



Leistungsbezogene und sprachliche Heterogenität in der Schule und deren erwartete Effekte auf die Leistung von Schüler*innen

Inauguraldissertation zur Erlangung des akademischen Grades
einer Doktorin der Sozialwissenschaften
der Universität Mannheim

Vorgelegt von
Stefanie Dotzel

Datum der Disputation: 17.12.2020

Dekan der Fakultät für Sozialwissenschaften:

Prof. Dr. Michael Diehl

BetreuerInnen der Dissertation:

Prof. Dr. Karina Karst

Prof. Dr. Stefan Münzer

Gutachter:

Prof. Dr. Stefan Münzer

Prof. Dr. Johannes Hartig

Danksagung

Mit der vorliegenden Dissertation schließe ich eine besondere Lebensphase ab; eine Lebensphase, in der ich mich nicht nur beruflich weiterentwickelt habe, sondern vor allem auch in persönlicher Hinsicht. Ich denke nun „anders“ über die Welt – und dafür bin ich allen Erfahrungen der letzten vier Jahre sehr dankbar. Zu diesen Erfahrungen hat nicht nur die Wissenschaft beigetragen – in meiner Promotionszeit haben mich viele Personen begleitet. Diesen möchte ich mit der folgenden Danksagung von ganzem Herzen mein Dankeschön aussprechen.

Mein besonderer Dank gilt *Prof. Dr. Karina Karst*, die mich fachlich mit fundiertem Wissen, hilfreichen Anmerkungen und Hinweisen in dem Gelingen meines Dissertationsvorhabens unterstützt hat. Zudem schätze ich ihre gelassene Art, die mir vor allem in den – sagen wir – „ambivalenten“ Phasen meiner Dissertationszeit geholfen hat, zuversichtlich zu bleiben.

Neben der Bereitschaft, die vorliegende Arbeit zu begutachten, möchte ich *Prof. Dr. Stefan Münzer* und *Prof. Dr. Johannes Hartig* vor allem für wertvolle Tipps im Rahmen von Forschungskolloquien, Konferenzen und persönlichen Gesprächen danken.

Darüber hinaus möchte ich allen Lehrstuhlinhaber*innen und Mitarbeiter*innen des *Lehrstuhls für Bildungspsychologie*, des *Lehrstuhls für Pädagogische Psychologie* und der *Juniorprofessur für Unterrichtsqualität in heterogenen Kontexten* danken. Ich habe mich als Teil dieses großen Teams ab Beginn an sehr wohl gefühlt und es genossen, mich wissenschaftlich mit euch allen auszutauschen. Ein besonderer Dank gilt an dieser Stelle *Dr. Benedict Fehringer* für die treue Rolle als Denksport-Partner – vor allem zu den späten und dunklen Stunden am Lehrstuhl. Zudem möchte ich mich bei ihm für dessen Unterstützung der nicht selten aufgetretenen Error-Messages in R bedanken. Durch seine hilfreichen Tipps konnte ich vor allem im Bereich der Methoden und der statistischen Programmierung große Fortschritte erzielen. Darüber hinaus möchte ich mich bei meinen Freundinnen *Dr. Meike Bonefeld* und *Anne-Sophie Waag* bedanken: für euer stetiges konstruktives Feedback und vor allem für eure emotionale Begleitung. Der Austausch mit euch war für mich vor allem für die Abschlussphase der Dissertation enorm wichtig. Mit eurer Hilfe ist es mir gelungen, auch in den letzten Monaten meine Zuversicht beizubehalten. Vielen lieben Dank dafür!

Außerdem möchte ich der Arbeitsgruppe *Diagnostik und Evaluation im schulischen Kontext* der Universität Münster mit besonderer Hervorhebung von *Dr. Natalie Förster*, *Dr. Karin Hebbecke* und *Prof. Dr. Elmar Souvignier* meinen Dank aussprechen – dafür, dass sie meine erste Forschungsidee mit Begeisterung unterstützt haben, für deren Kooperationsbereitschaft und die gelungene Zusammenarbeit.

Des Weiteren möchte ich mich beim ganzen Team der *Heideschule in Schwebheim* bedanken. Die ergänzende Tätigkeit in der pädagogisch-psychologischen Praxis hat mir während meiner

Promotionszeit sehr geholfen, das wissenschaftliche Arbeiten von einem anderen Blickwinkel zu betrachten und durch den Praxisbezug kritischer gegenüber meiner Arbeit zu werden. Insbesondere möchte ich an dieser Stelle *Ulrike Weber, Christina Bendorf, Lena Rechtenbacher* und *Marion Schröder* danken; für den fachlichen Austausch, für die emotionale Unterstützung und deren interessierter Art, mich hinsichtlich meines wissenschaftlichen Karriereweges auch weiterhin zu begleiten.

Überdies möchte ich mich bei der *Hanns-Seidel-Stiftung* und deren Förderung durch das *Bundesministerium für Bildung und Forschung* (BMBF) bedanken. Vor allem durch die ideelle Förderung konnte ich meine fachfremden Kompetenzen enorm weiterentwickeln und habe durch die Kontakte mit anderen Doktorand*innen einen gewinnbringenden Austausch erfahren können. Ein besonderer Dank gilt an dieser Stelle *Dr. Andreas Burtscheidt*, welcher mit seiner offenen und unterstützenden Art für jedes Anliegen ein Ohr hatte.

Des Weiteren möchte ich mich bei meinem Freund *Ralf Knaier* bedanken, mit welchem ich seit dem ersten Semester meines Studiums einen regen wissenschaftlichen Austausch führe. Die Freundschaft hat einen Teil dazu beigetragen, dass ich den Weg in die Wissenschaft überhaupt erst gewagt habe.

Zudem möchte ich mich bei meinen Eltern *Irmi und Edgar Dotzel* bedanken sowie bei meiner Schwester *Annika Dotzel*; dafür, dass sie mich bedingungslos bei meinen Lebensentscheidungen unterstützt haben und stets ein genuines Interesse an mir und meinem beruflichen Weg zeigen. Ein besonderer Dank gilt auch meinen bereits verstorbenen Großeltern *Opa Toni* und *Oma Reta*, dass sie mir schon als Kleinkind „die Welt“ aus einer pädagogisch-humanistischen Perspektive gezeigt haben und dadurch einen großen Beitrag dazu geleistet haben, wie ich heute denke und arbeite.

Der letzte Dank gebührt meinem besten Freund und treuesten Gefährten *Josua Neubert*. Er war derjenige, welcher mich den ganzen Weg über begleitete, mich nach Misserfolgen aufbaute und meine Erfolge gemeinsam mit mir feierte. Ein besonderer Dank gilt ihm, weil er an mich glaubt, wenn ich den Glauben an mich verliere – und weil er so lustig ist, dass er mich auch bei schlechter Laune zum Lachen bringt. Er ist mein engster Vertrauter und auch als „Nicht-Wissenschaftler“ mein bester Ratgeber.

Würzburg im August 2020

Zusammenfassung

In den letzten Jahrzehnten ist in deutschen Klassen ein Anstieg an sprachlicher und leistungsbezogener Heterogenität zu beobachten (Markic & Abels, 2014; Maehler & Brinkmann, 2016). Lehrkräfte stehen vor der Herausforderung, mit dieser Heterogenität professionell umzugehen und Schüler*innen sind durch diese Art der heterogenen Klassenzusammensetzung auch einer spezifischen Lernumwelt ausgesetzt. Vor diesem Hintergrund wird in der vorliegenden Arbeit das Thema *Leistungsbezogene und sprachliche Heterogenität in der Schule und deren erwarteten Effekte auf die Leistung von Schüler*innen* in den Blick genommen. Für die theoretische Einordnung wird eine Verknüpfung von zwei Unterrichtsmodellen vollzogen; dem Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts nach Helmke (2017) und dem Stufenmodell nach Pietsch (2010). Das übergeordnete Ziel dieser Arbeit ist, den wissenschaftlichen Diskurs und den Forschungsstand im Bereich leistungsbezogene und sprachliche Heterogenität in der Klasse und deren Effekte auf die Schüler*innenleistung zu erweitern. Um dieses Ziel zu erreichen, werden in der vorliegenden Arbeit drei Beiträge vorgestellt. *Beitrag 1* berichtet über die Effekte von Leistungsheterogenität der Klasse auf die Schüler*innenleistung unter Berücksichtigung verschiedener Heterogenitätsmaße. *Beitrag 2* untersucht die Effekte der sprachlichen Heterogenität der Klasse auf die Leistung von Schüler*innen. Und *Beitrag 3* analysiert die Einstellung zur Leistungsheterogenität von Schüler*innen und nimmt dabei unter anderem den Einfluss der tatsächlichen Leistungsheterogenität in den Blick. Die Ergebnisse zeigen, dass leistungsbezogene und sprachliche Heterogenität die Schüler*innenleistung beeinflussen. Jedoch unterscheidet sich die Art der Effekte: während bei der sprachlichen Heterogenität positive lineare Effekte gefunden werden, zeigen sich bei der leistungsbezogenen Heterogenität nur kurvilineare Effekte. Die Ergebnisse zeigen zudem, dass eine unterschiedliche Operationalisierung von leistungsbezogener und sprachlicher Heterogenität zwar aus pädagogisch-inhaltlicher Perspektive sinnvoll ist, jedoch keine bis kaum bedeutsame Unterschiede in den empirischen Ergebnissen bringt. Außerdem zeigen die Ergebnisse, dass leistungsbezogene Heterogenität zwar keinen direkten Einfluss auf die Leistung nimmt, jedoch auf die Einstellung zur Leistungsheterogenität bei Schüler*innen. Auf Basis der Ergebnisse kann festgehalten werden, dass Heterogenität positive Auswirkungen auf die Schüler*innenleistung haben kann. Deswegen ist es von Bedeutung, dass Lehrkräfte im Umgang mit Heterogenität eine fundierte und umfassende Ausbildung erhalten; so dass es diesen gelingt, die Chancen heterogener Lerngruppen zu nutzen und die Leistungssteigerung aller Schüler*innen zu fördern.

Inhaltsverzeichnis

1	Einleitung.....	3
2	Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts.....	6
3	Heterogenität in der Schule.....	9
3.1	Definition	9
3.2	Heterogenitätsdimensionen.....	9
3.3	Umgang mit Heterogenität in Schule und Unterricht.....	12
3.4	Zusammenfassung der theoretischen Einordnung.....	16
4	Heterogenität und Schüler*innenleistung	18
5	Forschungslücken und Ziele der vorliegenden Arbeit	22
6	Ableitung des Arbeitsmodells und dazugehöriger Fragestellungen.....	23
6.1	Leistungsheterogenitätsmaße	25
6.2	Sprachliche Heterogenität	27
6.3	Einstellung zur Leistungsheterogenität von Schüler*innen	29
7	Beiträge	32
7.1	Beitrag 1 – Leistungsheterogenitätsmaße	32
7.2	Beitrag 2 – Sprachliche Heterogenität	34
7.3	Beitrag 3 – Einstellung zur Leistungsheterogenität von Schüler*innen	35
8	Diskussion.....	37
8.1	Zusammenfassung und Synthese der Befunde.....	38
8.2	Limitationen.....	41
8.3	Implikationen für zukünftige Forschung.....	42
8.4	Implikationen für die Praxis.....	43
8.5	Fazit.....	45
	Literaturangaben	46
	Abbildungsverzeichnis.....	54
	Anhang A: Manuskript zu Beitrag 1 – Leistungsheterogenitätsmaße.....	55
	Anhang B: Manuskript zu Beitrag 2 – Sprachliche Heterogenität.....	89
	Anhang C: Manuskript zu Beitrag 3 – Einstellung zur Leistungsheterogenität von Schüler*innen.....	120
	Anhang D: Beitrag im Sammelband	148
	Anhang E: Erklärungen über die Koautor*innenschaft	172

1 Einleitung

Mit der PISA-Studie im Jahr 2000 wurde der Begriff Heterogenität ins öffentliche Interesse gerückt; seitdem wird das Thema *Heterogenität* kontrovers diskutiert (Scharenberg, 2012). Auch die pädagogisch-psychologische Forschung setzt sich seit dem sogenannten *PISA-Schock* stark mit dem Phänomen Heterogenität auseinander. Wirft man jedoch einen Blick in die Geschichte der Pädagogik, so stellt man fest, dass Heterogenität keinesfalls ein neuer Begriff ist. Bereits Johann Friedrich Herbart (1776-1841) sah in der „Verschiedenheit der Köpfe“ ein zentrales Problem des Unterrichts und auch Ernst Christian Trapp (1745-1810), der erste deutsche Pädagogikprofessor (1779, Halle) forderte eine Didaktik, in der sich der Unterricht auf die „Mittelköpfe“ konzentrieren sollte (zitiert nach Sandfuchs, 1994, S. 340; Tillmann, 2004, S. 33). Die Ausrichtung des Schulunterrichts auf ein mittleres Leistungsmaß hat in Deutschland also eine lange Tradition (Tillmann, 2004); und doch konnten sich in den letzten Jahrzehnten Ansätze herausbilden, die Heterogenität aus einer neuen Perspektive betrachten: Die Verschiedenheit von Schüler*innen wird begrüßt, die Heterogenität der Lerngruppen wird als Chance gesehen und Konzepte wie Individualisierung und Binnendifferenzierung zielen darauf ab, die Individualität der Schüler*innen zu fokussieren und die Unterschiede der Klasse pädagogisch fruchtbar zu machen (z.B. Biederbeck & Rothland, 2017; Boller, Rosowski & Stroot, 2007).

Diese *Lesart von Heterogenität* stellt in Deutschland jedoch immer noch große Herausforderungen dar: So zeigt sich, dass sich deutsche Lehrkräfte am häufigsten über eine zu große Heterogenität in den Klassen beklagen und das, obwohl in Deutschland die im internationalen Vergleich homogensten Lerngruppen vorzufinden sind (Baumert, 2002; vgl. Möller, 2006; Scholz, 2007; Tillmann, 2008). Auch werden im deutschen Schulsystem weiterhin Homogenisierungsmaßnahmen genutzt, wie beispielsweise die Zurückstellung von der Einschulung, Klassenwiederholungen, Überweisung an Förderschulen oder der Wechsel von Schulformen (van Ackeren & Kühn, 2017). Diese Beobachtungen unterstreichen, dass in Deutschland weiterhin an einer „fiktiven Homogenität“ (Tillmann, 2004, S. 38) festgehalten wird.

Die pädagogisch-psychologische Literatur macht an dieser Stelle deutlich, dass die Schulpraxis wie auch die Bildungspolitik von der Idee Abstand nehmen müssen, dass Schüler*innen mit unterschiedlichen Lernvoraussetzungen bei gleichem Unterricht gleiche Leistungen erbringen können (z.B. Trautmann & Wischer, 2011; Walgenbach, 2017). Ferner soll der Fokus darauf gerichtet werden, einen möglichst starken individuellen Kompetenzzuwachs anzustreben (Prenzel & Burba, 2007), auch wenn dies mit einer steigenden

Heterogenität einhergehen kann (vgl. Walgenbach, 2017). Das Thema *Heterogenität* spielt also in vielerlei Hinsicht eine Rolle für die Schulpraxis und die Lernumwelt von Schüler*innen – doch obwohl sich die Wissenschaft seit über 100 Jahren mit diesem Thema auseinandersetzt, weist der Forschungsstand bedeutsame Lücken auf.

Die vorliegende Arbeit hat deswegen zum Ziel, zu einem tieferen Verständnis über Heterogenität in der Schule beizutragen und den Fokus dabei auf die erwarteten Effekte auf die Schüler*innenleistung zu legen. Dieses Ziel verfolgt diese Arbeit mit drei Beiträgen, die jeweils einen spezifischen Aspekt von Heterogenität aufgreifen und analysieren. Beitrag 1 berichtet über die Effekte von Leistungsheterogenität auf die Schüler*innenleistung. Beitrag 2 untersucht die Effekte der sprachlichen Heterogenität in der Klasse auf die Leistung von Schüler*innen. Und Beitrag 3 analysiert die Einstellung zur Leistungsheterogenität von Schüler*innen und nimmt dabei unter anderem den Einfluss der Leistungsheterogenität in den Blick. Der Fokus der vorliegenden Dissertation liegt also auf den leistungsbedingten und den sprachlichen Unterschieden in Schulklassen und deren Effekte auf die Leistung und Einstellung von Schüler*innen. Die vorliegende Arbeit weist zwei Besonderheiten auf. Erstens werden in zweien der Beiträge (Beitrag 2 und Beitrag 3) noch nicht untersuchte Forschungsfelder exploriert. Und zweitens werden in der Arbeit nicht nur theoretische, sondern auch explorative Herangehensweisen genutzt (Beitrag 1). Diese Vorgehensweisen haben zum Ziel, neue Perspektiven und Ansätze in der Heterogenitätsforschung zu schaffen, den wissenschaftlichen Diskurs zum Thema Heterogenität anzuregen und dabei einen Neuigkeitswert zum bisherigen Forschungsstand zu generieren.

Um die drei Beiträge in einen erweiterten Zusammenhang zu betrachten, werden ausgehend vom Angebots-Nutzungsmodell unterrichtlicher Wirkungen nach Helmke (2017) insbesondere die Auswirkungen vom Kontextmerkmal *Heterogenität* auf die Leistung von Schüler*innen in den Blick genommen (Kapitel 2). Dazu wird in der vorliegenden Arbeit das Phänomen Heterogenität zunächst aus theoretischer Perspektive beleuchtet (Kapitel 3). Im Anschluss wird der bisherige Forschungsstand zum Thema *Heterogenität und Schüler*innenleistung* skizziert (Kapitel 4), um darauf aufbauend die Forschungslücken (Kapitel 5) zu verdeutlichen. Daran anknüpfend wird das Arbeitsmodell der Dissertation vorgestellt und die konkreten Fragestellungen abgeleitet (Kapitel 6). Danach werden die zentralen Befunde überblicksartig dargestellt (Kapitel 7), abschließend in einen Gesamtzusammenhang gebracht und vor dem Hintergrund wissenschaftlicher sowie praktischer Implikationen diskutiert (Kapitel 8).

Im Allgemeinen weist die vorliegende Dissertation also einen deduktiven Aufbau auf. In Kapitel 2 und Kapitel 3 werden *allgemeine Modelle und Kriterien* herangezogen, um das Thema zu umfassen und einordnen zu können. In Kapitel 4 und Kapitel 5 geht die Arbeit

in die *Empirie* über und identifiziert *spezielle* Forschungslücken. Das in Kapitel 6 abgeleitete Arbeitsmodell dient dann der Zusammenführung der sowohl allgemeinen Einordnung des Themas als auch deren konkreten Forschungslücken.

2 Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts

Schule hat den Auftrag, Schüler*innen zum Schulerfolg zu befähigen (KMK - Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2015a, 2015b, 2020). Schulerfolg ist jedoch ein komplexer Begriff und so hat Schule zeitgleich mehrere Bildungsziele. Neben dem Erwerb von beispielsweise Schlüsselkompetenzen oder sozialen Kompetenzen ist das zentrale Zielkriterium, die Schüler*innen bei der fachlichen Kompetenzentwicklung zu unterstützen (Weinert, 2000). Aus diesem Grund konzentriert sich die vorliegende Arbeit ausschließlich auf die fachlichen Kompetenzen von Schüler*innen. Zudem wird der Fokus auf den Fach- und Kompetenzbereich Deutsch und Lesen gelegt, denn die Lesekompetenz gilt als ein wichtiger Prädiktor für die Leistung in weiteren Fächern (Fuchs et al., 2006; Lee & Fradd, 1998).

Den fachlichen Schulerfolg von Schüler*innen zu beschreiben ist aufgrund der Vielzahl von beeinflussenden Faktoren sehr komplex (Hattie, 2009). So spielen neben schüler*innenbezogenen Einflussfaktoren (z.B. Intelligenz, Geschlecht) auch familiäre Merkmale (z.B. ethnische Herkunft, sozioökonomischer Status) und kontextuelle Merkmale, wie die der Klasse (z.B. leistungsbezogene Klassenzusammensetzung) und die der Schule (z.B. Schulform, leistungsbezogene Schulzusammensetzung) eine wichtige Rolle für die Schüler*innenleistung. Die Leistung von Schüler*innen ist also von vielen Faktoren abhängig und kann aus verschiedenen Perspektiven betrachtet werden. Deswegen existieren unterschiedliche wissenschaftliche Modelle mit verschiedenen Schwerpunkten, welche die Schüler*innenleistung zu erklären versuchen (z.B. Angebots-Nutzungsmodell der Wirkungsweise des Unterrichts, Helmke, 2017; Modell von Qualität und Qualitätssicherung, Ditton, 2000; Modell der Schuleffektivität, Scheerens & Bosker, 1997; vgl. Ökosystemischer Ansatz nach Bronfenbrenner; Bronfenbrenner, 1979). Für die vorliegende Arbeit soll das Angebots-Nutzungs-Modell nach Helmke (2017) aus vier Gründen herangezogen werden: Erstens wirft es einen besonderen Blick auf Kontextmerkmale (vgl. Ditton, 2000). Zweitens nimmt es dabei nicht nur schul- sondern auch klassenspezifische Merkmale in den Blick (vgl. Scheerens & Bosker, 1997). Drittens kann dessen Ursprung in der Pädagogik und pädagogischen Psychologie verortet werden (vgl. Bronfenbrenner, 1979). Und viertens, wird in Helmkes (2017) Modell eine Differenzierung von Schulerfolg vorgenommen, indem in fachliche Kompetenzen, fachübergreifende Kompetenzen und erzieherische

Wirkungen getrennt wird. Dadurch bietet Helmkes (2017) Modell eine geeignete Grundlage für die Einordnung von klassenspezifischer Heterogenität und deren Effekte auf die Schüler*innenleistung. Die folgende Darstellung lehnt sich an Helmke (2017, S. 69 ff.) an. Nach dem Angebots-Nutzungsmodell der Wirkungsweise des Unterrichts nach Helmke (2017) entspricht Unterricht einem Angebot, welches in unterschiedlicher Effektivität und Intensität genutzt werden kann (siehe Abbildung 1). Demnach hat Unterricht keinen direkten Einfluss auf die Leistung von Schüler*innen, sondern wirkt über die Nutzung von Lernaktivitäten im Unterricht als vermittelnder Prozess auf den Lernertrag ein. Vereinfacht kann der vermittelnde Prozess dargestellt werden als eine Kette aus:

Angebot (Unterricht) → Nutzung (Lernaktivitäten) → Ertrag (Wirkungen)

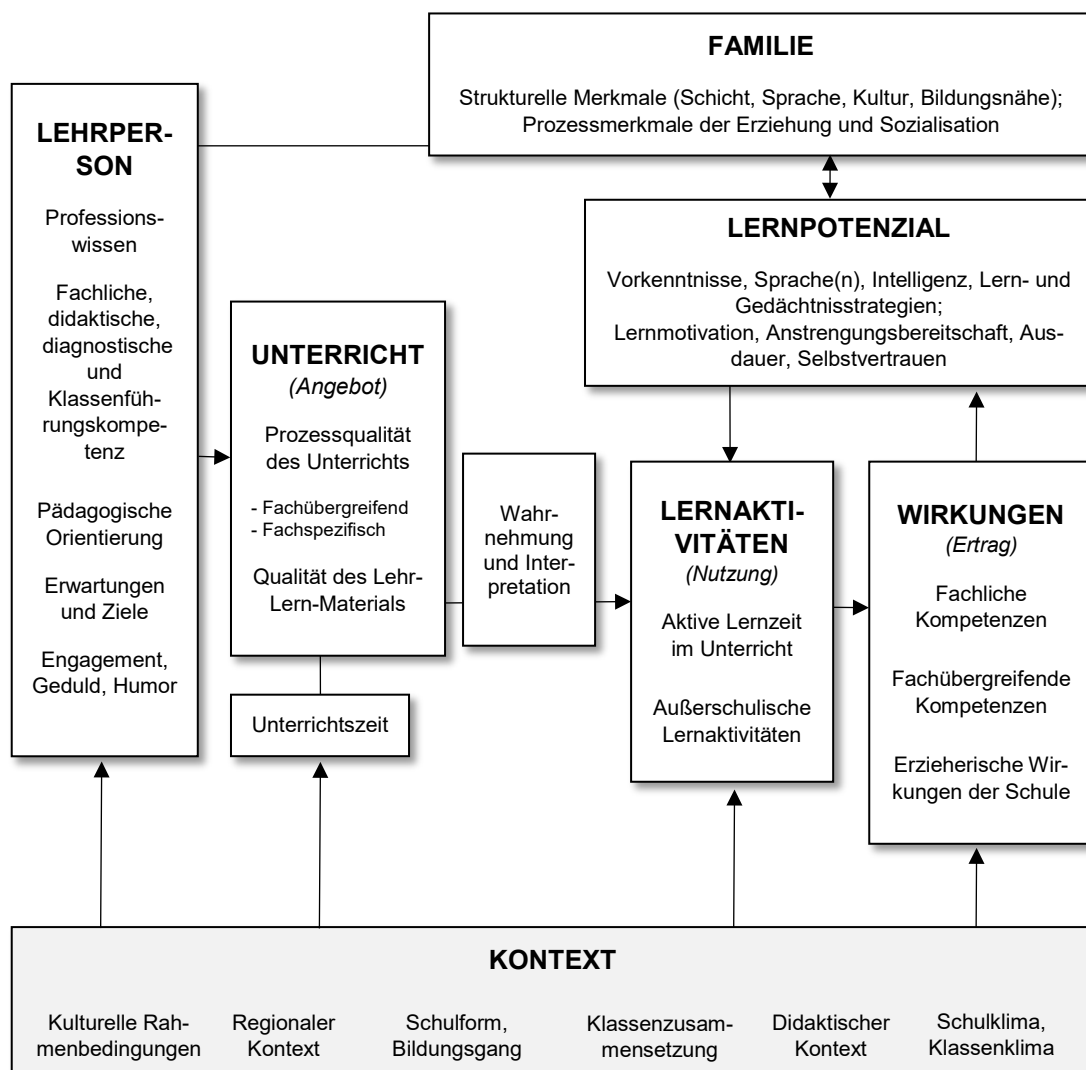


Abbildung 1: Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts (Helmke, 2017)

Solche vermittelnde Prozesse liegen jedoch auch zwischen den einzelnen Faktoren vor. Demnach beeinflusst ein komplexes Zusammenspiel von Faktoren das Angebot und die Nutzung von Unterricht, welches sich wiederum auf die Leistung von Schüler*innen auswirkt. Das Angebots-Nutzungsmodell integriert somit Faktoren der Unterrichtsqualität in ein umfassenderes Modell und beleuchtet dabei sowohl Zielkriterien des Unterrichts als auch die Wirkungsweisen zwischen diesen Zielkriterien und den Faktoren der Unterrichtsqualität. Neben Merkmalen der Lehrperson, des Unterrichts, der Familie, des individuellen Lernpotenzials, der Lernaktivitäten und der Wirkungen sind in dem Modell auch Merkmale des *Kontexts* prävalent. Das Modell lässt also die Bedeutsamkeit des *Kontexts* für die Schüler*innenleistung erkennen.

Kontext und Klassenzusammensetzung. Während die meisten Faktoren des Modells nur einen Faktor beeinflussen, wirkt sich der Kontext auf insgesamt vier Faktoren aus. So hat der Kontext einerseits eine Wirkung auf die Leistung (siehe *Fachliche Kompetenzen*) von Schüler*innen, indem er die Lehrperson, den Unterricht und die Lernaktivität beeinflusst. Andererseits wirkt sich der Kontext aber auch direkt auf die Leistung aus. Der Kontext, in dem sich Lernende befinden, kann somit einen komplexen Einfluss auf die Schüler*innenleistung haben, welcher sich sowohl direkt als auch indirekt äußern kann.

Neben klassenübergreifenden Kontextmerkmalen, wie dem regionalen Kontext oder der Schulform, können auch klassenspezifische Kontextmerkmale – wie die Klassenzusammensetzung – eine Wirkung auf den Lernerfolg von Schüler*innen haben. Die Zusammensetzung einer Klasse setzt sich selbst wiederum aus verschiedenen Aspekten zusammen. So subsumieren sich unter der Klassenzusammensetzung beispielsweise Faktoren der Klassengröße, des mittleren Eingangsniveaus und der Heterogenität.

Auf Basis von Helmkes Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts (2017) spielt Heterogenität für die Vorhersage der Schüler*innenleistung also eine Rolle. Wie der Effekt von Heterogenität auf die Leistung erklärt werden kann, ist aufgrund der Modellstruktur sehr komplex. So kann Heterogenität sowohl einen direkten Einfluss auf die Schüler*innenleistung haben, kann diese aber auch über die Lehrperson, den Unterricht oder die Lernaktivität beeinflussen. Helmke (2017) benutzt in seinem Modell ein sehr breites Begriffsverständnis für Heterogenität. Deswegen soll im nachfolgenden Kapitel die Komplexität des Begriffes aufgebrochen werden.

3 Heterogenität in der Schule

Im folgenden Kapitel wird erläutert, was Heterogenität im eigentlichen Sinne bedeutet, welche Komplexität mit dem Begriff einhergeht und welche Rolle Heterogenität für die Schulpraxis und die Lehrkräfte spielt. Dazu wird zuallererst eine Definition von dem Begriff Heterogenität vorgenommen. Danach werden die für die vorliegende Arbeit relevanten Heterogenitätsdimensionen abgeleitet, um darauf aufbauend auf den professionellen Umgang mit Heterogenität durch die Lehrkraft einzugehen. Jede der in diesem Kapitel vorgestellten theoretischen Inhalte spielt eine Rolle für das in Kapitel 6 elaborierte Arbeitsmodell.

3.1 Definition

Heterogenität lässt sich aus dem Altgriechischen ableiten (*heteros* = anders, verschieden; und *genos* = Klasse, Art) und beschreibt die Verschiedenheit der Teile in einem zusammengesetzten Ganzen (von der Groeben, 2003; Kluge, 2011). Heterogenität verweist somit auf die Verschiedenheit, Ungleichartigkeit oder Andersartigkeit bezogen auf Individuen, Gruppen oder pädagogische Organisationen (Walgenbach, 2017). Ein wichtiges Charakteristikum von Heterogenität ist die Relativität des Begriffes. Heterogenität existiert nämlich nicht an sich, sondern erst im Hinblick auf eine Bezugsgröße (c. q. Homogenität). Heterogenität braucht konstitutiv ein *tertium comparationis* (lateinisch für *Das Dritte des Vergleichs*) und ist somit untrennbar mit Homogenität verbunden (Wenning, 1999). Heterogenität und Homogenität sind somit das zunächst neutrale Ergebnis eines Vergleichs verschiedener Dinge und liegen beide nur bezogen auf einen Maßstab vor (Wenning, 2007). Dementsprechend kann sich die Beurteilung der Heterogenität durch das Anlegen einer neuen Bezugsgröße verändern. So würde dieselbe Klasse von einer Gesamtschullehrkraft als weniger leistungsheterogen eingeschätzt werden als von einer Gymnasiallehrkraft, da eine Gesamtschullehrkraft bezüglich der Beurteilung von Heterogenität in der Regel einen anderen – in diesem Fall heterogeneren – Vergleichsrahmen aufweist (Carl, 2017).

3.2 Heterogenitätsdimensionen

Heterogenität wird laut Definition mit Verschiedenheit, Ungleichheit oder Andersartigkeit beschrieben und ist somit ein breiter Begriff. Heterogenität, also Unterschiede zwischen

den Schüler*innen, kann sich demnach auf zahlreichen Dimensionen zeigen (z.B. nach Wenning, 2007):

1. Leistungsbedingte Unterschiede
2. Altersbedingte Unterschiede
3. Sozialkulturelle Unterschiede
4. Sprachliche Unterschiede
5. Migrationsbedingte Unterschiede
6. Gesundheits- sowie körperbedingte Unterschiede
7. Geschlechtsspezifische Unterschiede

zwischen Schüler*innen.

Heterogenität setzt sich also aus mehreren Heterogenitätsdimensionen zusammen und ist somit selbst ein heterogener Begriff. Auch wenn eine Aufgliederung in verschiedene Heterogenitätsdimensionen nötig ist, um Heterogenität messbar zu machen, sind die einzelnen Dimensionen stark miteinander verstrickt. So können sich beispielsweise sozialkulturelle oder migrationsbedingte Unterschiede in leistungsbezogenen oder sprachlichen Unterschieden manifestieren und können sich dadurch indirekt auf den Unterricht, die Lehrkraft oder die Leistung von Schüler*innen auswirken (Fischer, Rott & Veber, 2014). Überdies können sich auch sprachliche Unterschiede in der leistungsbezogenen Heterogenität manifestieren. Während leistungsbezogene Unterschiede in der Wissenschaft schon seit längerem diskutiert werden, rücken seit den letzten Jahren auch die sprachlichen Unterschiede immer mehr in den Fokus der wissenschaftlichen Debatte.

Dass vor allem die leistungs- und sprachlichen Unterschiede der Schüler*innenschaft im Fokus der wissenschaftlichen Aufmerksamkeit stehen, kann unter anderem auf gesellschaftliche Veränderungen zurückgeführt werden. Zum einen wächst der Anteil an Bundesländern, die ein offenes Übergangssystem von der Elementar- in die Sekundarstufe implementieren (KMK - Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2015c), was zu einer zunehmenden Leistungsheterogenität in den Schulen führt. Dies wird zusätzlich dadurch verstärkt, dass auch immer mehr Schüler*innen mit besonderen Förderbedürfnissen in Regelschulen integriert werden (Markic & Abels, 2014; Powell, 2006). Zum anderen steigt der Anteil an Schüler*innen mit Migrationshintergrund an den Schulen, was die sprachliche Heterogenität in den Klassen ansteigen lässt (Jensen & Rasmussen, 2011; Maehler & Brinkman, 2016). Schulen und deren Lehrkräfte sind also immer häufiger mit einer zunehmenden Leistungsheterogenität

in den Schulklassen sowie mit größeren Unterschieden in den Sprachkenntnissen konfrontiert.

Damit einhergehend zeigt sich, dass auch Lehrkräfte die Heterogenitätsdimensionen Leistung und Sprache als besonders relevant empfinden. Für Lehrkräfte sind vor allem die Heterogenitätsdimensionen von Bedeutung, die sie direkt betreffen und auf die sie direkt einwirken können (Böing, 1996; Scharenberg, 2012). Während Heterogenitätsdimensionen, wie beispielsweise gesundheitsbedingte Unterschiede oder altersbedingte Unterschiede nicht von der Lehrkraft veränderbar sind, kann eine Lehrkraft die leistungsbedingten und sprachlichen Unterschiede in einer Klasse – zumindest in einem gewissen Maße – beeinflussen. Die Heterogenitätsdimensionen Leistung und Sprache und das Handeln der Lehrkraft sind also nicht als unabhängig voneinander zu betrachten. Ferner stehen sie in einem dynamischen Verhältnis zueinander (vgl. Helmke, 2017). So kann das Ausmaß an Leistungs- oder sprachlicher Heterogenität einen Einfluss darauf haben, wie eine Lehrkraft mit dieser Heterogenität umgeht (z.B. Egalisierungsstrategie vs. Elitestrategie; Künstling, Post, Greb, Faust & Lipowsky, 2010; Opp & Speck-Hamdan, 2001; Treinies & Einsiedler, 1996); auf der anderen Seite kann aber auch das Handeln der Lehrkraft einen Einfluss auf die Heterogenität in der Klasse nehmen (z.B. zunehmende vs. abnehmende Heterogenität über ein Schuljahr hinweg; Walgenbach, 2017). Aufgrund der genannten Aspekte werden in der vorliegenden Arbeit vor allem die leistungsbezogene und sprachliche Heterogenität in den Blick genommen.

Leistungsbedingte Unterschiede. Leistungsheterogenität kann definiert werden als die divergenten leistungsbezogenen Ausgangsbedingungen einer Schüler*innenschaft (Scharenberg, 2012). Somit umfasst Leistungsheterogenität Unterschiede in den Lernvoraussetzungen, der Lerngeschwindigkeit und in den Lernergebnissen; und dementsprechend haben Schüler*innen in eher homogenen Klassen ähnliche Lernvoraussetzungen (o.Ä.) und Schüler*innen in eher heterogenen Klassen unterschiedliche Lernvoraussetzungen (o.Ä.).

Sprachliche Unterschiede. Sprachliche Heterogenität kann definiert werden als die Anzahl der verschiedenen Sprachen in einem bestimmten geografischen Gebiet, einschließlich zweier wichtiger Aspekte: erstens, dass Sprachformen als diskrete Sprachen klassifiziert werden können, und zweitens, dass diese diskreten Sprachen zählbar sind. Übertragen auf

den Bildungsbereich bedeutet dies, dass die sprachliche Heterogenität in den Klassen umso größer ist, je mehr Sprachen von den Schüler*innen gesprochen werden (Nettle, 1999).¹ Vor dem Hintergrund von Helmkes Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts (2017) bedeutet dies, dass sowohl die leistungsbezogene als auch sprachliche Heterogenität einen direkten sowie indirekten Einfluss auf die Schüler*innenleistung haben. In diesem Zusammenhang soll der interessante Aspekt hervorgehoben werden, dass für den direkten Effekt von Leistungsheterogenität auf die Schüler*innenleistung keine konkreten theoretischen Ansätze existieren, die den Effekt erklären; und auch für die Erklärung des Effekts der sprachlichen Heterogenität auf die Leistung von Schüler*innen existieren nur wenige theoretische Konzepte (siehe Kapitel 7.2; Anhang B). Es kann somit konstatiert werden, dass zwar angenommen wird, dass Heterogenität einen Effekt auf die Schüler*innenleistung hat – vor welchem Hintergrund sich dieser Effekt jedoch beschreiben lässt, bleibt weiterhin offen. Diese Frage kann in der vorliegenden Arbeit nicht empirisch geklärt werden; trotz allem ist es für die spätere Einordnung der Ergebnisse von Bedeutung (v.a. Kapitel 7.1) und soll für die nachfolgenden Kapitel berücksichtigt werden. Im Gegenzug zu den direkten Effekten von Heterogenität auf die Schüler*innenleistung existieren für die indirekten Effekte – vor allem hinsichtlich des Umgangs mit Heterogenität durch die Lehrkraft – mehrere theoretische Konzepte. Im nachfolgenden Kapitel soll auf den Umgang mit Heterogenität eingegangen werden, da dieser für das in Kapitel 6 vorgestellte Arbeitsmodell eine Rolle spielt.

3.3 Umgang mit Heterogenität in Schule und Unterricht

Unterricht und das erfolgreiche Lernen von Schüler*innen stellt Anforderungen an Lehrkräfte; und zwar insbesondere die, dass durch Unterricht kognitive Prozesse bei den Schüler*innen angestoßen werden. Solche Anforderungen steigen bei Lehrkräften in Klassen mit zunehmender Heterogenität (z.B. wegen stärkerer Unterschiede in den Lernvoraussetzungen der Schüler*innen). Aus der Forschung ist bekannt, dass erfolgreiches Lernen am besten bei hoher Unterrichtsqualität gelingt (z.B. Evertson, Emmer & Worsham, 2006;

¹ Der Begriff *sprachliche Heterogenität* wird sowohl in der deutschen als auch in der angloamerikanischen Forschung selten verwendet; genutzt wird zumeist der Begriff *sprachliche Diversität* oder *Diversity* (Walgenbach, 2017). *Diversity* ist dem angloamerikanischen Sprachraum entnommen und bedeutet Vielfalt (Mecheril & Plöber, 2011). Wenning (2007) betont jedoch, dass *Diversität* als vergleichbarer Begriff zur *Heterogenität* verwendet werden kann. In der vorliegenden Arbeit wird aus Gründen der sprachlichen Einheit der Begriff *sprachliche Heterogenität* für die *sprachliche Diversität* genutzt.

Hattie, 2009; Klieme, Lipowsky, Rakoczy & Ratzka, 2006). Die Befunde der Unterrichtsqualitätsforschung beruhen jedoch in der Regel auf einem eher lehrkraftzentrierten Unterricht und weniger auf einem individualisierten Unterricht (Bohl, 2017). Deswegen müssen die Kriterien für Unterrichtsqualität an heterogene Kontexte angepasst werden. Demnach soll im Folgenden beschrieben werden, welche Kriterien zur Messung von Unterrichtsqualität herangezogen werden können und wie eine Anpassung dieser Kriterien an heterogene Klassen vollzogen werden kann.

Unterrichtsqualität kann nach Klieme et al. (2006) mit drei Basisdimensionen beschrieben werden. Diese setzen sich zusammen aus Klassenführung, konstruktiver Unterstützung und kognitiver Aktivierung (z.B. Klieme, 2006; Klieme & Rakoczy, 2008; Klieme, Schümer & Knoll, 2001). Klassenführung hat zum Ziel, die verfügbare Lernzeit zu maximieren und eine intensive Auseinandersetzung mit den Lerninhalten zu gewährleisten (z.B. durch die Etablierung von Regelsystemen, einen konsequenten Umgang mit Störungen; Kounin, 1976; Marzano & Marzano, 2003; Seidel & Shavelson, 2007). Konstruktive Unterstützung hat zum Ziel, die Schüler*innen bei Schwierigkeiten gezielt und im Rahmen einer positiven Lehrkraft-Schüler*in-Beziehung zu unterstützen (z.B. durch schüler*innenorientiertes Feedback, individuelle Bezugsnormorientierung, positive Fehlerkultur; Cornelius-White, 2007; Dickhäuser, Janke, Praetorius & Dresel, 2017; Kunter & Voss, 2011; Pianta & Hamre, 2009; Rheinberg, 1980). Kognitive Aktivierung hat zum Ziel, die Schüler*innen zu einer aktiven Auseinandersetzung mit den Lerninhalten anzuregen (z.B. durch kognitiv aktivierende Aufgaben und Gespräche; Klieme & Rakoczy, 2008; Kunter & Voss, 2011; Lipowsky, 2009). Die Basisdimensionen gelten als wichtiger Teil der Tiefenstrukturen des Unterrichts (c. q. im Vergleich zu Sichtstrukturen nicht direkt sichtbare Merkmale des Unterrichts), welche für das Lernen von Schüler*innen von zentraler Bedeutung sind (z.B. Kunter & Trautwein, 2013; Reusser, Stebler, Mandel & Eckstein, 2013).

Um sich dieser Tiefenstruktur anzunähern, stehen Lehrkräfte heterogener Klassen anderen Anforderungen gegenüber als Lehrkräfte weniger heterogenen Klassen (z.B. gewinnen präventive Maßnahmen und ein effektiver Umgang mit Störungen in heterogenen Kontexten an Bedeutung, da Störungen in heterogenen Klassen häufiger auftreten können; Bohl, 2017; Dotzel & Karst, im Druck; Anhang D). Dies wird im Modell von Pietsch (2010) deutlich, der Stufen von Unterrichtsqualität abgeleitet hat.

Pietsch (2010) entwickelte das Stufenmodell der Unterrichtsqualität vor dem Hintergrund der Evaluation von Unterrichtsstandards. Er kritisiert, dass in Berichten zur Schul- und Unterrichtsinspektion Merkmale eines Kriterienkataloges nebeneinander gestellt werden oder spezifische Merkmale normativ gewichtet werden; und dies zumeist ohne zugrundeliegende empirische Evidenz (vgl. Meyer, 2006). Nach Pietsch (2010) kann die Darstellung

isolierter, einzelner Merkmale von Unterrichtsqualität aus theoretischer Perspektive jedoch als problematisch betrachtet werden; denn es existiert hinreichend Evidenz, dass einzelne Unterrichtsmerkmale in der Regel zusammenhängen und nicht isoliert betrachtet werden sollten (vgl. Brophy, 2000; Fraser, Walberg, Welch & Hattie, 1987; Helmke, 2003, 2006; Meyer, 2004; Seidel & Shavelson, 2007). Vor diesem Hintergrund entwickelte Pietsch (2010) ein abgestuftes Modell der Unterrichtsqualität.

Nach Bohl (2017) ist das Stufenmodell für Lehrkrafthandeln in heterogenen Klassen vor allem deswegen von Bedeutung, weil es den professionellen Aspekt des Umgangs mit Heterogenität in den Fokus rückt. So zeigt das Modell auf, dass eine erfolgreiche *Differenzierung* erst möglich ist, wenn die vorherigen drei Stufen *Lernklima und pädagogische Strukturen sichern*, *Klassen effizient führen und Methoden variieren* und *Schüler motivieren, aktives Lernen und Wissenstransfer ermöglichen* gesichert sind (Bohl, 2017). In Pietschs Stufenmodell (2010) werden also die Basisdimensionen guten Unterrichts adaptiert und erweitert, um gelingenden Unterricht in vier aufeinander aufbauende Stufen zu untergliedern (siehe Abbildung 2). Die folgenden Ausführungen beruhen auf der Modellbeschreibung von Pietsch (2010, S. 140).

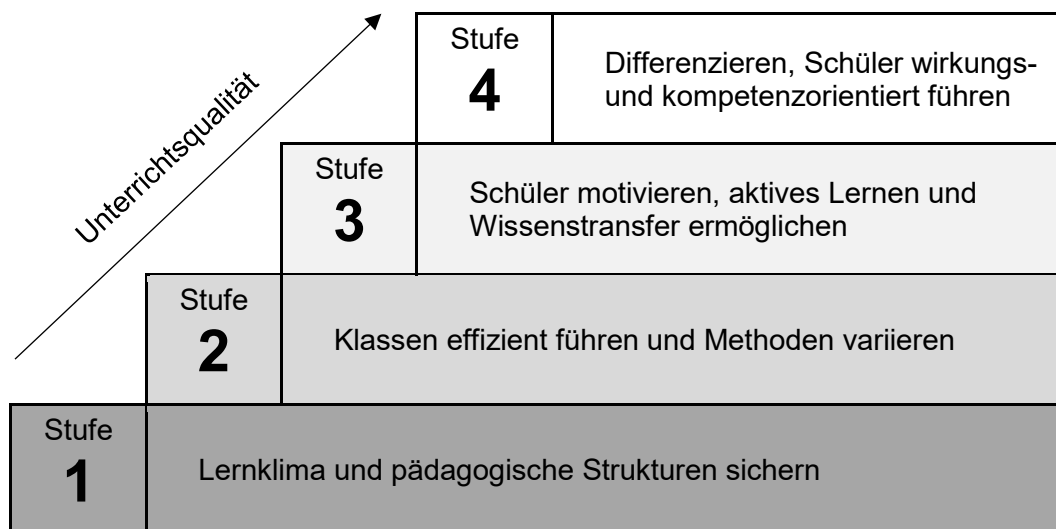


Abbildung 2: Stufenmodell der Unterrichtsqualität nach Pietsch (2010) (vgl. Bohl, 2017)

Auf Stufe 1 *Lernklima und pädagogische Strukturen sichern* wird die Grundlage für ein erfolgreiches und anspruchsvolles Lernen gelegt. Auf diesem Niveau werden Arbeitsauf-

träge von der Lehrkraft vorwiegend klar, präzise und an der unterrichteten Schüler*innen-schaft orientiert formuliert. Zudem wird ein lernförderliches Unterrichtsklima gesichert, indem erste, einfache Strukturen, Regeln und Rituale implementiert werden. Diese werden durch die Lehrkraft bei Bedarf an die jeweilige Situation adaptiert. Besonders in heterogenen Klassen kann die Art und Weise, wie Schüler*innen Regeln einhalten und Arbeitsverhalten verstehen, sehr unterschiedlich sein. Deswegen ist es wichtig, dass Lehrkräfte ihre Schüler*innen an die vereinbarten Regeln und Rituale erinnern, bis diese für die Schüler*innen zu einer Selbstverständlichkeit werden (Brophy, 1986). Die Verdeutlichung der Regeln mit detaillierter Anweisung und Modellierung reduziert Störungen im Unterricht und gibt den Schüler*innen eine klare Anleitung (z.B. Emmer, Evertson & Worsham, 2006; Evertson & Emmer, 1982; McGinnis, Frederick & Edwards, 1995; Van de Grift, 2007), was vor allem für den professionellen Umgang mit Heterogenität unerlässlich ist.

Stufe 2 *Klassen effizient führen und Methoden variieren* geht über die Strukturen auf Stufe 1 hinaus und stellt zusätzlich eine effiziente Klassenführung (z.B. vorausschauendes Handeln der Lehrkräfte, Maximierung der aktiven Lernzeit durch einen effektiven Umgang mit Störungen) und eine Variation von Methoden bereit. Auf dieser Stufe gelingt es Lehrkräften, unangemessenes Verhalten schnell zu unterbinden; und dies auf eine Art und Weise in den Unterricht einzubauen, ohne den Unterrichtsfluss zu unterbrechen (z.B. Anderson, Evertson & Brophy, 1979; Evertson, Emmer & Worsham, 2006; Pedota, 2007). Zudem kann auf Niveaustufe 2 das Unterrichtstempo an die Bedürfnisse der Schüler*innen und die Gegebenheiten der jeweiligen Lernsituationen angepasst werden. Niveaustufe 2 legt somit das Fundament für eine gelingende Individualisierung und Differenzierung (vgl. Stufe 4), welche für den professionellen Umgang mit Heterogenität von Bedeutung sind.

Auf Stufe 3 *Schüler motivieren, aktives Lernen und Wissenstransfer ermöglichen* werden Schüler*innen durch die Lehrkraft auf vielfältige Art und Weise zum Lernen motiviert. So arbeitet die Lehrkraft schüler*innenorientiert und lässt Schüler*innen das Unterrichtsgeschehen in angemessenem Rahmen mitbestimmen (z.B. McCombs & Whisler, 1997; Kunter & Voss, 2011; Pianta & Hamre, 2009). Auf Niveaustufe 3 gelingt es Lehrkräften aktives und selbstständiges Lernen der Schüler*innen anzuregen. In heterogenen Klassen sind Leistungsunterschiede stärker ausgeprägt und können von den Schüler*innen deswegen auch stärker wahrgenommen werden. Zudem können in leistungsheterogeneren Klassen Verständnisprobleme, Missverständnisse und falsche Antworten häufiger auftreten. Dies kann bei den betroffenen Schüler*innen ein Gefühl des Unbehagens auslösen und emotionales Unwohlsein verstärken (Kunter & Trautwein, 2013). Auf Niveaustufe 3 gelingt es Lehrkräften heterogener Klassen durch schüler*innenorientiertes Handeln den Fokus der Schüler*innen vom sozialen Vergleich auf ihr eigenes Lernwachstum zu verlagern (z.B.

durch eine individuelle Bezugsnormorientierung; Dickhäuser et al., 2017; Rheinberg, 1980). Zudem gelingt es Lehrkräften, die in heterogenen Klassen häufiger auftretenden Fehler zu integrieren (z.B. durch eine positive Fehlerkultur; Keith & Frese, 2008; Nordstrom, Wendland & Williams, 1998; Rach, Ufer & Heinze, 2013), indem diese Fehler zur Erläuterung von Gedankengängen und zur Anregung von Diskussionen genutzt werden (Tulis, 2013).

Lehrkräfte, welche auf der höchsten Stufe 4 *Differenzieren, Schüler wirkungs- und kompetenzorientiert führen* unterrichten, gelingt es, die Inhalte der ersten drei Niveaustufen angemessen zu implementieren und überdies die hohen Anforderungen bezüglich Schüler*innenorientierung, Binnendifferenzierung und Individualisierung nachzukommen. Dies spielt vor allem in heterogenen Klassen eine Rolle; denn lehrkraftzentrierter Unterricht kann in heterogenen Klassen zu einem höheren Maß an Unter- und Überforderung, Langeweile und Monotonie bei Schüler*innen führen (vgl. Larson & Richards, 1991; Perkins & Hill, 1985). Durch die Inhalte von Niveaustufe 4 können Lehrkräfte jede*n einzelne*n Schüler*in kognitiv aktivieren und erzielen dadurch einen nachhaltigen Kompetenzerwerb (z.B. durch Diskussionen und Gruppenarbeit zwischen Schüler*innen verschiedener Fähigkeitsstufen, Differenzierung der Einzelarbeit nach Schwierigkeitsgraden; Prediger & von Aufschnaiter, 2017). Zudem ziehen Lehrkräfte auf Niveaustufe 4 schüler*innenorientierte Diagnose- und Feedbackverfahren heran und unterstützen die Schüler*innen bei der Reflexion des eigenen Lernens, der eigenen Lernprozesse und der eigenen Lernentwicklung.

Auf Basis des Stufenmodells der Unterrichtsqualität nach Pietsch (2010) kann professionelles Handeln in heterogenen Klassen mit den Basisdimensionen gelingenden Unterrichts verknüpft werden. Dabei befinden sich Merkmale der Klassenführung vor allem auf Stufe 1 und Stufe 2, Merkmale der konstruktiven Unterstützung vor allem auf Stufe 3 und Merkmale der kognitiven Aktivierung weitgehend auf Stufe 4 – welche um Differenzierungsmaßnahmen erweitert werden. Durch die stufenförmige Anordnung unterstreicht Pietsch (2010), welche Ansprüche und Voraussetzungen mit differenziertem Unterricht verbunden sind und welche Komplexität mit dem professionellen Umgang mit Heterogenität einhergeht.

3.4 Zusammenfassung der theoretischen Einordnung

Vor dem Hintergrund der theoretischen Einordnung kann festgehalten werden, dass Heterogenität vor allem aus einer Perspektive leistungsbezogener und sprachlicher Unterschiede eine Rolle spielt. Zudem lassen Modelle zur Erklärung der Schüler*innenleistung

darauf schließen, dass Heterogenität für die Leistung von Schüler*innen bedeutsam ist. So ist Heterogenität mit vielen Faktoren der Unterrichtsumwelt verstrickt und hat dadurch nicht nur eine direkte sondern auch eine indirekte Wirkung auf die Leistung von Schüler*innen (siehe Kapitel 2: Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts nach Helmke, 2017). Als interessant zeigt sich an dieser Stelle, dass im Vergleich zu den indirekten Effekten nur wenige Erklärungsansätze für die direkten Effekte von Heterogenität auf die Schüler*innenleistung existieren. Für die indirekten Effekte wurde herausgearbeitet, dass ein konstruktiver Umgang mit Heterogenität (siehe Kapitel 3.3: Basisdimensionen guten Unterrichts nach Kunter & Trautwein, 2013; Stufenmodell der Unterrichtsqualität nach Pietsch, 2010) eine zentrale Komponente einnimmt. Deswegen findet dieser auch in der vorliegenden Arbeit Berücksichtigung.

4 Heterogenität und Schüler*innenleistung

Auf Basis des bisherigen Forschungsstandes kann konstatiert werden, dass der wissenschaftliche Schwerpunkt auf dem Thema *Leistungsheterogenität* liegt. Im Vergleich zu den Effekten der sprachlichen Heterogenität, zu welcher in der vorliegenden Arbeit die nach Recherche (Stand 10.09.2020) erste empirische Studie vorgestellt wird, werden die Effekte der Leistungsheterogenität auf die Schüler*innenleistung seit fast 100 Jahren erforscht.

Leistungsheterogenität spielt in einer Vielzahl von Disziplinen eine Schlüsselrolle. So untersuchten bereits Wissenschaftler*innen der Soziologie (z.B. Dreeben & Barr, 1988; Leiter, 1983), der Ökonomie (z.B. Burke & Sass, 2013; Carman & Zhang, 2012), der Politikwissenschaften (z.B. Vigdor & Neychbar, 2007) und der Bildungsforschung (z.B. Kuzmina & Ivanova, 2018; Lleras & Rangel, 2009) die Auswirkungen von Leistungsheterogenität auf die Leistung von Schüler*innen. Bemerkenswert ist, dass trotz dieses breiten Blickwinkels und der hohen Anzahl an Literatur der bisherige Forschungsstand kein klares Bild über die Effekte von Leistungsheterogenität auf die Schüler*innenleistung liefern konnte; und so berichten Studien über positive (z.B. Kuzmina & Ivanova, 2018; Vigdor & Neychbar, 2007; Zimmer & Toma, 2000), negative (z.B. Luyten & Van der Hoeven-van Doornum, 1995; Steenbergen-Hu, Makel & Olszewski-Kubilius, 2016) und Null-Effekte (z.B. Bellin, 2009; Carman & Zhang, 2012; Hattie, 2009).

Einige Studien, die Klassen der Elementar- und Sekundarstufe untersucht haben, konnten zeigen, dass sich Heterogenität positiv auf die Leistung von Schüler*innen auswirkt (z.B. Dar & Resh, 1986; Scharenberg, 2012; Vigdor & Neychbar, 2007). Zimmer und Toma (2000) untersuchten 16.560 Achtklässler*innen aus fünf Ländern und fanden einen positiven Effekt der Heterogenität auf das individuelle Lernen ($\beta = .23$). Ähnliche Effekte zeigten sich auch bei einer Studie von Kuzmina und Ivanova (2018), welche mit einer Stichprobe von 5.118 Erstklässler*innen berichten, dass sich Heterogenität positiv auf die Leistung der Schüler*innen im Leseverständnis auswirkt ($b = 2.81$). Eine weitere Studie untersuchte eine Stichprobe von etwa 700 Schüler*innen der 10. bis 12. Jahrgangsstufe aus 48 Klassen und berichtete über einen Leistungsvorteil in heterogenen Klassen für alle Schüler*innen ($b = .36$) (Dar & Resh, 1986). Im Gegensatz dazu haben mehrere Studien zeigen können, dass sich Heterogenität negativ auf die Leistung von Schüler*innen auswirkt. Luyten und Van der Hoeven-van Doornum (1995) untersuchten mit einer Stichprobe von 3.993 Schüler*innen den Effekt von Heterogenität beim IQ auf das Lernen und fanden einen negativen Effekt ($\beta = -.06$). Ähnliche Muster wurden auch von Meta-Analysen berichtet, welche die

Effekte der Fähigkeitsgruppierung (c. q. ability grouping) auf die Schüler*innenleistung untersuchten.² So berichteten Steenbergen-Hu, Makel und Olszewski-Kubilius (2016) in einer Meta-Analyse zweiter Ordnung mit qualitativ hochwertigen Studien über einen signifikanten positiven Effekt ($g = .15$) für Fähigkeitsgruppierung. Jahrzehnte zuvor hatten Kulik und Kulik (1982, 1984) bereits Ergebnisse vorgestellt, die den positiven Einfluss der Fähigkeitsgruppierung aufzeigen; sie berichteten von zwei Meta-Analysen, die eine mittlere Effektgröße von $ES = .19$ in der Grundschule und $ES = .10$ in der Sekundarschule ergaben. In einer weiteren Meta-Analyse wurden die Leistungseffekte einer homogenen mit einer heterogenen Fähigkeitsgruppierung verglichen (Lou, Abrami, Spence, Poulsen, Chambers, & D'Apollonia, 1996). Sie fanden eine mittlere Effektgröße von $ES = .12$, welche den positiven Einfluss von homogener Leistungsgruppierung unterstützt. Dennoch reichten die Effektgrößen von $-1.75 \leq ES \leq 1.12$ mit dreizehn Effektgrößen, die eine homogene Leistungsgruppierung begünstigten, eine Effektgröße gleich Null und sechs Effektgrößen, die eine heterogene Leistungsgruppierung begünstigten. Darüber hinaus berichten viele Studien Null-Befunde; Heterogenität nimmt hier keinen signifikanten Einfluss auf das Lernen der Schüler*innen (z.B. Bellin, 2009; Decristan, Fauth, Kunter, Büttner & Klieme, 2017; Leiter, 1983; Wright, Horn & Sanders, 1997). Carman und Zhang (2012) führten beispielsweise eine Analyse mit 923 Sekundarschüler*innen durch und stellten fest, dass die Heterogenität keinen Einfluss auf die Leistungen in Englisch, Chinesisch oder Mathematik hatte ($-.07 \leq \beta \leq .28$). Überdies führten Wright, Horn und Sanders (1997) eine Analyse mit Schüler*innen der dritten und vierten Klasse durch ($10477 \leq N \leq 14079$). Die Klassenheterogenität zeigte keinen Einfluss auf die Leseleistung der Schüler*innen ($.03 \leq z \leq 1.32$). Darüber hinaus fanden zwei Meta-Analysen zur Fähigkeitsgruppierung keine signifikanten Effekte in Grundschulen mit $ES = .00$, oder Sekundarschulen mit $ES = -.02$ (Slavin, 1987, 1990).

Es kann also festgehalten werden, dass die Effekte der Leistungsheterogenität auf die Schüler*innenleistung zwar bereits von vielen Wissenschaftler*innen untersucht wurden, der Forschungsstand jedoch trotzdem kein einheitliches Bild über diese Beziehung liefern kann. Ein Grund für den inkonsistenten Forschungsstand könnte in systematischen Unterschieden in den Stichproben liegen. Wie bereits erläutert, ist Heterogenität ein relatives Konstrukt, so dass die gleiche Klassenzusammensetzung je nach Merkmalen der Stichprobe einen unterschiedlichen Grad an Heterogenität aufweisen kann. So kann ein und dieselbe Klassenzusammensetzung in einer Stichprobe mit eher heterogenen Klassen homogen sein, während sie in einer Stichprobe mit eher homogenen Klassen heterogen sein kann.

² Ability Grouping ist eine Strategie zur Gestaltung der Heterogenität in der Klasse und in Schulen, bei der Schüler*innen nach deren Leistung selektiert werden (z.B. Huang, 2008; Slavin, 1987).

Systematischen Unterschiede in der Stichprobe können sich auch innerhalb eines Landes zeigen. In Deutschland zum Beispiel wechseln Schüler*innen nach der Grundschule je nach Leistungsniveau auf eine Schulform der Sekundarstufe (z.B. Gymnasium = höchste Schulform, Realschule = mittlere Schulform, Werkrealschule/Hauptschule = niedrigste Schulform).³ Dadurch können Stichproben mit Grundschulklassen eher heterogene Klassen und Stichproben mit Sekundarschulklassen eher homogene Klassen beinhalten. Außerdem können Stichproben mit Grundschulklassen durch die noch nicht vollzogene Leistungsselektion eine größere Heterogenitätsspanne aufweisen als Stichproben mit Sekundarschulklassen. Solche Unterschiede in den Stichproben führen zu der Überlegung, dass sich Heterogenität nicht unbedingt linear auf die Schüler*innenleistung auswirkt, sondern auch nichtlineare Zusammenhänge denkbar sind. Je nachdem, ob eine Stichprobe heterogene oder homogene Klassen aufweist, können positive oder negative Zusammenhänge gefunden werden; und in Stichproben mit einer großen Bandbreite an Heterogenität (z.B. Grundschulklassen, nicht-selektive Sekundarschulformen wie in England, den USA oder Australien; Carl, 2017; Dupriez, Dumay, & Vause, 2008; Postlethwaite, 19988) könnten nichtlineare Beziehungen vorliegen.

Die Erklärung der unterschiedlichen Befunde wird zudem durch die Tatsache erschwert, dass es keine theoretischen Annahmen darüber gibt, wie genau Heterogenität die Schüler*innenleistung beeinflusst. Wie bereits in Kapitel 3.2 erläutert, wird in der Theorie davon ausgegangen, dass Heterogenität die Schüler*innenleistung sowohl direkt als auch indirekt beeinflusst (Helmke, 2017). Während jedoch die indirekten Effekte auf theoretischen Annahmen beruhen (z.B. Lehrkraftverhalten), gibt es keine klaren Überlegungen, wie der direkte Effekt der Heterogenität auf die Schüler*innenleistung erklärt werden kann. Es bleibt also offen, ob Heterogenität positive oder negative direkte Effekte auf das Lernen der Schüler*innen hat und ob dieser Zusammenhang linear oder nichtlinear ist.

Auch der Effekt der sprachlichen Heterogenität auf die Leistung ist noch nicht ausreichend geklärt. Dies liegt jedoch nicht an widersprüchlichen Ergebnissen bisheriger Studien, sondern kann darauf zurückgeführt werden, dass die sprachliche Heterogenität und deren Effekte auf die Schüler*innenleistung bisher noch nicht empirisch untersucht wurden (Recherche Stand 09.09.2020). Die sprachliche Heterogenität ist nicht nur aus dem Grund des gesellschaftlichen und schulpraktischen Fokus von Bedeutung (siehe Kapitel 3.2) – ferner

³ Trotz der in Deutschland durchgeführten Homogenisierungsmaßnahmen, ist an dieser Stelle anzumerken, dass auch in Deutschland zunehmend schulstrukturelle Veränderungen der Bundesländer zu beobachten sind, die darauf hinweisen, dass Bildungsgänge flexibilisiert werden (z.B. Wahlmöglichkeit zwischen acht- und neunjährigem Gymnasium) und das gemeinsame Lernen aller Schüler*innen fördern (z.B. integrative Schulsysteme) (van Ackeren & Kühn, 2017).

kann davon ausgegangen werden, dass die sprachliche Heterogenität die Schüler*innenleistung positiv beeinflusst.

Hinsichtlich der Effekte der sprachlichen Heterogenität auf die Schüler*innenleistung lässt sich an der Forschung zum Thema *Ethnische Heterogenität* anknüpfen, denn die ethnische Komposition steht vermutlich in starkem Zusammenhang mit der sprachlichen Zusammensetzung der Klasse (Rjosk, Richter, Lüdtke & Eccles, 2017). An dieser Stelle muss jedoch angemerkt werden, dass auch zum Thema *Ethnische Heterogenität und deren Effekte auf die Schüler*innenleistung* nur wenige Studien vorhanden sind. Zudem untersuchen diese Studien nicht die klassenspezifische, sondern die schulspezifische Zusammensetzung. Trotz der geringen Anzahl an Studien berichten die Wissenschaftler*innen jedoch ähnliche Ergebnisse.

Demnach untersuchten Benner und Crosnoe (2011) die Leistung von 21.260 Grundschüler*innen und fanden heraus, dass Grundschüler*innen aus ethnisch heterogeneren Klassen eine höhere Leistung sowohl im Lesen ($\beta = .05$) als auch in Mathematik ($\beta = .04$) erzielten. Ähnliche Ergebnisse wurden auch von Rjosk et al. (2017) berichtet. Die Wissenschaftler*innen untersuchten 18.762 Grundschüler*innen aus 903 Schulen und fanden kleine positive Effekte der ethnischen Heterogenität auf die Mathematikleistung der Schüler*innen ($.03 < \beta < .04$), jedoch keine signifikanten Effekte der ethnischen Heterogenität auf die Leseleistung der Schüler*innen ($\beta = .00$). Eine weitere Studie, welche mit 1.661 Sekundarschüler*innen durchgeführt wurde, berichtete mit $\beta = .23$ positive Effekte der ethnischen Heterogenität auf den GPA der Schüler*innen im ersten Hochschulse semester (Tam & Bassett, 2004).

Neben den Effekten der sprachlichen Heterogenität auf die Schüler*innenleistung wurde von bisheriger Forschung auch noch nicht in den Blick genommen, welche Einstellungen Schüler*innen zur Leistungsheterogenität haben (Recherche Stand 09.09.2020). Bisher wurde das Thema *Einstellung zur Heterogenität/ Leistungsheterogenität* vor allem in lehrkraftzentrierter Forschung in den Blick genommen (z.B. Feyerer, 2014; Hartinger, Grittnner, Lang & Rehle, 2010; Ruhberg & Porsch, 2017). Studien unterstreichen jedoch, wie wichtig die individuellen Einstellungen für die Schüler*innenleistung sind (z.B. Bloom, 1976; Papanastasiou & Zembylas, 2002). Während die Einstellungen zum Schulfach und zu den eigenen Fähigkeiten die Leistung nachweislich beeinflussen (z.B. Else-Quest, Mineo & Higgins, 2013; Martinez, Aricak & Jewell, 2008), müssen die Einstellungen zum schulischen Umfeld – wie die Einstellung zur Leistungsheterogenität von Schüler*innen – noch untersucht werden.

5 Forschungslücken und Ziele der vorliegenden Arbeit

Vor diesem Hintergrund können folgende Forschungslücken abgeleitet werden: Erstens, welchen Einfluss hat die Leistungsheterogenität auf die Leistung von Schüler*innen und können nicht-lineare Analysen mit verschiedenen Heterogenitätsmaßen zusätzlichen Informationen liefern? Zweitens, welchen Einfluss hat die sprachliche Heterogenität auf die Leistung von Schüler*innen? Drittens, wie ist die Einstellung zur Leistungsheterogenität bei Schüler*innen ausgeprägt, welche Kontextmerkmale können sie beeinflussen und hat diese Einstellung von Schüler*innen auch einen Effekt auf deren Leistung?

Ziel dieser Arbeit ist es, diese Forschungslücken zu verringern und gegebenenfalls zu schließen. Um dieses Ziel zu erreichen, wird im Folgenden das Arbeitsmodell der vorliegenden Arbeit abgeleitet. Im Rahmen dieses nachfolgenden Kapitels werden drei Perspektiven vorgestellt, welche die hier genannten Forschungslücken aufgreifen und den wissenschaftlichen Diskurs und den Forschungsstand zum Thema *Leistungsbezogene und sprachliche Heterogenität in der Schule und deren erwartete Effekte auf die Leistung von Schüler*innen* erweitern sollen.

6 Ableitung des Arbeitsmodells und dazugehöriger Fragestellungen

In diesem Kapitel wird das Arbeitsmodell für die vorliegende Dissertation vorgestellt (siehe Abbildung 3). Ziel des Arbeitsmodells ist es, eine verknüpfende und überblicksartige Darstellung der theoretischen Einordnung (Kapitel 2 und 3) und der vorhandenen Forschungslücken (Kapitel 5) zu gewährleisten. Demnach werden zuallererst die Struktur und die Zusammensetzung des Arbeitsmodells erläutert. Daran anschließend werden in den drei Subkapiteln 6.1, 6.2 und 6.3 die abgeleiteten Forschungslücken aufgegriffen und konkrete Fragestellungen abgeleitet.

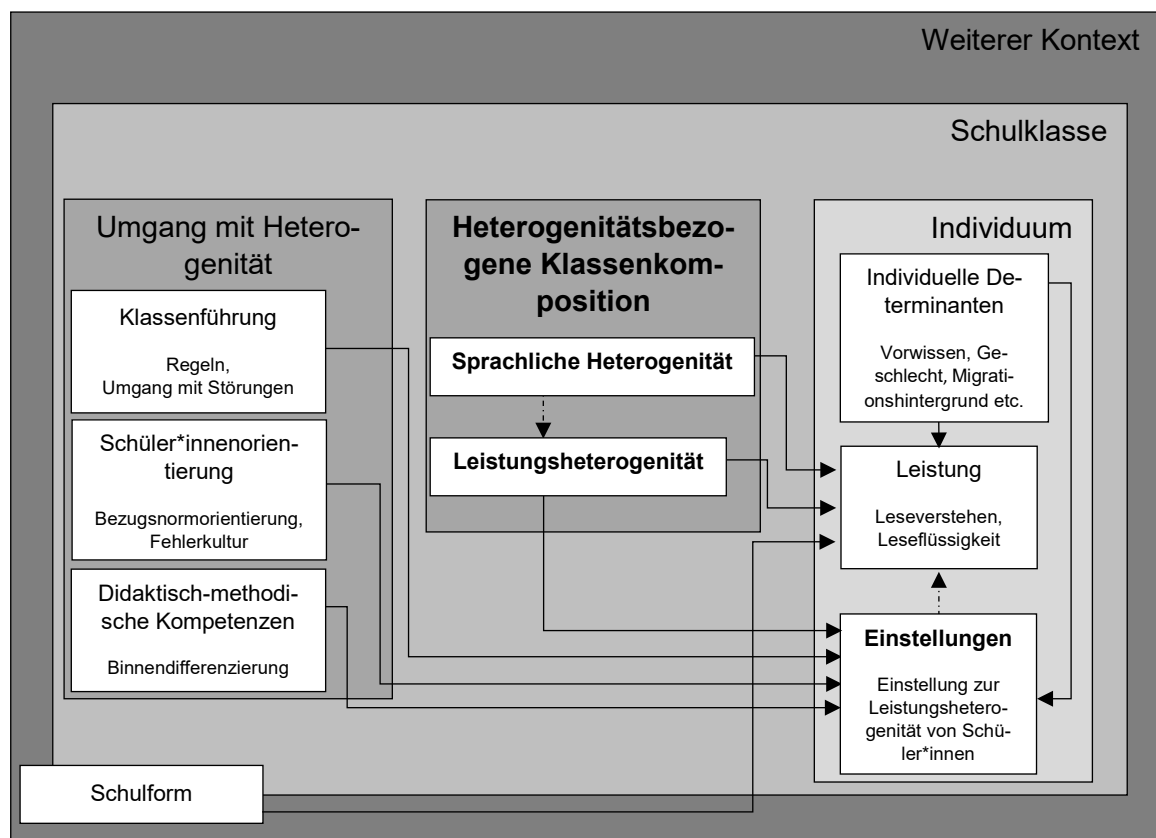


Abbildung 3: Arbeitsmodell der vorliegenden Arbeit

Das Arbeitsmodell weist insgesamt drei Ebenen auf: die des erweiterten Kontextes (z.B. regionaler, familiärer Kontext), die der Schulklasse und die des Individuums. Zudem weist das Arbeitsmodell drei Schwerpunkte auf: die sprachliche Heterogenität, die Leistungsheterogenität und die Einstellungen (im Arbeitsmodell deswegen fettgedruckt).

Aus Helmkes Modell (2017) werden die Schulform und die Klassenzusammensetzung ins Arbeitsmodell übernommen. Die Klassenzusammensetzung berücksichtigt Merkmale der *Heterogenitätsbezogenen Klassenkomposition* mit den Subkomponenten *Sprachliche Heterogenität* und *Leistungsheterogenität*. Ebenfalls aus Helmkes Modell (2017) adaptiert wurde der Aspekt *Individuelle Determinanten*. Hierunter subsummieren sich sowohl Merkmale des individuellen Lernpotenzials (z.B. Vorwissen), Merkmale der Familie (z.B. Migrationshintergrund, Sprache), als auch individuelle Schüler*innenmerkmale (z.B. Geschlecht, Alter). Die in Helmkes Angebots-Nutzungs-Modell (2017) dargestellten Wirkungen werden ebenfalls ins Arbeitsmodell übernommen. Diese werden in zwei Arten aufgliedert. Zum einem werden die fachlichen Kompetenzen durch die *Leistung* im Kompetenzbereich Lesen abgedeckt. Zum anderen werden die in Helmkes Modell (2017) dargestellten Wirkungen durch die *Einstellungen* von Schüler*innen untersucht. Hier wird der Blick auf die Einstellung zur Leistungsheterogenität von Schüler*innen gerichtet. Mit Hilfe von Pietschs Stufenmodell (2010) wird der Bereich *Umgang mit Heterogenität* erfasst. In diesem Bereich wird das professionelle Handeln einer Lehrkraft in heterogenen Klassen in die drei Subkomponenten *Klassenführung*, *Schüler*innenorientierung* und *didaktisch-methodische Kompetenzen* untergliedert. Dadurch sind alle vier Niveaustufen nach Pietsch (2010) vertreten (siehe Kapitel 3.3.).

Im Arbeitsmodell sind sowohl durchgezogene als auch gestrichelte Pfeile vorzufinden. Dabei stellen durchgezogene Pfeile zwischen den Subkomponenten Effekte dar, die in den jeweiligen Beiträgen tatsächlich untersucht und berichtet werden. Dagegen stellen gestrichelte Pfeile zwischen den Subkomponenten diejenigen Effekte dar, die auf Basis der bisherigen theoretischen Einordnung von Bedeutung sind, in der vorliegenden Arbeit jedoch nicht empirisch überprüft werden können. Vor dem Hintergrund dieses Arbeitsmodells werden in den nachfolgenden Subkapiteln die in Kapitel 5 aufgezeigten Forschungslücken aufgegriffen und dazu kompatible Forschungsfragen abgeleitet. In Kapitel 6.1 wird das Thema *Leistungsheterogenität und Schüler*innenleistung* in den Blick genommen, Kapitel 6.2 greift das Thema *Sprachliche Heterogenität und Schüler*innenleistung* auf und Kapitel 6.3 fokussiert die *Einstellung zur Leistungsheterogenität bei Schüler*innen*. Aufgrund der überblicksartigen Darstellung nimmt das Arbeitsmodell nicht in den Blick, welche über die im Arbeitsmodell genannten individuellen Kontrollvariablen hinaus (c. q. Vorwissen, Geschlecht, Migrationshintergrund) im jeweiligen Beitrag genutzt werden. Überdies nimmt das Arbeitsmodell nicht in den Blick, welche konkreten kontextuellen Faktoren in den jeweiligen Beiträgen als Kontrollvariablen mitaufgenommen werden. Dafür wird an dieser Stelle auf die einzelnen Manuskripte verwiesen (siehe Anhang A, Anhang B, Anhang C).

6.1 Leistungsheterogenitätsmaße

Im Arbeitsmodell befindet sich die Leistungsheterogenität als Teil der *Heterogenitätsbezogenen Klassenkomposition*. In diesem Kapitel wird in den Blick genommen, welchen Einfluss die *Leistungsheterogenität* auf die *Leistung* von Schüler*innen nimmt. Im Arbeitsmodell liegen nur diejenigen Variablen mit weißem Hintergrund vor, die auch für die konkrete Fragestellung von Bedeutung sind (siehe Abbildung 4).

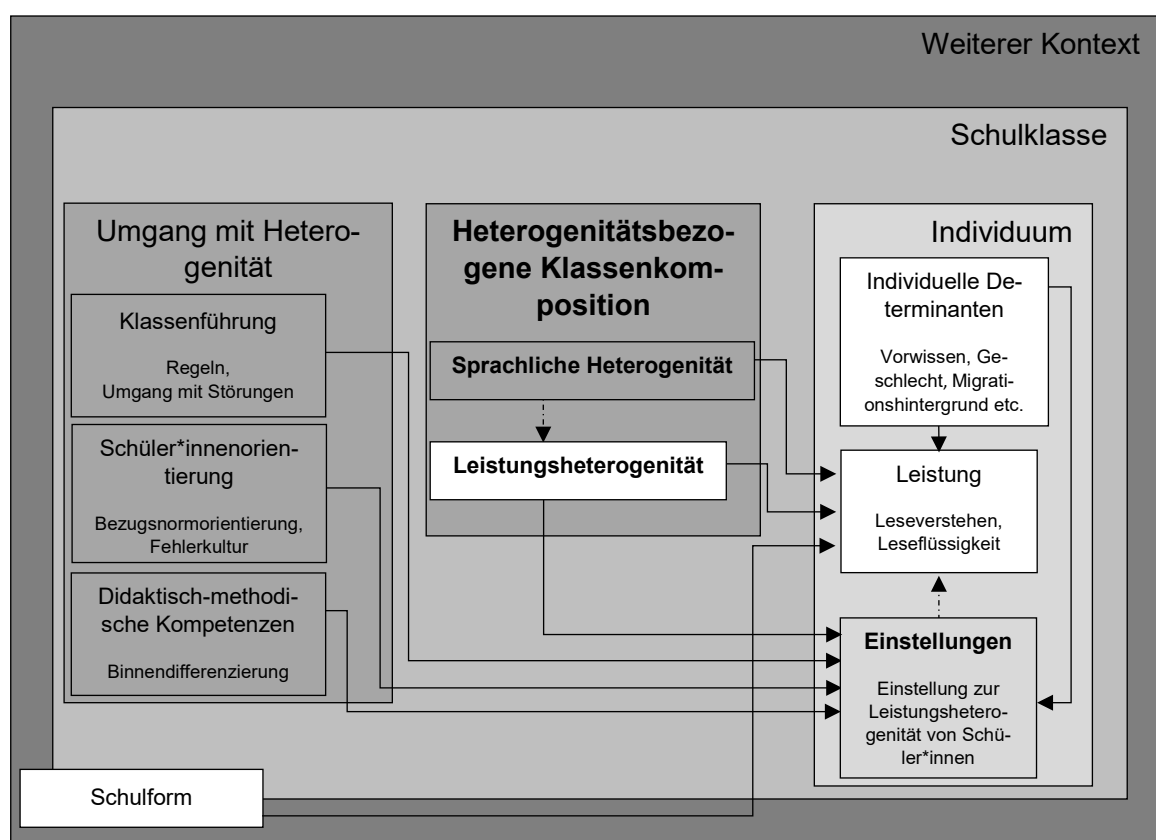


Abbildung 4: Arbeitsmodell Leistungsheterogenitätsmaße

Wie bereits in Kapitel 4 erläutert, weist der Forschungsstand unterschiedliche Ergebnisse darüber auf, wie Leistungsheterogenität die Schüler*innenleistung beeinflusst. Ziel von Beitrag 1 ist es deswegen Ansätze bereit zu stellen, die die Effekte von Leistungsheterogenität auf die Schüler*innenleistung aus verschiedenen Perspektiven betrachten. Diese Ansätze haben keine theoretische, sondern eine datengetriebene, explorative Fundierung. Bisher haben Studien für die Berechnung der Leistungsheterogenität die klasseninterne Standardabweichung verwendet und nur über lineare Analysen berichtet (mit Ausnahme von Scharenberg, 2012); Analysen mit verschiedenen Heterogenitätsmaßen und Analysen

nichtlinearer Effekte wurden bisher nicht durchgeführt. So sollen in diesem Beitrag zum einen potenzielle nichtlineare Beziehungen zwischen Leistungsheterogenität und der Schüler*innenleistung untersucht werden. Demnach soll getestet werden, ob lineare Analysen ausreichen, um Heterogenitätseffekte zu beschreiben, oder ob nichtlineare Analysen weitere Informationen liefern können. Die Verwendung nichtlinearer Analysen kann vor dem Hintergrund Helmkes Angebots-Nutzungs-Modell (2017) gewinnbringend sein. Da keine theoretischen Annahmen über die direkten Effekte von Leistungsheterogenität auf die Schüler*innenleistung bestehen, kann angenommen werden, dass direkte lineare Effekte ähnlich wahrscheinlich sind als direkte nichtlineare Effekte.

Zum anderen soll untersucht werden, ob und wie die verschiedenen Heterogenitätsmaße mit der Leistung von Schüler*innen zusammenhängen. Demnach soll getestet werden, ob die Standardabweichung ausreicht, um Heterogenitätseffekte zu beschreiben, oder ob zusätzliche Heterogenitätsmaße weitere Informationen liefern können. Die Verwendung unterschiedlicher Heterogenitätsmaße kann sowohl aus methodischer Sicht als auch aus inhaltlich-pädagogischer Sicht gewinnbringend sein. Aus methodischer Sicht ist die Verwendung verschiedener Heterogenitätsmaße sinnvoll, da die einzelnen Maße unterschiedliche methodische Stärken und Schwächen aufweisen (z.B. die Abhängigkeit von Ausreißern). Aus inhaltlich-pädagogischer Sicht ist die Analyse verschiedener Heterogenitätsmaße sinnvoll, weil die einzelnen Maße eine jeweils spezifische Perspektive einnehmen, wie Leistungsheterogenität die Schüler*innenleistung beeinflusst. Betrachtet man die Standardabweichung (SD) als Heterogenitätsmaß, so trägt die gesamte Schüler*innenschaft zur Heterogenität in der Klasse bei. Somit wird mit der SD ein Aspekt von Heterogenität fokussiert, der auf allen Schüler*innen innerhalb der Klasse basiert. Wenn man die SD als Heterogenitätsmaß interpretiert, würde dies bedeuten, dass jede/r Schüler*in einen ähnlichen Einfluss darauf hat, wie die Leistungsheterogenität die Schüler*innenleistung beeinflusst. Im Vergleich dazu fokussiert Heterogenität gemessen am Range die Leistungsränder der Klasse. Auf Basis des Ranges ist Heterogenität die Differenz zwischen dem*der leistungstärksten und der leistungsschwächsten Schüler*in; und auf Basis dessen wird davon ausgegangen, dass die Leistungsränder einer Klasse einen so starken Einfluss auf die Leistungsheterogenität haben, sodass das Leistungszentrum relativ bedeutungslos wird. Wenn man also den Range als Heterogenitätsmaß interpretiert, würde dies bedeuten, dass der Einfluss der Heterogenität auf die Schüler*innenleistung von den beiden Schüler*innen abhängt, die die leistungsbezogenen Ränder in der Klasse definieren. Im Gegensatz zum Range legt das Heterogenitätsmaß Interquartilsabstand (IQR) einen besonderen Schwerpunkt auf die Leistungsmittel einer Klasse. Gemäß dem IQR misst Heterogenität die Unterschiedlichkeit zwischen den durchschnittlich leistungsfähigen Schüler*innen; und er geht

davon aus, dass der Unterschied in der Leistungsmittel einen so starken Einfluss auf die Klassenheterogenität hat, dass die Ränder einer Klasse ziemlich bedeutungslos werden. Wenn man also den IQR als Heterogenitätsmaß interpretiert, würde dies bedeuten, dass der Einfluss der Heterogenität auf die Schüler*innenleistung von den Unterschieden im Leistungszentrum abhängt. Jedes der genutzten Heterogenitätsmaße hat somit seinen spezifischen methodischen wie auch inhaltlichen Fokus, wie Heterogenität charakterisiert ist und wie Heterogenität auf die Schüler*innenleistung einwirkt.

6.2 Sprachliche Heterogenität

Als nächstes wird das Untersuchungsfeld *sprachliche Heterogenität und Leistung* in den Blick genommen. Im Arbeitsmodell befindet sich die sprachliche Heterogenität als Teil der *Heterogenitätsbezogenen Klassenkomposition*. An dieser Stelle wird in den Blick genommen, welchen Einfluss die *sprachliche Heterogenität* auf die *Leistung* von Schüler*innen nimmt. Im Arbeitsmodell liegen nur diejenigen Variablen mit weißem Hintergrund vor, die auch für die konkrete Fragestellung von Bedeutung sind (siehe Abbildung 5).

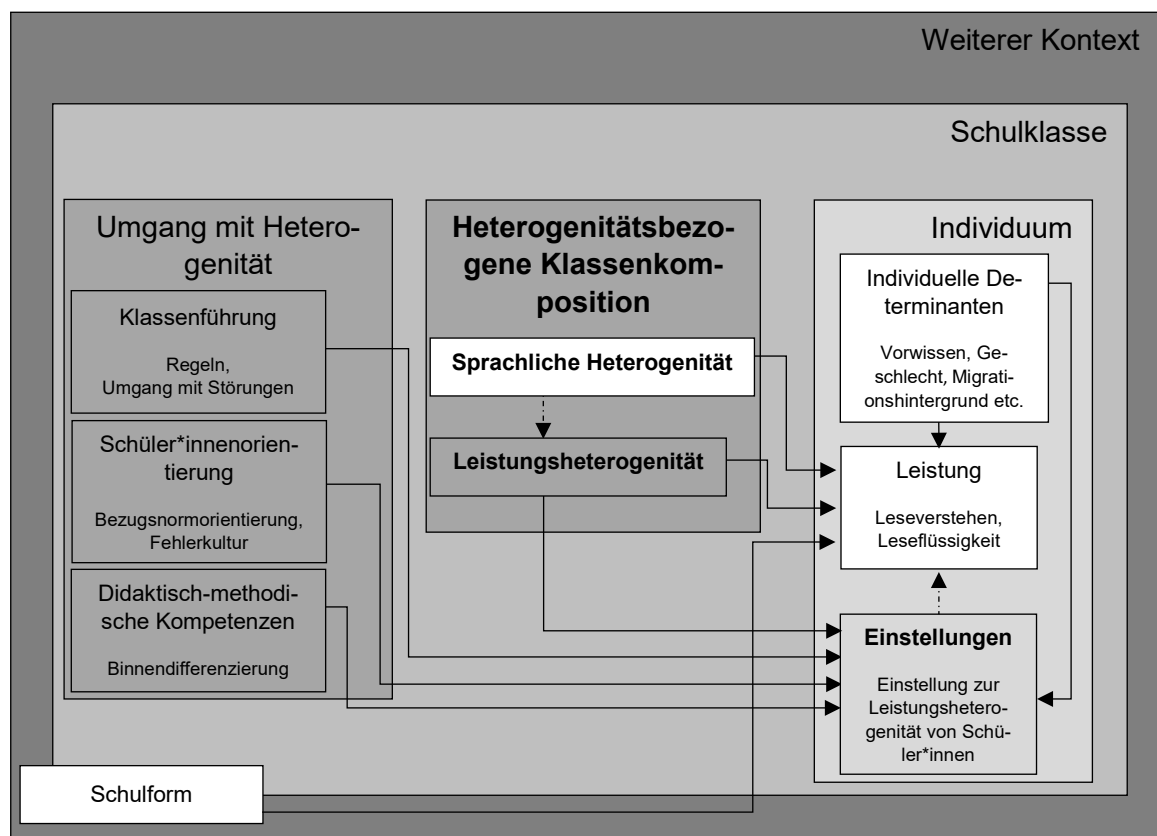


Abbildung 5: Arbeitsmodell Sprachliche Heterogenität

Dass der sprachliche Hintergrund von Schüler*innen ein wichtiger Prädiktor für die Leistung und Leistungsentwicklung von Schüler*innen ist, wurde bereits mehrfach nachgewiesen. Demnach zeigen Studien, dass die Schüler*innenleistung steigt, wenn Schüler*innen die Unterrichtssprache sprechen (z.B. Bellin, 2009; Esser, 2006; Hopf, 2005; Segeritz, Walter & Stanat, 2010) – und je qualitativ hochwertiger der sprachliche Austausch zwischen Schüler*innen ist, desto kompetenter werden Schüler*innen in der Unterrichtssprache und desto höher sind die schulischen Leistungen (siehe Basal-interpersonale Kommunikationsfähigkeiten BICS vs. Kognitiv-akademische Sprachfähigkeiten CALP; Cummins, 1979, 1981). Der sprachliche Hintergrund von Schüler*innen beeinflusst aber auch Merkmale der Klassenzusammensetzung, wie beispielsweise die sprachliche Heterogenität. In Deutschland war und ist sprachliche Heterogenität der Bevölkerung der Normalfall, wird aber als Ausnahme betrachtet (Wenning, 2004). In deutschen Schulen besteht mit der Unterrichtssprache Deutsch ein Anspruch auf sprachliche Gleichheit, wodurch die sprachliche Heterogenität in den Klassen vernachlässigt wird (Lutz & Wenning, 2001). Vor diesem Hintergrund unterstreicht Wenning (2004), dass Fragen des Umgangs mit Mehrsprachigkeit und sprachlicher Heterogenität stärker berücksichtigt werden müssen. Deswegen ist es das übergeordnete Ziel von Beitrag 2, die Effekte der sprachlichen Heterogenität auf die Leistung von Schüler*innen zu untersuchen.

Wie bereits in Kapitel 3.1 erläutert, kann sprachliche Heterogenität definiert werden als die Anzahl der verschiedenen Sprachen in einem bestimmten geografischen Gebiet (Nettle, 1999). Aus empirischer Sicht ist es jedoch unzureichend, die sprachliche Vielfalt einer Klasse nur anhand der Anzahl der Sprachen zu messen. Deswegen werden auch in diesem Beitrag mehrere Heterogenitätsmaße genutzt, um die sprachliche Heterogenität in der Klasse zu operationalisieren: Neben der Anzahl an Sprachen werden auch die Maße Simpson's D (Simpson, 1949) und Shannon's H (Shannon & Weaver, 1949) berechnet (vgl. Benner & Crosnoe, 2011; Rjosk et al., 2017).

Vor diesem Hintergrund soll in den Blick genommen werden, wie die sprachliche Heterogenität die Schüler*innenleistung beeinflusst und ob verschiedene Heterogenitätsmaße zusätzliche Informationen liefern können. Auf Basis von Piagets Konzept des Ungleichgewichts (1977) kann angenommen werden, dass mit zunehmender sprachlicher Heterogenität auch die Leistung der Schüler*innen steigt. Schüler*innen aus heterogeneren Klassen sind im Vergleich zu Schüler*innen aus homogenen Klassen stärkeren Diskrepanzen und Widersprüchen ausgesetzt, was den Lernprozess anregt.

Überdies soll untersucht werden, ob und wie sich der Einfluss der sprachlichen Heterogenität auf die Schüler*innenleistung unterscheidet, wenn Schüler*innen einen Migrationshintergrund aufweisen. Vor dem Hintergrund der Cognitive Load Theory (Sweller, van

Merriënboer & Paas, 1998) kann vermutet werden, dass der Effekt der sprachlichen Heterogenität für Schüler*innen mit und ohne Migrationshintergrund unterschiedlich ausfällt, da sich Schüler*innen mit Migrationshintergrund in der Regel von Schüler*innen ohne Migrationshintergrund in den sprachliche Vorkenntnissen unterscheiden (siehe Anhang B).

6.3 Einstellung zur Leistungsheterogenität von Schüler*innen

Während in Beitrag 1 und Beitrag 2 *Heterogenität und Leistung* im Fokus stehen, soll in Beitrag 3 das Augenmerk auf die *Einstellung zur Leistungsheterogenität* bei Schüler*innen gelegt werden. Im Arbeitsmodell liegen nur diejenigen Variablen mit weißem Hintergrund vor, die auch für die konkrete Fragestellung von Bedeutung sind (siehe Abbildung 6).

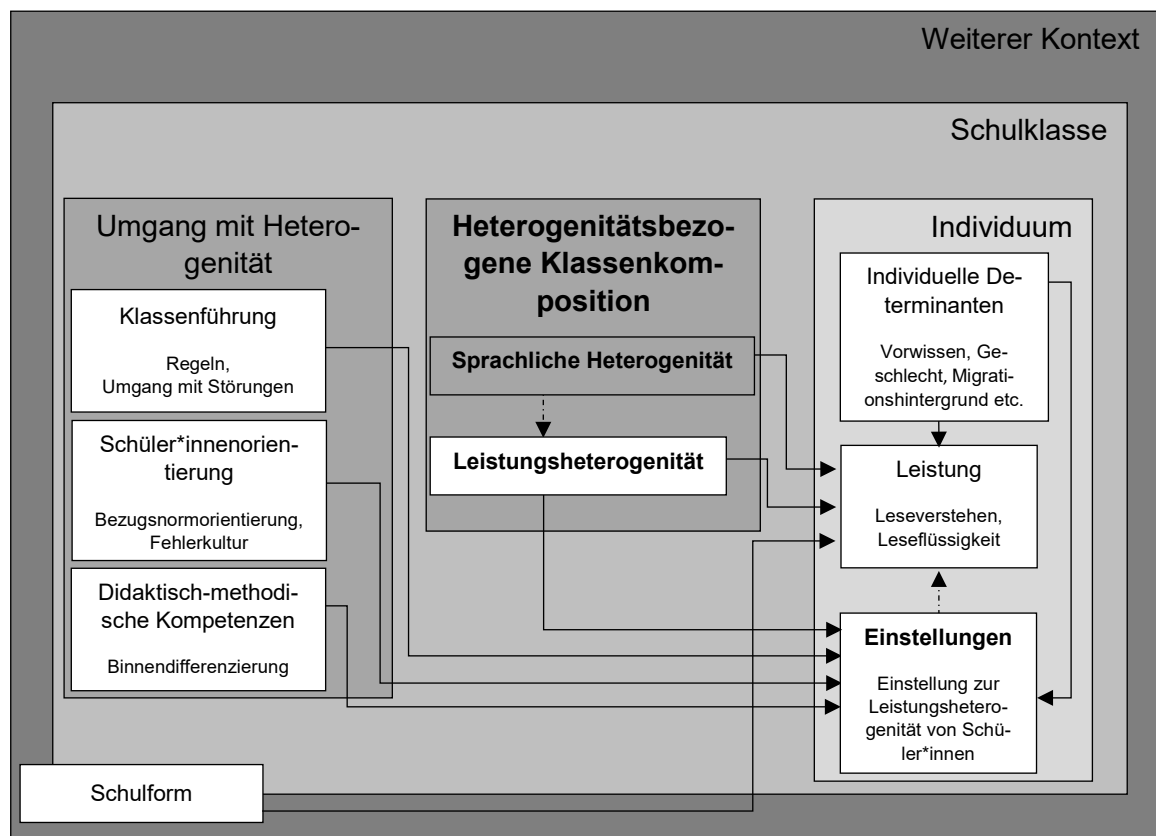


Abbildung 6: Arbeitsmodell Einstellung

Im Arbeitsmodell befindet sich die Einstellung zur Leistungsheterogenität als Teil des *Individuums*. Mit direktem Pfeil der Leistungsheterogenität auf die Einstellungen zur Leistungsheterogenität bei Schüler*innen wird in den Blick genommen, welchen Einfluss die *Leistungsheterogenität* auf diese *Einstellungen* nimmt. Zudem wird durch die direkten

Pfeile des *Umgangs mit Heterogenität* in den Blick genommen, welchen Einfluss das Handeln der Lehrkraft auf die *Einstellung zur Leistungsheterogenität* von Schüler*innen hat. Der gestrichelte Pfeil der *Einstellung* auf die *Leistung* der Schüler*innen deutet darauf hin, dass dieser Effekt nicht untersucht wird und deswegen nur vermutet werden kann. Analysiert wird dafür, ob das *Vorwissen*, also die Leistung zu Beginn des Schuljahres, einen Effekt auf die *Einstellung* hat.

In Kapitel 4 wurde bereits elaboriert, dass die Einstellungen zur Leistungsheterogenität von Schüler*innen in der Forschung zunächst unberücksichtigt blieben. Sich diesen Einstellungen von Schüler*innen zu widmen, kann jedoch aus mehreren Gründen zielführend sein. So zeigen Studien, dass positive Einstellungen der Schüler*innen zu Schule und Unterricht die Leistung positiv beeinflussen (z.B. Else-Quest, Mineo & Higgins, 2013; Martinez, Aricak & Jewell, 2008). Zudem kann die Rückmeldung der schüler*innenspezifischen Einstellung zur Heterogenität an die Lehrkraft von ebendieser bei der Unterrichtsgestaltung berücksichtigt werden (z.B. durch einen erhöhten Einsatz von Differenzierung). Und überdies werden im deutschen Bildungssystem politische Entscheidungen vollzogen, welche die Heterogenität in den Schulen stark beeinflussen (z.B. die Einführung eines offenen Übergangssystems vs. eines notengebundenen Übergangssystems). Diese Entscheidungen und deren Konsequenzen können die Einstellung zur Heterogenität beeinflussen; und dies tangiert nicht nur Lehrkräfte, sondern auch die Schüler*innen. Deswegen ist es von Bedeutung, die Einstellung zur Leistungsheterogenität nicht nur bei Lehrkräften, sondern auch bei Schüler*innen zu untersuchen.

In Beitrag 3 werden drei spezifische Ziele verfolgt. Erstens soll untersucht werden, wie die Einstellung zur Leistungsheterogenität von Schüler*innen ausgeprägt ist. Zweitens soll in den Blick genommen werden, welchen Einfluss die Leistungsheterogenität in der Klasse auf die Einstellung zur Leistungsheterogenität von Schüler*innen nimmt. Und drittens soll der Umgang der Lehrkraft mit Heterogenität und dessen Einfluss auf die Einstellung zur Leistungsheterogenität von Schüler*innen analysiert werden.

Für die theoretische Einbettung wird die Theorie nach Eagly und Chaiken (1993) herangezogen. In deren multidimensionalen Einstellungstheorie wird davon ausgegangen, dass eine Einstellung aus drei Komponenten besteht: der kognitiven, der affektiven und der behavioralen Komponente (z.B. Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960). Um die Einstellung zur Leistungsheterogenität von Schüler*innen nicht nur zu beschreiben, sondern auch deren Abhängigkeit von kontextuellen Merkmalen zu untersuchen, werden zwei potenzielle Faktoren, nämlich die Leistungsheterogenität in der Klasse selbst sowie

der Umgang der Lehrkraft mit Heterogenität, in den Blick genommen. Auf Basis der Kontakthypothese nach Allport (1954) wird angenommen, dass eine zunehmende Leistungsheterogenität in der Klasse zu positiven Veränderungen in der Einstellung zur Leistungsheterogenität der Schüler*innen führt (siehe Anhang C).

Überdies wird auf Basis der in Kapitel 3.3 und Kapitel 6 herausgearbeiteten Faktoren erwartet, dass ein professioneller Umgang mit Heterogenität die Einstellung zur Leistungsheterogenität von Schüler*innen positiv beeinflusst. An dieser Stelle werden folgende Aspekte des Umgangs mit Heterogenität untersucht: Erstens, eine effiziente Klassenführung durch die Implementation von Regeln und durch den effektiven Umgang mit Störungen (z.B. Emmer, Evertson & Worsham, 2006; Seidel & Shavelson, 2007). Zweitens, schüler*innenorientiertes Arbeiten durch eine individuelle Bezugsnormorientierung und durch eine positive Fehlerkultur (z.B. Cornelius-White, 2007; Dickhäuser et al., 2017; Pianta & Hamre, 2009). Und drittens, der Einsatz von binnendifferenzierenden Maßnahmen im Unterricht (z.B. Hattie 2009).

7 Beiträge

Im Folgenden werden die einzelnen Beiträge und deren zentrale Befunde dargestellt. Die Beiträge sind im Rahmen der Mannheimer Forschungsprojekte HEÄT (Herausforderung Heterogenität; Dotzel, Karst & Münzer, 2017) und VESPER (Verbesserung der Sprachkompetenz durch den Einsatz reziproker Lehre; Karst, Bonefeld, Dotzel, Fehringer, Stark & Thielmann, 2020) sowie im Rahmen eines Kooperationsprojekts der Universität Mannheim und Universität Münster (Quop – die Lernverlaufsdiagnostik; Souvignier, Förster & Salaschek, 2014) entstanden. Sie greifen die in Kapitel 6.1, 6.2 und 6.3 formulierten Forschungsfragen auf und versuchen, zur Schließung der in dieser Arbeit identifizierten Forschungslücken beizutragen.

7.1 Beitrag 1 – Leistungsheterogenitätsmaße

Zugehöriges Manuskript:

Dotzel, S., Förster, N., Hebbecker, K., Souvignier, E. & Karst, K. (in preparation for resubmission). Do students benefit from homogeneous or heterogeneous classes? An analysis of linear and non-linear effects of within-class heterogeneity on students' reading performance using different heterogeneity measures. *Frontline Learning Research*.

In diesem Manuskript werden die Effekte der Leistungsheterogenität von Klassen auf die Schüler*innenleistung untersucht (siehe Anhang A). Für die Operationalisierung von Leistungsheterogenität werden drei Heterogenitätsmaße genutzt: die klassenspezifische Standardabweichung (SD), der klassenspezifische Range und der klassenspezifische Interquartilsabstand (IQR). Die Effekte dieser drei Heterogenitätsmaße werden sowohl in linearen als auch nichtlinearen Modellen analysiert.

Die in diesem Beitrag verwendeten Daten wurden im Rahmen einer Längsschnittstudie zur Leseflüssigkeit und zum Leseverstehen in den dritten und vierten Klassenstufen ($N = 2.625$) in Deutschland erhoben (Förster, Kawohl & Souvignier, 2018; Förster & Souvignier, 2014, 2015). Zu Beginn des Schuljahres wurde das Vorwissen der Schüler*innen mit zwei standardisierten Leistungstests gemessen (*Salzburger Lese-Screening für die Klassenstufen 1-4* für Leseflüssigkeit; Mayringer & Wimmer, 2003; *Hamburger Lesetest für 3. und 4. Klassen* für Leseverstehen; Lehmann, Peek & Poerschke, 2006). Zudem nah-

men die Schüler*innen während des Schuljahres an Lernfortschrittserhebungen zur Leseflüssigkeit und zum Leseverstehen teil. Für die Lernfortschrittserhebungen absolvierten Schüler*innen acht kurze computergestützte Tests in Abständen von drei Wochen.

Um den Effekt der Heterogenität auf die Schüler*innenleistung zu untersuchen, wurden längsschnittliche Multilevel-Analysen mit vier Ebenen (Messzeitpunkt, Schüler*innen, Klasse, Schule) berechnet.

Die Ergebnisse zeigen drei wichtige Befunde auf. Erstens hat die Heterogenität, gemessen an der Standardabweichung, keinen signifikanten linearen Effekt auf die Schüler*innenleistung. Zweitens variieren die Ergebnisse durch die Operationalisierung der Heterogenität nur geringfügig. Und drittens zeigen nichtlineare Analysen beim Leseverstehen, dass die Schüler*innenleistung mit zunehmender oder abnehmender Heterogenität in der Klasse steigt.

Auch wenn die Ergebnisse der nichtlinearen Analysen explorativen Charakter haben und aufgrund dessen keine theoretische Einbettung zulassen, werden die nichtlinearen Befunde interpretiert (siehe Anhang A). Leistungsheterogenität hat nach Helmkes Modell (2017) nicht nur einen direkten Effekt auf die Leistung von Schüler*innen, sondern wirkt sich auch indirekt über den Unterricht und das Handeln der Lehrkraft auf die Schüler*innenleistung aus (vgl. Decristan et al., 2017). Dass Schüler*innen in eher heterogenen oder homogenen Klassen eine höhere Leistung aufzeigen, kann daran liegen, dass eine eindeutige Klassenzusammensetzung auch ein spezifisches Handeln der Lehrkraft erfordert (z.B. Differenzierung in eher heterogenen Klassen vs. Frontalunterricht in eher homogenen Klassen). Im Vergleich dazu ist die Wahl von geeigneten Unterrichtsmethoden in durchschnittlich heterogenen Klassen weniger klar und Lehrkräfte könnten Schwierigkeiten haben, an die Leistungsheterogenität angepasste Unterrichtsmethoden zu implementieren. Im Zusammenhang damit könnte ein häufigerer Wechsel zwischen Unterrichtsmethoden zu geringeren Übungseffekten bei Schüler*innen führen, welche gerade für die Effektivität von Gruppenarbeiten von Bedeutung sind (Krause, 2004).

Zusammenfassend lässt sich festhalten, dass die nichtlineare Modellierung beim Leseverstehen im Vergleich zur linearen Modellierung zu signifikanten Ergebnissen führt. Überdies führen auch die empirischen Analysen verschiedener Heterogenitätsmaße zu ähnlichen Ergebnissen. Dennoch trägt die Analyse mehrerer Heterogenitätsmaße sowohl methodisch als auch pädagogisch-inhaltlich zu einem ganzheitlicheren Wirkungsbild bei.

7.2 Beitrag 2 – Sprachliche Heterogenität

Zugehöriges Manuskript:

Dotzel, S., Bonefeld, M. & Karst, K. (Under Review). The impact of linguistic diversity on students' reading comprehension using different diversity measures and its' differential effects depending on students' migration background. *International Journal of Educational Research*.

Im Vergleich zu Beitrag 1 soll in diesem Manuskript nicht die leistungsbezogene Heterogenität und deren Effekte auf die Leistung in den Blick genommen werden, sondern die sprachliche Heterogenität und deren Einfluss auf die Schüler*innenleistung (siehe Anhang B). Für die Operationalisierung von sprachlicher Heterogenität werden drei Heterogenitätsmaße genutzt: die Anzahl an Sprachen, Simpson's D (Simpson, 1949) und Shannon's H (Shannon & Weaver, 1949).

Im Schuljahr 2018/2019 wurde eine längsschnittliche Erhebung mit $N = 1603$ Schüler*innen der fünften Jahrgangsstufe aus 65 Klassen und 27 Sekundarschulen (18 Gymnasien mit 45 Klassen, 4 Realschulen mit 11 Klassen, 3 Gemeinschaftsschulen mit 7 Klassen, 2 Werkrealschulen mit 2 Klassen) durchgeführt. Dabei nahmen die Schüler*innen sowohl im September 2018 als auch im Februar 2019 an einem Test zum Leserverstehen teil (Lernstand 5; Institut für Bildungsanalysen Baden-Württemberg, 2018). Im September 2018 füllten die Schüler*innen außerdem einen Fragebogen aus. In diesem wurde unter anderem erhoben, welche Sprache die Schüler*innen zu Hause mit ihren Eltern sprechen und welche Sprache sie mit ihren Freunden sprechen. Auf Basis dieser Angaben wurden die klassenspezifischen Heterogenitätsmaße berechnet.

Die Ergebnisse der Multilevel-Regressions-Modelle (Level 1 = Schüler*innen, Level 2 = Klasse, Level 3 = Schule) zeigen vier wichtige Befunde auf. Erstens kann die sprachliche Heterogenität die Schüler*innenleistung auch über den klassenspezifischen Migrationsanteil (genutzt als Kontrollvariable) hinaus erklären. Zweitens steigt mit einer zunehmenden sprachlichen Heterogenität in der Klasse das Leseverstehen der Schüler*innen. Drittens profitieren alle Schüler*innen von der sprachlichen Heterogenität in der Klasse; Schüler*innen ohne Migrationshintergrund profitieren jedoch stärker als Schüler*innen mit Migrationshintergrund. Und viertens sind die Ergebnisse sowohl unabhängig vom jeweiligen Heterogenitätsmaß als auch unabhängig davon, ob die Heterogenitätsmaße anhand der gesprochenen Sprache mit den Eltern oder mit Freunden berechnet wurden.

Zusammenfassend kann festgehalten werden, dass die sprachliche Heterogenität eine wichtige Rolle für die Schüler*innenleistung spielt. Zudem zeigen die Analysen den interessanten Befund, dass der Einfluss des Migrationsanteils einer Klasse einen stärkeren Einfluss auf die Schüler*innenleistung hat. Da dieser negativ ausfällt – die Schüler*innenleistung mit zunehmendem Migrationsanteil in der Klasse also sinkt – kann die sprachliche Heterogenität auch als diesen Effekt abschwächend interpretiert werden (siehe Anhang B).

7.3 Beitrag 3 – Einstellung zur Leistungsheterogenität von Schüler*innen

Zugehöriges Manuskript:

Dotzel, S., Bonefeld, M. & Karst, K. (Under Review). Students' attitudes towards performance heterogeneity and their differential change depending on contextual factors. *European Journal of Psychology of Education*.

In diesem Manuskript werden die Einstellungen von Schüler*innen zur Leistungsheterogenität analysiert (siehe Anhang C). Im Vergleich zu Beitrag 1 und 2 soll das Augenmerk also auf die Beziehung zwischen Leistungsheterogenität und Einstellung gelegt werden – und nicht auf Leistungsheterogenität und Leistung. Wie bereits elaboriert, kann jedoch theoretisch davon ausgegangen werden, dass auch die Einstellungen einen indirekten Einfluss auf die Leistung haben (siehe Kapitel 6.3); deswegen wird durch Beitrag 3 eine wichtige Ergänzung zu den Beiträgen 1 und 2 bereitgestellt.

Da die Einstellungen zur Leistungsheterogenität bisher mit Fokus auf die Lehrkraft untersucht wurden und bislang noch wenig über die Einstellung von Schüler*innen bekannt ist, wurden folgende (Pilotierungs-)Studien herangezogen: Im April 2018 wurden neun Leitfadeninterviews (sechs Mädchen, drei Jungen) mit Schüler*innen der fünften und sechsten Jahrgangsstufe durchgeführt. Die inhaltsanalytische Auswertung hat gezeigt, dass bereits Schüler*innen der frühen Sekundarstufe das Phänomen Leistungsheterogenität wahrnehmen (Dotzel & Karst, im Druck; Anhang D). Auf Basis der Auswertung wurde ein Fragebogen entwickelt, welcher im Juli 2018 in 13 Schulklassen der fünften und sechsten Jahrgangsstufe mit einer Stichprobe von $N = 290$ Schüler*innen pilotiert wurde. Die optimierte Version des Fragebogens wurde im Oktober 2018 in der Haupterhebung eingesetzt.

Für die Analysen werden die Daten der Haupterhebung mit $N = 784$ Schüler*innen der fünften Jahrgangsstufe aus 33 Klassen und 12 Schulen (8 Gymnasien mit 25 Klassen, 2 Realschulen mit 3 Klassen, 2 Gemeinschaftsschulen mit 5 Klassen) genutzt. Die Daten

enthalten Informationen zu den sprachlichen Kompetenzen der Schüler*innen (Duisburger Sprachstandstest; Theunissen, Schoppengerd, Witzke, Endell & Wenning, 2005; Karst, Dotzel, Stark, Derkau & Münzer, im Druck) und zu deren Einstellung zur Leistungsheterogenität. Außerdem wurden kontextuelle Unterrichtsvariablen durch die Schüler*innen (Implementation von Regeln im Unterricht, Umgang mit Störungen, individuelle Bezugsnormorientierung, positive Fehlerkultur), als auch durch die zugehörigen Lehrkräfte des Fachs Deutsch (Binnendifferenzierung) erhoben.

Die Ergebnisse der Multilevel-Strukturgleichungsmodelle decken drei wichtige Befunde auf. Erstens, die Einstellung zur Leistungsheterogenität von Schüler*innen ist positiv ausgeprägt. Zweitens, je heterogener die Klassenzusammensetzung (bei gleichem mittlerem Leistungsniveau), desto positiver ist die Einstellung zur Leistungsheterogenität. Drittens, der professionelle Umgang mit Heterogenität führt zu einer positiveren Einstellung zur Leistungsheterogenität bei Schüler*innen.

Zusammenfassend kann festgehalten werden, dass Schüler*innen tendenziell eine positive Einstellung zur Leistungsheterogenität haben. Diese positive Einstellung kann von den Lehrkräften genutzt werden, um eine angenehme Atmosphäre und ein positives Lernumfeld in der Klasse zu schaffen. Es ist von hoher Relevanz, dass Lehrkräfte die Einstellung zur Leistungsheterogenität ihrer Schüler*innen positiv verändern können, wenn sie in ihrem Unterricht Klassenführung, schüler*innenorientierten Unterricht und Binnendifferenzierung einsetzen. Zahlreiche Studien haben bereits zeigen können, dass die Anwendung dieser Methoden die Leistung der Schüler*innen verbessert (z.B. Kunter & Voss, 2011; Pianta & Hamre, 2009; Rakoczy, Klieme, Drollinger-Vetter, Lipowsky, Pauli & Reusser, 2007). Darüber hinaus verbessert sich der Lernprozess der Schüler*innen mit einer positiven schulbezogenen Einstellung der Schüler*innen (z.B. Else-Quest, Mineo & Higgins, 2013; Gaith & Bouzeineddine, 2003; Papanastasiou & Zembylas, 2002). Lehrkräfte können somit einen doppelten Effekt auf die Schüler*innenleistung haben: einmal durch direkte Faktoren, indem eine klare und strukturierte Lernumgebung geschaffen wird. Und daneben auch über indirekte Faktoren, indem Lehrkräfte durch die Anpassung der Strategien an heterogene Klassen die Einstellung zur Leistungsheterogenität von Schüler*innen positiv verändern.

8 Diskussion

Basierend auf der Frage, ob und wie Heterogenität die Leistung von Schüler*innen beeinflusst, wurden in der vorliegenden Arbeit drei Beiträge beschrieben. Jede dieser Beiträge nahm dabei einen spezifischen Blickwinkel ein: Beitrag 1 fokussierte, welche Heterogenitätsmaße zur Messung von Leistungsheterogenität herangezogen werden können und welche linearen und nichtlinearen Effekte auf Basis der verschiedenen Maße verzeichnet werden können. Beitrag 2 untersuchte die Effekte der sprachlichen Heterogenität auf die Schüler*innenleistung und nahm dabei in den Blick, welche Heterogenitätsmaße zur Messung von sprachlicher Heterogenität herangezogen werden können. Und Beitrag 3 analysierte die Einstellung zur Leistungsheterogenität von Schüler*innen und untersuchte, wie diese Einstellungen von heterogenitätsbezogenen Kontextmerkmalen beeinflusst werden.

Die drei Beiträge hatten zum Ziel, zu einem besseren Verständnis von Heterogenität und dessen Effekte auf die Schüler*innenleistung beizutragen und gleichzeitig identifizierte Lücken im Forschungsstand zu schließen. Demnach zeigt die vorliegende Arbeit eine methodische Vielfalt hinsichtlich der Operationalisierung von leistungsbezogener wie auch sprachlicher Heterogenität auf, erweitert den Forschungsstand hinsichtlich des Einflusses von leistungsbezogener und sprachlicher Heterogenität auf die Schüler*innenleistung und richtet einen ersten Blick auf die Einstellungen zur Leistungsheterogenität bei Schüler*innen. Im Folgenden werden die Ergebnisse der drei Beiträge vor dem Hintergrund des Arbeitsmodells (siehe Kapitel 6) diskutiert. Begonnen wird mit den Komponenten *Heterogenitätsbezogene Klassenkomposition* und *Einstellungen*, da in diesen die zentralen Fragestellungen der Arbeit gespiegelt sind. Danach wird die Komponente *Umgang mit Heterogenität* aufgegriffen. Abschließend werden zwei Variablen der Subkomponente *Individuellen Determinanten* diskutiert. Auch wenn die *Individuellen Determinanten* nicht im Zentrum der vorliegenden Arbeit stehen, zeigen sich einige hinsichtlich des individuellen Migrationshintergrundes und des individuellen Vorwissens von Schüler*innen beachtliche Ergebnisse, welche im Rahmen der vorliegenden Dissertation in den Blick genommen werden sollen. Mit Hilfe des Arbeitsmodells werden die Ergebnisse in einen Gesamtzusammenhang gebracht und in einem weiteren theoretischen Gefüge diskutiert.

8.1 Zusammenfassung und Synthese der Befunde

Die vorliegende Arbeit deckt vier wichtige Befunde auf. Erstens, sowohl die leistungsbezogene als auch die sprachliche Heterogenität haben einen Einfluss auf die Schüler*innenleistung, wobei die Effekte der sprachlichen Heterogenität systematischer ausfallen. Zweitens, die Effekte der leistungsbezogenen und sprachlichen Heterogenität sind nur geringfügig (siehe Anhang A) beziehungsweise überhaupt nicht (siehe Anhang B) von der jeweiligen Operationalisierung abhängig. Drittens, während die Effekte der Leistungsheterogenität auf die Schüler*innenleistung unsystematisch ausfallen, zeigen sich hinsichtlich der Effekte der Leistungsheterogenität auf die Einstellung zur Leistungsheterogenität von Schüler*innen stabile, positive Ergebnisse. Und viertens, der Umgang der Lehrkraft mit Heterogenität spielt für die Einstellung zur Leistungsheterogenität von Schüler*innen eine besonders wichtige Rolle; je professioneller eine Lehrkraft mit Heterogenität umgeht, desto höher ist auch die Einstellung zur Leistungsheterogenität.

Zuallererst soll die *heterogenitätsbezogene Klassenkomposition* des Arbeitsmodells in den Blick genommen werden. In der vorliegenden Arbeit kann gezeigt werden, dass sowohl die leistungsbezogene als auch die sprachliche Heterogenität einen Einfluss auf die Schüler*innenleistung hat. Während die sprachliche Heterogenität jedoch stabile, positive Effekte auf das Leseverstehen von Schüler*innen hat, haben sich bezüglich der leistungsbezogenen Heterogenität keine systematischen Effekte auf die Schüler*innenleistung gezeigt. Hier können weder lineare noch nichtlineare Effekte für die Leseflüssigkeit berichtet werden. Ein Anstieg der Schüler*innenleistung ist nur beim Leseverstehen mit der Analyse nichtlinearer Zusammenhänge zu beobachten. Demnach können nur geringe Verbindungslinien zwischen der leistungsbezogenen und der sprachlichen Heterogenität verzeichnet werden. In der theoretischen Herleitung wurde elaboriert, dass sich sprachliche Heterogenität in leistungsbezogener Heterogenität manifestiert. Vor diesem Hintergrund kann die Frage gestellt werden, wie stark die theoretisch angenommene Manifestierung von sprachlicher in leistungsbezogener Heterogenität wirklich vorliegt; denn wenn in der Theorie angenommen wird, dass mit einer Zunahme an sprachlicher Heterogenität auch die Zunahme leistungsbezogener Heterogenität einhergeht, könnte auch davon ausgegangen werden, dass die empirischen Effekte der sprachlichen und leistungsbezogenen Heterogenität auf die Schüler*innenleistung ähnlich ausfallen.

Im Vergleich zu den unsystematischen Effekten der Leistungsheterogenität auf die Schüler*innenleistung zeigen sich stabile, positive Effekte der Leistungsheterogenität auf die

Einstellung zur Leistungsheterogenität von Schüler*innen. Demnach weisen Schüler*innen aus leistungsheterogenen Klassen auch eine positivere Einstellung zur Leistungsheterogenität auf. Dies ist als sehr bedeutsam zu bewerten. Sowohl in Beitrag 1, als auch in einer Vielzahl vorheriger Studien (z.B. Bellin, 2009; Carman & Zhang, 2012; Decristan et al., 2017; Hattie, 2009; Leiter, 1983; Wright, Horn & Sanders, 1997) konnten nur geringe bis keine Effekte der Leistungsheterogenität auf die Leistung nachgewiesen werden; doch im selben Zuge spielt die Heterogenität über die Einstellung zur Leistungsheterogenität von Schüler*innen eine indirekte Rolle für die Schüler*innenleistung.

Daran anknüpfend zeigt die Komponente *Umgang mit Heterogenität* des Arbeitsmodells, dass die Einstellung zur Leistungsheterogenität von Schüler*innen nicht nur durch die Leistungsheterogenität in der Klasse erklärt werden kann. In Kapitel 3.3 und Kapitel 6 wurden mit Hilfe von Pietschs Stufenmodell (2010) die Basisdimensionen der Unterrichtsqualität (Kunter und Trautwein, 2013) in einen heterogenen Klassenkontext eingeordnet. Die Ergebnisse aus Beitrag 3 lassen darauf schließen, dass der Umgang mit Heterogenität durch die Lehrkraft einen positiven Einfluss auf die Einstellung zur Leistungsheterogenität von Schüler*innen hat. Zudem unterstreichen die Ergebnisse, dass die Strategien, die eine Lehrkraft im Umgang mit heterogenen Klassen nutzt, den Effekt der Leistungsheterogenität übersteigen.

Auch wenn die *individuellen Determinanten* des Arbeitsmodells für die zentralen Thesen der vorliegenden Arbeit eine geringere Rolle spielen, sollen diese aufgrund von zwei beachtlichen Ergebnissen diskutiert werden. Zuerst werden die Effekte des individuellen Migrationshintergrundes der Schüler*innen in den Blick genommen und anschließend die Effekte des individuellen Vorwissens diskutiert.

Mit Fokus auf den Migrationshintergrund von Schüler*innen zeigen sich in den Beiträgen 2 und 3 keine Haupteffekte des individuellen Migrationshintergrundes auf die Schüler*innenleistung. Dies ist beachtlich, da in großen nationalen, wie internationalen Bildungsstudien wie PISA, TIMSS oder IGLU immer wieder festgestellt wird, dass Schüler*innen mit Migrationshintergrund schlechtere Ergebnisse erzielen (z.B. Baumert & Schümer, 2001; Bos, Lankes, Prenzel, Schwippert, Walter & Valtin, 2003; Stanat, Rauch & Segeritz, 2010; Walter & Taskinen, 2007, 2008). Interessant zeigt sich an dieser Stelle, dass die genutzten Daten aus der Umgebung Mannheim herangezogen wurden, welche mit einem Migrationsanteil der Schüler*innenschaft von über 60% einhergehen. Auch wenn Vergleiche zu Klassen mit geringem Migrationsanteil deswegen schwer zu ziehen sind (z.B. Landschulen; siehe Kapitel 8.2), sind die Ergebnisse für die Heterogenitätsforschung relevant: unter der

Annahme, dass sich migrationsbedingte Unterschiede letzten Endes auch in sprachlich bedingten Unterschieden (vgl. Rjosk et al., 2017) und folglich in leistungsbedingten Unterschieden manifestieren (Fischer, Rott & Veber, 2014), ist der Migrationshintergrund der Schüler*innen für den Heterogenitätsdiskurs von Bedeutung. Mit besonderem Augenmerk auf Beitrag 2 konnte gezeigt werden, dass der Migrationshintergrund der einzelnen Schüler*innen keinen Einfluss auf die Leistung nimmt, der klassenspezifische Migrationsanteil die individuelle Schüler*innenleistung jedoch signifikant verringert (siehe Anhang B). Vor diesem Hintergrund kann festgehalten werden, dass – zumindest in Klassen mit hohem Migrationshintergrund – der individuelle Hintergrund kaum bis keine Effekte auf die Schüler*innenleistung hat, dies für die migrationsbedingte Klassenkomposition jedoch nicht der Fall ist. So kumulieren sich die Effekte der einzelnen Individuen zu einem klassenspezifischen Effekt, welcher folglich auch nicht nur einzelne Individuen beeinflusst, sondern die gesamte Schüler*innenschaft.

Überdies soll das individuelle Vorwissen und dessen enormer Einfluss auf die Schüler*innenleistung in den Blick genommen werden. Beitrag 1 und Beitrag 2 zeigen einen ähnlich hohen Effekt der Leistung von Schüler*innen zu Beginn des Schuljahres auf deren Leistung zu späteren Zeitpunkt(en); und im direkten Vergleich ist der Einfluss von sowohl leistungsbezogener als auch sprachlicher Heterogenität sehr gering. Dieser Befund ist beachtlich, wenn man berücksichtigt, welche Aufmerksamkeit das Thema *Heterogenität* sowohl in den Medien, in der Bildungspolitik als auch in der Schulpraxis erhält. Während das Thema *Leistungsheterogenität* kontrovers diskutiert wird (Scharenberg, 2012), zeigen sich die empirischen Effekte als verhältnismäßig schwach.

Um daran anzuschließen zeigen die Ergebnisse aus Beitrag 3, dass die Leistung von Schüler*innen keinen Effekt auf die Einstellung zur Leistungsheterogenität nimmt; und es in diesem Falle gerade die Heterogenität und die Maßnahmen im professionellen Umgang mit Heterogenität sind, die die Einstellung zur Leistungsheterogenität positiv verändern. Auch wenn der Einfluss der Einstellung zur Leistungsheterogenität auf die tatsächliche Leistung durch zukünftige Forschung geklärt werden muss, zeigt es sich als wichtigen Befund, dass sowohl leistungsschwache, als auch leistungsstarke Schüler*innen – welche von der leistungsbezogenen Heterogenität in der Klasse auf unterschiedliche Weise betroffen sind (z.B. Warten und Langeweile bei leistungsstärkeren Schüler*innen; Überforderung und Abschalten bei leistungsschwächeren Schüler*innen: Dotzel & Karst, 2020; Slavin, 1987) – eine ähnliche Einstellung aufweisen.

Zusammenfassend können die vorgestellten Beiträge zeigen, dass Heterogenität die Schüler*innenleistung positiv beeinflussen kann, der Einfluss von Heterogenität auf die Schüler*innenleistung jedoch von einer hohen Komplexität gekennzeichnet ist und deswegen differenziert betrachtet werden muss. Zudem unterstreichen die Ergebnisse, dass Heterogenität auch eine relevante Rolle für die Einstellungen von Schüler*innen spielt.

8.2 Limitationen

Bei der Interpretation der vorliegenden Arbeit müssen einige Limitationen berücksichtigt werden. Ziel dieses Kapitels ist die Einordnung in den Gesamtzusammenhang. Für eine detaillierte Diskussion und Betrachtung der Limitationen der einzelnen Beiträge wird daher auf die einzelnen Manuskripte verwiesen.

Zuallererst soll die Spezifität der Stichprobe einem kritischen Blick unterzogen werden. Vor allem in den Studien, die an der Universität Mannheim erhoben wurden (Beitrag 1 und Beitrag 2), liegt der Migrationshintergrund der Schüler*innenschaft bei über 60%. Dies ist im deutschlandweiten Vergleich beachtlich hoch, bei welchem nur rund ein Drittel der Schüler*innen einen Migrationshintergrund aufweisen (Statistisches Bundesamt, 2019). Dieser fast doppelt so hohe Anteil an Schüler*innen mit Migrationshintergrund könnte einen Einfluss auf die berichteten Ergebnisse haben, weswegen der statistische Schluss auf Klassen mit einem geringeren Migrationshintergrund kritisch gesehen werden kann. Überdies zeigt sich im Datensatz des Beitrags 1 ein Migrationsanteil von nur 23% (gemessen an der zu Hause gesprochenen Sprache), welcher den Vergleich der vorgestellten Beiträge erschwert. Überdies muss angemerkt werden, dass alle Untersuchungen als Sekundäranalysen durchgeführt wurden. Die Daten wurden also ursprünglich für die Analyse von inhaltlich anderen Fragestellungen erhoben, was ebenfalls einen Einfluss auf die berichteten Ergebnisse haben kann. Überdies sind die drei Studien bezüglich der Jahrgangsstufen stark eingeschränkt. So beziehen die jeweiligen Studien nur eine bis zwei Jahrgangsstufen mit ein. Zudem muss auch hier kritisch angemerkt werden, dass sich die Studien über die drei Jahrgangsstufen 3, 4 und 5 erstrecken, wobei Beitrag 1 Daten aus der Elementarstufe untersucht und Beitrag 2 und 3 Daten aus der Sekundarstufe. Die Ergebnisse können also auch durch altersbedingte und jahrgangsbedingte Unterschiede beeinflusst sein, was die Vergleichbarkeit erschwert. Trotz alledem sind Daten aus der Elementar- und frühen Sekundarstufe besonders geeignet, um die vorgestellten Forschungsfragen zu untersuchen. Da entweder noch keine Leistungsselektion vorgenommen wurde oder die Schüler*innen erst einen kurzen Zeitraum Teil einer gemeinsamen Klassengemeinschaft sind, ist gerade in diesen Jahrgangsstufen Heterogenität besonders prävalent. Vor diesem Hintergrund

kann die Spannweite von mehreren Jahrgangsstufen auch als gewinnbringend interpretiert werden, da das Thema *Heterogenität* dadurch aus einer übergreifenden Perspektive untersucht und diskutiert werden kann. Außerdem soll auch auf den fachspezifischen Fokus der Beiträge hingewiesen werden. Alle drei Studien wurden im Fach- und Kompetenzbereich Deutsch oder Lesen erhoben. Auf der einen Seite spricht dies für eine gute Vergleichbarkeit der Ergebnisse, jedoch ist fraglich, ob die Ergebnisse in anderen Fächern ähnlich ausfallen würden. Deswegen muss darauf hingewiesen werden, dass der statistische Schluss nur im Fach- und Kompetenzbereich Deutsch und Lesen vollzogen werden kann. Die Effekte von sprachlicher und leistungsbezogener Heterogenität auf die Mathematikleistung, sowie die Einstellung zur Leistungsheterogenität im Fach Mathematik könnten im Vergleich zu den Ergebnissen dieser Arbeit anders ausfallen (vgl. fachspezifische Unterschiede in den Effekten innerhalb einer Studie; z.B. Rjosk et al., 2017; Vigdor & Neychbar, 2007). Trotz alledem soll der Fokus auf das Fach Deutsch auch als Stärke der vorliegenden Arbeit hervorgehoben werden. Sprache ist für die Leistung von Schüler*innen ein wichtiges Medium; und so ist bereits mehrfach nachgewiesen, dass die Leistung im Fachbereich Deutsch – insbesondere im Leseverstehen – die Leistung in anderen Fachbereichen beeinflusst (Fuchs et al., 2006; Lee & Fradd, 1998). Zwar ist nicht bekannt, ob sich solche Verbindungslinien auch für die Leistungsheterogenität ziehen lassen, jedoch lässt sich hiermit begründen, wieso sich die vorliegende Arbeit vor allem vor dem Hintergrund der neuen Forschungsfragen mit dem Fach- und Kompetenzbereich Deutsch und Lesen befasst. Zu guter Letzt soll der Einfluss von Drittvariablen – insbesondere Eigenschaften der Lehrkraft – in den Blick genommen werden. Die Bedeutung des Umgangs mit Heterogenität durch die Lehrkraft wurde in dieser Arbeit bereits thematisiert. Dennoch wurden nur in Beitrag 3 Variablen des Unterrichts und der Lehrkraft miteinbezogen. Und auch in Beitrag 3 konnte die Variable *Einstellung zur Leistungsheterogenität der Lehrkraft* nicht erhoben werden, welche womöglich einen Effekt auf die Einstellung von Schüler*innen nimmt. Der mögliche Einfluss von Lehrkraftvariablen (z.B. Persönlichkeitsmerkmale, Stereotype) erschwert die Vergleichbarkeit zwischen den Beiträgen. Zudem wurde der Umgang mit Heterogenität der Lehrkraft in Beitrag 1 und 2 nicht erhoben, was jedoch auch einen Effekt auf die Schüler*innenleistung und die Leistungsheterogenität in der Klasse haben kann.

8.3 Implikationen für zukünftige Forschung

Die vorliegende Dissertation verdeutlicht die Relevanz für das Thema *Heterogenität in der Schule* und wirft in der Gesamtsynthese einige Fragen auf, die für zukünftige Forschung von Bedeutung sind. Diese werden im Folgenden vorgestellt.

Die vorliegenden Beiträge unterstreichen, wie komplex das Themenfeld Heterogenität ist und dass es sich nicht darauf begrenzen lässt, alleinig die Leistungsheterogenität zu untersuchen – auch wenn es das für die Schulpraxis relevanteste Heterogenitätsmerkmal darstellt (vgl. Scharenberg, 2012). Vor dem Hintergrund, dass gerade der Forschungsstand zum Thema *Leistungsheterogenität und Schüler*innenleistung* undurchsichtig ist, ist es von großer Wichtigkeit, dass sich die theoretische Wissenschaft dem Thema annähert und beschreibt, wie genau der Effekt von Leistungsheterogenität auf die Schüler*innenleistung begründet werden kann (vgl. Beitrag 2 mit Piaget, 1977). Zudem ist von Bedeutung, dass sich zukünftige Studien damit auseinandersetzen, wie die verschiedenen Heterogenitätsdimensionen miteinander zusammenhängen. Unter Einbezug der Ergebnisse aus Beitrag 1 und 2 muss explizit überprüft werden, ob die vermutete Manifestierung von sprachlicher Heterogenität in leistungsbezogener Heterogenität auch empirisch nachweisbar ist. Zukünftige Studien müssen also komplexe Struktur-Gleichungs-Modelle heranziehen, in denen die aufgestellten Thesen beeinflussender und beeinflusster Heterogenitätsdimensionen getestet werden können. Mit besonderem Blick auf Beitrag 2 stellt sich auch die Frage, ob und wie die sprachliche Heterogenität von ethnischer Heterogenität bedingt ist (Rjosk et al., 2017; siehe Anhang B). Des Weiteren ist es wichtig, die gefundenen Ergebnisse weiter zu untermauern (z.B. Replikationsanalysen, zusätzliche konfirmatorische Prüfungen), da in allen Beiträgen neue Ansätze oder Forschungsfelder exploriert wurden. Vor dem Hintergrund der Einstellung zur Heterogenität bei Schüler*innen ist es für zukünftige Studien interessant herauszufinden, welchen Einfluss die Einstellung zur Leistungsheterogenität der Lehrkraft auf die der Schüler*innen hat. Zudem muss explizit untersucht werden, ob – wie bei Bloom (1976) angenommen – die positive Einstellung zur Leistungsheterogenität von Schüler*innen tatsächlich zu einer höheren Leistung führt.

8.4 Implikationen für die Praxis

Die Befunde der vorliegenden Arbeit sind für die Praxis relevant. Deswegen werden im folgenden Kapitel praktische Implikationen abgeleitet und erläutert.

Wie in Kapitel 8.1 konstatiert, ist das Verhältnis der Gewichtung von Heterogenität in der Praxis vs. Wissenschaft unausgewogen. So scheint das Thema *Heterogenität* für die Bildungspraxis von größerer Bedeutung zu sein, als sich empirisch nachweisen lässt. Lehrkräfte weisen die Tendenz auf, ein eher negatives Bild gegenüber heterogenen Klassen zu haben (Baumert et al., 1997). Zudem haben Lehrkräfte Schwierigkeiten im Umgang mit heterogenen Klassen; und so scheint es eine häufige Vorstellung in der Bildungspraxis zu sein, dass heterogene Klassen auch *die schwierigen Klassen* sind (z.B. Gebauer, McElvany

& Klukas, 2013; Kastirke & Jennessen, 2006; van Ackeren & Kühn, 2017). Die vorliegenden Beiträge zeigen jedoch, dass Heterogenität positive Effekte haben kann. Vor dem Hintergrund von Beitrag 1 und Beitrag 2 zeigt sich, dass sich Heterogenität positiv auf die Schüler*innenleistung auswirken kann. Vor dem Hintergrund von Beitrag 3 zeigt sich, dass sich Heterogenität positiv auf die Einstellung zur Leistungsheterogenität auswirkt – und dies wiederum kann einen Effekt auf die Leistung haben (z.B. Else-Quest, Mineo & Higgins, 2013; Martinez, Aricak & Jewell, 2008). Diese Ergebnisse sind vor allem für die Lehramtsausbildung relevant. Lehrkräfte und angehende Lehrkräfte müssen darin geschult werden, dass Heterogenität – vor allem die sprachliche Heterogenität – mit einer Leistungssteigerung der Schüler*innen einhergehen kann. Dies kann zum einen dazu beitragen, dass auch Lehrkräfte eine positivere Einstellung zur Heterogenität entwickeln (denn demnach geht Heterogenität nicht nur mit einer persönlichen Belastung der Lehrkräfte einher, sondern auch mit einer Unterstützungsfunktion für das Lernen der Schüler*innen). Zum anderen kann dies auch dazu führen, dass sich Lehrkräfte intensiver mit dem Umgang mit Heterogenität auseinandersetzen. Studien konnten zeigen, dass sich Studierende nicht genug auf die sprachliche Heterogenität in den Schulen vorbereitet fühlen (Becker-Mrotzek, Rosenberg, Schroeder & Witte, 2017; Gebauer, McElvany & Klukas, 2013). Vor dem Hintergrund der positiven Ergebnisse der sprachlichen Heterogenität aus Beitrag 2 ist es deswegen von Bedeutung, dass Lehrkräfte in Strategien zum Umgang damit geschult werden (z.B. bilinguale Bildungsprogramme; Gogolin, 2011; Markic & Abels, 2014; Miller, 2009; Salleh, Venville & Treagust, 2007; geführter Übergang vom Erlernen der Muttersprache zum Erlernen einer anderen Sprache; Tsung, 2012; Vertiefung des Sprachkontakts; Driesen, 2002; Entwisle & Alexander, 1994; Peetsma, van der Veen, Koopman & van Schooten, 2006; Scheele, Leseman & Mayo, 2010). Eine weitere praktische Implikation, die in der Lehrer*innenbildung Anklang finden soll, ist die ganzheitliche Betrachtung von Heterogenität. So wurde in Beitrag 1 und 2 hergeleitet, welche Perspektive durch die jeweiligen Heterogenitätsmaße eingenommen wird. Dies sollte auch mit Lehrkräften diskutiert werden, da es den Begriff Heterogenität für diese greifbarer machen könnte (z.B. Was bekomme ich als Lehrkraft zurückgemeldet, wenn landesweit angelegte Studien die Standardabweichung berichten?). Zudem wurde in der vorliegenden Arbeit auch die Komplexität des Themas *Heterogenität* verdeutlicht. Auch dies sollte in der Lehramtsausbildung stärker eingebaut werden, indem nicht nur Maßnahmen im Umgang mit leistungsbezogener Heterogenität erlernt werden, sondern auch solche für den Umgang mit weiteren Heterogenitätsdimensionen. An dieser Stelle wäre es außerdem von Bedeutung, dass fachübergreifende, interdisziplinäre Lernangebote von den Universitäten angeboten werden. Die

Betrachtung von Heterogenität und der Umgang damit aus mehreren fachlichen Perspektiven, wie der Psychologie, der Fachdidaktik und der Pädagogik, wäre an dieser Stelle besonders gewinnbringend, um das Handlungsfeld für Lehrkräfte besser greifbar zu machen.

8.5 Fazit

Zusammenfassend zeigt die vorliegende Arbeit, dass eine vielfältige Betrachtung des Forschungsgegenstandes Heterogenität für die pädagogisch-psychologische Wissenschaft und Praxis gewinnbringend ist. Diesbezüglich kann die vorliegende Arbeit folgende Ableitungen zum Thema *Leistungsbezogene und sprachliche Heterogenität in der Schule und deren erwartete Effekte auf die Leistung von Schüler*innen* vornehmen:

- Leistungsbezogene Heterogenität hat keine linearen, aber nichtlineare Effekte auf die Schüler*innenleistung; wenn auch unsystematisch
- Sprachliche Heterogenität hat positive Effekte auf die Schüler*innenleistung
- Unterschiedliche Heterogenitätsmaße von leistungsbezogener und sprachlicher Heterogenität sind pädagogisch-inhaltlich sinnvoll, führen jedoch empirisch zu ähnlichen Ergebnissen
- Leistungsbezogene Heterogenität hat zwar keinen direkten Einfluss auf die Leistung, jedoch auf die Einstellung zur Leistungsheterogenität
- Im Mittel zeigen Schüler*innen eine positive Einstellung zur Leistungsheterogenität

Schule ist ein zentraler Bildungsort und bestimmt zu großen Teilen die Bildungsbiografie. Vor diesem Hintergrund ist Heterogenität von großer Bedeutung – vor allem im Hinblick darauf, dass Heterogenität die Schullaufbahn von Schüler*innen im Sinne des Besuchs verschiedener Schulformen systemisch bedingt. Dadurch hat Heterogenität auch eine Bedeutung in Bezug auf den Zugang zu Bildung sowie zur weiterführenden Bildung, wie die Zulassung zu verschiedenen Studiengängen oder Berufswegen. Durch die vorliegende Arbeit wurden einige Forschungslücken verringert und gleichzeitig neue Forschungsfragen aufgeworfen. Das Thema *Heterogenität in der Schule* bleibt also nicht nur eine Herausforderung für die Schulpraxis und die Lehrkräfte, sondern auch für die pädagogisch-psychologische Wissenschaft.

Literaturangaben

- Allport, G. (1954). *The nature of prejudice*. Garden City, NJ: Doubleday Anchor Books.
- Anderson, L. M., Evertson, C. M. & Brophy, J. E. (1979). An experimental study of effective teaching in first grade groups. *Elementary School Journal*, 79(1), 193–223. doi:10.1086/461151
- Baumert, J. (2002). Umgang mit Heterogenität. Ein Gespräch mit Professor Jürgen Baumert, wissenschaftlicher Leiter des deutschen Teils der PISA-Studie. *Forum Schule. Magazin für Lehrerinnen und Lehrer*, 1, 72–75.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. et al. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: deskriptive Befunde*. Opladen: Leske und Budrich.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–410). Opladen: Leske und Budrich.
- Becker-Mrotzek, M., Rosenberg, P., Schroeder, C. & Witte, A. (Hrsg.) (2017). *Deutsch als Zweitsprache in der Lehrerbildung* (Sprachliche Bildung, Bd. 2, S. 203–215). Münster: Waxmann.
- Bellin, N. (2009). *Klassenkomposition, Migrationshintergrund und Leistung. Mehrebenenanalyse zum Sprach- und Leseverständnis von Grundschulern*. Wiesbaden: VS.
- Benner, A. D. & Crosnoe, R. (2011). The racial/ethnic composition of elementary schools and young children's academic and socioemotional functioning. *American Educational Research Journal*, 48, 621–646. doi:10.3102/0002831210384838
- Biederbeck, I. & Rothland, M. (2017). Professionalisierung des Umgangs mit Heterogenität. In T. Bohl, J. Budde & M. Rieger-Ladich (Hrsg.), *Umgang mit Heterogenität in Schule und Unterricht* (S. 223–236). Bad Heilbrunn: Klinkhardt.
- Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: McGraw-Hill.
- 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* (S. 257–274). Bad Heilbrunn: Klinkhardt.
- Böing, A.M. (1996). Mit Unterschieden umgehen lernen. Heterogenität im Unterricht. Abschlussbericht des Forschungsschwerpunktes „Heterogenität im Unterricht“ am Oberstufen-Kolleg Bielefeld. In L. Huber & A. Wenzel (Hrsg.), *Wir sind alle gleich. Wir sind alle verschieden. Erfahrungen im Umgang mit Heterogenität in der Sekundarstufe II* (S. 43–78). Bielefeld: Oberstufen-Kolleg des Landes NW.
- Boller, S., Rosowski, E. & Stroot, T. (Hrsg.) (2007). *Heterogenität in Schule und Unterricht. Handlungsansätze zum pädagogischen Umgang mit Vielfalt*. Weinheim und Basel: Beltz.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walter, G. & Valtin, R. (Hrsg.) (2003): *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA Harvard University Press. doi:10.1080/00131728109336000
- Brophy, J. (1986). Classroom Management Techniques. *Education and Urban Society*, 18(2), 182–194. doi:10.1177/0013124586018002005
- Brophy, J. (2000). *Teaching*. Genf: IBE.
- Burke, M. A. & Sass, T. R. (2013). Classroom Peer Effects and Student Achievement. *Journal of Labor Economics*, 31(1), 51–82. doi:1086/666653
- Carl, F. (2017). *Gymnasium ohne Sitzenbleiben – Wie Lehrpersonen mit leistungsschwachen Schülerinnen und Schülern umgehen*. Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-658-15416-5_3
- Carman, K. G. & Zhang, L. (2012). Classroom Peer Effects and Academic Achievement: Evidence from a Chinese Middle School. *China Economic Review*, 23, 223–237. doi:10.1016/j.chieco.2011.10.004
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77(1), 113–143. doi:10.3102/003465430298563

- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 121-129.
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.), *Schooling and Language Minority Students: A Theoretical Framework* (pp. 16-62). Los Angeles: Evaluation, Dissemination and Assessment Center California State University.
- Dar, Y. & Resh, N. (1986). Classroom Intellectual Composition and Academic Achievement. *American Educational Research Journal*, 23, 357-374.
- Decristan, J., Fauth, B., Kunter, M., Büttner, G. & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research*, 86, 109-121. doi:10.1016/j.ijer.2017.09.004
- Dickhäuser, O., Janke, S., Praetorius, A.-K. & Dresel, M. (2017). The Effects of Teachers' Reference Norm Orientations on Students' Implicit Theories and Academic Self-Concepts. *Zeitschrift für Pädagogische Psychologie*, 31(3-4), 205-219. doi:10.1024/1010-0652/a000208
- Ditton, H. (2000). Elemente eines Systems der Qualitätssicherung im schulischen Bereich. In H. Weis- haupt (Hrsg.), *Qualitätssicherung im Bildungswesen – Probleme und aktuelle Forschungsbefunde* (S. 13-36). Erfurt: PH Erfurt.
- Dotzel, S. & Karst, K. (im Druck). Einstellung zur Leistungsheterogenität bei Schülerinnen und Schül- ern. Wie können sie gemessen werden, wie sind sie ausgeprägt und wie können sie beeinflusst wer- den?. In K. Karst, D. Thoma, J. Derkau, J. Seifried & S. Münzer (Hrsg.), *Lehrer*innenbildung im Kontext leistungsbezogener Heterogenität und Mehrsprachigkeit. Die Leuchtturm-Projekte der Uni- versität Mannheim*. Münster: Waxmann.
- Dotzel, S., Karst, K. & Münzer, S. (2017). *Heterogenität kognitiver Leistungsdispositionen nach dem Übergang in die Sekundarstufe I. Abschlussbericht über die Sprachstandsdiagnostik 2017*. Verfüg- bar unter: https://www.sowi.uni-mannheim.de/media/Lehrstuehle/sowi/Karst/Dateien/HEAT_Projektbericht_2017_FINAL.pdf
- Dreeben, R. & Barr, R. (1988). Classroom Composition and the Design of Instruction. *Sociology of Ed- ucation*, 61(3), 129-142. doi:10.2307/2112622
- Driessen, G. (2002). School composition and achievement in primary education: A large-scale multi- level approach. *Studies in Educational Evaluation*, 28, 347-368. doi:10.1016/S0191-491X(02)00043-3
- Dupriez, V., Dumay, X. & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review*, 52(2), 245-273. doi:10.1086/528764
- Eagly, A.H. & Chaiken, S. (1993). *Psychology of Attitudes*. Fort Worth: Harcourt Brace Jovanovich College Publishers.
- Else-Quest, N.M., Mindeo, C.C. & Higgins, A. (2013). Math and Science Attitudes and Achievement at the Intersection of Gender and Ethnicity. *Psychology of Women Quarterly*, 37(3), 293-309. doi:10.1177/0361684313480694
- Emmer, E. T., Evertson, C. M. & Worsham, M. E. (2006). *Classroom management for middle and high school teachers* (7. Aufl.). Boston, MA: Allyn & Bacon.
- Entwisle, D. & Alexander, K. (1994). Winter setback: the racial composition of schools and learning to read. *American Sociological Review*, 59, 446-460. doi:10.2307/2095943
- Esser, H. (2006). *Sprache und Integration. Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt: Campus Verlag.
- Evertson, C. M. & Emmer, E. T. (1982). Preventive classroom management. In D. Duke (Ed.), *Helping Teachers manage Classrooms* (pp. 2-31). Alexandria, VA: Association for Supervision and Curricu- lum Development.
- Evertson, C. M., Emmer, E. T. & Worsham, M. E. (2006). *Classroom management for elementary teach- ers* (7. Aufl.). Boston, MA: Allyn & Bacon.
- Feyerer, E. (2014). Einstellungen und Haltungen zur inklusiven Schule. *Erziehung und Unterricht*, 3-4, 219-227.
- Fischer, C., Rott, D. & Veber, M. (2014). Diversität von Schüler/-innen als mögliche Ressource für in- dividuelles und wechselseitiges Lernen im Unterricht. *Lehren & Lernen*, 40(8/9), 22-28.

- Förster, N., Kawohl, E. & Souvignier, E. (2018). Short- and long-term effects of assessment-based differentiated reading instruction in general education on reading fluency and reading comprehension. *Learning and Instruction*, 56, 98-109. doi:10.1016/j.learninstruc.2018.04.009
- Förster, N. & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction*, 32, 91-100. doi:10.1016/j.learninstruc.2014.02.002
- Förster, N. & Souvignier, E. (2015). Effects of providing teachers with information about their students reading progress. *School Psychology Review*, 44, 60-75. doi:10.17105/SPR44-1.60-75
- Fraser, B. J., Walberg, H. J., Welch, W. W. & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11(2), 145-252.
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M. et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, 98(1), 29-43. doi:10.1037/0022-0663.98.1.29
- Gaith, G. & Bouzeineddine, A., R. (2003). Relationships between Reading Attitudes, Achievement, and Learners' Perception of their Jigsaw II Cooperative Learning Experience. *Reading Psychology*, 24, 105-121. doi:10.1080/02702710390197444
- Gebauer, M.M., McElvany, N. & Klukas, S. (2013). Einstellungen von Lehramtsanwärterinnen und Lehramtsanwärtern zum Umgang mit heterogenen Schülergruppen in Schule und Unterricht. In N. McElvany, M.M. Gebauer, W. Bos & G. Holtappels (Hrsg.), *Jahrbuch der Schulentwicklung* (S. 191-216). Weinheim, Basel: Juventa.
- Gogolin, I. (2011). The challenge of super-diversity for education in Europe. *Education Inquiry*, 2(2), 239-249. doi:10.3402/edui.v2i2.21976
- Hartinger, A., Grittner, F., Lang, E. & Rehle, C. (2010). Ein Vergleich der Einstellung zur Heterogenität von Lehrkräften in jahrgangsgemischten und jahrgangshomogenen Klassen. In K.-H. Arnold, K. Hauenschild, B. Schmidt & Ziegenmeyer, B. (Hrsg.), *Zwischen Fachdidaktik und Stufendidaktik. Perspektiven für die Grundschulforschung* (S. 77-80). Berlin: VS Verlag für Sozialwissenschaften.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge. doi:10.1080/00071005.2011.584660
- Helmke, A. (2003). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyer.
- Helmke, A. (2006). Was wissen wir über guten Unterricht? Über die Rückbesinnung auf den Unterricht als Kerngeschäft der Schule. *Pädagogik*, 2, 42-45.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (7. Aufl.). Seelze: Klett-Kallmeyer.
- Hopf, D. (2005). Zweisprachigkeit und Schulleistung bei Migrantenkindern. *Zeitschrift für Pädagogik*, 51, 236-251.
- Huang, M.-H. (2008). Classroom Homogeneity, Family Background, and Math Performance: A Cross-National Difference-in-Differences Analysis. *Taiwanese Journal of Sociology* 40, 1-44.
- Institut für Bildungsanalysen Baden-Württemberg (2018). *Lernstand 5*. Unveröffentlichtes Testinstrument. Verfügbar unter: <https://ibbw.kultus-bw.de/Lde/Startseite/Kompetenzmessung/Lernstand+5>
- Jensen, P. & Rasmussen, A. W. (2011). The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review*, 30, 1503-1515. doi:10.1016/j.econedurev.2011.08.002 doi:10.1016/j.econedurev.2011.08.002
- Karst, K., Bonefeld, M., Dotzel, S., Fehrer, B., Stark, M. & Thielmann, M. (2020). *Ergebnisbericht. Verbesserung der Sprachkompetenz durch den Einsatz reziproker Lehre (VESPER)*. Verfügbar unter: https://www.sowi.uni-mannheim.de/media/Lehrstuehle/sowi/Karst/Dateien/Ergebnisbericht_VESPER_260820.pdf
- Karst, K., Dotzel, S., Stark, M., Derkau, J. & Münzer, S. (im Druck). Diagnostik im Unterricht unterstützt durch Kooperationen in Campus Community-Partnerschaften. In K. Karst, D. Thoma, J. Derkau, J. Seifried & S. Münzer (Hrsg.), *Lehrer*innenbildung im Kontext leistungsbezogener Heterogenität und Mehrsprachigkeit. Die Leuchtturm-Projekte der Universität Mannheim*. Münster: Waxmann.
- Kastirke, N. & Jennessen, S. (2006): *Die neue Schuleingangsphase als Thema der Schulentwicklung. Forschung, Stolpersteine, Praxisempfehlungen*. Kronach: Schneider Verlag.

- Keith, N. & Frese, M. (2008). Effectiveness of error management training: a meta-analysis. *Journal of Applied Psychology*, 93(1), 59-69. doi:10.1037/0021-9010.93.1.59
- Klieme, E. (2006). Empirische Unterrichtsforschung: aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einleitung in den Thementeil. *Zeitschrift für Pädagogik*, 52(6), 765-773.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts "Pythagoras". In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule, Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127-146). Münster: Waxmann.
- 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. In Bundesministerium für Bildung und Forschung (BMBF) (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43-57). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Kluge, F. (2011). *Etymologisches Wörterbuch der deutschen Sprache* (25. Aufl.). Berlin/New York: De Gruyter.
- KMK - Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2015a). *Empfehlungen zur Arbeit in der Grundschule*. Verfügbar unter: https://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2015/Empfehlung_350_KMK_Arbeit_Grundschule_01.pdf
- KMK - Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2015b). *Förderstrategie für leistungsstarke Schülerinnen und Schüler*. Verfügbar unter: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Foerderstrategie-leistungsstarke-Schueler.pdf
- KMK - Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2015c). *Übergang von der Grundschule in Schulen des Sekundarbereichs I und Förderung, Beobachtung und Orientierung in den Jahrgangsstufen 5 und 6 (sog. Orientierungsstufe)*. Verfügbar unter: http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_02_19-Uebergang_Grundschule-SI-Orientierungsstufe.pdf
- KMK - Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2020). *Bericht zum Stand der Umsetzung der Förderstrategie für leistungsschwächere Schülerinnen und Schüler*. Verfügbar unter: https://www.kmk.org/fileadmin/Dateien/pdf/Bildung/AllgBildung/Individuelle_Foerderung/2020-05-14-Bericht_Foerderstrategie-Leistungsschwache.pdf
- Kounin, J. S. (1976). *Techniken der Klassenführung*. Bern: Huber.
- Krause, A. (2004). Erhebung aufgabenbezogener Belastungen im Unterricht – ein Untersuchungskonzept. *Zeitschrift für Arbeits- und Organisationspsychologie*, 48, 139-147. doi:10.1026/0932-4089.48.3.139
- Kulik, C.-L. C. & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research*, 19(3), 415-428. doi:10.3102/0002831201900341
- Kulik, C.-L. C. & Kulik, J. A. (1984). *Effects of ability grouping on primary school pupils: A meta-analysis*. Paper presented at the annual meeting of the American Psychological Association August, ERIC No. ED 255 329.
- Künsting, J., Post, S., Greb, K., Faust, G. & Lipowsky, F. (2010). Leistungsheterogenität im mathematischen Anfangsunterricht. Ein Risiko für die Leistungsentwicklung? *Zeitschrift Für Grundschulforschung*, 3(1), 46–64.
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Paderborn: Schöningh.
- Kunter, M. & Voss, T. (2011): Das Modell der Unterrichtsqualität in COACTIV: eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Kusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften – Ergebnisse des Forschungsprogramms COACTIV* (S. 85-113). Münster: Waxmann.

- Kuzmina, Y. & Ivanova, A. (2018). The effects of academic class composition on academic progress in elementary school for students with different levels of initial academic abilities. *Learning and Individual Differences*, 64, 43-53. doi:10.1016/j.lindif.2018.04.004
- Larson, R. & Richards, M. (1991). Boredom in the middle school years: blaming schools versus blaming students. *American Journal of Education*, 99(4), 418-443. doi:10.1086/443992
- Lee, O. & Fradd, S. H. (1998). Science for all, including students from non-English-language backgrounds. *Educational Researcher*, 27(4), 12-21. doi:10.3102/0013189X027004012.
- Lehmann, R. H., Peek, R. & Poerschke, J. (2006). *HAMLET 3-4-Hamburger Lesetest für 3. und 4. Klassen*. Göttingen: Hogrefe.
- Leiter, J. (1983). Classroom composition and achievement gains. *Sociology of Education*, 56(3), 126-132. doi:10.2307/2112381
- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 73-102). Heidelberg: Springer Medizin.
- Lleras, C. & Rangel, C. (2009). Ability Grouping Practices in Elementary School and African American/Hispanic Achievement. *American Journal of Education*, 115, 279-304. doi:10.1086/595667
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B. & D'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, 66(4), 423-458. doi:10.3102/00346543066004423
- Lutz, H. & Wenning, N. (2001): Differenzen über Differenz – Einführung in die Debatten. In H. Lutz, H. & N. Wenning (Hrsg.), *Unterschiedlich verschieden. Differenz in der Erziehungswissenschaft* (S. 11-24). Opladen: Leske und Budrich.
- Luyten, H. & Van der Hoeven-van Doornum, A. (1995). Classroom composition and individual achievement: Effects of classroom composition and teacher goals in Dutch elementary education. *Tijdschrift voor Onderwijsresearch*, 20(1), 42-62.
- Maehler, D. & Brinkman, H. U. (2016). *Methoden der Migrationsforschung - Ein interdisziplinärer Forschungsleitfaden*. Wiesbaden: Springer VS. doi:10.1007/978-3-658-10394-1
- Markic, S. & Abels, S. (2014). Heterogeneity and Diversity. A Growing Challenge or Enrichment for Science Education in German Schools?. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(4), 271-283. doi:10.12973/eurasia.2014.1082a
- Martinez, R.S., Aricak, O.T. & Jewell, J. (2008). Influence of Reading Attitudes on Reading Achievement: A Test of the temporal-interaction Model. *Psychology in the Schools*, 45(10), 1010-1022. doi:10.1002/pits.20348
- Marzano, R. J. & Marzano, J. S. (2003). The key to classroom management. *Educational leadership*, 61(1), 6-13.
- Mayringer, H., & Wimmer, H. (2003). *SLS 1-4. Salzburger Lese-Screening für die Klassenstufen 1-4*. Bern: Huber.
- McCombs, B. L. & Whisler, J. S. (1997). *The learner-centered classroom and school: Strategies for increasing student motivation and achievement*. San Francisco: Jossey-Bass.
- McGinnis, J. C., Frederick, B. P. & Edwards, R. (1995). Enhancing classroom management through proactive rules and procedures. *Psychology in the Schools*, 32, 220-224. doi:10.1002/1520-6807(199507)32:3<220::AID-PITS2310320309>3.0.CO;2-4
- Mecheril, P. & Plöber, M. (2011). Diversity und soziale Arbeit. In H.-U. Otto & H. Thiersch (Hrsg.), *Handbuch Soziale Arbeit* (S. 278-287). München: Reinhardt.
- Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelsen.
- Meyer, H. (2006). *Schulinspektion führt nicht automatisch zu Qualitätssicherung: Interview mit der westfälisch-lippischen Direktorenvereinigung*. Verfügbar unter: <http://www.westfaelischedirektorenvereinigung.de/PDF/Jahrestagung/2006/Interview/Schulinspektion.pdf>
- Miller, J. (2009). Teaching refugee learners with interrupted education in science: vocabulary, literacy and pedagogy. *International Journal of Science Education*, 31, 571-591. doi:10.1080/09500690701744611
- Möller, G. (2006): Wie heterogen sind deutsche Schulen im internationalen Vergleich? Sehnsucht nach Leistungshomogenität im deutschen Schulsystem. *SchulVerwaltung Spezial*, 1, 11-12.
- Nettle, D. (1999). *Linguistic diversity*. Oxford: Oxford University Press.

- Nordstrom, C. R., Wendland, D. & Williams, K. B. (1998). "To err is human": an examination of the effectiveness of error management training. *Journal of Business and Psychology*, 12(3), 269-282. doi:10.1023/A:1025019212263
- Opp, G. & Speck-Hamdan, A. (2001). Heterogenität der Schulanfänger - Herausforderung für die Schule. In G. Faust-Siehl & A. Speck-Hamdan (Hrsg.), *Beiträge zur Reform der Grundschule: Schulanfang ohne Umwege. Mehr Flexibilität im Bildungswesen* (S. 175–189). Frankfurt a. M.: Grundschulverband.
- Papanastasiou, E. C. & Zembylas, M. (2002). The effect of Attitudes on Science Achievement: a Study conducted among High school pupils in Cyprus. *International Review of Education*, 48(6), 469-484. doi:10.1023/A:1021334424571
- Peetsma, T., van der Veen, I., Koopman, P. & van Schooten, E. (2006). Class Composition Influences on Pupils' Cognitive Development. *School Effectiveness and School Improvement*, 17(3), 275-302. doi:10.1080/13803610500480114
- Pedota, P. (2007). Strategies for effective Classroom Management in the Secondary Setting. *The Clearing House*, 80(4), 163-166. doi:10.3200/TCHS.80.4.163-168
- Perkins, R. E. & Hill, A. B. (1985). Cognitive and affective aspects of boredom. *British Journal of Psychology*, 76, 221-234. doi:10.1111/j.2044-8295.1985.tb01946.x
- Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. Oxford, England: Viking.
- 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. doi:10.3102/0013198X09332374
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft*, 1(13), 121-148. doi:10.1007/s11618-010-0113-z
- Powell, J. (2006). Special education and the risk of becoming less educated. *European Societies* 8(4), 577-599. doi:10.1080/14616690601002673
- Postlethwaite, T. N. (1988). *The Encyclopedia of comparative Education and National Systems of Education*. Oxford: Pergamon Press.
- Prediger, S. & von Aufschnaiter, C. (2017). Umgang mit heterogenen Lernvoraussetzungen aus fachdidaktischer Perspektive: Fachdidaktische Anforderungs- und Lernstufungen berücksichtigen. In T. Bohl, J. Budde & M. Rieger-Ladich (Hrsg.), *Umgang mit Heterogenität in Schule und Unterricht* (S. 291-308). Bad Heilbrunn: Klinkhardt.
- Prenzel, M. & Burba, D. (2006). PISA-Befunde zum Umgang mit Heterogenität. In G. Opp, T. Hellbrügge & L. Stevens (Hrsg.), *Kindern gerecht werden. Kontroverse Perspektiven auf Lernen in der Kindheit* (S. 23-33). Bad Heilbrunn: Klinkhardt.
- Rach, S., Ufer, S. & Heinze, A. (2013). Learning from errors: effects of teachers' training on students' attitudes towards and their individual use of errors. *PNA*, 8(1), 21-30. doi:10.30827/pna.v8i1.6122
- Rakoczy, K., Klieme, E., Drollinger-Vetter, B., Lipowsky, F., Pauli, C. & Reusser, K. (2007). Structure as quality feature in mathematics instruction of the learning environment vs. a structured presentation of learning content. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report of the DFG Priority Programme* (pp. 101-120). Muenster: Waxmann.
- Reusser, K., Stebler, R., Mandel, D. & Eckstein, B. (2013). *Erfolgreicher Unterricht in heterogenen Lerngruppen auf der Volksschulstufe des Kantons Zürich*. Universität Zürich: Zürich. Verfügbar unter: <https://www.fachportal-paedagogik.de/literatur/vollanzeige.html?FId=1022484#vollanzeige>
- Rheinberg, F. (1980). *Leistungsbewertung und Lernmotivation*. Göttingen: Hogrefe.
- Rjosk, C., Richter, D., Lüdtke, O. & Eccles, J. S. (2017). Ethnic composition and heterogeneity in the classroom: Their measurement and relationship with student outcomes. *Journal of Educational Psychology*, 109, 1188–1204. doi:10.1037/edu0000185
- Rosenberg, M. J. & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change* (pp. 1-14). New Haven: Yale University Press.
- Ruhberg, C. & Porsch, R. (2017). Einstellungen von Lehramtsstudierenden und Lehrkräften zur schulischen Inklusion. *Zeitschrift für Pädagogik*, 63(4), 393-415.

- Salleh, R., Venville, G. J. & Treagust, D. F. (2007). When a bilingual child describes living things: an analysis of conceptual understanding from a language perspective. *Research in Science Education*, 37, 291-312. doi:10.1007/s11165-006-9027-4
- Sandfuchs, U. (1994). Unterricht. In R. Keck & U. Sandfuchs (Hrsg.), *Wörterbuch Schulpädagogik* (S. 339-340). Bad Heilbrunn: Klinkhardt.
- Scharenberg, K. (2012). *Leistungsheterogenität und Kompetenzentwicklung*. Münster: Waxmann.
- Scheele, A. F., Leseman, P. P. M. & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31(1), 117-140. doi:10.1017/S0142716409990191
- Scheerens, J. & Bosker, R. (1997). *The foundations of educational effectiveness*. Berlin: BMBF.
- Scholz, I. (2007). Es ist normal, verschieden zu sein – Unterrichten in heterogenen Klassen. In I. Scholz (Hrsg.), *Der Spagat zwischen Fördern und Fordern: Unterrichten in heterogenen Klassen* (S. 7-23). Göttingen: Vandenhoeck & Ruprecht.
- Segeritz, M., Walter, O. & Stanat, P. (2010). Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland: Evidenz für segmentierte Assimilation? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62, 113-138. doi:10.1007/s11577-010-0094-1
- 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. doi:10.3102/0034654307310317
- Shannon, C. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press. doi:10.1002/j.1538-7305.1948.tb01338.x
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163, 688. doi:10.1038/163688a0
- Slavin, R. E. (1987). Ability grouping and student achievement in primary schools: A best-evidence synthesis. *Review of Educational Research*, 57(3), 293-336. doi:10.3102/00346543057003293
- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60(3), 471-499. doi:10.3102/00346543060003471
- Souvignier, E., Förster, N. & Salaschek, M. (2014). quop: Ein Ansatz internetbasierter Lernverlaufsdiagnostik mit Testkonzepten für Lesen und Mathematik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (Tests und Trends, Bd. 12, S. 239-256). Göttingen: Hogrefe.
- Stanat, P., Rauch, D. P. & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200-230). Waxmann: Münster.
- Statistisches Bundesamt (2019). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2018*. Verfügbar unter: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publicationen/Downloads-Migration/migrationshintergrund-2010220187004.pdf?__blob=publicationFile#page=46
- Steenbergen-Hu, S., Makel, M. C. & Olszewski-Kubilius, 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 Meta-Analyses. *Review of Educational Research*, 86(4), 849-899. doi:10.3102/0034654316675417
- Sweller, J., van Merriënboer, J. J. G. & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296. doi:10.1023/a:1022193728205
- Tam, M. Y. S. & Bassett, G. W. (2004). Does diversity matter? Measuring the impact of high school diversity on freshman GPA. *Policy Studies Journal: The Journal of the Policy Studies Organization*, 32, 129-143. doi:10.1111/j.0190-292X.2004.00056.x
- Theunissen, U., Schoppengerd, E., Witzke, F., Endell, A. & Wenning, F. (2005). *Duisburger Sprachstandstest. Gesamtschule Duisburg-Meiderich*. Unveröffentlichtes Testinstrument.
- Tillmann, K.-J. (2004). System jagt Fiktion. Die homogene Lerngruppe. *Friedrich Jahresheft*, 22, 6-9. doi:10.1007/978-3-531-90972-1_2
- Tillmann, K.-J. (2008). Viel Selektion – wenig Leistung: Erfolg und Scheitern in deutschen Schulen. In R. Lehberger & U. Sandfuchs (Hrsg.), *Schüler fallen auf. Heterogene Lerngruppen in Schule und Unterricht* (S. 62-78). Bad Heilbrunn: Klinkhardt.

Literatur

- Trautmann, M. & Wischer, B. (2011). *Heterogenität und Schule. Eine kritische Einführung*. Wiesbaden: Springer.
- Treinius, G. & Einsiedler, W. (1996). Zur Vereinbarkeit von Steigerung des Lernleistungsniveaus und Verringerung von Leistungsunterschieden in Grundschulklassen. *Unterrichtswissenschaft*, 24(4), 290–311.
- Tsung, L. (2012). Rethinking Multilingual Education for Minority Students in China. In A. Yiakoumetti (Ed.), *Harnessing Linguistic Variation to Improve Education*. 5th ed. (pp. 115-138). Bern: Peter Lang.
- Tulis, M. (2013). Error Management Behavior in Classrooms. Teachers responses to student mistakes. *Teaching and Teacher Education*, 33, 56-68. doi:10.1016/j.tate.2013.02.003
- van Ackeren, I. & Kühn, S. (2017). Homogenität und Heterogenität im Schulsystem. In T. Bohl, J. Budde & M. Rieger-Ladich (Hrsg.), *Umgang mit Heterogenität in Schule und Unterricht* (S. 175-190). Bad Heilbrunn: Klinkhardt.
- Van de Grift, W. (2007). Quality of teaching in four European countries: a review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127-152. doi:10.1080/00131880701369651
- Vigdor, J. & Nechyba, T. (2007). Peer effects in North Carolina public schools. In P. E. Peterson & L. Woessmann (Eds.), *Schools and the equal opportunity problem* (pp. 73-101). Cambridge, MA: MIT Press.
- von der Groeben, A. (2003). Lernen in heterogenen Gruppen. Chancen und Herausforderung. *Pädagogik*, 55(9), 6–9.
- Walgenbach, K. (2017). *Heterogenität – Intersektionalität – Diversity in der Erziehungswissenschaft* (2. Aufl.). Opladen & Toronto: Barbara Budrich.
- Walter, O. & Taskinen, P. (2007). Kompetenzen und bildungsrelevante Einstellungen von Jugendlichen mit Migrationshintergrund in Deutschland: Ein Vergleich mit ausgewählten OECD-Staaten. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 337-366). Münster: Waxmann.
- Walter, O. & Taskinen, P. (2008). Der Bildungserfolg von Jugendlichen mit Migrationshintergrund in den deutschen Ländern. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2006 in Deutschland: Die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 343-374). Münster: Waxmann.
- Weinert, F. E. (2000). Lehren und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule. *Pädagogische Nachrichten Rheinland-Pfalz*, 2, 1-16.
- Wenning, N. (1999). *Vereinheitlichung und Differenzierung. Zu den ‚wirklichen‘ gesellschaftlichen Funktionen des Bildungswesens im Umgang mit Gleichheit und Verschiedenheit*. Opladen: Leske & Budrich.
- Wenning, N. (2004). Heterogenität als neue Leitidee der Erziehungswissenschaft? Zur Berücksichtigung von Gleichheit und Verschiedenheit. *Zeitschrift für Pädagogik*, 4, 565-582.
- Wenning, N. (2007). Heterogenität als Dilemma für Bildungseinrichtungen. In S. Boller, E. Rosowski & T. Stroot (Hrsg.), *Heterogenität in Schule und Unterricht. Handlungsansätze zum pädagogischen Umgang mit Vielfalt* (S. 21-31). Weinheim/Basel: Beltz.
- Wright, S. P., Horn, S. P. & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57-67. doi:10.1023/A:1007999204543
- Zimmer, R. W. & Toma, E. F. (2000). Peer Effects in Private and Public Schools across Countries. *Journal of Policy Analysis and Management*, 19(1), 75-92. doi:10.1002/(SICI)1520-6688(200024)19:1<75::AID-PAM5>3.0.CO;2-W

Abbildungsverzeichnis

Abbildung 1	Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts (Helmke, 2017)	7
Abbildung 2	Stufenmodell der Unterrichtsqualität nach Pietsch (2010) (vgl. Bohl, 2017)	14
Abbildung 3	Arbeitsmodell der vorliegenden Arbeit	23
Abbildung 4	Arbeitsmodell Leistungsheterogenitätsmaße	25
Abbildung 5	Arbeitsmodell Sprachliche Heterogenität	27
Abbildung 6	Arbeitsmodell Einstellung	29

Anhang A: Manuskript zu Beitrag 1 – Leistungsheterogenitätsmaße

Do students benefit from homogeneous or heterogeneous classes? An analysis of linear and non-linear effects of within-class heterogeneity on students' reading performance using different heterogeneity measures

1 Introduction

Teachers find it highly challenging to manage a class in which students vary greatly in their ability. However, such challenges have increased in recent years because classrooms became more heterogeneous for two main reasons. First, schools are seeing a growing proportion of students with a migration background, which led to greater differences in language skills within classrooms (Jensen & Rasmussen, 2011; Maehler & Brinkman, 2016), and second, students with special educational needs are more often being included in general education, which increased the classroom heterogeneity of learning prerequisites (Markic & Abels, 2014; Powell, 2006). In the near future, this increase in heterogeneity might be further enhanced by effects of the COVID-19 pandemic. According to recent representative data, 82% of the teachers call the compensation for learning differences and possible learning deficits to be by far the greatest challenge in the reopening phase of the schools (Forsa, 2020).

Regardless of teachers' challenges with heterogeneous classrooms, practitioners, politicians, and researchers, have hotly debated whether students encounter better learning conditions in homogeneous or heterogeneous classes. While some argue that grouping students according to their ability results in educational inequality and impairments in students' learning, others emphasize benefits of homogeneous learning groups for all students (e.g., Francis et al., 2017; Kerckhoff, 1986; Oakes, 2005; Slavin, 1987, 1990). These opposite positions are mirrored by the state of empirical findings. Despite this topic being studied for almost one hundred years, the state of research is still contradictory (e.g., Hattie, 2009; Luyten & Van der Hoeven-van Doornum, 1995; Slavin, 1987; Steenbergen-Hu et al., 2016;

Vigdor & Nechbya, 2007; Wright et al., 1997). To date, it is unclear whether heterogeneity itself has positive or negative effects on student learning or whether it is influential at all.

Therefore, the goal of our study is to better understand the relationship between heterogeneity and student learning by expanding the existing approaches used in heterogeneity research. So far, studies have used the intra-class standard deviation to calculate heterogeneity and reported linear analyses only (except Scharenberg, 2012); addressing heterogeneity using various measures and analyzing nonlinear relationships has been left out. However, using several heterogeneity measures can expand the present understanding of how heterogeneity affects student learning. Furthermore, the relation between heterogeneity and student learning might be non-linear with beneficial effects for either an average level of heterogeneity (inverted U-shaped relationship) or with beneficial effects of either very high or very low heterogeneity (U-shaped relationship). Therefore, this study compares various measures of heterogeneity using both, linear and nonlinear analyses in order to gain deeper insights into the meaning of heterogeneity and its' effects on learning. For this purpose, we use several exploratory approaches, which is an exceptional feature about our study.

Our first aim is to investigate whether and how the different heterogeneity measures are related to student achievement. That is, we want to find out whether the standard deviation is sufficient to describe effects of heterogeneity or whether different heterogeneity measures provide different information. Our second aim is to investigate potential nonlinear relationships between the heterogeneity measures and student learning across classrooms. That is, we want to find out whether linear analyzes are sufficient to describe effects of heterogeneity or whether nonlinear analyzes provide different information. To examine these aims we use longitudinal data with eight measurement points and a sample of $N = 2,605$ students from 124 classes.

2 Heterogeneity and its' Relevance for Student Learning

2.1 Defining Heterogeneity

Heterogeneity can be derived from ancient Greek (*héteros* = different; and *génos* = class, species) and refers to the differences and variations between individuals of a group (von der Groeben, 2003). An important characteristic of heterogeneity is its relativity. Heterogeneity does not exist per se, but only

with regard to a reference value (i.e. homogeneity). Heterogeneity constitutively needs a *tertium comparationis* (Wenning, 1999), meaning that it is inseparably connected with homogeneity.

Heterogeneity relates to a number of factors, such as social, cultural, ethnic, or performance-related differences (e.g., Gröhlich, Scharenberg & Bos, 2009). Performance-related heterogeneity can be defined as the distribution of “population elements” (Liebersohn, 1969, p. 851; cited in Budescu & Budescu, 2012) along a continuum of homogeneity to heterogeneity with respect to an ability or performance-related variable (cf. Teachman, 1980). Accordingly, students in rather homogeneous classes have similar cognitive abilities or performance levels, and students in rather heterogeneous classes have different cognitive abilities or performance levels. In the following, we use the term *heterogeneity* to express performance-related heterogeneity, the within-class variability of students’ performance.

2.2 Effects of Heterogeneity on Student Learning

Research on heterogeneity as well as research on ability grouping (i.e. a strategy for schools to systematically shape heterogeneity in their classes; e.g. Huang, 2008; Slavin, 1987) plays a key role in a large number of disciplines, such as economics (e.g., Burke & Sass, 2013; Carman & Zhang, 2012), sociology (Dreeben & Barr, 1988; Leiter, 1983), political (e.g., Vigdor & Nychbar, 2007) and educational research (e.g., Kuzmina & Ivanova, 2018; Lleras & Rangel, 2009); and its’ effect on student learning has been investigated for nearly 100 years all over the world. Despite the impressive body of literature on heterogeneity and ability-grouping, the state of research has not provided clear outcomes: studies have reported positive effects (e.g., Kuzmina and Ivanova, 2018; Scharenberg, 2012; Vigdor, & Nychbar, 2007), negative effects (e.g., Kulik & Kulik, 1982, 1984; Lou et al., 1996; Luyten & Van der Hoeven-van Doornum, 1995; Steenbergen-Hu, Makel & Olszewski-Kubilius, 2016), and non-significant effects (e.g., Bellin, 2009; Carman & Zhang, 2012; Decristan et al., 2017; Leiter, 1983; Slavin, 1987, 1990; Wright, Horn, & Sanders, 1997).

One reason for the inconsistent state of research may be due to systematic differences in the samples. As explained above, heterogeneity is a relative construct, so that the same class composition can have a different degree of heterogeneity depending on characteristics of the sample. One and the same class composition can be homogeneous in a sample with rather heterogeneous classes, while it can

be heterogeneous in a sample with rather homogeneous classes. In Germany, for example, after primary school, students join a school form according to their achievement level (e.g., *Gymnasium* = highest form of school, *Realschule* = middle form of school, *Werkrealschule/Hauptschule* = lowest form of school). Thus, samples with elementary school classes are more likely to include rather heterogeneous classes and samples with secondary school classes are more likely to include rather homogeneous classes. Furthermore, samples with elementary school classes may have a wider range of heterogeneity than samples with secondary school classes. This leads to the consideration that heterogeneity does not necessarily have a linear effect on student learning; a nonlinear relationship is also conceivable. Depending on whether a sample has heterogeneous or homogeneous classes, positive or negative correlations can be found; and samples with a wide range of heterogeneity (e.g. elementary school classes, non-selective secondary school forms such as in England, the U.S. or Australia; Carl, 2017; Dupriez, Dumay, & Vause, 2008; Postlethwaite, 19988) could find nonlinear relationships.

The explanation of the inconsistency is additionally complicated by the fact that there are no theoretical assumptions about how exactly heterogeneity should affect student learning. In theory, it is assumed that heterogeneity influences student learning both directly and indirectly (Helmke, 2017). However, while the indirect effects are based on theoretical assumptions (e.g. teacher behavior, student motivation), there is no clear consideration of how the direct effect of heterogeneity on student learning can be explained. Thus, it remains open whether heterogeneity has positive or negative direct effects on student learning and whether this relationship is linear or non-linear.

2.3 Heterogeneity Measures

To date, heterogeneity has always been operationalized using the standard deviation. However, there are no conceptual or methodological reasons that state that heterogeneity is measured most adequately by the standard deviation. On the contrary: different measures of heterogeneity have different methodological strengths as well as different conceptual focuses. Three measures of heterogeneity are presented below: the within-class standard deviation (SD), the range, and the inter-quartile range (IQR).

From a methodological perspective, the SD – as the sum of the squared differences – represents the average extent to which each student differs from the average performance level in class (Welkowitz,

Ewen, & Cohen, 1971). The SD is a solid measure, as it refers to the whole distribution of values and it is robust against outliers given enough power on the class level (Bühner & Ziegler, 2017). The SD captures the entire student body. Even if students who highly differ from the average performance level in the class affect the SD more than students who differ less, each students' performance contributes to the SD. Thus, the SD assesses an aspect of heterogeneity, that is based on all students within the classroom. When interpreting the SD as a heterogeneity measure, this would mean, that each student has a similar influence on how the classes' heterogeneity affects learning.

A more parsimonious heterogeneity measure is the students' range of performance levels. The range reflects the difference between the highest and the lowest achieving student in a class. Compared to the SD, the range is calculated by only two values, which decreases the likelihood of biased results due to missing values (however, if missing values were needed to calculate the range, the range is more affected by outliers than other heterogeneity measures). Heterogeneity measured by the range focuses on the extreme values of the class. According to the range, heterogeneity is defined as the differences between the highest and the lowest performing student; and it assumes, that these two students within the class have such a strong influence on class heterogeneity, that the more average performing students become less meaningful. Consider two classes A and B with each including 20 students, who are participating in the same test with a maximum of 100 points: In both classes the highest performing student has 90 points and the lowest one 10 points. Accordingly, the range of both classes, A and B, would be 80 points. However, while in class A all of the remaining 18 students had a test-score of 50 points, in class B the remaining students differ in their scores from 20 to 80 points. Regardless, the range views heterogeneity as having the same effect – independent of how the rest of the class deviates in performance. Therefore, when interpreting the range as a heterogeneity measure this would mean, that the influence of heterogeneity on students' learning depends on the two students with the most extreme values.

The third heterogeneity measure, the IQR, is defined by the difference between the average 50% of a class, i.e. the difference between the students of the 75% and 25% quartile. Like the range, the IQR is a parsimonious heterogeneity measure and its' calculation is based on only two values, making it

fairly robust towards outliers and missing values. In comparison to the range, the IQR puts a particular focus on the average 50% of the students in the class. It assumes, that the variability between the average performing students affects learning and that very low or very high performing students in a class are less meaningful for effects of heterogeneity. Again, consider two classes A and B each including 20 students participating in the same test with a maximum of 100 points: In both classes, the students who mark the 25% quartile have 30 points; and those who mark the 75% quartile have 70 points. Accordingly, in both classes A and B the IQR is 40 points. However, while in class A the highest performing student has 80 points and the lowest one 20 points, in class B the highest performing student has 90 points and the lowest one 10 points. Regardless, the IQR views heterogeneity as having the same effect – independent of how the rest of the class deviates in performance. Therefore, when interpreting the IQR as a heterogeneity measure, this would mean, that the influence of heterogeneity on students' learning depends on the differences in a classes center.

In summary, each of the heterogeneity measures has its' own methodological strengths and from an empirical research perspective, it might be reasonable to provide several heterogeneity measures in order to increase the reliability of results. Furthermore, using different heterogeneity measures is also reasonable from a from a conceptual point of view, because not only should research consider other measurements as valid and constructive in the discussion of the topic, but it can also be questioned whether the standard deviation is the most suitable measure for a classes heterogeneity.

3 The Present Study

In this study, we aim to increase our understanding of the effects of heterogeneity on student learning by expanding existing approaches used in educational research. Therefore, we analyze whether the effects of heterogeneity on student learning vary depending on different heterogeneity measures and explore those effects not only in linear but also in nonlinear models.

We investigate heterogeneity effects within grade three and grade four on reading fluency and reading comprehension using two different standardized assessments, respectively: one standardized paper-pencil test to assess prior ability and one computer-based learning progress assessment (LPA) to measure student reading progress (Förster & Souvignier, 2011). The latter assesses students reading

fluency and reading comprehension at eight measurement points across one school year, leading to highly reliable estimates of student learning. We focus our analyses on reading, a domain that represents one of the core learning aims in primary education, as it is highly relevant for future learning and performance in other subjects (Fuchs et al., 2006; Lee & Fradd, 1998).

We examine the following research questions. First, is students' reading performance affected by the heterogeneity in class using the linear SD as a heterogeneity measure? Second, do effects vary using different heterogeneity measures? And third, do effects vary depending on using linear or nonlinear modeling?

Based on the state of research, there is no definite direction of how heterogeneity is supposed to influence students' learning; and given that prior research has operationalized heterogeneity exclusively as the SD using linear modeling, our research questions are exploratory in nature.

4 Method

4.1 Participants & Design

Data were obtained through a large longitudinal study on reading fluency and reading comprehension in grade three and four in Germany (Author, 2014; Author, 2015b; Author, 2018b). The initial sample consisted of 4,350 students from 195 classes. The studies included learning progress assessment (LPA), but not all classes participated in LPA. Since LPA data are needed for the examination of our research questions, we excluded the non-LPA classes ($N = 1,408$) from the initial sample. We also excluded data from 59 students whose parents did not provide consent for study participation, data from 5 classes in which the parental consent rate was lower than 90%, and 8 intergrade classes, in which students from grade three and four studied together.¹ Thus, data from $N = 2,605$ students from 124 classes were analyzed (49% female; $M_{\text{age}} = 9.16$, $SD = .69$; 63% third grade). Of this subsample with only LPA classes, some classes received further treatment. This is covered in chapter Limitations and future Prospects (see chapter 6.1) and controlled for in our analyses.

¹ We rerun all analyses including the intergrade classes and provide the results in the *SM 01 – Detailed Results* in the Supplementary Material available online. No differences depending on the sample were found (except: non-significant linear effect of IQR on reading comprehension).

Data were assessed in 52 schools in and around a medium-sized German town with an average class size of 21 students. Regarding language, 77% of the students spoke only German at home, 14% spoke German and another language at home and for 9% of the students their family language was not German. At the beginning of each school year, students' prior ability in reading fluency and reading comprehension was assessed with two standardized paper-pencil achievement tests. During the school year students participated in LPA. LPA required students to complete eight short computer-based tests at intervals of three weeks. The LPA-tests measured both, progress in reading fluency and progress in reading comprehension.

4.2 Measures

4.2.1 Independent Variables: Reading Fluency and Reading Comprehension

Initial reading fluency was assessed using the Salzburger Reading-Screening (SLS) for grade 1-4 (Mayringer & Wimmer, 2003). The SLS measures the level of students' basal reading skills, i.e. whether students read words correctly, relatively fast, and effortlessly. The SLS requires students to read a list of easy sentences (e.g. "bananas are blue") and to decide whether the sentence is correct. The raw test value is the number of correctly rated sentences within three minutes.

Initial reading comprehension was assessed using an adapted version of the Hamburger Reading Test (HAMLET 3 – 4; Lehmann, Peek, & Poerschke, 2006) for grades three and four. This short version (HAMLET-S) consisted of one non-continuous and two continuous texts and a total of 22 multiple-choice items (Author, 2014; Author, 2018a). Questions either addressed the surface structure, the text-base or the situation model.

4.2.2 Dependent Variables: Reading Fluency and Reading Comprehension

During the school year, students' reading performance in reading fluency and reading comprehension was assessed using LPA (Author, 2014; Author 2015a). Every three weeks, students completed one out of eight computer-based tests. Each test started with a maze task assessing students' reading fluency. In the maze task every seventh word was missing and replaced with a choice of three words. All tests in grade three had about 170 words and 17 gaps, while in grade four, the texts consisted of about 240 words and 24 gaps. Students were instructed to read as quickly and as accurately as possible. Following the

maze task, students answered 12 comprehension questions (16 in fourth grade) asking for information explicitly mentioned in the text or requiring the students to make inferences from the text. When answering the comprehension questions, the complete text was visible.

For each item in the maze task and for each answer to the comprehension questions, accuracy and response time (in milliseconds) were assessed and integrated following the procedure described by Richter and van Holt (2005). We first applied a reciprocal transformation by dividing 1 by the response times to obtain a normal distribution for the response times. Next, we computed the sum of the reciprocally transformed response times for those items that were answered correctly. Thus, any item that was answered correctly increased the scale score even more the faster the correct answer was given. Items that were answered incorrectly did not contribute to the scale value, independent of the time needed to make the decision. Consequently, scores and scale values are higher the more efficient the reading process was. We computed the mean of the accuracy-response-time quotients to account for the different numbers of items for third and fourth grade and finally multiplied the scores by 60 to gain higher scores for further analyses and thereby prevent convergence problems. Accuracy and response time data from the maze task were used to compute a measure of reading fluency, and the respective information from the comprehension questions were used to compute a measure of reading comprehension.

4.3 Heterogeneity Measures

For each class, we calculated the three heterogeneity measures SD, range and IQR using the initial reading fluency (SLS) and reading comprehension (HAMLET-S) data.

4.4 Data Analysis

At first, we ran the likelihood ratio test to investigate whether accounting for autocorrelation and heteroscedasticity is necessary in our models (Bliese & Ployhart, 2002; Pinheiro & Bates, 2000). To account for the multilevel structure of the data, we applied a four-level longitudinal multilevel regression analysis with the four levels being measurement points, students, classes, and schools. We estimated all longitudinal multilevel models using the package nlme of the software R (Pinheiro et al., 2018). Given that our research questions address the individual development of students, class-level variables were z-

standardized, and individual test scores were centered at the group mean and afterwards z-standardized (Enders & Tofighi, 2007).

For both, reading fluency and reading comprehension we investigated seven models to test the effect of heterogeneity on students' reading performance (see Tables 4 and 5). Model 0 is the baseline model and examines the impact of students' background variables, namely gender, migration background (i.e. spoken language at home; 0 = German, 1 = German and other languages or other languages only), and students' pre-test scores. On class-level we included the variables grade and average class achievement. Also, five dummy variables were added to control for the respective further treatment (for detailed results see Supplementary Material 01). The remaining six models included the different heterogeneity measures as linear and non-linear predictors of student learning to investigate effects of heterogeneity and explore possible differences depending on the respective operationalization (Research Question 1-3). Formula 1 shows the complete measurement model:

$$[Level\ 1:] \quad Y_{ijkl} = \beta_{0jkl} + \beta_{1jkl}X_1 + u_{ijkl}$$

$$[Level\ 2:] \quad \beta_{0jkl} = \beta_{00kl} + \beta_{01kl}W_1 + u_{jkl}$$

$$[Level\ 3:] \quad \beta_{00kl} = \beta_{000l} + \beta_{001l}V_1 + u_{kl}$$

$$[Level\ 4:] \quad \beta_{000l} = \beta_{0000} + u_1$$

With:

Level-1 variables (i.e., X_1) = the eight measurement points.

Level-2 variables (i.e., W_1) = students' gender, migration background and pretest achievement.

Level-3 variables (i.e., V_1) = the classes' average achievement, linear heterogeneity measure (SD, Range, IQR), non-linear heterogeneity measure (SD², Range², IQR²), grade and five further treatments.

[Formula 1]

The measurement model considers that the eight measurement points (Level 1) are nested within students (Level 2), which are nested within classes (Level 3), which are nested within schools (Level 4). No Level-4 variables were included.

4.5 Missing Data

An analysis of missing data reported a missing value average of 5.6%. We decided to impute for reading variables, since the missing values in this data exceeded 5%. We singly imputed using the R package norm (Novo & R Core Team, 2013). The prediction was based on students' values in gender, age, grade, and migration background (i.e. spoken language at home; 0 = German, 1 = German and other languages, other languages only).

5 Results

5.1 Descriptive Results

Tables 1 and 2 provide descriptive statistics and correlations for reading fluency and reading comprehension for the three levels measurement point, student, and class. Average scores on the student ($M = 35.04$) and class level ($M = 35.07$) are similar, indicating that the average student represents the average class regarding reading fluency and comprehension. A comparison of the standard deviations at the student ($SD = 11.40$) and class level ($SD = 7.09$) shows that the standard deviation at class level is smaller than at student level. Thus, students' reading ability is more similar to those of their classmates than to students from different classes, which highlights the importance of applying multilevel modeling. Table 1 reveals significant positive correlations between the different heterogeneity measures and a classes' mean achievement in reading fluency. Thus, more heterogeneous classes are more likely to perform better in reading fluency. In contrast, mean achievement in reading comprehension shows no significant correlation with any of the heterogeneity measures (see Table 2). Furthermore, the heterogeneity measures showed strong, but also medium correlations (Cohen, 1988), indicating that they represent different aspects of heterogeneity.

Insert Table 1 and 2 about here

5.2 Effects of Heterogeneity on Students' Reading Progress

The likelihood ratio test revealed that accounting for autocorrelation and heteroscedasticity in our models gives the best data fit for both reading fluency and reading comprehension (see Table 3, Model 1-4).

Moreover, the analyses of variance (ANOVAs) revealed that accounting for school affiliation as a fourth level in the analyses fitted the data significantly better for both performance in reading fluency and reading comprehension than a three-level model (see Table 3, Model 5-8).

 Insert Table 3 about here

The unconditional model for reading fluency shows an intraclass correlation coefficient ICC 1 = 19% for classes within schools. Thus, 19% of the variance in reading fluency is explained by belonging to a particular class within a particular school. For reading comprehension, we found an ICC 1 = 14% for classes within schools.

Tables 4 and 5 show the results for the eight models predicting students' performance in reading fluency and reading comprehension, respectively.² Both baseline models show that students' reading fluency and reading comprehension increases throughout the school year (Table 4 and 5, Model 0). As implied by the positive effect of students' pretest scores (reading fluency: $\beta_{\text{Fluency}} = .58, p < .001$; reading comprehension: $\beta_{\text{Comprehension}} = .46, p < .001$), the higher students' reading ability at the beginning of the school year, the more they increased their ability during the year. Our analyses revealed a similar pattern for a classes' mean achievement ($\beta_{\text{Fluency}} = .34, \beta_{\text{Comprehension}} = .28, p < .001$). The higher the average reading ability of a class, the higher the students' learning during the year. Performance in students' reading fluency and reading comprehension was lower for students with migration background ($\beta_{\text{Fluency}} = -.11, \beta_{\text{Comprehension}} = -.15, p < .001$).

 Insert Table 4 and 5 about here

Overall, within-class heterogeneity showed no effects on students' reading fluency using different heterogeneity measures ($\beta_{\text{SD}} = -.02, p = .172$; $\beta_{\text{range}} = -.01, p = .309$; $\beta_{\text{IQR}} = -.02, p = .258$; Table 4, Model

² We provide the coefficients and *p*-values of the respective treatments in the Supplementary Material online. Please see *SM 01 - Detailed Results*.

1-3). Furthermore, results were independent of linear or nonlinear modeling ($\beta_{SD^2} = .02, p = .835$; $\beta_{range^2} = -.06, p = .494$; $\beta_{IQR^2} = .09, p = .174$; Table 4, Model 4-6).

For reading comprehension, however, we found different results for different heterogeneity measures and nonlinear modeling. Linear analyses showed no significant effects of heterogeneity on students' performance in reading comprehension; with one exception, as the IQR positively predicted students' performance in reading comprehension ($\beta_{SD} = .01, p = .403$; $\beta_{range} = -.00, p = .642$; $\beta_{IQR} = .04, p = .047$; Table 5, Model 1-3). Nonlinear analyses revealed significant effects when heterogeneity was assessed as SD ($\beta_{SD^2} = .32, p = .013$; Table 5, Model 4) or IQR ($\beta_{IQR^2} = .23, p = .014$; Table 5, Model 6), but not as range ($\beta_{range^2} = .12, p = .406$; Table 5, Model 5). Figure 1 shows a functional graph of the nonlinear relationship between students' performance in reading comprehension and heterogeneity measured as SD and IQR.

 Insert Figure 1 about here

Heterogeneity is represented along the x-axis, students' performance in reading comprehension along the y-axis. The span of the axes was adapted to the descriptive statistics of the z-transformed variables. The graph shows that performance in reading comprehension is lowest in classes of moderate heterogeneity and increases with increasing and decreasing values of heterogeneity.

6 Discussion

Our analyses revealed three important results. First, heterogeneity measured by the standard deviation does not linearly predict students' performance in reading achievement. Second, the pattern of results was quite unaffected by the operationalization of heterogeneity. Third, curvilinear analyses for reading comprehension revealed that higher learning is associated with an increasing or decreasing amount of heterogeneity.

Students' reading performance is not affected by heterogeneity measured by the standard deviation in linear models, which is in line with several previous studies (e.g., Carman & Zhang, 2012; Decristan et al., 2017; Leiter, 1983; Wright, Horn, & Sanders, 1997). Using different heterogeneity

measures supported the finding, that heterogeneity does not play a significant role for students' learning when modeled linearly (except IQR for reading comprehension). However, modeling non-linear relations between heterogeneity and student learning lead to significant results for reading comprehension. Accordingly, students' performance has its minimum in classes of moderate heterogeneity and grows with increasing or decreasing values of heterogeneity.

However, non-linear effects were only found for reading comprehension and did not appear for reading fluency. As reading fluency and reading comprehension represent important but different aspects of reading ability (e.g., Watson et al., 2003), we assume that these differences might explain the differences in heterogeneity effects. While reading fluency is an automatized and rather unconscious process, reading comprehension is characterized by the active construction of a situational model (i.e. a mental representation of the information described in the text). To comprehend a text, the reader needs to extract relevant information from the text and connect it to prior knowledge to build up the situational model. Consequently, methods to foster reading fluency and reading comprehension differ. Reading fluency is efficiently increased by methods that require students to (repeatedly) read aloud to automatize the word recognition (e.g., repeated reading, paired reading; Topping, 2006). Reading comprehension is fostered by cognitive and metacognitive reading strategies that help to identify relevant information, to connect new information to prior knowledge, and to monitor one's own reading comprehension (e.g., Author, 2006). Since the constructs of fluency and comprehension are fundamentally different, heterogeneity might also differ in its' impact on students' learning in reading fluency and students' learning in reading comprehension.

Furthermore, nonlinear effects are only found for heterogeneity measured by the nonlinear SD and the nonlinear IQR, but not for the nonlinear range. As stated above, the SD and IQR are less biased heterogeneity measures, because they are less dependent on outliers. In contrast, the range is strongly affected by outliers. This could lead to a bias in its' heterogeneity measure and thus, to a bias in its' influence on students learning. Furthermore, the range – which defines heterogeneity by the most two extreme values within a class – might not be a proper measure to represent a classes' heterogeneity.

According to the range, two students are the determining factor for heterogeneity and this kind of conceptual perspective may not be appropriate for everyday school life. Even though it is important to consider especially high and low performing students, it may not be enough to only focus on those two students having the highest performance-related differences. Therefore, effects of heterogeneity on students' learning might be represented best, if the measure take into account each student of a class (SD) or the differences within the class center (IQR).

To date, no nonlinear effects of heterogeneity on student learning have been investigated (except Scharenberg, 2012). Our results indicate that heterogeneity nonlinearly affects students' performance in reading comprehension – with lowest learning in moderately heterogeneous classes. Although our research is exploratory in nature, we want to interpret those nonlinear results more deeply. In heterogeneity research most of the studies do not provide teacher variables (e.g. extent of differentiation used while teaching). However, heterogeneity might not only directly affect student learning, but also influence instruction, which in turn affects students' learning (cf. Decristan et al., 2017; Helmke, 2017). We assume that teachers of moderately heterogeneous classes face difficulties in making instructional decisions because the heterogeneity of the class does not demand an obvious course of action. It is easier for a teacher to decide which teaching methods to choose with a class composition that is clearly homogeneous or heterogeneous. Students of homogeneous classes show similar learning prerequisites. Teaching in such classes would support choosing teacher-managed instruction, such as inquiry-based teaching or direct instruction (Decristan et al., 2015; Minner, Levy, & Century, 2009). Additionally, even if those methods have been found to be very effective in general (Hattie, 2009; Tarelli et al., 2012) they might be more effective in homogeneous classes instead of moderately heterogeneous ones. Although those methods can be adapted by the teacher, they do not include methods of differentiation between students of different performance levels. Therefore, teachers of moderately heterogeneous classes might have more difficulties keeping each student cognitively active. On the contrary, students of heterogeneous classes show different learning prerequisites, which underlines the need of differentiated instruction. Teachers of highly heterogeneous classes might be more likely to recognize, that they need to use dif-

differentiated instruction (e.g., cooperative learning methods such as reciprocal teaching or learning tandems; Helmke, 2017; Palincsar & Brown, 1984; Slavin, 1987) in order to meet individual needs. Furthermore, even if differentiated instruction has been found to promote similar or even higher reading achievement in general (Connor et al., 2013; Little, McCoach, & Reis, 2014), it might be more effective in heterogeneous classes instead of moderately heterogeneous ones: research showed that teachers struggle with the implementation of differential instruction (Author, 2018a; Fuchs & Vaughn, 2012; Philipp, 2013) and revealed that collaborative learning is not effective until settings are frequently practiced, and processes are clear (Krause, 2004). Teachers of moderately heterogeneous classes may be less likely to recognize the necessity of differential instruction and even if – they might not use it often enough to actually have a positive impact on students' learning.

Although our study highlights the effect of heterogeneity on students' learning, the strongest predictor was previous individual reading achievement. Accordingly, high performing students showed greater learning within one school year than low performing students. Considering a classes' mean achievement, these so-called Matthew effects not only occurred on the individual level but also on the class level. This finding is in line with studies showing that mean achievement positively predicts students' learning (e.g., Bellin, 2009; Dar & Resh, 1986; Burke & Sass, 2013). In this regard, the comparison of the two beta weights mean achievement and heterogeneity is of particular interest: mean achievement has a much higher impact on student learning for both reading fluency and reading comprehension than heterogeneity. Furthermore, heterogeneity increased the proportion of explained variance by only about 2%. This is remarkable, as discussions about today's educational challenges and about the challenges in terms of the COVID-19 pandemic concern mainly heterogeneity – neglecting the average level of achievement within classes.

Finally, we have one more point that we would like to raise in this discussion. Even if our research questions investigate heterogeneity effects on students' reading performance we also ran several models to investigate heterogeneity effects on students learning progress, adding the respective interaction term between measurement point and heterogeneity. Results revealed, that students' learning progress does neither depend on any of the linear heterogeneity measures nor on any of the nonlinear heterogeneity

measures. Thus, heterogeneity does only affect students' average performance but does not affect students' learning progress. This finding has a specific value for research, as most study results are based on two measurement points (and studies that used longitudinal data did not calculate growth models; e.g. Vigdor & Neychba, 2007); but still some authors interpret the results in the light of learning progress and growth.

6.1 Limitations and future Prospects

When interpreting the results of our study, some limitations need to be considered. First, all classes whose data were reanalyzed in this study applied LPA, thus, teachers received information about students' learning eight times a year. On the one hand, teachers received more assessment information than teachers usually have. On the other hand, however, the LPA data enabled us to estimate student learning based on longitudinal data with eight points of measurement and thus with increased reliability (e.g., Singer & Willett, 2003; Willett, 1989). Thus, using the multiple LPA assessments to model learning is a clear advantage of our study. Second, some of the classes received further treatment next to LPA via either a student goal-setting procedure (13 classes) or additional worked-out reading material to differentiate instruction (36 classes). Given that these treatments might have affected student reading performance and the latter might have changed teachers' response to within-class heterogeneity, we checked the influence of the additional treatments in two ways: When analyzing the data, we controlled for the different conditions by including dummy variables into our models. Moreover, we reran the same analyses excluding all classes that received worked-out material to differentiate instruction. Analyses indicated no effects of the treatment groups: like the presented results, the subset consisting of 1884 students from 88 classes showed the same pattern of results with no linear effects of within-class heterogeneity (except: non-significant linear effect of IQR on reading comprehension) and a non-linear effect for SD and IQR for reading comprehension ($\beta_{SD^2} = .35, p < .05$; $\beta_{IQR^2} = .24, p < .05$). Third, we analyzed effects of class composition on individual reading growth with regard to mean achievement and heterogeneity, but other class composition variables that were not included in our study (e.g. social skills) might play a role, too. Also, since effects of class composition do not only affect student learning directly but are mediated or moderated by instruction, including information about instruction would help to understand

how class composition affects student learning. Fourth, we used data from primary education because those schools encounter the highest amount of heterogeneity in the German school system. Nevertheless, effects of class composition should also be evaluated in secondary education. Finally, our study is limited to composition effects within one school year. Effects of heterogeneity might increase over time or have a cumulative impact in later school years.

6.2 Conclusion and Practical Implications

In summary, this study strengthens the previous approach to measure heterogeneity as SD. However, it states, that future research should consider nonlinear relationships between heterogeneity and student learning. Furthermore, this study underlines, that using several measures leads to a more holistic conceptual picture of heterogeneity, but the empirical analysis of different heterogeneity measures leads to similar results. With regard to heterogeneity being considered as one of the greatest challenges in education, our findings reveal that heterogeneity is empirically less important than previous performance and a classes' mean achievement. Analyzing nonlinear relationships our results indicate that – in contrast to public discussion and teacher perceptions – the biggest challenge for teachers might not be teaching in highly heterogeneous classes but in classes where the average ability level is low and in classes with a moderate level of heterogeneity.

Moreover, our study provides implications for teacher education. To date, university students are taught that heterogeneity is measured by the standard deviation; in addition, international comparative studies such as PISA or TIMSS use the standard deviation as a heterogeneity measure and thus, teachers receive feedback about the heterogeneity of their class through the heterogeneity measure SD. Given the methodological and conceptual differences between the heterogeneity measures identified in this study, it would be important to systematically analyze the construct heterogeneity with teachers and to discuss why heterogeneous classes can achieve higher performance than moderately heterogeneous classes.

Table 1

Means, Standard Deviations, and Correlations for Reading Fluency

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Level 1 (time)											
(1) RF (LPA) ^a	20840	5.29	2.27	-							
Level 2 (student)											
(2) RF (SLS)	2605	35.04	11.40	.70*	-						
Level 3 ^b (class)											
(3) Mean RF	124	35.07	7.09	.38*	.62*	-					
(4) SD RF	124	9.02	1.75	.10*	.20*	.30*	-				
(5) SD ² RF	124	84.41	33.87	.10*	.20*	.30*	.99*	-			
(6) R RF	124	34.69	8.22	.08*	.17*	.26*	.85*	.85*	-		
(7) R ² RF	124	1270.18	629.92	.07*	.16*	.25*	.84*	.84*	.99*	-	
(8) IQR RF	124	11.31	3.47	.06*	.13*	.21*	.69*	.68*	.34*	.32*	-
(9) IQR ² RF	124	139.88	84.3	.06*	.13*	.22*	.71*	.70*	.37*	.35*	.98*

Note. RF = Reading fluency; LPA = learning progress assessment; SLS = pretest scores in the standardized reading fluency test; SD = standard deviation; R = range; IQR = interquartile range.

^aAverage test scores measured over the eight tests for reading fluency; ^bLevel 3 data concern pretest scores.

* $p < .05$.

Table 2
Means, Standard Deviations, and Correlations for Reading Comprehension

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Level 1 (time)											
(1) RC(LPA) ^a	20840	3.24	1.82	-							
Level 2 (student)											
(2) RC (HAMLET)	2605	9.10	5.20	.56*	-						
Level 3 ^a (class)											
(3) Mean RC	124	9.11	2.63	.29*	.50*	-					
(4) SD RC	124	4.52	0.8	.04*	.07*	.14	-				
(5) SD ² RC	124	21.05	7.23	.04*	.07*	.14	.99*	-			
(6) R RC	124	16.44	2.97	-.01	.04*	.07	.82*	.80*	-		
(7) R ² RC	124	278.92	95.18	-.01	.03	.05	.81*	.79*	.99*	-	
(8) IQR RC	124	5.95	1.68	.06*	.08*	.17	.74*	.76*	.43*	.42*	-
(9) IQR ² RC	124	38.17	21.98	.07*	.08*	.17	.73*	.75*	.41*	.40*	.98*

Note. RC = Reading comprehension; LPA = learning progress assessment; SLS = pretest scores in the standardized reading comprehension test; SD = standard deviation; R = range; IQR = interquartile range.

^aAverage test scores measured over the eight tests for reading comprehension; ^bLevel 3 data concern pretest scores.

* $p < .05$.

Table 3

Model Fit using the Likelihood Ratio Test (Model 1-3) and ANOVAS (Model 4-7)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
	Fluency	Fluency	Comprehension	Comprehension	Fluency	Fluency	Comprehension	Comprehension
	Standard Model	Autocorrelation and Heteroscedas- ticity Model	Standard Model	Autocorrelation and Heteroscedas- ticity Model	Null-Model nested for Classes	Null-Model nested for Classes within Schools	Null-Model nested for Classes	Null-Model nested for Classes within Schools
df	6	14	6	14	3	4	3	4
AIC	55915.02	55255.83	52364.19	52174.87	40686	40666	41243	41222
BIC	55960.74	55362.52	52409.92	52281.55	40709	40697	41266	41252
Likelihood ratio test		(vs. Model 1)		(vs. Model 3)		(vs. Model 5)		(vs. Model 7)
		L.Ra- tio = 675.19 $p < .001$		L.Ra- tio = 205.33 $p < .001$		$\chi^2 = 22.14$ $p < .001$		$\chi^2 = 22.82$ $p < .001$

Table 4

Longitudinal Multilevel Regression Analysis predicting Students' Performance in Reading Fluency

Predictor variables	Model 0 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Intercept	-.58 (.20)	-.64 (.20)	-.64 (.20)	-.61 (.20)	-.64 (.20)	-.65 (.20)	-.61 (.20)
Measurement Point	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)
Within class level							
Female	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)
Mig	-.11* (.02)	-.11* (.02)	-.11* (.02)	-.11* (.02)	-.11* (.02)	-.11* (.02)	-.11* (.02)
Pre-test Fluency	.58* (.03)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)
Between class level							
Grade	.06 (.06)	.07 (.06)	.07 (.06)	.06 (.06)	.07 (.06)	.07 (.06)	.06 (.06)
Mean Achievement	.34* (.03)	.34* (.03)	.34* (.03)	.34* (.03)	.34* (.03)	.34* (.03)	.34* (.03)
Heterogeneity Measures							
SD		-.02 (.01)			-.04 (.11)		
Range			-.01 (.01)			.04 (.09)	
IQR				-.02 (.01)			-.11(.07)
SD ²					.02 (.11)		
Range ²						-.06 (.09)	
IQR ²							.09 (.07)
R^2 (marginal/conditional)	.52/.70	.54/.73	.54/.73	.54/.73	.54/.73	.54/.73	.54/.73
AIC	35320.95	35062.82	35063.80	35063.56	35067.61	35068.27	35066.97
BIC	35502.85	35307.99	35308.73	35321.35	35308.97	35320.69	35320.05

Note. Mig = Migration Background (0 = speaks German language at home; 1 = speaks German and other language or other language only at home); Grade (0 = 3rd grade; 1 = 4th grade).

* $p < .05$

Table 5

Longitudinal Multilevel Regression Analysis predicting Students' Performance in Reading Comprehension

Predictor variables	Model 0 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Intercept	-.14 (.23)	-.15 (.23)	-.15 (.24)	-.09 (.23)	-.21 (.23)	-.16 (.23)	-.17 (.23)
Measurement Point	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)
Within class level							
Female	.03 (.02)	.03 (.02)	.03 (.02)	.04 (.02)	.04 (.02)	.03 (.02)	.03 (.02)
Mig	-.15* (.02)	-.15* (.02)	-.15* (.02)	-.15* (.02)	-.15* (.02)	-.15* (.02)	-.15* (.02)
Pre-test Comprehension	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)
Between class level							
Grade	-.09 (.06)	-.09 (.06)	-.09 (.06)	-.11 (.06)	-.08 (.06)	-.09 (.06)	-.09 (.06)
Mean Achievement	.28* (.03)	.27* (.03)	.27* (.03)	.26* (.03)	.26* (.03)	.28* (.03)	.26* (.03)
Heterogeneity Measures							
SD		.01 (.02)			-.30* (.13)		
Range			-.00 (.02)			-.12 (.14)	
IQR				.04* (.02)			-.17* (.09)
SD ²					.32* (.13)		
Range ²						.12 (.14)	
IQR ²							.23* (.09)
R ² (marginal/conditional)	.34/.62	.36/.65	.36/.65	.36/.65	.36/.65	.36/.65	.36/.65
AIC	40876.85	40807.12	40807.69	40848.02	40804.98	40811.52	40802.47
BIC	41058.76	41052.29	41048.77	41064.6	41052.87	41058.06	41055.55

Note. Mig = Migration Background (0 = speaks German language at home; 1 = speaks German and other language or other language only at home); Grade (0 = 3rd grade; 1 = 4th grade).

* $p < .05$

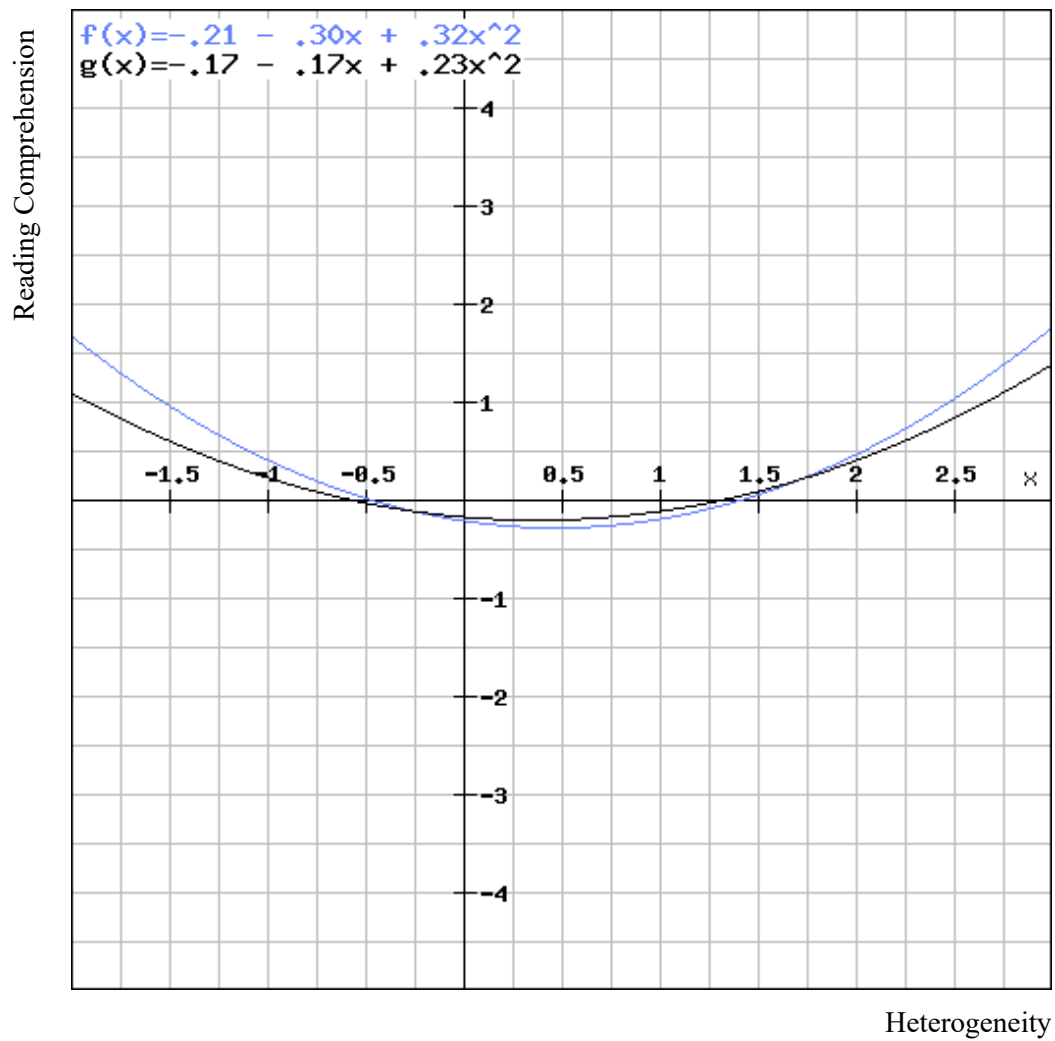


Figure 1: Reading Comprehension. Blue line = Squared Standard Deviation Model (Model 4), Black line = Squared Interquartile Range Model (Model 6).

References

- Author, 2006. [details removed for peer review]
 Author, 2011. [details removed for peer review]
 Author, 2014. [details removed for peer review]
 Author, 2015a. [details removed for peer review]
 Author, 2015b. [details removed for peer review]
 Author, 2018a. [details removed for peer review]
 Author, 2018b. [details removed for peer review]
- Bellin, N. (2009). *Klassenkomposition, Migrationshintergrund und Leistung. Mehrebenenanalyse zum Sprach- und Leseverständnis von Grundschulern* [Class composition, ethnic background, and achievement. A multi-level analysis of word and reading comprehension in primary school]. Wiesbaden: VS.
- Bliese, P.D., & Ployhart, R. E. (2002). Growth Modeling using Random Coefficient Models: Model Building, Testing, and Illustrations. *Organizational Research Methods*, 5(4), 362-387. doi:10.1177/109442802237116
- Budescu, D. V., & Budescu, M. (2012). How to measure diversity when you must. *Psychological Methods*, 17, 215–227. doi:10.1037/a0027129
- Burke, M. A., & Sass, T. R. (2013). Classroom Peer Effects and Student Achievement. *Journal of Labor Economics*, 31(1), 51-82. doi:1086/666653
- Bühner, M., & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* [Statistics for Psychologists and Social Scientists]. Hallbergmoos: Pearson.
- Carl, F. (2017). Heterogenität in der Schule [Heterogeneity in school]. In *Gymnasium ohne Sitzenbleiben – Wie Lehrpersonen mit leistungsschwachen Schülerinnen und Schülern umgehen* [Gymnasium without failing class – How teachers deal with underachieving pupils] (pp. 45–66). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-658-15416-5_3
- Carman, K. G., & Zhang, L. (2012). Classroom Peer Effects and Academic Achievement: Evidence from a Chinese Middle School. *China Economic Review*, 23, 223-237. doi:10.1016/j.chieco.2011.10.004
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third Grade. *Psychological Science*, 24, 1408-1419. doi:10.1177/0956797612472204
- Dar, Y., & Resh, N. (1986). Classroom Intellectual Composition and Academic Achievement. *American Educational Research Journal*, 23, 357-374.
- Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., et al. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *The Journal of Educational Research*, 108, 358-370. doi:10.1080/00220671.2014.899957
- Decristan, J., Fauth, B., Kunter, M., Gerhard Büttner, G., & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research*, 86, 109-121. doi:10.1016/j.ijer.2017.09.004
- Dreeben, R., & Barr, R. (1988). Classroom Composition and the Design of Instruction. *Sociology of Education*, 61(3), 129-142.
- Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review*, 52(2), 245–273. doi:10.1086/528764
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138. doi:10.1037/1082-989X.12.2.121
- Forsa (2020). Die Arbeitssituation von Lehrkräften nach den Schulöffnungen. Ergebnisse einer bundesweiten Repräsentativbefragung [The work situation of teachers after school openings. Results of a nationwide representative survey]. Retrieved from https://www.vbe.de/fileadmin/user_upload/VBE/Service/Meinungsumfragen/2020-06-09_-_forsa-VBE_Corona_Schuloeffnungsphase.pdf
- Francis, B., Archer, L., Hodgen, J., Pepper, D., Tylor, B., & Travers, M.-C. (2017). Exploring the relative lack of impact of research on 'ability grouping' in England: a discourse analytic account. *Cambridge Journal of Education*, 47, 1-17. doi:10.1080/0305764X.2015.1093095
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., ... Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, 98(1), 29–43. doi:10.1037/0022-0663.98.1.29

- Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-Intervention: A decade later. *Journal of Learning Disabilities*, 45, 195-203. doi:10.1177/0022219412442150
- Gröhlich, C., Scharenberg, K. & Bos, W. (2009). Wirkt sich Leistungsheterogenität in Schulklassen auf den individuellen Lernerfolg in der Sekundarstufe aus? [Does Classroom Heterogeneity in Secondary Schools affect Student Achievement?]. *Journal for Educational Research Online*, 1(1), 86-105.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts (7. Aufl.)* [Teaching quality and teacher professionalisation. Diagnosis, evaluation and enhancement of instruction (Vol. 7)]. Bobingen: Klett Kallmeyer.
- Hopkins, K. D., Hopkins, B. R., & Glass, G. V. (1996). *Basic Statistics for the Behavioral Sciences*. Boston: Allyn & Bacon.
- Huang, M.-H. (2008). Classroom Homogeneity, Family Background, and Math Performance: A Cross-National Difference-in-Differences Analysis. *Taiwanese Journal of Sociology* 40, 1-44.
- Jensen, P., & Rasmussen, A. W. (2011). The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review*, 30, 1503-1515. doi:10.1016/j.econedurev.2011.08.002
- Kerckhoff, A. (1986). Effects of Ability Grouping in British Secondary Schools. *American Sociological Review*, 51(6), 842-858. doi:10.2307/2095371
- Krause, A. (2004). Erhebung aufgabenbezogener Belastungen im Unterricht – ein Untersuchungskonzept [Measuring task-related disturbances in class – an examination concept]. *Zeitschrift für Arbeits- und Organisationspsychologie*, 48, 139-147. doi:10.1026/0932-4089.48.3.139
- Kulik, C.-L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research*, 19(3), 415-428. doi:10.3102/0002831201900341
- Kulik, C.-L. C., & Kulik, J. A. (1984). *Effects of ability grouping on primary school pupils: A meta-analysis*. Paper presented at the annual meeting of the American Psychological Association August, ERIC No. ED 255 329.
- Kuzmina, Y., & Ivanova, A. (2018). The effects of academic class composition on academic progress in elementary school for students with different levels of initial academic abilities. *Learning and Individual Differences*, 64, 43-53. doi:10.1016/j.lindif.2018.04.004
- Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English-language backgrounds. *Educational Researcher*, 27(4), 12-21. doi:10.3102/0013189X027004012.
- Lehmann, R. H., Peek, R., & Poerschke, J. (2006). *HAMLET 3-4-Hamburger Lesetest für 3. und 4. Klassen* [Hamburger reading test for grade 3 and 4]. Göttingen: Hogrefe.
- Leiter, J. (1983). Classroom composition and achievement gains. *Sociology of Education*, 56(3), 126-132. doi:10.2307/2112381
- Little, C. A., McCoach, D. B., & Reis, S. M. (2014). Effects of Differentiated Reading Instruction on Student Achievement in Middle School. *Journal of Advanced Academics*, 25, 384-402. doi:10.1177/1932202X14549250
- Liebertson, S. (1969). Measuring population diversity. *American Sociological Review*, 34, 850-862. <http://dx.doi.org/10.2307/2095977>
- Lleras, C. & Rangel, C. (2009). Ability Grouping Practices in Elementary School and African American/Hispanic Achievement. *American Journal of Education*, 115, 279-304. doi:10.1086/595667
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B. & D'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, 66(4), 423-458. doi:10.3102/00346543066004423
- Luyten, H., & Van der Hoeven-van Doornum, A. (1995). Classroom composition and individual achievement: Effects of classroom composition and teacher goals in Dutch elementary education. *Tijdschrift voor Onderwijsresearch*, 20(1), 42-62.
- Maehler, D., & Brinkman, H. U. (2016). *Methoden der Migrationsforschung - Ein interdisziplinärer Forschungsleitfaden* [Methods of research of migration – an interdisciplinary research guide]. Wiesbaden: Springer VS. doi:10.1007/978-3-658-10394-1
- Markic, S., & Abels, S. (2014). Heterogeneity and Diversity. A Growing Challenge or Enrichment for Science Education in German Schools?. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(4), 271-283. doi:10.12973/eurasia.2014.1082a
- Mayringer, H., & Wimmer, H. (2003). *SLS 1-4. Salzburger Lese-Screening für die Klassenstufen 1-4* [Salzburger reading-screening for grade 1-4]. Bern: Huber.
- Minner, D. D., Levy, A. J., & Century, J. (2009). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47, 474-496. doi:10.1002/tea.20347
- Novo, A., & R Core Team (2013). *norm: Analysis of multivariate normal datasets with missing values*. Retrieved from <https://CRAN.R-project.org/package=norm>

- Oakes, J. (2005). *Keeping Track: How Schools Structure Inequality*. 2nd ed. New Haven, CT: Yale University Press.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. [Empirical Study]. *Cognition and Instruction*, 1, 117-175. doi:10.1207/s1532690xci0102_1
- Philipp, M. (2013). *Lese- und Schreibunterricht* [reading and writing lessons]. Tübingen: Francke.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.
- Pinheiro J., Bates D., DebRoy S., Sarkar D., & R Core Team (2018). *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-137*. Retrieved from <https://CRAN.R-project.org/package=nlme>
- Powell, J. (2006). Special education and the risk of becoming less educated. *European Societies* 8(4), 577-599. doi:10.1080/14616690601002673
- Postlethwaite, T. N. (1988). *The Encyclopedia of comparative Education and National Systems of Education*. Oxford: Pergamon Press.
- Richter & van Holt (2005). ELVES: Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von Teilprozessen des Leseverstehens [ELVES. A computer-based diagnostic tool to assess efficiency of reading comprehensions' sub-processes]. *Diagnostica*, 51, 169-182. doi:10.1026/0012-1924.51.4.169
- Scharenberg, K. (2012). *Leistungsheterogenität und Kompetenzentwicklung* [class heterogeneity and the development of competences]. Münster, Germany: Waxmann.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Slavin, R. E. (1987). Ability grouping and student achievement in primary schools: A best-evidence synthesis. *Review of Educational Research*, 57(3), 293-336. doi:10.3102/00346543057003293
- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60(3), 471-499. doi:10.3102/00346543060003471
- Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, 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 Meta-Analyses. *Review of Educational Research*, 86(4), 849-899. doi:10.3102/0034654316675417
- Tarelli, I., Lankes, E.-M., Drossel, K., & Gegenfurtner, A. (2012). Lehr- und Lernbedingungen an Grundschulen im internationalen Vergleich. [An international comparison of teaching and learning conditions in primary schools.]. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Eds.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [IGLU 2011. Reading skills of primary school children in Germany compared internationally] (pp. 137-173). Münster: Waxmann.
- Teachman, J. D. (1980). Analysis of population diversity: Measures of qualitative variation. *Sociological Methods & Research*, 8, 341–362. <http://dx.doi.org/10.1177/004912418000800305>
- Topping, K. J. (2006) Building reading fluency: Cognitive, behavioral, and socio-emotional factors and the role of peer-mediated learning. In S. J. Samuels, & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 106-129). Newark, DE: International Reading Association.
- Vigdor, J., & Nechyba, T. (2007). Peer effects in North Carolina public schools. In P. E. Peterson, & L. Woessmann (Eds.), *Schools and the equal opportunity problem* (pp. 73-101). Cambridge, MA: MIT Press.
- von der Groeben, A. (2003). Lernen in heterogenen Gruppen. Chance und Herausforderung [Learning in heterogeneous groups. Opportunity and challenge]. *Pädagogik*, 55(9), 6-9.
- Watson, C. S., Kidd, G. R., Horner, D. G., Lowther, C. A., Eddins, D. A., Krueger, G. et al. (2003). Sensory, Cognitive, and Linguistic Factors in the Early Academic Performance of Elementary School Children: The Benton-IU Project. *Journal of Learning Disabilities*, 36(2), 165-197. doi:10.1177/002221940303600209
- Welkowitz, J., Ewen, R. B., & Cohen, J. (1971). *Introductory Statistics for the behavioral sciences*. London: Academic Press.
- Wenning, N. (1999). *Vereinheitlichung und Differenzierung. Zu den 'wirklichen' gesellschaftlichen Funktionen des Bildungswesens im Umgang mit Gleichheit und Verschiedenheit* [Unification and differentiation. The 'real' social functions of education in dealing with equality and diversity]. Opladen: Leske + Budrich.
- Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement*, 49(3), 587–602. doi:10.1177/001316448904900309
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57-67. doi:10.1023/A:1007999204543
- Zimmer, R. W., & Toma, E. F. (2000). Peer Effects in Private and Public Schools across Countries. *Journal of Policy Analysis and Management*, 19(1), 75-92. doi:10.1002/(SICI)1520-6688(200024)19:1<75::AID-PAM5>3.0.CO;2-W

Supplementary Material 01: Detailed Results

Here, we provide detailed results for effects of within-class heterogeneity on students' performance in reading fluency and reading comprehension (sample without intergrade classes: Tables 1 and 2; sample with intergrade classes: Tables 3 and 4).

Analyses were based on the following measures:

Dependent variables:

Performance of students on the fluency questions from the eight LPA tests (Table 1)

Performance of students on the comprehension questions from the eight LPA tests (Table 2)

Independent variables at the time level:

Measurement Point = Point of assessment; operationalized as 0 = first measurement point, 1 = second measurement point, (...), 7 = eighth measurement point.

Independent variables at the within-class level:

Gender = sex of student; operationalized as 0 = male and 1 = female.

Mig = Migration background of student; operationalized using the spoken language at home, as 0 = speaks German language at home; 1 = speaks German and other language or other language only at home.

Pretest = Pretest at the beginning of the respective school year.

Pretest fluency = Students' performance on the reading fluency test (SLS)

Pretest comprehension = Students' performance on the reading comprehension test (HAM-LET-S)

Independent variables at the between-class level:

LPA-G = Implementation of LPA combined with a student goal-setting procedure (Study 1; Sample A).

1 = LPA with goal setting condition; 0 = other

Anhang A

LPA-RS-3 = Implementation of LPA combined with the Reading Sportsman reading condition in Grade 3 (Study 2; Sample B₃).

1 = LPA with reading sportsman condition; 0 = other

LPA-RS-4 = Implementation of LPA combined with the Reading Sportsman reading condition in Grade 4 (Study 3; Sample B₄).

1 = LPA with reading sportsman condition; 0 = other

LPA-FB = Implementation of LPA in combination with feedback material (Study 3; Sample C).

1 = LPA with feedback material condition; 0 = other

LPA-FB-RS = Implementation of LPA in combination with feedback material and a Reading Sportsman condition (Study 4; Sample C).

1 = LPA with feedback material and Reading Sportsman condition; 0 = other

LPA = Implementation of LPA without further treatment

1 = LPA without further treatment; 0 = other

Grade = Grade students belong to; operationalized as factor variables

Sample without intergrade classes: 0 = 3rd grade and 1 = 4th grade.

Sample with intergrade classes: 0 = Grade_3, 1 = Grade_4, 2 = Intergrade_3_4 (Intergrade class mixed between 3rd and 4th grade students)

Average class achievement and each measure of heterogeneity was operationalized using the respective pretest for reading fluency or reading comprehension.

Mean = within-class average achievement

SD = Heterogeneity operationalized as the within-class standard deviation

Range = Heterogeneity operationalized as the within-class range

IQR = Heterogeneity operationalized as the within-class interquartile range

SD² = Heterogeneity operationalized as the squared within-class standard deviation

Range² = Heterogeneity operationalized as the squared within-class range

IQR² = Heterogeneity operationalized as the squared within-class interquartile range

Table 1
Longitudinal Multilevel Regression Analysis Predicting Reading Fluency

Predictor variables	Model 0 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Intercept	-.58 (.20)	-.64 (.20)	-.64 (.20)	-.61 (.20)	-.64 (.20)	-.65 (.20)	-.61 (.20)
Measurement Point	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)
Within class level							
Female	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)
Mig	-.11* (.02)	-.11* (.02)	-.11* (.02)	-.11* (.02)	-.11* (.02)	-.11* (.02)	-.11* (.02)
Pre-test Fluency	.58* (.03)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)
Between class level							
LPA-G	-.15* (.06)	-.14* (.06)	-.14* (.06)	-.14* (.06)	-.14* (.06)	-.14* (.06)	-.14* (.06)
LPA-RS-3	.11 (.05)	.10 (.05)	.10 (.05)	.10 (.05)	.10 (.05)	.10 (.05)	.10 (.05)
LPA-RS-4	-.04 (.06)	-.05 (.06)	-.05 (.06)	-.05 (.07)	-.05 (.07)	-.05 (.07)	-.04 (.06)
LPA-FB	-.01 (.05)	-.00 (.05)	-.00 (.05)	-.01 (.05)	-.00 (.05)	.00 (.05)	-.02 (.06)
LPA-FB-RS	.04 (.06)	.05 (.06)	.05 (.06)	.05 (.06)	.05 (.06)	.05 (.06)	.04 (.06)
Grade	.06 (.06)	.07 (.06)	.07 (.06)	.06 (.06)	.07 (.06)	.07 (.06)	.06 (.06)
Mean Achievement	.34* (.03)	.34* (.03)	.34* (.03)	.34* (.03)	.34* (.03)	.34* (.03)	.34* (.03)
Heterogeneity Measures							
SD		-.02 (.01)			-.04 (.11)		
Range			-.01 (.01)			.04 (.09)	
IQR				-.02 (.01)			-.11(.07)
SD ²					.02 (.11)		
Range ²						-.06 (.09)	
IQR ²							.09 (.07)
R ² (marginal/conditional)	.52/.70	.54/.73	.54/.73	.54/.73	.54/.73	.54/.73	.54/.73
AIC	35320.95	35062.82	35063.80	35063.56	35067.61	35068.27	35066.97
BIC	35502.85	35307.99	35308.73	35321.35	35308.97	35320.69	35320.05

Note. Mig = Migration Background (0 = speaks German language at home; 1 = speaks German and other language or other language only at home); LPA-G = Goal setting; LPA-RS-3 = Worked-out material Grade 3; LPA-RS-4 = Worked-out material Grade 4; LPA-FB = Feedback; LPA-FB-RS = Worked-out material and feedback; Grade (0 = 3rd grade; 1 = 4th grade).

* $p < .05$

Table 2
Longitudinal Multilevel Regression Analysis Predicting Reading Comprehension

Predictor variables	Model 0 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Intercept	-.14 (.23)	-.15 (.23)	-.15 (.24)	-.09 (.23)	-.21 (.23)	-.16 (.23)	-.17 (.23)
Measurement Point	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)
Within class level							
Female	.03 (.02)	.03 (.02)	.03 (.02)	.04 (.02)	.04 (.02)	.03 (.02)	.03 (.02)
Mig	-.15* (.02)	-.15* (.02)	-.15* (.02)	-.15* (.02)	-.15* (.02)	-.15* (.02)	-.15* (.02)
Pre-test Comprehension	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)
Between class level							
LPA-G	-.22* (.07)	-.22* (.07)	-.23* (.07)	-.19* (.07)	-.22* (.07)	-.23* (.07)	-.20* (.07)
LPA-RS-3	.03 (.06)	.02 (.06)	.03 (.06)	.04 (.06)	.03 (.06)	.03 (.06)	.02 (.06)
LPA-RS-4	-.02 (.08)	-.01 (.08)	-.03 (.08)	.01 (.08)	-.02 (.08)	-.03 (.08)	-.01 (.08)
LPA-FB	-.01 (.06)	-.01 (.06)	-.01 (.06)	-.01 (.06)	-.01 (.06)	-.01 (.06)	-.00 (.06)
LPA-FB-RS	.01 (.07)	.00 (.07)	.01 (.07)	-.01 (.07)	.02 (.07)	.00 (.07)	.01 (.07)
Grade	-.09 (.06)	-.09 (.06)	-.09 (.06)	-.11 (.06)	-.08 (.06)	-.09 (.06)	-.09 (.06)
Mean Achievement	.28* (.03)	.27* (.03)	.27* (.03)	.26* (.03)	.26* (.03)	.28* (.03)	.26* (.03)
Heterogeneity Measures							
SD		.01 (.02)			-.30* (.13)		
Range			-.00 (.02)			-.12 (.14)	
IQR				.04* (.02)			-.17* (.09)
SD ²					.32* (.13)		
Range ²						.12 (.14)	
IQR ²							.23* (.09)
R ² (marginal/conditional)	.34/.62	.36/.65	.36/.65	.36/.65	.36/.65	.36/.65	.36/.65
AIC	40876.85	40807.12	40807.69	40848.02	40804.98	40811.52	40802.47
BIC	41058.76	41052.29	41048.77	41064.6	41052.87	41058.06	41055.55

Note. Mig = Migration Background (0 = speaks German language at home; 1 = speaks German and other language or other language only at home); LPA-G = Goal setting; LPA-RS-3 = Worked-out material Grade 3; LPA-RS-4 = Worked-out material Grade 4; LPA-FB = Feedback; LPA-FB-RS = Worked-out material and feedback; Grade (0 = 3rd grade; 1 = 4th grade).

* $p < .05$

Table 3

Longitudinal Multilevel Regression Analysis Predicting Reading Fluency using a sample with additional eight intergrade classes

Predictor variables	Model 0 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Intercept	-.42 (.07)	-.43 (.06)	-.43 (.06)	-.43 (.06)	-.43 (.06)	-.44 (.06)	-.43 (.06)
Measurement Point	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)	.09* (.01)
Within class level							
Female	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)
Mig	-.11* (.02)	-.11* (.02)	-.10* (.02)	-.10* (.02)	-.10* (.02)	-.10* (.02)	-.10* (.02)
Pre-test Fluency	.59* (.01)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)	.58* (.01)
Between class level							
LPA-G	-.15* (.06)	-.14* (.06)	-.14* (.06)	-.14* (.06)	-.14* (.06)	-.14* (.06)	-.14* (.06)
LPA-RS-3	.11* (.05)	.11 (.05)	.10 (.05)	.10 (.05)	.10 (.05)	.11 (.05)	.09 (.05)
LPA-RS-4	-.04 (.07)	-.05 (.07)	-.05 (.07)	-.05 (.07)	-.05 (.07)	-.04 (.07)	-.04 (.07)
LPA-FB	-.02 (.06)	-.00 (.06)	-.00 (.05)	-.02 (.05)	-.00 (.06)	-.00 (.06)	-.03 (.05)
LPA-FB-RS	.03 (.06)	.05 (.06)	.05 (.06)	.05 (.06)	.05 (.06)	.05 (.06)	.04 (.06)
Grade_4	.06 (.06)	.05 (.06)	.05 (.06)	.04 (.06)	.05 (.06)	.06 (.06)	.04 (.06)
Intergrade_3_4	.19* (.09)	.19* (.09)	.20* (.09)	.19* (.09)	.19* (.09)	.20* (.09)	.19* (.09)
Mean Achievement	.34* (.03)	.35* (.03)	.34* (.03)	.35* (.03)	.35* (.03)	.34* (.03)	.35* (.03)
Heterogeneity Measures							
SD		-.02 (.01)			-.03 (.11)		
Range			-.01 (.01)			.09 (.09)	
IQR				-.02 (.01)			-.11(.07)
SD ²					.02 (.11)		
Range ²						-.11 (.09)	
IQR ²							.09 (.07)
R ² (marginal/conditional)	.53/.72	.55/.74	.55/.74	.55/.74	.55/.74	.55/.74	.55/.74
AIC	38732.95	38501.66	38500.97	38500.91	38505.77	38504.67	38504.43
BIC	38925.27	38758.09	38757.39	38757.34	38770.20	38769.11	38768.87

Note. Mig = Migration Background (0 = speaks German language at home; 1 = speaks German and other language or other language only at home); LPA-G = Goal setting; LPA-RS-3 = Worked-out material Grade 3; LPA-RS-4 = Worked-out material Grade 4; LPA-FB = Feedback; LPA-FB-RS = Worked-out material and feedback; Grade (0 = 3rd grade (reference group); 1 = 4th grade; 2= Intergrade class).

* $p < .05$

Table 4

Longitudinal Multilevel Regression Analysis Predicting Reading Comprehension using a sample with additional eight intergrade classes

Predictor variables	Model 0 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 3 β (SE)	Model 4 β (SE)	Model 5 β (SE)	Model 6 β (SE)
Intercept	-.42 (.08)	-.42 (.08)	-.42 (.08)	-.42 (.08)	-.43 (.08)	-.42 (.08)	-.43 (.08)
Measurement Point	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)	.11* (.01)
Within class level							
Gender	.04 (.02)	.03 (.02)	.04 (.02)	.04 (.02)	.04 (.02)	.03 (.02)	.04 (.02)
Mig	-.14* (.02)	-.14* (.02)	-.14* (.02)	-.14* (.03)	-.14* (.02)	-.14* (.02)	-.14* (.02)
Pre-test Comprehension	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)	.46* (.01)
Between class level							
LPA-G	-.22* (.07)	-.22* (.07)	-.24* (.07)	-.21* (.07)	-.23* (.07)	-.24* (.07)	-.22* (.07)
LPA-RS-3	.03 (.06)	.03 (.06)	.03 (.06)	.04 (.06)	.03 (.06)	.03 (.06)	.04 (.06)
LPA-RS-4	-.02 (.08)	-.01 (.08)	-.03 (.08)	.01 (.08)	-.02 (.08)	-.03 (.08)	-.02 (.08)
LPA-FB	-.00 (.06)	-.01 (.06)	-.01 (.06)	-.01 (.06)	-.01 (.06)	-.02 (.06)	-.01 (.06)
LPA-FB-RS	.00 (.07)	.00 (.07)	.01 (.07)	-.01 (.07)	.01 (.07)	.00 (.07)	.01 (.07)
Grade_4	-.13* (.06)	-.13 (.06)	-.13 (.06)	-.14* (.06)	-.11 (.06)	-.13 (.06)	-.11 (.06)
Intergrade_3_4	.12 (.11)	.11 (.11)	.13 (.11)	.08 (.11)	.08 (.11)	.12 (.11)	.06 (.11)
Mean	.29* (.03)	.28* (.03)	.29* (.03)	.28* (.03)	.28* (.03)	.30* (.03)	.26* (.03)
Heterogeneity Measures							
SD		.01 (.02)			-.29* (.12)		
Range			-.01 (.02)			-.11 (.14)	
Range				.03 (.02)			-.14 (.07)
Range ²					.31* (.12)		
IQR						.10 (.14)	
IQR ²							.17* (.07)
R ² (marginal/conditional)	.36/.64	.38/.67	.38/.67	.38/.67	.38/.67	.38/.67	.38/.67
AIC	45219.21	45116.02	45115.78	45113.38	45114.62	45119.50	45114.07
BIC	45411.53	45372.45	45372.21	45369.81	45379.06	45383.94	45378.51

Note. Mig = Migration Background (0 = speaks German language at home; 1 = speaks German and other language or other language only at home); LPA-G = Goal setting; LPA-RS-3 = Worked-out material Grade 3; LPA-RS-4 = Worked-out material Grade 4; LPA-FB = Feedback; LPA-FB-RS = Worked-out material and feedback; Grade (0 = 3rd grade (reference group); 1 = 4th grade; 2 = Intergrade class).

* $p < .05$

Anhang B: Manuskript zu Beitrag 2 – Sprachliche Heterogenität

The impact of linguistic diversity on students' reading comprehension using different diversity measures and its' differential effects depending on students' migration background

1 Introduction

Currently, there are between 6000 and 8000 different languages spoken in the world (Hammarström, 2016; Kallens, Dale, & Christiansen, in press) and approximately 175 nationalities within the borders of the European Union (Eurostat., 2014, 2019). The increasing foreign population is bringing major changes to the European linguistic landscape and, as a result, to European classrooms (Bruen & Kelly, 2016). In Germany, the topic of linguistic diversity has been discussed in recent years, as well. Teachers are challenged with an increasing linguistic diversity of the student body; and even if teachers tend to neglect its' existence in the classroom or simply don't have strategies to deal with linguistically heterogeneous classes (e.g., Cho and McDonnough, 2009; Mazak & Herbas-Donoso, 2014), they acknowledge the importance of linguistic diversity for their teaching practice and for students' learning (Gkaintartzi, Kiliari, & Tsokolidou, 2015). As assumed by Gogolin (2011), a teacher with a deficit perspective on linguistic diversity – or a monolingual habitus – does not recognize the linguistic diversity of students as part of the process of learning. Or in other words, as Gogolin (2011) states in a broader sense, a deficit perspective on linguistic diversity leads to school failure, with students with a migration background considerably lagging behind students without a migration background with regard to reading skills, mathematics, and science (OECD, 2010).

Accordingly, both teachers and educational theory suggest that linguistic diversity is important for students' learning. However, this importance can only be theoretically substantiated using Piaget's concept of disequilibrium (1977), and empirical evidence is rare. Therefore, the aim of this study is to empirically investigate linguistic diversity and its effect on students' learning. Using a sample of 1603

5th grade students of 65 classes, we analyzed whether linguistic diversity can explain students' learning in reading comprehension. We calculated different diversity measures based on the language(s) spoken with parents as well as the language(s) spoken with friends to examine possible methodological differences and to provide a more valid picture of the expected effects.

2 Linguistic Diversity

Diversity is a key ingredient of education. In addition to diversity factors – such as cognitive ability, gender, and disability – ethnic and linguistic diversity are gaining in importance; learning and teaching in multicultural classrooms poses a major challenge to both students and teachers (Phalet et al., 2004).

According to theoretical linguistics (Nettle, 1999), linguistic diversity can be defined from three perspectives: language diversity, phylogenetic diversity, and structural diversity. The most relevant perspective for educational research is the language diversity perspective. It defines linguistic diversity as the number of different languages in a given geographical area, including two important aspects: first, speech forms can be classified as discrete languages; and second, these discrete languages are countable. Transferred to education, this perspective means that classes are more likely to be linguistically diverse, the more languages are spoken by students.

However, from an empirical point of view, it would be insufficient to measure the linguistic diversity of a class based only on the number of languages spoken in the class. First, the meaning of linguistic diversity depends on class size. Larger classes tend to feature a higher number of languages. It is a different scenario for both teachers and students if the number of languages is $N = 10$ in a class of 15 or in a class of 25 students. Second, the weight of the respective language is not taken into account. A class with a total of 20 students and three languages can be composed in very different ways: Class A, in which each language is spoken by only one student, is very different from Class B, in which each language is spoken by five students. In Class A the language of instruction is different from the mother tongue for three students, in Class B for 15 students. Moreover, in Class A foreign students do not have any classmates who speak the same language, whereas in Class B there are five students speaking the mutual foreign language, respectively.

Thus, linguistic diversity is clearly more complex than the number of language categories spoken in class. It is therefore important to investigate the effects of linguistic diversity using measures that consider both the class size and the weight of the number of languages in a classroom.

3 The Relation between Ethnic Class Composition and Linguistic Diversity

To date there is no empirical literature about how linguistic diversity affects students' learning. Nevertheless, there are findings regarding ethnic diversity (which is considered to be a related construct to linguistic diversity; see Chapter 3.3). A large number of studies have investigated the impact of ethnic class composition on student learning using the ethnic minority share – with some authors interpreting their results in the light of diversity (cf. Rjosk et al., 2017). In the last decade, however, educational researchers have extended their view of diversity beyond the ethnic minority share and have made first attempts to investigate ethnic diversity with actual diversity measures. In addition to the number of ethnic categories, researchers have also used complex indices (e.g., Simpson's *D*; Shannon's *H*) provided by other disciplines (e.g., ecology; McCann, 2000) to examine the effects of ethnic diversity on students' academic achievement (see Chapter 3.2).¹

The following chapters provide insights into research on ethnic class composition and its relevance for research on linguistic diversity. First, we will outline how the ethnic minority share affects students' learning. Second, we will present studies on ethnic diversity that have already implemented multiple, complex diversity measures. And third, we will establish the link between ethnic diversity and linguistic diversity.

3.1 Ethnic Minority Share

Before ethnic diversity was measured with multiple, complex measures, studies distinguished only very broadly between ethnic minority and majority. The ethnic minority share, a characteristic of class composition, measures the relative number of students with a migration background in a classroom. The body of research regarding the ethnic minority share is wide-ranging. Theoretical accounts assume that a class's ethnic minority share negatively affects students' academic performance – an assumption

¹ For a detailed overview of existing diversity measures, see Rjosk et al. (2017).

which is typically supported by empirical research (e.g., Cortes, 2006; Driessen, 2002; Gould, Lavy, & Paserman, 2009; Jensen & Rasmussen, 2011; Mickelson, Bottia, & Lambert, 2013; Peetsma et al., 2006; Schnepf, 2007; Van Ewijk & Slegers, 2010). However, researchers suggest that the ethnic minority share does not directly impact academic performance but assumes mediating processes: next to school resources, instructional quality, and learning culture, the linguistic background plays a role when interpreting the negative relationship between the ethnic minority share and students' learning (Driessen, 2002; Goldsmith, 2011; Raudenbush, Fotiu, & Cheong, 1998; Stipek, 2004).

According to the language contact hypothesis (Driessen, 2002), the higher the ethnic minority share, the higher the chance that minority students mainly speak their mother tongue. Classrooms with a high ethnic minority share tend to have high proportions of students who do not speak the language of instruction with their family at home. Furthermore, students may not speak the language of instruction with each other in situations such as school breaks or cooperative learning sessions (cf. Cummins, 1979; 1981; basic interpersonal communicative skills). Consequently, they learn the national language less well and thus they may be less able to follow the teaching process and to support each other by explaining learning materials in the language of instruction (e.g., Driessen, 2002; Entwisle & Alexander, 1994; Peetsma et al., 2006; Rjosk et al., 2017; Van Ewijk & Slegers, 2010). This may undermine the quality of learning because cognitive processes are less stimulated; and as a consequence, this might negatively affect students' achievement.

When interpreting the ethnic minority share, researchers take into account the linguistic background. This again underlines the importance of considering language and linguistic diversity as a relevant factor for student learning.

3.2 Ethnic Diversity

As stated above, researchers on ethnic diversity have made first attempts to operationalize diversity with multiple, complex measures. Therefore, research on ethnic diversity – as opposed to research on linguistic diversity – has generated some empirical findings about how diversity influences students' learning. Nevertheless, the state of research is extremely limited.

Researchers on ethnic diversity have assumed that ethnic diversity positively influences students' learning (e.g., Benner & Crosnoe, 2011; Gurin, Dey, Gurin, & Hurtado, 2003; Peetsma et al., 2006; Tam & Bassett, 2004) and typically, this assumption has been supported by empirical findings. Benner and Crosnoe (2011) used the diversity measure Simpson's D (Simpson, 1949) to predict learning among elementary school students. Their analyses revealed that, beyond the ethnic minority share, students in more ethnically diverse elementary schools achieved significantly higher test scores in both reading ($\beta = .05$) and mathematics ($\beta = .04$). Another study by Rjosk et al. (2017) found similar results using several diversity measures. Controlling for the ethnic minority share, the three measures Number of Ethnic Categories, Simpson's D (Simpson, 1949), and Shannon's H (Shannon & Weaver, 1949) showed small, but positive effects on students' mathematics achievement ($\beta_{\text{EthnicCategories}} = .04$; $\beta_{\text{Simpson's } D} = .03$; $\beta_{\text{Shannon's } H} = .03$), but not on students' reading achievement ($\beta_{\text{EthnicCategories}} = .00$; $\beta_{\text{Simpson's } D} = .00$; $\beta_{\text{Shannon's } H} = .00$). Tam & Bassett (2004) used the Absolute Diversity Index (similar to Simpson's D ; Meyer & McIntosh, 1992) to investigate whether high school diversity can predict students' GPA at university. Their analyses revealed that the expected GPA in the first semester was about one-fourth higher for students from more diverse high schools than for students from less diverse high schools ($\beta = .23$).

In summary, although the impact of ethnic diversity on academic performance has received scant attention, research suggests that ethnic diversity tends to increase students' learning.

3.3 Linguistic Diversity as a Result of Ethnic Diversity

From a theoretical perspective, it is reasonable that linguistic diversity is a result of ethnic diversity. In ethnically diverse classrooms, there should be a larger number of different linguistic backgrounds (e.g., Rjosk et al., 2017; Terry & Irving, 2010). Even if linguistic diversity is assumed to be a result of ethnic diversity – and, thus, the two components should be positively related – it is important to focus on linguistic diversity for three reasons.

First, a theoretical point of view. Ethnic diversity comes with high theoretical complexity and captures – compared to linguistic diversity – several characteristics of the student body. Ethnicity is

based on a shared understanding of the history and territorial origins of an ethnic group as well as on particular cultural characteristics: language, religion, specific customs, and ways of life (United Nations Economic Commission for Europe, 2006). Thus, ethnic diversity is defined widely and does not only refer to the variety of ethnic groups; it also involves diverse values, behaviors, and attitudes experienced and accepted by those groups (Nemetz & Christensen, 1996).

Second, a school practical point of view related to teaching. Teachers need different strategies for dealing with ethnic diversity and linguistic diversity. When considering ethnic diversity, teachers have to focus on a wide spectrum of a student's background, including religion, attitudes, and values. In comparison, linguistic diversity is much more closely linked to cognitive aspects of learning and to instruction. The literature has provided teachers with various methods that are directly related to a class's linguistic composition, such as the implementation of language contact (e.g., Driessen, 2002; Entwisle & Alexander, 1994; Peetsma et al., 2006; Scheele, Leseman, & Mayo, 2010; see also Time-on-Task Hypothesis: Carroll, 1963, 1989), team teaching for second language learners (e.g., Sierens & Ramaut, 2018), or structuring bilingual classrooms (e.g., Gogolin, 2011; Markic & Abels, 2014; Miller, 2009; Salleh, Venville and Treagust, 2007).

And third, an empirical point of view. There is evidence that a student's migration background is not fully comparable to a student's linguistic background. According to research on the operationalization of migration background, a student's migration status depends on how it is measured: either based on whether the student's parents are born abroad or based on the language spoken at home (Dubowy et al., 2011). If there are substantial differences between the student-level measures, there are also supposed to be differences between the class-level measures. This means that the basic descriptive characteristics of a class's ethnic diversity and a class's linguistic diversity can be different, which in turn can be pivotal for the results of empirical analyses.

In summary, the three aspects suggest that linguistic and ethnic diversity are similar in many ways but are also different, and that it is therefore important that research focuses on both ethnic diversity and linguistic diversity.

4 The Role of Linguistic Diversity for Students' Learning

According to Piaget's concept of disequilibrium (1977), diverse learning environments promote students' learning. Students in diverse learning environments encounter contradictions and discrepancies; and tension between what students know and what they are encountering leads to a state of cognitive conflict. This state of cognitive conflict – of disequilibrium – is induced by information that does not fit into one's schemas.

A schema is a cognitive construct that organizes and stores knowledge in long-term memory (cognitive load theory; Sweller, van Merriënboer, & Paas, 1998; van Merriënboer & Sweller, 2005). Being faced with multiple linguistic backgrounds can lead to students' encountering information, that does not fit into their schemas. This drives their learning process through assimilation and accommodation of new information and leads to cognitive stimulation: existing schemata are extended, or new schemata are developed. The aim of learning is the formation and storage of automated schemata because the presence of schemata reduces the cognitive load (e.g., Sweller, 2005). Even an overly complex schema can be dealt with as one element in working memory. Thus, a diverse classroom composition is stimulating the modification and the development of schemata; and it is therefore supposed to enhance students' learning.

Using Piaget's concept of disequilibrium as a theoretical rationale, all students should make academic progress – independent of whether they have a migration background or not. Both students with and without migration backgrounds encounter contradictions in diverse classes and thus, both groups should have similar opportunities to expand their cognitive skills. However, it is possible that students with and without a migration background benefit differently from linguistic diversity.

According to the language-contact hypothesis, academic performance is supposed to be better if students have frequent contact with the language of instruction (e.g., Driessen, 2002; Peetsma et al., 2006). Outside the classroom, students spend most of their time with their families, with friends, and friends. But even if several cultural groups are present in school, most friendships

remain segregated along ethnic lines (Kumar, 2003). Therefore, students without a migration background might be more likely to stay in environments with people who mainly speak the national language – the language of instruction. Students with a migration background may be more likely to stay in environments with people who mainly speak a mother tongue different to the national language (e.g., Walter & Taskinen, 2007; Tiedemann & Billmann-Mahecha, 2007).

Previous knowledge plays an important role in cognitive load theory. The higher the level of knowledge in a content area and the more familiar the learners are with this content, the easier it is for them to process new information. Furthermore, an overload of cognitive resources can lead to problems in understanding content and storing information (e.g., Chandler & Sweller, 1992; Kalyuga, Ayres, & Sweller, 2003). A class's linguistic diversity drives cognitive activation through adopting and developing new schemata. However, the less proficient students are in the language of instruction, the more complex and difficult the learning content becomes. As a result, more cognitive resources have to be used (cf. intrinsic load; e.g., van Merriënboer & Sweller, 2005; Sweller, 1994). Students with lower proficiency in the language of instruction might be more likely to face a cognitive overload as the linguistic diversity in class increases. Hence, proficiency in the language of instruction might be crucial when investigating linguistic diversity and its effects on students' learning.

5 Research Questions

To date, there has been no empirical research regarding linguistic diversity and students' learning. However, theoretical accounts indicate that students' learning is expected to be higher as linguistic diversity increases. The aim of our study is to provide clarification by testing the following research questions (RQ):

1. Does linguistic diversity have an impact on students' reading comprehension, beyond the ethnic minority share?

According to Piaget's concept of disequilibrium (1977), we assumed that a class's linguistic diversity would have a positive effect on students' reading comprehension. In order to examine possible meth-

odological differences, we used three diversity measures: relative number of languages N_{Lan} , Simpson's D , and Shannon's H . Furthermore, the analyses were calculated twice – one time with the diversity measures operationalized by the language(s) spoken with parents; and one time by language(s) spoken with friends. This allowed us to provide a more valid picture of the effects of linguistic diversity.

2. Do linguistic diversity and students' migration background have a differential effect on students' reading comprehension?

To date, there has been no empirical research on differential effects of linguistic diversity and students' migration background. However, theoretical accounts suggest that students with a migration background have less contact to the language of instruction, and students with lower proficiency in the language of instruction are more likely to face problems with understanding content and storing information. Therefore, we assumed that the impact of linguistic diversity on learning in reading comprehension might differ between students with and without a migration background.

6 Method

5.1 Sample and Design

In the school year 2018/2019, a longitudinal study on students' reading comprehension skills was conducted. Thirteen classes were excluded because parental consent rate was lower than 80%. The sample consisted of $N = 1603$ 5th grade students (50% female; $M_{age} = 10.20$; $SD = .52$) from 65 classes and 27 secondary schools (18 Gymnasien with 45 classes, 4 Realschulen with 11 classes, 3 Gemeinschaftsschulen with 7 classes, 2 Werkrealschulen with 2 classes).² The average class size was 25 students. A subsample participated in a training on reading strategies.³ This is covered in the Limitations Chapter (see Chapter 7.1), and it was controlled for in our analyses.

² Gymnasium = highest secondary school track; Realschule = second-highest secondary school track; Werkrealschule = lowest secondary school track; Gemeinschaftsschule = mixed form of secondary school tracks.

³ Participation in the training in reading strategies did not significantly predict students' reading comprehension performance ($\beta = -.09$; $p = .23$).

The data were collected at two measurement points: Each student took part in a reading comprehension pretest in September 2018 and in a posttest in February 2019. Both tests were conceptualized by the Institute for Educational Analysis Baden-Württemberg (IBBW) in Germany. In addition to the reading comprehension tests, students also filled out a questionnaire, giving information about their parents' country of birth, the language(s) they speak with their parents, and the language(s) they speak with their friends. In the sample, 59% of the students had a migration background, which was measured based on their parents' country of birth (i.e., at least one parent born abroad; c.f. Statistisches Bundesamt, 2018). The total number of languages spoken by students was $N = 61$, with 61 languages spoken with parents and 29 languages spoken with friends. Moreover, 49% of students spoke only German with their parents at home; 17% spoke German and another language, and 34% spoke another language only. Furthermore, 85% of students spoke only German with their friends; 2% spoke German and another language, and 13% spoke another language only.

5.2 Instruments

5.2.1 Reading Comprehension

Pretest and posttest reading comprehension were measured with two different tests of Lernstand 5 (IBBW, 2018), which is an established, state-wide measure of 5th grade students' reading ability. Both tests consisted of four texts (two factual texts and two literary texts) of varying length and difficulty. The maximum score for the pretest was 37 points and for the posttest 35 points. Raw data (dichotomously coded) were used for scaling according to the IRT model in order to create ability scores. The item parameters were fixed for both measurement points to keep the ability scores comparable. Expected a posteriori (EAP/PV) reliability was .85 for the pretest and .86 for the posttest.

5.2.2 Diversity Measures

For each class, we used the language(s) spoken with parents and the language(s) spoken with friends to calculate three diversity measures, which were: relative Number of Languages N_{Lan} , Simpson's D , and Shannon's H (see Table 1). Together, a total of six diversity measures were calculated – three on the basis of language(s) spoken with parents and three on the basis of language(s) spoken with friends.

$NLan$ is a proportion-based measure which interprets variety as a property of diversity (see Stirling, 2007). Its minimum is $NLan = 1/n$ and its maximum is $NLan = x/n$ with n = number of students and x = number of languages. Because students can speak more than one language, its maximum is theoretically infinite. The higher the diversity measure $NLan$, the greater the diversity in a class.

Simpson's D is a probability-based measure which is interpreted as the probability that randomly selected students from a classroom speak different languages (see Rjosk et al., 2017). Its minimum is $D = 0$ (i.e., all students in a classroom speak the same language), and its maximum $D = (c - 1)/c$ is achieved when the distribution across c languages in the classroom is uniform (i.e., all students are equally likely to meet a classmate who speaks another language). The higher the diversity measure Simpson's D , the greater the diversity in a class.

Shannon's H is a logarithm-based measure (see Budescu & Budescu, 2012). Its minimum is $H = 0$ (i.e., all students in a classroom speak the same language), and its maximum $H = \ln(c)$ is achieved when the distribution across the c languages in the classroom is uniform. The higher the measure Shannon's H , the greater the diversity in a class.

 Insert Table 1 about here

5.3 Data Analysis

We calculated multilevel regression models to predict students' learning in reading comprehension according to a class's linguistic diversity. Due to the nested structure, our analyses included students' class affiliation and school affiliation. Individual test performance was centered at the group mean because of its relationship to class affiliation (Enders and Tofighi, 2007). All other independent continuous variables were z-standardized. We used the R package lme4 to run our models (Bates et al., 2015).

5.3.1 Analyses for Research Questions

On student level 1, we included students' reading comprehension pretest score and individual demographics, which were defined as control variables (i.e., gender; 0 = male; 1 = female), migration background (0 = both parents born in Germany; 1 = at least one parent born abroad), language(s) spoken with parents/friends (0 = only other language(s); 1 = national language and other language(s); 2 = only national language), cultural capital (how many books do you have at home? 1 = none or only very few; 5 = over 200 books). On class level 2, we controlled for participation in a reading strategy training (0 = no participation; 1 = participation). On school level 3, we included school form as an additional control variable and used dummy-coded variables with Gymnasium (i.e., highest secondary school track) as the reference group.

To test RQ 1, we ran four multilevel regression models for both linguistic diversity based on the language(s) spoken with parents and linguistic diversity based on the language(s) spoken with friends. Model 0, the baseline model, examined the effect of a class's ethnic minority share on students' learning in reading comprehension. Model 1-3 analyzed the effects of a class's linguistic diversity using *NLan* in Model 1, Simpson's *D* in Model 2, and Shannon's *H* in Model 3.

To answer RQ 2, we investigated differential effects using Simpson's *D* as an exemplary diversity measure. We calculated two cross-level multilevel regression models: one with Simpson's *D* measured by language(s) spoken with parents and one by Simpson's *D* measured by language(s) spoken with friends.

5.3.2 Missing Data

The proportion of missing values in the data set for students' general background variables was 3.9% for gender and 10.8% for cultural capital. For the ethnic- and language-related variables, the proportion of missing values was 4.2% for migration background, 10.5% for language spoken with parents, and 11.2% for language spoken with friends. The proportion of missing values for the reading comprehension test scores was 9.7% for the pretest and 13.9% for the posttest.

In all models we imputed for missing data using multiple imputation via mice (Multiple Imputation by Chained Equations; van Buuren & Groothuis-Oudshoorn, 2011). Multiple imputations are considered an appropriate procedure for dealing with missing values (see Graham 2009). We generated one hundred imputations with five iterations, respectively. The results of the multiple data sets were integrated according to the formulas proposed by Rubin (1987), so that the between- and within-imputation variance was adequately taken into account.

6 Results

6.1 Descriptive Analyses

Table 2 presents the descriptive statistics which reveal high correlations (Cohen, 1988) between diversity measures being either calculated with spoken language(s) with parents ($.82 < r < .98$) or spoken language(s) with friends ($.94 < r < .99$). Correlations, which were measured between diversity measures being calculated with spoken language(s) with parents and friends showed medium to high coefficients ($.41 < r < .57$). The average ethnic minority share was 59 percent with a minimum of 26 percent and a maximum of 100 percent. The ethnic minority share showed high correlations with the diversity measures ($.51 < r < .79$), indicating that linguistic diversity increases with a higher ethnic minority share in class. In addition, the diversity measures calculated with the language(s) spoken with parents showed higher values than the diversity measures calculated based on the language(s) spoken with friends, while their standard deviations were quite similar (e.g., Simpson's D : $M = .59$, $SD = .14$ for parents; $M = .22$, $SD = .14$ for friends).

 Insert Table 2 about here

To present the diversity measures in a more understandable way, Figure 1 depicts the absolute number of languages spoken in class. The means of the overlaid histogram (A) show that, in an average class, 3.5 languages are spoken with friends, and 8.6 languages are spoken with parents, including the German language. The range between language(s) spoken with friends and language(s) spoken with

parents is 13 (classes speaking only German with friends vs. classes speaking 14 languages with parents). Compared to the relatively similar distribution of the diversity measures *relative NLan* ($SD = .11$ for parents; $SD = .07$ for friends; see Table 2), the overlaid density plot (B) shows differences in the distribution of language(s) spoken with friends ($SD = 1.8$) and of language(s) spoken with parents ($SD = 2.7$). Those differences emphasize the importance of considering the respective class size.

Insert Figure 1 about here

6.2 Effects of Diversity Measures

Table 3 shows the effects on students’ learning in reading comprehension based on diversity measures using the language(s) spoken with parents, and Table 4 shows the effects based on diversity measures using the language(s) spoken with friends. Model 0, the baseline model, determined the influence of the control variables (i.e., students’ individual demographics), students’ pretest performance, and the ethnic minority share. It revealed that students with a high pretest performance also showed a high posttest performance ($\beta = .59$; $p < .001$), and that female students showed higher learning than male students ($\beta = .22$; $p < .001$). Neither the language spoken with parents ($\beta = -.02$; $p = .52$) nor friends ($\beta = .00$; $p = .76$), nor the migration background ($\beta = -.08$; $p = .16$ for parents; $\beta = -.08$; $p = .26$ for friends) significantly influenced students’ reading comprehension. However a high ethnic minority share in class led to a decrease in students’ learning progress in reading comprehension ($\beta = -.17$; $p < .001$).

Insert Table 3 and 4 about here

RQ 1 examined whether a class’s linguistic diversity predicts students’ reading comprehension, beyond the ethnic minority share (see Tables 3 and 4; Models 1-3). The results revealed that

linguistic diversity led to a positive increase in students' reading comprehension. The significance of results was independent of the chosen diversity measure and of whether the diversity measure was calculated based on the languages spoken with parents or with friends. The respective models did not show noteworthy differences in the proportion of explained variance.

RQ 2 examined differential effects with a cross-level interaction between linguistic diversity and students' migration background (see Table 5, Models 1-2).⁴ The results revealed that students with a migration background had lower gains from linguistic diversity than their classmates without a migration background. This held true regardless of whether linguistic diversity was calculated based on the language(s) spoken with parents or based on the language(s) spoken with friends. As an example, Figure 2 depicts the results for Simpson's D calculated using the spoken language(s) with parents, showing that students with a migration background had lower gains from linguistic diversity than students without a migration background.

Insert Table 5 about here

Insert Figure 2 about here

7 Discussion

Our analyses reveal three important findings. First, the linguistic diversity in a class positively influences students' learning, beyond the ethnic minority share. Second, students without a migration background benefit more from linguistic diversity in the classroom than students with a migration background. And third, the results are independent of the diversity measure used and independent of whether the diversity measures are calculated using the language spoken with parents or with friends.

⁴ The results presented are based on Simpson's D as an exemplary diversity measure. The significance of the results did not differ depending on the chosen diversity measure. For the complete results, please refer to Supplementary Material *SM01_Complete Results Interaction Effects Migration Background*.

According to the descriptive findings, students speak more languages with their parents than with their friends. Thus, linguistic diversity is higher when it is calculated based on the language(s) spoken with parents. The diversity measure *N_{Lan}* revealed that, in an average class, every third student differs in which language he/she speaks with his/her parents. In contrast, only about one out of 10 students differ in which language he/she speaks with his/her friends. Accordingly, the diversity measure Simpson's *D* showed higher diversity when it was calculated by the language(s) spoken with parents; indeed, some classes even have zero linguistic diversity when it is measured by the language(s) spoken with friends. Interestingly, even if the descriptive statistics revealed differences between the diversity measures calculated by depending on whether they were calculated using the language(s) spoken with either parents or friends, the results of the multi-level analyses hardly differ. Both approaches confirm that linguistic diversity in class leads to an increase in reading comprehension achievement. Thus, from an empirical point of view, it does not appear to be important to use several measures of diversity or to examine the language(s) spoken with parents and those spoken with friends.

It is, however, crucial that at least one of the diversity measures is used, because it is not enough to consider only the ethnic minority share in a class. Our analyses show that students achieve lower outcomes in reading comprehension in classrooms with a higher ethnic minority share. Our analyses also show that, beyond the ethnic minority share, linguistic diversity increases students' reading comprehension performance. These findings are in line with international research on the ethnic minority share (e.g., Mickelson et al., 2013; Van Ewijk & Sleegers, 2010) and in line with Piaget's concept of disequilibrium (1977) and partly in line with previous research regarding ethnically diverse classrooms (Benner & Crosnoe, 2011; Rjosk et al., 2017; Tam & Bassett, 2004). Even if linguistic diversity and ethnic minority share are positively correlated, they are clearly different in terms of content. While a higher share of ethnic minorities is expected to negatively influence students' learning – attributed to various mediators – diversity is in many ways a cognitive exercise (Gurin et al., 2002), and it is therefore expected to foster students' learning. Linguistically diverse classrooms may also be characterized

by further beneficial aspects, such as school resources in terms of learning materials, teacher competencies and cultural understanding, organizational structure, and pedagogical concepts (Eccles & Roesser, 2011), reinforcing the positive effect of diversity.

Even if the two composition variables can be interpreted as antagonistic, our results indicate that the ethnic minority share plays a greater role. All beta coefficients of the ethnic minority share exceed those of the diversity measures. This sheds a different light on linguistic diversity, so that beyond the ethnic minority share, linguistic diversity has a rather mitigating effect than an actual positive one.

Assuming that a diverse student body increases cognitive stimulation in a class, the question arises as to whether diversity benefits all students: students with and without a migration background equally? Previous research regarding school composition has highlighted the benefits of ethnic diversity for all ethnic groups (Borman et al., 2004; Gurin et al., 2002; Gurin et al., 2003). However, Benner and Crosnoe (2011) found differential effects of ethnic diversity and ethnic group using stepwise multigroup comparisons. Their analyses revealed that white students, but not those from minority populations, tended to score higher in both reading and mathematics when attending schools with greater ethnic diversity. According to our analyses, students with a migration background also benefited less from linguistic diversity in the classroom than students without a migration background. As stated in the theoretical chapter, it could be the case that cognitive overload reduces the positive effect for students with a migration background.

If students whose mother tongue is not German do not have sufficient language skills, this represents a major challenge which can lead to learning deficits which are difficult to close (Gogolin, Neumann and Roth, 2004; cf. Johnstone & Selepeng, 2001). Students with a migration background might have less contact to the German language (c.f., Walter & Taskinen, 2007; Tiedemann & Billmann-Mahecha, 2007) and might therefore speak the language of instruction less well and need more concentration to process information (cf., van Merriënboer & Sweller, 2005; Sweller, 1994). The higher the linguistic diversity in class, the higher the state of disequilibrium – of contradictions and

discrepancies. In such classes there should be a high level of cognitive activity because students have to constantly compare, modify, or create new schemata. Students, which need to concentrate because of both processing discrepancies and decoding a language that they are not completely confident with, might face a cognitive overload. Thus, a high linguistic diversity in class is less beneficial for students with a migration background and less likely to foster their learning in reading comprehension.

However, the explanation via language only is not sufficient. A post-hoc analysis using the spoken language(s) instead of students' migration background revealed that the effects are not significant.⁵ Nevertheless, these non-significant effects have to be interpreted carefully. The tendency remains that students with more frequent contact to the German language show higher gains from linguistic diversity (see Figure 3). With a higher power on class level 2, the results might be significant – at least for the language(s) spoken with parents (already marginal significance with $p = .05$).

 Insert Figure 3 about here

7.1 Limitations

When interpreting our results, the following limitations have to be considered. First, our data only cover 5th grade students. Although our results are quite consistent with results of previous studies, future research has to show whether our results can be replicated for other grade levels. Studies in elementary school, in particular, might yield different results because linguistic diversity might affect early learning processes more strongly. Second, we must consider the specificity of our sample; 59% of the students had a migration background. On the one hand, our sample was highly suitable for the analysis since a high proportion of students with a migration background leads to diversity measures having a higher variance, rendering the analyses more stable. However, transferability to regions with

⁵ The results revealed $\beta = -.06$, with $p = .05$ for parents, and $\beta = -.01$, with $p = .39$ for friends using Simpson's D as a diversity measure. For the complete results, please refer to Supplementary Material *SM02_Complete Results Interaction Effects Language*.

fewer students with a migration background in schools (e.g., rural schools) can be seen critically. Furthermore, a subsample of the study participated in a training in readings strategies. Given that this program might have affected the impact of linguistic diversity on students' learning, we checked its influence in two ways. First, when analyzing the data, we controlled for participation in the program by including a dummy variable (i.e., training in reading strategies TRS) in our models. Second, we reran the same analyses adding an interaction term between linguistic diversity and program participation. The analyses indicated no effects for any of the models. Accordingly, there were no differences in the impact of linguistic diversity on students' reading comprehension between classes that did and did not participate in the program ($0.01 < \beta < .03$, $.67 < p < .89$). Another important limitation was the lack of teacher variables. A study by Decristan et al. (2017) showed that instructional quality has an influence on how performance heterogeneity impacts students' conceptual understanding. The relationship between performance heterogeneity and linguistic diversity has not been investigated so far and thus we do not know whether the mechanisms are similar. Future research has to show, if instructional quality plays a significant role in this regard.

7.2 Practical Implications

Our study shows that linguistic diversity plays an important role for students' reading comprehension, beyond the ethnic minority share of class. When in a state of disequilibrium, students have to use strategies to integrate new information into their understanding (cf. Hattie, 2012). There are many possible strategies (e.g., discussing with a teacher or with a learned, more experienced classmate) and students need to develop proficiency in choosing and using these strategies. Lee et al. (2009) showed that elementary teachers rarely discuss students' diversity in their own class or with colleagues in their schools. Furthermore, as already mentioned, teachers tend to neglect linguistic diversity in the classroom or simply do not have the methods to deal with it (e.g., Cho and McDonnough, 2009; Mazak & Herbas-Donoso, 2014). If teachers are not used to dealing with linguistic diversity, students might not have much practice in using such strategies. Despite this, our results show that linguistic diversity leads to an increase in reading comprehension performance and that it can mitigate the negative effect of a

class's ethnic minority share. Hence, it is important that teachers are trained in strategies for dealing with linguistic diversity (e.g., bilingual education programs, cf. Gogolin, 2011; Markic & Abels, 2014; Miller, 2009; Salleh, Venville and Treagust, 2007; guided transition by learning to use one's mother tongue to learn another language, cf. Tsung, 2012; implementation of language contact, cf. Driessen, 2002; Entwisle & Alexander, 1994; Peetsma et al., 2006; Scheele, Leseman, & Mayo, 2010). If a teacher actively uses the advantages of linguistic diversity in the classroom, it can additionally strengthen the positive effect for students' learning.

Our study also shows that linguistic diversity generally fosters the learning of all students, with students without a migration background benefiting even more. Even though the differences are less prominent when considering students' language(s) spoken with parents or friends, it is important to focus on these differential effects. Classes with a linguistically diverse student body clearly yield benefits. The linguistic diversity of a class depends on the ethnic minority share and thus, on each individual student with a migration background. To narrow the gap of educational inequalities among students with a migration background, it is essential that additional inequalities are not allowed. In order to compensate the differences, it is important that schools and their teachers promote interactions between students with and without migration backgrounds (see also Time-on-Task Hypothesis: Carroll, 1963, 1989). In relation to reading performance, in particular, teachers need to draw on linguistic diversity by encouraging conversations and engaging students in discussions in order to promote cultural and linguistic exchange (May, 2012).

7.3 Conclusion

In conclusion, our study shows that linguistic diversity promotes students' learning. Therefore, future empirical research should investigate the effects of linguistic diversity for additional grades and subjects. It is important that teachers become aware of the effectiveness of linguistic diversity and acquire strategies on how to use it in the classroom. In this respect, it is important that theoretical research examines the positive effects of linguistic diversity in greater depth, also beyond Piaget's concept of disequilibrium (1977), in order to develop suitable methods for teachers and school practice.

Table 1
Operationalization of diversity measures

Diversity Measure	Operationalization
Number of Languages <i>NLan</i> (e.g., Driessen, 2002)	$NLan = \frac{N(\text{languages per class})}{\text{class size}}$
Simpson's <i>D</i> (Simpson, 1949)	$D = 1 - \sum_{i=1}^C p_i^2$
Shannon's <i>H</i> (Shannon & Weaver, 1949)	$H = - \sum_{i=1}^C p_i \ln(p_i)$

Table 2

Descriptive statistics and correlations of the linguistic diversity measures

Variable	<i>n</i>	min	max	<i>M</i>	<i>SD</i>	skew	kurtosis	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) EthnMin	65	.26	1	.59	.17	.15	-.20	-						
(2) NLan_Pa	65	.15	.60	.35	.11	.14	-.66	.64*	-					
(3) SimD_Pa	65	.19	.80	.59	.14	-.98	.71	.79*	.82*	-				
(4) ShaH_Pa	65	.45	1.98	1.38	.38	-.61	-.14	.76*	.89*	.98*	-			
(5) NLan_Fr	65	.04	.31	.14	.07	.27	-.65	.51*	.47*	.55*	.54*	-		
(6) SimD_Fr	65	.00	.53	.22	.14	.04	-.89	.54*	.42*	.57*	.55*	.94*	-	
(7) ShaH_Fr	65	.00	1.23	.48	.31	.14	-.82	.52*	.41*	.56*	.55*	.96*	.99*	-

Note. EthnMin = ethnic minority share; NLan_Pa = relative number of languages spoken with parents; SimD_Pa = Simpson's *D* calculated based on language(s) spoken with parents; ShaH_Pa = Shannon's *H* calculated based on language(s) spoken with parents; NLan_Fr = relative number of languages spoken with friends; SimD_Fr = Simpson's *D* calculated based on language(s) spoken with friends; ShaH_Fr = Shannon's *H* calculated based on language(s) spoken with friends.

* $p < .05$

Table 3

Multilevel regression analyses predicting students' reading comprehension according to linguistic diversity measured based on spoken language with parents

Variables	Model 0 β (<i>SE</i>)	Model 1 β (<i>SE</i>)	Model 2 β (<i>SE</i>)	Model 3 β (<i>SE</i>)
Intercept	1.02* (.15)	.94* (.16)	.80* (.17)	.85* (.17)
Student Level				
Gender	.22* (.04)	.21* (.04)	.22* (.04)	.22* (.04)
Cultural Capital	.03 (.02)	.03 (.02)	.03 (.02)	.03 (.02)
Pretest Reading Comprehension	.59* (.02)	.60* (.02)	.59* (.02)	.59* (.02)
Migration Background	-.08 (.06)	-.07 (.06)	-.07 (.06)	-.07 (.06)
Language(s) Parents	-.02 (.03)	-.02 (.03)	-.00 (.03)	-.01 (.03)
Class Level				
TRS	-.09 (.07)	-.12 (.07)	-.11 (.07)	-.12 (.07)
EthnMin	-.17* (.03)	-.24* (.04)	-.26* (.05)	-.27* (.04)
NLan_Pa		.10* (.04)		
SimD_Pa			.11* (.05)	
ShaH_Pa				.13* (.04)
School Level				
WRS	-1.36* (.20)	-1.67* (.19)	-1.35* (.20)	-1.36* (.21)
RS	-.92* (.09)	-.86* (.08)	-.90* (.09)	-.86* (.09)
GS	-1.13* (.11)	-1.12* (.11)	-1.11* (.11)	-1.11* (.10)
R^2	.49	.50	.50	.49
AIC	3777.07	3784.83	3798.35	3793.42
BIC	3853.38	3865.52	3879.04	3874.12

Note. TRS = training in reading strategies; EthnMin = ethnic minority share; NLan_Pa = relative number of languages spoken with parents; SimD_Pa = Simpson's D calculated based on language(s) spoken with parents; ShaH_Pa = Shannon's H calculated based on language(s) spoken with parents; WRS = Werkrealschule; RS = Realschule; GS = Integrierte Gesamtschule.

* $p < .05$

Table 4

Multilevel regression analyses predicting students' reading comprehension according to linguistic diversity measured based on spoken language with friends

Variables	Model 0 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 3 β (SE)
Intercept	.95* (.15)	.88* (.15)	.95* (.16)	.97* (.15)
Student Level				
Gender	.22* (.04)	.22* (.04)	.22* (.04)	.22* (.04)
Cultural Capital	.03 (.02)	.03 (.02)	.03 (.02)	.03 (.02)
Pretest Reading Comprehension	.59* (.02)	.59* (.02)	.59* (.02)	.59* (.02)
Migration Background	-.06 (.05)	-.05 (.05)	-.05 (.05)	-.06 (.05)
Language(s) Friends	.00 (.03)	.02 (.03)	.02 (.03)	.01 (.03)
Between Level				
TRS	-.08 (.07)	-.12 (.07)	-.12 (.07)	-.12 (.07)
EthnMin	-.17* (.03)	-.22* (.03)	-.22* (.03)	-.22* (.03)
NLan_Fr		.10* (.03)		
SimD_Fr			.09* (.03)	
ShaH_Fr				.09* (.03)
School Level				
WRS	-1.36* (.21)	-1.32* (.21)	-1.30* (.21)	-1.29* (.20)
RS	-.92* (.09)	-.86* (.09)	-.90* (.09)	-.89* (.09)
GS	-1.13* (.11)	-1.12* (.11)	-1.11* (.11)	-1.10* (.11)
R^2	.49	.51	.49	.51
AIC	3768.49	3768.36	3817.40	3789.19
BIC	3861.80	3849.05	3898.09	3869.88

Note. TRS = training in reading strategies; EthnMin = ethnic minority share; NLan_Fr = relative number of languages spoken with friends; SimD_Fr = Simpson's D calculated based on language(s) spoken with friends; ShaH_Fr = Shannon's H calculated based on language(s) spoken with friends; WRS = Werkrealschule; RS = Realschule; GS = Integrierte Gesamtschule.

* $p < .05$

Table 5

Cross-level interaction between students' migration background and linguistic diversity using Simpson's *D* as an example.

Variables	Model 1 β (<i>SE</i>)	Model 2 β (<i>SE</i>)
Intercept	.33* (.10)	.30* (.09)
Student Level		
Gender	.21* (.04)	.22* (.04)
Cultural Capital	.03 (.02)	.03 (.02)
Pretest Reading Comprehension	.60* (.02)	.60* (.02)
Migration Background	-.05 (.05)	-.03 (.05)
Language(s) Parents	.02 (.03)	
Language(s) Friends		.01 (.02)
Between Level		
TRS	-.08 (.07)	-.04 (.06)
EthnMin	-.30* (.03)	-.26* (.04)
SimD_Pa	.18* (.06)	
SimD_Fr		.16* (.05)
School Level		
WRS	-1.70* (.20)	-1.59* (.21)
RS	-.88* (.09)	-.89* (.09)
GS	-1.08* (.11)	-1.07* (.11)
Cross-Level Interaction		
Migration Background*SimD_Pa	-.12* (.05)	
Migration Background*SimD_Fr		-.09* (.04)
<i>R</i> ²	.52	.52
AIC	2835.84	2824.52
BIC	2917.49	2906.08

Note. TRS = training in reading strategies; EthnMin = ethnic minority share; SimD_Pa = Simpson's *D* calculated based on language(s) spoken with parents; SimD_Fr = Simpson's *D* calculated based on language(s) spoken with friends; WRS = Werkrealschule; RS = Realschule; GS = Integrierte Gesamtschule.

**p* < .05

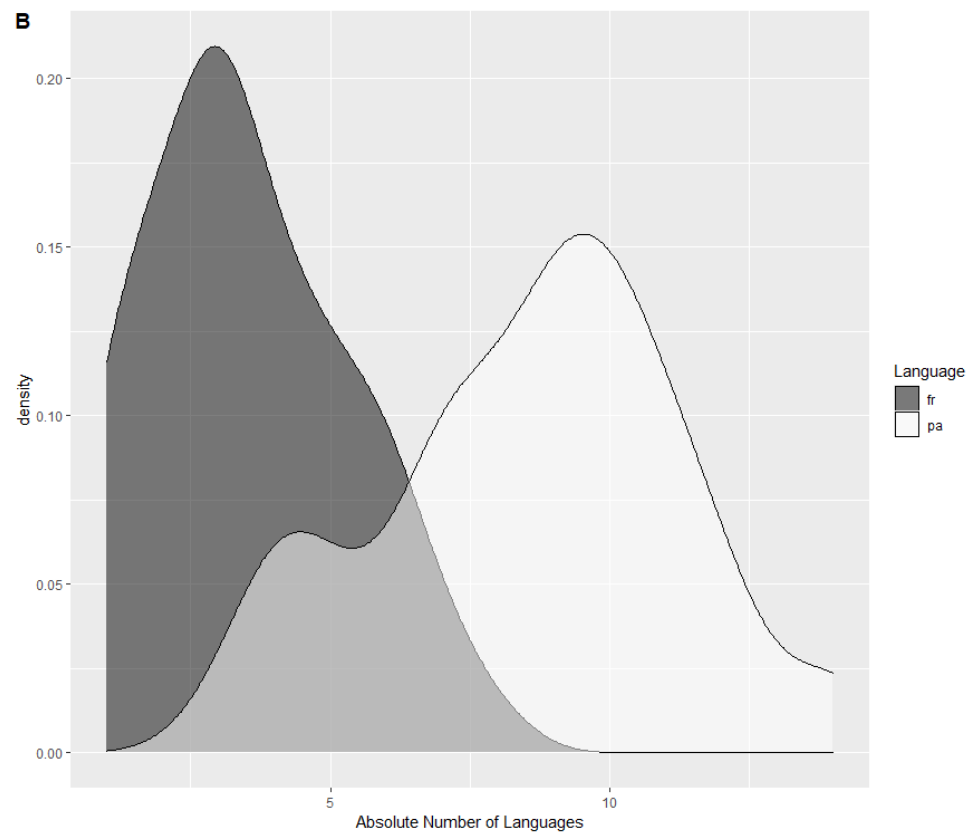
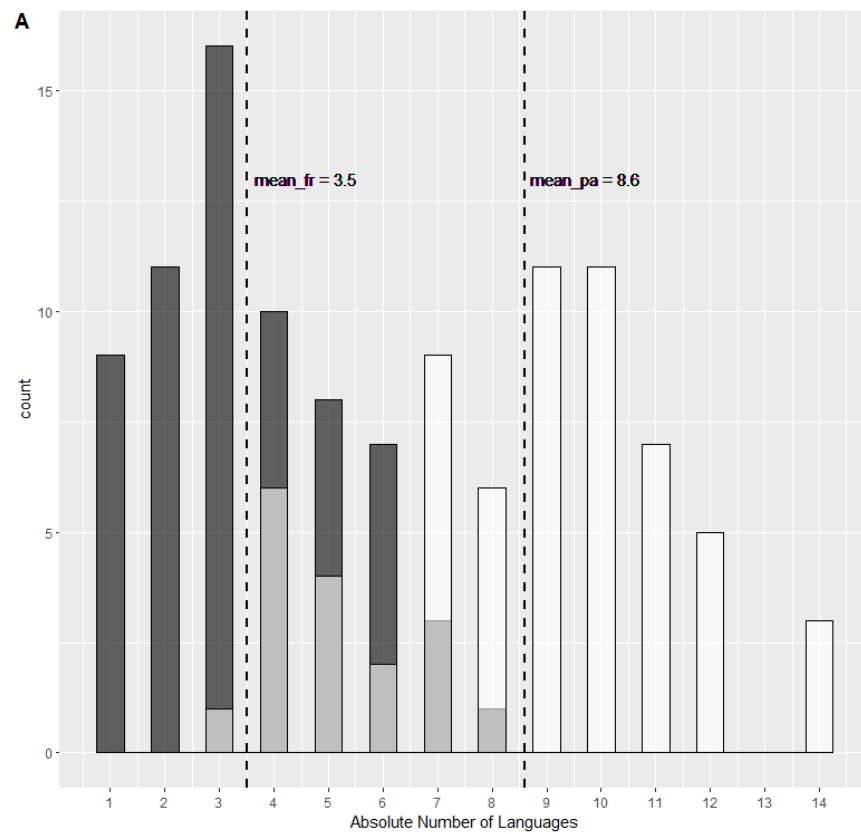


Figure 1. Descriptive characteristics of linguistic diversity using the absolute number of languages spoken with parents/friends as an example. (A) Overlaid histograms with means; (B) Overlaid density plots.

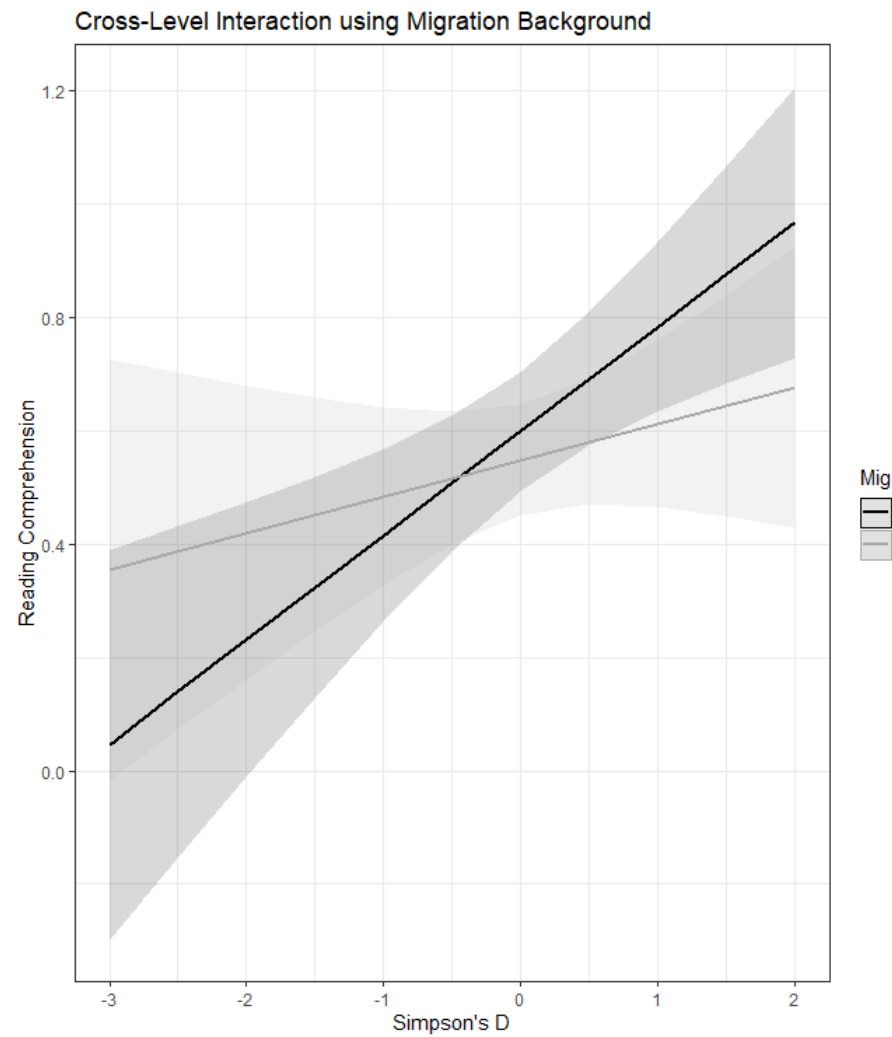


Figure 2. Cross-level interaction plot using migration background (0 = both parents born in Germany; 1 = at least one parent born abroad) and Simpson's D calculated using spoken language with parents.

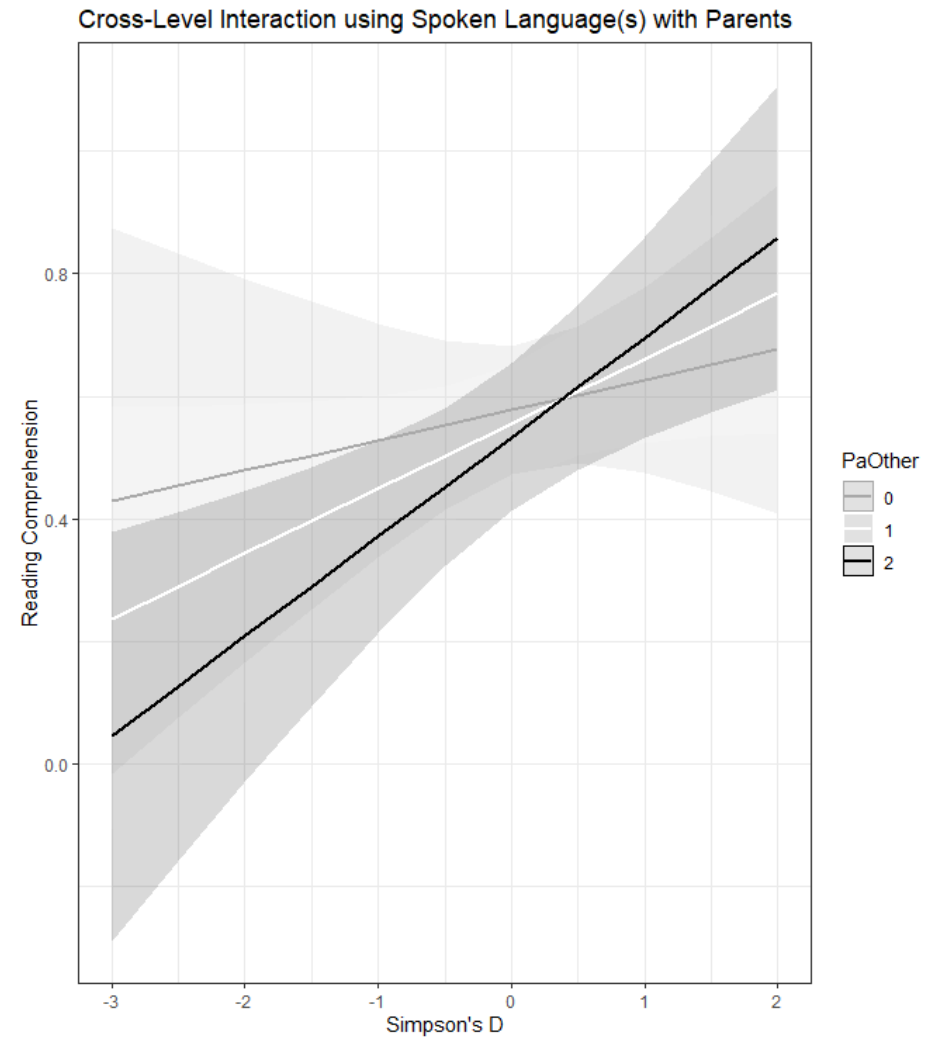


Figure 3. Cross-level interaction plot using spoken language with parents (0 = only other language(s); 1 = German language and other language(s); 2 = only German language) and Simpson's D calculated using spoken language with parents.

Literature

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Benner, A. D., & Crosnoe, R. (2011). The racial/ethnic composition of elementary schools and young children's academic and socioemotional functioning. *American Educational Research Journal*, 48, 621-646. doi:10.3102/0002831210384838
- Borman, K. M., Eitle, T. M., Michael, D., Eitle, D. J., Lee, R., Johnson, L., . . . Shircliffe, B. (2004). Accountability in a postdesegregation era: The continuing significance of racial segregation in Florida schools. *American Educational Research Journal*, 41, 605-631. doi:10.3102/00028312041003605
- Bruen, J., & Kelly, N. (2016). Language teaching in a globalised world: harnessing linguistic super-diversity in the classroom. *International Journal of Multilingualism*, 13(3), 333-352. doi:10.1080/14790718.2016.1142548
- Budescu, D. V., & Budescu, M. (2012). How to measure diversity when you must. *Psychological Methods*, 17, 215-227. doi:10.1037/a0027129
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.
- Carroll, J. B. (1989). The Carroll Model: a 25-year retrospective and prospective view. *Educational Researcher*, 18, 3-26. doi:10.3102/0013189X018001026
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62(2), 233-246. doi:10.1111/j.2044-8279.1992.tb01017.x
- Cho, S., & McDonnough, J. T. (2009). Meeting the needs of high school science teachers in English language learner instruction. *Journal of Science Teacher Education*, 20, 385-402. doi:10.1007/s10972-009-9136-9
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Cortes, K. E. (2006). The effects of age at arrival and enclave schools on the academic performance of immigrant children. *Economics of Education Review*, 25, 121-132. doi:10.1016/j.econedurev.2004.12.001
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 121-129.
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.), *Schooling and Language Minority Students: A Theoretical Framework*. Los Angeles: Evaluation, Dissemination and Assessment Center California State University.
- Decristan, J., Fauth, B., Kunter, M., Büttner, G., & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research*, 86, 109-121. doi:10.1016/j.ijer.2017.09.004
- Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. *Studies in Educational Evaluation*, 28, 347-368. doi:10.1016/S0191-491X(02)00043-3
- Dubowy, M., Duzy, D., Pröscholdt, M. V., Schneider, W., Souvignier, E., & Gold, A. (2001): Was macht den "Migrationshintergrund" bei Vorschulkindern aus? Ein Vergleich alternativer Klassifikationskriterien und ihr Zusammenhang mit deutschen Sprachkompetenzen [What constitutes the "migration background" among pre-school children? A comparison of alternative classification criteria and their relation to German language skills]. *Schweizerische Zeitschrift für Bildungswissenschaften*, 33(3), 355-376.
- Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence*, 21, 225-241. doi:10.1111/j.1532-7795.2010.00725.x
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138. doi:10.1037/1082-989X.12.2.121
- Entwisle, D., & Alexander, K. (1994). Winter setback: the racial composition of schools and learning to read. *American Sociological Review* (59), 446-460. doi:10.2307/2095943
- Eurostat. (2014). *Europe in figures: Eurostat yearbook*. Luxembourg: European Commission.
- Eurostat. (2019). *Migration and migrant population statistics*. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics#Migration_flows:_Immigration_to_the_EU_from_non-member_countries_was_2.4_million_in_2017

- Gkaintartzi, A., Kiliari, A., & Tsokalidou, R. (2015). 'Invisible' bilingualism – 'invisible' language ideologies: Greek teachers' attitudes towards immigrant pupils' heritage languages. *International Journal of Bilingual Education and Bilingualism*, 18(1), 60–72. doi:10.1080/13670050.2013.877418
- Gogolin, I. (2011). The challenge of super-diversity for education in Europe. *Education Inquiry*, 2(2), 239–249. doi:10.3402/edui.v2i2.21976
- Gogolin, I., Neumann, U., & Roth, H.J. (2004). *Förderung von Kindern und Jugendlichen mit Migrationshintergrund* [Promoting children and adolescents with a migration background]. Bonn: Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung.
- Goldsmith, P. R. (2011). Coleman revisited: School segregation, peers, and frog ponds. *American Educational Research Journal*, 48, 508–535. doi:10.3102/0002831210392019
- Gould, E. D., Lavy, V., & Paserman, M. D. (2009). Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *Economic Journal*, 119, 1243–1269. doi:10.1111/j.1468-0297.2009.02271.x
- Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60, 549–576. doi:10.1146/annurev.psych.58.110405.085530
- Gurin, P. Y., Dey, E. L., Gurin, G., & Hurtado, S. (2003). How does racial/ethnic diversity promote education? *The Western Journal of Black Studies*, 27, 20–29.
- Gurin, P., Dey, E. L., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review*, 72, 330–366. doi:10.17763/haer.72.3.01151786u134n051
- Hammarström, H. (2016). Linguistic diversity and language evolution. *Journal of Language Evolution*, 1(1), 19–29. doi:10.1093/jole/lzw002
- Hattie, J. (2012). *Visible Learning for Teachers*. Maximizing Impact on Learning. New York: Routledge.
- IBBW (2018). *Lernstand 5* [Learning Level 5]. Unveröffentlichtes Testinstrument [Unpublished test instrument]. Retrieved from <https://ibbw.kultus-bw.de/Lde/Startseite/Kompetenzmessung/Lernstand+5>
- Jensen, P., & Rasmussen, A. W. (2011). The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review*, 30, 1503–1515. doi:10.1016/j.econedurev.2011.08.002
- Johnstone, A. H., & Selepeng, D. (2001). A language problem revisited. *Chemistry Education Research and Practice*, 2, 19–29. doi:10.1039/B0RP90028A
- Kallens, P. C., Dale, R., & Christiansen, M. H. (in press). Linguistic diversity and individual variation: Comment on "Rethinking foundations of language from a multidisciplinary perspective" by T. Gong et al. *Physics of Life Reviews*. doi:10.1016/j.plrev.2018.04.004
- Kalyuga, S., Ayres, P., Chandler, P. and Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, 38(1), 23–31. doi:10.1207/S15326985EP3801_4
- Kumar, R. (2003). Multicultural education and achievement goal theory. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in Motivation and Achievement: Vol. 13. Motivating students, improving schools*. Greenwich, CT: JAI Press.
- Lee, O., Maaerten-Rivera, J., Buxton, C., Penfield, R., & Secada, W. G. (2009). Urban elementary teachers' perspectives on teaching science to English language learners. *Journal of Science Teacher Education*, 20, 263–286. doi:10.1007/s10972-009-9133-z
- Markic, S., & Abels, S. (2014). Heterogeneity and Diversity. A Growing Challenge or Enrichment for Science Education in German Schools?. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(4), 271–283. doi:10.12973/eurasia.2014.1082a
- Mazak, C., & Herbas-Donoso, C. (2014). Translanguaging practices at a bilingual university: A case study of a science classroom. *International Journal of Bilingual Education and Bilingualism*, 18(6), 1–17. doi:10.1080/13670050.2014.939138
- May, S. (2012). Educational Approaches to Minorities: Context, Contest and Opportunities. In A. Yia-koumetti (Ed.), *Harnessing Linguistic Variation to Improve Education* (5th ed.). Bern: Peter Lang.
- McCann, K. S. (2000). The diversity-stability debate. *Nature*, 405, 228–233. doi:10.1038/35012234
- Meyer, P., & McIntosh, S. (1992). The USA Today index of ethnic diversity. *International Journal of Public Opinion Research*, 4(1), 56–58. doi: 10.1093/ijpor/4.1.51

- Mickelson, R. A., Bottia, M. C., & Lambert, R. (2013). Effects of school racial composition on K-12 mathematics outcomes: A metaregression analysis. *Review of Educational Research*, 83, 121–158. doi:10.3102/0034654312475322
- Miller, J. (2009). Teaching refugee learners with interrupted education in science: vocabulary, literacy and pedagogy. *International Journal of Science Education*, 31, 571–591. doi:10.1080/09500690701744611
- Nemetz, P. L., & Christensen, S. L. (1996). The Challenge of Cultural Diversity: Harnessing a Diversity of Views to Understand Multiculturalism. *The Academy of Management Review*, 21(2), 434–462. doi:10.5465/amr.1996.9605060218
- Nettle, Daniel. 1999. *Linguistic diversity*. Oxford: Oxford University Press.
- OECD (2010). *PISA 2009 Results: What Students Know and Can Do*. Paris, OECD.
- Peetsma, T., van der Veen, I., Koopman, P., & van Schooten, E. (2006). Class Composition Influences on Pupils' Cognitive Development. *School Effectiveness and School Improvement*, 17(3), 275–302. doi:10.1080/13803610500480114
- Phalet, K., Andriessen, I., & Lens, W. (2004). How future goals enhance motivation and learning in multicultural classrooms. *Educational Psychology Review*, 16(1), 59–89. doi:10.1023/B:EDPR.0000012345.71645.d4
- Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. Oxford, England: Viking.
- Raudenbush, S. W., Fotiu, R. P., & Cheong, Y. F. (1998). Inequality of access to educational resources: A national report card for eighth-grade math. *Educational Evaluation and Policy Analysis*, 20, 253–267. doi:10.3102/01623737020004253
- Rjosk, C., Richter, D., Lüdtke, O., & Eccles, J. S. (2017). Ethnic composition and heterogeneity in the classroom: Their measurement and relationship with student outcomes. *Journal of Educational Psychology*, 109, 1188–1204. doi:10.1037/edu0000185
- Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley. doi:10.1002/9780470316696
- Salleh, R., Venville, G. J. & Treagust, D. F. (2007). When a bilingual child describes living things: an analysis of conceptual understanding from a language perspective. *Research in Science Education*, 37, 291–312. doi:10.1007/s11165-006-9027-4
- Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31(1), 117–140. doi:10.1017/S0142716409990191
- Schnepf, S. V. (2007). Immigrants' educational disadvantage: An examination across ten countries and three surveys. *Journal of Population Economics*, 20, 527–545. doi:10.1007/s00148-006-0102-y
- Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press. doi:10.1002/j.1538-7305.1948.tb01338.x
- Sierens, S., & Ramaut, G. (2018). Breaking Out of L2-Exclusive Pedagogies: Teachers Valorizing Immigrant Pupils' Multilingual Repertoire in Urban Dutch-Medium Classrooms. In P. V. Avermaet, S. Slem-brouck, K. Van Gorp, S. Sierens & K. Maryns (Eds.), *The Multilingual Edge of Education* (pp. 285–312). London: Macmillan Publishers. doi:10.57/978-1-137-54856-6
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163, 688. doi:10.1038/163688a0
- Statistisches Bundesamt (2018). *Bevölkerung und Erwerbstätigkeit, Bevölkerung mit Migrationshintergrund, Ergebnisse des Mikrozensus 2017* [Population and employment, population with migration background, results of the microcensus 2017]. Wiesbaden: destatis.
- Stipek, D. (2004). Teaching practices in kindergarten and first grade: Different strokes for different folks. *Early Childhood Research Quarterly*, 19, 548–568. doi:10.1016/j.ecresq.2004.10.010
- Sweller, J. (1994). Cognitive Load Theory, Learning Difficulty, and Instructional Design. *Learning and Instruction*, 4, 295–312. doi:10.1016/0959-4752(94)90003-5
- Sweller, J. (2005). Implications of Cognitive Load Theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). Cambridge, MA: Cambridge University Press.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296. doi:10.1023/a:1022193728205

- Tam, M. Y. S., & Bassett, G. W. (2004). Does diversity matter? Measuring the impact of high school diversity on freshman GPA. *Policy Studies Journal: The Journal of the Policy Studies Organization*, 32, 129–143. doi:10.1111/j.0190-292X.2004.00056.x
- Terry, N. P., & Irving, M. A. (2010). Cultural and Linguistic Diversity: Issues in Education. In R. P. Colarusso & C. M. Oourke (Eds.), *Special Education for All Teachers* (pp. 110-132). Iowa: Kendall Hunt Publishing.
- Tiedemann, J. & Billmann-Mahecha, E. (2007). Leseverständnis, Familiensprache und Freizeitsprache. Ergebnisse aus der Hannoverschen Grundschulstudie [Reading comprehension, family language and free-time language. Results from the Hannover primary school study]. *Zeitschrift für Pädagogische Psychologie*, 21, 41-49.
- Tsung, L. (2012). Rethinking Multilingual Education for Minority Students in China. In A. Yiakoumetti (Ed.), *Harnessing Linguistic Variation to Improve Education* (5th ed.). Bern: Peter Lang.
- United Nations Economic Commission for Europe (2006). Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing. New York: United Nations
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. doi:10.18637/jss.v045.i03
- Van Ewijk, R., & Sleegers, P. (2010). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, 21, 237–265. doi:10.1080/09243451003612671
- Van Merriënboer, J., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), 147-177. doi:10.1007/s10648-005-3951-0
- Walter, O. & Taskinen, P. (2007). Kompetenzen und bildungsrelevante Einstellungen von Jugendlichen mit Migrationshintergrund in Deutschland: Ein Vergleich mit ausgewählten OECD-Staaten [Competencies and educationally relevant attitudes of young people with a migration background in Germany: A comparison with selected OECD countries]. In PISA-Konsortium Deutschland (Ed.), *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie* [The results of the third international comparative study] (pp. 337-366). Münster: Waxmann.

Anhang C: Manuskript zu Beitrag 3 – Einstellung zur Leistungsheterogenität von Schüler*innen**Students' attitudes towards performance heterogeneity and their differential change
depending on contextual factors****1 Introduction**

The question of how to deal with performance heterogeneity is not a new topic. In recent years, schools have had to deal with a growing proportion of students with a migration background and, in turn, with greater differences in language skills within classrooms (Jensen and Rasmussen 2011; Maehler and Brinkman 2016). Furthermore, students with special educational needs are more often included in general education, increasing the heterogeneity of learning prerequisites (Markic and Abels 2014; Powell 2006).

Teachers' attitudes play a major role in how this growing heterogeneity is dealt with in the classroom and have been the focus of the prevailing literature (e.g., Dar 1985; Hartinger et al. 2010; Kumar and Hamer 2012; Linchevski and Kutscher 1998, Youngs and Youngs 2001). However, the class composition (e.g., performance heterogeneity, mean class performance level) does not only affect teachers, but also students. Regarding students' school-related attitudes, studies show that positive attitudes increase learning motivation, joy, and performance (e.g., Else-Quest et al. 2013; Gaith and Bouzeineddine 2003; Papanastasiou and Zembylas 2002). Accordingly, students perform especially well if they have a positive attitude towards school, a particular subject, or teacher.

Therefore, attitudes of students play an important role for educational practice and policy. The aim of this study was to analyze the *attitudes towards performance heterogeneity* (APH) among students. Based on a sample of 784 5th grade students, we examined how students' APH can be described and how students' APH can be influenced by individual and contextual factors.

2 Attitudes as psychological constructs

Attitudes can conceptually be defined as a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor (Eagly and Chaiken 1993; Krosnick and Petty 1995). According to the multi-dimensional theory, attitudes can be distinguished into three classes: cognition, affect, and behavior (Katz and Stotland 1959; Rosenberg and Hovland 1960). Combined, these three components represent the attitude construct. The cognitive category contains beliefs people have about the attitude object. The affective component consists of emotions people have related to the attitude object. The behavioral category encompasses people's intentions or actions with respect to the attitude object.

3 Attitude towards performance heterogeneity

It is already known that teachers perceive increasing performance heterogeneity as a professional difficulty (e.g., Baumert et al. 1997; Cho and DeCastro-Ambrosetti 2005; Dooly 2005; Kumar and Hamer 2012). Among students themselves, this question is still unanswered. An investigation of students' attitudes is of high relevance, because it is assumed that 25% of the variance of students' performance can be explained by attitudes towards the subject, towards self-belief, and towards the school environment (Bloom 1976). While the influence of students' attitudes towards the subject (e.g., Martinez et al. 2008; Papanastasiou and Zembylas 2002, 2004) and towards self-belief (e.g., Else-Quest et al. 2013; Gaith and Bouzeineddine 2011) have already been proven to positively influence students' performance, students' attitudes towards the school environment, such as performance heterogeneity, have played a minor role in previous research.

3.1 Attitude object: performance heterogeneity

In attitude theory, anything that can be discriminated can be evaluated and used as an attitude object (Bem 1972), whereby attitude objects can be either concrete or abstract (Eagly and Chaiken 1993). The attitude object *performance heterogeneity* can be defined as the distribution of population elements along a continuum of homogeneity to heterogeneity with respect to an ability or performance-related variable (cf. Lieberman 1969; Teachman 1980). Accordingly, students in rather homogeneous classes

have rather similar cognitive abilities or performance levels, and students in rather heterogeneous classes have divergent cognitive abilities or performance levels.

3.2 The impact of performance heterogeneity on students' APH

Performance heterogeneity plays an important role in educational research and practice, because teachers face the challenge of dealing with performance heterogeneity professionally. However, performance heterogeneity affects both teachers and students. Performance heterogeneity is a contextual circumstance in which students learn and interact with each other. Students in classes with different compositions have different experiences with performance heterogeneity; these experiences are expected to influence students' APH. Even if previous research has not focused on students' APH, the literature reports findings regarding teachers' APH and students' attitudes towards inclusion, which can be used as a theoretical foundation.

Studies investigating teachers' APH reveal that teachers show a more positive APH if they have already had contact with heterogeneity. Accordingly, a study by Linchevski and Kutscher (1998) showed that teacher participation in a project regarding mixed-ability grouping had a positive influence on their attitude towards teaching mixed-ability classes. Their analyses also show that teachers' attitudes became more positive the longer they participated in the project. Additionally, a study by Hartinger et al. (2010) compared the APH of teachers in mixed-age (heterogeneous) and same-age (homogeneous) classes in primary school. Their results revealed that teachers from mixed-age classes were significantly less likely to pursue a homogenization goal and more likely to view heterogeneity as an enrichment. Another study showed that teaching in heterogeneous classes led teachers to have a more positive APH (measured by a teachers' preference to teach a heterogeneous class in future) (Dar 1985). Furthermore, Byrnes and Kiger (1997) found similar results for attitudes towards language diversity, proving that such attitudes can be predicted by a teacher's previous experience. If teachers had already been in contact with linguistic-minority students, they had a more positive attitude towards language diversity.

Studies investigating students' attitudes towards inclusion of children with disability follow a similar pattern. Hence, a meta-analysis by Nowicki and Sandieson (2002) showed that in general students preferred their classmates to be children without disabilities, but students from integrative classes had more positive attitudes towards children with disabilities compared to their classmates in regular classes. Another meta-analysis, which was recently conducted by Armstrong et al. (2016), found that direct, extended, and guided imagined contact with students with disabilities (physical disability, learning disability, specific diagnoses) led to more positive attitudes towards students with disabilities among students.

Some of these studies claim that their results can be explained by the contact hypothesis according to Allport (1954) (e.g., Byrnes and Kiger 1997; Armstrong et al. 2016). The contact hypothesis states that frequent contact with groups or individuals towards which or whom prejudices have been formed reduces these prejudices and helps individuals develop a positive attitude. Contact is assumed to be particularly effective if people in the contact situation pursue cooperative goals, are of the same status, interact with each other, and if contact is supported by the authorities. As already stated, performance heterogeneity is a contextual circumstance that students are confronted with. According to the contact hypothesis, this could mean that students have a more positive APH if they are part of a heterogeneous class. Because students in a class pursue common goals (e.g., passing the class, acceptance for colleges), interact with each other (e.g., discussions, group work, joint assignments), and are supported by their teacher, performance heterogeneity is assumed to positively influence students' APH.

3.3 The impact of teacher behavior on students' APH

In addition to performance heterogeneity, a teacher's behavior of how to establish the learning environment might be significant for developing a positive APH in students. If a teacher cannot ensure that the disadvantages of heterogeneous classes (e.g., disturbances, under- and overstraining students) are limited and that the advantages (e.g. students of different cognitive abilities working together) are used for social and cognitive learning (Slavin 1990), students have no opportunity to gain positive experiences with performance heterogeneity and therefore cannot develop a positive APH.

Professionally dealing with performance heterogeneity is quite complex and difficult to implement. Teachers encounter several challenges with performance heterogeneity and must implement strategies in order to professionally manage it (Pietsch 2010). The state of research and the existing literature provide teachers with the following strategies: Firstly, teachers must establish a simple, well-structured learning environment through efficient classroom management (e.g., Emmer et al. 2006). Secondly, teachers must cultivate a learner-centered teacher-student relationship (e.g., Cornelius-White 2007). And, thirdly, teachers need to differentiate instruction in order to assure adequate learning support for students of different ability groups (e.g., Hattie 2009).

3.3.1 Classroom management

In order to positively influence students' APH, stable classroom management by teachers is vital. Classroom management is relevant for effective teaching and can be defined as "the ability to establish, maintain and (when necessary) restore the classroom as an effective environment for teaching and learning (Brophy 1986)". Literature suggests two strategies for effective classroom management: First, it is important to instruct students in rules (e.g., Emmer et al. 2006; McGinnis et al. 1995; Van de Grift 2007) and, second, teachers need to monitor student compliance with the rules by effectively managing disruptions (e.g., Anderson et al. 1979; Brophy and Evertson 1976; Evertson et al. 2006; Pedota 2007).

Clarification of rules with detailed instructions and modeling prevents a high amount of disturbance and gives students clear guidance (Evertson and Emmer 1982). This is essential for professionally managing heterogeneity and is assumed to positively influence students' APH. In addition, rapid intervention when students behave inappropriately is assumed to promote students' APH. Especially in heterogeneous classes the way students understand compliance and working behavior might highly vary; and it is important that teachers remind students about rules and provide extra instruction until students follow rules automatically (Brophy 1986). If teachers succeed in implementing both strategies – preventing and effectively managing disturbances – they create a fruitful basis for successful learning, which enables students to develop a positive APH.

3.3.2 Learner-centered approach

Furthermore, students' APH might be positively influenced by teachers using a learner-centered teaching approach. The learner-centered approach connects a focus on individual learners with a focus on learning (McCombs and Whisler 1997). It involves teacher and students interacting successfully – interaction that is characterized by sensitivity, respect, and quality of relationship (e.g., Cornelius-White 2007; Kunter and Voss 2011; Pianta and Hamre 2009). To ensure a learner-centered environment, two strategies are particularly appropriate: First, teachers orient themselves towards temporal reference norms (e.g., Dickhäuser et al. 2017; Rheinberg 1980) and, second, teachers foster a positive error culture in the classroom (e.g., Keith and Frese 2008; Nordstrom et al. 1998; Rach et al. 2013).

A reference norm orientation (RNO) applies to the assessment of students' performance and distinguishes between three approaches: criterion-based RNO (comparing individual performance with an absolute standard), social RNO (comparing individual performance with the performance of other students), and temporal RNO (comparing individual performance with past results) (Rheinberg 1980). In order to positively influence students' APH, teachers should apply temporal RNO. In heterogeneous classes, differences in performance are highly pronounced and, therefore, more easily noticed by students. This can lead to a larger gap between students and tensions in the class, negatively influencing students' APH. However, if teachers highlight individual learning development, they blur differences in ability, shifting students' focus from social comparison to their own learning growth; and thus, students are more likely to develop a positive APH. Another learner-centered approach which supports a positive APH among students is an error-friendly feedback culture. Because in heterogeneous classes students' cognitive abilities vary substantially, comprehension problems, misconceptions, and wrong answers may occur more frequently. Especially the latter trigger a feeling of unpleasantness in students and increases emotional discomfort (Kunter and Trautwein 2013). To promote students' APH, it might be crucial that teachers integrate these errors in order to explain thought processes and stimulate discussions (Tulis 2013). In doing so, teachers derive a benefit from class heterogeneity and provide students with the opportunity to develop a positive APH.

3.3.3 Differentiated instruction

A final aspect for fostering a positive APH among students is differentiated instruction. Differentiated instruction adapts to the individual ability level of each student, with teaching materials designed in a differentiated and stimulating way (e.g., Hattie 2009; King-Sears 2008; Vaughn et al. 2001).

Students are frequently under- and overstrained in everyday school life, leading to boredom and monotony, which are perceived as the strongest school-related problems for students (e.g., Larson and Richards 1991; Perkins and Hill 1985). These factors play a role especially in heterogeneous classes, where students are more often under- and overstrained. By didactically focusing on the individual learner (e.g., enabling discussions and groupwork between students of different ability levels, differentiating individual work in difficulty levels), teachers deal professionally with heterogeneity and cognitively activate each individual student. In this way, each student benefits from the heterogeneity of the class, which is assumed to positively influence APH.

4 Hypotheses

To date, there has been no literature regarding students' APH. However, in Bloom's Theory of School Learning (1976) it is assumed that students' attitudes towards the school environment influence their learning and performance. Therefore, the aim of our study was to examine students' APH and its change depending on contextual factors.

In the following, we present our hypothesis concerning performance heterogeneity:

1. Performance heterogeneity in the classroom positively influences students' APH

Based on the contact hypothesis according to Allport (1954), we assumed that contact with an attitude object would have a positive influence on the attitude itself. Studies on attitudes towards APH among teachers (e.g., Byrnes and Kiger 1997; Dar 1985) as well as studies on attitudes towards inclusion among students (e.g., Armstrong et al. 2016; Avramidis and Norwich 2002) suggest this pattern, and we therefore expected the same pattern for APH among students. We assumed that students from classes with higher heterogeneity would show a more positive APH.

In the following, the hypotheses regarding classroom management are presented:

2a. Implementing rules positively influences students' APH

2b. Effectively managing disturbances positively influences students' APH

Both strategies contribute to effective classroom management, which is an important foundation for students to develop a positive APH. If teachers prevent disturbances using rules (Hypothesis 2a) and respond to disturbances effectively (Hypothesis 2b), students should have a more positive APH.

In the following, the hypotheses related to the learner-centered approach are presented:

3a. A temporal RNO positively influences students' APH

3b. A positive error culture positively influences students' APH

Both strategies contribute to a learner-centered environment. Teachers applying a temporal RNO (Hypothesis 3a) level out differences in performance between students and focus on individual learning development. Teachers fostering a positive error culture (Hypothesis 3b) integrate errors into their instruction. Both strategies are expected to lead to a more positive APH among students.

In the following, we present our hypothesis regarding differentiated instruction:

4. Differentiated instruction positively influences students' APH

Differentiated instruction reduces monotonous teaching phases. Furthermore, orientation towards individual performance levels increases the cognitive activity of students. We thus assumed that professionally managing performance heterogeneity by differentiating instruction has a positive effect on students' APH.

5 Method

5.1 Sample and design

At the beginning of the school year 2018/2019, a survey of students and their teachers was conducted. Each student took part in a language proficiency test and subsequently completed a questionnaire. In the student questionnaire, data on individual demographics, students' APH, and contextual teaching variables (i.e., implementation of rules, disturbance management, temporal RNO, positive error culture) were collected. The sample consisted of $N = 784$ 5th grade students (50% female; $M_{\text{Age}} = 10.20$; $SD = .51$) from 33 classes. Their teachers (65% female; on average between 35 and 45 years old) were

interviewed on their individual demographics and the extent to which they differentiate their instruction ($N = 21$ because of low response rate).¹ The data were collected in 12 secondary schools (eight Gymnasien, two Realschulen, two Gemeinschaftsschulen). The average class consisted of 25 students; 65% of the students had a migration background, which means that at least one parent was born abroad (Statistisches Bundesamt 2018).

5.2 Instruments

5.2.1 Attitude towards performance heterogeneity

APH was measured multidimensionally using three separate scales for the cognitive, affective, and behavioral components. To measure the attitude components, an agreement-rejection dimension was used with respect to the attitude object *performance heterogeneity*. The items are based on a six-level Likert scale and have a bipolar structure. The two respective poles are provided with response categories and differ in their valence. The higher the agreement, the more positive the attitude of students towards performance heterogeneity. Students' APH was measured based on their attitude in the subject German. The cognitive attitude dimension was measured with a total of 13 items reflecting students' thoughts about performance heterogeneity (e.g.: I think a class in which the students' performance in German differs is... [bad] - [good]). The affective component was measured using 15 items reflecting students' emotions towards performance heterogeneity (e.g.: In a class in which students' performance in German differed, I would be... [stressed] - [relaxed]). The behavioral component was measured using five items to capture the intentional behavior of students (e.g.: If I had the choice, I would prefer to be in a class in which students perform... [similarly to me] - [differently to me]).

5.2.2 Performance

An adapted version of the Duisburg language proficiency test was used to collect performance data (Theunissen et al. 2005). This adapted version comprises four subscales: listening comprehension, grammar, sentence structure, and vocabulary. Listening comprehension ($\alpha = .73$) was assessed using an audio recording which the students listened to once. Afterwards students were allowed to read and

¹ Results of a t-test revealed that there were no significant differences in performance heterogeneity ($p = .23$) between the classes for which the teachers responded ($M = 9.88$) and those for which the teachers did not respond ($M = 10.89$).

answer ten single-choice questions (each with four possible answers) within eight minutes. The students' grammar performance ($\alpha = .82$) was measured by combining two different tests. In Test A, students had to complete a sentence with the grammatically correct form using one or several given words. Test B consisted of ten single-choice questions (each with four possible answers) which required students to complete a sentence using the grammatically correct form. The sentence construction test ($\alpha = .75$) consisted of 40 items in a binary 0-1 format. Students had to decide whether the sentence structure was correct or incorrect. In the vocabulary test ($\alpha = .86$), the students had to find a synonym for a given word. It was based on a single-choice format with four possible answers for a total of 20 items.

The four subtests were weighted to allow a maximum of 20 points each. Together, they represented students' final language proficiency score. Thus, the maximum test score was 80 points. Raw data (dichotomously coded) formed the basis for scaling according to the 2-PL IRT model. Scaling was performed with R (R Core Team 2019). No items had to be excluded due to poor item fits ($WMNSQ > 1.2$ or < 0.8).

5.2.3 Performance heterogeneity

Performance heterogeneity was calculated with the intra-class standard deviation *SD* based on the final score of the Duisburg language proficiency test.

5.2.4 Teacher behavior

Students assessed the extent that rules apply in class ($n_{\text{Items}} = 4$; $\alpha = .80$; $ICC1 = 5.99\%$; $ICC2 = .61$; e.g.: With my teacher, everyone knows the rules that must be followed. 1 = not true at all, 4 = absolutely true) and the teacher's effective management of disturbances ($n_{\text{Items}} = 4$; $\alpha = .71$; $ICC1 = 13.15\%$; $ICC2 = .79$; e.g.: My teacher ends disturbances by students quickly and without discussion. 1 = not true at all, 4 = absolutely true) (Piwowar 2013). Students also assessed teachers' use of temporal RNO ($n_{\text{Items}} = 4$; $\alpha = .76$; $ICC1 = 0.94\%$; $ICC2 = .20$; e.g.: If I make a special effort, my teacher usually praises me, even if other students are better than me. 1 = not true at all, 4 = absolutely true) (Jerusalem et al. 2009) and fostering of a positive error culture by the teacher ($n_{\text{Items}} = 4$; $\alpha = .78$; $ICC1 = 3.61\%$;

ICC2 = .48; e.g.: If I make mistakes in class, my teacher discusses them with me in such a way that it really helps me. 1 = not true at all, 4 = absolutely true) (Piwovar 2013). Differentiated instruction was assessed by the teachers themselves. Teachers were asked to assess the extent to which they differentiate instruction in class ($n_{\text{Items}} = 11$; $\alpha = .91$; e.g.: When students work individually, I vary the task to meet the needs of students with varying abilities. 1 = do not agree at all, 4 = agree absolutely) (based on: Baumert et al. 2008; Rheinberg 1980).

5.3 Data analysis

First of all, a confirmatory factor analysis (CFA) was performed to uncover basic relationships between the attitude items and to analyze the structure of students' APH. In this process, we controlled for class affiliation. The respective items model the three components: cognitive, affective, and behavioral attitudes. Together the three subcomponents represent the latent construct *attitudes towards performance heterogeneity* (APH) among students. Secondly, descriptive analyses were conducted to identify the value of students' APH and to analyze the relationship between the three attitude components. Finally, to predict students' APH depending on contextual factors (Hypotheses 1-4), we calculated seven multi-level structural equation models. The loadings of the attitude variables and attitude components were fixed at level 1 and level 2.

Model 0, the baseline model, examined the effects of the individual demographics, which we defined as control variables – gender (0 = male; 1 = female), migration background (0 = both parents born in Germany; 1 = at least one parent born abroad), and cultural capital (how many books do you have at home? 1 = None or only very few; 5 = Over 200 books) – on students' APH. Model 1 addressed Hypothesis 1: the assumption that performance heterogeneity in the classroom would have a positive influence on APH. Here, individual performance and average class performance were used as additional control variables. Furthermore we controlled for school type by including binary school type variables as predictors for performance heterogeneity and average class performance (see Figure 1). Models 2-6 examined the differential change of students' APH depending on the teachers' professional management of performance heterogeneity. Models 2-6 extended Model 1, adding one predictor each.

Models 2-5 included the respective predictors on student level 1 and class level 2, Model 6 on class level 2 only.

Individual test performance was centered at the group mean because it depended on the respective class (Enders and Tofighi 2007). All other independent continuous variables were z-standardized. We used the R package lavaan to run our models (Rosseel 2012). In all models we imputed for missing data using *fmil*.

6 Results

Based on the CFA, we excluded one item (*c_10*) because of its poor fit ($\lambda = .19$).² The CFA showed an acceptable model fit with a Comparative Fit Index (CFI) = .928, Tucker Lewis Index (TLI) = .922, Root Mean Square Error of Approximation (RMSEA) = .047 (Browne and Cudeck 1993; Byrne 1994), and a relative Chi²-Index with $\chi^2/df = 2.593$ (Kline 1998; Ullman 2001). The calculation of the attitude components showed very good reliabilities with $\alpha_{\text{cog}} = .91$, $\alpha_{\text{aff}} = .95$ and $\alpha_{\text{beh}} = .90$ (see Table 1).

6.1 Descriptive analyses

The descriptive statistics revealed a positive value for students' attitudes towards performance heterogeneity (Table 1). With a theoretical mean of 3.5 (1 = negative attitude; 6 = positive attitude), the cognitive ($M = 4.1$) and affective ($M = 4.3$) components were in the positive area of the scale. Hence, students showed rather positive beliefs and emotions towards performance heterogeneity. The behavioral component had a mean value of $M = 3.5$, the exact mean of the scale. Furthermore, Figure 2 shows the distribution function of the respective attitude component and reveals differences between the cognitive ($SD = 1.1$) and affective ($SD = 1.1$) components and the behavioral ($SD = 1.6$) component. We found positive correlations between the attitude components (see Table 1). Students who have positive beliefs about performance heterogeneity are more likely to show positive emotions about performance heterogeneity ($r_{\text{cog_aff}} = .64$; $p < .001$) and are more likely to choose a heterogeneous class ($r_{\text{cog_beh}} = .22$; $p < .001$; $r_{\text{aff_beh}} = .27$; $p < .001$).

² We provide a table with the factor loadings of the confirmatory factor analysis in the supplementary material online. Please see *SM01_Standardized Factor Loadings*.

6.2 Contextual effects

Table 2 shows the effects of contextual factors on students' APH. Model 0, the baseline model, determined the influence of the control variables (i.e., students' individual demographics) on APH. Models 0-6 revealed that none of the students' individual background characteristics influenced APH (except students' migration background in Model 0 ($\beta = .11$; $p < .05$) and Model 6 ($\beta = .22$; $p < .05$)).

6.2.1 Effect of performance heterogeneity

Hypothesis 1 examined whether a classes performance heterogeneity predicts students' APH (see Model 1, Table 2). The results revealed that performance heterogeneity in a class leads to a positive increase in students' APH ($\beta = .09$; $p < .05$). This effect remains stable when controlling for individual performance ($\beta = -.00$; $p = .607$) and average class performance ($\beta = .01$; $p = .877$), both of which do not significantly influence APH.

6.2.2 Effects of teacher behavior

Hypotheses 2a and 2b examined how effective classroom management influences students' APH. Model 2 showed that if teachers implement rules in the classroom, students show a higher APH ($\beta = .17$; $p < .001$). Model 3 revealed that if teachers apply an effective approach to manage disturbances, students' APH increases ($\beta = .10$; $p < .01$). Hypotheses 3a and 3b investigated the impact of a learner-centered teaching approach on students' APH. Model 4 showed that a temporal RNO increases students' APH ($\beta = .13$; $p < .001$). Model 5 revealed that students show a more positive APH if their teacher fosters a positive error culture ($\beta = .07$; $p < .05$). Alongside the positive effects on level 1, the results revealed additional compositional effects. Thus, beyond students' individual perception, a classes shared perception of their teacher implementing rules ($\beta = .22$; $p < .10$), managing disruptions ($\beta = .17$; $p < .10$), adopting a temporal RNO ($\beta = .20$; $p < .05$), and implementing a positive error culture ($\beta = .25$; $p < .05$) increases students' APH. Model 6 focused on the professional management of performance heterogeneity through differentiated instruction, which was found to have a positive effect on students' APH. Hence, if teachers differentiate their instruction more, students show a more positive APH ($\beta = .12$; $p < .05$).

In Model 6, performance heterogeneity no longer had a positive influence ($\beta = .05$; $p = .486$). To examine this diminishing influence, we conducted a post-hoc mediation analysis (see Figure 3). In this analysis, we assumed that performance heterogeneity would influence the extent to which a teacher differentiated instruction, which in turn predicts students' APH. The results confirmed that higher performance heterogeneity in a class increases the extent of differentiated instruction ($\beta = .66$; $p < .01$) and that differentiated instruction leads to a more positive APH ($\beta = .13$; $p < .01$).

7 Discussion

Our analyses reveal four important findings. First, students have a rather positive APH. Second, compared to contextual factors, individual variables play a minor role in explaining students' APH. Third, a heterogeneous class composition leads to a more positive APH among students. Fourth, if teachers apply a professional approach managing performance heterogeneity, it increases students' APH.

Our findings show that students have a rather positive APH. Accordingly, students exhibit positive beliefs, positive emotions, and an average behavioral intention regarding performance heterogeneity. The three components correlate positively and highly significantly. Thus, students with more positive beliefs about performance heterogeneity also show more positive emotions and are more positive about choosing a heterogeneous class. This confirms the basic assumption of the consistency theorem in multidimensional attitude theory (Eagly and Chaiken 1993) that the different components are generally consistent.

However, the behavioral component differs slightly from the cognitive and affective component with regard to its mean and distribution. The descriptive characteristics show that the response pattern changes when students assess their intentional behavior regarding a heterogeneous class. The distribution function shows that students do not clearly prefer a moderately heterogeneous class – there rather is a similar number of students who would choose a homogeneous, heterogeneous, or moderately heterogeneous class. The difference between the behavioral component and the cognitive and affective component is also revealed by the respective correlations. Although positive correlations were found between the attitude components, the cognitive and affective components were much more strongly

related than the behavioral component. Attitude research has extensively discussed the fact that the behavioral component deviates from the cognitive and affective component (e.g., Eagly and Chaiken 1993; Fishbein and Aizen 1975; Rosenberg 1960). The Construal Level Theory assumes that psychological distances and mental abstractions influence each other. Individuals think more abstractly about distant objects and more concretely about close objects (Libermann and Trope 2008). *Believing* that a heterogeneous class is good or bad, and *feeling* that a heterogeneous class is good or bad is a distant, abstract construct. As soon as students are asked to *decide* on a certain class composition, the attitude object *performance heterogeneity* is evaluated based on their current individual state. As a result, students have higher attitude values for the cognitive and affective component, and some students would choose a homogeneous class despite their positive cognitive and affective attitude.

Another central finding of our analyses is the influence of performance heterogeneity itself on students' APH. In accordance with the contact hypothesis, being part of a heterogeneous class positively affects students' APH (Allport 1954). Heterogeneity is a critically discussed topic in the media and everyday school life. Students from heterogeneous classes might be able to weigh up the advantages of heterogeneity and be less affected by negative stereotypes. Our results reveal one exception: Performance heterogeneity does not have a significant impact if the extent of differentiated instruction is taken into account. A post-hoc mediation analysis provided the deeper insight that performance heterogeneity has a positive impact on the extent of differentiated instruction, which leads to a growth of students' APH.

An additional central finding of our analysis is the influence of teachers' behavior on students' APH. Especially students in heterogeneous classes strongly differ in performance. Disturbances might arise more often, because students feel bored or overstrained. Furthermore, incorrect answers might occur more frequently, comprehension questions might be asked more often, and the time to finish an assignment might vary considerably. Each of these factors can lead to a negative APH among students, because students have difficulties dealing with such characteristics of heterogeneity. Therefore, it is important that teachers themselves implement strategies to professionally deal with heterogeneity. In

doing so, they stress the advantages derived from a heterogeneous context and increase the likelihood that students develop a positive APH.

In that respect, our results prove that professional management of performance heterogeneity by teachers – classroom management, a learner-centered teaching approach, and differentiated instruction – increases students' APH. Our results reveal that students' APH increases if teachers effectively manage their classrooms by implementing rules and dealing with disturbances. In this way, teachers have more capacity for performing instruction, supporting individual learning, mobilizing groups, and facilitating social negotiation processes (Doyle 2006; Evertson et al. 2006). As already stated, disturbances might negatively influence students' APH. However, if teachers keep this disadvantage to a minimum (e.g., by dealing with disturbances using eye contact or by involving the respective students), students might not be irritated by heterogeneity and be more likely to develop a positive APH. In addition, our results show that a learner-centered teaching approach integrating a temporal RNO and a positive error culture positively influences students' APH. Performance-related differences between students can increase the separation between high and low-performing students and tense the classroom atmosphere, leading to a negative APH among students. If teachers give performance-related feedback using temporal RNO, they focus on the development of students' individual performance instead of inter-individual comparisons; this feedback approach is assumed to result in students developing a positive APH. However, the results regarding RNO must be interpreted carefully: the ICC1 and ICC2 are very low, hinting that students within a class do not resemble one another regarding their perception of a teachers RNO. The effect of RNO on class level 2 on students' APH might be estimated inaccurately, and a careful interpretation is needed. Besides RNO, teachers can actively incorporate incorrect answers into instruction using a positive error culture to support students' APH. By fostering a positive error culture, teachers convey their appreciation of misunderstandings and misconceptions to students, enabling students to feel comfortable in a heterogeneous class. Furthermore, differentiated instruction was found to positively influence students' APH. When applying differentiated instruction, teachers

keep in mind the complexity and variety of teaching-learning environments. This is especially important in heterogeneous classes, in which student performance highly differs. Some studies show that differentiated instruction has advantages for social learning compared to traditional teaching (e.g., Lüders and Rauin 2004; Trautmann and Wischer 2011). Especially cooperative approaches such as peer-assisted learning strategies or reciprocal teaching involving tutors and tutees of different ability levels (Chatoupis and Vagenas 2018; Fuchs et al. 1997; Hattie 2009; Palincsar and Brown 1984; Pilten 2016) stimulate classroom settings and lead students to perceive performance-related heterogeneity in the classroom positively.

Besides baseline Model 0, Model 6 reveals a significant positive effect of students' migration background on APH. This effect must be interpreted carefully. The sample's composition regarding migration background in this study might differ to that of other studies, because students' countries of origin are very diverse in our study. Researchers have already shown that when analyzing students' migration background differential effects must be considered, since the effects of subgroups from different countries of origin differ (e.g., Müller and Stanat 2006; Segeritz et al. 2010). In our study, origin-related effects could not be identified because of the small sample size; and before putting forward alternative explanations, the effect should be replicated in future studies.

7.1 Limitations

When interpreting our results, the following limitations have to be considered. First, our data were collected in the first quarter of the school year; thus, students had been in the same class for only about two months. However, the variance of the attitude scale and the systematic correlations indicate that students had already gained an impression of the performance heterogeneity in their classroom. Secondly, our data only cover 5th grade students. Although a preliminary study with guided interviews (Authors, unpublished) has shown that even young students can grasp the performance heterogeneity construct, it remains questionable if older students show the same patterns. Furthermore, we must consider the specificity of our sample, in which 65% of the students had a migration background and transferability to regions with fewer students with migration backgrounds in schools (e.g., rural

schools) can be seen critically. Another limitation was the focus on the subject German. Although studies show that language proficiency is a good predictor for school success and performance in other subjects (Watson et al. 2003), we cannot assume that students' APH would be similar for mathematics or a foreign language. Future analyses are needed to reveal whether the results are comparable for these subjects. Another important limitation is the lack of the variable *teachers' APH*. Studies have already shown that teachers of heterogeneous classes show more positive attitudes (e.g., Dar 1985; Hartinger et al. 2010; Linchevski and Kutscher 1998). A teachers' APH can reinforce the effect of performance heterogeneity on students' APH, since the teacher plays an important role in dealing with this heterogeneity. Finally, the relatively small number of classes should be considered when interpreting our results. The effects of performance heterogeneity and differentiated instruction are affected by class size. In comparison to Models 2-5, there are no student level 1 data available (accordingly, averaging outliers is not possible). Hence, follow-up analyses with a larger sample at class level are necessary to replicate our findings.

7.2 Practical implications and conclusions

Our study shows that APH tends to be positive among students. This positive attitude can be used by teachers to cultivate a pleasant atmosphere and positive learning environment in the classroom. It is of high relevance that individual background characteristics are not as important for students' APH as their teacher's behavior. This is an advantage for learning and instruction, because individual characteristics cannot be influenced by teachers and therefore cannot be used to change students' APH. Furthermore, it is of great importance that teachers can positively change their students' APH if they implement classroom management, learner-centered teaching, and internal differentiation in their lessons. Numerous studies have already proven that the application of these methods in the classroom enhances students' learning (e.g., Kunter and Voss 2011; Pianta and Hamre 2009; Rakoczy et al. 2007). Furthermore, students' learning improves with positive school-related attitudes among students (e.g., Else-Quest et al. 2013; Gaith and Bouzeineddine 2003; Papanastasiou and Zembylas 2002). Teachers can thus have a dual effect on students' learning: first, the direct effect of creating a clear, positive, and

differentiated learning environment; second, the indirect effect of these factors, through which teachers support learning by positively changing students' APH.

Table 1

Psychometric properties and correlations for the cognitive, affective, and behavioral attitude component

	n	min	max	mean	sd	skew	kurtosis	α	cog	aff	beh
cog	720	1	6	4.11	1.06	-.14	-.24	.91	-		
aff	724	1	6	4.30	1.12	-.50	-.34	.95	.64*	-	
beh	719	1	6	3.52	1.58	-.01	-1.06	.90	.22*	.27*	-

Note. cog = cognitive component; aff = affective component; beh = behavioral component.* $p < .001$

Table 2

Multilevel structural equation models predicting students' attitudes towards performance heterogeneity

Path		Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Within Class Level								
APH	← Gender	.06 (.04)	.05 (.06)	.07 (.05)	.08 (.05)	.04 (.05)	.06 (.05)	.05 (.08)
APH	← Mig	.11* (.06)	.10 (.06)	.06 (.06)	.04 (.06)	.05 (.06)	.05 (.06)	.22* (.09)
APH	← CultCap	.00 (.02)	.00 (.03)	.01 (.02)	.01 (.02)	-.00 (.02)	.01 (.03)	-.01 (.04)
APH	← Performance		-.00 (.00)	-.00 (.00)	-.00 (.00)	-.03 (.03)	-.05 (.03)	-.00 (.01)
APH	← Rules			.17* (.05)				
APH	← Disruptions				.10* (.03)			
APH	← TempRNO					.13* (.03)		
APH	← Error Culture						.07* (.03)	
Between Class Level								
APH	← PH		.09* (.05)	.10* (.04)	.10* (.04)	.11* (.04)	.08~ (.04)	.05 (.07)
APH	← PAVg		.01 (.05)	.00 (.04)	.01 (.04)	.02 (.04)	-.03 (.04)	-.01 (.07)
APH	← Rules			.22~ (.12)				
APH	← Disruptions				.17~ (.09)			
APH	← TempRNO					.20* (.10)		
APH	← Error Culture						.25* (.10)	
APH	← Differentiation							.12* (.05)
PH	← Realschule		-.05 (.25)	-.06 (.25)	-.06 (.25)	-.06 (.25)	-.06 (.25)	-.14 (.30)
PAvg	← Realschule		-.05 (.25)	-.06 (.25)	-.06 (.25)	-.06 (.25)	-.06 (.25)	-.14 (.30)
PH	← Gesamtschule		-.32 (.23)	-.32 (.23)	-.32 (.23)	-.32 (.23)	-.32 (.23)	-.28 (.30)
PAvg	← Gesamtschule		-.32 (.23)	-.32 (.23)	-.32 (.23)	-.32 (.23)	-.32 (.23)	-.28 (.30)
Covariances								
PH	↔ PAVg		-.72* (.21)	-.72* (.21)	-.72* (.21)	-.72* (.21)	-.72* (.21)	-.66* (.26)
AIC		71062.6	118233.2	114917.0	115951.5	62448.8	61816.6	66172.5
BIC		71541.8	118874.0	115561.9	116597.2	63094.0	62460.7	66736.9

Note. APH = students' attitudes towards performance heterogeneity; Mig = students' migration background (0 = both parents born in Germany; 1 = at least one parent born abroad); Gender (0 = male; 1 = female), CultCap = cultural capital; TempRNO = temporal reference norm orientation; PH = performance heterogeneity; PAVg = average performance.

Reference group for school type = Gymnasium.

~ $p < .10$; * $p < .05$

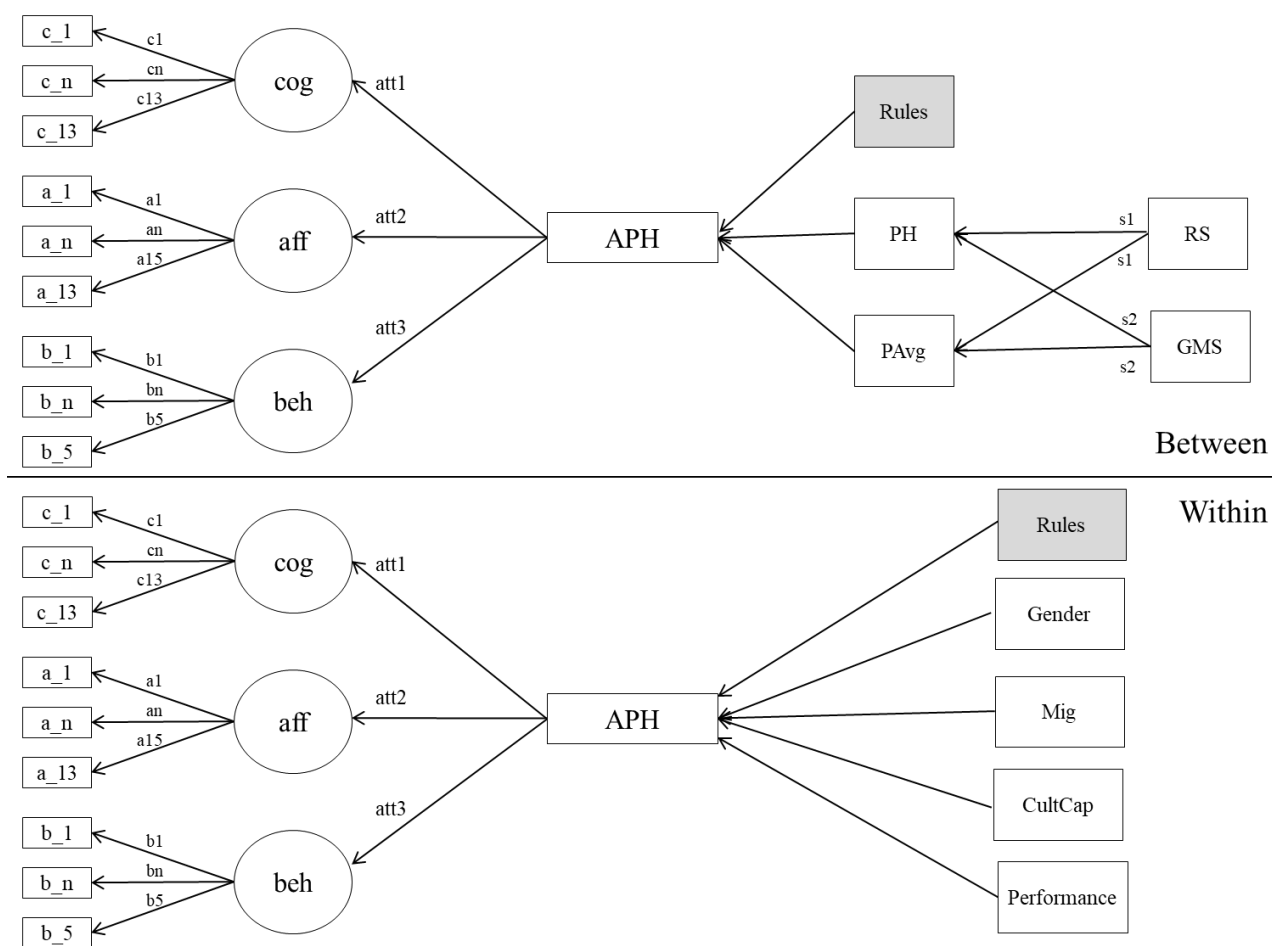


Figure 1: Multilevel structural equation model using Hypothesis 2a as an example

Note. cog = cognitive component; aff = affective component; beh = behavioral component; APH = students' attitudes towards performance heterogeneity; Mig = students' migration background (0 = both parents born in Germany; 1 = at least one parent born abroad); Gender (0 = male; 1 = female), CultCap = cultural capital; PH = performance heterogeneity; PAvg = average performance, RS = Realschule; GMS = Gemeinschaftsschule.

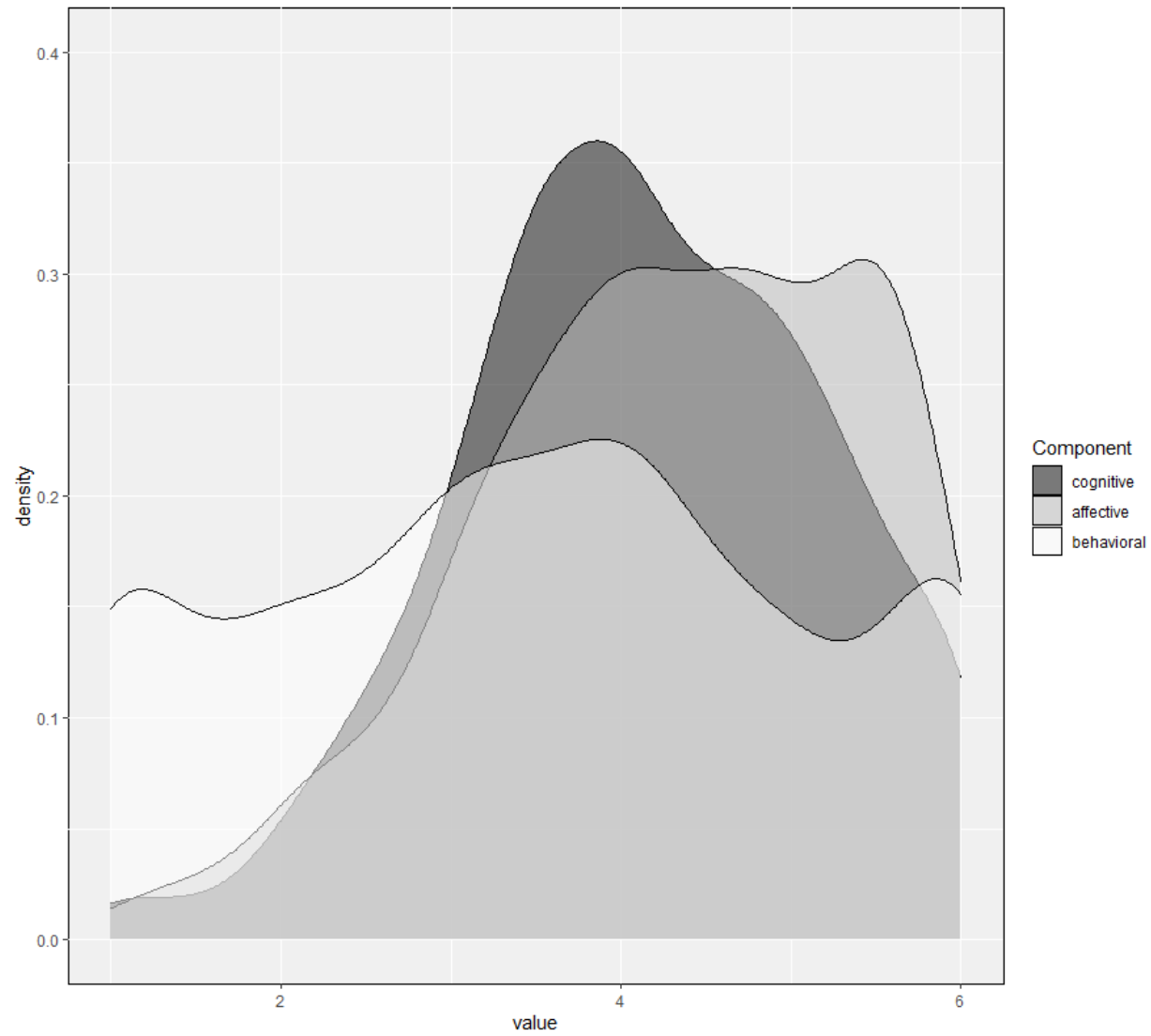


Figure 2: Density plot for the attitude components

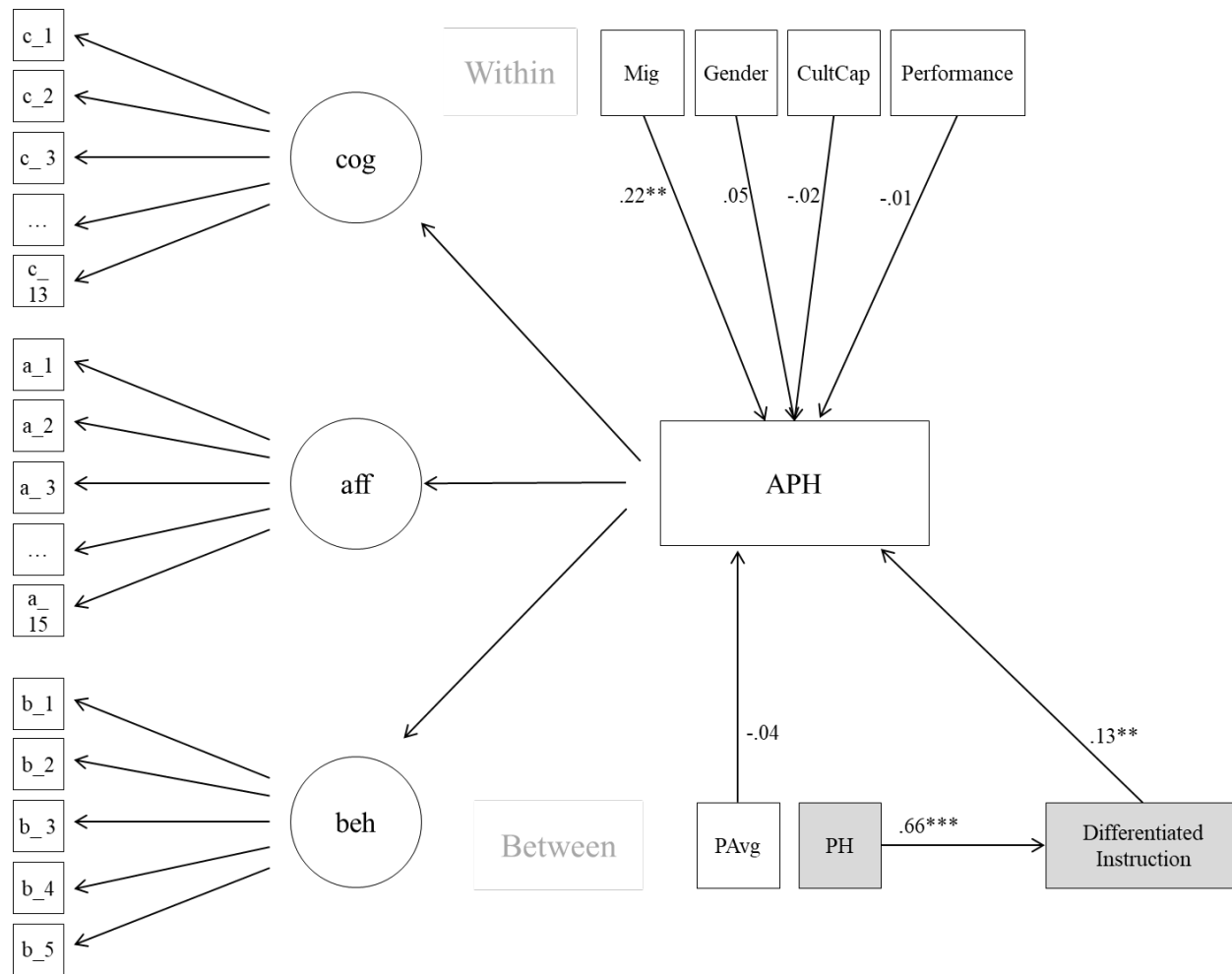


Figure 3: Paths and results for the post-hoc mediation analysis (simplified figure)

Note. cog = cognitive component; aff = affective component; beh = behavioral component; APH = students' attitudes towards performance heterogeneity; Mig = students' migration background (0 = both parents born in Germany; 1 = at least one parent born abroad); Gender (0 = male; 1 = female), CultCap = cultural capital; PH = performance heterogeneity; PAvg = average performance.

Literature

Authors (unpublished)

- Allport, G. (1954). *The nature of prejudice*. Garden City, NJ: Doubleday Anchor Books.
- Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first grade groups. *Elementary School Journal*, 79(1), 193–223. doi:10.1086/461151
- Armstrong, M. M., Morris, C., Abraham, C., & Tarrant, M. (2016). Interventions utilising contact with people with disabilities to improve children's attitudes towards disability: a systematic review and meta-analysis. *Disability and Health Journal*, 10(1), 11–22. doi:10.1016/j.dhjo.2016.10.003
- Avramidis, E., & Norwich, B. (2002). Teachers attitudes towards integration inclusion a review of the literature. *European Journal of Special Needs Education*, 17(2), 129–147. doi:10.1080/08856250210129056
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. et al. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: deskriptive Befunde* [TIMSS - Teaching mathematics and science in international comparison: descriptive findings]. Opladen: Leske + Budrich. doi: 10.1007/978-3-322-95096-3
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U. et al. (2008): *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente* [Professional knowledge of teachers, cognitively activating mathematics teaching and the development of mathematical competence (COACTIV): documentation of the survey instruments]. Berlin: Max-Planck-Institut für Bildungsforschung.
- Bem, D.J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). San Diego, CA: Academic Press. doi:10.1016/S0065-2601(08)60024-6
- Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: McGraw-Hill.
- Brophy, J. (1986). Classroom Management Techniques. *Education and Urban Society*, 18(2), 182–194. doi:10.1177/0013124586018002005
- Brophy, J. E., & Evertson, C. M. (1976). *Learning from teaching: a developmental perspective*. Boston, MA: Allyn & Bacon. doi:10.1002/1520-6807(197710)14:4<525::AID-PITS2310140430>3.0.CO;2-Z
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage. doi:10.1177/0049124192021002005
- Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage Publications.
- Byrnes, D. A., & Kiger, G. (1997). Teachers Attitudes about Language Diversity. *Teaching and Teacher Education*, 13(6), 637–644. doi:10.1016/S0742-051X(97)80006-6
- Chatoupis, C., & Vagenas, G. (2018). Effectiveness of the practical style and reciprocal style of teaching: A meta-analysis. *The Physical Educator* (75), 175–194. doi: 10.18666/TPE-2018-V75-I2-7920
- Cho, G., & DeCastro-Ambrosetti, D. (2005). Is Ignorance Bliss? Preservice-Teachers' Attitudes toward Multicultural Education. *The High School Journal*, 89(2), 24–28. doi:10.1353/hsj.2005.0020
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77(1), 113–143. doi:10.3102/003465430298563
- Dar, Y. (1985). Teachers attitudes towards ability grouping: Educational Considerations and Social Organizational Influences. *Interchange*, 16(2), 17–28. doi:10.1007/BF01807206
- Dickhäuser, O., Janke, S., Praetorius, A.-K., & Dresel, M. (2017). The Effects of Teachers' Reference Norm Orientations on Students' Implicit Theories and Academic Self-Concepts. *Zeitschrift für Pädagogische Psychologie*, 31(3–4), 205–219. doi:10.1024/1010-0652/a000208
- Dooly, M. (2005). How aware are they? Research into Teachers Attitudes about Linguistic Diversity. *Language Awareness*, 14(2/3), 97–111. doi:10.1080/09658410508668827
- Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson, & C. S. Weinstein (Eds.), *Handbook of Classroom Management* (pp. 97–125). Mahwah, NJ: Lawrence Erlbaum.
- Eagly, A.H., & Chaiken, S. (1993). *Psychology of Attitudes*. Fort Worth: Harcourt Brace Jovanovich College Publishers. doi:10.1002/mar.4220120509
- Else-Quest, N.M., Mindeo, C.C., & Higgins, A. (2013). Math and Science Attitudes and Achievement at the Intersection of Gender and Ethnicity. *Psychology of Women Quarterly*, 37(3), 293–309. doi:10.1177/0361684313480694
- Emmer, E. T., Evertson, C. M., & Worsham, M. E. (2006). *Classroom management for middle and high school teachers* (Vol. 7). Boston, MA: Allyn & Bacon.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138. doi:10.1037/1082-989X.12.2.121
- Evertson, C. M., & Emmer, E. T. (1982). Preventive classroom management. In D. Duke (Ed.), *Helping Teachers manage Classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Evertson, C. M., Emmer, E. T., & Worsham, M. E. (2006). *Classroom management for elementary teachers* (Vol. 7). Boston, MA: Allyn & Bacon.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, behavior*. Reading, MASS: Addison-Wesley.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34(1), 174-206. doi:10.3102/00028312034001174
- Gaith, G., & Bouzeineddine, A., R. (2003). Relationships between Reading Attitudes, Achievement, and Learners' Perception of their Jigsaw II Cooperative Learning Experience. *Reading Psychology*, 24, 105-121. doi:10.1080/02702710390197444
- Hartinger, A., Grittner, F., Lang, E., & Rehle, C. (2010). Ein Vergleich der Einstellung zur Heterogenität von Lehrkräften in jahrgangsgemischten und jahrgangshomogenen Klassen [A comparison of attitudes towards heterogeneity of teachers in mixed and homogeneous classes]. In K.-H. Arnold, K. Hauen-schild, B. Schmidt, & B. Ziegenmeyer (Eds.), *Zwischen Fachdidaktik und Stufendidaktik. Perspektiven für die Grundschulforschung* [Between subject didactics and grade didactics. Perspectives for primary school research] (pp. 77-80). Berlin: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-92475-5_1
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge. doi:10.1080/00071005.2011.584660
- Jensen, P., & Rasmussen, A. W. (2011). The effect of immigrant concentration in schools on native and im-migrant children's reading and math skills. *Economics of Education Review*, 30, 1503-1515. doi:10.1016/j.econedurev.2011.08.002
- Jerusalem, M., Drössler, S., Kleine, D., Klein-Heßling, J., Mittag, W., & Röder, B. (2009). *Skalen zur Er-fassung von Lehrer- und Schülermerkmalen. Schülerperzipierte Lehrerbezugsnormorientierung (SPLB)* [Scales for measuring teacher and student characteristics. Student-perceived teacher reference norm ori-entation (SPLB)]. Retrieved from https://www.erziehungswissenschaften.hu-ber-lin.de/de/paedpsych/forschung/Skalenbuch_FoSS.pdf
- Katz, D., & Stotland, E. (1959). A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3, pp. 423-457). New York: McGraw-Hill.
- Keith, N., & Frese, M. (2008). Effectiveness of error management training: a meta-analysis. *Journal of Ap-plied Psychology*, 93(1), 59-69. doi:10.1037/0021-9010.93.1.59
- King-Sears, M. E. (2008). *Differentiation and the Curriculum. Facts and Fallacies: Differentiation and the general education curriculum for students with special educational needs*. Oxford: Blackwell Publish-ing. doi:10.1111/j.1467-9604.2008.00371.x
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. NY: Guilford Press.
- Krosnick, J.A., & Petty, R.E. (1995). *Attitude Strength: Antecedents and Consequences*. Lawrence Erl-baum, Mahwah.
- Kumar, R., & Hamer, L. (2012). Preservice teachers' attitudes and beliefs toward student diversity and pro-posed instructional practices: a sequential design study. *Journal of Teacher Education*, 64(2), 162-177. doi:10.1177/0022487112466899
- Kunter, M., & Trautwein, U. (2013). *Psychologie des Unterrichts* [Psychology of teaching]. Paderborn: Schöningh.
- Kunter, M., & Voss, T. (2011): Das Modell der Unterrichtsqualität in COACTIV: eine multikriteriale Ana-lyse [The model of teaching quality in COACTIV: a multi-criteria analysis]. In M. Kunter, J. Baumert, W. Blum, U. Kusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften – Ergebnisse des Forschungsprogramms COACTIV* [Professional competence of teachers - results of the COACTIV research programme] (pp. 85-113). Münster: Waxmann.
- Larson, R., & Richards, M. (1991). Boredom in the middle school years: blaming schools versus blaming students. *American Journal of Education*, 99(4), 418-443. doi:10.1086/443992
- Lieberman, N., & Trope, Y. (2008). The psychology of transcending the here and now. *Science*, 322(5905), 1201-1205. doi:10.1126/science.1161958
- Lieberson, S. (1969). Measuring population diversity. *American Sociological Review*, 34, 850-862. <http://dx.doi.org/10.2307/2095977>
- Linchevski, L., & Kutscher, B. (1998). Tell Me with Whom You're Learning, and I'll Tell You How Much You've Learned: Mixed-Ability versus Same-Ability Grouping in Mathematics. *Journal for Research in Mathematics Education*, 29(5), 533-554. doi:10.2307/749732
- Lüders, M., & Rauin, U. (2004). Unterrichts- und Lehr-Lernforschung [Research in teaching and learning]. In: W. Helsper, & Böhme, J. (Eds.), *Handbuch der Schulforschung* [Handbook of school research] (pp. 691-720). Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-91095-6
- Maehler, D., & Brinkman, H. U. (2016). *Methoden der Migrationsforschung - Ein interdisziplinärer For-schungsleitfaden* [Methods of migration research - An interdisciplinary research guide]. Wiesbaden: Springer VS. doi:10.1007/978-3-658-10394-1

- Markic, S., & Abels, S. (2014). Heterogeneity and Diversity. A Growing Challenge or Enrichment for Science Education in German Schools?. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(4), 271-283. doi:10.12973/eurasia.2014.1082a
- Martinez, R.S., Aricak, O.T., & Jewell, J. (2008). Influence of Reading Attitudes on Reading Achievement: A Test of the temporal-interaction Model. *Psychology in the Schools*, 45(10), 1010-1022. doi:10.1002/pits.20348
- McCombs, B. L., & Whisler, J. S. (1997). *The learner-centered classroom and school: Strategies for increasing student motivation and achievement*. San Francisco: Jossey-Bass.
- McGinnis, J. C., Frederick, B. P., & Edwards, R. (1995). Enhancing classroom management through proactive rules and procedures. *Psychology in the Schools*, 32, 220-224. doi:10.1002/1520-6807(199507)32:3<220::AID-PITS2310320309>3.0.CO;2-4
- Müller A.G., & Stanat P. (2006). Schulischer Erfolg von Schülerinnen und Schülern mit Migrationshintergrund: Analysen zur Situation von Zuwanderern aus der ehemaligen Sowjetunion und aus der Türkei [Academic success of students with migration background: Analyses of the situation of immigrants from the former Soviet Union and Turkey]. In J. Baumert, P. Stanat, R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* [Origin-related disparities in education: Differential educational processes and problems of distributive equity] (pp. 221-255). VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-531-90082-7_6
- Nordstrom, C. R., Wendland, D., & Williams, K. B. (1998). "To err is human": an examination of the effectiveness of error management training. *Journal of Business and Psychology*, 12(3), 269-282. doi:10.1023/A:1025019212263
- Nowicki, E.A. & Sandieson, R. (2002): A meta-analysis of school-age children's attitudes towards persons with physical or intellectual disabilities. *International Journal of Disability, Development and Education*, 49,243-265. doi:10.1080/1034912022000007270
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. [Empirical Study]. *Cognition and Instruction*, 1, 117-175. doi:10.1207/s1532690xci0102_1
- Papanastasiou, E. C., & Zembylas, M. (2002). The effect of Attitudes on Science Achievement: a Study conducted among High school pupils in Cyprus. *International Review of Education*, 48(6), 469-484. doi:10.1023/A:1021334424571
- Papanastasiou, E. C., & Zembylas, M. (2004). Differential effects of science attitudes and science achievement in Australia, Cyprus, and the USA. *International Journal of Science Education*, 26(3), 259-280. doi:10.1080/0950069022000038277
- Pedota, P. (2007). Strategies for effective Classroom Management in the Secondary Setting. *The Clearing House*, 80(4), 163-166. doi:10.3200/TCHS.80.4.163-168
- 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. doi:10.3102/0013198X09332374
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards [Evaluation of classroom teaching standards]. *Zeitschrift für Erziehungswissenschaft* 1(13), 121-148. doi:10.1007/s11618-010-0113-z
- Pilten, G. (2016). The Evaluation of Effectiveness of Reciprocal Teaching Strategies on Comprehension of Expository Texts. *Journal of Education and Training Studies*, 4(10), 232-247. doi:10.11114/jets.v4i10.1791
- Piwowar, V. (2013). Multidimensionale Erfassung von Kompetenzen in Klassenmanagement. Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens [The Multidimensional Assessment of Classroom Management Competencies: Construction and Validation of an Observer Rating and a Student Rating for Secondary Teachers]. *Zeitschrift für Pädagogische Psychologie*, 27 (4), 215–228. doi:10.1024/1010-0652/a000108
- Perkins, R. E., & Hill, A. B. (1985). Cognitive and affective aspects of boredom. *British Journal of Psychology*, 76, 221-234. doi:10.1111/j.2044-8295.1985.tb01946.x
- Powell, J. (2006). Special education and the risk of becoming less educated. *European Societies* 8(4), 577-599. doi:10.1080/14616690601002673
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rach, S., Ufer, S., & Heinze, A. (2013). Learning from errors: effects of teachers' training on students' attitudes towards and their individual use of errors. *PNA*, 8(1), 21-30. doi:10.30827/pna.v8i1.6122
- Rakoczy, K., Klieme, E., Drollinger-Vetter, B., Lipowsky, F., Pauli, C., & Reusser, K. (2007). Structure as quality feature in mathematics instruction of the learning environment vs. a structured presentation of learning content. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report of the DFG Priority Programme* (pp. 101-120). Muenster: Waxmann.

- Rheinberg, F. (1980). *Leistungsbewertung und Lernmotivation* [Performance evaluation and motivation for learning]. Göttingen: Hogrefe.
- Rosenberg, M. J. (1960). An analysis of affective-cognitive consistency. In C. I. Hovland, & M. J. Rosenberg (Eds.), *Attitude organization and change*. New Haven: Yale University Press.
- Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I. Hovland, & M. J. Rosenberg (Eds.), *Attitude organization and change*. New Haven: Yale University Press.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. doi:10.18637/jss.v048.i02
- Segeritz, M., Walter, O., & Stanat, P. (2010). Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland: Evidenz für segmentierte Assimilation? [Patterns of academic success of adolescents migrants in Germany: Evidence for segmented assimilation?]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62(1), 113–138. doi:10.1007/s11577-010-0094-1
- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60(3), 471–499. doi:10.3102/00346543060003471
- Statistisches Bundesamt (2018). *Bevölkerung und Erwerbstätigkeit, Bevölkerung mit Migrationshintergrund, Ergebnisse des Mikrozensus 2017* [Population and employment, population with migration background, results of the microcensus 2017]. Wiesbaden: destatis.
- Teachman, J. D. (1980). Analysis of population diversity: Measures of qualitative variation. *Sociological Methods & Research*, 8, 341–362. <http://dx.doi.org/10.1177/004912418000800305>
- Theunissen, U., Schoppengerd, E., Witzke, F., Endell, A., & Wenning, F. (2005). *Duisburger Sprachstandtest. Gesamtschule Duisburg-Meiderich*. [Duisburg language level test. Duisburg-Meiderich Comprehensive School]. Unveröffentlichtes Testinstrument [Unpublished test instrument].
- Trautmann, M., & Wischer, B. (2011). *Heterogenität und Schule. Eine kritische Einführung* [Heterogeneity and school. A critical introduction]. Wiesbaden: Springer.
- Tulis, M. (2013). Error Management Behavior in Classrooms. Teachers responses to student mistakes. *Teaching and Teacher Education*, 33, 56–68. doi:10.1016/j.tate.2013.02.003
- Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick, & L. S. Fidell (Eds.) *Using Multivariate Statistics* (pp. 653–771). Needham Heights, MA: Allyn & Bacon. doi:10.1371/journal.pone.0071264
- Van de Grift, W. (2007). Quality of teaching in four European countries: a review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127–152. doi:10.1080/00131880701369651
- Vaughn, S., Klingner, J. K., & Bryant, D. P. (2001). Collaborative Strategic Reading as a means to enhance peer-mediated instruction for reading comprehension and content area learning. *Remedial and Special Education*, 22, 66–74. doi:10.1177/074193250102200201
- Watson, C. S., Kidd, G. R., Horner, D. G., Lowther, C. A., Eddins, D. A., Krueger, G. et al. (2003). Sensory, Cognitive, and Linguistic Factors in the Early Academic Performance of Elementary School Children: The Benton-IU Project. *Journal of Learning Disabilities*, 36(2), 165–197. doi:10.1177/002221940303600209
- Youngs, C. S., & Youngs, G. A. (2001). Predictors of mainstream teachers attitudes toward ESL students. *TESOL Quarterly* 35(1), 97–120. doi:10.2307/3587861

Anhang D: Beitrag im Sammelband

Dotzel, S. & Karst, K. (im Druck). Einstellung zur Leistungsheterogenität bei Schülerinnen und Schülern. Wie können sie gemessen werden, wie sind sie ausgeprägt und wie können sie beeinflusst werden?. In K. Karst, D. Thoma, J. Derkau, J. Seifried & S. Münzer (Hrsg.), *Lehrer*innenbildung im Kontext leistungsbezogener Heterogenität und Mehrsprachigkeit. Die Leuchtturm-Projekte der Universität Mannheim*. Münster: Waxmann.

Einstellung zur Leistungsheterogenität bei Schülerinnen und Schülern

Wie können sie gemessen werden, wie sind sie ausgeprägt und wie können sie beeinflusst werden?

1 Hintergrund

Der Frage nach dem schulischen Umgang mit Heterogenität wird nicht erst in den letzten Jahren Aufmerksamkeit geschenkt. Neben den Forderungen nach Inklusion und individueller Förderung (z.B., Markic & Abels, 2014) begegnen Schulen auch einer steigenden Anzahl von Schüler*innen mit Migrationshintergrund (z.B., Maehler & Brinkman, 2016). Lehrkräfte haben zur Aufgabe mit der damit einhergehenden Heterogenität so umzugehen, dass Schüler*innen einer förderlichen Lernumgebung begegnen. Für den professionellen Umgang mit Heterogenität spielt neben der Unterrichtsgestaltung (z.B. binnendifferenzierende Maßnahmen) und den diagnostischen Kompetenzen der Lehrkraft (z.B. von Schüler*innen mit besonderem Förderbedarf) auch die Einstellung zur Heterogenität eine Rolle (Trautmann & Wischer, 2011). Aus diesem Grund wurde die Einstellung zur Heterogenität bei Lehrkräften bereits mehrfach untersucht (z.B. Hartinger et al., 2010; Ruhberg & Porsch, 2017). Die Einstellungen von Schüler*innen blieben zunächst unberücksichtigt.

Sich den Einstellungen von Schüler*innen zu widmen, kann aber aus drei Gründen zielführend sein. Erstens zeigen Studien, dass positive Einstellungen der Schüler*innen zu Schule und Unterricht das Lernen wiederum positiv beeinflussen (z.B. Else-Quest, Mineo &

Higgins, 2013; Martinez, Aricak & Jewell, 2008). Zweitens kann die Rückmeldung der schülerspezifischen Einstellung zur Heterogenität an die Lehrkraft von ebendieser bei der Unterrichtsgestaltung berücksichtigt werden (z.B. durch einen erhöhten Einsatz von Differenzierung). Und drittens werden im deutschen Bildungssystem politische Entscheidungen vollzogen, welche die Heterogenität in den Schulen stark beeinflussen (z.B. die Einführung eines offenen Übergangssystems versus eines notengebundenen Übergangssystems). Diese Entscheidungen und deren Konsequenzen können die Einstellung zur Heterogenität beeinflussen; und dies tangiert nicht nur Lehrkräfte, sondern auch die Schüler*innen.

Die Einstellungen von Schüler*innen spielen also eine wichtige Rolle. Deswegen ist es Ziel im Rahmen des Projekts HEÄT (Herausforderung Heterogenität), die *Einstellung zur Leistungsheterogenität* (ELH) von Schüler*innen zu untersuchen. Mit einer Stichprobe von 784 Schüler*innen der fünften Klasse soll in den Blick genommen werden, wie die ELH von Schüler*innen gemessen werden kann, wie diese ausgeprägt ist und wie sie für die Unterrichtsgestaltung der Lehrkraft von Bedeutung sein kann.

2 Einstellung zur Leistungsheterogenität

In der multidimensionalen Einstellungstheorie wird davon ausgegangen, dass eine Einstellung aus drei Komponenten besteht: der kognitiven, der affektiven und der behavioralen Komponente (z.B. Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960) (siehe Abbildung 1). Die kognitive Komponente repräsentiert dabei Überzeugungen einer Person gegenüber einem Einstellungsobjekt, die affektive Komponente repräsentiert die Emotionen gegenüber dem Einstellungsobjekt und die behaviorale Komponente repräsentiert die verhaltensbezogenen Intentionen und Reaktionen gegenüber dem Einstellungsobjekt.

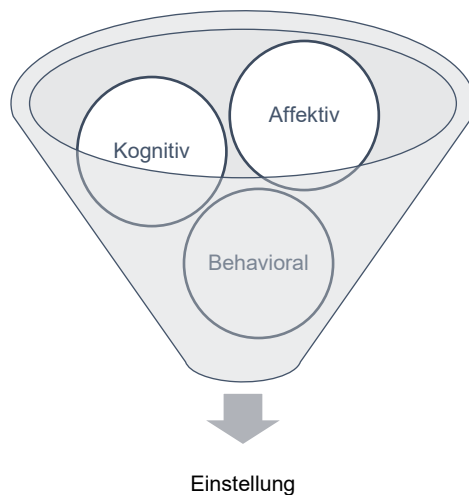


Abbildung 1. Die drei Komponenten einer Einstellung.

Als Einstellungsobjekt bezeichnet man das Objekt, gegenüber dem die Einstellung ausgerichtet ist; und grundsätzlich gibt es bei der Art des Einstellungsobjekts keine Einschränkungen. So kann das Einstellungsobjekt eine bestimmte Person (z.B. die Schulleiter*in, eine gewisse Schüler*in) sowie eine Personengruppe (z.B. Lehrkräfte, Migrant*innen) sein; ein Einstellungsobjekt kann aber auch Verhalten (z.B. Rauchen), konkrete Objekte (z.B. ein Stuhl) und abstrakte Objekte (z.B. Heterogenität) umfassen (Bem, 1972). Als Einstellungsobjekt kann also auch die Leistungsheterogenität in der Klasse herangezogen werden. Die Einstellung zur Leistungsheterogenität umfasst somit, welche Überzeugungen (kognitive Komponente), Gefühle (affektive Komponente) und Verhaltensintention (behaviorale Komponente) eine Person gegenüber Leistungsheterogenität hat.

3 Kontextuelle Einflussfaktoren auf die Einstellung zur Leistungsheterogenität

In Blooms Theorie schulischen Lernens (1976) wird angenommen, dass 25% der schulischen Leistung auf positive Einstellungen zum Schulfach, positive Einstellungen zu eigenen Fähigkeiten und positive Einstellungen zum schulischen Umfeld zurückgeführt werden können (vgl. Papanastasiou & Zembylas, 2002). Während die Einstellungen zum Schulfach (z.B. Martinez, Aricak & Jewell, 2008) und die Einstellungen zu den eigenen Fähigkeiten (z.B. Else-Quest, Mineo & Higgins, 2013) bereits untersucht wurde, spielten die Einstellungen zum schulischen

Umfeld – wie bspw. die Einstellung zur Leistungsheterogenität – eine geringere Rolle. Damit blieb ebenso unberücksichtigt, welche Faktoren die Einstellung zur Leistungsheterogenität beeinflussen können. Zwei potenzielle Faktoren, nämlich die Leistungsheterogenität in der Klasse selbst sowie das unterrichtliche Handeln der Lehrkraft, werden im Folgenden in den Blick genommen.

3.1 Leistungsheterogenität

Heterogenität lässt sich aus dem Griechischen ableiten (*heteros* = der Andere, und *gignomai* = entstehen) und beschreibt die Verschiedenheit der Teile in einem zusammengesetzten Ganzen (von der Groeben, 2003). Leistungsheterogenität kann definiert werden als die divergenten leistungsbezogenen Ausgangsbedingungen einer Schülerschaft (Scharenberg, 2010). Dementsprechend haben Schüler*innen in eher homogenen Klassen ähnliche Lernvoraussetzungen und Lernstände und Schüler*innen in eher heterogenen Klassen unterschiedliche Lernvoraussetzungen und Lernstände.

Leistungsheterogenität ist ein kontextueller Umstand, mit dem Schüler*innen einer Klassengemeinschaft konfrontiert sind. Schüler*innen in Klassen mit unterschiedlicher Leistungszusammensetzung haben unterschiedliche Erfahrungen mit Leistungsheterogenität; und es wird erwartet, dass diese Erfahrungen die *Einstellung zur Leistungsheterogenität* (ELH) von Schüler*innen beeinflussen. Auch wenn die ELH von Schüler*innen bisher nicht in den Blick genommen wurde, können Studien zu verwandten Fragestellungen einen Einblick in den Wirkmechanismus geben. So zeigen Studien zur ELH von Lehrkräften, dass diese die zunehmenden Begabungsunterschiede der Schüler*innen als eine Beruferschwernis empfinden (z.B. Baumert et al., 1997; Trautmann und Wischer, 2011), jedoch führt das Unterrichten in heterogenen Klassen zu einer positiveren ELH. Haben Lehrkräfte bereits Erfahrungen mit Leistungsheterogenität, scheint sich dies in einer positiveren ELH zu manifestieren (z.B. Hartinger et al., 2010; Linchevski & Kutscher, 1998; vgl. Byrnes & Kiger, 1997). Studien, die die Einstellung zur Inklusion von Schüler*innen untersuchen, folgen einem ähnlichen Muster. Schüler*innen, die Kontakt zu Mitschüler*innen mit Behinderung haben und somit bereits

Erfahrungen sammeln konnten, zeigen positivere Einstellungen als Schüler*innen ohne einen solchen Kontakt (z.B. Armstrong et al., 2016).

Einige Wissenschaftler*innen nehmen an, dass ihre Ergebnisse durch die Kontakthypothese nach Allport (1954) erklärt werden können (z.B. Byrnes und Kiger 1997; Armstrong et al. 2016). Die Kontakthypothese besagt, dass der häufige Kontakt mit Gruppen oder Einzelpersonen (gegenüber denen Vorurteile bestehen) Vorurteile reduziert und Personen unterstützt, eine positive Einstellung zu entwickeln. Der Kontakt soll besonders wirksam sein, wenn Menschen in der Kontaktsituation kooperative Ziele verfolgen, gleichgestellt sind, miteinander interagieren und wenn der Kontakt von Autoritäten unterstützt wird. Auch Schüler*innen in einer „heterogenen Kontaktsituation“ – in einer heterogenen Klassengemeinschaft – verfolgen gemeinsame Ziele (z.B. Versetzung in die nächsthöhere Jahrgangsstufe, gemeinsame Leistungserbringung in Gruppenreferaten), interagieren miteinander (z.B. in Diskussionen) und werden in ihrem Kontakt von einer Autorität (z.B. der Lehrkraft) unterstützt. Gemäß der Kontakthypothese lässt sich ableiten, dass Schüler*innen eine positivere ELH haben sollten, wenn sie Teil einer heterogenen Klasse sind; wenn sie im Kontakt mit Leistungsheterogenität stehen.

3.2 Störungen im Unterricht und das Handeln der Lehrkraft

Lehrkräfte haben unterschiedliche Strategien, um mit heterogenen Klassen umzugehen (z.B. Differenzierung: Bohl, 2017; Pietsch, 2010; Individualisierter Unterricht: Burmeister et al., 2016; Anforderungsstufung: Prediger & von Aufschnaiter, 2017). Ein professioneller Umgang mit Heterogenität durch die Lehrkraft wirkt sich wiederum auf die Lernumwelt der Schüler*innen aus; dadurch kann eine Lehrkraft durch ihr Handeln auch einen Einfluss auf die ELH von Schüler*innen nehmen.

In der Debatte um einen professionellen Umgang mit Heterogenität wird unter anderem das Thema Störung und störungsfreier Unterricht in den Blick genommen (Rieger-Ladich, 2017). Lehrkräfte haben die Aufgabe, den Unterricht herausfordernd zu gestalten – und das für alle Schüler*innen. Während Lehrkräfte in (eher) homogenen Klassen einen Großteil der

Klasse mittels Frontalunterricht oder direkter Instruktion kognitiv angemessen aktivieren können, stehen Lehrkräfte aus (eher) heterogenen Klassen bei der Nutzung solcher Methoden vor größeren Schwierigkeiten. Je größer die Unterschiede in den Lernvoraussetzungen sind, desto stärker müssen Lehrkräfte darauf achten, dass der Unterricht weder zu einer Über- noch Unterforderung bei Schüler*innen führt (z.B. Slavin, 1987; vgl. Scharenberg, 2010). Über- oder Unterforderung im Unterricht kann nämlich mit Langeweile oder Monotonie einhergehen (z.B. Czerwenka et al., 1990; Helmke, 1993), woraus Störungen resultieren können (bspw. weil Schüler*innen schneller sind, als der Rest der Klasse; weil Schüler*innen um eigene Schwächen wissen und „abtauchen“). Deswegen ist es insbesondere in heterogenen Klassen wichtig, dass Lehrkräfte auf einen professionellen Umgang mit Störungen achten und Störungen gar nicht erst aufkommen lassen.

Um das Ausmaß an Störungen im Unterricht zu reduzieren, ist ein effektiver Umgang mit Störungen durch die Lehrkraft notwendig. Der Umgang mit Störungen gilt als eine wichtige Facette im Bereich Classroom Management, der „Fähigkeit, den Klassenraum als eine effektive Umgebung für das Lehren und Lernen zu schaffen, zu erhalten und (wenn nötig) wiederherzustellen" (Brophy 1986, vgl. Kunter & Trautwein, 2013). Um Störungen effektiv zu bewältigen, überwachen Lehrkräfte die Einhaltung von Regeln, sanktionieren Störungen angemessen und greifen zügig in störende Situationen ein (z.B. Anderson et al. 1979; Evertson, Emmer, & Worsham. 2006). Durch ein solches Vorgehen können Lehrkräfte der Heterogenität positiv begegnen und mit stringenter Klassenführung die Heterogenität berücksichtigen (vgl. Bohl, 2017; Pietsch, 2010). Dies kann dazu führen, dass die Heterogenität in der Klasse nicht nur mit Negativereignissen (bspw. ein hohes Ausmaß an Störungen) verbunden wird, sondern auch positiv wahrgenommen werden kann. Somit sollte ein effektiver Umgang mit Störungen durch die Lehrkraft eine positive ELH von Schüler*innen fördern.

4 Forschungsfragen

Auf Basis des Forschungsstandes kann davon ausgegangen werden, dass Einstellungen zur Schule und Unterricht das Lernen und die Leistung von Schüler*innen beeinflussen. Insofern

ist es von Bedeutung, die *Einstellung zur Leistungsheterogenität* (ELH) bei Schüler*innen in den Blick zu nehmen und der Frage nachzugehen, welche kontextuellen Faktoren die ELH beeinflusst.

Im Rahmen dieses Kapitels werden folgende Forschungsfragen untersucht:

1. Wie kann die ELH von Schüler*innen gemessen werden?
2. Wie ist die ELH von Schüler*innen ausgeprägt?
3. Welche kontextuellen Faktoren können die ELH von Schüler*innen erklären?
 - 3a. Welchen Einfluss hat die Leistungsheterogenität auf die ELH von Schüler*innen?
 - 3b. Welchen Einfluss haben Störungen im Unterricht auf die ELH von Schüler*innen?
 - 3c. Welchen Einfluss hat der Umgang der Lehrkraft mit Störungen auf die ELH von Schüler*innen?

5 Methode

5.1 Stichprobe und Design

Da bislang noch nichts über die ELH von Schüler*innen bekannt ist, wurden im April 2018 zunächst neun Leitfadeninterviews (6 Mädchen, 3 Jungs) mit Schüler*innen der 5. und 6. Jahrgangsstufe durchgeführt. Auf Basis einer inhaltsanalytischen Auswertung wurde ein Fragebogen entwickelt, welcher im Juli 2018 pilotiert¹ ($N_{\text{Schüler*innen}} = 290$, $N_{\text{Klasse}} = 13$; 5. und 6. Jahrgangsstufe) und im Herbst 2018 in der Hauptuntersuchung eingesetzt wurde.

In der Hauptuntersuchung nahmen dann $N = 784$ Schüler*innen (50% weiblich; $M_{\text{Alter}} = 10.20$; $SD = .51$) aus 33 Klassen der 5. Jahrgangsstufe teil. In dieser Studie wurden sowohl die sprachlichen Kompetenzen als auch die ELH der Schüler*innen erhoben. Außerdem wurden demographische Hintergrundvariablen (Migrationshintergrund, Geschlecht, Kulturelles Kapital) sowie kontextuelle Unterrichtsvariablen (Ausmaß an Störungen im Unterricht, Um-

¹ Die Ergebnisse der Pilotierung werden nicht berichtet.

gang der Lehrkraft mit Störungen) erhoben. Die Befragung wurde an 12 Stadt- und Landschulen des Einzugsgebiets Mannheim durchgeführt (8 Gymnasien, 2 Realschulen, 2 Gemeinschaftsschulen). Die durchschnittliche Klassengröße beträgt 25 Schüler*innen. 65% der Schüler*innen haben einen Migrationshintergrund.

5.2 Instrumente

In der vorliegenden Studie kamen drei Instrumente zum Einsatz. Erstens wurde ein Interview-Leitfaden zur ELH von Schüler*innen entwickelt. Zweitens wurde ein Fragebogen mit geschlossenem Antwortformat zur ELH von Schüler*innen konzipiert. Drittens wurde eine adaptierte Version des Duisburger Sprachstandtests (Karst et al., in diesem Band) eingesetzt, um leistungsbezogene Variablen zu erfassen.

5.2.1 Leitfadeninterviews

Mit den Interviews wurden drei Ziele verfolgt: Erstens, die Wahrnehmung der Leistungsheterogenität in der Klasse zu erfassen; zweitens, die affektive Einstellungskomponente der Schüler*innen zu erfassen; und drittens, die behaviorale Einstellungskomponente der Schüler*innen zu erfassen. Die kognitive Einstellungskomponente wurde in den Leitfadeninterviews nicht erfasst, da diese nur über spezifische Fragen, die dichotome Antworten im Sinne von Ja und Nein zulassen, gemessen werden kann. Solche Fragen sind für den Anlass von Leitfadeninterviews ungeeignet und wurden deswegen nicht miteinbezogen. Da Schüler*innen der frühen Sekundarstufe der Fachterminus Leistungsheterogenität wohl fremd ist, wurde dieser Begriff inhaltlich umrissen und im gesamten Interview mit „Leistungsunterschiede (in der Klasse)“ ersetzt. Nachdem die wichtigsten Leitfragen des Interviews thematisiert wurden, wurde mit einer Zusammenfassung (Heterogenität im Kosten-Nutzen-Abgleich) abgerundet. Ein Beispielvorgehen für ein Leitfadeninterview ist in Tabelle 1 zu erkennen. Die Leitfadeninterviews dauerten im Mittel 14 Minuten und wurden von einer Wissenschaftler*in durchgeführt, um Versuchsleiter-Effekte (Diekmann, 1995) möglichst konstant zu halten.

Tabelle 1. Fragen des Leitfadeninterviews.

Themenblock	Leitfrage
Kurzabriss Alltag Deutschstunde	„Wie sieht denn so eine typische Deutschstunde bei dir aus? Erzähl mal!“
Wahrnehmung der Leistungsheterogenität	„Jetzt gehen wir mal ein bisschen weg vom Unterricht hin zu deiner Klassengemeinschaft. Es gibt Klassen, in denen sind die Schüler eigentlich alle gleich gut in Deutsch. Dann gibt es aber auch Klassen, da sind die Schüler in der Deutschleistung sehr unterschiedlich; da gibt es sehr starke, aber auch sehr schwache Schüler. Wie ist das bei dir so? Erzähl mal!“ „Woran merkst du, dass es so ist? Nenn mir mal ein paar Beispiele!“
Affektive Einstellungskomponente	„Wenn du das jetzt so erzählst, dann frage ich mich, wie du dich in deiner Klasse in der Regel fühlst?“ „Gibt es Situationen, in denen du dich besonders gut oder schlecht fühlst, genau weil deine Klasse in der Deutschleistung so unterschiedlich/ gleich ist?“
Behaviorale Einstellungskomponente	„Mal angenommen, du könntest dir aussuchen, wie deine Klasse von der Leistungszusammensetzung in Deutsch wäre. Was würdest du dir wünschen? Wie würde deine Klasse aussehen?“
Heterogenität im Kosten-Nutzen Abgleich	„Wenn du jetzt noch einmal überlegst, wie sehr sich deine Klasse in der Deutschleistung unterscheidet, kannst du noch einmal zusammenfassen, was für dich die Vorteile sind und was die Nachteile?“
Hintergrundmerkmale	„Was war denn deine letzte Note in Deutsch?“ „Findest du, du gehörst in deiner Klasse in Deutsch eher zu den Guten oder zu den Schlechten in deiner Klasse?“ „In welche Klasse gehst du?“ „Auf welcher Schule bist du? Haupt-, Realschule oder Gymnasium?“

5.2.2 Schüler*innenfragebogen

Der Schüler*innenfragebogen bestand aus drei Teilen, die im Folgenden vorgestellt werden.

5.2.2.1 Einstellung zur Leistungsheterogenität

Die ELH wurde über drei separate Skalen zur kognitiven, affektiven und behavioralen Komponente multidimensional erfasst. Zur Messung der Einstellungskomponenten wurde eine Zustimmung-Ablehnungsdimension bezüglich des Einstellungsobjektes *Leistungsheterogenität* verwendet. Die Items liegen als sechsstufige Likert-Skala vor und weisen eine inhaltlich bipolare Struktur auf (siehe Abbildung 2). Die beiden jeweiligen Pole sind mit Antwortkategorien versehen und unterscheiden sich in deren Valenz. Je höher die Zustimmung, desto positiver ist die Einstellung der Schüler*innen zur Leistungsheterogenität. Die kognitive Einstellungsdimension wurde mit insgesamt 13 Items erfasst und misst die Überzeugungen der Schüler*innen gegenüber Leistungsheterogenität (z.B.: Ich finde eine Klasse, in der die Schüler*innen in

ihrer Leistung in Deutsch unterschiedlich sind... [schlecht] – [gut]). Die affektive Komponente wurde mit 15 Items erfasst und misst die Emotionen der Schüler*innen gegenüber Leistungsheterogenität (z.B.: In einer Klasse, in der die Schüler unterschiedlich sind, wäre ich... [gestresst] – [entspannt]). Die behaviorale Komponente wurde mit 5 Items erfasst und misst das intentionale Verhalten der Schüler*innen (z.B.: Wenn ich die Wahl hätte, wäre ich am liebsten in einer Klasse, in der die Schüler in Deutsch... [gleich gut sind wie ich] – [unterschiedlich gut sind wie ich]).

Nun stell dir mal vor, du hättest die Wahl zwischen einer Klasse, bei der alle Schüler gleich gut oder unterschiedlich gut in Deutsch sind. Wie würdest du dich entscheiden?

Wenn ich die Wahl hätte, wäre ich am liebsten in einer Klasse, in der die Schüler ...

gleich gut sind wie ich	☐	☐	☐	☐	☐	☐	unterschiedlich gut sind wie ich
genauso schnell lernen wie ich	☐	☐	☐	☐	☐	☐	unterschiedlich schnell lernen wie ich
gleich viel wissen wie ich	☐	☐	☐	☐	☐	☐	unterschiedlich viel wissen wie ich
genauso schlau sind wie ich	☐	☐	☐	☐	☐	☐	unterschiedlich schlau sind wie ich
genauso viel können wie ich	☐	☐	☐	☐	☐	☐	unterschiedlich viel können wie ich

Abbildung 2. Skalenformat am Beispiel der behavioralen Komponente.

5.2.2.2 Individuelle Hintergrundmerkmale

Migrationshintergrund. Schüler*innen wurden zu dem Herkunftsland ihrer Eltern befragt. Ein Migrationshintergrund liegt vor, sobald eines der Elternteile im Ausland geboren wurde (0 = kein Migrationshintergrund; 1 = Migrationshintergrund) (Statistisches Bundesamt, 2018).

Kulturelles Kapital. Das kulturelle Kapital wurde anhand der Bücherfrage erhoben (Paulus, 2009). Schüler*innen wurden danach befragt, wie viele Bücher sie zu Hause haben (1 = keine oder sehr wenige, 2 = genug, um ein Regalbrett zu füllen, 3 = genug, um ein Regal zu füllen,

4 = genug, um drei Regale zu füllen, 5 = über 200 Bücher). Die jeweiligen Antworten wurden mit dazu passenden Bildern visuell unterstützt.

5.2.2.3 Ausmaß an Störungen und Umgang mit Störungen

Ausmaß an Störungen. Um das Ausmaß an Störungen zu messen, wurde die Skala „Störungen im Unterricht“ aus dem Schüler*innen-Fragebogen KODEK-S genutzt ($n_{\text{Items}} = 4$; $\alpha = .61$; $M = 2.21$; $SD = 0.59$; Bsp.: Bei diesem Lehrer wird der Unterricht oft sehr gestört. 1 = Stimmt gar nicht, 4 = Stimmt genau) (Piwowar, 2013).

Umgang der Lehrkraft mit Störungen. Für den Umgang der Lehrkraft mit Störungen wurde ebenfalls die Einschätzung der Schüler*innen herangezogen. Genutzt wurde die Skala „Umgang des Lehrers mit Störungen“ aus dem KODEK-S ($n_{\text{Items}} = 4$; $\alpha = .71$; $M = 2.98$; $SD = 0.63$; Bsp.: Dieser Lehrer beendet Störungen von Schülern schnell und ohne Diskussion. 1 = Stimmt gar nicht, 4 = Stimmt genau) (Piwowar, 2013).

Die beiden Variablen *Ausmaß an Störungen* und *Umgang der Lehrkraft mit Störungen* wurden auf Klassenebene (Level 2) aggregiert. So kann für jede Klasse ein Durchschnittswert ermittelt werden, der Auskunft über die geteilte Wahrnehmung von Störungen und den Umgang der Lehrkraft damit gibt. Der ICC2 verweist auf eine zufriedenstellende Reliabilität für sowohl das *Ausmaß an Störungen* ($ICC2 = .79$) als auch den *Umgang der Lehrkraft mit Störungen* ($ICC2 = .60$).

5.2.3 Schülerleistung

Individuelle Leistung. Zur Erhebung wurde eine adaptierte Form des Duisburger Sprachstandstests (Jahr 2018/ 2019) genutzt (Karst et al., in diesem Band; Theunissen et al., 2005). Rohdaten bilden die Basis für die Skalierung gemäß dem 2-PL IRT Modell. Die Skalierung wurde mit R (R Core Team, 2019) durchgeführt. Es mussten keine Items wegen schlechter Item-Fits ($WMNSQ > 1.2$ oder < 0.8) ausgeschlossen werden. Die Reliabilitäten (Cronbachs Alpha) der Skalen sind als gut zu bezeichnen ($0.73 < \alpha < 0.86$).

Mittlere Klassenleistung. Zur Berechnung der mittleren Klassenleistung wurde der klassenspezifische Mittelwert (M) der Gesamtpunktzahl des Duisburger Sprachstandstests genutzt.

Leistungsheterogenität. Zur Berechnung der Leistungsheterogenität wurde die klassenspezifische Standardabweichung (*SD*) der Gesamtpunktzahl des Duisburger Sprachstandstests genutzt. Eine hohe *SD* verweist auf eine hohe Leistungsheterogenität in der Klasse.

5.3 Auswertungsstrategie

Für die Analyse der neun Leitfadeninterviews wurde die qualitative Inhaltsanalyse nach Mayring (2015) verwendet. Die transkribierten Interviews wurden mittels Frequenzanalyse ausgewertet, um darauf aufbauend ein Kategoriensystem abzuleiten. Die Ergebnisse dieser Analyse (siehe Kapitel 6.1) sind in die Konstruktion des Fragebogens für die Hauptstudie eingeflossen.

Die quantitativ erhobenen Fragebogendaten der Hauptstudie wurden dann für die statistische Analyse der ELH von Schüler*innen verwendet. Mittels konfirmatorischer Faktorenanalyse wurde getestet, ob die ELH bei Schüler*innen über die mehrdimensionale Einstellungstheorie gemessen werden kann (Forschungsfrage 1). Hierbei wurde nach Klassenzugehörigkeit geclustert, um die geschachtelte Datenstruktur zu berücksichtigen. Um zu ermitteln, wie die ELH bei Schüler*innen ausgeprägt ist – also ob diese eher positiv oder negativ ausfällt – wurden deskriptive Analysen herangezogen (Forschungsfrage 2). Darüber hinaus wird mit Korrelationen gezeigt, wie die einzelnen Einstellungskomponenten zusammenhängen. Um die ELH auf Basis von kontextbedingten Faktoren vorhersagen zu können (Forschungsfrage 3), wurden mehrere Strukturgleichungsmodelle berechnet. Diese wurden mehrebenenanalytisch realisiert, um auch hier die geschachtelte Datenstruktur zu berücksichtigen. Die Ladungen der Einstellungsvariablen und Einstellungskomponenten wurden für Level 1 und Level 2 fixiert (siehe Abbildung 3).

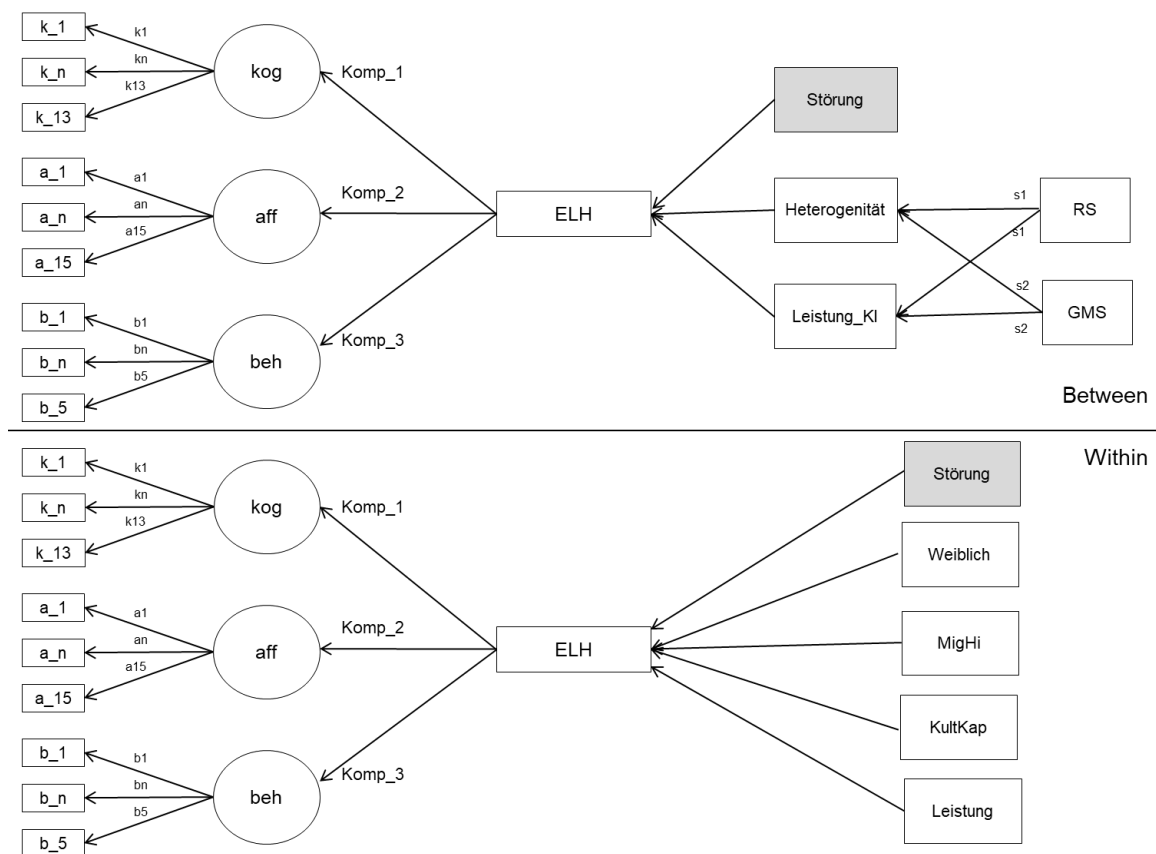


Abbildung 3. Modellierung des Multilevel-Strukturgleichungsmodells am Beispiel von Modell 3b.

Anmerkung. kog = kognitive Komponente; aff = affektive Komponente; beh = behaviorale Komponente; ELH = Einstellung zur Leistungsheterogenität bei Schüler*innen; MigHi = Migrationshintergrund (0 = kein Migrationshintergrund; 1 = Migrationshintergrund); KultKap = kulturelles Kapital; Leistung_Kl = mittlere Klassenleistung, RS = Realschule; GMS = Gemeinschaftsschule.

Modell 0 gilt als Baseline-Modell und untersucht die Effekte von den individuellen Hintergrundvariablen (i.e. Geschlecht (0 = männlich; 1 = weiblich), Migrationshintergrund (0 = beide Elternteile in Deutschland geboren; 1 = Mindestens eines der beiden Elternteile im Ausland geboren), Kulturelles Kapital (Wie viele Bücher habt ihr zu Hause?)) auf die ELH bei Schüler*innen. Modell 1 untersucht, ob die tatsächliche Leistungsheterogenität in der Klasse die ELH beeinflusst (Forschungsfrage 3a). Als zusätzliche Kontrollvariablen werden die individuelle und die mittlere Klassenleistung herangezogen. Überdies wird für die Schulform kontrolliert, indem diese als Prädiktoren für die Leistungsheterogenität und die mittlere Klassenleistung miteinbezogen werden. In Modell 2 wird der Einfluss von Störungen auf die ELH von Schüler*innen untersucht (Forschungsfrage 3b). In Modell 3 wird das Handeln der

Lehrkraft integriert, indem getestet wird, ob der Umgang mit Störungen durch die Lehrkraft einen Einfluss auf die ELH nimmt (Forschungsfrage 3c).

Aufgrund der Abhängigkeit zur jeweiligen Klasse wurden die individuelle Testleistung group-mean zentriert (Enders & Tofighi, 2007). Alle weiteren unabhängigen kontinuierlichen Variablen wurden z-standardisiert.

6 Ergebnisse

6.1 Leitfadeninterviews

Auf Basis einer Frequenzanalyse der transkribierten Leitfadeninterviews wurde ein Kategoriensystem erstellt (siehe Tabelle 2). Um die Wahrnehmung der Leistungsheterogenität bei Schüler*innen zu erfassen, wurden die Interviewten danach befragt, ob und woran sie Leistungsheterogenität in der eigenen Klasse erkennen. Die inhaltsanalytische Auswertung zeigt, dass bereits Schüler*innen der frühen Sekundarstufe das Phänomen Leistungsheterogenität wahrnehmen. Wahrnehmungsaspekte für Heterogenität umfassen die Kategorien Meldeverhalten, Verständnis und Wissen, Noten und das Verhalten bei Leistungsrückmeldung. Bei den Interviews zeigte sich, dass die Schüler*innen eine Tendenz dazu haben, die Klasse als eher heterogen wahrzunehmen. Von den insgesamt 9 Schüler*innen bezeichneten 7 Schüler*innen ihre Klasse als heterogen und nur 2 Schüler*innen als homogen. Die Wahrnehmung der Leistungsheterogenität ist jedoch sehr subjektiv. So unterschieden sich zwei Schüler*innen derselben Klasse in deren Wahrnehmung: ein*e Schüler*in bezeichnete die Zusammensetzung als homogen, eine*r als heterogen. Um die affektive Einstellungskomponente zu erfassen, wurden die Schüler*innen dazu befragt, welche Gefühle sie mit der Leistungsheterogenität in der eigenen Klasse verbinden. Hier konnten die Kategorien positive Emotion und negative Emotion gebildet werden. Bei der behavioralen Komponente berichteten manche Schüler*innen, die sich nach deren Wahrnehmung in einer homogenen bzw. heterogenen Klasse befanden, dass sie auch eine solche Zusammensetzung bevorzugen, manche jedoch auch, dass sie die jeweils andere Zusammensetzung bevorzugen. Die Begründungen für die Wahl können in die Kategorien sozial, leistungsspezifisch und unterrichtsspezifisch eingeteilt werden.

Tabelle 2. Kategoriensystem der Interviews.

Konstrukt	Definition	Kategorien	Ankerbeispiel
Wahrnehmung Leistungsheterogenität	Gründe für die eigene Wahrnehmung der Leistungsheterogenität in der Klasse	Meldeverhalten	„Manche melden sich halt sehr oft und so und manche nicht so viel.“
		Verständnis und Wissen	„Also es gibt ein paar, die manche Sachen noch nicht so gut verstehen und manche wissen einfach alles.“
		Noten	„Manche schreiben halt nur Einser und manche mehr so Fünfer.“
		Verhalten bei Leistungsrückmeldung	„Ein paar weinen immer bei ihren Noten [...] und ein paar freuen sich richtig.“
Affektive Einstellungskomponente	Gründe für auftretende Affekte in der eigenen Klasse	Positive Emotionen	„Ich fühl mich ganz wohl in meiner Klasse.“
		Negative Emotionen	„Über die Schlechteren ärgert man sich vielleicht öfter, weil [...], die fragen halt immer wieder, ob man das wiederholen kann und dann nervt das halt einen auch.“
Behaviorale Einstellungskomponente	Gründe für die Entscheidung, eine eher heterogene oder eher homogene Klasse zu bevorzugen	Soziale Faktoren	„Wenn ich etwas nicht verstehe, dann ärgert sich halt nicht die ganze Klasse über mich.“
		Leistungsspezifische Faktoren	„Dann kann ich von den allen was lernen.“
		Unterrichtsspezifische Faktoren	„Dann kann der Lehrer auch ein bisschen schneller machen.“

6.2 Forschungsfrage 1

Mit der konfirmatorischen Faktorenanalyse soll Forschungsfrage 1 beantwortet werden: Wie kann die ELH von Schüler*innen gemessen werden? Bei der Berechnung der konfirmatorischen Faktorenanalyse wurde für die Klassenzugehörigkeit kontrolliert. Die jeweiligen Items modellieren die kognitive, die affektive und die behaviorale Einstellungskomponente. Die drei Subkomponenten bilden dann gemeinsam das latente Konstrukt *Einstellung zur Leistungsheterogenität* bei Schüler*innen ab (siehe Tabelle 3). Ausgenommen Item k_10 ($\lambda = .19$) können die Faktorladungen der einzelnen Komponenten als gut interpretiert werden (Awang, 2014). Item k_10 wird aufgrund unzureichender Faktorladung für die zukünftige Modellierung und Analysen ausgeschlossen.

Tabelle 3. Faktorladungen der konfirmatorischen Faktorenanalyse.

Item	kog	aff	beh	ELH
k_1	.59			
k_2	.78			
k_3	.76			
k_4	.75			
k_5	.67			
k_6	.67			
k_7	.45			
k_8	.65			
k_9	.78			
k_11	.66			
k_12	.79			
k_13	.52			
a_1		.70		
a_2		.55		
a_3		.74		
a_4		.70		
a_5		.80		
a_6		.86		
a_7		.85		
a_8		.68		
a_9		.76		
a_10		.85		
a_11		.76		
a_12		.76		
a_13		.88		
a_14		.69		
a_15		.66		
b_1			.73	
b_2			.71	
b_3			.84	
b_4			.86	
b_5			.86	
cog				.75
aff				.92
beh				.32

Anmerkung. kog = Kognitive Einstellungskomponente; aff = Affektive Einstellungskomponente; beh = Behaviorale Einstellungskomponente; ELH = Einstellung zur Leistungsheterogenität bei Schüler*innen.

Das Modell der mehrebenenanalytischen konfirmatorischen Faktorenanalyse zeigt einen akzeptablen Model-Fit mit den robusten Fit-Indices Comparative Fit Index (CFI) = .928, Tucker Lewis Index (TLI) = .922, Root Mean Square Error of Approximation (RMSEA) = .047 (Byrne, 1994). Auch der relative Chi²-Index erreichte mit $\chi^2/df = 2.59$ die Akzeptanzgrenze von $\chi^2/df < 3.00$ (Kline, 1998).

6.3 Forschungsfrage 2

Mittels deskriptiver Analysen soll Forschungsfrage 2 beantwortet werden: Wie ist die ELH von Schüler*innen ausgeprägt? Tabelle 4 und Abbildung 4 zeigen deskriptive Ergebnisse zu den drei Einstellungskomponenten. Die Reliabilitäten sind mit $\alpha_{kog} = .91$, $\alpha_{aff} = .95$ und

$\alpha_{beh} = .90$ als sehr gut zu bezeichnen. Bei einem theoretischen Mittelwert von 3.5 (1 = negative Einstellung; 6 = positive Einstellung) liegen die kognitive Komponente ($M = 4.11$) und die affektive Komponente ($M = 4.30$) im positiven Bereich (siehe Abbildung 4A). Hiermit haben Schüler*innen eher positive Überzeugungen (kognitive Einstellungskomponente) sowie positive Emotionen (affektive Einstellungskomponente) gegenüber Leistungsheterogenität im Fach Deutsch. Die behaviorale Komponente liegt mit einem Mittelwert von $M = 3.5$ exakt auf dem Mittel der Skala. Ein Blick in die Verteilungseigenschaften der Einstellungskomponenten (siehe Abbildung 4B) zeigt, dass sich das Antwortmuster der Schüler*innen deutlich ändert, sobald die behaviorale Komponente herangezogen wird. Schüler*innen haben also keine klare Präferenz für eine durchschnittlich heterogene Klasse, wie der Mittelwert von $M = 3.5$ zuerst vermuten lässt. Ferner gibt es eine ähnliche Anzahl an Schüler*innen, welche sich für eine homogene, heterogene oder moderat heterogene Klasse entscheiden würden. Dieses Muster lässt sich auch bei den Korrelationen wiedererkennen, denn die kognitive und affektive Komponente hängen stärker miteinander zusammen als mit der behavioralen Komponente. Grundsätzlich zeigen sich zwischen den Einstellungskomponenten aber positive Korrelationen: Schüler*innen, welche positive Überzeugungen gegenüber Leistungsheterogenität haben, zeigen häufig positive Emotionen gegenüber Leistungsheterogenität ($r_{kog_aff} = .64$; $p < .001$) und würden mit höherer Wahrscheinlichkeit eine heterogene Klasse wählen ($r_{kog_beh} = .21$; $p < .001$; $r_{aff_beh} = .27$; $p < .001$).

Tabelle 4. Deskriptive Ergebnisse für die kognitive, affektive und behaviorale Einstellungskomponente.

	n	min	max	<i>M</i>	<i>SD</i>	skew	kurtosis	α	kog	aff	beh
kog	734	1	6	4.11	1.06	-.25	-.26	.91	-		
aff	738	1	6	4.30	1.12	-.50	-.34	.95	.64*	-	
beh	733	1	6	3.51	1.58	-.01	-1.06	.90	.21*	.27*	-

Anmerkung. kog = Kognitive Einstellungskomponente; aff = Affektive Einstellungskomponente; beh = Behaviorale Einstellungskomponente.

* $p < .001$

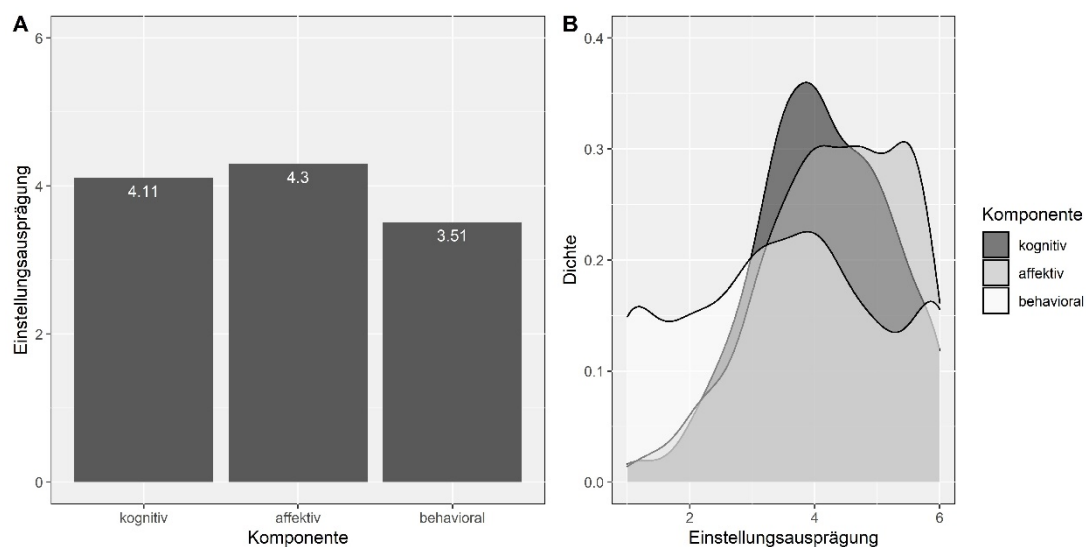


Abbildung 4. Deskriptive Analysen. (A) Balkendiagramm; (B) Überlagertes Dichtediagramm.

6.4 Forschungsfrage 3

Mittels inferenzstatistischer Analysen soll Forschungsfrage 3 beantwortet werden: Welche kontextuellen Faktoren können die ELH von Schüler*innen beeinflussen? In Tabelle 5 werden die inferenzstatistischen Ergebnisse zum Einfluss von kontextuellen Merkmalen auf die ELH von Schüler*innen dargestellt. Modell 0 ist das Baseline-Modell und berichtet über individuelle Hintergrundmerkmale und deren Einfluss auf die ELH von Schüler*innen. Es zeigt sich, dass die ELH unabhängig vom Geschlecht der Schüler*innen ($\beta = .06$; $p = .167$) und vom kulturellen Kapital ($\beta = .00$; $p = .833$) ist. Schüler*innen mit Migrationshintergrund zeigen eine positivere ELH als Schüler*innen ohne Migrationshintergrund ($\beta = .11$; $p < .05$). Dieser Effekt ist jedoch nur im Baseline-Modell signifikant.

In Modell 1 wird untersucht, ob die Leistungsheterogenität die ELH von Schüler*innen vorhersagen kann. Es zeigt sich, dass die Leistungsheterogenität in Deutsch einen positiven Effekt auf die ELH nimmt ($\beta = .09$; $p < .05$). Der Effekt ist unter Kontrolle der individuellen Leistung und der Klassenleistung stabil. An dieser Stelle zeigt sich, dass weder die Leistung der Schüler*innen selbst ($\beta = -.00$; $p = .607$), noch die mittlere Klassenleistung ($\beta = .01$; $p = .877$) einen Einfluss auf die ELH nimmt.

In Modell 2 wird getestet, wie sich das Ausmaß an Störungen in der Klasse auf die ELH von Schüler*innen auswirkt. Hier zeigt sich, dass Schüler*innen in Klassen, die ein geringeres Störungsausmaß berichten, eine positivere ELH aufweisen ($\beta = -.07$; $p < .05$). Dieser Effekt ist nur auf der Individualebene signifikant; die Klassenebene bietet keine weitere Aufklärung ($\beta = -.08$; $p = .37$).

Darüber hinaus nimmt Modell 3 in den Blick, wie der Umgang mit Störungen durch die Lehrkraft die ELH beeinflusst. Hier zeigt sich, dass ein professioneller Umgang mit Störungen die ELH von Schüler*innen positiv beeinflusst ($\beta = .08$; $p < .01$). Zudem hat die Variable auch auf Klassenebene einen positiven Effekt ($\beta = .18$; $p = .06$ mit marginaler Signifikanz). Der negative Effekt des Störungsausmaßes verschwindet ($\beta = -.04$; $p = .24$), sobald für den Umgang der Lehrkraft mit Störungen kontrolliert wird.

Tabelle 5. Multilevel-Strukturgleichungsmodell zur Vorhersage der Einstellung zur Leistungsheterogenität bei Schüler*innen.

Pfad	Model 0 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 3 β (SE)
Within Class Level				
ELH $\leftarrow$ Weiblich	.06 (.04)	.05 (.06)	.08 (.06)	.09 (.05)
ELH $\leftarrow$ MigHi	.11* (.06)	.10 (.06)	.07 (.06)	.04 (.06)
ELH $\leftarrow$ KultKap	.00 (.02)	.00 (.03)	.01 (.03)	.01 (.02)
ELH $\leftarrow$ Leistung		-.00 (.00)	-.00 (.00)	-.00 (.00)
ELH $\leftarrow$ Störung			-.07* (.03)	-.04 (.03)
ELH $\leftarrow$ Umgang				.08* (.03)
Between Class Level				
ELH $\leftarrow$ Heterogenität		.09* (.05)	.09* (.04)	.10* (.04)
ELH $\leftarrow$ Leistung_Kl		.01 (.05)	.00 (.05)	.01 (.04)
ELH $\leftarrow$ Störung			-.08 (.09)	.01 (.09)
ELH $\leftarrow$ Umgang				.18~ (.09)
Heterogenität $\leftarrow$ Realschule		-.05 (.25)	-.05 (.25)	-.05 (.25)
Leistung_Kl $\leftarrow$ Realschule		-.05 (.25)	-.05 (.25)	-.05 (.25)
Heterogenität $\leftarrow$ Gesamtschule		-.32 (.23)	-.32 (.23)	-.32 (.23)
Leistung_Kl $\leftarrow$ Gesamtschule		-.32 (.23)	-.32 (.23)	-.32 (.23)
Kovarianzen				
Heterogenität $\leftrightarrow$ Leistung_Kl		-.72* (.21)	-.72* (.21)	-.72* (.21)
AIC	71062.6	118233.2	116634.4	116771.3
BIC	71541.8	118874.0	117279.9	117424.8

Anmerkung. ELH = Einstellung zur Leistungsheterogenität bei Schüler*innen; MigHi = Migrationshintergrund (0 = kein Migrationshintergrund; 1 = Migrationshintergrund); KultKap = kulturelles Kapital; Leistung_Kl = mittlere Klassenleistung.

~ $p < .10$; * $p < .05$

7 Diskussion

Unsere Analysen decken drei wichtige Befunde auf. Erstens, die ELH von Schüler*innen kann über die multidimensionale Einstellungstheorie gemessen werden. Zweitens, die ELH von Schüler*innen ist positiv ausgeprägt. Und drittens, kontextuelle Merkmale der Unterrichtsumwelt spielen bei der ELH von Schüler*innen eine Rolle.

Unsere Befunde können zeigen, dass die ELH von Schüler*innen gemessen werden kann. In den Leitfadeninterviews wurde festgestellt, dass bereits Schüler*innen in der frühen Sekundarstufe das Konstrukt Leistungsheterogenität wahrnehmen. Die Auswertung des auf den Interviews basierenden Fragebogens mittels konfirmatorischer Faktorenanalyse zeigt gute Modell-Eigenschaften. Hier zeigt sich aber auch, dass vor allem die affektive Komponente einen großen Einfluss auf das Gesamtkonstrukt ELH hat. Im Vergleich dazu hat die behaviorale Komponente einen sehr geringen Aufklärungsanteil. Dieser Befund stimmt mit bisherigen Analysen und Ansätzen zur Definition von Einstellungen überein. So nehmen einige Wissenschaftler*innen an, dass die Einstellung allein mit der affektiven Komponente gemessen werden könne (z.B. Fishbein & Ajzen, 1975). Und andere Wissenschaftler*innen betonen, dass die behaviorale Komponente – als Komponente für die Vorhersage von Verhalten und Verhaltensintentionen – von dem Gesamtkonstrukt getrennt zu betrachten sei (z.B. Fishbein, 1967; Rosenberg, 1960). Auch wenn sich die Ungleichheit zwischen den einzelnen Komponenten in unseren Daten zeigt, können wir trotz allem einen akzeptablen Modell-Fit berichten, welcher die multidimensionale Herangehensweise unterstützt.

Unsere Befunde unterstreichen darüber hinaus, dass Schüler*innen eine positive ELH haben. Zudem zeigen sich stark signifikante, positive Korrelationen zwischen den einzelnen Komponenten. Dies bestätigt die grundsätzliche Annahme des Konsistenztheorems in der mehrdimensionalen Einstellungstheorie (Garzmann, 1982), bei der von einer prinzipiellen Konsistenz zwischen den verschiedenen Komponenten ausgegangen wird. Demnach haben Schüler*innen mit positiven Überzeugungen gegenüber Leistungsheterogenität auch positive Emotionen und würden wahrscheinlicher eine heterogene Klasse wählen.

Ein weiterer zentraler Befund unserer Analysen ist die Bedeutsamkeit von kontextuellen Faktoren für die ELH. So zeigt sich auf der einen Seite ein positiver Einfluss der tatsächlichen Leistungsheterogenität, aber auch das Handeln der Lehrkraft kann die ELH von Schüler*innen beeinflussen. Wie durch die Kontakthypothese angenommen, zeigt sich in allen Modellen ein positiver Effekt der Leistungsheterogenität auf die ELH. Befinden sich Schüler*innen also bereits in heterogenen Klassen und können einschätzen, welche Vor- und Nachteile eine solche Klasse mit sich bringt, ist die ELH positiver ausgeprägt. Die Analysen aus Modell 2 zeigen, dass ein hohes Ausmaß an Störungen die ELH von Schüler*innen negativ beeinflusst. Schüler*innen in Klassen, in denen viele Störungen während des Unterrichts auftreten, haben somit eine negativere ELH. In Modell 3 zeigt sich jedoch, dass dieser Einfluss verschwindet, wenn es der Lehrkraft gelingt, mit diesen Störungen umzugehen. Zudem hat der professionelle Umgang mit Störungen einen positiven Effekt auf die ELH von Schüler*innen. Unterbindet eine Lehrkraft Störungen im Unterricht schnell und reagiert angemessen auf Störungen (bspw. nicht zu mild und nicht zu streng), dann haben Schüler*innen eine positivere ELH. Die Analysen aus Modell 2 und Modell 3 unterstreichen hiermit, wie wichtig das Handeln der Lehrkraft für die ELH der Schüler*innen ist.

7.1 Limitationen

Bei der Interpretation der Ergebnisse müssen folgende Limitationen berücksichtigt werden. Die Stichprobe umfasst nur Schüler*innen der fünften Jahrgangsstufe, welche sich erst seit ungefähr zwei Monaten in der Klassengemeinschaft befinden. Die Varianz der Einstellungsskala und die systematischen Zusammenhänge weisen jedoch darauf hin, dass sich Schüler*innen bereits nach dieser relativ kurzen Zeit einen Eindruck über die Leistungsheterogenität der eigenen Klasse gemacht haben. Auch das Alter der Schüler*innen sollte bei der Interpretation der Ergebnisse in den Blick genommen werden. Zwar kann durch eine Vorstudie mit Leitfadeninterviews belegt werden, dass bereits junge Schüler*innen das Konstrukt Heterogenität fassen können, fraglich bleibt jedoch, wie sich diese Einstellung im Laufe der Schulzeit verändert. Eine weitere Limitation unserer Studie ist die Besonderheit der Stichprobe, denn 65%

der Schüler*innen weisen einen Migrationshintergrund auf. So kann die Übertragbarkeit auf Regionen mit geringem Migrationshintergrund der Schüler*innen (bspw. Landschulen) kritisch gesehen werden, weswegen die Fragestellungen auch in weiteren Datensätzen überprüft werden sollten.

7.2 Praktische Implikationen

Bisher stand vor allem die ELH von Lehrkräften im Fokus; und nach Trautmann & Wischer (2011) gilt eine positive ELH auf Seiten der Lehrkraft als eine Grundbedingung für einen erfolgreichen Umgang mit Heterogenität. Unsere Studie kann jedoch zeigen, dass die Leistungsheterogenität auch bei Schüler*innen eine wichtige Rolle im Schulsetting spielt. Trotz der teilweise negativen Affekte, die mit heterogenen Klassen einhergehen, zeigen die meisten Schüler*innen positive Tendenzen in deren ELH.

Unsere Befunde unterstreichen, dass Lehrkräfte die ELH von Schüler*innen positiv verändern können, wenn diese gezielt mit den Störungen im Unterricht umgehen. Deswegen ist es wichtig, die Ergebnisse der Studie auch in die Ausbildung von Lehramtsstudierenden miteinfließen zu lassen. An dieser Stelle ist es von Bedeutung, angehenden Lehrkräften ihre eigene ELH aufzuzeigen (bspw. durch Fragebögen für Lehrkräfte), aber auch die Bedeutung der ELH von Schüler*innen für deren Leistung zu thematisieren. Dass der professionelle Umgang mit Störungen im Unterricht das Lernen von Schüler*innen fördert, ist bereits durch zahlreiche Studien nachgewiesen worden (Helmke, 2012; Kunter & Trautwein, 2013). Lehrkräfte können durch ihr Handeln aber einen doppelten Effekt auf die Lernentwicklung von Schüler*innen nehmen. Einmal durch direkte Faktoren, indem eine klare und strukturierte Lernumgebung geschaffen wird. Und daneben auch über indirekte Faktoren, indem Lehrkräfte durch die Anpassung der Strategien an heterogene Klassen die ELH von Schüler*innen positiv verändern. Deswegen ist es von Bedeutung, dass angehenden Lehrkräften in der Ausbildung verdeutlicht wird, dass neben direkten Faktoren auch indirekte Faktoren (wie bspw. Einstellungen) eine wichtige Rolle für das Lernen spielen.

Literatur:

- Allport, G. (1954). *The nature of prejudice*. Garden City, NJ: Doubleday Anchor Books.
- Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first grade groups. *Elementary School Journal*, 79(1), 193–223.
- Armstrong, M. M., Morris, C., Abraham, C., & Tarrant, M. (2016). Interventions utilising contact with people with disabilities to improve children's attitudes towards disability: a systematic review and meta-analysis. *Disability and Health Journal*, 10(1), 11–22.
- Awang, Z. (2014). *A Handbook on SEM*. Selangor: MPWS Publisher.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. et al. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: deskriptive Befunde*. Opladen: Leske + Budrich.
- Bem, D.J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). San Diego, CA: Academic Press.
- Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: McGraw-Hill.
- 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* (S. 257–173). Regensburg: utb.
- Brophy, J. (1986). Classroom Management Techniques. *Education and Urban Society*, 18(2), 182–194.
- Burmeister, P., Bohl, T., & Meissner, S. (2016). Individualisierung im Englischunterricht. Forschungsstand, Wahrnehmung fremdsprachendidaktischer Spezifika und konzeptioneller Veränderungen durch Lehrkräfte. In T. Bohl & Wacker, A. (Hrsg.), *Die Einführung der Gemeinschaftsschule in Baden-Württemberg. Abschlussbericht der wissenschaftlichen Begleitforschung (WissGem)* (247–254). Münster: Waxmann.
- Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage Publications.
- Byrnes, D. A., & Kiger, G. (1997). Teachers Attitudes about Language Diversity. *Teaching and Teacher Education*, 13(6), 637–644.
- Czerwenka, K., Nölle, K., Pause, G., Schlotthaus, W., Schmidt, H. J. & Tesloff, J. (1990). *Schülerurteile über die Schule. Bericht über eine internationale Untersuchung*. Frankfurt: Peter Lang.
- Dickmann, A. (1995). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Hamburg: Rowohlt.
- Eagly, A.H. & Chaiken, S. (1993). *Psychology of Attitudes*. Fort Worth: Harcourt Brace Jovanovich College Publishers.
- Else-Quest, N.M., Mindeo, C.C. & Higgins, A. (2013). Math and Science Attitudes and Achievement at the Intersection of Gender and Ethnicity. *Psychology of Women Quarterly*, 37(3), 293–309.
- Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138.
- Evertson, C. M., Emmer, E. T. & Worsham, M. E. (2006). *Classroom management for elementary teachers* (7. Aufl.). Boston, MA: Allyn & Bacon.
- Fishbein, M. (1967). Attitudes and the prediction of behavior. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley.
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, behavior*. Reading, MASS: Addison-Wesley.
- Garzmann, R. (1982). Einstellungen. In H. Werbik & H. J. Kaiser (Hrsg.), *Kritische Stichwörter zur Sozialpsychologie* (S. 73–87). München: Fink
- Harteringer, A., Grittner, F., Lang, E. & Rehle, C. (2010). Ein Vergleich der Einstellung zur Heterogenität von Lehrkräften in jahrgangsgemischten und jahrgangshomogenen Klassen. In K.-H. Arnold, K. Hauenschild, B. Schmidt & Ziegenmeyer, B. (Hrsg.), *Zwischen Fachdidaktik und Stufendidaktik. Perspektiven für die Grundschulforschung* (S. 77–80). Berlin: VS Verlag für Sozialwissenschaften.
- Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5. Klassenstufe. *Zeitschrift für Pädagogische Psychologie*, 7, 77–86.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (4. Aufl.). Seelze: Klett-Kallmeyer.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. NY: Guilford Press.
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Paderborn: Schöningh.
- Linchevski, L. & Kutscher, B. (1998). Tell Me with Whom You're Learning, and I'll Tell You How Much You've Learned: Mixed-Ability versus Same-Ability Grouping in Mathematics. *Journal for Research in Mathematics Education*, 29(5), 533–554.
- Maehler, D. & Brinkman, H. U. (2016). *Methoden der Migrationsforschung - Ein interdisziplinärer Forschungsleitfaden*. Wiesbaden: Springer VS.

- Markic, S. & Abels, S. (2014). Heterogeneity and Diversity. A Growing Challenge or Enrichment for Science Education in German Schools? *Eurasia Journal of Mathematics, Science & Technology Education*, 10(4), 271–283.
- Martinez, R.S., Aricak, O.T. & Jewell, J. (2008). Influence of Reading Attitudes on Reading Achievement: A Test of the temporal-interaction Model. *Psychology in the Schools*, 45(10), 1010–1022.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.
- Papanastasiou, E., C. & Zembylas, M. (2002). The effect of Attitudes on Science Achievement: A Study conducted among High school pupils in Cyprus. *International Review of Education*, 48(6), 469–484.
- Paulus, C. (2009). *Die "Bücheraufgabe" zur Bestimmung des kulturellen Kapitals bei Grundschulern*. Saarbrücken: Universitäts- und Landesbibliothek.
- Piwowar, V. (2013). Multidimensionale Erfassung von Kompetenzen in Klassenmanagement. Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens. *Zeitschrift für Pädagogische Psychologie*, 27(4), 215–228.
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft* 1(13), 121–148.
- 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 Unterricht* (S. 257–173). Regensburg: utb.
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Verfügbar unter: <https://www.R-project.org/>.
- Rieger-Ladich, M. (2017). Ordnung stiften, Differenzen markieren. Machttheoretische Überlegungen zur Rede von Heterogenität. In T. Bohl, J. Budde & M. Rieger-Ladich (Hrsg.), *Umgang mit Heterogenität in Schule und Unterricht* (S. 257–173). Regensburg: utb.
- Rosenberg, M. J. & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I.
- Hovland & M. J. Rosenberg (Hrsg.), *Attitude organization and change* (S. 1-14). New Haven: Yale University Press.
- Ruhberg, C. & Porsch, R. (2017). Einstellungen von Lehramtsstudierenden und Lehrkräften zur schulischen Inklusion. *Zeitschrift für Pädagogik*, 63(4), 393–415.
- Scharenberg, K. (2012). *Leistungsheterogenität und Kompetenzentwicklung. Zur Relevanz klassenbezogener Kompositionsmerkmale im Rahmen der KESS-Studie* (Empirische Erziehungswissenschaft, Bd. 36). Münster: Waxmann.
- Slavin, R. E. (1987). Ability grouping and student achievement in primary schools: A best-evidence synthesis. *Review of Educational Research*, 57(3), 293–336.
- Statistisches Bundesamt (2018). *Bevölkerung und Erwerbstätigkeit, Bevölkerung mit Migrationshintergrund, Ergebnisse des Mikrozensus 2017*. Wiesbaden: destatis.
- Theunissen, U., Schoppengerd, E., Witzke, F., Endell, A. & Wenning, F. (2005). *Duisburger Sprachstandstest. Gesamtschule Duisburg-Meiderich*. Unveröffentlichtes Testinstrument.
- Trautmann, M. & Wischer, B. (2011). *Heterogenität und Schule. Eine kritische Einführung*. Wiesbaden: Springer.
- von der Groeben, A. (2003). Lernen in heterogenen Gruppen. Chancen und Herausforderung. *Pädagogik*, 55(9), 6–9.

Anhang E: Erklärungen über die Koautor*innenschaft

E.1: Stefanie Dotzel

E.2: Natalie Förster

E.3: Karin Hebbecker

E.4: Elmar Souvignier

E.5: Karina Karst

E.6: Meike Bonefeld