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 299–313). Springer: Wiesbaden. https://doi.org/10.1007/978-3-662-56320-5_18

Steffensky, M., Scholz, L. A., Kasper, D. & Köller, O. (2020). Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 115–168). Münster: Waxmann. https://doi.org/10.31244/9783830993193

Steinmayr, R. & Spinath, B. (2009). The Importance of Motivation as a Predictor of School Achievement. *Learning and Individual Differences*, *19*, 80–90. https://doi.org/10.1016/j.lindif.2008.05.004

Stern, E. & Grabner, R. H. (2014). Die Erforschung menschlicher Intelligenz. In L. Ahnert (Hrsg.), *Theorien in der Entwicklungspsychologie* (S. 174–201). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-642-34805-1_7

Stern, E. & Guthke, J. (2001). Vorwort. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 9–12). Lengerich: Pabst. https://doi.org/10.17104/9783406682971

Stern, E. & Neubauer, A. (2016). Intelligenz: kein Mythos, sondern Realität. *Psychologische Rundschau*, *67*(1), 15–27. https://doi.org/10.1026/0033-3042/a000290

Stone, B. M. (2021). The Ethical Use of Fit Indices in Structural Equation Modeling: Recommendations for Psychologists. *Frontiers in Psychology*, *12*, 1–4. https://doi.org/10.3389/fpsyg.2021.783226

Stubbe, T. C., Krieg, M., Beese, C. & Jusufi, D. (2020). Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 263–289). Münster: Waxmann. https://doi.org/10.31244/9783830993193

Subotnik, R., Olszewski-Kublilius, P. & Worrell, F. C. (2011). Rethinking Giftedness and Gifted Education: A Proposed Direction Forward Based on Psychological Science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056

Suprayogi, M. N., Valcke, M. & Godwin, R. (2017). Teachers and Their Implementation of Differentiated Instruction In The Classroom. *Teaching and Teacher Education*, *67*, 291–301. https://doi.org/10.1016/j.tate.2017.06.020

Swisher, L. L., Beckstead, J. W. & Bebeau, M. J. (2004). Factor Analysis as a Tool for Survey Analysis Using a Professional Role Orientation Inventory as an Example. *Physical Therapy*, *84*(9), 784-799. https://doi.org/10.1093/ptj/84.9.784

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tachtsoglou, S. & König, J. (2017). *Statistik für Erziehungswissenschaftlerinnen und Erziehungswissenschaftler. Konzepte, Beispiele und Anwendungen in SPSS und R*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-13437-2

Tausendpfund, M. (2020). Mehrebenenanalyse. In M. Tausendpfund (Hrsg.), *Fortgeschrittene Analyseverfahren in den Sozialwissenschaften* (S. 102–140). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-30237-5

Teig, N. & Nilsen, T. (2022). Profiles of Instructional Quality in Primary and Secondary Education: Patterns, Predictors, and Relations to Student Achievement and Motivation in Science. *Studies in Educational Evaluation*, *74*, 1–17. https://doi.org/10.1016/j.stueduc.2022.101170

Tella, A. (2008). Teacher Variables As Predictors of Academic Achievement of Primary School Pupils Mathematics. *International Electronic Journal of Elementary Education*, *1*(1), 16–33. https://doi.org/10.4314/ifep.v16i2.23806

Thommen, D., Grob, U., Lauermann, F., Klassen, R. M. & Praetorius, A.-K. (2022). Different Levels of Context-Specificity of Teacher Self-Efficacy and Their Relations With Teaching Quality. *Frontiers in Psychology*, *13*, 857526. https://doi.org/10.3389/fpsyg.2022.857526

Thoonen, E. E. J., Sleegers, P. J. C., Peetsma, T. T. D. & Oort, F. J. (2011). Can Teachers Motivate Students to Learn? *Educational Studies*, *37*(3), 345–360. https://doi.org/10.1080/03055698.2010.507008

Trautmann, M. & Wischer, B. (2011). *Heterogenität in der Schule. Eine kritische Einführung*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-92893-7

Trentham, L., Silvern, S. & Brogdon, R. (1985). Teacher Efficacy and Teacher Competency Ratings. *Psychology in Schools*, *22*(3), 343–352.

Trültzsch-Wijnen, C. W. (2020). *Medienhandeln zwischen Kompetenz, Performanz und Literacy*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-29534-9

Tschannen-Moran, M. & McMaster, P. (2009). Sources of Self-Efficacy: Four Professional Development Formats and Their Relationship to Self-Efficacy and Implementation of a New Teaching Strategy. *The Elementary School Journal*, *110*(2), 228–245. https://doi.org/10.1086/605771

Tschannen-Moran, M., Woolfolk Hoy, A. & Hoy, W. K. (1998). Teacher Efficacy: Its Meaning and Measure. *Review of Educational Research*, *68*(2), 202–248. https://doi.org/10.3102/00346543068002202

Tuner, M. & Yilmaz, Ö. (2020). Relations Attitude Towards Mathematics Lessons: Anxiety and Academic Success. *Journal of Research in Mthematics Education*, *9*(2), 173–195. https://doi.org/10.17583/redimat.2020.4061

Tuorón, J., López-González, E., Hernandéz, L. L., García San Pedro, M. J. & Navarro Ascencio, E. (2018). Spanish High and Low Achievers in Science in PISA 2015: Impact Analysis of Some Contextual Variables. *Revista de Educación*, *380*, 156–184. https://doi.org/10.4438/1988-592X-RE-2017-380-376

Twele, N. (2022). *Effects of Educational Investments Within and Beyond a Framework of Home-Based Parental Involvement. Qualitative and Quantitative Analyses on South African Subpopulations*. Universität Hamburg: univ. Diss.

Upmeier zu Belzen, A. (2007). Einstellungen im Kontext Biologieunterricht. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 21–31). Wiesbaden: Springer. https://doi.org/10.1007/978-3-540-68166-3_3

Urban, D. (2022). *Mehrebenenanalyse: Die wichtigsten 23 Modelle zur Analyse von dichotomen und metrischen Zwei- und Drei-Ebenen-Effekten (unter Verwendung von Mplus)* (2. Aufl.). SISS – Schriftenreihe des Instituts f. Sozialwissenschaften der Universität Stuttgart. https://doi.org/10.18419/opus-12146

Vachon, J. (1986). Attitudes: Their Relationship With Intelligence and 385 achievement and their Evaluation. In S. E. Newstead, S. H. Irvine & P. L. Dann (Hrsg.), *Human Assessment: Cognition and Motivation* (S. 385). Dordrecht: Nijhoff. https://doi.org/10.1007/978-94-009-4406-0_54

Valtin, R. (2012). Noten oder verbale Beurteilungen: Was ist ein gutes Zeugnis? In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Leistungsbeurteilung* (S. 89–106). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-18846-1_5

Valtin, R., Wagner, C. & Schwippert, K. (2005). Schülerinnen und Schüler am Ende der vierten Klasse – schulische Leistungen, lernbezogene Einstellungen und außerschulische Lernbedingungen. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 187–238). Münster: Waxmann.

Vieluf, S., Kunter, M. & van de Vijver, F. J. R. (2013). Teacher Self-Efficacy in Cross-National Perspective. *Teaching and Teacher Education*, *35*, 92–103. https://doi.org/10.1016/j.tate.2013.05.006

Virtanen, T. E., Sørensen Vaaland, G. & Ertesvåg, S. K. (2019). Associations Between Observed Patterns of Classroom Interactions and Teacher Wellbeing in Lower secondary School. *Teaching and Teacher Education*, *77*, 240–252. https://doi.org/10.1016/j.tate.2018.10.013

Vock, M. & Gronostaj, A. (2017). *Umgang mit Heterogenität in Schule und Unterricht*. Friedrich Ebert Stiftung.

Vock, M., Gronostaj, A. & Preckel, F. (2017). Begabungsdiagnostik und -forschung mit dem Berliner Intelligenzstrukturtest für Jugendliche (BIS-HB). In U. Trautwein & M. Hasselhorn (Hrsg.), *Begabung und Talente* (S. 49–64). Göttingen: Hogrefe.

Voland, E. (2006). Lernen – Die Grundlegung der Pädagogik in evolutionärer Charakterisierung. *Zeitschrift für Erziehungswissenschaft*, *Beiheft 5*, 103–116. https://doi.org/10.1007/978-3-531-90607-2_8

Voss, T., Kunter, M., Seiz, J., Hoehne, V. & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität. *Zeitschrift für Pädagogik*, *60*(2), 184–201. https://doi.org/10.25656/01:14653

Wagner, W., Göllner, R., Helmke, A., Trautwein, U. & Lüdtke, O. (2013). Construct Validity of Student Perceptions of Instructional Quality is High, But Not Perfect: Dimensionality and Generalizability of Domain-Independent Assessments. *Learning and Instruction*, *28*, 1–11. https://doi.org/10.1016/j.learninstruc.2013.03.003

Wang, L. (2022). Exploring the Relationship Among Teacher Emotional Intelligence, Work Engagement, Teacher Self-Efficacy, and Student Academic Achievement: A Moderated Mediation Model. *Frontiers in Psychology*, *12*, 1–8. https://doi.org/10.3389/fpsyg.2021.810559

Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a Knowedge Base for School Learning. *Review of Educational Research*, *63*(3), 249–294. https://doi.org/10.3102/00346543063003249

Warwas, J., Hertel, S. & Labuhn, A. S. (2011). Bedingungsfaktoren des Einsatzes von adaptiven Unterrichtsformen im Grundschulunterricht. *Zeitschrift für Pädagogik*, *57*(6), 854–867. https://doi.org/10.25656/01:8785

Wechsler, D. (1944). *The Measurement of Adult Intelligence* (3. Aufl.). Baltimore: Williams & Wilkins.

Weidinger, A. F., Steinmayr, R. & Spinath, B. (2017). Changes in the Relation Between Competence Beliefs and Achievement in Math Across Elementary School Years. *Child Development*, *89*(2), 138–156. https://doi.org/10.1111/cdev.12806

Weigand, G. (2020). »Leistung macht Schule« – Eine Einführung. In G. Weigand, C. Fischer, F. Käpnick, C. Perleth, F. Preckel, M. Vock & H.-W. Wollersheim (Hrsg.), *Leistung macht Schule. Förderung leistungsstarker und potenziell besonders leistungsfähiger Schülerinnen und Schüler* (S. 13–22). Weinheim: Beltz.

Weinburgh, M. (1995). Gender Differences in Student Attitudes Toward Science: A Meta-Analysis of The Literature From 1970 to 1991. *Journal of Research in Science Education*, *32*(4), 387–398. https://doi.org/10.1002/tea.3660320407

Weinert, A. B. (2015). *Organisations- und Personalpsychologie*. Weinheim: Beltz.

Weinert, F. E. (1999). *Konzepte der Kompetenz*. Paris: OECD.

Weinert, F. E. (2001). Leistungen der Schule oder der Schüler. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 73–86). Weinheim: Beltz.

Weinert, F. E. & Helmke, A. (1996). Der gute Lehrer: Person, Funktion oder Fiktion? In A. Leschinsky (Hrsg.), *Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule* (S. 223–233). Weinheim: Beltz. https://doi.org/10.25656/01:9799

Weinert, F. E., Helmke, A. & Schneider, W. (1989). Individual Differences in Learning Performance and In School Achievement: Plausible Parallels and Unexplained Discrepancies. In H. Mandl, E. De Corte, N. Bennet & H. Friedrich (Hrsg.), *Learning and Instruction* (S. 461–479). Oxford: Pergamon Press.

Weinert, F. E., Schrader, F.-W. & Helmke, A. (1989). Quality of Instruction and Achievement Outcomes. *International Journal of Educational Research*, *13*(8), 895–914. https://doi.org/10.1016/0883-0355(89)90072-4

Weißeno, G. & Landwehr, B. (2015). Effektiver Unterricht über die Europäische Union – Ergebnisse einer Studie zur Schülerperzeption von Politikunterricht. In M. Oberle (Hrsg.), *Die Europäische Union erfolgreich vermitteln* (S. 99–109). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-06817-2_7

Wendt, H., Bos, W., Kasper, D., Walzeburg, A., Goy, M. & Jusufi, D. (2016). Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2015). In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 31–77). Münster: Waxmann. https://doi.org/10.25656/01:14022

Wendt, H., Schwippert, K., Stubbe, T. C. & Jusufi, D. (2020). Kompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 291–313). Münster: Waxmann. https://doi.org/10.31244/9783830993193

Wendt, H., Willems, A. S., Tarelli, I., Euen, B. & Bos, W. (2013). Ausreichend geförderte Talente? Zu den deutschen Ergebnissen von leistungsstarken Viertklässlerinnen und Viertklässlern in IGLU 2011 und TIMSS 2011. In C. Fischer (Hrsg.), *Schule und Unterricht adaptiv gestalten. Fördermöglichkeiten für benachteiligte Kinder und Jugendliche*. Münster: Waxmann. https://doi.org/10.5159/IQB_IGLU_2011_v1

Wentzel, K. R. (1998). Social Relationships and Motivation in Middle School: The Role of Parents, teachers, and Peers. *Journal of Educational Psychology*, *90*(2), 202–209. https://doi.org/10.1037/0022-0663.90.2.202

Willems, A. S. (2016). Unterrichtsqualität und professionelles Lehrerhandeln. Prozesse und Wirkungen guten Unterrichts aus dem Blickwinkel der empirischen Schul- und Unterrichtsforschung. In R. Porsch (Hrsg.), *Einführung in die Allgemeine Didaktik* (S. 289–337). Münster: Waxmann.

Willson, V. L. (1983). A Meta-Analysis of The Relationship Between Science Achievement and Science Attitude: Kindergarten Through College. *Journal of Research in Science Education*, *20*(9), 839–850. https://doi.org/10.1002/tea.3660200906

Wolf, C. & Best, H. (2010). *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-92038-2

Wolke, A. (2008). Wenn Schüler nicht mehr in die Schule gehen. *AJS Forum*, *32*(2-3), 4–6. https://doi.org/10.1055/s-2008-1077187

Woolfolk Hoy, A. & Burke Spero, R. (2005). Changes in Teacher Efficacy During the Early Years of Teaching: A Comparison of four Measures. *Teaching and Teacher Education*, *21*(4), 343–356. https://doi.org/10.1016/j.tate.2005.01.007

Wüsten, S. (2010). *Allgemeine und fachspezifische Merkmale der Unterrichtsqualität im Fach Biologie. Eine Video- und Interventionsstudie*. Berlin: Logos.

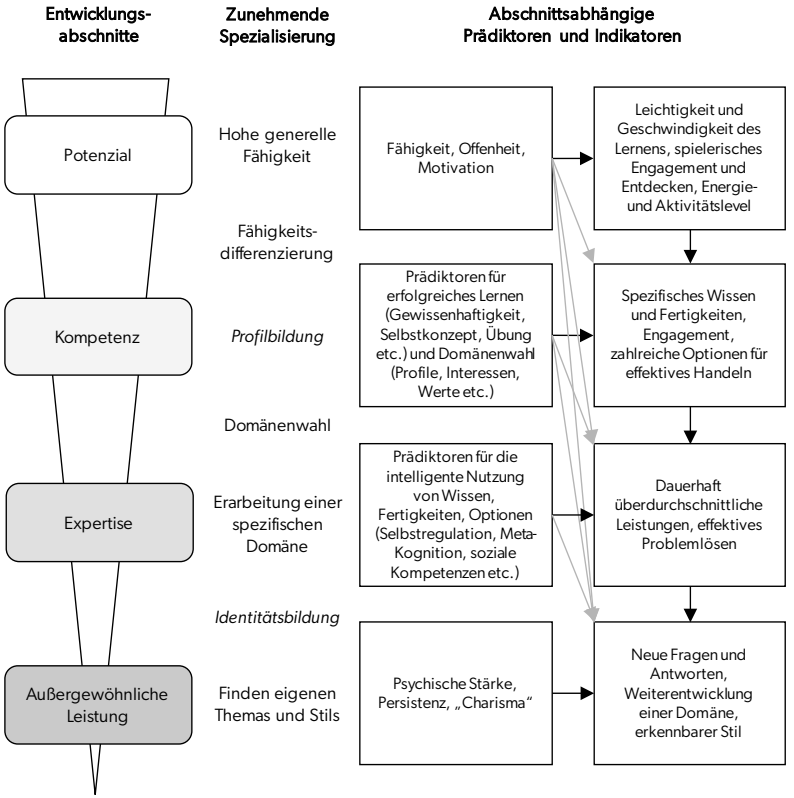
Zee, M. & Koomen, H. M. Y. (2016). Teacher Self-Efficacy and Its Effects on Classroom Processes, Student Academic Adjustment, and Teacher Well-Being: A Synthesis of 40 Years of Research. *Review of Educational Research*, *86*(4), 981–1015. https://doi.org/10.3102/0034654315626801

Zeidner, M. & Schleyer, E. J. (1999). The Effects of Educational Context on Individual Difference Variables, Self-Perceptions of Giftedness, and School Attitudes in Gifted Adolescents. *Journal of Youth and Adolescence*, *28*(6), 687–703. https://doi.org/10.1023/A:1021687500828

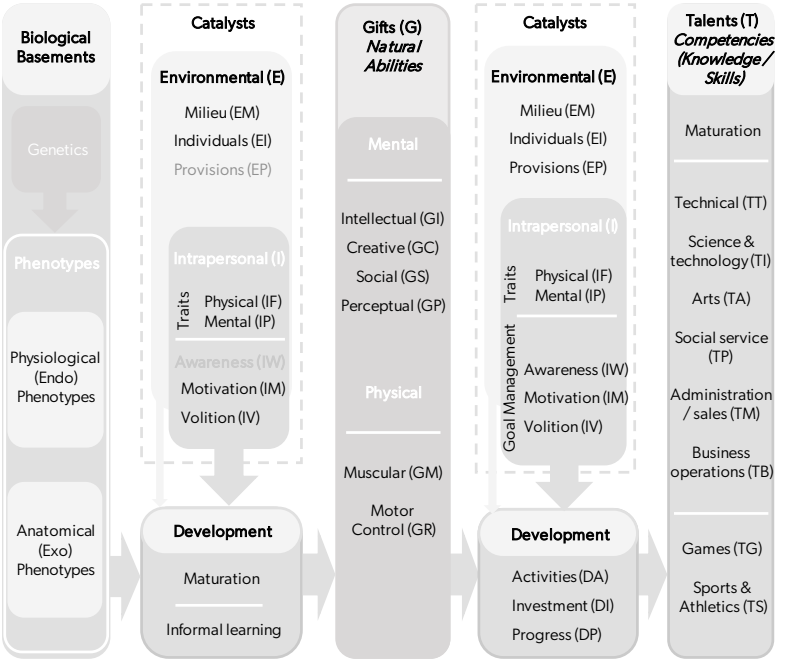
Anhang

Talentmodelle

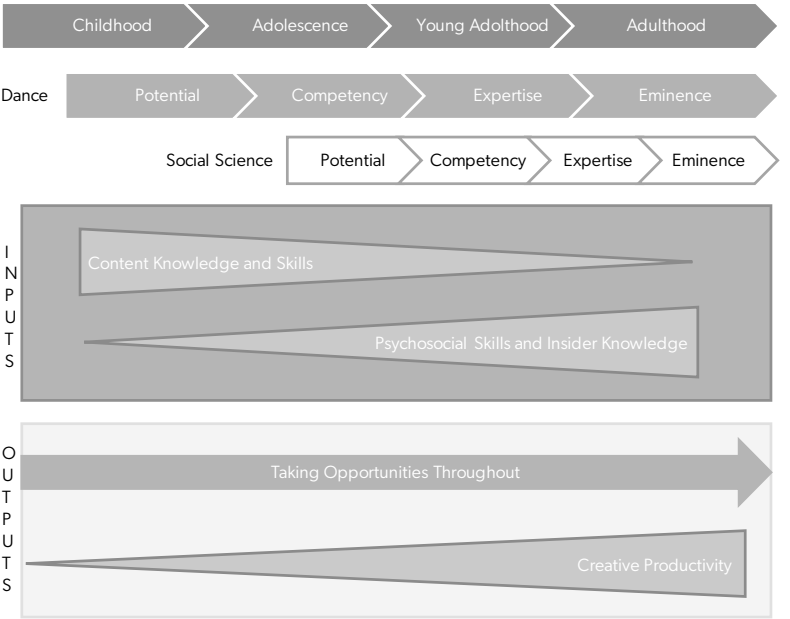
Talent-Development-in-Achievement-Domains (TAD) Framework nach Preckel et. al. (2020)



Extended Model of Giftedness and Talent (EMGT) nach Gagné (2013)



Talent Development Megamodell nach Subotnik, Olszewski-Kubilius & Worrell (2011)



Deskriptive Statistiken

Die deskriptiven Statistiken zu den in TIMSS 2019 eingesetzten Skalen wurden im Rahmen des Skalenhandbuches zu TIMSS 2019 berechnet und veröffentlicht. Da somit bereits eine umfassende Dokumentation der Deskriptiva vorliegt, wurde in dieser Arbeit auf eine erneute Berechnung verzichtet. Die Angaben in diesem Anhang entstammen deshalb der Skalendokumentation von TIMSS 2019 (Beese et al., 2020)[24].

Skala: Fachbezogenes Selbstkonzept Sachunterricht I und intrinsische Werte Sachunterricht

Frage: Wie sehr stimmst du diesen Aussagen über das Lernen im naturwissenschaftlichen Sachunterricht zu?

Kodierung: 1 = stimme völlig zu; 2 = stimme eher zu; 3 = stimme eher nicht zu; 4 = stimme überhaupt nicht zu

Item: Ich lerne gerne im Sachunterricht. (ASBS07A)

Kodierung	1	2	3	4	Gesamt
Absolut	1284	1025	376	194	2879
Prozent	45.10	35.69	12.79	6.42	100.0
(SE)	(1.30)	(0.84)	(0.63)	(0.59)	

Item: Ich wünschte, ich hätte keinen Sachunterricht. (ASBS07B)

Kodierung	1	2	3	4	Gesamt
Absolut	274	326	668	1600	2868
Prozent	9.30	11.19	22.88	56.63	100.0
(SE)	(0.66)	(0.77)	(1.09)	(1.36)	

Item: Sachunterricht ist langweilig. (ASBS07C)

Kodierung	1	2	3	4	Gesamt
Absolut	263	342	722	1492	2819
Prozent	8.81	11.94	25.83	53.42	100.0
(SE)	(0.62)	(0.69)	(0.83)	(1.35)	

24 Beese, C., Scholz, L. A., Jentsch, A., Jusufi, D. & Schwippert, K. (2023). *Skalenhandbuch zur Dokumentation der Erhebungsinstrumente und Arbeit mit den Datensätzen*. Münster: Waxmann.

Item: Ich lerne viele interessante Dinge im Sachunterricht. (ASBS07D)

Kodierung	1	2	3	4	Gesamt
Absolut	1770	727	234	139	2870
Prozent	62.32	25.06	8.08	4.54	100.0
(SE)	(1.36)	(0.92)	(0.64)	(0.50)	

Item: Ich mag Sachunterricht. (ASBS07E)

Kodierung	1	2	3	4	Gesamt
Absolut	1431	873	334	231	2869
Prozent	50.69	30.14	11.32	7.84	100.0
(SE)	(1.43)	(0.83)	(0.74)	(0.64)	

Item: Ich freue mich auf den Sachunterricht in der Schule. (ASBS07F)

Kodierung	1	2	3	4	Gesamt
Absolut	1206	1030	410	222	2868
Prozent	42.87	35.67	14.04	7.43	100.0
(SE)	(1.38)	(0.93)	(0.80)	(0.67)	

Item: Im Sachunterricht lerne ich, wie die Dinge in der Welt funktionieren. (ASBS07G)

Kodierung	1	2	3	4	Gesamt
Absolut	1774	856	158	80	2868
Prozent	62.23	29.57	5.56	2.64	100.0
(SE)	(1.21)	(0.97)	(0.43)	(0.37)	

Item: Ich mache gerne Versuche im Sachunterricht. (ASBS07H)

Kodierung	1	2	3	4	Gesamt
Absolut	1848	596	268	131	2843
Prozent	65.55	20.62	9.39	4.45	100.0
(SE)	(1.26)	(0.80)	(0.66)	(0.55)	

Item: Sachunterricht gehört zu meinen Lieblingsfächern. (ASBS07I)

Kodierung	1	2	3	4	Gesamt
Absolut	1146	737	591	386	2860
Prozent	40.95	25.77	20.44	12.83	100.0
(SE)	(1.34)	(0.87)	(1.00)	(0.89)	

Skala: Fachbezogenes Selbstkonzept Sachunterricht II

Frage: Wie sehr stimmst du diesen Aussagen über den naturwissenschaftlichen Sachunterricht zu?

Kodierung: 1 = stimme völlig zu; 2 = stimme eher zu; 3 = stimme eher nicht zu; 4 = stimme überhaupt nicht zu

Item: Normalerweise bin ich gut im Sachunterricht. (ASBS09A)

Kodierung	1	2	3	4	Gesamt
Absolut	1321	1107	331	86	2845
Prozent	46.65	38.98	11.37	3.00	100.0
(SE)	(1.08)	(0.91)	(0.50)	(0.33)	

Item: Sachunterricht fällt mir schwerer als vielen meiner Mitschüler. (ASBS09B)

Kodierung	1	2	3	4	Gesamt
Absolut	205	452	821	1351	2829
Prozent	7.08	15.90	29.20	47.82	100.0
(SE)	(0.57)	(0.77)	(0.91)	(1.18)	

Item: Ich bin einfach nicht gut im Sachunterricht. (ASBS09C)

Kodierung	1	2	3	4	Gesamt
Absolut	237	359	751	1438	2785
Prozent	8.46	12.78	26.76	52.00	100.0
(SE)	(0.59)	(0.71)	(0.93)	(1.27)	

Item: Ich lerne schnell im Sachunterricht. (ASBS09D)

Kodierung	1	2	3	4	Gesamt
Absolut	1142	1107	393	134	2776
Prozent	41.00	40.33	13.96	4.71	100.0
(SE)	(1.07)	(0.93)	(0.70)	(0.50)	

Item: Meine Lehrerin/mein Lehrer sagt, dass ich gut im Sachunterricht bin. (ASBS09E)

Kodierung	1	2	3	4	Gesamt
Absolut	706	1229	572	218	2725
Prozent	26.05	45.45	20.64	7.85	100.0
(SE)	(1.01)	(1.09)	(0.70)	(0.53)	

Item: Ich finde Sachunterricht schwerer als alle anderen Fächer. (ASBS09F)

Kodierung	1	2	3	4	Gesamt
Absolut	227	353	644	1602	2826
Prozent	7.79	12.45	22.89	56.88	100.0
(SE)	(0.51)	(0.78)	(0.89)	(1.24)	

Item: Sachunterricht bringt mich durcheinander. (ASBS09G)

Kodierung	1	2	3	4	Gesamt
Absolut	226	296	630	1687	2839
Prozent	7.66	10.43	21.81	60.10	100.0
(SE)	(0.62)	(0.69)	(0.80)	(1.12)	

Skala: Wahrnehmung individueller Förderung im Sachunterricht

Frage: Wie oft passieren diese Dinge in eurem naturwissenschaftlichen Sachunterricht?

Kodierung: 1 = immer oder fast immer; 2 = häufig; 3 = manchmal; 4 = selten; 5 = nie

Item: Im Unterricht gibt meine Lehrerin/mein Lehrer Schülern Zusatzaufgaben, wenn sie etwas noch nicht verstanden haben. (ASXS22A)

Kodierung	1	2	3	4	5	Gesamt
Absolut	350	252	450	771	958	2781
Prozent	12.39	9.04	16.55	27.61	34.40	100.0
(SE)	(0.93)	(0.53)	(0.78)	(0.81)	(1.19)	

Item: Meine Lehrerin/mein Lehrer gibt den guten Schülern zusätzlich schwierige Aufgaben, bei denen sie sich wirklich anstrengen müssen. (ASXS22B)

Kodierung	1	2	3	4	5	Gesamt
Absolut	440	419	519	636	758	2772
Prozent	15.57	15.09	18.56	23.28	27.50	100.0
(SE)	(1.06)	(0.88)	(0.69)	(0.90)	(1.18)	

Item: Bei der Stillarbeit gibt meine Lehrerin/mein Lehrer den Schülern unterschiedlich schwierige Aufgabenarten, die zu jedem Schüler passen. (ASXS22C)

Kodierung	1	2	3	4	5	Gesamt
Absolut	335	291	420	616	1090	2752
Prozent	12.41	10.48	15.21	22.32	39.57	100.0
(SE)	(0.98)	(0.59)	(0.72)	(0.89)	(1.41)	

Item: Bei Gruppenarbeiten verteilt meine Lehrerin/mein Lehrer unterschiedlich schwierige Aufgaben an die verschiedenen Gruppen. (ASXS22D)

Kodierung	1	2	3	4	5	Gesamt
Absolut	353	354	536	651	850	2744
Prozent	13.12	12.89	19.56	23.82	30.61	100.0
(SE)	(0.96)	(0.71)	(0.84)	(0.82)	(1.37)	

Item: Wenn wir alleine oder in Gruppen arbeiten, bekommen wir unterschiedlich schwierige Aufgaben. (ASXS22E)

Kodierung	1	2	3	4	5	Gesamt
Absolut	305	300	466	663	949	2683
Prozent	11.41	11.27	17.76	24.57	34.99	100.0
(SE)	(0.81)	(0.65)	(0.91)	(0.96)	(1.39)	

Item: Im Unterricht erhalten die Schüler unterschiedliche Aufgaben, die gut zu ihren Interessen passen. (ASXS22F)

Kodierung	1	2	3	4	5	Gesamt
Absolut	352	308	446	570	1019	2695
Prozent	12.96	11.46	16.20	21.33	38.04	100.0
(SE)	(0.82)	(0.71)	(0.71)	(0.94)	(1.40)	

Item: Im Unterricht erhalten die Schüler unterschiedliche Aufgaben, je nachdem, wofür sich die Schüler interessieren. (ASXS22G)

Kodierung	1	2	3	4	5	Gesamt
Absolut	305	259	396	591	1173	2724
Prozent	11.12	9.51	14.37	21.91	43.09	100.0
(SE)	(0.72)	(0.70)	(0.77)	(0.96)	(1.51)	

Skala: Wahrgenommene Klassenführung (Sachunterricht)

Frage: Wie oft passieren diese Dinge in eurem naturwissenschaftlichen Sachunterricht?

Kodierung: 1 = in jeder Stunde; 2 = in den meisten Stunden; 3 = in manchen Stunden; 4 = nie oder fast nie

Item: Die Schüler hören der Lehrerin/dem Lehrer nicht zu. (ASXS27A)

Kodierung	1	2	3	4	Gesamt
Absolut	390	592	1164	497	2643
Prozent	14.84	21.86	44.00	19.30	100.0
(SE)	(1.05)	(1.09)	(1.27)	(1.15)	

Item: Es ist laut und unruhig. (ASXS27B)

Kodierung	1	2	3	4	Gesamt
Absolut	413	652	1040	401	2506
Prozent	16.58	25.56	41.38	16.48	100.0
(SE)	(1.14)	(1.12)	(1.25)	(1.18)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer muss lange warten, bis alle Schüler leise sind. (ASXS27C)

Kodierung	1	2	3	4	Gesamt
Absolut	539	634	1054	401	2628
Prozent	20.44	23.67	40.07	15.82	100.0
(SE)	(1.27)	(0.99)	(1.29)	(1.05)	

Item: Die Schüler können nicht gut arbeiten. (ASXS27D)

Kodierung	1	2	3	4	Gesamt
Absolut	269	497	1000	833	2599
Prozent	10.31	18.97	37.92	32.80	100.0
(SE)	(0.85)	(0.84)	(1.16)	(1.44)	

Item: Die Schüler fangen erst an zu arbeiten, wenn der Unterricht schon lange begonnen hat. (ASXS27E)

Kodierung	1	2	3	4	Gesamt
Absolut	372	441	846	966	2625
Prozent	14.08	16.81	31.95	37.16	100.0
(SE)	(0.98)	(0.78)	(1.19)	(1.51)	

Skala: Wahrgenommene Unterstützung und kognitive Aktivierung durch die Lehrperson (Sachunterricht)

Frage: Wie sehr stimmst du den folgenden Aussagen zum naturwissenschaftlichen Sachunterricht zu?

Kodierung: 1 = stimme völlig zu; 2 = stimme eher zu; 3 = stimme eher nicht zu; 4 = stimme überhaupt nicht zu

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer ist auch dann nett zu mir, wenn ich einen Fehler mache. (ASXS28A)

Kodierung	1	2	3	4	Gesamt
Absolut	1572	740	169	145	2626
Prozent	59.62	28.20	6.59	5.60	100.0
(SE)	(1.37)	(0.98)	(0.62)	(0.48)	

Item: Ich bin unserer Sachunterrichtslehrerin/unserem Sachunterrichtslehrer wichtig. (ASXS28B)

Kodierung	1	2	3	4	Gesamt
Absolut	826	1026	503	165	2520
Prozent	33.06	40.69	19.85	6.41	100.0
(SE)	(1.30)	(1.00)	(1.02)	(0.59)	

Item: Wenn ich einen Fehler mache, sagt unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer mir, wie ich es besser machen kann. (ASXS28C)

Kodierung	1	2	3	4	Gesamt
Absolut	1461	821	216	99	2597
Prozent	56.60	31.51	8.22	3.67	100.0
(SE)	(1.40)	(1.11)	(0.58)	(0.46)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer mag mich. (ASXS28D)

Kodierung	1	2	3	4	Gesamt
Absolut	1047	982	309	140	2478
Prozent	41.99	39.83	12.47	5.71	100.0
(SE)	(1.53)	(1.17)	(0.96)	(0.54)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer glaubt, dass ich schwierige Aufgaben lösen kann. (ASXS28E)

Kodierung	1	2	3	4	Gesamt
Absolut	855	1060	466	157	2538
Prozent	33.54	42.10	18.08	6.27	100.0
(SE)	(1.12)	(0.94)	(0.77)	(0.49)	

Item: Wir bearbeiten im Sachunterricht Aufgaben, über die ich sehr gründlich nachdenken muss. (ASXS28F)

Kodierung	1	2	3	4	Gesamt
Absolut	683	996	687	192	2558
Prozent	26.48	38.89	27.24	7.39	100.0
(SE)	(0.90)	(0.97)	(1.11)	(0.54)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer fragt mich, was ich verstanden habe und was nicht. (ASXS28G)

Kodierung	1	2	3	4	Gesamt
Absolut	919	919	496	227	2561
Prozent	35.67	36.14	19.29	8.90	100.0
(SE)	(1.22)	(1.04)	(0.85)	(0.68)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer stellt Fragen, über die ich sehr gründlich nachdenken muss. (ASXS28H)

Kodierung	1	2	3	4	Gesamt
Absolut	688	962	691	212	2553
Prozent	26.80	37.79	27.32	8.09	100.0
(SE)	(1.09)	(1.24)	(1.25)	(0.52)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer stellt uns Aufgaben, die auf den ersten Blick schwierig wirken. (ASXS28I)

Kodierung	1	2	3	4	Gesamt
Absolut	756	1065	559	163	2543
Prozent	29.60	42.13	21.91	6.35	100.0
(SE)	(0.99)	(1.06)	(0.79)	(0.49)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer fragt uns bei einem neuen Thema, was wir schon wissen. (ASXS28J)

Kodierung	1	2	3	4	Gesamt
Absolut	1236	837	338	117	2528
Prozent	48.96	33.08	13.24	4.72	100.0
(SE)	(1.16)	(0.91)	(0.77)	(0.49)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer stellt uns Aufgaben, über die ich gerne nachdenke. (ASXS28K)

Kodierung	1	2	3	4	Gesamt
Absolut	841	986	497	195	2519
Prozent	33.64	39.46	19.37	7.53	100.0
(SE)	(1.06)	(0.94)	(0.95)	(0.60)	

Item: Unsere Sachunterrichtslehrerin/unser Sachunterrichtslehrer möchte, dass ich meine Antworten erkläre. (ASXS28L)

Kodierung	1	2	3	4	Gesamt
Absolut	1026	1011	370	97	2504
Prozent	40.82	40.73	14.60	3.86	100.0
(SE)	(1.15)	(1.19)	(0.87)	(0.38)	

Skala: Zufriedenheit mit Lehrberuf

Frage: Wie häufig empfinden Sie folgendermaßen in Bezug auf Ihre Tätigkeit als Lehrerin/Lehrer?

Kodierung: 1 = sehr häufig; 2 = häufig; 3 = manchmal; 4 = nie

Item: Ich bin mit meinem Beruf als Lehrerin/Lehrer zufrieden. (ATBG08A)

Kodierung	1	2	3	4	Gesamt
Absolut	122	135	20	0	277
Prozent	46.12	46.91	6.97	0.00	100.0
(SE)	(3.24)	(3.09)	(1.69)	(-)	

Item: Ich empfinde meine Arbeit als sehr sinnvoll und wichtig. (ATBG08B)

Kodierung	1	2	3	4	Gesamt
Absolut	186	85	6	0	277
Prozent	67.95	29.56	2.49	0.00	100.0
(SE)	(3.11)	(3.22)	(1.12)	(-)	

Item: Mein Beruf begeistert mich. (ATBG08C)

Kodierung	1	2	3	4	Gesamt
Absolut	106	142	28	0	276
Prozent	38.30	51.72	9.99	0.00	100.0
(SE)	(3.98)	(3.96)	(1.95)	(-)	

Item: Meine Arbeit inspiriert mich. (ATBG08D)

Kodierung	1	2	3	4	Gesamt
Absolut	94	133	49	0	276
Prozent	35.36	46.22	18.42	0.00	100.0
(SE)	(3.54)	(3.44)	(2.75)	(-)	

Item: Ich bin stolz auf meine Arbeit. (ATBG08E)

Kodierung	1	2	3	4	Gesamt
Absolut	115	130	28	2	275
Prozent	45.01	42.99	10.93	1.07	100.0
(SE)	(3.37)	(3.08)	(2.25)	(0.10)	

Skala: Sicherer Umgang beim Unterrichten im Sachunterricht

Frage: Wie sicher fühlen Sie sich im naturwissenschaftlichen Sachunterricht in dieser Klasse bei den folgenden Tätigkeiten?

Kodierung: 1 = sehr sicher; 2 = sicher; 3 = einigermaßen sicher; 4 = unsicher

Item: Schüler für das Lernen in Naturwissenschaften zu begeistern (ATXS17A)

Kodierung	1	2	3	4	Gesamt
Absolut	43	98	44	0	185
Prozent	23.93	54.00	22.07	0.00	100.0
(SE)	(3.73)	(3.61)	(3.02)	(-)	

Item: naturwissenschaftliche Konzepte oder Prinzipien durch naturwissenschaftliche Experimente erklären (ATXS17B)

Kodierung	1	2	3	4	Gesamt
Absolut	20	73	85	7	185
Prozent	10.21	40.75	45.46	3.58	100.0
(SE)	(2.18)	(3.91)	(3.83)	(1.49)	

Item: den leistungsfähigsten Schülern herausfordernde Aufgaben stellen (ATXS17C)

Kodierung	1	2	3	4	Gesamt
Absolut	20	71	75	19	185
Prozent	10.21	39.86	39.86	10.07	100.0
(SE)	(2.39)	(4.00)	(3.74)	(2.45)	

Item: meinen Unterricht so gestalten, dass das Interesse der Schüler geweckt wird (ATXS17D)

Kodierung	1	2	3	4	Gesamt
Absolut	40	114	29	2	185
Prozent	20.68	63.75	14.70	0.88	100.0
(SE)	(2.88)	(4.10)	(2.85)	(0.69)	

Item: den Schülern vermitteln, wie wichtig es ist, Naturwissenschaften zu lernen (ATXS17E)

Kodierung	1	2	3	4	Gesamt
Absolut	32	105	43	4	184
Prozent	16.94	56.33	24.57	2.16	100.0
(SE)	(2.82)	(3.67)	(3.59)	(1.10)	

Item: das naturwissenschaftliche Verständnis der Schüler bewerten (ATXS17F)

Kodierung	1	2	3	4	Gesamt
Absolut	18	101	59	7	185
Prozent	8.98	55.04	32.21	3.77	100.0
(SE)	(2.20)	(3.99)	(3.73)	(1.44)	

Item: Schüler, die Schwierigkeiten haben, darin unterstützen, den Lernstoff zu verstehen (ATXS17G)

Kodierung	1	2	3	4	Gesamt
Absolut	23	109	52	2	186
Prozent	11.02	59.25	28.81	0.92	100.0
(SE)	(2.29)	(4.13)	(3.77)	(0.65)	

Item: die Bedeutsamkeit und Relevanz von Naturwissenschaften verdeutlichen (ATXS17H)

Kodierung	1	2	3	4	Gesamt
Absolut	26	103	53	2	184
Prozent	13.84	54.90	30.15	1.10	100.0
(SE)	(2.57)	(3.71)	(3.32)	(0.79)	

Item: Förderung der Denkfähigkeiten höherer Ordnung der Schüler (ATXS17I)

Kodierung	1	2	3	4	Gesamt
Absolut	10	61	91	15	177
Prozent	4.79	34.44	51.90	8.87	100.0
(SE)	(1.25)	(3.27)	(3.68)	(2.23)	

Item: Naturwissenschaft mithilfe der Methode des forschend-entdeckenden Lernens unterrichten (ATXS17J)

Kodierung	1	2	3	4	Gesamt
Absolut	17	83	74	8	182
Prozent	8.41	46.05	40.29	5.25	100.0
(SE)	(1.86)	(3.58)	(3.30)	(1.78)	

Faktorladungen der Vor- und Hauptmodelle

Vormodell I

Berufszufriedenheit		Lehrkräfteselbstwirksamkeit	
Item	Ladung	Item	Ladung
A08Ain	.666	A17Ain	.637
A08Bin	.601	A17Cin	.731
A08Cin	.813	A17Din	.666
A08Din	.810	A17Ein	.709
A08Ein	.792	A17Fin	.691
		A17Gin	.603
		A17Hin	.642
		A17Iin	.691

Vormodell II

Individuelle Förderung		Klassenführung		Konstruktive Unterstützung		Kognitive Aktivierung	
Item	Ladung	Item	Ladung	Item	Ladung	Item	Ladung
mA22Ain	.420	mA27An	.714	mA28Ain	.435	mA28Fin	.236
mA22Bin	.459	mA27Bn	.843	mA28Bin	.645	mA28Gin	.676
mA22Cin	.749	mA27Cn	.838	mA28Cin	.571	mA28Hin	.375
mA22Din	.775	mA27Dn	.556	mA28Din	.625	mA28Iin	.539
mA22Ein	.792	mA27En	.680	mA28Ein	.521	mA28Kin	.517
mA22Fin	.780						
mA22Gin	.792						

Hauptmodell I

Fachbezogene Einstellung		Fachbezogenes Selbstkonzept	
Item	Ladung	Item	Ladung
A07Ain	.847	A09Ain	.787
A07Bn	.687	A09Din	.789
A07Din	.705	A09Ein	.666
A07Ein	.910	A09Bn	.749
A07Fin	.884	A09Cn	.743
A07Iin	.821	A09Fn	.803
		A09Gn	.758
		SKpositiv	.871
		SKnegativ	.700

Hauptmodell II

Fachbezogene Einstellung		Fachbezogenes Selbstkonzept	
Item	Ladung	Item	Ladung
A07Ain	.829	A09Ain	.775
A07Bn	.661	A09Din	.779
A07Din	.680	A09Ein	.652
A07Ein	.898	A09Bn	.741
A07Fin	.870	A09Cn	.734
A07lin	.802	A09Fn	.795
		A09Gn	.749
		SKpositiv	.868
		SKnegativ	.683

Individuelle Förderung		Klassenführung		Konstruktive Unterstützung		Kognitive Aktivierung	
Item	Ladung	Item	Ladung	Item	Ladung	Item	Ladung
mA22Ain	.425	mA27An	.714	mA28Ain	.452	mA28Fin	.194
mA22Bin	.465	mA27Bn	.843	mA28Bin	.618	mA28Gin	.607
mA22Cin	.751	mA27Cn	.839	mA28Cin	.586	mA28Hin	.290
mA22Din	.773	mA27Dn	.554	mA28Din	.610	mA28lin	.463
mA22Ein	.790	mA27En	.680	mA28Ein	.551	mA28Kin	.602
mA22Fin	.781						
mA22Gin	.793						

Hauptmodell III

Fachbezogene Einstellung		Fachbezogenes Selbstkonzept	
Item	Ladung	Item	Ladung
A07Ain	.829	A09Ain	.775
A07Bn	.661	A09Din	.779
A07Din	.680	A09Ein	.652
A07Ein	.898	A09Bn	.741
A07Fin	.870	A09Cn	.734
A07Iin	.802	A09Fn	.795
		A09Gn	.749
		SKpositiv	.868
		SKnegativ	.683

Individuelle Förderung		Klassenführung		Konstruktive Unterstützung		Kognitive Aktivierung	
Item	Ladung	Item	Ladung	Item	Ladung	Item	Ladung
mA22Ain	.424	mA27An	.716	mA28Ain	.450	mA28Fin	.193
mA22Bin	.463	mA27Bn	.842	mA28Bin	.617	mA28Gin	.607
mA22Cin	.749	mA27Cn	.840	mA28Cin	.586	mA28Hin	.290
mA22Din	.772	mA27Dn	.555	mA28Din	.610	mA28Iin	.463
mA22Ein	.790	mA27En	.680	mA28Ein	.551	mA28Kin	.601
mA22Fin	.781						
mA22Gin	.793						

Berufszufriedenheit		Lehrkräfteselbstwirksamkeit	
Item	Ladung	Item	Ladung
A08Ain	.663	A17Ain	.642
A08Bin	.599	A17Cin	.721
A08Cin	.815	A17Din	.668
A08Din	.809	A17Ein	.725
A08Ein	.792	A17Fin	.684
		A17Gin	.594
		A17Hin	.645
		A17Iin	.679

Hauptmodell IV / Gesamtmodell

Fachbezogene Einstellung		Fachbezogenes Selbstkonzept	
Item	Ladung	Item	Ladung
A07Ain	.831	A09Ain	.791
A07Bn	.666	A09Din	.775
A07Din	.680	A09Ein	.655
A07Ein	.903	A09Bn	.747
A07Fin	.877	A09Cn	.739
A07lin	.803	A09Fn	.791
		A09Gn	.743
		SKpositiv	.905
		SKnegativ	.721

Individuelle Förderung		Klassenführung		Konstruktive Unterstützung		Kognitive Aktivierung	
Item	Ladung	Item	Ladung	Item	Ladung	Item	Ladung
mA22Ain	.423	mA27An	.709	mA28Ain	.467	mA28Fin	.216
mA22Bin	.463	mA27Bn	.844	mA28Bin	.644	mA28Gin	.701
mA22Cin	.748	mA27Cn	.838	mA28Cin	.619	mA28Hin	.367
mA22Din	.777	mA27Dn	.550	mA28Din	.631	mA28lin	.537
mA22Ein	.792	mA27En	.676	mA28Ein	.550	mA28Kin	.495
mA22Fin	.780						
mA22Gin	.800						

Berufszufriedenheit		Lehrkräfteselbstwirksamkeit	
Item	Ladung	Item	Ladung
A08Ain	.666	A17Ain	.638
A08Bin	.602	A17Cin	.736
A08Cin	.818	A17Din	.671
A08Din	.808	A17Ein	.731
A08Ein	.796	A17Fin	.698
		A17Gin	.595
		A17Hin	.650
		A17lin	.685

Über die Autorin

Luise Annemarie Scholz (geboren 1997 in Rostock) studierte Erziehungs- und Bildungswissenschaft und Biologie mit den Schwerpunkten Sozialpädagogik/Kinder- und Jugendbildung sowie Empirische Bildungsforschung und Bildungstheorie an der Universität Hamburg. Nach dem Masterabschluss arbeitete sie zunächst als wissenschaftliche Angestellte im KMK- und BMBF-geförderten Projekt *Trends in International Mathematics and Science Study* (TIMSS 2019) an der Universität Hamburg. Seit 2021 ist sie in den Nachfolgeprojekten TIMSS 2023 und TIMSS 2027 Teil der nationalen Projektleitung und promovierte in diesem Rahmen im Jahr 2024 bei Prof. Dr. Knut Schwippert.