

# Selbstbezogene Fähigkeitskognitionen und Schulleistungen: Kausale Ordnung und Urteilsgenauigkeit der Lehrkräfte

## **Dissertation**

Zur Erlangung des Doktorgrades der Philosophischen Fakultät der  
Christian-Albrechts-Universität zu Kiel

vorgelegt von

Christian Schöber

Kiel, im April 2017

Erstgutachterin: Prof. Dr. Friederike Zimmermann

Zweitgutachter: Prof. Dr. Olaf Köller

Tag der mündlichen Prüfung: 27.06.2017

Durch den zweiten Prodekan Prof. Dr. Elmar Eggert zum Druck genehmigt: 05.07.2017

Die Forschungsarbeiten, die in dieser Dissertation berichtet werden, sind im Rahmen der Projekte *Entwicklung und Implementierung eines neuen Konzeptes zur Eingliederung Jugendlicher in die Berufs- und Arbeitswelt in Schulen mit erhöhtem Förderbedarf (EIKA)*, gefördert durch die Bremer Senatorin für Bildung und Wissenschaft sowie aus dem Europäischen Fonds für regionale Entwicklung (EFRE) und *Selbstwirksamkeit bei Kindern mit und ohne Migrationshintergrund: Differenzielle Erwerbsformen, diagnostische Fähigkeiten der Lehrkräfte und Bedeutung für die Leistungsentwicklung (Se-Mig)*, gefördert durch das Bundesministerium für Bildung und Forschung (BMBF; FKZ: 01JC1118B) entstanden.

## Danksagung

An dieser Stelle danke ich allen Personen, die auf unterschiedliche Weise zum Gelingen dieser Arbeit beigetragen haben. Zunächst geht mein Dank an Olaf Köller, der mich als Projektmitarbeiter ausgewählt hat und mir bei meiner Arbeit stets hilfreich zur Seite stand. Seine Freude an der wissenschaftlichen Arbeit und seine natürliche Gelassenheit haben viel zu meiner Motivation für diese Arbeit beigetragen. Friederike Zimmermann danke ich für die gemeinsame Arbeit an den *EIKA*-Daten zu Beginn meiner Promotionszeit sowie für die Übernahme des Erstgutachtens dieser Arbeit. Kerstin Schütte und Jan Marten Ihme danke ich herzlich für ihren Einsatz als Mentorin und Mentor während meiner Promotionszeit, sowie darüber hinaus gehende Gespräche zwischendurch. Anja Mikowski und Dennis Föste danke ich für hilfreiche Kommentierungen von Teilen dieser Arbeit.

Nele McElvany, Hanna Ferdinand, Miriam Gebauer und Thorben Huelmann gilt mein Dank für die gemeinsame Arbeit im Projekt *Se-Mig* und den Forschungsaufenthalt zur gemeinsamen Arbeit an Studie 3 dieser Arbeit. Meiner Radfahrergruppe sei herzlich für die entspannenden Flussradtouren gedankt, die mich zwischendurch immer wieder völlig abschalten ließen.

Schließlich gilt mein großer Dank meinen Eltern, die mich immer bedingungslos unterstützt haben und vor allem Anke, die in allen Phasen dieser Arbeit für mich da war.

## Zusammenfassung

Höhere selbstbezogene Fähigkeitskognitionen (Fähigkeitsselbstkonzepte und Selbstwirksamkeitsüberzeugungen) gehen mit stärkeren späteren Leistungen einher; starke Leistungen fördern höhere selbstbezogene Fähigkeitskognitionen. Bezüglich der kausalen Ordnung dieser Zusammenhänge liegen jedoch insbesondere für die Selbstwirksamkeitsüberzeugungen nur wenige Befunde vor. Zudem stellt sich die Befundlage insgesamt über Fachbereiche different dar, was eine Generalisierbarkeit in Frage stellt.

Für den Bereich des Lesens ist die empirische Unterstützung der kausalen Ordnung sehr gering. Hinzu kommt, dass in Deutschland wiederholt deutliche Leseleistungsunterschiede zwischen Schulkindern mit und ohne Migrationshintergrund bei vergleichbarem Niveau lesespezifischer Fähigkeitskonzepte vorgefunden wurden. Ein möglicher moderierender Einfluss des Migrationshintergrunds auf die kausale Ordnung von Leseleistungen und selbstbezogenen Fähigkeitskognitionen wurde jedoch nur unzureichend untersucht.

Die Selbstwirksamkeitsüberzeugungen von Schulkindern können durch Lehrkräfte beeinflusst werden, um ihren potenziell leistungsfördernden Effekt zu nutzen. Dazu sollten Lehrkräfte in der Lage sein, die Selbstwirksamkeitsüberzeugungen und die Leistungen ihrer Schülerinnen und Schüler möglichst genau zu beurteilen. Die Urteilsgenauigkeit von Lehrkräften in Bezug auf lesespezifische Selbstwirksamkeitsüberzeugungen ihrer Schulkinder und die Rolle des Migrationshintergrunds der Schulkinder bei der Beurteilung blieben bisher empirisch nahezu unbeachtet.

Im Rahmen dieser Dissertation wurde daher in zwei empirischen Arbeiten mit zwei Messzeitpunkten die kausale Ordnung zwischen selbstbezogenen Fähigkeitskognitionen und Schulleistungen mittels Strukturgleichungsmodellen evaluiert. Während in der ersten Studie der reziproke Zusammenhang zwischen dem verbalen Fähigkeitsselbstkonzept und den Leistungen im Lesen und in Rechtschreiben vom fünften bis zum siebten Schuljahr überprüft wurde, fokussierte die zweite Studie den Zusammenhang zwischen Selbstwirksamkeitsüberzeugungen und Leistungen im Lesen und in Mathematik innerhalb des siebten Schuljahres. Hierbei wurde jeweils eine mögliche Moderation durch den Migrationshintergrund durch die Invarianz der Mess- und Strukturmodelle getestet.

In einer dritten empirischen Studie wurde querschnittlich die Urteilsgenauigkeit von Lehrkräften in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen ihrer Schulkinder mehrbenenanalytisch unter Berücksichtigung von Prädiktoren auf der Individualebene und der Schulklassenebene untersucht. Im Rahmen dieser Studie wurde ein möglicher differenzieller Einfluss des Migrationshintergrundes auf die Urteile von Lehrkräften durch die Berücksichtigung unterschiedlicher Herkunftsgruppen untersucht.

Während die erste Studie die Annahme reziproker Effekte zwischen sprachlichen Leistungen und dem verbalen Fähigkeitsselbstkonzept bestätigte, wiesen die Ergebnisse der zweiten Studie auf kausale Effekte in unterschiedlichen Richtungen hin. Beide Studien weisen jedoch auf die kulturelle Übertragbarkeit der Befunde hin. In der dritten Studie zeigten die Lehrkräfte eine schwache Urteilsgenauigkeit und eine differenzielle Beurteilung insbesondere für Kinder mit einem türkischen Migrationshintergrund.

Die Ergebnisse der drei Studien werden abschließend zusammengefasst und mit Blick auf offene Forschungsfragen und praktische Implikationen diskutiert.

## Inhaltsverzeichnis

1. Einleitung und theoretischer Rahmen der Arbeit .....	9
2. Studie 1: Verbales schulisches Selbstkonzept und sprachliche Leistungen in Gruppen mit und ohne Migrationshintergrund .....	113
3. Studie 2: Reciprocal effects between self-efficacy and achievement within mathematics and reading .....	157
4. Studie 3: Teachers' judgments of students' self-efficacy in reading: Accuracy and students' migrant origins .....	203
5. Gesamtdiskussion .....	255





# **1** **Einleitung und theoretischer Rahmen der Arbeit**

## 1.1 Einleitung

One of the thorniest issues in the research of the relationship between academic self-beliefs and academic achievement is the chicken-and-egg question of causality. Succinctly put, the question asks “Do students’ academic self-beliefs determine their academic achievement, or does academic achievement determine their self-beliefs?” (Pajares & Schunk, 2005, S. 112).

Dieses mittlerweile gut zehn Jahre alte Zitat verdeutlicht, dass bezüglich selbstbezogener Fähigkeitskognitionen, wie Fähigkeitsselbstkonzepten (Shavelson, Hubner & Stanton, 1976) und Selbstwirksamkeitsüberzeugungen (Bandura, 1977), die kausale Ordnung mit Leistungen auch nach Jahrzehnten mit besonderen Schwierigkeiten verbunden ist. Dies liegt vor allem daran, dass insbesondere Fähigkeitsselbstkonzepte, aber auch Leistungen, einer experimentellen Manipulation schwer zugänglich sind (Pajares & Schunk, 2001). Deshalb wurden methodische Vorgaben zur Überprüfung der kausalen Ordnung erarbeitet (Marsh, Byrne & Yeung, 1999; Marsh & Martin, 2011), deren Einhaltung insbesondere Längsschnittdaten mit großen Stichproben voraussetzt.

Während Fähigkeitsselbstkonzepte auf Bewertungen von in Schulfächern gesammelten Kompetenzerfahrungen beruhende generalisierte fachspezifische Fähigkeitseinschätzungen darstellen (Eckert, Schilling & Stiensmeier-Pelster, 2006; Möller & Köller, 2004), bezeichnet die Selbstwirksamkeitsüberzeugung die wahrgenommene Fähigkeit, Handlungen auf einem bestimmten Niveau lernen oder ausführen zu können (Bandura, 1997; Schunk & Meece, 2006; Schunk & Pajares, 2009). Die Fähigkeitsselbstkonzepte stellen Kompetenzbewertungen, die Selbstwirksamkeitsüberzeugungen hinge-

gen Kompetenzbeschreibungen dar (Parker, Marsh, Ciarrochi, Marshall & Abduljabbar, 2014). Beide rekurren jedoch auf die subjektive Wahrnehmung der eigenen Leistungen (Lee, 2009). Die Wahrnehmungen eigener Leistungen tragen entscheidend zur Entwicklung von Fähigkeitsselbstkonzepten und Selbstwirksamkeitsüberzeugungen bei (Bandura, 1993; 1997; Möller & Köller, 2004; Pajares & Schunk, 2005). Umgekehrt konnten für Fähigkeitsselbstkonzepte und Selbstwirksamkeitsüberzeugungen positive Zusammenhänge mit der Auswahl herausfordernder Aufgaben sowie mit der Anstrengung und Ausdauer bei deren Bearbeitung gefunden werden, wodurch sie, vermittelt über diese Zusammenhänge, positiv mit gezeigten Leistungen assoziiert waren (Bong & Skaalvik, 2003). Die Förderung selbstbezogener Fähigkeitskognitionen, um positive Effekte auf die Leistungsentwicklung nutzbar zu machen, sind daher Gegenstand vieler Empfehlungen in der rezenten Literatur (siehe z. B. Pajares & Schunk, 2001; Schunk & Meece, 2006).

Für die Bildungsbeteiligung und den Bildungserfolg werden insbesondere sprachliche Leistungen, beispielsweise im Lesen, als zentral angesehen (Autorengruppe Bildungsberichterstattung, 2016; Baumert, Watermann & Schümer, 2003; Esser, 2006; Hopf, 2005). So werden durch Lesen die fachspezifischen Inhalte der meisten anderen Schulfächer erlernt und dadurch ein erfolgreicher Übergang in die berufliche Ausbildung oder der Zugang zur Hochschule mit anschließendem Übergang in den Arbeitsmarkt ermöglicht (Artelt et al., 2007; Weis et al., 2016). In großen Vergleichsstudien mit repräsentativen Stichproben zeigten sich jedoch insbesondere in Deutschland große Leseleistungsnachteile von Schulkindern mit Migrationshintergrund im Vergleich zu Schulkindern ohne Migrationshintergrund (Schwippert, Wendt & Tarelli, 2012; Stanat, Rauch & Segeritz, 2010). Dabei war die Ausprägung des lesespezifischen Fähigkeitsselbstkon-

zeptes in beiden Gruppen vergleichbar (Roebbers, Mecheril & Schneider, 1998; Shajek, Lüdtke & Stanat, 2006). Obwohl damit die Kalibrierung der Leseleistung und des lese-spezifischen Fähigkeitsselbstkonzeptes in den Gruppen mit und ohne Migrationshintergrund unterschiedlich zu sein scheint, blieb ein möglicher moderierender Einfluss des Migrationshintergrunds auf die Beziehung zwischen beiden Variablen und ihre kausale Ordnung unberücksichtigt und ist noch immer ein Desiderat der Forschung (Valentine, Dubois & Cooper, 2004).

Insbesondere für Fähigkeitsselbstkonzepte und Leistungen existiert viel Forschung zur kausalen Ordnung, die bereits in einem Überblicksartikel zusammengefasst wurde (Marsh & Craven, 2006). Dennoch liegen nur wenige Erkenntnisse zur kausalen Ordnung im sprachlichen Bereich vor (für eine Ausnahme im Bereich des Lesens siehe Retelsdorf, Köller & Möller, 2014). Für Selbstwirksamkeitsüberzeugungen liegen sogar insgesamt nur sehr wenige Befunde zur kausalen Ordnung mit Leistungen vor (Hannula, Bofah, Tuohilampi & Metsämuuronen, 2014; Hwang, Choi, Lee, Culver & Hutchison, 2016; Mittag, Kleine & Jerusalem, 2002; Williams & Williams, 2010). Diese entsprechen zudem nicht immer den methodischen Richtlinien von Marsh et al. (1999) und ließen den wichtigen sprachlichen Bereich bisher unbeachtet (siehe Abschnitt 1.3.4). Damit ist bisher nicht klar, ob und wie sich ein Migrationshintergrund von Schulkindern moderierend auf die kausale Ordnung zwischen selbstbezogenen Fähigkeitskognitionen und Leistungen im sprachlichen Bereich auswirkt.

Die Beantwortung der kausalen Ordnung von Leistungen und selbstbezogenen Fähigkeitskognitionen geht mit wichtigen Implikationen für Schule und Unterricht einher. So sollten positive Effekte selbstbezogener Fähigkeitskognitionen auf spätere Leis-

tungen durch unterrichtliche Maßnahmen zur direkten Förderung dieser Kognitionen genutzt werden (Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005; Pajares & Schunk, 2001; Schunk & Meece, 2006). Eine wichtige Voraussetzung für die individuelle Förderung von Schulkindern sind jedoch möglichst genaue Urteile seitens der Lehrkraft in Bezug auf die zu fördernden Schulkindmerkmale (Praetorius, Karst, Dickhäuser & Lipowsky, 2011; Ready & Wright, 2011). Während zur Urteilsgenauigkeit in Bezug auf Schulleistungen bereits umfangreiche Befunde und Metaanalysen vorliegen (Hoge & Coladarci, 1989; Südkamp, Kaiser & Möller, 2012), liegen zur Urteilsgenauigkeit in Bezug auf selbstbezogene Fähigkeitskognitionen nur wenige Befunde vor, die zudem selten im sprachlichen Bereich verortet sind (für Ausnahmen siehe Coddington & Guthrie, 2009; Praetorius et al., 2011).

Die vorliegende Dissertation greift die genannten Forschungslücken mit drei empirischen Arbeiten auf. In der ersten Studie wurde das Ziel verfolgt, die kausale Ordnung zwischen verbalem Fähigkeitsselbstkonzept und Leistungen in den Bereichen Lesen und Rechtschreiben für Schulkinder mit und ohne Migrationshintergrund über zwei Jahre (fünfte bis siebte Klassenstufe) empirisch zu unterfüttern. Dabei war, neben den gegenseitig fördernden Effekten zwischen den Leistungen und dem Fähigkeitsselbstkonzept, auch die Frage nach der Entwicklung von Leistungen und verbalem Fähigkeitsselbstkonzept in beiden Gruppen über zwei Schuljahre hinweg zu beantworten. Im Rahmen der zweiten Studie wurde die kausale Ordnung mit Blick auf die Entwicklung von Leseleistungen und lesespezifischen Selbstwirksamkeitsüberzeugungen sowie zum Vergleich die Entwicklung von Mathematikleistungen und mathematikspezifische Selbstwirksamkeitsüberzeugungen innerhalb der siebten Klassenstufe überprüft, wobei hier ebenfalls der

Zuwanderungshintergrund der Schulkinder berücksichtigt wurde. Die dritte empirische Arbeit beschäftigte sich, in Anknüpfung an die zweite Studie, mit der Genauigkeit der Lehrkrafturteile in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen ihrer Schulkinder zu Beginn der siebten Klasse. Auf die Genauigkeit des Lehrkrafturteils könnten sich individuelle Schulkindmerkmale, wie beispielsweise deren Migrationshintergrund, differenziell auswirken (siehe Abschnitt 1.4.3). Lehrkräfte könnten also die lesespezifischen Selbstwirksamkeitsüberzeugungen ihrer Schulkinder in Abhängigkeit der ethnischen Herkunft des Schulkindes systematisch unterschiedlich beurteilen. Deshalb wurden in der dritten Studie neben den Schulkindern ohne Migrationshintergrund die Schulkinder mit Migrationshintergrund getrennt nach Herkunftsgruppen (Türkei, Staaten der ehemaligen Sowjetunion und Polen, sonstige Staaten) betrachtet. Die verwendeten Herkunftsgruppen orientieren sich an den größten Zuwanderungsgruppen in Deutschland (Rauch, Mang, Härtig & Haag, 2016).

Zunächst sollen in einer theoretischen Einführung die Fähigkeitsselbstkonzepte und Selbstwirksamkeitsüberzeugungen konzeptuell betrachtet sowie empirische Befunde zur kausalen Ordnung dieser beiden selbstbezogenen Fähigkeitskognitionen und Leistungen berichtet werden, bevor näher auf das Konzept der Urteilsgenauigkeit von Lehrkräften eingegangen wird. Auf den theoretischen Teil folgt der empirische Teil mit den drei skizzierten Studien. Schließlich werden die drei Studien zusammengefasst und im Hinblick auf ihre theoretischen und praktischen Konsequenzen diskutiert.

## 1.2 Fähigkeitsselbstkonzepte und Selbstwirksamkeitsüberzeugungen

Selbstbezogene Fähigkeitskognitionen wie Fähigkeitsselbstkonzepte und Selbstwirksamkeitsüberzeugungen haben eine hohe praktische Bedeutsamkeit im Bildungskontext, da sie positiv mit dem Verhalten in schulischen Leistungssituationen zusammenhängen (Pajares & Schunk, 2001). Das heißt, Schulkinder mit höheren selbstbezogenen Fähigkeitskognitionen zeigen stärkere Leistungszuwächse als ihre Klassenkameradinnen und Klassenkameraden mit geringeren selbstbezogenen Fähigkeitskognitionen (Arens, Marsh, Pekrun et al., 2016; Caprara et al., 2008; Huang, 2011; Marsh & Craven, 2006; Möller, Retelsdorf, Köller & Marsh, 2011; Valentine et al., 2004). Im Gegenzug gehen stärkere Leistungen mit höheren späteren Fähigkeitskognitionen einher (Guay, Marsh & Boivin, 2003; Hwang et al., 2016; Marsh et al., 1999; Marsh & Craven, 2006; Valentine & Dubois, 2005; Williams & Williams, 2010). Trotz dieser Gemeinsamkeit weisen Fähigkeitsselbstkonzepte und Selbstwirksamkeitsüberzeugungen ein paar entscheidende konzeptuelle wie psychometrische Unterschiede auf (Bong & Clark, 1999; Bong & Skaalvik, 2003; Zimmerman, 2000). Gegenstand dieses Abschnittes ist daher zunächst der Vergleich von Fähigkeitsselbstkonzepten und Selbstwirksamkeitsüberzeugungen, bevor auf das Zusammenspiel zwischen Schulleistungen und selbstbezogenen Fähigkeitskognitionen eingegangen wird.

Fähigkeitsselbstkonzepte sind definiert als auf Bewertungen von in Schulfächern gesammelten Kompetenzerfahrungen beruhende generalisierte fachspezifische Fähigkeitseinschätzungen (Eckert et al., 2006; Möller & Köller, 2004). Selbstwirksamkeitsüberzeugungen bezeichnen die wahrgenommene Fähigkeit, Handlungen auf einem bestimmten Niveau lernen oder ausführen zu können (Bandura, 1997; Schunk & Meece,

2006; Schunk & Pajares, 2009). Durch diese beiden Definitionen wird deutlich, dass das zentrale Element beider Konstrukte durch die subjektive Wahrnehmung der Schulkinder gebildet wird. Während bei Fähigkeitsselbstkonzepten die wahrgenommene Kompetenz das zentrale Element abbildet, fokussieren schulische Selbstwirksamkeitsüberzeugungen auf der wahrgenommenen Erfolgswahrscheinlichkeit (Bong & Skaalvik, 2003). Dieser konzeptuelle Unterschied manifestiert sich in der Komposition der beiden Konstrukte. Während die Kompetenzwahrnehmung im Rahmen des Fähigkeitsselbstkonzeptes aus kognitiven wie affektiven Bewertungen zusammengesetzt ist, besteht die wahrgenommene Erfolgswahrscheinlichkeit als Selbstbeschreibung ausschließlich aus einer kognitiven Komponente (Bong & Clark, 1999; Bong & Skaalvik, 2003). Um ihre Fähigkeitsselbstkonzepte zu erschließen, muss sich eine Person Fragen über das *Sein* und das *Fühlen* stellen, mit deren Beantwortung klar wird wie positiv oder negativ sie sich selbst in der abgefragten Situation fühlen (Pajares & Schunk, 2001). Demgegenüber sind Selbstwirksamkeitsüberzeugungen allein Fragen um das *Können* gewidmet, also der jeweiligen Zuversicht eine bestimmte Aufgabe mit den eigenen vorhandenen Kompetenzen erfolgreich zu bewältigen (Pajares & Schunk, 2001). Beim Fähigkeitsselbstkonzept geht es also beispielsweise um das Gefühl über sich selbst als lesende Person, während die Selbstwirksamkeitsüberzeugung zum Beispiel die Zuversicht ausdrückt, einen Zeitungsartikel zu verstehen. Obwohl die Selbstwirksamkeitsüberzeugung einen Zeitungsartikel zu verstehen sehr stark ausgeprägt sein kann, kann dennoch das lesespezifische Selbstkonzept durch affektive Bewertungen (z. B. „Wer liest, lebt zurückgezogen und ist langweilig!“) trotz hoher Kompetenzbewertung insgesamt gering ausgeprägt sein.



Während die Selbstwirksamkeitsüberzeugungen als rein kognitive Einschätzungen eigener Kompetenzen in Bezug auf die Bewältigung bestimmter Herausforderungen unstrittig erscheint (Pajares, 1996), wird der affektive Bewertungsanteil der Fähigkeitsselbstkonzepte in einigen Arbeiten nicht als bedeutsam anerkannt (z. B. Schöne, Dickhäuser, Spinath & Stiensmeier-Pelster, 2003; siehe jedoch Möller & Köller, 2004; Pajares & Miller, 1994). Begründet wird diese Fokussierung auf kognitive Bewertungsanteile mit einer größeren theoretischen Klarheit durch die Abgrenzung zum Selbstwert (siehe Abschnitt 1.2.1) und der Möglichkeit, gezieltere Interventionen durchführen zu können (Schöne et al., 2003). Bereits für Kinder im Vor- und Grundschulalter konnte jedoch bestätigt werden, dass Fähigkeitsselbstkonzepte sowohl affektbezogene als auch kompetenzbezogene Bestandteile enthalten, wobei die affektbezogenen Bestandteile geringere Zusammenhänge mit Leistungen aufwiesen (Arens, Marsh, Craven et al., 2016; Arens, Yeung, Craven & Hasselhorn, 2011). Diese empirische Bestätigung wird auch theoretisch unterstützt durch die Annahme, dass die eigene Bewertung der Person als gut (oder schlecht) zum Beispiel im Lesen nicht ohne affektive Auswirkungen bleiben wird (Möller & Trautwein, 2015). Obwohl die wissenschaftliche Debatte über die Bestandteile der Kompetenzbewertung im Rahmen der Fähigkeitsselbstkonzepte keineswegs abgeschlossen ist, wird in dieser Arbeit dem Ansatz gefolgt, dass Fähigkeitsselbstkonzepten neben der kognitiven Bewertung auch eine affektive Bewertung der Person inhärent ist.

Da Fähigkeitsselbstkonzepte vergangenheitsbezogene Kompetenzbewertungen sind (Bong & Skaalvik, 2003), gelten diese oftmals als konstituierend für die zukunftsgerichtete Erfolgsoptimismus, die im Rahmen der Selbstwirksamkeitsüberzeugungen erfasst wird (Schwarzer & Jerusalem, 2002). Dementgegen wird auch angenommen, dass die

wahrgenommene Erfolgszuversicht ein Teil der Kompetenzbewertung sei, wodurch dann die Selbstwirksamkeitsüberzeugungen zum konstituierenden Teil des Fähigkeitsselbstkonzeptes würden (Bong & Clark, 1999; Pajares, 1996). Um diesen Konflikt aufzulösen verglichen Ferla, Valcke & Cai (2009) die Modellgüte zweier Pfadmodelle mit jeweils unidirektionalen Pfaden zwischen mathematikspezifischer Selbstwirksamkeitsüberzeugung und mathematikspezifischem Fähigkeitsselbstkonzept sowie einem dritten Modell mit reziproker Beziehung zwischen beiden Konstrukten. Im Ergebnis wies das Modell mit dem reziproken Zusammenhang zwischen mathematikspezifischer Selbstwirksamkeitsüberzeugung und mathematikspezifischem Fähigkeitsselbstkonzept die beste Modellpassung auf (Ferla et al., 2009). Demnach ist anzunehmen, dass beide Konstrukte sowohl Ursache als auch Effekt des jeweils anderen sind. Einschränkend zeigte sich, dass der Pfad mit dem direkten Einfluss der mathematikspezifischen Selbstwirksamkeitsüberzeugungen auf das mathematikspezifische Fähigkeitsselbstkonzept in diesem dritten Modell fast nicht mehr signifikant war. Ferla et al. schlossen daraus, dass der Einfluss des mathematikspezifischen Fähigkeitsselbstkonzepts auf die mathematikspezifischen Selbstwirksamkeitsüberzeugungen insgesamt stärker ist als in die andere Richtung.

In der Studie von Ferla et al. (2009) wurde zudem die empirische Trennbarkeit von mathematikspezifischen Fähigkeitsselbstkonzepten und mathematikspezifischen Selbstwirksamkeitsüberzeugungen über konfirmatorische Faktorenanalysen mit einem beziehungsweise zwei Faktoren durch die bessere Modellpassung des Zwei-Faktoren-Modells bestätigt. Die Trennbarkeit der Konstrukte wurde zusätzlich dadurch untermauert, dass keine Kreuzladungen beobachtbar waren, alle Ladungen Werte über .50 aufwiesen und die internen Konsistenzen jeder Skala mit  $\alpha > .80$  gut waren (Ferla et al., 2009).

Allerdings wurden beide Konstrukte auf unterschiedlichen Ebenen der Spezifität gemessen. Während das mathematikspezifische Fähigkeitsselbstkonzept auf einer breiteren mathematikspezifischen Ebene erhoben wurde, wurden für die Erfassung der mathematikspezifischen Selbstwirksamkeitsüberzeugungen der Kinder spezifischere Itemformulierungen verwendet, die einen konkreten Aufgabenbezug enthielten. Dieses Vorgehen wurde auch in weiteren Studien zur empirischen Trennbarkeit von Fähigkeitsselbstkonzepten und Selbstwirksamkeitsüberzeugungen aus den Bereichen Mathematik (Parker et al., 2014) und Naturwissenschaften (Jansen, Scherer & Schroeders, 2015) gewählt. Entsprechend weisen Jansen et al. (2015) auch darauf hin, dass die gefundene empirische Trennbarkeit der Konstrukte bei Verwendung von Items mit ähnlicher Spezifität nicht in gleichem Maße gegeben sein muss.

Die Spezifität ist ein zentrales Merkmal der hierarchischen und multidimensionalen Struktur beider selbstbezogener Fähigkeitskognitionen. Dabei sind Fähigkeitsselbstkonzepte vergleichsweise strenger an hierarchischen Ordnungen strukturiert als Selbstwirksamkeitsüberzeugungen, was in den folgenden beiden Abschnitten thematisiert wird.

### **1.2.1 Hierarchie und Multidimensionalität – Ausführungen zur Struktur**

Der strukturelle Aufbau von Fähigkeitsselbstkonzepten und Selbstwirksamkeitsüberzeugungen weist Gemeinsamkeiten und Unterschiede auf. Sowohl Fähigkeitsselbstkonzepte als auch die Selbstwirksamkeitsüberzeugungen sind Bestandteile multidimensionaler und mehr oder weniger hierarchischer Strukturen (Bong & Clark, 1999; Bong & Skaalvik, 2003). Die Annahme einer hierarchischen Struktur des Selbstkonzepts basiert auf der Arbeit von Shavelson et al. (1976), in der die bis dahin eher unsystemati-

sche Selbstkonzeptforschung zusammengefasst und in ein hypothetisches Konstrukt mit einem allgemeinen Selbstkonzept an der Spitze der Hierarchie zusammengefasst wurde. In diesem Modell (siehe Abbildung 1.1) werden unterhalb des allgemeinen Selbstkonzepts sowohl ein Fähigkeitsselbstkonzept als auch mehrere nichtakademische Selbstkonzepte angenommen (das soziale, das emotionale und das physische Selbstkonzept). Unterhalb des Fähigkeitsselbstkonzeptes befinden sich fachspezifische Fähigkeitsselbstkonzepte beispielsweise für Mathematik, Naturwissenschaften, Sprachen und Geschichte, die auf einer darunterliegenden Ebene durch spezifischere Erfahrungen und schlussendlich auf unterster Ebene durch die Bewertung des eigenen Verhaltens in spezifischen Situationen komplettiert werden.

Shavelson et al. nahmen an, dass das auch als Selbstwert (Baumeister, Campbell, Krueger & Vohs, 2005; Pelham & Swann, 1989; Rosenberg, 1989) bezeichnete allgemeine Selbstkonzept sehr stabil sei und diese Stabilität auf den unteren Ebenen abnehme. So brauche es viele einzelne situationsspezifische Verhaltensweisen auf unterster Ebene, die konträr zum allgemeinen Selbstkonzept wahrgenommen würden, um eine Veränderung des allgemeinen Selbstkonzepts auslösen zu können. Obwohl einige Annahmen des Modells bezüglich der Hierarchie mittlerweile auf der Basis empirischer Befunde modifiziert werden mussten, blieben die grundlegenden Erkenntnisse zur Struktur des Selbstkonzeptes gültig. Damit bildet diese grundlegende Arbeit einen historischen Startpunkt für eine theoriegeleitete Forschung zum Selbstkonzept.

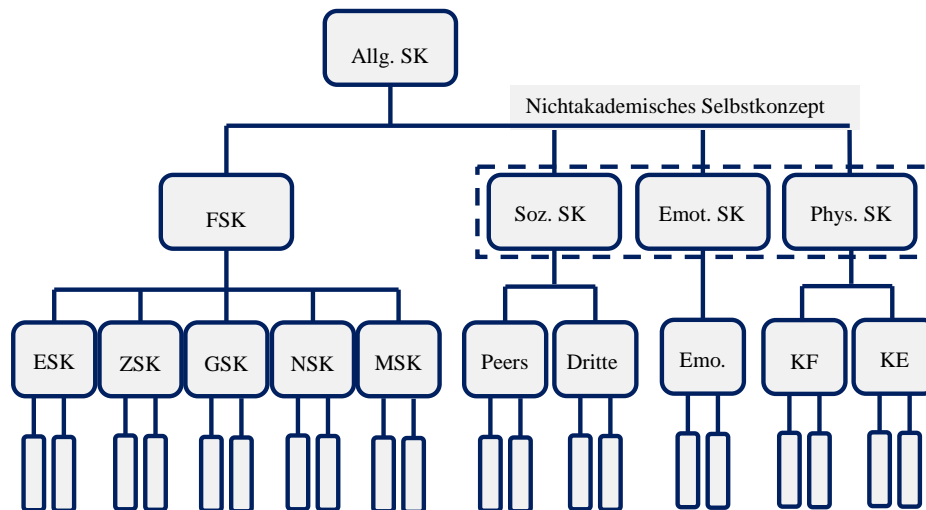


Abbildung 1.1. Hierarchisches Selbstkonzeptmodell (modifiziert nach Shavelson, Hubner & Stanton, 1976). Allg. SK = allgemeines Selbstkonzept; FSK = Fähigkeitsselbstkonzept; Soz. SK = soziales Selbstkonzept; Emot. SK = emotionales Selbstkonzept; Phys. SK = physisches Selbstkonzept; ESK = erstsprachliches Fähigkeitsselbstkonzept; ZSK = zweitsprachliches Fähigkeitsselbstkonzept; GSK = geschichtsspezifisches Fähigkeitsselbstkonzept; NSK = naturwissenschaftsspezifisches Fähigkeitsselbstkonzept; MSK = mathematikspezifisches Fähigkeitsselbstkonzept. Peers = Gleichaltrige; Dritte = weitere wichtige Personen außer Peers; Emo. = bestimmte emotionale Zustände; KF = körperliche Fähigkeiten; KE = körperliches Erscheinungsbild.

Beim Fähigkeitsselbstkonzept besteht mittlerweile Einigkeit darüber, dass in der hierarchischen Ordnung zumindest zwischen verbalem und mathematikspezifischem Fähigkeitsselbstkonzept zu trennen ist. Diese Erkenntnis resultierte aus dem Befund, dass verbale und mathematische Leistungen substantielle Korrelationen aufwiesen, während das verbale und das mathematikspezifische Fähigkeitsselbstkonzept nahezu unkorreliert miteinander waren (Marsh, 1986; 1990a; Möller & Köller, 2004; Möller & Marsh, 2013). Möller und Köller (2004) berichten zusammenfassend von Korrelationen der Leistungen

zwischen  $r = .31$  und  $r = .94$  ( $Md = .63$ ), während die Fähigkeitsselbstkonzepte Korrelationen von  $r = -.13$  und  $r = .22$  aufwiesen ( $Md = .10$ ).

Obwohl die Trennung zwischen mathematikspezifischem und verbalem Fähigkeitsselbstkonzept als kulturübergreifend gültig angenommen werden kann (Marsh & Hau, 2004), existieren auch zu dieser als *Marsh/Shavelson-Modell* (Marsh & Shavelson, 1985) bekannt gewordenen revidierten Modellformulierung empirische Befunde, die mit den Modellannahmen unvereinbar sind. Dazu gehört zum Beispiel die geringe Korrelation von Fähigkeitsselbstkonzepten in unterschiedlichen Sprachen (Marsh, Kong & Hau, 2001; Marsh & Yeung, 2001). So zeigten sich in einer Längsschnittstudie an einer Stichprobe chinesischer Schulkinder die Selbstkonzepte in der Erstsprache Chinesisch und der Fremdsprache Englisch nahezu unkorreliert miteinander (zwischen  $r = -.07$  und  $r = .19$ ). Bei den sprachlichen Leistungen zeigten sich negative Effekte auf die differenten sprachlichen Fähigkeitsselbstkonzepte, also negative Effekte der Englischleistung auf die Fähigkeitsselbstkonzepte in Chinesisch und ebenso negative Effekte der Chinesischleistung auf die Fähigkeitsselbstkonzepte in Englisch (Marsh, Kong et al., 2001). Mit diesem Befund ließ sich die Einteilung in nur ein verbales Fähigkeitsselbstkonzept unter Berücksichtigung unterschiedlicher Sprachen nur schwer aufrechterhalten. Diese Inkonsistenz in der hierarchischen Struktur lässt vermuten, dass die Entwicklung des erstsprachlichen Fähigkeitsselbstkonzeptes nur geringe Zusammenhänge mit der Entwicklung fremdsprachlicher Fähigkeitsselbstkonzepte aufweist. Insofern könnte die Interpretation von Befunden erschwert sein, wenn beispielsweise nur ein verbales Fähigkeitsselbstkonzept in gemischten Schulkindgruppen erfasst wird, bei denen für einen Teil die Unterrichtssprache und die Erstsprache identisch und für den anderen Teil different sind. Hier bietet

sich die Inspektion der Entwicklung verbaler Fähigkeitsselbstkonzepte in Abhängigkeit des sprachlichen Hintergrundes der Schulkinder an.

Ein in der rezenten Literatur diskutierter Ansatz zur Auflösung der Inkonsistenz des Marsh/Shavelson-Modells entstand durch die empirische Prüfung einer verschachtelten Struktur des Fähigkeitsselbstkonzeptes (*nested Marsh/Shavelson Model*, NMS-Modell; Brunner et al., 2010). Das NMS-Modell ist in Abbildung 1.2 schematisch dargestellt. Hier wird die ursprüngliche Auffassung von Shavelson et al. (1976) mit nur einem Fähigkeitsselbstkonzept aufgegriffen. Dieses hat im NMS-Modell jedoch direkt Einfluss auf die einzelnen fachspezifischen Erfahrungen innerhalb der Schulfächer sowie auf allgemein schulbezogene Erfahrungen. Die fachspezifischen Erfahrungen wiederum sind Bestandteil fachspezifischer Selbstkonzeptfacetten, womit klar wird, dass das Fähigkeitsselbstkonzept im NMS-Modell nur die domänenübergreifende Varianz beinhaltet (Brunner et al., 2010). Zwischen den fachspezifischen Selbstkonzeptfacetten werden zusätzlich Korrelationen erlaubt, wobei deren Ausprägung nicht vorgegeben ist. Diese darf somit für einige Facetten positiv, für andere negativ oder auch nicht signifikant sein. Damit wäre es auch möglich, dass die Fähigkeitsselbstkonzepte zweier sprachlicher Schulfächer, beispielsweise Deutsch und Französisch oder Chinesisch und Englisch, negativ miteinander korreliert sind. Entsprechend zeigte das NMS-Modell im Vergleich der drei hier vorgestellten Strukturmodelle die beste Anpassung an die empirischen Daten (Brunner et al., 2010).

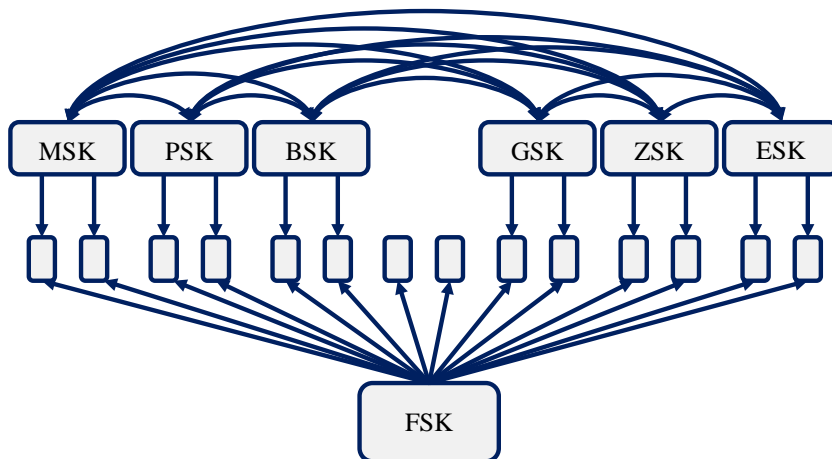


Abbildung 1.2. Genestetes Marsh/Shavelson-Modell (modifiziert nach Brunner et al., 2010); MSK = mathematikspezifisches Fähigkeitsselbstkonzept; PSK = physikspezifisches Fähigkeitsselbstkonzept; BSK = biologiespezifisches Fähigkeitsselbstkonzept; GSK = geschichtsspezifisches Fähigkeitsselbstkonzept; ZSK = zweitsprachliches Fähigkeitsselbstkonzept; ESK = ersprachliches Fähigkeitsselbstkonzept; FSK = Fähigkeitsselbstkonzept.

Inwieweit sich dieses noch recht neue Modell in der Selbstkonzeptforschung bewähren oder durchsetzen wird, bleibt abzuwarten. Für die Struktur des Fähigkeitsselbstkonzeptes ist jedoch aktuell unstrittig, dass dieses Teil einer umfassenderen hierarchischen Ordnung ist und zumindest eine Trennung in ein verbales und ein mathematikspezifisches Fähigkeitsselbstkonzept im Vergleich zur Annahme eines allgemeinen Fähigkeitsselbstkonzeptes notwendig ist.

### 1.2.2 Spezifität und Passung

Auch die Selbstwirksamkeitsüberzeugungen werden als multidimensional strukturiert angenommen (Bandura, 1997; Bong & Skaalvik, 2003; Schunk & Pajares, 2009). Es wird jedoch von einer lockeren hierarchischen Struktur ausgegangen, die durch eine



unterschiedliche Spezifität der Messungen bestimmt wird (Bong & Skaalvik, 2003). Dabei stellt die Spezifität aktuell eine der großen Herausforderungen in der Forschung zu Selbstwirksamkeitsüberzeugungen dar (Schunk & Pajares, 2009). Die Spezifität der Messungen reicht von globalen Messungen der Selbstwirksamkeitsüberzeugungen in allen Lebensbereichen (Reitz, Motti-Stefanidi & Asendorpf, 2014) bis hin zur Selbstwirksamkeitsüberzeugung in Bezug auf ganz konkrete Aufgaben (Niemivirta & Tapola, 2007). Pajares (1996) stellte heraus, dass die Bewertung eigener Kompetenz im Rahmen der Selbstwirksamkeitsüberzeugungen stets domänen- oder aufgabenspezifisch erfolge. Eine globale Erfassung der Selbstwirksamkeitsüberzeugungen führe hingegen zu geringerer prädiktiver Nutzbarkeit in Bezug auf die Umsetzung bestimmter Handlungen. Im Rahmen der pädagogisch-psychologischen Anwendung sei bei globaler schulbezogener Selbstwirksamkeitsüberzeugung die fehlende klar abgegrenzte Aktivität oder Anforderung bei der Erfassung der Erfolgszuversicht in schulischen Leistungssituationen problematisch, da eine Orientierung für die Schulkinder zur Bewertung ihrer Erfolgszuversicht fehle. Daher sei in diesem Kontext die Erfassung auf domänenspezifischer Ebene gängige Praxis. Die domänenspezifische Erfassung werde durch die eingeschränkt spezifischen Ergebniskriterien (Schulnoten bzw. Leistungstests) verursacht, welche schlecht zu einer aufgabenspezifischen Erfassung der Selbstwirksamkeitsüberzeugungen passten (Pajares, 1996). Die sowohl inhaltliche als auch in Bezug auf die Spezifität korrespondierende Erfassung von Selbstwirksamkeitsüberzeugungen und Leistungen trägt jedoch entscheidend zur Güte der Vorhersage zukünftiger Leistung bei (Bandura, 1997; Pajares, 1996). Im Bereich des Lesens dürfte beispielsweise die Selbstwirksamkeitsüberzeugung, einen Zeitungsartikel zu verstehen, eine geringe Vorhersagekraft für die Deutschnote haben. Hin-

gegen sollte die Selbstwirksamkeitsüberzeugung, schwierige Sachtexte im Deutschunterricht zu verstehen, eine hohe Vorhersagekraft für Leistungen in einem Test zum Leseverstehen haben. Die Bevorzugung der domänenspezifischen Erfassung vor aufgabenspezifisch erfassten Selbstwirksamkeitsüberzeugungen dürfte sicherlich auch darin begründet sein, dass die Stichproben im pädagogisch-psychologischen Kontext zumeist aus Schulkindern bestehen und die an diese Schulkinder gestellten Leistungsanforderungen durch Reifung und Beschulung typischerweise mit der Zeit ansteigen. Durch diese stetigen Leistungsverbesserungen sollten auch die Selbstwirksamkeitsüberzeugungen für spezifische Aufgaben relativ schnell ansteigen. Im Rahmen längsschnittlicher Erfassung bietet die domänenspezifische Operationalisierung gegenüber aufgabenspezifischen Operationalisierungen den Vorteil, dass das in der Definition geforderte „bestimmte Niveau“ gerade mit einem gewissen Interpretationsspielraum im Itemstamm bestimmt werden kann. Konkret formulierte aufgabenspezifische Itemstämme sollten besser unter Beachtung curricularer und altersgemäßer Anwendbarkeit verwendet und entsprechend angepasst werden. Durch Anpassungen der Items an den sich durch Reifung und Unterricht üblicherweise positiv entwickelnden Leistungsstand der Schulkinder gäbe man die Invarianz der Messungen auf und gefährdete die Validität der Schlussfolgerungen, die aufgrund des Antwortverhaltens auf die angepassten Items gezogen werden (Ployhart & Vandenberg, 2010; Vandenberg & Lance, 2000; Wainer, 2010). Dadurch erscheinen feinkörnige aufgabenspezifische Messungen von Selbstwirksamkeitsüberzeugungen und Leistung eher für die kurzfristige Vorhersage spezifischer Leistungen geeignet, während für die mittel- bis langfristige längsschnittliche Verfolgung der Entwicklung breiter formulierte domänenspezifische Items empfehlenswert erscheinen. Gegen die noch abstraktere Erfassung

auf einer allgemein schulbezogenen Spezifitätsebene sprechen höhere Korrelationen zwischen Selbstwirksamkeitsüberzeugungen und Leistungen auf domänenspezifischer Ebene im Vergleich zur abstrakteren Ebene (Bong, Cho, Ahn & Kim, 2012; Pajares & Schunk, 2005). Die hier getroffenen Ausführungen zur Spezifität und Passung der Selbstwirksamkeitsüberzeugungen decken sich *cum grano salis* mit Befunden aus der Forschung zu Fähigkeitsselbstkonzepten (Marsh, 1992; Marsh & Craven, 2006).

Neben unterschiedlicher Spezifität der Messung und Passung zum Leistungskriterium wird die Struktur der Selbstwirksamkeitsüberzeugungen durch Unterscheidungen in den Bereichen Niveau (*level* oder *magnitude*), Allgemeingültigkeit (*generality*) und Stärke (*strength*) bestimmt (Bandura, 1997; Maddux, 1995; Pajares, 1996; Zimmerman, 2000). Das Niveau der Selbstwirksamkeitsüberzeugungen wird durch die Schwierigkeit der Anforderungssituation vorgegeben, die beispielsweise beim Lesen auf das Verstehen von Texten mit unterschiedlicher Komplexität bezogen sein kann. Die Allgemeingültigkeit betrifft die Übertragbarkeit der Selbstwirksamkeitsüberzeugungen auf andere Bereiche, wie zum Beispiel das Lesen eines modernen Romans auf ein klassisches Werk. Diese beiden Bereiche werden maßgeblich durch die Operationalisierung der Selbstwirksamkeitsüberzeugungen bestimmt. Schließlich bildet die Stärke der Selbstwirksamkeitsüberzeugungen die Zuversicht ab, mit der die Person glaubt, die gestellte Anforderung erfolgreich bewältigen zu können, und wird entsprechend durch das Verhalten der Person (Itemantwort) bestimmt. Insbesondere im Falle von auftretenden Hindernissen und Schwierigkeiten hing die Stärke der Selbstwirksamkeitsüberzeugungen positiv mit der Persistenz bei der Aufgabenbearbeitung unter Anwendung bereits erlernter Fähigkeiten zusammen (Bandura, 1986).

Zusätzlich erschien eine konzeptuelle Trennung für bekannte und unbekannte Aufgaben und damit zwischen Selbstwirksamkeitsüberzeugungen in Bezug auf Fähigkeitsanwendung (*self-efficacy for performance*) und Selbstwirksamkeitsüberzeugungen in Bezug auf Fähigkeitserwerb (*self-efficacy for learning*) notwendig (Schunk, 1989). Die Selbstwirksamkeitsüberzeugungen in Bezug auf den Fähigkeitserwerb sollten kaum vom Fähigkeitsselbstkonzept unterscheidbar sein, da keine eigenen Erfahrungen mit bereits ausgeführten Handlungen zur Einschätzung der Erfolgszuversicht herangezogen werden können (Pajares, 1996). Stattdessen werden eher soziale Vergleiche genutzt (vgl. Abschnitt 1.2.3). Um beide Konstrukte besser voneinander unterscheidbar zu machen, sollte deshalb die Selbstwirksamkeitsüberzeugung möglichst in Bezug auf bekannte Aufgaben oder übliche Barrieren (z. B. nach einer Erkrankung) und den aus der Bekanntheit heraus erschließbaren Fähigkeitserfordernissen erfasst werden.

Zusammenfassend lässt sich festhalten, dass sowohl Fähigkeitsselbstkonzepte als auch Selbstwirksamkeitsüberzeugungen hierarchisch organisiert sind. Für beide gilt zudem, dass im pädagogisch-psychologischen Kontext die domänenspezifische Erfassung einer globaleren oder noch spezifischeren Erfassung vorzuziehen ist. In Bezug auf die Struktur des Fähigkeitsselbstkonzeptes existiert weiterhin rege Forschungsaktivität, wobei einheitlich zumindest eine Trennung zwischen dem verbalen und dem mathematikspezifischen Fähigkeitsselbstkonzept empfohlen wird. Es existieren jedoch Hinweise, dass Fähigkeitsselbstkonzepte und Leistungen in unterschiedlichen Sprachen nicht oder negativ miteinander zusammenhängen können. Eine getrennte Betrachtung von Personengruppen, deren Erstsprache nicht mit der des interessierenden verbalen Fähigkeitsselbstkonzeptes identisch ist, könnte daher sinnvoll sein. Im Bereich der Selbstwirksam-

keitsüberzeugungen finden strukturelle Diskussionen wie bei Fähigkeitsselbstkonzepten nicht statt. Selbstwirksamkeitsüberzeugungen werden hinsichtlich des Niveaus, der Allgemeingültigkeit und der Stärke unterschieden, wobei dadurch eher operationale als strukturelle Merkmale bestimmt sind.

Im folgenden Abschnitt wird die Entwicklung von Fähigkeitsselbstkonzepten und Selbstwirksamkeitsüberzeugungen beleuchtet. Während bei Fähigkeitsselbstkonzepten überwiegend normative Vergleiche mit sozialem, dimensionalem und temporalem Bezugsrahmen eine Rolle spielen, werden für Selbstwirksamkeitsüberzeugungen eher zielbezogene Leistungswahrnehmungen als relevante Einflussquellen angesehen (Bong & Skaalvik, 2003).

### 1.2.3 Bezugsrahmenmodelle

Bei sozialen Vergleichsprozessen (Festinger, 1954) in der Schule ziehen Schülerinnen und Schüler die eigenen Einschätzungen der Leistungen ihrer Mitschülerinnen und Mitschüler heran, um diese mit ihren Einschätzungen der eigenen Leistung zu vergleichen und in der Folge eine Aktualisierung des eigenen Selbstkonzepts vorzunehmen. In Festingers (1954) Theorie werden für soziale Vergleiche bevorzugt Personen mit ähnlichen Leistungen und Einstellungen herangezogen (*Ähnlichkeitshypothese*), um die Selbstbewertung zu aktualisieren. Ergänzt wurde die Ähnlichkeitshypothese durch den Ansatz, die *relevanten Attribute* potenzieller Vergleichspersonen zu berücksichtigen (Goethals & Darley, 1977). Personen wählen diesem Ansatz zufolge nicht allein Personen mit ähnlichen Leistungen aus, sondern beziehen ebenfalls leistungsrelevante Attribute wie Anstrengung und Erfahrung, aber auch beispielsweise das Alter ein. Zu diesen relevanten Attributen kann insbesondere im sprachlichen Bereich der identische Migrations-

hintergrund zählen, da hier aufgrund der gemeinsamen Erstsprache eine leistungsrelevante Ähnlichkeit angenommen werden kann. Bei Gültigkeit dieser Annahme vergleichen sich Schulkinder mit Migrationshintergrund demnach eher mit ihren Klassenkameradinnen und Klassenkameraden, die ebenfalls einen Migrationshintergrund aufweisen. Da diese Gruppe insgesamt eher leistungsschwach ist (vgl. Abschnitt 1.3.1), sollten die sozialen Vergleiche geringere Kosten für das Selbstkonzept mit sich bringen, was sich trotz schwächerer Leistungen in ähnlichen Ausprägungen des Selbstkonzepts im Vergleich zur Gruppe von Schulkindern ohne Migrationshintergrund ausdrücken könnte. Diese Vermutung wird unterstützt durch entsprechende Befunde zum Niveau des verbalen und mathematikspezifischen Selbstkonzepts von Schulkindern mit und ohne Migrationshintergrund (Shajek et al., 2006; vgl. Abschnitt 1.3.2).

Für die Entwicklung der Selbstwirksamkeitsüberzeugungen wird den sozialen Vergleichen nur eine untergeordnete Rolle zugeschrieben (Bong & Skaalvik, 2003). Durch den Rekurs auf die Erfolgszuversicht bei der Bewältigung bestimmter Aufgaben ist die Leistungsstärke in der Schulklasse vergleichsweise unbedeutend. Dennoch fließen auch soziale Vergleiche als eine von vielen Informationsquellen in die Entwicklung der Selbstwirksamkeitsüberzeugungen ein (Bandura, 1993; 1997; Schunk, 1985). Zum Beispiel ließ Bouffard-Bouchard (1990) in einer experimentellen Studie zwei Gruppen von Studierenden eine ihnen unbekannte Aufgabe bearbeiten, in der in einzelnen Sätzen jeweils ein Wort durch ein Fantasiewort ersetzt war. Die Aufgabe bestand darin, das ursprüngliche Wort herauszufinden. Nach einem Vortest, der die identische Leistungsstärke in beiden Gruppen zeigte, wurde den Mitgliedern der einen Gruppe die Rückmeldung gegeben, besser als die anderen abgeschnitten zu haben. In der anderen Gruppe wurde hin-

gegen mitgeteilt, man habe schlechter abgeschnitten. Nach diesen Rückmeldungen wurden die Selbstwirksamkeitsüberzeugungen der Studierenden sowie die Leistung in einem Nachtest gemessen. Es zeigte sich, dass die Rückmeldung, man sei besser als eine Vergleichsgruppe, zu höherer Selbstwirksamkeitseinschätzung und stärkerer Leistung führte. Dieses Ergebnis stützt die Annahme, dass soziale Vergleiche auch bei der Entwicklung von Selbstwirksamkeitsüberzeugungen eine relevante Einflussgröße darstellen können, zumindest bei unbekanntem Aufgaben. Bei bekannten Herausforderungen wird jedoch bevorzugt die Erfolgsvorsicht aus der eigenen vorhergegangenen Bewältigung ähnlicher Herausforderungen extrapoliert (Bong & Skaalvik, 2003; Schunk, 1985). Diese Bevorzugung eines temporalen Vergleichs gilt natürlich nur dann, wenn solche Bewältigungserfahrungen bereits vorhanden sind, was im Experiment von Bouffard-Bouchard (1990) aufgrund der unbekanntem Aufgaben nicht der Fall war.

Temporale Vergleiche spielen auch bei der Genese von Fähigkeitsselbstkonzepten eine Rolle, auch wenn diese vergleichsweise selten untersucht wurden. Die Theorie temporaler Vergleiche ist konzeptuell aus der Theorie sozialer Vergleiche hervorgegangen. Anders als bei sozialen Vergleichsprozessen werden temporale Vergleiche ange stellt, um die Selbstwahrnehmungen aus unterschiedlichen Zeitpunkten zu integrieren und so zu einem aktualisierten Selbstbild zu gelangen (Albert, 1977). Im Gegensatz zu Einschätzungen des sich eher verschlechternden Gesundheitszustandes älterer Personen (Suls, Marco & Tobin, 1991) sind bei Schulkindern durch einen in aller Regel stetigen Leistungszuwachs während der Beschulung im Rahmen temporaler Vergleiche stets positive Einflüsse auf das Selbstkonzept zu erwarten. Dies mag erklären, warum der temporale Bezugsrahmen in der pädagogisch-psychologischen Selbstkonzeptforschung bezüglich

der Anzahl veröffentlichter Literatur gegenüber der Literatur zu den sozialen und zu den dimensional Vergleichenden vergleichsweise unterrepräsentiert erscheint. Im Gegensatz dazu spielen eigene Leistungen aus der Vergangenheit bei der zielbezogenen Einschätzung der Selbstwirksamkeitsüberzeugungen die wichtigste Rolle (Bandura, 1997; Usher & Pajares, 2008; vgl. Abschnitt 1.2.4).

Explizit Bedeutung für die Selbstkonzeptgenese haben dimensionale Vergleiche, welche im *Internal/External-Frame-of-Reference-Modell* (I/E-Modell; Marsh, 1986; 1990b) beschrieben wurden. Nachdem empirische Befunde zeigten, dass Leistungen in verbalen und mathematischen Schulfächern hoch miteinander korreliert waren, während die entsprechenden Fähigkeitsselbstkonzepte in diesen beiden Fächern nahezu keinen Zusammenhang aufwiesen (Marsh, 1986; Möller, Pohlmann, Köller & Marsh, 2009), wurden mehrere Vergleichsprozesse zur Erklärung angenommen. Bereits thematisiert wurden die sozialen Vergleichsprozesse (*external frame of reference*), bei denen Schulkinder mit einer relativ hohen Leistung zum Beispiel in Mathematik gemessen an der sozialen Bezugsgruppe ein entsprechend hohes mathematikspezifisches Fähigkeitsselbstkonzept entwickelten. Dies wurde ergänzt durch die Annahme, dass hohe Mathematikleistungen über dimensionale Vergleiche (*internal frame of reference*) mit Schulfächern der verbalen Domäne zu einem relativen Absinken des verbalen Fähigkeitsselbstkonzepts und damit zu einem Kontrasteffekt führen (Marsh, 1986). Zwei Schulkinder mit identischen Leistungen in Deutsch und in derselben Schulklasse würden entsprechend über soziale Vergleiche ein identisches verbales Fähigkeitsselbstkonzept ausbilden. Ist jedoch die Mathematikleistung des einen Schulkindes besser als dessen Deutschleistung und beim anderen Schulkind die Mathematikleistung schwächer als dessen Deutschleistung, so verrin-



gert sich über den dimensionalen Aufwärtsvergleich das verbale Fähigkeitsselbstkonzept des Schulkindes mit der stärkeren Mathematikleistung, während das Schulkind mit der vergleichsweise schwächeren Mathematikleistung hinsichtlich des verbalen Fähigkeitsselbstkonzeptes durch den dimensionalen Abwärtsvergleich profitiert. Dieser Kontrasteffekt konnte auch experimentell bestätigt werden (Möller & Köller, 2001). Er war jedoch geringer als der über soziale Vergleiche erzielte positive Einfluss relativ starker Leistungen im selben Fach (Möller et al., 2009). Obwohl dimensionale Vergleiche mit Kosten für das Fähigkeitsselbstkonzept verbunden sein können, scheinen sie insgesamt eher einen vorteilhaften Einfluss auf die Selbstkonzeptentwicklung auszuüben. So waren die Kosten für das Fähigkeitsselbstkonzept bei Aufwärtsvergleichen mit starken Leistungen in einem anderen Fach geringer als der Profit für das Fähigkeitsselbstkonzept durch Abwärtsvergleiche mit schwächeren Leistungen in anderen Fächern (Pohlmann & Möller, 2009). Von den beiden Schulkindern aus dem oben genannten Beispiel muss also das Schulkind mit der starken Mathematikleistung durch den Aufwärtsvergleich aus Sicht der Deutschleistung geringere Einbußen beim verbalen Fähigkeitsselbstkonzept hinnehmen, als das Schulkind mit den schwächeren Mathematikleistungen im verbalen Fähigkeitsselbstkonzept hinzugewinnt.

Für die Genese akademischer Selbstwirksamkeitsüberzeugungen hingegen scheinen dimensionale Vergleiche keine Rolle zu spielen. Dies wird einerseits dadurch deutlich, dass zwischen fachspezifischen Selbstwirksamkeitsüberzeugungen im Gegensatz zu den fachspezifischen Selbstkonzepten Korrelationen von  $r = .50$  nachweisbar waren (Möller et al., 2009). Andererseits wies das I/E-Modell, wenn Selbstwirksamkeitsüberzeugungen als selbstbezogene Kompetenzeinschätzung verwendet wurden, eine

schlechte Passung an empirische Daten auf (Möller et al., 2009). Neben diesen beiden rein datengestützten Argumenten gegen dimensionale Vergleiche bei fachspezifischen Selbstwirksamkeitsüberzeugungen können die positiven Zusammenhänge der fachspezifischen Selbstwirksamkeitsüberzeugungen auch die Annahme stützen, dass die Schulkinder in Bezug auf ihre Fähigkeiten, bestimmte Aufgaben zu bewältigen, vergleichenden Bewertungen eher geringere Beachtung schenken. Wer gute Leistungen in einem Fach zeigt, wird diese mit einer hohen Wahrscheinlichkeit auch in anderen Fächern zeigen können, selbst wenn für ein spezifisches Fach eine besondere Begabung vorhanden ist. Insofern scheinen fachspezifische Selbstwirksamkeitsüberzeugungen im Vergleich zu fachspezifischen Fähigkeitsselbstkonzepten weniger verzerrte Einschätzungen der tatsächlichen eigenen Leistungsfähigkeit abzubilden.

Den drei skizzierten Bezugsrahmen ist gemein, dass die Genese von Fähigkeitsselbstkonzepten und Selbstwirksamkeitsüberzeugungen stets durch die Leistungen beeinflusst wird. Wie in den vorigen Ausführungen angemerkt unterscheidet sich die Rolle der Bezugsrahmen bei der Genese von Selbstwirksamkeitsüberzeugungen im Vergleich zur Genese von Fähigkeitsselbstkonzepten. Zumeist werden zur Entwicklung der Erfolgszuversicht temporale Vergleiche durchgeführt, auch soziale Vergleiche kommen vor. Dimensionale Vergleiche scheinen hingegen keine Rolle zu spielen. Hingegen werden für die Entwicklung der Selbstwirksamkeitsüberzeugungen vier Quellen angenommen. Diese sind Gegenstand des folgenden Abschnitts.

#### **1.2.4 Quellen von Fähigkeitsselbstkonzepten und Selbstwirksamkeitsüberzeugungen**

Für die Entwicklung der Selbstwirksamkeitsüberzeugungen werden vier Quellen als maßgeblich angesehen, deren Informationen Schulkinder bewerten, um ihre akademi-

schen Selbstwirksamkeitsüberzeugungen zu aktualisieren: (a) eigene Bewältigungserfahrungen, (b) stellvertretende Bewältigungserfahrungen, (c) verbale Bekräftigungen und (d) affektive und physiologische Zustände (Bandura, 1986; 1997). Wichtig zur Entfaltung ihrer selbstwirksamkeitsrelevanten Wirkung ist, dass diese Quellen wahrgenommen und entsprechend bewertet werden müssen (Bandura, 1997). Die Wahrnehmung selbstwirksamkeitsförderlicher Informationen und deren entsprechende Bewertung können im Kontext von Schule durch die Lehrkräfte gesteuert werden, indem diese ihren Schulkindern zum Beispiel relevante Hinweisreize wie Leistungsergebnisse des Schulkindes oder leistungsbeeinflussende situationelle Rahmenbedingungen kognitiv salient machen und Attributionen verarbeiten helfen (Schunk, 1985). So können Lehrkräfte auch durch geeignete Aufgabenauswahl ihren Schulkindern einen Lernfortschritt und somit eigene positive Bewältigungserfahrung induzieren und durch erfolgsbasierte Rückmeldungen und substantielle verbale Bekräftigung selbstwirksamkeitsförderliche Quelleninformationen erlebbar machen und bei deren Bewertung unterstützen (Schunk & Zimmerman, 2007).

Die eigene Bewältigungserfahrung als Information über Erfolgserlebnisse oder Misserfolgserlebnisse mit ähnlichen Anforderungssituationen in der Vergangenheit wird als die mächtigste der vier Quellen angenommen (Bandura, 1997; Joët, Usher & Bressoux, 2011; Maddux, 1995; Pajares, 1996; Pajares, Johnson & Usher, 2007; Schunk & Meece, 2006; Schunk & Pajares, 2009; Usher & Pajares, 2008). Eigene Erfolgserlebnisse wirken selbstwirksamkeitserhöhend, während wiederholte Misserfolge die Selbstwirksamkeitsüberzeugung verringern (Bandura, 1986; Schunk & Zimmerman, 2007). So können verbale Unterstützungen und die Beobachtung erfolgreicher Aufgabenbewältigung bei einem relevanten Dritten schon durch ein eigenes Misserfolgserlebnis bereits wir-

kungslos werden oder sogar eine negative Wirkung auf die Selbstwirksamkeitsüberzeugung entfalten (Bandura, 1997). Jedoch fließen auch hier subjektive Wahrnehmungen insbesondere der Umstände des Leistungsergebnisses mit ein, so dass ein Misserfolg, der beispielsweise auf Pech oder unüblich hohe Aufgabenschwierigkeit zurückgeführt wird, möglicherweise keine oder nur sehr abgeschwächte negative Auswirkungen auf die Selbstwirksamkeitsüberzeugung auslösen wird (Bandura, 1997; Joët et al., 2011; Usher & Pajares, 2008). Eigene Bewältigungserfahrungen aus der Vergangenheit sind dem temporalen Vergleichsrahmen ähnlich, wie er auch bei der Genese von Fähigkeitsselbstkonzepten eine Rolle spielt. Diese Ähnlichkeit sollte bei Bewertungen der Selbstwirksamkeitsüberzeugung zur Fähigkeitsanwendung auf Basis bekannter Handlungsalternativen besonders groß sein, da bei unbekanntem Handeln auch soziale Vergleiche als Referenz herangezogen werden (Bouffard-Bouchard, 1990).

Die stellvertretende Erfahrung, also das Beobachten erfolgreicher oder erfolgloser Aufgabenbewältigung bei relevanten Dritten, ist dem sozialen Bezugsrahmen der Genese von Fähigkeitsselbstkonzepten zwar ähnlich, jedoch nicht mit ihm identisch (Bandura, 1982). So werden oft Modelle mit erstrebenswerten Kompetenzen ausgewählt, um von diesen zu lernen (Bandura, 1997). Diese Auswahl deckt sich mit der für die Genese von Fähigkeitsselbstkonzepten relevanten Theorie sozialer Vergleiche (Festinger, 1954). Beim Selbstkonzept hingegen lösten erfolgreiche Dritte in der unmittelbaren Vergleichsgruppe einen Kontrasteffekt aus, der zu einem Absinken des eigenen Selbstkonzeptes führte (Marsh, Trautwein, Lüdtke & Köller, 2008). Die stellvertretende Erfahrung kann durch die Auswahl erfolgreicher Mitschülerinnen und Mitschüler als Modell zur Verbesserung der eigenen Erfolgszuversicht führen („Wenn die anderen aus meiner Klasse das

schaffen, dann sollte ich das auch schaffen!“). Entsprechend zeigte sich nur ein vernachlässigbarer negativer Effekt der mittleren Klassenleistung auf die allgemeine schulbezogene Selbstwirksamkeitsüberzeugung (-.08), während das akademische Selbstkonzept in starken Klassen um 0.20 Standardabweichungen geringer war als in schwachen Klassen (Marsh et al., 2008). Ähnliche Ergebnisse wurden für den naturwissenschaftlichen Bereich gefunden (Jansen et al., 2015), wobei dort die Erfassung der Selbstwirksamkeitsüberzeugung aufgabenspezifisch vorgenommen wurde und damit die Effekte des sozialen Bezugsrahmens im Vergleich zum naturwissenschaftsspezifisch erfassten Fähigkeits-selbstkonzept weiter abgemildert worden sein dürften. Die selbstwirksamkeitserhöhende Wirkung von Modellen, die von Anfang an erfolgreich agieren, ist geringer als die selbstwirksamkeitserhöhende Wirkung durch die Beobachtung von Modellen, die zunächst an einer Aufgabe scheitern und diese nach erhöhter Anstrengung doch erfolgreich bewältigen (Schunk, 1983). Durch solche Beobachtungen können sogar die eigentlich stärker wirkenden direkten Erfahrungen eigenen Scheiterns in ihrem negativen Effekt auf die Selbstwirksamkeitsüberzeugungen abgemildert werden (Bandura, 1997).

Die dritte Quelle von Selbstwirksamkeitsüberzeugungen sind verbale Beeinflussungen durch relevante Dritte (Bandura, 1997). Im Gegensatz zu relevanten Modellen, die vornehmlich aus der Gruppe der Gleichaltrigen entstammen, werden als relevante Dritte solche Personen angesehen, denen vertraut wird. Dadurch können Erwachsene (Lehrkräfte, Eltern) direkten Einfluss auf fachspezifische Selbstwirksamkeitsüberzeugungen von Schulkindern ausüben (Usher & Pajares, 2008). Insbesondere Lehrkräfte können den Einfluss einer selbstwirksamkeitserhöhenden verbalen Beeinflussung durch entsprechend abgestimmte Unterrichtsmaßnahmen verstärken, die zu einem anschließenden ei-

genen Erfolgserleben im Sinne der ersten Quelle führen (Usher & Pajares, 2008). Zudem können die Erfolgseinschätzungen durch das Schulkind insoweit modifiziert werden, dass ein Ergebnis dann als Erfolg angesehen wird, wenn dieses zu einer Verbesserung der eigenen Leistung beitrug. Dies gilt selbst dann, wenn es relevante Modelle gibt, die immer noch bessere Leistungen zeigen (Usher & Pajares, 2008). Hierzu ist es jedoch notwendig, dass die Handlungen der Lehrkraft an das Schulkind adaptiert sind (Helmke, Hosenfeld & Schrader, 2004). Für eine adaptive Unterrichtsgestaltung müsste die Lehrkraft die lern- und leistungsrelevanten Merkmale des Schulkindes gut einschätzen können (Schrader, 2009), hier also dessen Leistung und dessen Selbstwirksamkeitsüberzeugungen. Anderenfalls könnte das Schulkind die verbale Beeinflussung der Lehrkraft fehlinterpretieren oder die Unterrichtsmaßnahme erfolglos bleiben. Beispielsweise könnte gutes Zureden der Lehrkraft bei einer vom Schulkind als leicht empfundenen Aufgabe als Hinweis auf eine schlechte Bewertung der eigenen Leistungen durch die Lehrkraft und damit zur generellen Herabsetzung eigener Selbstwirksamkeitsüberzeugungen führen. Andererseits kann bei der flankierenden Unterrichtsmaßnahme möglicherweise eine für den Erfolg nötige Leistungsfähigkeit seitens der Lehrkraft als bereits vorhanden eingeschätzt worden sein, die tatsächlich jedoch noch förderungswürdig ist, was in einem selbstwirksamkeitsreduzierenden Misserfolgserlebnis enden wird.

Als vierte Quelle zur Entwicklung von Selbstwirksamkeitsüberzeugungen werden die Bewertungen der wahrgenommenen physiologischen Reaktionen des Körpers genutzt (Bandura, 1986; 1997; Usher & Pajares, 2008). Ein Zustand der Erregung oder Aufgeregtheit wie zum Beispiel Schwitzen, Erröten und Zittern wird dabei üblicherweise als Unsicherheit interpretiert und sich selbstwirksamkeitsreduzierend auswirken; Gefühle

von Angst oder Beklemmung bei einer Aufgabenbearbeitung werden mit einer geringen wahrgenommenen Kompetenz und damit einhergehend mit einer geringeren zukünftigen Erfolgszuversicht verknüpft (Bandura, 1986). Auch diese körperlichen Reaktionen sind jedoch Gegenstand der Interpretation und können bei entsprechender bereits vorhandener Erfolgszuversicht auch als selbstwirksamkeitsunschädliches Anzeichen für die Bereitschaft zum Angehen der herausfordernden Aufgabe gedeutet werden (Schwarzer & Jerusalem, 2002).

Die vielfältigen Informationen, die zudem von mehreren Personen stammen können (z. B. Rückmeldungen durch Lehrkräfte, Eltern, Freundinnen und Freunde sowie Modellfunktionen Gleichaltriger) müssen zunächst gewichtet und dann zu einem Gesamtbild integriert werden. Diese komplexe kognitive Aufgabe gelingt insbesondere jüngeren Kindern noch nicht, da diese die unterschiedliche Wichtigkeit der Informationen sowie deren Passung zu bestimmten Anforderungen nur mit Schwierigkeiten zuordnen können und somit deren Selbstwirksamkeitsüberzeugungen oft überhöht und wenig stabil sind (Bandura, 1997). Bei Kindern sind zudem die hierarchische Struktur und die Dimensionalität der Selbstwirksamkeitsüberzeugungen nur gering ausgeprägt, es handelt sich bei ihnen also eher um eine eindimensionale, allgemeine Erfolgszuversicht (Schunk & Meece, 2006). Dadurch haben Kinder den Eindruck, dass sie bei ersten Erfolgen in einer Anforderungssituation auch alle anderen Anforderungssituationen erfolgreich mit den ihnen zur Verfügung stehenden Mitteln bewältigen könnten. Diese Ansichten werden im Verlaufe der Schulzeit unter anderem durch Lehrkraftrückmeldungen zu den Leistungen ausdifferenziert. Mit dem Wechsel von der Grundschule in die Sekundarstufe I ist eine erste Korrektur der Selbstwirksamkeitsüberzeugungen festzustellen (Schunk & Meece,

2006). In der Zeit nach diesem Wechsel der Umwelt sind daher Unterstützungen der Selbstwirksamkeitsentwicklung besonders hilfreich zum Aufbau einer leistungsförderlichen und differenzierten Erfolgszuversicht (Pintrich & Schunk, 1996).

Die Informationen aus den vier Quellen und die notwendige Verarbeitung der Informationen eignen sich zur Beeinflussung durch die Lehrkräfte (Schunk & Meece, 2006; Schunk & Pajares, 2009; Schunk & Zimmerman, 2007). So kann durch eine entsprechende Herausforderung von Schulkindern mit schwierigen, jedoch durch das Schulkind lösbarer Aufgaben die eigene Erfahrung zur Bewältigung von Herausforderungen im Unterrichtsfach erzeugt werden. Der Hinweis auf die positiven Leistungen relevanter Klassenkameradinnen und Klassenkameraden kann eine entsprechende Modellfunktion ausüben. Ebenso können durch entsprechende Rückmeldungen und aufmunternde Bekräftigungen während des Auftretens von Schwierigkeiten die Selbstwirksamkeitsüberzeugungen verbessert werden. Falls ein Schulkind eine Leistung in bestimmten Situationen Unbehagen bereitet, so könnte die Lehrkraft diese Situation zunächst zu vermeiden versuchen. Ein Beispiel hierfür wäre ein Kind mit hervorragender Lesefähigkeit, das aber in der sozialen Klassensituation Angst hat laut vorzulesen oder dabei hohen Stress erlebt und dieses negative Empfinden mit Kosten für die lesespezifische Selbstwirksamkeitsüberzeugung auf alle Lesesituationen generalisiert. Hier könnte ein erfolgreiches Modell, also eine sichere Mitschülerin oder ein sicherer Mitschüler mit hoher Ähnlichkeit zum unsicheren Schulkind, dabei helfen, dieses Unbehagen zu reduzieren, indem dieses Modell erfolgreich und laut einen Text in der Klasse vorliest (Schunk & Zimmerman, 2007).



### 1.2.5 Zwischenfazit – Konzeptuelle Unterschiede

In den vorigen Abschnitten wurden konzeptuelle Unterschiede zwischen Fähigkeitsselbstkonzepten und Selbstwirksamkeitsüberzeugungen betrachtet. Zusammengefasst sind Fähigkeitsselbstkonzepte normative, über mehrere Ebenen aggregierte, hierarchisch strukturierte und vergangenheitsbezogene Kompetenzbewertungen, die aufgrund der im Vergleich zu Selbstwirksamkeitsüberzeugungen stärker ausgeprägten Verallgemeinerbarkeit stabiler sind. Demgegenüber sind Selbstwirksamkeitsüberzeugungen kognitive, auf konkrete Ziele bezogene, zukunftsorientierte und eher kontextspezifische Kompetenzbeschreibungen, die im Vergleich zu Fähigkeitsselbstkonzepten aufgrund der stärkeren Abhängigkeit von der jeweiligen Anforderungssituation leichter veränderbar sind (Bong & Skaalvik, 2003; Schunk & Pajares, 2009).

Die globale Annahme höherer Stabilität der Fähigkeitsselbstkonzepte im Vergleich zu Selbstwirksamkeitsüberzeugungen könnte durch die unterschiedliche Spezifität der Messung (Selbstwirksamkeitsüberzeugungen auf Aufgabenebene und Fähigkeitsselbstkonzepte auf Domänenebene) begründet sein, die in diversen vergleichenden Studien verwendet wurde (Ferla et al., 2009; Jansen et al., 2015; Parker et al., 2014). Die höhere Stabilität der Fähigkeitsselbstkonzepte mag jedoch auch auf einer identischen domänenspezifischen Ebene gelten, bilden doch die für die Genese von Fähigkeitsselbstkonzepten wichtigen sozialen und dimensional Vergleiche sehr stabile Faktoren, solange die soziale Bezugsgruppe und die Vergleichsdimension unverändert bleiben. Selbstwirksamkeitsüberzeugungen entstehen hingegen durch die Interpretation von Informationen aus verschiedenen Quellen, die schneller angepasst werden können. So könnten gezielte Unterrichtsmaßnahmen kurzfristig Erfolgserlebnisse auslösen und dadurch die Selbst-

wirksamkeitsüberzeugungen fördern, ohne dass eine deutliche relative Leistungsverbesserung im Vergleich zur sozialen Bezugsgruppe oder der Vergleichsdimension stattgefunden hätte.

Selbstwirksamkeitsüberzeugungen können besser auf unterschiedliche Kontexte abgestimmt erfasst werden (Pajares & Schunk, 2005). Dadurch ist es möglich situationsbedingte spezifische Stärken und Schwächen zu erkennen. Eine Schülerin, die sehr hohe Selbstwirksamkeitsüberzeugungen in Bezug auf das Verstehen klassischer und moderner Romane berichtet, mag gleichzeitig eine geringe Erfolgszuversicht hinsichtlich ihres Leseverstehens in Bezug auf Gedichte oder naturwissenschaftliche Fachbücher empfinden. Hingegen bietet die Erfassung von Fähigkeitsselbstkonzepten neben der kognitiven Bewertung der Person in Bezug auf Leistungssituationen auch die emotionale Bewertung, die eine Person mit sich selbst verbindet. Leistungsschwierigkeiten könnten abseits rein kognitiver Bewertungen der eigenen Fähigkeiten stattdessen Erkenntnisse über affektive negative Verknüpfungen beispielsweise mit einem Schulfach zutage bringen. Eine gute Leserin, die auch um ihre guten Lesefähigkeiten weiß, mag sich vielleicht nicht gerne mit Lesen beschäftigen, weil sprachliche Leistungen im Vergleich zu Mathematik und Naturwissenschaften gesellschaftlich weniger beachtet erscheinen (beispielsweise wegen der verstärkten Aufmerksamkeit auf die Qualifikation in den sogenannten MINT-Fächern Mathematik, Informatik, Naturwissenschaften und Technik). Dies kann dazu führen, dass, obwohl beide Konstrukte bei gemeinsamer Erfassung hoch ( $r > .50$ ) miteinander korreliert sind ((Bong et al., 2012; Jansen et al., 2015; Marsh, Walker & Debus, 1991), durch die konzeptuellen Unterschiede zwischen den Konstrukten ein Zusammenhang jedoch nicht notwendig ist (Pajares & Schunk, 2005). So kann jemand besonders hohe Er-

folgserwartungen aufgrund eigener erfolgreicher Leseleistungen aufweisen und daher eine hohe lesespezifische Selbstwirksamkeitsüberzeugung berichten. Gleichzeitig kann sich diese Person aber aufgrund stärkerer Leistungen in anderen Bereichen für allgemein wenig kompetent im Lesen halten oder aufgrund negativer Reaktionen der Umwelt auf die guten Leseleistungen eher schlechte Gefühle in Bezug auf die eigenen Leseleistung empfinden und entsprechend ein geringes Leseselbstkonzept berichten.

Hinzu kommt, dass schulische Selbstwirksamkeitsüberzeugungen in unterschiedlichen Fächern oder Domänen oft hoch miteinander korreliert sind, während fachspezifische Fähigkeitsselbstkonzepte zum Teil unkorreliert miteinander sind und durch verbesserte Leistungen in einem anderen Fach über dimensionale Vergleiche in ihrem Niveau abgesenkt werden könnten. Die hohen Korrelationen der Selbstwirksamkeitsüberzeugungen in unterschiedlichen Domänen können dazu führen, dass die Steigerung der Selbstwirksamkeitsüberzeugungen in einer Domäne unter Umständen mit einer Verbesserung der Selbstwirksamkeitsüberzeugungen auch in anderen Domänen einherginge. Jedoch könnte eine durch höhere Selbstwirksamkeitsüberzeugungen möglicherweise verursachte Leistungssteigerung im Gegenzug über dimensionale Vergleichsprozesse auf Kosten des Fähigkeitsselbstkonzeptes in anderen Domänen gehen. Dennoch lohnt es sich auch fachspezifische Fähigkeitsselbstkonzepte im Blick zu behalten, da hier eine gewisse Stabilität von Fördereffekten zu erwarten ist. Zudem scheinen beide Konstrukte mit relativ einfachen Mitteln, wie der Vergabe herausfordernder aber lösbarer Aufgaben, dem Hinweis auf relevante Modelle oder der verbalen Unterstützung im Unterricht gefördert werden zu können (Schunk & Pajares, 2009).

Trotz aller konzeptuellen Unterschiede haben beide Konstrukte gemeinsam, dass diese in unterschiedlicher Spezifität von global bis aufgabenspezifisch erfasst werden können. Je höher die Spezifität und je genauer die inhaltliche Übereinstimmung zwischen selbstbezogener Fähigkeitskognition und Leistung, desto größer ist auch der vorzufindende Zusammenhang zwischen beiden (Bandura, 1997; Marsh, 1990a; Pajares, 1996). Hier scheint es weniger darauf anzukommen, ob man Fähigkeitsselfkonzepte oder Selbstwirksamkeitsüberzeugungen als Prädiktor für Leistungen einsetzt, sondern viel mehr darauf, wie feinkörnig man beides erfasst und wie gut die Passung ist. Das Zusammenspiel zwischen Leistungen und Fähigkeitskognitionen bildet einen Schwerpunkt dieser Dissertation und ist Gegenstand des folgenden Abschnitts.

### **1.3 Schulleistungen und selbstbezogene Fähigkeitskognitionen**

Nachdem im vorigen Abschnitt die beiden für diese Arbeit relevanten selbstbezogenen Fähigkeitskognitionen hinsichtlich ihrer Konzeption eingeführt wurden, beschäftigt sich dieser Abschnitt mit der Fragestellung des Zusammenspiels dieser Kognitionen mit Schulleistungen. Dazu werden im Folgenden zunächst empirische Befunde zum Niveau und der Entwicklung von Schulleistungen und selbstbezogenen Fähigkeitskognitionen in Deutschland betrachtet, wobei ein Fokus auf Schulkinder mit und ohne Migrationshintergrund gelegt wird. Daran anschließend werden drei Annahmen zur kausalen Ordnung von Leistungen und selbstbezogenen Fähigkeitskognitionen kontrastiert und die empirische Befundlage aus den Bildungsvergleichsstudien skizziert, die große und repräsentative Datensätze zur Verfügung hatten. Der Abschnitt schließt mit einem Fazit über die Forschungslage zu diesem Thema.

### 1.3.1 Befunde zu Leistungen in Deutschland

Durch die Ergebnisse großer internationaler Vergleichsstudien wie *Programme for International Student Assessment* (PISA; Klieme et al., 2010; Prenzel, Sälzer, Klieme & Köller, 2013; Reiss, Sälzer, Schiepe-Tiska, Klieme & Köller, 2016), *Internationale Grundschul-Lese-Untersuchung* (IGLU; (Bos, Hornberg, Arnold, Faust & Fried, 2008; Bos, Tarelli, Bremerich-Vos & Schwippert, 2012) und *Trends in International Mathematics and Science Study* (TIMSS; Bos, Bensen et al., 2008; Bos, Wendt, Köller & Selzer, 2012; Wendt et al., 2016) wurden für Deutschland große Leistungsdifferenzen zuungunsten von Kindern mit Migrationshintergrund im Vergleich zu ihren Mitschülerinnen und Mitschülern ohne Migrationshintergrund aufgedeckt. Eine solche Leistungsdifferenz war gerade in der Lesefähigkeit evident. In Deutschland verließen Schulkinder mit Migrationshintergrund die Grundschule mit einem Leseleistungsrückstand gegenüber ihren Mitschülerinnen und Mitschülern ohne Migrationshintergrund, der mehr als einem Lernjahr entsprach (Schwippert, Hornberg & Goy, 2008; Schwippert et al., 2012). In der deutschen PISA-2009-Stichprobe, an der die Lesefähigkeit als Schwerpunktkompetenz erhoben worden war, wiesen Fünfzehnjährige mit Migrationshintergrund einen Rückstand in der Leseleistung von 44 Punkten gegenüber ihren Mitschülerinnen und Mitschülern ohne Migrationshintergrund auf, der ebenfalls größer als der Lernfortschritt in einem Schuljahr war (Stanat et al., 2010). Dies kann, trotz unterschiedlicher Stichproben, ein Indiz dafür sein, dass sich die Leseleistungen von Schulkindern mit und ohne Migrationshintergrund in Deutschland ähnlich entwickeln, die Leistungsdifferenz zwischen diesen Schulkindern also über die Sekundarstufe I hinweg stabil bleibt. Insbesondere für

Schulkinder an Hauptschulen bedeuten die PISA-Befunde jedoch auch, dass dieser Leseleistungsrückstand nicht mehr bis zum Ende der Pflichtschulzeit aufgeholt werden kann.

Durch die kompetente Verwendung der Unterrichtssprache werden die fachspezifischen Inhalte der meisten anderen Schulfächer erlernt (Artelt et al., 2007; Weis et al., 2016). Der Vergleich der Viertklässlerinnen und Viertklässler in Deutschland, die immer oder fast immer zu Hause die Unterrichtssprache Deutsch sprechen, mit denjenigen, die nie oder fast nie zu Hause Deutsch sprechen, zeigte einen Vorsprung der Schulkinder mit häufigem Gebrauch der Unterrichtssprache von 33 Punkten in Mathematik auf, der etwa einem Schuljahr entsprach. Die Leistungsdifferenz zwischen diesen Kindern in den Naturwissenschaften fiel mit 48 Punkten sogar noch deutlicher aus als in Mathematik (Tarelli, Schwippert & Stubbe, 2012). Auch bei den im Rahmen von PISA getesteten Fünfzehnjährigen zeigten sich Leistungsrückstände der Jugendlichen mit Migrationshintergrund im Vergleich zu ihren Klassenkameradinnen und Klassenkameraden ohne Migrationshintergrund. Betrachtet man aus den beiden letzten PISA-Erhebungen die jeweiligen Bereiche, welche als Schwerpunkt erhoben wurden (2012: Mathematik, 2015: Naturwissenschaften), so zeigte sich ein Leistungsungleichgewicht zuungunsten der Fünfzehnjährigen mit Migrationshintergrund in Höhe von 46 Punkten im Schwerpunkt Mathematik, der einem Rückstand von mehr als einem Schuljahr entsprach (Gebhardt, Rauch, Mang, Sälzer & Stanat, 2013). Im naturwissenschaftlichen Schwerpunkt drei Jahre später fiel die Leistungsdifferenz zuungunsten der Fünfzehnjährigen mit Migrationshintergrund mit 61 Punkten noch deutlicher aus als in Mathematik (Rauch et al., 2016).

Längsschnittliche Befunde zur Leistungsentwicklung innerhalb der Sekundarstufe I liegen jedoch leider kaum vor, was insbesondere für den Bereich der sprachlichen

Leistungen gilt. Für die Leistungen in Mathematik und den Naturwissenschaften konnte (Walter, 2006) unter Nutzung der deutschen längsschnittlichen Ergänzungsstichprobe zur PISA-2003-Erhebung für Jugendliche in der zehnten Schulklasse zeigen, dass die Entwicklung der Mathematikleistungen im Verlaufe des Schuljahres weder für eine Zunahme noch für eine Abnahme der Leistungsdisparitäten spricht. In den Naturwissenschaften zeigten sich hingegen sehr uneinheitliche Ergebnisse. So konnten die selbst zugewanderten Jugendlichen (erste Generation) zwar den Rückstand zu den Jugendlichen ohne Migrationshintergrund signifikant aufholen. Andererseits wurde der Leistungsnachteil der Jugendlichen mit Migrationshintergrund, die nicht selbst zugewandert sind (zweite Generation) leicht größer (Walter, 2006). Dieses Ergebnis ist jedoch vor dem Hintergrund zu betrachten, dass durch die Messwiederholung in der zehnten Klasse nur die Jugendlichen berücksichtigt werden konnten, die nicht an Hauptschulen unterrichtet wurden und denen die Versetzung in die zehnte Klasse erfolgreich gelang, die also insgesamt zu den erfolgreicheren Jugendlichen aus der PISA-2003-Stichprobe in Deutschland gehörten. Dadurch waren die Kompetenzunterschiede in der Stichprobe mit Messwiederholung bereits geringer als in der deutschen Gesamtstichprobe. Leider liegen Ergebnisse zur Leistungsentwicklung innerhalb der Sekundarstufe I für den ebenfalls im Rahmen der PISA-Erhebungen erfassten Bereich des Lesens bisher nicht vor.

Diese deutlichen Leistungs Nachteile der Schulkinder mit Migrationshintergrund in den zentralen Kompetenzen für eine erfolgreiche gesellschaftliche Teilhabe sind für Deutschland, das einen Anteil an Schulkindern mit Migrationshintergrund von 32% in der Altersgruppe von 6 bis unter 10 Jahren und von 27% in der Altersgruppe von 10 bis unter 15 Jahren aufweist (Autorengruppe Bildungsberichterstattung, 2016), nur schwer

hinnehmbar. Die Entwicklung der Anteile an Schulkindern mit Migrationshintergrund in Deutschland sowie die aktuellen Migrationsbewegungen weltweit deuten zudem an, dass die Unterstützung von Schulkindern mit Migrationshintergrund auch zukünftig eine Herausforderung des Bildungssystems bleiben wird.

### **1.3.2 Befunde zu Fähigkeitskognitionen in Deutschland**

Obwohl Jugendliche mit Migrationshintergrund deutlich geringere Leistungen sowohl im Lesen (Schwippert et al., 2012; Stanat et al., 2010), als auch in Mathematik (Gebhardt et al., 2013; Tarelli et al., 2012) und den Naturwissenschaften (Rauch et al., 2016; Tarelli et al., 2012) aufwiesen, zeigten diese nur geringe Unterschiede in Bezug auf ihre fach- beziehungsweise domänenspezifischen Fähigkeitsselbstkonzepte (Roebbers et al., 1998; Shajek et al., 2006). Für die vorliegende Arbeit interessanter als das Niveau der Fähigkeitsselbstkonzepte ist hingegen deren längsschnittliche Entwicklung. Hier zeigten sich bei Roebbers et al. (1998) im Verlauf über zweieinhalb Jahre und fünf Messzeitpunkte die fachspezifischen Fähigkeitsselbstkonzepte bei Schulkindern mit und ohne Migrationshintergrund ähnlich stabil. Die Entwicklung der mathematikspezifischen und verbalen Fähigkeitsselbstkonzepte verlief entsprechend in beiden Gruppen vergleichbar.

In der Sekundarstufe I, wo leider nur Befunde zum Niveau der Fähigkeitsselbstkonzepte vorliegen, zeigten sich ähnlich wie in der Grundschule nur geringe Unterschiede im verbalen und mathematikspezifischen Fähigkeitsselbstkonzept zwischen den Jugendlichen mit und ohne Migrationshintergrund (Shajek et al., 2006). Hier ging bei den Jugendlichen mit Migrationshintergrund jeweils im Vergleich zu Jugendlichen ohne Migrationshintergrund eine geringere Ausprägung des verbalen Fähigkeitsselbstkonzeptes mit



einer leicht höheren Ausprägung des mathematikspezifischen Fähigkeitsselbstkonzeptes einher (Shajek et al., 2006).

Ein anderes Bild ergab sich für mathematikspezifische Selbstwirksamkeitsüberzeugungen an Sekundarschulen in Deutschland. Hier wiesen die Jugendlichen mit Migrationshintergrund ein signifikant geringeres Niveau auf als deren Mitschülerinnen und Mitschüler ohne Migrationshintergrund (Stanat & Christensen, 2006). Zugleich erwies sich die Selbstwirksamkeitsüberzeugung als im Vergleich zum mathematikspezifischen Fähigkeitsselbstkonzept stärkerer Prädiktor der Mathematikleistung und klärte in den Gruppen mit und ohne Migrationshintergrund jeweils etwa ein Viertel der Varianz auf (Stanat & Christensen, 2006). Hierbei sollte jedoch erwähnt werden, dass die Selbstwirksamkeitsüberzeugungen im Rahmen von PISA aufgabenspezifisch erfasst wurden und das mathematikspezifische Fähigkeitsselbstkonzept domänenspezifisch, was den Vorhersagevorteil der Selbstwirksamkeitsüberzeugung begünstigte (vgl. Abschnitt 1.2.2). Dennoch bleibt festzustellen, dass mathematikspezifische Selbstwirksamkeitsüberzeugungen und das verbale Fähigkeitsselbstkonzept bei Jugendlichen mit Migrationshintergrund trotz deutlich schwächerer Leistungen nur leicht geringer ausgeprägt zu sein scheinen als bei ihren Mitschülerinnen und Mitschülern ohne Migrationshintergrund und gleichzeitig die Höhe ihres mathematikspezifischen Fähigkeitsselbstkonzeptes sogar leicht über dem Niveau ihrer Mitschülerinnen und Mitschüler ohne Migrationshintergrund liegt. Leider liegen keine vergleichbaren Befunde aus den großen, repräsentativen Bildungsvergleichsstudien für lesespezifische Selbstwirksamkeitsüberzeugungen vor.

Das relativ hohe Niveau des mathematikspezifischen Fähigkeitsselbstkonzeptes könnte ein Resultat dimensionaler Vergleiche sein (Möller & Köller, 2004; vgl. Abschnitt

1.2.3). Bei Jugendlichen mit Migrationshintergrund ist zu erwarten, dass die sprachlichen Leistungen in Deutsch (als Unterrichtssprache, die nicht mit der Erstsprache identisch ist, vgl. Abschnitt 1.2.1) schlechter bewertet werden als die Mathematikleistungen, was über dimensionale Vergleiche zu einer Aufwertung des mathematikspezifischen Fähigkeits-selbstkonzepts auf Kosten des verbalen Fähigkeitsselbstkonzeptes führen sollte. Der Befund, dass das verbale Fähigkeitsselbstkonzept bei Jugendlichen mit Migrationshintergrund schwächer ausgeprägt war als bei Jugendlichen ohne Migrationshintergrund kann als ein Hinweis in diese Richtung aufgefasst werden. Der insgesamt geringe Unterschied in den Niveaus der Fähigkeitsselbstkonzepte im Vergleich zu den Leistungsdifferenzen in den Gruppen der Jugendlichen mit und ohne Migrationshintergrund trotz der unterschiedlichen relativen Position im sozialen Vergleich (Möller & Köller, 2004; vgl. Abschnitt 1.2.3) könnte Ausdruck des Ansatzes zur Verwendung relevanter Attribute (Goethals & Darley, 1977) bei der Auswahl der sozialen Vergleichspersonen sein. Unter der Annahme, dass die Jugendlichen den Migrationshintergrund als relevantes Attribut für die sozialen Vergleiche ansehen, würden Jugendliche mit Migrationshintergrund sich bevorzugt mit ihren Mitschülerinnen und Mitschülern mit Migrationshintergrund vergleichen – einer Gruppe, die insgesamt leistungsschwächer ist. Dadurch nähmen sie eine im Vergleich zur Gesamtklasse bessere relative Position ein, was geringere Kosten des sozialen Vergleichs für die Genese des jeweiligen Fähigkeitsselbstkonzepts mit sich brächte.

Für Selbstwirksamkeitsüberzeugungen sind dimensionale Vergleiche hingegen irrelevant (vgl. Abschnitt 1.2.3). Das Niveau der mathematikspezifischen Selbstwirksamkeitsüberzeugungen bei Jugendlichen mit Migrationshintergrund sollte hier vornehmlich durch die eigenen Mathematikleistungen zustande gekommen sein, weshalb das geringere

Niveau der mathematikspezifischen Selbstwirksamkeitsüberzeugungen im Vergleich zu den Jugendlichen ohne Migrationshintergrund auch ein direktes Resultat des Unterschieds der Mathematikleistungen zwischen beiden Gruppen ist. Der vergleichsweise geringere Niveauunterschied bei den Selbstwirksamkeitsüberzeugungen gegenüber den Mathematikleistungen könnte aus der Nutzung erfolgreicher Modelle in der Klasse resultieren, aber auch aus selbstwirksamkeitserhöhender Beeinflussung durch die Lehrkraft (vgl. Abschnitt 1.2.4).

Das unterschiedliche Niveau der Leistungen im Vergleich zum vergleichbaren Niveau der Selbstwirksamkeitsüberzeugungen deutet auf eine mögliche falsche Kalibrierung hin. Eine gute Kalibrierung auf Aufgabenebene wäre gegeben, wenn sich Personen eine Aufgabe zutrauen und diese dann auch lösen können oder wenn sie sich die Aufgabe nicht zutrauen und diese dann auch nicht lösen können (Schunk & Pajares, 2009). Wird hingegen eine hohe Selbstwirksamkeitsüberzeugung geäußert, obwohl nur schwache Leistungsergebnisse vorliegen, dann spricht dies ebenso für eine schlechte Kalibrierung wie der umgekehrte Fall, dass geringe Selbstwirksamkeitsüberzeugungen an starke Leistungsergebnisse gekoppelt sind. Die besondere Herausforderung der Kalibrierung ist das Verständnis der internen Prozesse der Wahrnehmung und Gewichtung von Informationen aus den einzelnen Quellen zur Entwicklung der Selbstwirksamkeitsüberzeugungen und wie dies an entsprechende Leistungsniveaus angekoppelt ist (Schunk & Pajares, 2009). Eine schlechte Kalibrierung könnte beispielsweise durch fehlende Informationen über die Anforderungen zur erfolgreichen Aufgabebearbeitung begründet sein (Schunk & Pajares, 2009). Im Ergebnis könnte eine übermäßige Selbstwirksamkeitsüberzeugung durch geringe Kenntnisse der Aufgabenanforderungen zu einer mangelhaften Auseinanderset-

zung mit der Aufgabe und dadurch schlechteren Leistungsergebnissen führen (Klassen, 2006). Hier sind daher insbesondere die Lehrkräfte gefordert die Schulkinder mit den entsprechenden Informationen zu versorgen, damit diese ihre Kalibrierung anpassen können, wobei die zusätzliche Herausforderung besteht, dies möglichst ohne Kosten für die Zuversicht und den Optimismus der Schulkinder zu bewältigen (Pajares, 1996).

Leider wurde, trotz der großen Bedeutung der Leseleistung für den Erwerb von Kompetenzen in den meisten anderen Schulfächern, im Rahmen der großen Bildungsvergleichsstudien weder deren Zusammenhang mit Selbstwirksamkeitsüberzeugungen noch deren längsschnittliche Entwicklung untersucht. Die längsschnittliche Überprüfung der Zusammenhänge zwischen selbstbezogenen Fähigkeitskognitionen und Leistungen wäre jedoch notwendig zur Lösung der in der Einleitung aufgeworfenen Frage nach der Kausalität („chicken-and-egg question of causality“; Pajares & Schunk, 2005, S. 112). Die hierbei vertretenen verschiedenen Annahmen zur kausalen Ordnung sind Gegenstand des folgenden Abschnitts.

### **1.3.3 Kausale Ordnung selbstbezogener Fähigkeitskognitionen und Leistungen**

Selbstbezogene Fähigkeitskognitionen sind nicht nur durch Leistungen beeinflusst (vgl. Abschnitt 1.2), sie zeigten sich ebenfalls prädiktiv für spätere Leistungen; höhere selbstbezogene Fähigkeitskognitionen gehen demnach mit einer günstigeren Leistungsentwicklung einher (Huang, 2011; Marsh & Craven, 2006; Valentine et al., 2004). Zur kausalen Ordnung zwischen Leistungen und Fähigkeitsselbstkonzepten wurden jedoch zunächst zwei widerstreitende Annahmen vertreten, welche die mittleren bis hohen Korrelationen insbesondere zwischen domänen- oder fachspezifischen Fähigkeitsselbstkonzepten und den entsprechenden Schulleistungen erklären sollten (Calsyn & Kenny,

1977). Die *Skill-Development-Hypothese* nimmt eine kausale Beziehung von Leistungen auf das entsprechende Fähigkeitsselbstkonzept an, während die *Self-Enhancement-Hypothese* eine kausale Beziehung vom Fähigkeitsselbstkonzept auf die dazugehörige Leistung postuliert. Später wurden beide Annahmen im Modell reziproker Effekte (*Reciprocal Effects Model*; REM; Marsh et al., 1999; Marsh & Craven, 2006; Marsh & Martin, 2011) integriert. Das REM nimmt kausale Beziehungen in beide Richtungen an, also sowohl von Leistungen auf Fähigkeitsselbstkonzepte als auch von Fähigkeitsselbstkonzepten auf Leistungen und bildet einen inhaltlichen Schwerpunkt dieser Dissertation.

Im Abschnitt 1.2.3 wurde die Entwicklung des Fähigkeitsselbstkonzeptes als subjektive Interpretation der Informationen beschrieben, die sich durch die Auseinandersetzung einer Person mit ihrer Umwelt und den daraus resultierenden sozialen, dimensional und temporalen Vergleichen ergeben. Bei relativ hoher Ausprägung beispielsweise der eigenen Leseleistung im Vergleich zur sozialen Bezugsgruppe, sollte die Entwicklung des lesespezifischen Fähigkeitsselbstkonzepts durch soziale Vergleichsprozesse positiv beeinflusst sein. Dieser Einfluss auf das Fähigkeitsselbstkonzept wurde in der Skill-Development-Hypothese aufgegriffen, die gerade einen positiven kausalen Einfluss der Leistung auf das zeitlich nachgeordnete Fähigkeitsselbstkonzept annimmt. Die Self-Enhancement-Hypothese hingegen geht von einer Schlüsselrolle der Fähigkeitsselbstkonzepte aus (Byrne, 1986). Hier kann ein positives Fähigkeitsselbstkonzept mit angenehmeren Gefühlen und positiven Einstellungen gegenüber Schule einhergehen, was wiederum weniger Schulabsentismus und häufigere Bearbeitung von Hausaufgaben sowie aktivere Unterrichtsteilnahme mit sich bringe, wodurch sich spätere Schulleistungen verbesserten (Green et al., 2012). Schülerinnen und Schüler mit höheren Fähigkeitsselbstkonzepten

sollten bei Gültigkeit der Self-Enhancement-Hypothese zeitlich nachgeordnet stärkere Leistungen zeigen.

Das Modell reziproker Effekte greift die beiden Annahmen auf und verbindet diese, indem es kausale Beziehungen sowohl von Fähigkeitsselbstkonzepten auf spätere Leistungen (Self-Enhancement-Hypothese) als auch von Leistungen auf spätere Fähigkeitsselbstkonzepte (Skill-Development-Hypothese) zulässt. Die Klärung der Frage nach der kausalen Ordnung hat nicht nur eine theoretische Bedeutung, insbesondere für die schulische Praxis kann es entscheidend sein, an welcher Stelle im Idealfall die Förderung von Schulkindern angesetzt werden sollte (Marsh et al., 2005; Pajares & Schunk, 2001). Bei Gültigkeit der Skill-Development-Hypothese wäre der Ansatz direkt die Leistung der Schulkinder zu fördern, anstatt auf Fähigkeitsselbstkonzepte zu achten. Im Gegensatz dazu wäre es bei Gültigkeit der Self-Enhancement-Hypothese ratsam, die Förderung der Fähigkeitsselbstkonzepte in die Instrukionsstrategie mit aufzunehmen, um so die Schulleistung zusätzlich verbessern zu können. Gilt hingegen das REM, so könnte durch eine gemeinsame Förderung von beiden in Kombination mit deren gegenseitigen positiven Effekten eine Art Aufwärtsspirale nutzbar werden (Marsh et al., 2005; Pajares & Schunk, 2001).

Die angemessene empirische Überprüfung des REM ist jedoch anspruchsvoll, so dass methodische Richtlinien vorgeschlagen wurden (Marsh et al., 1999; Marsh & Martin, 2011; Pajares & Schunk, 2001). Ein Erfordernis ist die Nutzung von Längsschnittdaten, bei denen beide Konstrukte zu mindestens zwei Messzeitpunkten und möglichst gleichzeitig erhoben wurden. Dies ermöglicht die Kontrolle der Entwicklung beider Konstrukte um die jeweiligen Ausgangswerte. Im REM werden hohe Stabilitäten des Fä-

higkeitsselbstkonzepts und der Leistung angenommen. Zudem werden positive Korrelationen zwischen Leistung und Fähigkeitsselbstkonzept innerhalb jedes Messzeitpunktes erwartet. Die Annahmen des REM werden gestützt, sofern trotz der hohen Stabilität zusätzlich ein positiver Effekt des Fähigkeitsselbstkonzepts auf die spätere Leistung wie auch ein positiver Effekt der Leistung auf das spätere Fähigkeitsselbstkonzept nachweisbar ist. Einzelne positive Effekte von Leistungen aufs Fähigkeitsselbstkonzept (Skill-Development-Hypothese) oder vom Fähigkeitsselbstkonzept auf die Leistung (Self-Enhancement-Hypothese) würden hingegen auf die Gültigkeit einer der anderen Annahmen hindeuten. In Abbildung 1.3 ist die Spezifikation des Strukturmodells des REM über zwei Messzeitpunkte dargestellt.

Eine methodisch empfehlenswerte Prüfung des REM sollte das Fähigkeitsselbstkonzept und die Leistung als latente Faktoren mit mindestens drei Indikatoren zur Kontrolle der Messfehler umfassen (Marsh et al., 1999). Die Indikatoren sollten dabei zu allen Zeitpunkten möglichst unverändert eingesetzt werden, wobei eine Korrelation der Residuen zu beachten ist. Um diese anspruchsvollen Methoden umsetzen zu können, ist es erforderlich eine ausreichend große und diverse Stichprobe zur Verfügung zu haben, die zudem die Generalisierbarkeit der Befunde erleichtere (siehe für weitere Empfehlungen und eine ausführlichere Beschreibung Marsh et al., 1999).

Die Umsetzung dieser Vorgaben und Empfehlungen wurde erst durch jüngere methodische Entwicklungen ermöglicht. So ist für die Prüfung des REM entsprechend der von (Marsh et al., 1999) formulierten Vorgaben die Verwendung von Strukturgleichungsmodellen mit autoregressiven und kreuzverzögerten Pfaden (Selig & Little, 2012) erforderlich. Dennoch bleibt nach wie vor die Erhebung von Längsschnittdaten an großen Stichproben eine Herausforderung, so dass aktuell noch nicht viele Datensätze vorliegen, mit denen die Empfehlungen und Vorgaben umgesetzt werden können.

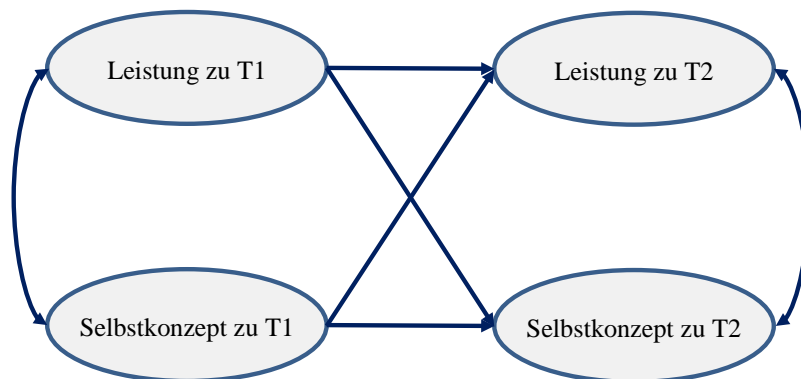


Abbildung 1.3. Modell reziproker Effekte des Selbstkonzepts auf spätere Leistung und der Leistung auf das spätere Selbstkonzept.

Während in Deutschland zumindest die mathematischen und naturwissenschaftlichen Leistungen Jugendlicher mit und ohne Migrationshintergrund eine ähnliche Entwicklung zu nehmen scheinen (Walter, 2006), ist über die Entwicklung sprachlicher Leistungen sowie der Fähigkeitsselbstkonzepte und Selbstwirksamkeitsüberzeugungen in der Sekundarstufe I wenig bekannt. Denkt man an die deutlichen Leistungsdifferenzen und die nur geringen Unterschiede in den Fähigkeitsselbstkonzepten der Gruppen von Schul-



kindern mit und ohne Migrationshintergrund, so ergibt sich ein gewisser Widerspruch, da die Fähigkeitsselbstkonzepte in der Gruppe mit Migrationshintergrund entsprechend des relativen Leistungsstandes über soziale Vergleiche ebenfalls ein geringeres Niveau als bei Jugendlichen ohne Migrationshintergrund aufweisen sollten. Gleichwohl scheinen soziale und dimensionale Vergleichsprozesse in unterschiedlichen Ländern, also in kulturell unterschiedlichen Gruppen, identisch abzulaufen (Marsh & Hau, 2003; Marsh & Hau, 2004). Überträgt man diese Befunde auf Jugendliche innerhalb eines Landes, so sollten sich vergleichbare Prozesse auch zwischen Jugendlichen mit und ohne Migrationshintergrund in Deutschland abspielen. Inwieweit sich jedoch ein differenzielles Zusammenhangsmuster zwischen Leistungen und Fähigkeitsselbstkonzepten in unterschiedlichen ethnischen Gruppen innerhalb eines Landes ergeben könnte, ist weiterhin ein Desiderat der Forschung (Valentine et al., 2004). Dies gilt insbesondere für den sprachlichen Bereich, in dem es Hinweise auf differenzielle Fähigkeitsselbstkonzepte für die Erstsprache und Fremdsprachen gibt (Marsh, Kong et al., 2001; Marsh & Yeung, 2001; vgl. Abschnitt 1.2.1).

#### **1.3.4 Empirische Befunde zur kausalen Ordnung**

Die Frage nach der kausalen Ordnung hat eine rege Forschungsaktivität für das Zusammenspiel zwischen Leistungen und Fähigkeitsselbstkonzepten angeregt. So können reziproke Zusammenhänge zwischen Fähigkeitsselbstkonzepten und Leistungen sowohl auf einem allgemeinen Spezifitätsniveau (Guay et al., 2003; Marsh et al., 1999; Preckel, Niepel, Schneider & Brunner, 2013) als auch für Mathematikleistungen und mathematikspezifische Fähigkeitsselbstkonzepte (Marsh et al., 2005; Pinxten, Marsh, De Fraine, Van Den Noortgate & Van Damme, 2014) mittlerweile als gut belegt gelten. Seltener unter-

sucht wurden Leseleistungen und lesespezifische Fähigkeitsselbstkonzepte (Retelsdorf et al., 2014).

Betrachtet man die vorliegenden Befunde zum reziproken Zusammenspiel zwischen fach- und domänenspezifischen Leistungen und Fähigkeitsselbstkonzepten für Schülerinnen und Schüler an Sekundarschulen in Deutschland, so sind insbesondere die Arbeiten von Marsh et al. (2005) im Bereich Mathematik und von Retelsdorf et al. (2014) im Bereich Lesen interessant. Marsh et al. konnten mit zwei Messzeitpunkten innerhalb des siebten Schuljahres reziproke Effekte zwischen Mathematiktestleistungen und dem mathematikspezifischen Fähigkeitsselbstkonzept finden, die jeweils sehr gering ausfielen ( $< .10$ ). Der Effekt im Sinne der Self-Enhancement-Hypothese, also vom Mathematikselbstkonzept auf die spätere Mathematiktestleistung, war dabei etwas größer als der Effekt im Sinne der Skill-Development-Hypothese. In einer vier etwa gleichabständige Messzeitpunkte umfassenden Studie konnten Retelsdorf et al. demgegenüber nur für den Zeitraum zwischen der ersten Erhebung in der fünften Klasse und der zweiten Erhebung etwa 18 Monate später reziproke Effekte zwischen dem lesespezifischen Fähigkeitsselbstkonzept und Lesetestleistungen nachweisen. Innerhalb der nachfolgenden Intervalle wurden nur noch Pfadkoeffizienten gefunden, die Unterstützung für die Skill-Development-Hypothese boten. Damit konnten die von Marsh et al. für den Bereich Mathematik gefundenen reziproken Effekte bei Siebtklässlerinnen und Siebtklässlern im Bereich Lesen nicht bestätigt werden. Die Existenz reziproker Zusammenhänge kann also nicht ohne weiteres über alle Schulfächer oder Leistungsdomänen hinweg generalisiert werden.

In ihrer Metaanalyse zum Einfluss selbstbezogener Fähigkeitskognitionen auf Leistungen unter Kontrolle der Vorleistungen haben Valentine et al. (2004) den Einfluss möglicher Moderatoren untersucht. Dabei zeigte sich weder ein Einfluss des Geschlechts noch des Alters. Für die Beurteilung einer möglichen Moderation des Zusammenhangs durch den ethnischen Hintergrund lagen und liegen zu wenig Studien vor. So haben Valentine et al. in ihrer Metaanalyse amerikanischer Studien den Anteil an weißen Schülerinnen und Schülern als möglichen Moderator untersucht und keinen Einfluss auf den Zusammenhang zwischen selbstbezogenen Fähigkeitskognitionen und Leistungen unter Kontrolle der Vorleistungen aufdecken können.

Obwohl das REM im Kontext von Fähigkeitsselbstkonzepten entwickelt und ausführlich beforscht wurde, sollte die Übertragung auf andere selbstbezogene Fähigkeitskognitionen und hier insbesondere die Selbstwirksamkeitsüberzeugungen theoretisch leicht zu begründen und unmittelbar plausibel sein. Schon bei der Konzeption der Selbstwirksamkeitsüberzeugungen im Rahmen der sozial kognitiven Theorie wurde angenommen, dass ein reziproker Einfluss zwischen Person (Kognitionen, Affekte, physiologische Reaktionen), menschlichem Verhalten (Leistung) und der Umwelt existiert (Bandura, 1978; 1986; 1997). In Abschnitt 1.2.4 wurde der Einfluss der vier Quellen auf die Aktualisierung der akademischen Selbstwirksamkeitsüberzeugungen dargestellt. Diese Quellen sind ebenfalls Teil der Person (physiologische oder affektive Reaktionen während der Leistungssituation), der Informationen aus der Umwelt (stellvertretende Bewältigungserfahrungen oder verbale Überzeugungen relevanter Dritter, Leistungsrückmeldungen auf gezeigtes eigenes Verhalten) oder des eigenen Verhaltens (selbstbewertete eigene Leistungserfahrung). Gemeinsam mit der Annahme, dass die aktualisierten Selbst-

wirksamkeitsüberzeugungen wiederum die Auswahl und Bearbeitung von Aufgaben und damit spätere Leistungshandlungen beeinflussen (Bandura, 1997; Schunk, 1989) sind die reziproken Effekte des REM bereits Teil der Selbstwirksamkeitstheorie.

Die empirische Forschung blieb bis jetzt jedoch sparsam mit einer Überprüfung dieser Theorie, was nicht zuletzt am Mangel an Datensätzen liegen dürfte, die den Empfehlungen von Marsh et al. (1999) entsprechen. Einen ersten Ansatz lieferten Mittag et al. (2002) über vier Messzeitpunkte im Abstand von einem Jahr. Die auf einer fachübergreifenden Ebene ermittelten Ergebnisse konnten die theoretisch anzunehmenden reziproken Effekte nicht stützen, belegten jedoch Effekte im Sinne der Self-Enhancement-Hypothese von Klasse 7 zu Klasse 8 und von Klasse 9 zu Klasse 10. Höhere schulische Selbstwirksamkeitsüberzeugungen gingen also in diesen Zeiträumen mit höheren Schulleistungen einher. Ebenfalls auf einer fachunspezifischen Ebene konnten in Korea reziproke Effekte zwischen Selbstwirksamkeitsüberzeugungen und Schulleistungen innerhalb jeweils eines Jahres von Klasse 8 bis Klasse 12 aufgedeckt werden, wobei die Pfadkoeffizienten einen stärkeren Effekt im Sinne der Skill-Development-Hypothese andeuteten (Hwang et al., 2016). Eine Berücksichtigung der für pädagogisch-psychologische Forschung sinnvollen fach- beziehungsweise domänenspezifischen Ebene (Pajares, 1996; vgl. Abschnitt 1.2.2) blieb in diesen Studien allerdings aus.

Befunde zu reziproken Effekten zwischen Leistungen und Selbstwirksamkeitsüberzeugungen, die den methodischen Empfehlungen zur Überprüfung des REM entsprechen, sind rar und, wenn überhaupt, nur für den Bereich Mathematik zu finden. Unter der Annahme, dass Fünfzehnjährige realistische Annahmen ihrer Selbstwirksamkeitsüberzeugungen treffen, modellierten Williams und Williams (2010) das REM mit Quer-

schnittdaten aus der PISA-2003-Erhebung (Organisation for Economic Co-operation and Development, 2005). Sie fanden dabei reziproke Effekte zwischen Leistungen und aufgabenspezifisch erfassten Selbstwirksamkeitsüberzeugungen in 24 der 33 im Datensatz enthaltenen Länder. Allerdings räumten Williams und Williams ein, dass ein Desiderat der Forschung die Replikation dieser Befunde unter der Nutzung längsschnittlicher Daten sei.

Mit einer großen, repräsentativen Stichprobe aus Finnland konnten Hannula et al. (2014) diesem Desiderat nachkommen und reziproke Effekte zwischen Leistungen und Selbstwirksamkeitsüberzeugungen in Mathematik über Zeiträume von mindestens drei Jahren nachweisen. In der Stichprobe mit drei Messzeitpunkten vom Beginn der dritten bis zum Ende der neunten Klasse zeigte sich zudem eine Verschiebung des Zusammenspiels von einem anfangs eher leistungsdominierten reziproken Verhältnis hin zu einem reziproken Zusammenspiel ohne einen favorisierten Pfad. Dieses Ergebnis kontrastiert damit die Ergebnisse auf fachunspezifischer Ebene von Hwang et al. (2016), die bei jährlichen Erhebungen von der achten bis zur zwölften Klasse noch eine Dominanz des Pfades von Leistung auf Selbstwirksamkeitsüberzeugungen herausarbeiteten.

Zusammenfassend liegen nunmehr erste Befunde vor, die eine Bestätigung der theoretisch begründeten reziproken Effekte zwischen Selbstwirksamkeitsüberzeugungen und Leistungen liefern. Teilweise beruhen diese Ergebnisse jedoch auf der Nutzung von Querschnittdaten oder wurden fachunspezifisch erhoben und verstoßen damit gegen die methodischen Empfehlungen von Marsh et al. (1999) oder gegen die empfohlene Spezifität der Messung (Pajares, 1996). Längsschnittliche Befunde zum Zusammenspiel zwischen mathematikspezifischen Selbstwirksamkeitsüberzeugungen und Mathematikleistungen zeigten bei Messintervallen von mindestens drei Jahren einen Wechsel vom eher

leistungsdominierten Effekt hin zu mehr Reziprozität (Hannula et al., 2014), was sich jedoch in fachunspezifischen Befunden mit kürzeren Messintervallen nicht wiederfinden ließ (Hwang et al., 2016). Weitere Arbeiten zur Überprüfung des REM im Bereich Mathematik mit kürzeren Messintervallen stehen daher noch aus. Ebenso deuten die Unterschiede in den Befunden zwischen der mathematikspezifischen und der fachunspezifischen Ebene an, dass die Gültigkeit des REM nicht einfach über Fächer hinweg generalisiert werden kann. Weitere fach- oder domänenspezifische Arbeiten sollten die Befundlage zum REM erweitern. Dies gilt auch für die Überprüfung der kulturellen Übertragbarkeit des REM unter der Nutzung von Längsschnittdaten.

### **1.3.5 Kulturelle Übertragbarkeit reziproker Effekte**

Insgesamt lagen und liegen zur Untersuchung eines moderierenden Einflusses auf das Zusammenspiel zwischen selbstbezogenen Fähigkeitskognitionen und Schulleistungen nicht genügend Befunde vor (Valentine et al., 2004). Für die Struktur und Genese von Fähigkeitsselbstkonzepten liegen zwar diverse Studien aus unterschiedlichen Kulturen und Kontinenten vor (Marsh, 1990a; Marsh & Craven, 2006; Marsh & Hau, 2003, 2004; Marsh, Köller & Baumert, 2001; Marsh, Kong et al., 2001), aber die Forschung zu Selbstwirksamkeitsüberzeugungen fand bisher meistens anhand von Stichproben aus amerikanischen Schulen statt (Schunk & Pajares, 2009). Mit Blick auf die Genese von Fähigkeitsselbstkonzepten kann immerhin die kulturelle Übertragbarkeit der Annahmen zu den Vergleichsrahmen (siehe Abschnitt 1.2.3) als empirisch gestützt gelten (Marsh & Hau, 2003; Marsh & Hau, 2004; Marsh & Köller, 2004; Marsh, Köller et al., 2001). Obwohl sich das REM im Bereich Mathematik in der Sekundarstufe I in Deutschland robust über die Zeit, unterschiedliche Schulformen und Leistungskriterien (Leistungstests bzw.

Schulnoten) zeigte (Arens, Marsh, Pekrun et al., 2016), liegen entsprechende Untersuchungen zur kulturellen Übertragbarkeit des REM besonders im sprachlichen Bereich hingegen kaum vor. Dies dürfte auch an einem Mangel an entsprechenden Längsschnittdaten liegen. So bleibt für Hinweise auf die kulturelle Übertragbarkeit des REM nur die Betrachtung einzelner Studien aus unterschiedlichen Ländern. Hierfür liegt neben Studien aus dem westlichen Kulturraum (Marsh et al., 1999; Marsh & O'Mara, 2008; Marsh et al., 2005) auch eine Studie mit Schulkindern in Hongkong vor (Marsh, Hau & Kong, 2002), mit der die methodischen Empfehlungen zur Untersuchung des REM erfüllt wurden. Hier konnten reziproke Effekte zwischen dem Fähigkeitsselbstkonzept und Schulleistungen gefunden werden, allerdings auf einer fachunspezifischen Ebene (Marsh et al., 2002). Die Evaluation der kulturellen Übertragbarkeit auf fach- oder domänenspezifischer Ebene ist damit noch immer eine offene Forschungsfrage.

Für Selbstwirksamkeitsüberzeugungen deuten vorliegende Befunde an, dass trotz eines unterschiedlichen Niveaus der Selbstwirksamkeitsüberzeugungen in unterschiedlichen Kulturen die Vorhersagekraft für Leistungen vergleichbar ist (Klassen, 2004). Der querschnittliche Zusammenhang zwischen Selbstwirksamkeitsüberzeugungen und Leistungen sollte demnach kulturell invariant sein. In Bezug auf die Entwicklung akademischer Selbstwirksamkeitsüberzeugungen in Gruppen mit und ohne Migrationshintergrund gibt es in der rezenten Literatur Befunde aus den Niederlanden, die sich auf eine Stichprobe aus Grundschulkindern beziehen (Hornstra, van der Veen, Peetsma & Volman, 2013). Hier zeigten sich ebenfalls keine Unterschiede in der Entwicklung.

Insgesamt bleibt festzustellen, dass Selbstwirksamkeitsüberzeugungen trotz eines differenten Niveaus in unterschiedlichen Kulturen eine vergleichbare Entwicklung zu

nehmen scheinen. Dies gilt auch für den Zusammenhang mit Leistungen. Die Forschungslage zur kausalen Ordnung zwischen selbstbezogenen Kognitionen und Leistungen unter Berücksichtigung kultureller Hintergründe ist jedoch noch sehr dünn und es liegen insbesondere kaum Studien vor, die einen moderierenden Einfluss anhand einer Stichprobe evaluierten.

### **1.3.6 Zwischenfazit – Kausale Ordnung**

Die Frage der kausalen Ordnung zwischen Fähigkeitsselbstkonzepten und Leistungen sowie zwischen Selbstwirksamkeitsüberzeugungen und Leistungen geht mit theoretisch wie praktisch interessanten Implikationen einher. Aus theoretischer Perspektive ist es interessant, die bereits bei der Einführung der Selbstwirksamkeitsüberzeugungen formulierte Reziprozität (Bandura, 1977, 1978) zwischen Verhalten (Leistungen) und Kognitionen (Selbstwirksamkeitsüberzeugungen) auch empirisch unter Verwendung angemessener Daten und Methoden zu bestätigen. Obwohl bereits eine Reihe von Arbeiten vorliegen, die ein reziprokes Zusammenspiel zwischen Leistungen und Fähigkeitsselbstkonzepten unterstützen, ist die Befundlage im verbalen Bereich überraschend gering. Die wenigen Befunde deuten jedoch an, dass die kausale Ordnung im Lesen eher leistungsdominiert ist (Retelsdorf et al., 2014), wohingegen das mathematikspezifische Fähigkeitsselbstkonzept einen stärkeren Einfluss auf die Mathematikleistungen ausübte (Marsh et al., 2005).

Aus praktischer Perspektive bringen Arbeiten zur kausalen Ordnung zwischen selbstbezogenen Fähigkeitenkognitionen und Leistungen wichtige Implikationen für die Gestaltung von Unterricht mit sich. Sollte die kausale Ordnung im Sinne der Skill-Development-Hypothese bestimmt sein und somit allein die Leistungen einen Effekt auf



spätere selbstbezogene Fähigkeitskognitionen zeigen, ist im Rahmen des Unterrichts primär auf die Förderung der Leistungen zu achten (Pajares & Schunk, 2001). Zeigen sich hingegen allein Effekte im Sinne der Self-Enhancement-Hypothese, also ein positiver Effekt selbstbezogener Fähigkeitskognitionen auf die zeitlich nachgeordnete Leistung, so erscheint die zusätzliche direkte Förderung der selbstbezogenen Fähigkeitskognitionen besonders ratsam. Ergibt sich Evidenz für reziproke Effekte zwischen Leistungen und selbstbezogenen Fähigkeitskognitionen, so sollte dies in eine fortwährende unterrichtliche Förderung von beidem zur Nutzung einer sich gegenseitig verstärkenden Aufwärtsbewegung hin zu gesteigertem Erfolg und Wohlbefinden der Schulkinder münden.

Betrachtet man die Befunde aus großen, repräsentativen Stichproben in Deutschland, so fallen die deutlichen Leistungsunterschiede zuungunsten der Schulkinder mit Migrationshintergrund auf. Diese sind zum Teil so gravierend, dass die spätere erfolgreiche Teilhabe am gesellschaftlichen und beruflichen Leben mit Schwierigkeiten verbunden sein dürfte. Damit erscheint gerade für Schulkinder und Jugendliche mit Migrationshintergrund eine Leistungsförderung nötig zu sein. Sieht man sich die entsprechenden Befunde zum Niveau der selbstbezogenen Fähigkeitskognitionen an, so fällt die vergleichbare Ausprägung bei Kindern und Jugendlichen mit und ohne Migrationshintergrund auf, steht diese doch in einem gewissen Widerspruch zu den Leistungen. Dies deutet auf eine mangelhafte Kalibrierung der selbstbezogenen Kognitionen hin, die von den Lehrkräften durch entsprechende Informationen über Aufgabenanforderungen angegangen werden sollte, um eine Überzüversicht und damit einhergehende schwache Leistungen abzubauen (Pajares, 1996; Schunk & Pajares, 2009). In kulturell vergleichenden Studien zeigte sich jedoch, dass trotz unterschiedlichem Niveau die Zusammenhänge mit

Leistungen vergleichbar ausfallen (Klassen, 2004; Marsh et al., 2002). Allerdings steht hierzu noch die Übertragbarkeit solcher Befunde auf die fach- oder domänenspezifische Ebene aus.

#### 1.4 Urteilsgenauigkeit von Lehrkräften

Gegenstand der vorangehenden Abschnitte war die Beschreibung von Fähigkeitsselfkonzepten und Selbstwirksamkeitsüberzeugungen als selbstbezogenen Fähigkeitskognitionen von Schulkindern und deren reziprokem Zusammenspiel mit Schulleistungen. Demnach sind höhere Leistungen förderlich für die Entwicklung höherer selbstbezogener Fähigkeitskognitionen, während gleichzeitig eine höhere Ausprägung selbstbezogener Fähigkeitskognitionen mit einer stärkeren Leistungsentwicklung einhergeht (Pajares & Schunk, 2001). Dieser gegenseitige positive Effekt sollte durch die Lehrkräfte im Rahmen des Unterrichts über die entsprechende Förderung von selbstbezogenen Fähigkeitskognitionen und Leistungen in einer sich selbst verstärkenden Aufwärtsbewegung nutzbar gemacht werden (Marsh et al., 2005; Pajares & Schunk, 2001). Um jedoch Leistungen und selbstbezogene Fähigkeitskognitionen optimal fördern zu können, sollte Lehrkräften eine möglichst genaue Einschätzung von beidem gelingen. Die Urteilsgenauigkeit von Lehrkräften ist ein Teil ihrer *diagnostischen Kompetenz* und Gegenstand dieses Abschnitts.

Den Kern der sogenannten diagnostischen Kompetenz bildet die Fähigkeit von Lehrkräften, lern- und leistungsrelevante Merkmale von Schulkindern möglichst zutreffend einzuschätzen (Schrader, 2009). Lern- und leistungsrelevante Schulkindmerkmale sind dabei nicht nur ihre Leistungen, sondern auch ihre kognitive Grundfähigkeit, sowie affektive und motivationale Voraussetzungen (Spinath, 2005). Differenzen zwischen von

der Lehrkraft erwarteter und vom Schulkind gezeigter Leistung können dann auf Basis korrekter Einschätzungen dieser vorauslaufenden Bedingungen für die gezeigten Leistungen angegangen werden (Spinath, 2005). Beispielsweise könnte ein Schulkind eine sehr geringe Selbstwirksamkeitserwartung in Bezug auf lautes Vorlesen vor der Klasse haben und dadurch, trotz grundsätzlich ausreichender Lesefähigkeit an dieser Aufgabe scheitern. Hier wäre dann der Ansatzpunkt nicht in der Verbesserung der Leseleistungen sondern in der Förderung der entsprechenden Selbstwirksamkeitsüberzeugung zu suchen. Ebenso könnten auch überhöhte Selbstwirksamkeitsüberzeugungen des Schulkinds zu geringer Leistung führen (Bandura, 1997; Schunk & Pajares, 2009).

Die Bedeutung der Beurteilung selbstbezogener Fähigkeitskognitionen von Schulkindern durch deren Lehrkräfte wurde bereits erkannt und in diversen Studien umgesetzt (Huelmann et al., 2014; Praetorius, Berner, Zeinz, Scheunpflug & Dresel, 2013; Praetorius, Greb, Lipowsky & Gollwitzer, 2010; Praetorius et al., 2011; Spinath, 2005). Nachfolgend werden zunächst die Komponenten der Urteilsgenauigkeit eingeführt, sowie empirische Befunde zur Urteilsgenauigkeit zusammengefasst. Anschließend werden mögliche Einflüsse auf die Urteilsgenauigkeit benannt. Nach einer Zusammenfassung schließt der Abschnitt mit einem Fazit.

#### **1.4.1 Messung der Urteilsgenauigkeit von Lehrkräften**

Die Urteilsgenauigkeit von Lehrkräften wird klassisch durch drei Komponenten beschrieben, die *Niveauelemente*, die *Differenzierungskomponente* und die *Rangkomponente* (Schrader & Helmke, 1987). Jede für sich bildet einen Teil der Urteilsgenauigkeit ab, die gesamte Betrachtung aller drei Komponenten liefert ein komplettes Bild über die Genauigkeit der Urteile von Lehrkräften. In jüngerer Zeit wurden vermehrt moderne

Ansätze zur Messung der Urteilsgenauigkeit unter Verwendung mehrbenenanalytischer Methoden verwendet, wobei deren Ergebnisse analog zu klassischen Komponenten der Urteilsgenauigkeit interpretiert werden können (Leucht, Tiffin-Richards, Vock, Pant & Köller, 2012; Ready & Wright, 2011).

Die Niveauekomponente der Urteilsgenauigkeit beschreibt die Differenz, welche sich zwischen der mittleren Ausprägung des Zielkonstrukts in der Gruppe der Schulkinder einer Klasse und der mittleren Einschätzung der Lehrkräfte in Bezug auf dieses Zielkonstrukt ergibt. Das optimale Ergebnis wäre entsprechend, wenn keine Differenz vorhanden ist, die Niveauekomponente also Null beträgt. Ein Ergebnis im positiven Bereich indiziert eine Überschätzung der mittleren Ausprägung des Konstruktes durch die Lehrkraft und ein Ergebnis im negativen Bereich demzufolge die durchschnittliche Unterschätzung der Ausprägung des Merkmales bei den Schulkindern durch die Lehrkräfte. Die Niveauekomponente kann auch individuell für einzelne Schulkinder ermittelt werden.

Bei der Differenzierungskomponente steht die Einschätzung der Varianz des Zielkonstrukts in der Schulklasse im Fokus. Sie ergibt sich als Quotient der Streuung der Lehrkräfteeinschätzungen geteilt durch die tatsächliche Streuung in der Schulklasse. Bei akkurater Einschätzung der Varianz des Konstrukts in der Schulklasse durch die Lehrkraft ergibt sich eine Differenzierungskomponente von Eins. Werte größer als Eins kennzeichnen eine Überschätzung durch die Lehrkraft, während Werte kleiner als Eins eine Unterschätzung der Varianz des Konstrukts in der Schulklasse indizieren. Hier zeigt sich bereits, dass die Unterschätzung nur auf dem Wertebereich zwischen 0 und 1 abgebildet wird, während die Überschätzung Werte im Bereich von 1 und unendlich annehmen kann. Eine absolute Abweichung um 0.5 vom Optimum 1 kennzeichnet also bei einer

Differenzierungskomponente von 0.5 eine stärkere Unterschätzung als eine Überschätzung bei einer Differenzierungskomponente von 1.5 gegeben wäre. Stattdessen müsste für einen Vergleich die Multiplikation der beiden Kehrwerte verwendet werden (Baumert et al., 2003), oder um die Entsprechung in die jeweils andere Richtung zu ermitteln einfach der Kehrwert gebildet werden. Bei einer Differenzierungskomponente von 0.5, also  $1/2$ , wäre die entsprechende Überschätzung damit  $2/1$ , also 2. Die Differenzierungskomponente ist erst dann definiert, wenn eine Varianz in der Merkmalsverteilung vorliegt, wozu mindestens zwei Schulkinder mit unterschiedlicher Ausprägung des Zielkonstrukts in der Stichprobe vorhanden sein müssen.

Schließlich wird die Rangkomponente durch die Korrelation der Einschätzungen durch die Lehrkräfte mit den Werten der Schulkinder ermittelt. Die Rangkomponente liefert dementsprechend Aussagen über die Güte der Lehrkrafturteile hinsichtlich der Platzierungen der Schulkinder in der korrekten Reihenfolge. Bei der Rangkomponente ergeben sich entsprechend Werte zwischen -1 und 1. Eine Korrelation von  $r = 1.00$  ergibt sich bei perfekter Übereinstimmung zwischen den Lehrkrafturteilen und den Positionen der Schulkinder in der Klasse. Eine Korrelation von  $r = -1.00$  indiziert entsprechend, dass die Reihenfolge der Merkmalseinschätzungen durch die Lehrkraft genau entgegengesetzt zur objektiv gemessenen Verteilung der Merkmale bei den Schulkindern erfolgte. Liegt die Korrelation nahe Null, so bedeutet dies, dass die Einschätzungen durch die Lehrkraft eher zufällig erfolgten und nahezu keinen linearen Zusammenhang mit den objektiv gemessenen Ausprägungen der Merkmale bei den Schulkindern aufweisen.

Betrachtet man die drei Komponenten gemeinsam, so ergibt sich ein vollständiges Bild hinsichtlich der Lehrkräfteeinschätzungen hinsichtlich des Niveaus, der Streuung

und der Rangfolge des interessierenden Schulkindmerkmals. Dennoch wird die Rangkomponente oft als die zentrale Komponente der Urteilsgenauigkeit angesehen (Schrader & Helmke, 1987). Eine ähnliche Bevorzugung der relativen Position zeigte sich bereits bei der Genese von Fähigkeitsselbstkonzepten, wo ebenfalls für die Vergleichsrahmen die relative Position im Vergleich zur sozialen Bezugsgruppe, zu eigenen Leistungen in anderen Schulfächern und zur eigenen früheren Leistung entscheidender war, als die absolute Leistung oder deren Varianz (vgl. Abschnitt 1.2.3). Entsprechend der Ansicht, dass die Rangkomponente das zentrale Element der Bestimmung der Urteilsgenauigkeit von Lehrkräften ausmache, blieben die Niveauelemente und die Differenzierungskomponente seltener beachtet (Spinath, 2005; Urhahne, Timm, Zhu & Tang, 2013). Allerdings hat die Beschränkung auf die Rangkomponente den Nachteil, dass systematische Urteilstendenzen wie eine deutliche Über- oder Unterschätzung (Niveauelemente) oder einer systematischen Fehleinschätzung bezüglich der Streuung (Differenzierungskomponente) unerkannt bleiben. So könnte einer Lehrkraft eine perfekte Rangfolgeeinschätzung der Schulkindmerkmale gelingen (Rangkomponente mit  $r = 1.00$ ) und ergänzend mag auch die Streuung perfekt eingeschätzt sein (Differenzierungskomponente = 1), aber dennoch könnte die mittlere Ausprägung des Lehrkrafturteils deutlich über dem Niveau in der Klasse liegen (Niveauelemente  $> 0$ ). Eine andere Lehrkraft mag hingegen sehr schwache diagnostische Leistungen zugeschrieben bekommen, da die Rangfolge der Schulkinde exakt falsch herum eingeschätzt wurde (Rangkomponente mit  $r = -1.00$ ), obwohl hinsichtlich der Niveauelemente und der Differenzierungskomponente perfekte Urteile gefällt worden sein mögen (für numerische Beispiele siehe Schrader & Helmke, 1987). Obgleich es optimale Ausprägungen für die jeweiligen Komponenten gibt, wurden noch

keine einheitlichen Richtlinien für die Bewertung der Abweichungen vom Optimum festgelegt (Leucht et al., 2012; Praetorius et al., 2011; Spinath, 2005). Für die zentrale Rangkomponente können die üblichen von (Cohen, 1992) vorgeschlagenen Richtwerte angenommen werden, denen zufolge eine Rangkomponente ab  $r = .10$  eine geringe Urteils-genauigkeit kennzeichnet. Rangkomponenten ab  $r = .30$  stünden entsprechend für eine mittlere und ab  $r = .50$  für eine hohe Urteils-genauigkeit. Hingegen merkten (Praetorius et al., 2011) an, dass sich diese Richtwerte gut für die Beurteilung von Zusammenhängen inhaltlich verschiedener Konstrukte eignen. Bei inhaltlich identischen Konstrukten, wie sie im Rahmen der Forschung zur Urteils-genauigkeit von Lehrkräften typischerweise erfasst werden, sollten jedoch höhere Werte im Sinne einer konvergenten Validität angelegt werden (Praetorius et al., 2011). Um dies zu beurteilen wäre jedoch eine umfassendere Untersuchung nötig, welche nicht nur die Verwendung unterschiedlicher Methoden (z. B. Selbsteinschätzung und Fremdeinschätzung), sondern auch mehrere Konstrukte einschließt (Campbell & Fiske, 1959).

Die eher deskriptive Betrachtung der Urteils-genauigkeit wurde im Rahmen methodisch fortgeschrittener Modellierung um mehrebenenanalytische Ansätze ergänzt. Im Rahmen der Forschung zur Urteils-genauigkeit von Lehrkräften liegen üblicherweise Datensätze vor, die eine geschachtelte Struktur aufweisen – die Schulkinder sind geschachtelt in Klassen. Dadurch ist zu erwarten, dass das Zielkonstrukt, zum Beispiel die Leistung innerhalb von Klassen, systematisch zwischen unterschiedlichen Klassen variiert. Die mittlere Leistungsstärke einer Klasse kann sich deutlich von der mittleren Leistungsstärke einer anderen Schulklasse unterscheiden und die Leistung eines Schulkindes ist somit auch abhängig von der Klasse, in der es sich befindet. Diesem Umstand muss durch

die Trennung zwischen der individuellen Schulkindebene und der übergeordneten Klassebene Rechnung getragen werden. Durch die Nutzung von Mehrebenenmodellen wird es möglich, Variablen auf der individuellen Ebene (z. B. Schulkind) und auf der übergeordneten Ebene (z. B. Schulklasse) simultan auf ihren Einfluss auf das Kriterium (z. B. Lehrkrafturteil) zu untersuchen (Raudenbush & Bryk, 2002). Eine Vorstellung unterschiedlicher Einflüsse auf das Lehrkrafturteil ist Gegenstand des folgenden Abschnitts.

#### **1.4.2 Empirische Befunde zur Urteilsgenauigkeit von Lehrkräften**

Die Rangkomponente wird als der zentrale Indikator der Urteilsgenauigkeit angesehen (Schrader & Helmke, 1987) und wurde nicht allein deshalb in den meisten Studien als einziger Indikator berichtet. Aufgrund dessen wird nach einem knappen Überblick über Befunde zur Niveauebene und der Differenzierungskomponente der Hauptteil dieses Abschnittes dem Bericht korrelativer Befunde zur Urteilsgenauigkeit gewidmet.

Bei der Beurteilung von Schulkindleistungen neigen Lehrkräfte häufig zu einer Überschätzung (Bates & Nettelbeck, 2001; Madelaine & Wheldall, 2005). Diese Tendenz zeigt sich besonders in Gruppen von Schulkindern mit schwachen Leseleistungen (Bates & Nettelbeck, 2001; Begeny, Eckert, Montarello & Storie, 2008; Feinberg & Shapiro, 2009). Der Befund aus den PISA-2000-Daten, dass Lehrkräfte 90% der Schulkinder, die im Rahmen von PISA schwache Leseleistungen zugeschrieben bekamen, selbst nicht als Schulkinder mit schwachen Leseleistungen identifizieren konnten (Artelt, Stanat, Schneider & Schiefele, 2001), bildete entsprechend einen Ausgangspunkt für das seitdem starke Interesse an Forschung zur Urteilsgenauigkeit von Lehrkräften. Das fachunspezifische Fähigkeitsselbstkonzept von Grundschulkindern wurde hingegen im Mittel ebenso unter-



schätzt (Spinath, 2005), wie deren mathematikspezifische, lesespezifische und schreibspezifische Fähigkeitsselbstkonzepte (Praetorius et al., 2011). Auch die mathematikspezifischen Selbstwirksamkeitsüberzeugungen von Sekundarschulkindern wurden unterschätzt (Huelmann et al., 2014).

Interessant wird ein Bezug der Befunde zur Überschätzung der Leistungen und zur Unterschätzung selbstbezogener Fähigkeitenkognitionen zu den Befunden zum Niveau der Leistungen und zum Niveau selbstbezogener Fähigkeitenkognitionen in Gruppen von Schulkindern mit und ohne Migrationshintergrund. Dort zeigten sich deutliche Leistungsunterschiede zuungunsten der Schulkinder mit Migrationshintergrund (vgl. Abschnitt 1.3.1) bei vergleichbarem Niveau der selbstbezogenen Fähigkeitenkognitionen (vgl. Abschnitt 1.3.2). Die Leistungsergebnisse aus standardisierten Mathematikleistungstests zeigten sich in einer Stichprobe aus Grundschulkindern in Deutschland als stärkerer Prädiktor zur Beurteilung des mathematikspezifischen Fähigkeitsselbstkonzepts als das mathematikspezifische Fähigkeitsselbstkonzept selbst (Praetorius et al., 2010). Verallgemeinert man diesen Befund, dann könnten Lehrkräfte die Einschätzung selbstbezogener Fähigkeitenkognitionen bevorzugt auf Leistungen als Informationsquelle basieren. Damit wäre zu erwarten, dass Lehrkräfte das Niveau der selbstbezogenen Kognitionen von Jugendlichen mit Migrationshintergrund trotz der Überschätzung gerade leistungsschwächerer Jugendlicher immer noch besonders deutlich unterschätzen. Hingegen könnte die Beurteilung der Jugendlichen ohne Migrationshintergrund hinsichtlich des Niveaus ihrer selbstbezogenen Fähigkeitenkognitionen nur in geringem Ausmaß verzerrt sein. Insgesamt könnte die Unterschätzung der selbstbezogenen Fähigkeitenkognitionen durch Lehrkräfte also aus der Diskrepanz in den Leistungen beider Gruppen bei vergleichbarem Niveau

der selbstbezogenen Fähigkeitskognitionen resultieren. Die empirische Bestätigung dieser Annahme steht allerdings noch aus.

Für die Differenzierungskomponente liegen insgesamt sehr wenige Befunde vor, die jedoch auf eine Unterschätzung der Leistungsvarianz (Helmke et al., 2004; Spinath, 2005) und eine Überschätzung der Streuung des fachunspezifischen Fähigkeitsselbstkonzepts (Spinath, 2005) hinweisen. Die Varianz spezifischer Fähigkeitsselbstkonzepte von Grundschulkindern im Lesen, im Schreiben und in Mathematik wurde hingegen akkurat eingeschätzt (Praetorius et al., 2011), ebenso die mathematikspezifischen Selbstwirksamkeitsüberzeugungen von Sekundarschulkindern (Huelmann et al., 2014). Erfreulich ist, dass die selbstbezogenen Fähigkeitskognitionen auf der für die pädagogisch-psychologische Forschung empfehlenswerten fachspezifischen Ebene zumindest hinsichtlich der Streuung akkurat beurteilt zu werden scheinen. Insgesamt deutet die Uneinheitlichkeit dieser Befunde jedoch noch weiteren Klärungsbedarf an, was insbesondere die Beurteilung von Selbstwirksamkeitsüberzeugungen im sprachlichen Bereich betrifft.

Bezüglich der häufig berichteten Rangkomponente liegen zur Urteilsgenauigkeit von Lehrkräften in Bezug auf die Leistungen von Schulkindern bereits metaanalytische Befunde vor, die einen Median in Höhe von  $r = .66$  (Hoge & Coladarci, 1989) beziehungsweise von  $r = .53$  (Südkamp et al., 2012) berichten. Zu beachten ist jedoch, dass die Spannweite der Urteilsgenauigkeit in den betrachteten Einzelstudien erheblich ist und von Korrelationen nahe Null bis  $r > .90$  reicht (Hoge & Coladarci, 1989; Südkamp et al., 2012). Für kognitive Schulkindmerkmale wie kognitiver Grundfähigkeit und Kreativität konnten Machts, Kaiser, Schmidt & Möller (2016) in einer jüngeren Metaanalyse eine

mittlere Urteilsgenauigkeit von  $r = .43$  aufdecken, wobei die Spannweite von  $r = -.18$  bis  $r = .72$  reichte.

Gegenüber diesen metaanalytischen Befunden fällt die Urteilsgenauigkeit motivationaler und affektiver Schulkindmerkmale in den wenigen veröffentlichten Einzelstudien zum Teil deutlich ab. Bezüglich der für diese Dissertation zentralen selbstbezogenen Fähigkeitskognitionen liegen bis heute vor allem Befunde zur Urteilsgenauigkeit in Bezug auf Fähigkeitsselbstkonzepte vor. So fanden (Marsh & Craven, 1991) Korrelationen zwischen Schulkindselbstberichten und Lehrkrafturteilen von  $r = .39$  für das lesespezifische,  $r = .47$  für das mathematikspezifische und  $r = .62$  für das allgemein schulische Fähigkeitsselbstkonzept. In einer deutschen Stichprobe zeigte sich jedoch nur eine Rangkomponente von  $r = .39$  für die Urteilsgenauigkeit des allgemein schulischen Fähigkeitsselbstkonzepts von Grundschulkindern (Spinath, 2005). Die Spannweite der Rangkomponente reichte in dieser Studie sogar von  $r = -.39$  bis  $r = .82$ , was darauf hindeutet, dass einige Lehrkräfte die Rangfolge der Schulkinder in Bezug auf das schulische Fähigkeitsselbstkonzept weniger genau beurteilen konnten, als wenn sie einfach zufällige Urteile abgegeben hätten. Anderen Lehrkräften gelang hingegen eine sehr akkurate Rangeinschätzung des schulischen Fähigkeitsselbstkonzeptes ihrer Schulkinder. In einer deutschen Studie zur Urteilsgenauigkeit in Bezug auf domänenspezifische Fähigkeitsselbstkonzepte wurden Rangkomponenten in Höhe von  $r = .25$  für das schreibspezifische,  $r = .52$  für das lesespezifische und  $r = .55$  für das mathematikspezifische Fähigkeitsselbstkonzept gefunden (Praetorius et al., 2011). Die Urteilsgenauigkeit von allgemein schulbezogenen Selbstwirksamkeitsüberzeugungen lag mit einer Rangkomponente von  $r = .21$  an deutschen Sekundarschulen etwas unter den Befunden für Fähigkeitsselbstkon-

zepte (Huelmann et al., 2014). Im amerikanischen Sprachraum zeigte sich für die lese-spezifischen Selbstwirksamkeitsüberzeugungen von Grundschulkindern eine Korrelation von  $r = .37$  mit den Lehrkräfteeinschätzungen (Coddington & Guthrie, 2009).

Insgesamt zeigten sich in den bisherigen Befunden zum Teil deutlich geringere Rangkorrelationen in den Urteilen zu selbstbezogenen Fähigkeitskognitionen als in den Leistungsurteilen. Eine Ursache dieser geringeren Rangkomponenten könnte darin liegen, dass Leistungen den stärksten Prädiktor der Lehrkrafturteile zumindest in Bezug auf das mathematikspezifische Fähigkeitsselbstkonzept von Grundschulkindern darstellen (Praetorius et al., 2010). In den Ausführungen zur Niveaueinschätzung wurde argumentiert, dass durch die Verwendung der Leistungen als primäre Informationsquelle eine stärkere Unterschätzung des Niveaus der selbstbezogenen Fähigkeitskognitionen insbesondere für Jugendliche mit Migrationshintergrund zustande kommen könnte, während diese Unterschätzung bei Jugendlichen ohne Migrationshintergrund geringer ausfallen dürfte. Da jedoch das Niveau selbstbezogener Fähigkeitskognitionen zwischen diesen beiden Gruppen geringere Differenzen aufweist als das Niveau der Leistungen, sollte eine solche Verzerrung der Niveaueinschätzung in der Folge auch zu einer geringeren Rangkomponente führen, da durch die Leistungsunterschiede systematisch die eingeschätzte Rangfolge zugunsten der Jugendlichen mit Migrationshintergrund verändert wird. Obwohl diese Vermutung bisher nicht weiter untersucht wurde, könnte eine geringere Rangkomponente in Stichproben mit einem hohen Anteil an Kindern und Jugendlichen mit Migrationshintergrund erste Hinweise bieten. Die bei Huelmann et al. (2014) berichtete sehr geringe Rangkomponente der Urteile in Bezug auf die fachunspezifische Selbstwirksamkeitsüberzeugung von  $r = .21$  in einer Stichprobe mit hohem Anteil Jugendlicher mit Migrati-

onshintergrund im Vergleich zu den Ergebnissen in Bezug auf die Urteilsgenauigkeit fachunspezifischer Fähigkeitsselfstkonzepte von  $r = .39$  (Spinath, 2005) beziehungsweise  $r = .62$  (Marsh & Craven, 1991) lässt die Inspektion eines moderierenden Einflusses durch den Migrationshintergrund von Schulkindern empfehlenswert erscheinen.

### 1.4.3 Einflüsse auf die Urteilsgenauigkeit von Lehrkräften

Metaanalytische Befunde zur Urteilsgenauigkeit von Lehrkräften in Bezug auf die Leistungen ihrer Schul Kinder zeigten große Spannweiten der Rangkomponente von  $r = .28$  bis  $r = .92$  (Hoge & Coladarci, 1989). Dies nahmen Südkamp et al. (2012) zum Anlass, in einer weiteren Metaanalyse mögliche Moderatoren der Urteilsgenauigkeit zu untersuchen. Im Ergebnis schlugen sie in ihrer Arbeit ein heuristisches Modell der Urteilsgenauigkeit von Lehrkrafturteilen in Bezug auf die Leistungen ihrer Schul Kinder vor, in dem potenzielle Einflussfaktoren aufgegriffen werden (siehe Abbildung 1.4). Dieses Modell bildet die Basis der nachfolgenden Ausführungen.

Der Kern des Modells wird durch die Urteilsgenauigkeit der Lehrkraft gebildet. Diese wird im Wesentlichen durch die Lehrkrafturteile sowie die Testleistung des Schul Kindes beeinflusst. Zusätzlich wirken sich die Eigenschaften der Tests (z. B. Länge, Schwierigkeit, Spezifität) und des Urteils (z. B. Spezifität) sowohl auf das Lehrkrafturteil beziehungsweise die Testleistung als auch auf die Urteilsgenauigkeit aus. Diese eher technischen Einflüsse sind bezüglich der Testmerkmale die Spezifität des Tests und seine Gütekriterien (Machts et al., 2016; Südkamp et al., 2012). Jedoch zeigte sich metaanalytisch bisher kein Vorteil einer höheren Spezifität durch Erfassung spezifischer Fähigkeiten (z. B. nur Algebra) im Vergleich zu einer fachspezifischen Erfassung der mathematischen Fähigkeiten in mehreren Bereichen (Südkamp et al., 2012) und ebenso war eine

höhere Reliabilität des Tests nicht mit einer höheren Urteilsgenauigkeit bezüglich der kognitiven Fähigkeiten von Schulkindern assoziiert (Machts et al., 2016). Wurde hingegen die Spezifität des Urteils der Spezifität des Tests angepasst, also die Kongruenz zwischen der Spezifität des Tests und des Urteils erhöht, so ging dies mit einer höheren Urteilsgenauigkeit einher (Südkamp et al., 2012). Dieser Zusammenhang ist durch die gestrichelte Linie auf der rechten Seite in Abbildung 1.4 symbolisiert.

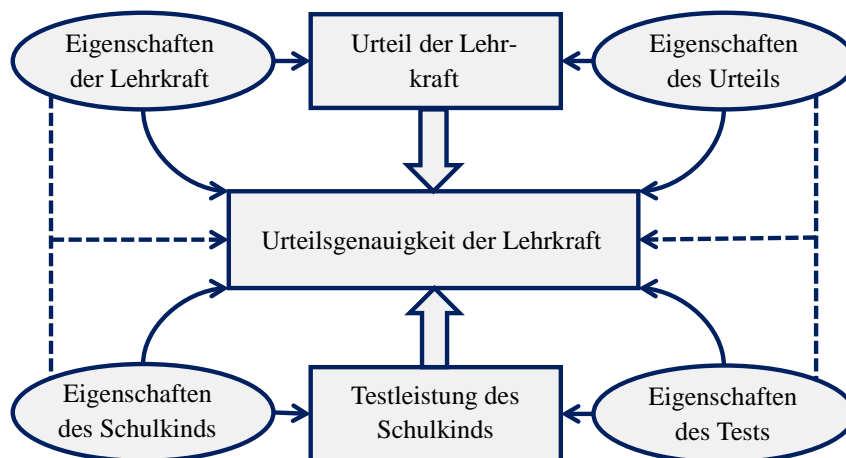


Abbildung 1.4. Heuristisches Modell zu Einflussfaktoren auf die Urteilsgenauigkeit der Lehrkraft (modifiziert nach Südkamp, Kaiser & Möller, 2012).

Ein weiteres technisches Urteilsmerkmal, das mit einer Verbesserung der Urteilsgenauigkeit bei der Leistungseinschätzung einherging, waren informierte Urteile im Vergleich zu uninformierten Urteilen (Südkamp et al., 2012). Informierte Urteile bezeichnen beispielsweise die Einschätzung der Anzahl korrekt gelöster Items eines Tests, nachdem einige Beispielitems gezeigt wurden, während uninformierte Urteile durch die Leistungseinschätzung anhand einer Ratingskala gewonnen werden. Auf der technischen

Seite empfehlen sich somit hinsichtlich ihrer Spezifität abgestimmte Erfassungen des zu beurteilenden Schulkindmerkmals und entsprechende Erfassungen des Lehrkrafturteils, welches zudem möglichst als informiertes Urteil abzugeben ist.

Persönliche Einflüsse wie Merkmale der Lehrkraft (z. B. Berufserfahrung, Vorurteile) oder des Schulkindes (z. B. Vorwissen, Intelligenz) wirken sich ebenfalls auf die Urteilsgenauigkeit der Lehrkraft aus (Südkamp et al., 2012). Die persönlichen Schulkindmerkmale wie auch die Testmerkmale können zudem einen indirekten Einfluss über die Schulkindleistungen auf die Urteilsgenauigkeit nehmen. Ebenso wirken sich Urteilsmerkmale und persönliche Lehrkraftmerkmale über die Lehrkrafturteile auf die Urteilsgenauigkeit aus. Hinzu kommen mögliche Einflüsse aus dem Zusammenspiel zwischen den persönlichen Merkmalen des Schulkindes und der Lehrkraft sowie Einflüsse aus dem Zusammenspiel zwischen Urteilsmerkmal und Testmerkmal, die sich ebenfalls auf die Urteilsgenauigkeit auswirken können. Ein solches Zusammenspiel kann zwischen persönlichen Merkmalen von Lehrkräften und Schulkindern wie beispielsweise deren ethnischen Hintergrund angenommen werden. Hier können Stereotype potenziell verzerrenden Einfluss auf die Urteile von Lehrkräften nehmen. In einer experimentellen Studie mit Lehramtsstudierenden in Deutschland beschrieben (Glock & Krolak-Schwerdt, 2013) fiktive Schulkinder entweder als unterdurchschnittlich oder zu den Besten in der Klasse gehörend und ließen das Sprachverständnis, das Leseverständnis und die Buchstabierfähigkeit einschätzen. In den Beschreibungen wurden zusätzlich jeweils der Vorname und die zu Hause gesprochene Sprache so variiert, so dass es sich bei der Beschreibung entweder um ein deutsches Kind oder um ein türkisches Kind handelte. Die unterdurchschnittlichen türkischen Kinder entsprachen damit dem Stereotyp des leistungsschwachen türkischen

Schulkindes, die leistungsschwachen deutschen Schulkinder waren entsprechend als nicht stereotypkonform eingestuft. Ebenso galten die leistungsstarken deutschen Schulkinder als stereotypkonform, während die leistungsstarken türkischen Schulkinder unvereinbar mit dem Stereotyp waren. Die stereotypkonformen Schulkinder mit türkischem Vornamen und Türkisch als zu Hause gesprochener Sprache wurden von den Lehramtsstudierenden im Sprachverständnis, Leseverständnis und der Buchstabierfähigkeit schwächer eingeschätzt als die nicht stereotypkonformen als leistungsschwach beschriebenen Schulkinder mit deutschem Vornamen und Deutsch als zu Hause gesprochener Sprache. Die Kompetenzeinschätzungen unterschieden sich jedoch nicht bei den als leistungsstark beschriebenen Schulkindern.

In einer amerikanischen Stichprobe konnten Alexander, Entwisle & Thompson (1987) zeigen, dass Schulkinder der ersten Grundschulklasse mit afrikanischem Migrationshintergrund (Minorität) und einem geringen sozioökonomischen Status (erfasst als dreistufige Variable für den elterlichen Schulabschluss mit High-School-Abschluss als mittlerem Wert) von ihren Lehrerinnen (es befanden sich keine Lehrer in der Stichprobe) als weniger reif angesehen wurden und geringere Leistungserwartungen an diese gestellt wurden als ihre Mitschülerinnen und Mitschüler, die Mitglied der Majorität waren und einen höheren sozioökonomischen Status aufwiesen. Diese Verzerrung war besonders stark ausgeprägt bei Lehrerinnen, die ihrerseits einen hohen sozioökonomischen Status, operationalisiert über den beruflichen Status des Vaters während des Kindes- und Jugendalters der Lehrerin über oder unter dem Stichprobenmittelwert, hatten (Alexander et al., 1987). Die Unterschiede in der Erwartungshaltung und Leistungseinschätzung der Lehrkräfte machten Alexander et al. als eine Ursache dafür aus, dass als benachteiligt



eingeschätzte Schulkinder mit vergleichbaren Testleistungen im ersten Quartal nach dem vierten Quartal mit den Testleistungen deutlich hinter den als nicht benachteiligt eingeschätzten Schulkindern zurückgefallen waren. Diese Ergebnisse legen nahe, dass ein moderierender Einfluss des Migrationshintergrundes von Schulkindern bei der Einschätzung der Urteilsgenauigkeit ihrer Lehrkräfte Berücksichtigung finden sollte.

In vier experimentellen Studien unter Nutzung des simulierten Klassenraums untersuchten Kaiser, Südkamp & Möller (2016) bei Trennung von Leistungsvermögen, ethnischer Zugehörigkeit und Minoritätsstatus, ob die Leistungsurteile von Lehrkräften eher in Richtung stereotyper Leistungserwartungen verzerrt waren oder ob die Zugehörigkeit zu einer Minderheit allein einen Einfluss auf das Urteil der Lehrkräfte hat. So wurden in drei Studien jeweils unterschiedliche ethnische Minoritäten präsentiert, die allgemein als unterschiedlich leistungsstark gelten. In der vierten Studie wurde das Geschlecht als Kriterium für die Zugehörigkeit zur Minorität oder zur Majorität operationalisiert. In allen vier Studien zeigte sich, dass die Mitglieder der Minorität hinsichtlich ihrer Leistungen akkurater eingeschätzt wurden als die Mitglieder der Majorität. Hingegen zeigte sich keine Verzerrung der Leistungseinschätzungen in die Richtung erwarteter Leistungstendenzen. Kaiser et al. bemerkten zwar, dass die Ergebnisse aus diesen experimentellen Ansätzen nicht direkt auf die Situation im echten Klassenzimmer übertragen werden können, empfehlen jedoch, dass dennoch die Komposition von Klassen bei der Einschätzung der Urteilsgenauigkeit beachtet werden sollte. Insgesamt sind psychologische Merkmale von Lehrkräften, wie zum Beispiel Stereotype, jedoch noch insgesamt selten Gegenstand der Forschung zur Urteilsgenauigkeit (Südkamp et al., 2012).

Die Ergebnisse von Kaiser et al. (2016) können auch als Hinweis darauf gedeutet werden, dass die Salienz, verbunden mit einer besseren Wahrnehmbarkeit, ein entscheidender Faktor für die Genauigkeit von Lehrkrafturteilen in Bezug auf die Leistungsstärke von Schulkindern ist. Die Wahrnehmbarkeit von Merkmalen kann abseits der vorab diskutierten technischen und persönlichen Faktoren einen hilfreichen Ansatz bieten, um Unterschiede in der Urteilsgenauigkeit zu erklären. Einen Ansatz hierfür bietet das *Realistic-Accuracy-Model* (RAM; Funder, 1995; siehe Abbildung 1.5).

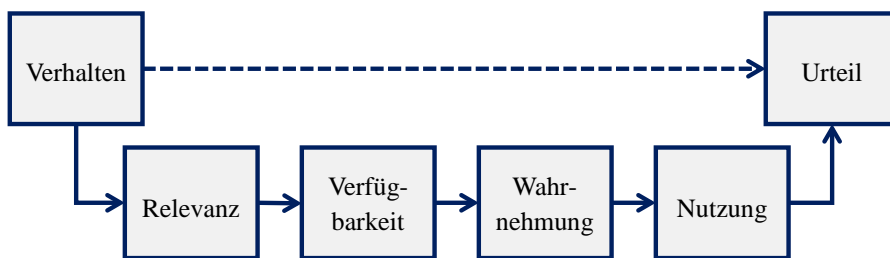


Abbildung 1.5. Prozessmodell einer akkuraten Persönlichkeitsmerkmalsbeurteilung (modifiziert nach Funder, 1995).

Im RAM werden Voraussetzungen für akkurate Urteile in Bezug auf Persönlichkeitsmerkmale spezifiziert. Die Grundannahme dabei ist, dass das einzuschätzende Persönlichkeitsmerkmal merkmalsrelevante Verhaltensweisen bei der Person auslöst, die eine beurteilende Person zur Urteilsbildung verwenden kann. Dieses Verhalten muss (1) für das einzuschätzende Persönlichkeitsmerkmal relevant sein, (2) für die beurteilende Person verfügbar sein, (3) von der beurteilenden Person wahrgenommen werden und (4) von der beurteilenden Person im Beurteilungsprozess genutzt werden (Funder, 1995). Ist eine dieser vier Voraussetzungen nicht gegeben, so fließt das vom zu beurteilenden Per-

sönlichkeitsmerkmal ausgelöste Verhalten nicht in die Beurteilung ein. Während zur Bestimmung der Urteilsgenauigkeit methodisch die Regression des Lehrkrafturteils auf das Schulkindmerkmal vorgenommen wird, die in Abbildung 1.5 durch den gestrichelten Pfeil im oberen Pfad symbolisiert ist, wird über den unteren Pfad in Abbildung 1.5 der beschriebene Prozess abgebildet, der für ein akkurates Urteil komplett durchlaufen werden müsste.

Das RAM bietet somit einen weiteren Zugang zur Erklärung der Diskrepanz zwischen den Rangkomponenten in den Leistungsurteilen und den Urteilen zu selbstbezogenen Fähigkeitskognitionen. Während zur Bewertung von Leistungen im täglichen Unterricht eine Vielzahl relevanter, verfügbarer, wahrnehmbarer und nutzbarer Verhaltensweisen durch das Schulkind auftreten sollte, ist dies bei selbstbezogenen Fähigkeitskognitionen sicherlich seltener der Fall. Da selbstbezogene Fähigkeitskognitionen stark von den subjektiven Bewertungen der Schulkinder abhängen (vgl. Abschnitt 1.2), wird der Prozess der Urteilsentstehung oft dadurch unterbrochen, dass dieser interne Bewertungsprozess für die Lehrkraft nicht verfügbar ist (zweite Stufe), sofern dadurch überhaupt ein zur Beurteilung selbstbezogener Fähigkeitskognitionen relevantes Verhalten (Stufe 1) angenommen werden kann. Auch auf der Seite der beurteilenden Lehrkraft können während der Stufe 3 und 4 des Prozesses Unterbrechungen auftreten. Bei Praetorius et al. (2011) zeigte sich insbesondere für das schreibspezifische Fähigkeitsselbstkonzept eine schwache Urteilsgenauigkeit ( $r = .25$ ) im Vergleich zu den Rangkomponenten der lesespezifischen und mathematikspezifischen Fähigkeitsselbstkonzepte. Um das schreibspezifische Fähigkeitsselbstkonzept akkurat einschätzen zu können, müssten beispielsweise auch Schreibaktivitäten außerhalb der Schulzeit nicht nur wahrgenommen werden

(Stufe 3), sondern insbesondere auch zur Beurteilung des schreibspezifischen Fähigkeitsselbstkonzepts genutzt werden (Stufe 4). Freizeitliche Schreibaktivitäten könnten nicht nur Ausdruck hoher Leistungsfähigkeit sein, sondern auch ein Hinweis darauf, dass sich das Kind auch affektiv positiv als schreibende Person sieht, was eine der Komponenten des schreibspezifischen Fähigkeitsselbstkonzepts darstellt (vgl. Abschnitt 1.2). Bleibt die Nutzung der relevanten, verfügbaren und wahrgenommenen Information über freizeitliche Schreibaktivität zur Beurteilung des Fähigkeitsselbstkonzeptes aus und wird stattdessen ausschließlich zur Aktualisierung des Schreibleistungsurteils verwendet, so resultiert daraus nach dem RAM eine geringere Urteilsgenauigkeit für das schreibspezifische Fähigkeitsselbstkonzept im Vergleich zur Schreibleistungsbeurteilung.

#### **1.4.4 Zwischenfazit zur Urteilsgenauigkeit von Lehrkräften**

In diesem Abschnitt wurde die Urteilsgenauigkeit von Lehrkräften mit einem Schwerpunkt auf dem Bereich der selbstbezogenen Fähigkeitenkognitionen betrachtet. Diese ist wichtig, um den Unterricht adaptiv gestalten und den individuellen Förderungsbedarf der Schulkinder optimal zu adressieren zu können. Insgesamt zeigte sich, dass die Lehrkrafturteile in Bezug auf Fähigkeitsselbstkonzepte schwächer sind als für Leistungen. Noch schwächer fielen die Korrelationen mit Selbstwirksamkeitsüberzeugungen aus.

Es wird eine Vielzahl von technischen und persönlichen Einflüssen auf die Urteilsgenauigkeit von Lehrkräften angenommen. Einer dieser möglichen Einflüsse ist der Migrationshintergrund von Schulkindern. Durch diesen könnten möglicherweise die Unterschätzung der Niveauelemente und die schwache Rangkomponente bei den Urteilen in Bezug auf selbstbezogene Fähigkeitenkognitionen erklärt werden. Allerdings existiert in diesem Bereich noch kaum Forschung mit Feldstudien.

Experimentelle Forschung legt nahe, dass Mitglieder von Minoritäten, unabhängig von deren ethnischem Hintergrund, genauer beurteilt werden als Mitglieder der Majorität. Dies könnte in einer besseren Wahrnehmbarkeit relevanter Hinweise zur Beurteilung des Zielmerkmals begründet sein. Allerdings ist die Befundlage insbesondere mit Feldstudien zu diesem Bereich äußerst dünn.

## **1.5 Zusammenfassung und Bewertung der Forschungslage**

Kern dieser Arbeit ist die Überprüfung der kausalen Ordnung zwischen Fähigkeitsselbstkonzepten beziehungsweise Selbstwirksamkeitsüberzeugungen und Leistungen an Schulkindern der Sekundarstufe I in Deutschland unter besonderer Berücksichtigung einer möglichen Moderation der Zusammenhänge durch den Migrationshintergrund der Schulkinder. Zudem soll die Urteilsgenauigkeit von Sekundarschullehrkräften in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen evaluiert werden, wobei ebenfalls der Migrationshintergrund der Schulkinder berücksichtigt wird.

Im theoretischen Hintergrund wurden zunächst Fähigkeitsselbstkonzepte und Selbstwirksamkeitsüberzeugungen auf einer konzeptuellen Ebene miteinander verglichen. Obwohl noch rege Forschungsaktivität auf diesem Gebiet besteht, kann grundsätzlich festgehalten werden, dass beide Konstrukte sowohl multidimensional wie hierarchisch aufgebaut sind (Bong & Clark, 1999; Bong & Skaalvik, 2003). Die Erfassung in der pädagogisch-psychologischen Forschung gestaltet sich ebenfalls für beide auf einer fachbeziehungsweise domänenspezifischen Ebene als besonders empfehlenswert (Marsh, 1990a; Pajares, 1996), was insbesondere bei Längsschnittstudien gelten sollte. Unterschiede zwischen beiden Konstrukten bestehen unter anderem darin, dass Selbstwirksamkeitsüberzeugungen fragiler als Fähigkeitsselbstkonzepte sind und auf spezifischere An-

forderungssituationen bezogen sind. Dies macht sie im Vergleich zu Fähigkeitsselbstkonzepten gut geeignet für eine Intervention beispielsweise durch Lehrkräfte im Unterricht. Mögliche Ansatzpunkte für Interventionen ergeben sich in den vier Quellen der Selbstwirksamkeitsüberzeugung. So können durch herausfordernde aber lösbare Aufgaben Erfolgserlebnisse bei Schulkindern erlebbar gemacht werden, auf die Lehrkräfte direkt hinweisen können. In der Folge sollte die spezifische Selbstwirksamkeitsüberzeugung des Schulkinds passend zu seiner Leistungsfähigkeit erhöht sein und wiederum positiv mit zeitlich nachfolgenden Leistungen zusammenhängen.

Obwohl die Frage der kausalen Ordnung zwischen Leistungen und selbstbezogenen Fähigkeitskognitionen eine Vielzahl empirischer Arbeiten nach sich zog, bestehen insbesondere im sprachlichen Bereich noch Lücken. Die bisherigen Befunde über verschiedene Leistungsbereiche hinweg sind zudem uneinheitlich. Arbeiten zur kausalen Ordnung von Selbstwirksamkeitsüberzeugungen und Leistungen im Bereich Lesen fehlen bislang gänzlich. Trotz deutlicher Leistungsdifferenzen zuungunsten von Kindern und Jugendlichen mit Migrationshintergrund bei vergleichbaren Niveaus der selbstbezogenen Fähigkeitskognitionen wurde ein möglicher moderierender Einfluss des Migrationshintergrunds auf das Zusammenspiel zwischen Leistungen und selbstbezogenen Fähigkeitskognitionen nicht ausreichend evaluiert. Die vergleichsweise hohe Ausprägung der selbstbezogenen Fähigkeitskognitionen bei den leistungsschwächeren Kindern und Jugendlichen mit Migrationshintergrund deuten auf eine schlechte Kalibrierung der selbstbezogenen Fähigkeitskognitionen hin. Diese zu verbessern ist eine Herausforderung an die Lehrkräfte. Insbesondere müsste dazu eine akkurate Einschätzung der selbstbezogenen Fähigkeitskognitionen gelingen.

Forschungsergebnisse zur Urteilsgenauigkeit von Lehrkräften weisen auf eine schwache Urteilsgenauigkeit in Bezug auf selbstbezogene Fähigkeitskognitionen hin. Eine mittlere Unterschätzung des Niveaus geht hier mit einer schwachen bis moderaten Korrelation der Lehrkrafturteile mit den Schulkindselbsteinschätzungen einher. Insbesondere bei Stichproben mit hohem Anteil an Jugendlichen mit Migrationshintergrund zeigten sich dabei geringe Rangkorrelationen, was die Inspektion eines Einflusses des Migrationshintergrundes auf die Urteilsgenauigkeit impliziert.

## **1.6 Forschungsfragen und Vorstellung der Teilstudien**

Die theoretische Konzeption der Selbstwirksamkeitsüberzeugungen gibt im Grunde das Gerüst vor, von reziproken Effekten zwischen Selbstwirksamkeitsüberzeugungen und Schulleistungen auszugehen. Die empirische Befundlage dazu ist jedoch bisher äußerst dünn und entspricht methodisch oft nicht den aktuellen Empfehlungen. Für Fähigkeitsselbstkonzepte ist die empirische Befundlage für das Vorliegen reziproker Effekte hingegen überzeugend. Gleichwohl liegen hier wenige Befunde im Bereich verbaler Fähigkeitsselbstkonzepte vor. Zudem blieb bisher die Frage nach einer möglichen Moderation der reziproken Effekte insbesondere im sprachlichen Bereich durch den Migrationshintergrund von Schulkindern unbeantwortet.

Obwohl die Förderung von schulischen Selbstwirksamkeitsüberzeugungen einen substanziellen Teil der pädagogisch-psychologischen Literatur einnimmt, beschränkt sich die Auseinandersetzung mit diesem Thema auf Hinweise, welche Unterrichtsmaßnahmen hilfreich sein könnten. Die Frage danach, ob Lehrkräfte diese Maßnahmen erfolgreich umsetzen können und an die individuellen Lernvoraussetzungen der Schulkin-

der durch eine optimale Einschätzung der Selbstwirksamkeitsüberzeugungen von Schulkindern anpassen können, blieb ebenfalls unbeachtet.

Vor diesem Hintergrund leistet die vorliegende Arbeit aus theoretischer wie praktischer Sicht mit drei empirischen Forschungsarbeiten einen ersten Schritt zur Füllung der bestehenden Forschungslücken. Die theoretische Leistung dieser Arbeit besteht darin, die in der Theorie angenommenen reziproken Effekte zwischen Selbstwirksamkeitsüberzeugungen und Schulleistungen in einer angemessenen methodischen Zugangsweise empirisch zu bestätigen, sowie die empirische Befundlage für reziproke Effekte zwischen dem verbalen Fähigkeitsselbstkonzept und sprachlichen Leistungen zu unterfüttern.

Zunächst wurde mit der ersten Studie der noch selten erforschte reziproke Zusammenhang zwischen dem verbalen Fähigkeitsselbstkonzept und den sprachlichen Leistungen in den Bereichen Lesen und Rechtschreiben mittels Strukturgleichungsmodellen unter Beachtung der methodischen Empfehlungen für solche Arbeiten untersucht. Es wurden sowohl die Schulform als auch der Migrationshintergrund als mögliche Moderatoren berücksichtigt, indem die metrische Invarianz der Mess- und Strukturmodelle überprüft wurde. Entsprechend der bisherigen Befunde zu Leistungen und Fähigkeitsselbstkonzepten in Gruppen mit und ohne Migrationshintergrund wurde erwartet, dass Schulkinder mit Migrationshintergrund deutlich geringere sprachliche Leistungen bei nur leicht geringerer Höhe des verbalen Fähigkeitsselbstkonzeptes als ihre Mitschülerinnen und Mitschüler aufweisen. Die sprachlichen Leistungen und die verbalen Selbstkonzepte sollten in beiden Gruppen eine ähnliche Entwicklung nehmen und dabei eine hohe Stabilität aufweisen. Im Fokus dieser ersten Studie stand jedoch die Frage, ob sowohl bei Schul-



kindern mit Migrationshintergrund als auch bei Schulkindern ohne Migrationshintergrund reziproke Effekte für beide sprachlichen Leistungen und dem verbalen Fähigkeitsselbstkonzept nachweisbar sind, was auf Basis der rezenten Literatur erwartet wurde. Ein höheres verbales Fähigkeitsselbstkonzept sollte also in beiden Schulkindgruppen einen vergleichbaren leistungsfördernden Effekt haben, wie auch höhere Leistungen in beiden Schulkindgruppen zu einem höheren verbalen Fähigkeitsselbstkonzept führen sollten.

Gegenstand der zweiten Studie war die empirische Überprüfung reziproker Effekte zwischen Selbstwirksamkeitsüberzeugungen und Leistungen in Lesen und Mathematik. Wie in Studie 1 wurden Strukturgleichungsmodelle mit zwei Messzeitpunkten zur Beantwortung der fokussierten Fragestellung gewählt, wobei ebenfalls entsprechend der methodischen Empfehlungen vorgegangen wurde. Ebenso wie in Studie 1 wurden auch in Studie 2 sowohl die Schulform als auch der Migrationshintergrund als mögliche Moderatoren durch Überprüfung der metrischen Invarianz der Mess- und Strukturmodelle berücksichtigt. Auch in dieser Studie wurde entsprechend der theoretischen Konzeption der Selbstwirksamkeitsüberzeugungen und den relevanten empirischen Befunden erwartet, dass sich reziproke Effekte sowohl zwischen mathematikspezifischen Selbstwirksamkeitsüberzeugungen und Mathematikleistungen als auch zwischen lesespezifischen Selbstwirksamkeitsüberzeugungen und Leseleistungen zeigen und diese Befunde in Gruppen von Schulkindern mit und ohne Migrationshintergrund gleichermaßen auftreten. Diese Studie gehört zu den ersten, in denen reziproke Effekte zwischen lesespezifischen Selbstwirksamkeitsüberzeugungen und Leseleistungen entsprechend aktueller methodischer Empfehlungen untersucht wurden. Zudem bietet sie durch die Erfassung von Selbstwirksamkeitsüberzeugungen und Leistungen in zwei unterschiedlichen Bereichen

anhand einer Stichprobe zu identischen Zeitpunkten die Möglichkeit des Vergleichs der Effekte über die Bereiche Lesen und Mathematik. Die hier gewonnen Befunde aus dem Bereich der Mathematik bieten zudem Anschluss an bereits vorliegende empirische Befunde.

Die dritte Studie ergänzt schließlich die Studie 2, indem die Urteilsgenauigkeit von Lehrkräften in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen ihrer Schulkinder untersucht wird. In dieser Querschnittstudie wurden zunächst die klassischen Komponenten (Niveauelemente, Differenzierungskomponente und Rangkomponente; siehe Abschnitt 1.4.1) der fokussierten Lehrkrafturteile in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen der Schulkinder ermittelt. Hier wurde auf Grundlage der Befunde zu Lehrkräfteeinschätzungen von motivationalen Lernvoraussetzungen ihrer Schulkinder erwartet, dass das Niveau der lesespezifischen Selbstwirksamkeitsüberzeugung in der Klasse unterschätzt sowie deren Streuung überschätzt wird. Bezüglich der Rangfolge sollte sich eine positive, wenn auch schwache Korrelation zeigen. Ergänzend werden diese Komponenten auch für die Lehrkrafturteile in Bezug auf die Leseleistungen der Schulkinder berichtet. Diese Vorgehensweise erlaubt einerseits den Vergleich der Ergebnisse zum Leistungsurteil mit bereits bekannten empirischen Befunden zur Urteilsgenauigkeit von Leseleistungen und andererseits den Vergleich der Ergebnisse zur Einschätzung der lesespezifischen Selbstwirksamkeitsüberzeugungen mit empirischen Befunden zur Urteilsgenauigkeit bezüglich anderer individueller Schulkindmerkmale, wie beispielsweise der Urteilsgenauigkeit in Bezug auf Fähigkeitsselbstkonzepte. Zusätzlich wurden mit einem mehrebenenanalytischen Ansatz unter Verwendung von Prädiktoren auf der Schulkindebene (z. B. Geschlecht und Migrationshintergrund des

Schulkind, Schulnote) und der Klassenebene (z. B. Schulform und Lehrkraftefahrung) mögliche Einflussfaktoren auf die Lehrkräfteeinschätzungen der Schulkindselbstwirksamkeitsüberzeugungen explorativ untersucht. Während die Selbstwirksamkeitsüberzeugungen auf intern ablaufenden Wahrnehmungen der Schulkinder beruhen und damit kaum direkt durch die Lehrkräfte zu beurteilen sind, liegen im täglichen Unterricht diverse Indikatoren für die Schulkindleistungen ebenso wie deren Migrationshintergrund vor. Entsprechend wurde insbesondere auch eine Interaktion zwischen dem Migrationshintergrund des Schulkindes und der Leseleistungseinschätzung durch die Lehrkraft auf die Leseselbstwirksamkeitsüberzeugungseinschätzung der Lehrkraft überprüft.

## 1.7 Literatur

- Albert, S. (1977). Temporal comparison theory. *Psychological Review*, 84, 485-503.
- Alexander, K. L., Entwisle, D. R. & Thompson, M. S. (1987). School performance, status relations, and the structure of sentiment: Bringing the teacher back in. *American Sociological Review*, 52, 665-682.
- Arens, A. K., Marsh, H. W., Craven, R. G., Yeung, A. S., Randhawa, E. & Hasselhorn, M. (2016). Math self-concept in preschool children. Structure, achievement relations, and generalizability across gender. *Early Childhood Research Quarterly*, 36, 391-403.
- Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K. & vom Hofe, R. (2016). Math self-concept, grades, and achievement test scores. Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*. Advance Online Publication.
- Arens, A. K., Yeung, A. S., Craven, R. G. & Hasselhorn, M. (2011). The twofold multidimensionality of academic self-concept. Domain specificity and separation between competence and affect components. *Journal of Educational Psychology*, 103, 970-981.
- Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J. et al. (2007). *Förderung von Lesekompetenz - Expertise* (Bundesministerium für Bildung und Forschung (BMBF), Hrsg.) (Bildungsforschung Nr. 17), Bonn.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele & W.

Schneider (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-137). Opladen: Leske + Budrich.

Autorengruppe Bildungsberichterstattung (Hrsg.). (2016). *Bildung in Deutschland 2016. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann.

Bandura, A. (1977). Self-efficacy. Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist*, 33, 344-358.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122-147.

Bandura, A. (1986). *Social foundations of thought and action. A social cognitive theory*. Englewood Cliffs: Prentice-Hall.

Bandura, A. (1993). Perceived self-efficacy in cognitive functioning and development. *Educational Psychologist*, 28, 117-148.

Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York, NY: W. H. Freeman.

Bates, C. & Nettelbeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, 21, 177-187.

Baumeister, R. F., Campbell, J. D., Krueger, J. I. & Vohs, K. D. (2005). Exploding the self-esteem myth. *Scientific American Mind*, 16 (4), 50-57.

- Baumert, J., Watermann, R. & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6, 46-72.
- Begeny, J. C., Eckert, T. L., Montarello, S. A. & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23, 43-55.
- Bong, M., Cho, C., Ahn, H. S. & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *Journal of Educational Research*, 105, 336-352.
- Bong, M. & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, 34, 139-153.
- Bong, M. & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15, 1-40.
- Bos, W., Bensen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G. & Fried, L. (Hrsg.). (2008). *IGLU-E 2006. Die Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Tarelli, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.). (2012). *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.

- Bos, W., Wendt, H., Köller, O. & Selter, C. (Hrsg.). (2012). *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bouffard-Bouchard, T. (1990). Influence of Self-Efficacy on Performance in a Cognitive Task. *Journal of Social Psychology, 130*, 353-363.
- Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A. et al. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology, 102*, 964-981.
- Byrne, B. M. (1986). Self-concept/academic achievement relations. An investigation of dimensionality, stability, and causality. *Canadian Journal of Behavioural Science, 18*, 173-186.
- Calsyn, R. J. & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology, 69*, 136-145.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.
- Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C. et al. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology, 100*, 525-534.
- Coddington, C. S. & Guthrie, J. T. (2009). Teacher and student perceptions of boys' and girls' reading motivation. *Reading Psychology, 30*, 225-249.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

- Eckert, C., Schilling, D. & Stiensmeier-Pelster, J. (2006). Einfluss des Fähigkeitsselbstkonzepts auf die Intelligenz- und Konzentrationsleistung. *Zeitschrift für Pädagogische Psychologie*, 20, 41-48.
- Esser, H. (2006). *Sprache und Integration. Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt am Main: Campus.
- Feinberg, A. B. & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *Journal of Educational Research*, 102, 453-462.
- Ferla, J., Valcke, M. & Cai, Y. (2009). Academic self-efficacy and academic self-concept. Reconsidering structural relationships. *Learning and Individual Differences*, 19, 499-505.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652-670.
- Gebhardt, M., Rauch, D., Mang, J., Sälzer, C. & Stanat, P. (2013). Mathematische Kompetenz von Schülerinnen und Schülern mit Migrationshintergrund. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 275-308). Münster: Waxmann.
- Glock, S. & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Social Psychology of Education*, 16, 111-127.



- Goethals, G. R. & Darley, J. M. (1977). Social comparison theory: An attributional approach. In J. M. Suls & R. L. Miller (Hrsg.), *Social comparison processes: Theoretical and empirical perspectives* (S. 259-278). Washington, DC: Hemisphere.
- Green, J., Liem, G. A. D., Martin, A. J., Colmar, S., Marsh, H. W. & McInerney, D. (2012). Academic motivation, self-concept, engagement, and performance in high school: Key processes from a longitudinal perspective. *Journal of Adolescence*, 35, 1111-1122.
- Guay, F., Marsh, H. W. & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, 95, 124-136.
- Hannula, M. S., Bofah, E. A., Tuohilampi, L. & Metsämuuronen, J. (2014). A longitudinal analysis of the relationship between mathematics-related affect and achievement in Finland. In *Proceedings of the Joint Meeting of PME 38 and PME-NA 36* (S. 249-256). Vancouver, Canada: PME. Zugriff am 04.10.2016. Verfügbar unter [https://www.unikassel.de/fb10/fileadmin/datas/fb10/mathematik/didaktik/YERME\\_Summer\\_School/Experts/RR\\_Bofah.pdf](https://www.unikassel.de/fb10/fileadmin/datas/fb10/mathematik/didaktik/YERME_Summer_School/Experts/RR_Bofah.pdf)
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Grieser (Hrsg.), *Schulmanagement und Schulentwicklung* (S. 119-144). Hohengehren: Schneider.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement. A review of literature. *Review of Educational Research*, 59, 297-313.

- Hopf, D. (2005). Zweisprachigkeit und Schulleistung bei Migrantenkindern. *Zeitschrift für Pädagogik*, 51, 236-251.
- Hornstra, L., van der Veen, I., Peetsma, T. & Volman, M. (2013). Developments in motivation and achievement during primary school: A longitudinal study on group-specific differences. *Learning and Individual Differences*, 23, 195-204.
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49, 505-528.
- Huelmann, T., Ferdinand, H. D., Gebauer, M. M., McElvany, N., Bos, W., Köller, O. et al. (2014). Urteilsgenauigkeit von Lehrkräften bezüglich der Selbstwirksamkeitsüberzeugungen von Schülerinnen und Schülern. In H.-G. Holtappels, M. Pfeifer, A. S. Willems, W. Bos & N. McElvany (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Veröffentlichungen des Instituts für Schulentwicklungsforschung, Bd. 18, 1. Aufl., S. 298-320). Weinheim: Beltz Juventa.
- Hwang, M. H., Choi, H. C., Lee, A., Culver, J. D. & Hutchison, B. (2016). The relationship between self-efficacy and academic achievement. A 5-year panel analysis. *Asia-Pacific Education Researcher*, 25, 89-98.
- Jansen, M., Scherer, R. & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences. Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, 41, 13-24.
- Joët, G., Usher, E. L. & Bressoux, P. (2011). Sources of self-efficacy: An investigation of elementary school students in France. *Journal of Educational Psychology*, 103, 649-663.

- Kaiser, J., Südkamp, A. & Möller, J. (2016). The effects of student characteristics on teachers' judgment accuracy. Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*. Advance Online Publication.
- Klassen, R. M. (2004). Optimism and realism: A review of self-efficacy from a cross-cultural perspective. *International Journal of Psychology*, 39, 205-230.
- Klassen, R. M. (2006). Too much confidence? The self-efficacy of adolescents with learning disabilities. In F. Pajares & T. Urdan (Hrsg.), *Adolescence and education. Vol. 5: Self-efficacy and adolescence* (Adolescence and education, Bd. 5, S. 181-200). Greenwich, CT: Information Age Publishing.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M. et al. (Hrsg.). (2010). *Pisa 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, 19, 355-365.
- Leucht, M., Tiffin-Richards, S., Vock, M., Pant, H. A. & Köller, O. (2012). Diagnostische Kompetenz von Englischlehrkräften. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 44, 163-177.
- Machts, N., Kaiser, J., Schmidt, F. T. & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities. A meta-analysis. *Educational Research Review*, 19, 85-103.
- Maddux, J. E. (1995). Self-efficacy theory. An introduction. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment. Theory, research, and application* (The Plenum series in social/clinical psychology, pp. 3-33). New York, NY: Plenum Press.

- Madelaine, A. & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education*, 52(1), 33-42.
- Marsh, H. W. (1986). Verbal and math self-concepts. An internal/external frame of reference model. *American Educational Research Journal*, 23, 129-149.
- Marsh, H. W. (1990a). A multidimensional, hierarchical model of self-concept. Theoretical and empirical justification. *Educational Psychology Review*, 2, 77-172.
- Marsh, H. W. (1990b). Influences of internal and external frames of reference on the formation of math and English self-concepts. *Journal of Educational Psychology*, 82, 107-116.
- Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology*, 84, 35-42.
- Marsh, H. W., Byrne, B. M. & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement. Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, 34, 155-167.
- Marsh, H. W. & Craven, R. G. (1991). Self-other agreement on multiple dimensions of preadolescent self-concept. Inferences by teachers, mothers, and fathers. *Journal of Educational Psychology*, 83, 393-404.
- Marsh, H. W. & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1, 133-163.

- Marsh, H. W. & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*, 364-376.
- Marsh, H. W. & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*, 56-67.
- Marsh, H. W., Hau, K.-T. & Kong, C.-K. (2002). Multilevel causal ordering of academic self-concept and achievement. Influence of Language of Instruction (English Compared With Chinese) for Hong Kong Students. *American Educational Research Journal*, *39*, 727-763.
- Marsh, H. W. & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations. Reunification of east and west German school systems after the fall of the Berlin wall. *Contemporary Educational Psychology*, *29*, 264-282.
- Marsh, H. W., Köller, O. & Baumert, J. (2001). Reunification of east and west German school systems: Longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal*, *38*, 321-350.
- Marsh, H. W., Kong, C.-K. & Hau, K.-T. (2001). Extension of the internal/external frame of reference model of self-concept formation. Importance of native and nonnative languages for Chinese students. *Journal of Educational Psychology*, *93*, 543-553.
- Marsh, H. W. & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*, 59-77.

- Marsh, H. W. & O'Mara, A. (2008). Reciprocal Effects Between Academic Self-Concept, Self-Esteem, Achievement, and Attainment Over Seven Adolescent Years: Unidimensional and Multidimensional Perspectives of Self-Concept. *Personality and Social Psychology Bulletin*, 34, 542-552.
- Marsh, H. W. & Shavelson, R. (1985). Self-concept. Its multifaceted, hierarchical structure. *Educational Psychologist*, 20 (3), 107-123.
- Marsh, H. W., Trautwein, U., Lüdtke, O. & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs. Role of generalized and specific others. *Journal of Educational Psychology*, 100, 510-524.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76, 397-416.
- Marsh, H. W., Walker, R. & Debus, R. (1991). Subject-Specific Components of Academic Self-Concept and Self-Efficacy. *Contemporary Educational Psychology*, 16, 331-345.
- Marsh, H. W. & Yeung, A. S. (2001). An Extension of the Internal/External Frame of Reference Model: A Response to Bong (1998). *Multivariate Behavioral Research*, 36, 389-420.
- Mittag, W., Kleine, D. & Jerusalem, M. (2002). Evaluation der schulbezogenen Selbstwirksamkeit von Sekundarschülern. *Zeitschrift für Pädagogik, Beiheft*, 44, 145-173.
- Möller, J. & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology*, 93, 826-835.

- Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte. Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau*, 55, 19-27.
- Möller, J. & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, 120, 544-560.
- Möller, J., Pohlmann, B., Köller, O. & Marsh, H. W. (2009). A Meta-Analytic Path Analysis of the Internal/External Frame of Reference Model of Academic Achievement and Academic Self-Concept. *Review of Educational Research*, 79, 1129-1167.
- Möller, J., Retelsdorf, J., Köller, O. & Marsh, H. W. (2011). The Reciprocal Internal/External Frame of Reference Model: An Integration of Models of Relations Between Academic Achievement and Self-Concept. *American Educational Research Journal*, 48, 1315-1346.
- Möller, J. & Trautwein, U. (2015). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 177-199). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Niemivirta, M. & Tapola, A. (2007). Self-efficacy, interest, and task performance. Within-task changes, mutual relationships, and predictive effects. *Zeitschrift für Pädagogische Psychologie*, 21, 241-250.
- Organisation for Economic Co-operation and Development. (2005). *PISA 2003 technical report* (Organisation for Economic Co-operation and Development, Hrsg.). Paris, France: Organisation for Economic Co-operation and Development. Zugriff am 14.07.2016.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66, 543-578.

- Pajares, F., Johnson, M. J. & Usher, E. L. (2007). Sources of writing self-efficacy beliefs of elementary, middle, and high school students. *Research in the Teaching of English, 42*, 104-120.
- Pajares, F. & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*, 193-203.
- Pajares, F. & Schunk, D. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. J. Riding & S. Rayner (Hrsg.), *International perspectives on individual differences* (S. 239-266). Westport, CT: Ablex.
- Pajares, F. & Schunk, D. H. (2005). Self-efficacy and self-concept beliefs. Jointly contributing to the quality of human life. In H. W. Marsh, R. Craven & D. M. McInerney (Eds.), *New frontiers for self-research* (International advances in self-research, vol. 2, pp. 95-121). Greenwich, CT: Information Age Publishing.
- Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S. & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology, 34*, 29-48.
- Pelham, B. W. & Swann, W. B. (1989). From self-conceptions to self-worth. On the sources and structure of global self-esteem. *Journal of Personality and Social Psychology, 57*, 672-680.
- Pintrich, P. R. & Schunk, D. H. (1996). *Motivation in Education. Theory, Research and Applications*. Englewood Cliffs: Prentice-Hall.
- Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W. & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on math-



ematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, 84, 152-174.

Ployhart, R. E. & Vandenberg, R. J. (2010). Longitudinal Research. The Theory, Design, and Analysis of Change. *Journal of Management*, 36, 94-120.

Pohlmann, B. & Möller, J. (2009). On the benefit of dimensional comparisons. *Journal of Educational Psychology*, 101, 248-258.

Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A. & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging self-concepts of students. *Journal of Educational Research*, 106, 64-76.

Praetorius, A.-K., Greb, K., Lipowsky, F. & Gollwitzer, M. (2010). Lehrkräfte als Diagnostiker. Welche Rolle spielt die Schülerleistung bei der Einschätzung von mathematischen Selbstkonzepten? *Journal for Educational Research Online*, 2(1), 121-144.

Praetorius, A.-K., Karst, K., Dickhäuser, O. & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 58, 81-91.

Preckel, F., Niepel, C., Schneider, M. & Brunner, M. (2013). Self-concept in adolescence: a longitudinal study on reciprocal effects of self-perceptions in academic and social domains. *Journal of Adolescence*, 36, 1165-1175.

Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (Hrsg.). (2013). *PISA 2012. Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.

Rauch, D., Mang, J., Härtig, H. & Haag, N. (2016). Naturwissenschaftliche Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund. In K. Reiss, C. Sälzer, A.

- Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (1. Aufl., S. 317-348). Münster: Waxmann.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (Advanced quantitative techniques in the social sciences, Vol. 1, 2nd ed.). Thousand Oaks: Sage Publications.
- Ready, D. D. & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, 48, 335-360.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E. & Köller, O. (Hrsg.). (2016). *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (1. Aufl.). Münster: Waxmann.
- Reitz, A. K., Motti-Stefanidi, F. & Asendorpf, J. B. (2014). Mastering developmental transitions in immigrant adolescents: the longitudinal interplay of family functioning, developmental and acculturative tasks. *Developmental Psychology*, 50, 754-765.
- Retelsdorf, J., Köller, O. & Möller, J. (2014). Reading achievement and reading self-concept – Testing the reciprocal effects model. *Learning and Instruction*, 29, 21-30.
- Roebbers, C. M., Mecheril, A. & Schneider, W. (1998). Migrantenkinder in deutschen Schulen. Eine Studie zur Persönlichkeitsentwicklung. *Zeitschrift für Pädagogik*, 44, 723-736. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:0111-opus-68373>
- Rosenberg, M. (1989). *Society and the adolescent self-image* (Wesleyan paperback, Rev. ed., 1st Wesleyan ed.). Middletown, CT: Wesleyan University Press.
- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2003). Das Fähigkeitsselbstkonzept und seine Erfassung. In J. Stiensmeier-Pelster & F. Rheinberg (Hrsg.), *Diagnostik*

von *Motivation und Selbstkonzept* (Tests und Trends Jahrbuch der pädagogisch-psychologischen Diagnostik Neue Folge, Bd. 2, S. 3-14). Göttingen: Hogrefe.

Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23, 237-245.

Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27-52.

Schunk, D. H. (1983). Developing children's self-efficacy and skills. The roles of social comparative information and goal setting. *Contemporary Educational Psychology*, 8, 76-86.

Schunk, D. H. (1985). Self-efficacy and classroom learning. *Psychology in the Schools*, 22, 208-223.

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review*, 1, 173-208.

Schunk, D. H. & Meece, J. L. (2006). Self-efficacy development in adolescences. In F. Pajares & T. Urdan (Hrsg.), *Adolescence and education. Vol. 5: Self-efficacy and adolescence* (Adolescence and education, Bd. 5, S. 71-96). Greenwich, CT: Information Age Publishing.

Schunk, D. H. & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (Educational Psychology Handbook Series, pp. 35-54). New York, NY: Routledge.

Schunk, D. H. & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading and Writing Quarterly*, 23, 7-25.

- Schwarzer, R. & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. *Zeitschrift für Pädagogik, Beiheft, 44*, 28-53.
- Schwippert, K., Hornberg, S. & Goy, M. (2008). Lesekompetenzen von Kindern mit Migrationshintergrund im nationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust & L. Fried (Hrsg.), *IGLU-E 2006. Die Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Schwippert, K., Wendt, H. & Tarelli, I. (2012). Lesekompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 191-208). Münster: Waxmann.
- Selig, J. P. & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little & N. A. Card (Hrsg.), *Handbook of developmental research methods* (S. 265-278). New York: Guilford Press.
- Shajek, A., Lüdtke, O. & Stanat, P. (2006). Akademische Selbstkonzepte bei Jugendlichen mit Migrationshintergrund. *Unterrichtswissenschaft, 34*, 125-145.
- Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research, 46*, 407-441.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie, 19*, 85-95.
- Stanat, P. & Christensen, G. (2006). *Schulerfolg von Jugendlichen mit Migrationshintergrund im internationalen Vergleich. Eine Analyse von Voraussetzungen und Erträgen schulischen*

*Lernens im Rahmen von PISA 2003* (Bundesministerium für Bildung und Forschung, Hrsg.) (Bildungsforschung Nr. 19), Bonn.

- Stanat, P., Rauch, D. & 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). Münster: Waxmann.
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*, 743-762.
- Suls, J., Marco, C. A. & Tobin, S. (1991). The role of temporal comparison, social comparison, and direct appraisal in the elderly's self-evaluations of health. *Journal of Applied Social Psychology*, *21*, 1125-1144.
- Tarelli, I., Schwippert, K. & Stubbe, T. C. (2012). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 247-268). Münster: Waxmann.
- Urhahne, D., Timm, O., Zhu, M. & Tang, M. (2013). Sind unterschätzte Schüler weniger leistungsmotiviert als überschätzte Schüler? *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, *45*, 34-43.
- Usher, E. L. & Pajares, F. (2008). Sources of self-efficacy in school. Critical review of the literature and future directions. *Review of Educational Research*, *78*, 751-796.

- Valentine, J. C. & Dubois, D. L. (2005). Effects of self-beliefs on academic achievement and vice versa. Separating the chicken from the egg. In H. W. Marsh, R. Craven & D. M. McInerney (Eds.), *New frontiers for self-research* (International advances in self-research, vol. 2, pp. 53-77). Greenwich, CT: Information Age Publishing.
- Valentine, J. C., Dubois, D. L. & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39, 111-133.
- Vandenberg, R. J. & Lance, C. E. (2000). A Review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-70.
- Wainer, H. (2010). 14 conversations about three things. *Journal of Educational and Behavioral Statistics*, 35, 5-25.
- Walter, O. (2006). Die Entwicklung der mathematischen und der naturwissenschaftlichen Kompetenzen von Jugendlichen mit Migrationshintergrund im Verlauf eines Schuljahres. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 249-275). Münster: Waxmann.
- Weis, M., Zehner, F., Sälzer, C., Strohmaier, A., Artelt, C. & Pfof, M. (2016). Lesekompetenz in PISA 2015: Ergebnisse, Veränderungen und Perspektiven. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (1. Aufl., S. 249-283). Münster: Waxmann.

- Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K. & Kasper, D. (Hrsg.). (2016). *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Williams, T. & Williams, K. (2010). Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology, 102*, 453-466.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*, 82-91.





## **2** **Studie 1: Verbales schulisches Selbstkonzept und sprachliche Leistungen in Gruppen mit und ohne Migrationshintergrund**

Schöber, C., Retelsdorf, J. & Köller, O. (2015). Verbales schulisches Selbstkonzept und sprachliche Leistungen in Gruppen mit und ohne Migrationshintergrund. *Psychologie in Erziehung und Unterricht*, 62(2), 89-105. Verfügbar unter <http://www.reinhardt-journals.de/index.php/peu/article/view/2336>

## Zusammenfassung

In der vorliegenden Arbeit analysieren wir das Zusammenspiel von sprachlichen Leistungen (Leseverstehen und Rechtschreibkompetenzen) und dem verbalen schulischen Selbstkonzept bei Schülerinnen und Schülern mit und ohne Migrationshintergrund. In zwei Stichproben (Kohorte 1:  $N = 1001$  Jugendliche aus 44 Klassen; Kohorte 2:  $N = 855$  Jugendliche aus 38 Klassen) aus unterschiedlichen Schulformen (Gymnasium vs. andere) wurden die Rechtschreib- und Leseleistungen zu Beginn der 5. (T1) und der 7. Jahrgangsstufe (T2) erhoben. Zusätzlich wurde das verbale schulische Selbstkonzept erfasst. Im Einklang mit der existierenden Forschungsliteratur konnten wir zunächst nach Kontrolle der Schulform zu beiden Zeitpunkten signifikant niedrigere Leistungen in der Gruppe mit Migrationshintergrund nachweisen. Keine signifikanten Differenzen ergaben sich im verbalen Selbstkonzept. Die Entwicklung des verbalen Selbstkonzepts und der Leistungen verlief in beiden Gruppen ähnlich. Weitere Auswertungen ergaben reziproke Effekte von sprachlichen Leistungen und verbalem Selbstkonzept, d.h. sprachliche Leistungsmaße zeigten positive Effekte auf die Veränderung des verbalen Selbstkonzepts und ein höheres verbales Selbstkonzept ging mit einer günstigeren Entwicklung sprachlicher Leistungen einher. Die Effekte waren robust in unterschiedlichen Schulformen und in Gruppen mit und ohne Migrationshintergrund. Implikationen für die schulische Praxis werden diskutiert.

## 2.1 Einleitung

Domänenspezifische schulische Selbstkonzepte gelten als zentrale Einflussfaktoren auf schulische Leistungen, d.h. Schülerinnen und Schüler mit einem höheren domänenspezifischen Selbstkonzept weisen üblicherweise höhere Leistungen in den entsprechenden Domänen auf als ihre Klassenkameradinnen und Klassenkameraden und zeigen darüber hinaus in einer längsschnittlichen Betrachtung auch größere Leistungszuwächse (z. B. Marsh & Köller, 2004; Möller, Retelsdorf, Köller & Marsh, 2011). Nicht zuletzt wegen ihrer großen Bedeutung für erfolgreiches Lernen gilt die Steigerung der domänenspezifischen Selbstkonzepte der Schülerinnen und Schüler auch als direktes Ziel schulischer Arbeit (Marsh & Craven, 2006), wobei wenige Arbeiten existieren, die auf schulische Selbstkonzepte in Zuwanderungsgruppen Rücksicht nehmen. In der vorliegenden Arbeit gehen wir daher der Frage nach, wie sich das verbale schulische Selbstkonzept als eine zentrale selbstbezogene Kognition im Kontext von Schule und die sprachlichen Leistungen (Lesen und Rechtschreiben) von Schülerinnen und Schülern unterschiedlicher ethnischer Gruppen (mit Migrationshintergrund vs. ohne Migrationshintergrund) in deutschen Schulen über einen Zweijahreszeitraum entwickeln. Wir prüfen, ob verbale Selbstkonzepte und sprachliche Leistungen bei Schülerinnen und Schülern mit Migrationshintergrund niedriger ausfallen als bei Schülerinnen und Schülern ohne Migrationshintergrund. In einer längsschnittlichen Betrachtung wird zudem die Entwicklung des verbalen Selbstkonzepts wie auch der sprachlichen Leistungen zwischen den Gruppen mit und ohne Migrationshintergrund verglichen. Schließlich analysieren wir die gewonnenen Daten vor dem Hintergrund des Modells reziproker Zusammenhänge (*Reciprocal Effects Model*, REM; vgl. Marsh & Martin, 2011; Retelsdorf, Köller & Möller, 2014), in dem sowohl

positive Effekte von domänenspezifischen Selbstkonzepten auf schulische Leistungen als auch positive Effekte von schulischen Leistungen auf domänenspezifische Selbstkonzepte prognostiziert werden. Wir untersuchen in diesem Zusammenhang, ob die wechselseitigen Effekte zwischen verbalem Selbstkonzept und sprachlichen Leistungen in ihrer Höhe durch den Migrationsstatus (mit vs. ohne Migrationshintergrund) moderiert werden.

Im Folgenden betrachten wir zunächst empirische Ergebnisse zu Schulleistungen in Deutschland mit besonderer Berücksichtigung der Gruppe von Kindern und Jugendlichen mit Migrationshintergrund. Ebenfalls mit Bezug auf Gruppen unterschiedlichen Migrationshintergrunds wird daran anschließend der für die vorliegende Untersuchung relevante Forschungsstand zu domänenspezifischen schulischen Selbstkonzepten zusammengefasst. Zusammenhänge zwischen schulischen Leistungen und schulischen Selbstkonzepten sind Gegenstand des dritten Abschnittes.

### **2.1.1 Schulleistungen von Kindern und Jugendlichen mit Migrationshintergrund**

Vor allem die großen Schulleistungsstudien wie *Programme for International Student Assessment* (PISA; Klieme et al., 2010; Prenzel, Sälzer, Klieme & Köller, 2013), *Internationale Grundschul-Lese-Untersuchung* (IGLU; Bos, Tarelli, Bremerich-Vos & Schwippert, 2012), *Trends in International Mathematics and Science Study* (TIMSS; Bos, Wendt, Köller & Selter, 2012) und die Ländervergleiche des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB; Köller, Knigge & Tesch, 2010; Pant et al. 2013) haben in Deutschland auf der Basis großer, repräsentativer Stichproben von Schülerinnen und Schülern der Grundschule und Sekundarstufe I Leistungsunterschiede zwischen den Gruppen mit und ohne Migrationshintergrund herausgearbeitet. Für die Sekundarstufe I, die im Zentrum des vorliegenden Artikels steht, zeigt sich in PISA 2009 für 15-Jährige,

dass die mittleren Leseleistungen der Jugendlichen mit Migrationshintergrund 44 Punkte unter denen der Jugendlichen ohne Migrationshintergrund liegen, eine Differenz, die einem Leistungsrückstand von über einem Schuljahr entspricht (Stanat, Rauch & Segeritz, 2010). Im IQB-Ländervergleich liegt der Rückstand sogar bei fast zwei Schuljahren (Böhme, Tiffin-Richards, Schipolowski & Leucht, 2010). Die Leistungsdifferenzen zwischen Zuwanderern der 1. und 2. Generation sind dabei gering (leichte Vorteile der Schülerinnen und Schüler der 2. Generation).<sup>1</sup> Unterscheidet man die Zuwanderungsgruppe nach Herkunftsland, so fallen beispielsweise die Leistungen der Schülerinnen und Schüler mit Migrationshintergrund in der ehemaligen Sowjetunion deutlich höher aus als die der Jugendlichen mit türkischem Hintergrund (Differenz von rund 50 Punkten, was einer halben Standardabweichung entspricht). Vertiefte Analysen der aus den Large-Scale-Assessments gewonnenen Daten weisen darauf hin, dass die Sprachpraxis zuhause (wird vorwiegend die Herkunftssprache gesprochen?), die soziale Situation der Familien, das kulturelle Kapital und das Ausmaß der kulturellen Identität (Identifikation mit bzw. Assimilation an das Zuwanderungsland) die Minderleistungen der Schülerinnen und Schüler mit Migrationshintergrund erklären können (z. B. Edele, Stanat, Radmann & Segeritz, 2013). Welche Rolle institutionelle Effekte (z. B. die besuchte Schulform) spielen, ist selten untersucht; wenig untersucht wird auch, ob die Disparitäten im Laufe der Schulzeit zu- oder abnehmen (siehe aber u.a. Mücke & Schröder-Lenzen, 2008 für die Grundschule; Walter, 2006 für die Sekundarstufe I). Gerade für den Bereich der Sekundarstufe I lie-

---

<sup>1</sup>Zuwanderer der 1. Generation: Schülerinnen und Schüler, die selbst zugewandert sind; Zuwanderer der 2. Generation: Schülerinnen und Schüler, die in Deutschland geboren sind, deren Eltern aber zugewandert sind.

gen leider nur wenige empirische Studien zur Entwicklung migrationsbedingter Disparitäten in Deutschland in verbalen Leistungen vor, so dass für die vorliegende Untersuchung auch die Ergebnisse aus anderen Fachbereichen interessant sind, insbesondere die Analysen von Walter (2006). Walter hat die Messwiederholung von PISA 2003 (vgl. PISA-Konsortium Deutschland, 2006) genutzt, um die Leistungsveränderungen im Laufe der 10. Klasse in den Fächern Mathematik und Naturwissenschaften zu untersuchen. Für Mathematik zeigten sich keine Effekte des Migrationshintergrunds auf den Kompetenzzuwachs, in den Naturwissenschaften lagen Migranten der 2. Generation in ihrem Kompetenzzuwachs deutlich zurück, allerdings war die Zugehörigkeit zur 2. Generation erheblich mit dem Zuwanderungsland (überwiegend Türkei) konfundiert. Festzuhalten bleibt die Uneinheitlichkeit der Befunde bei Walter (2006). Inwieweit sich also migrationsbedingte Disparitäten im Laufe der Sekundarstufe I in Deutschland erhöhen, ist damit eine nicht abschließend beantwortete Forschungsfrage. Dies gilt insbesondere für den Bereich der Rechtschreibleistungen.

### **2.1.2 Domänenspezifische Selbstkonzepte und Migrationsstatus**

Domänenspezifische schulische Selbstkonzepte stellen im Wesentlichen das Resultat sozialer, dimensionaler und temporaler schulischer Leistungsvergleiche dar (Marsh, 1984, 1986; Möller & Marsh, 2013; Möller & Köller, 2004). Bei sozialen Vergleichen werden die eigenen Leistungen mit denen der Mitschülerinnen und Mitschüler verglichen; bei dimensionalen Vergleichen vergleicht eine Person ihre eigenen Leistungen in einer Domäne mit denen in einer anderen Domäne; schließlich ist der zeitliche Verlauf von erreichten Leistungen in einer Domäne Gegenstand des temporalen Vergleichs. Schaut man sich die von Marsh und Hau (2003, 2004) durchgeführten Sekundäranalysen

der PISA-2000-Daten oder die Meta-Analyse von Möller, Pohlmann, Köller und Marsh (2009) an, so scheinen diese Prozesse universell in allen OECD-Staaten gleichermaßen aufzutreten, d.h. unabhängig vom kulturellen Kontext, in dem Schülerinnen und Schüler aufwachsen. Überträgt man diese Befunde auf die Situation der Schülerinnen und Schüler mit und ohne Migrationshintergrund in Deutschland, so sollten sich grundlegende Prozesse der Selbstkonzeptgenese in beiden Gruppen ähnlich abspielen. Die wenigen Befunde zu domänenspezifischen Selbstkonzepten von Schülerinnen und Schülern mit Migrationshintergrund (vgl. Roebbers, Mecheril & Schneider, 1998; Shajek, Stanat & Lüdtke, 2006) lassen folgende Schlussfolgerungen zu: Schülerinnen und Schüler mit Migrationshintergrund haben im Vergleich zu ihren Klassenkameradinnen und Klassenkameraden ohne Migrationshintergrund etwas niedrigere Ausprägungen des verbalen Selbstkonzepts bei gleichzeitig leicht höheren Ausprägungen des mathematischen Selbstkonzepts. Diese insgesamt geringen Selbstkonzeptdifferenzen stehen in einem gewissen Widerspruch zu den großen Leistungsdifferenzen zuungunsten der Kinder und Jugendlichen mit Migrationshintergrund.

Das Modell des internen und externen Referenzrahmens (I/E Modell; vgl. Marsh, 1984, 1986) scheint trotz unterschiedlich ausfallender Höhe einzelner Pfadkoeffizienten für deutsche und nicht-deutsche Schülerinnen und Schüler grundsätzlich gleichermaßen zu gelten, d.h. beide Schülergruppen basieren ihre domänenspezifischen Selbstkonzepte in der Tat auf dimensional (interne Referenz) und sozialen (externe Referenz) Leistungsvergleichen (Marsh & Hau, 2003). Das I/E-Modell mag auch erklären, dass die mathematischen Selbstkonzepte bei Schülerinnen und Schülern mit Migrationshintergrund höher ausgeprägt sind als ihre verbalen Selbstkonzepte, dürften die verbalen Leis-

tungen doch im dimensionalen Vergleich mit den mathematischen Leistungen schlechter eingeschätzt werden.

Die insgesamt geringen Differenzen in den Selbstkonzepten zwischen Jugendlichen mit und ohne Migrationshintergrund sind mit der klassischen Theorie der sozialen Vergleichsprozesse von Festinger (1954) allerdings kaum vereinbar, müssten doch die substanziell geringeren Leistungen in den verschiedenen Fächern auch zu niedrigeren Ausprägungen der Selbstkonzepte in der Gruppe mit Migrationshintergrund führen. Plausibel werden die Ergebnisse erst wieder unter Berücksichtigung der relevanten Attribute von Vergleichspersonen (Goethals & Darley, 1977). In diesem Ansatz ergibt sich eine erbrachte Leistung aus den (mehr oder weniger unveränderlichen) Fähigkeiten einer Person und leistungsrelevanten Attributen wie Anstrengung, Erfahrung, Alter usw. Folgt dementsprechend eine Person ihrem Bedürfnis nach Selbstbewertung und will einschätzen, wie hoch ihre eigenen Fähigkeiten ausgeprägt sind, so wählt sie Vergleichspersonen nicht nur nach deren Leistungen sondern auch über andere leistungsrelevante Attribute aus. Sofern der Migrationsstatus ein für Vergleiche relevantes Attribut ist, sollten sich Schülerinnen und Schüler mit Migrationshintergrund auch eher Vergleichspersonen aus der Gruppe mit Zuwanderungsgeschichte zu suchen. Da diese Gruppe insgesamt leistungsschwächer ist, kann dies letztendlich zu einer Höhe der Selbstkonzepte führen, wie sie auch in der Gruppe ohne Migrationshintergrund aufzufinden ist. Stimmt diese Annahme, so sollte sich auch ein Zusammenhangsmuster zwischen Leistungen und Selbstkonzepten zeigen, das in den Gruppen mit und ohne Migrationsgeschichte identisch ist.

Wiederholt wurde untersucht, ob sich das Zusammenhangsmuster zwischen fachspezifischen Selbstkonzepten und Leistungen robust in Subpopulationen zeigt. Hin-



sichtlich des moderierenden Effektes des Geschlechts ergab die Metaanalyse von Valentine, DuBois und Cooper (2004) keine Evidenz, dass die Zusammenhänge zwischen selbstbezogenen Kognitionen und Leistungen zwischen Jungen und Mädchen variieren. Valentine et al. haben auch versucht, den potenziell moderierenden Effekt des ethnischen Hintergrundes der Schülerinnen und Schüler zu untersuchen. Allerdings lagen und liegen zu wenige Studien vor, die es erlauben, der Frage nachzugehen, ob die Zusammenhangsmuster zwischen selbstbezogenen Kognitionen und Leistungen in unterschiedlichen ethnischen Gruppen gleichermaßen gelten. Die Autoren haben sich dann in ihren Analysen der amerikanischen Studien darauf beschränkt, den Anteil der weißen Schülerinnen und Schüler als Moderator aufzunehmen. Hier ergab sich kein Effekt auf das untersuchte Zusammenhangsmuster.

### **2.1.3 Das Modell reziproker Zusammenhänge**

Die Frage nach dem Zusammenspiel zwischen domänenspezifischen Selbstkonzepten und schulischen Leistungen beschäftigt die Forschung seit vielen Jahrzehnten. Vor bald 40 Jahren haben Calsyn und Kenny (1977) die Self-Enhancement- und die Skill-Development-Hypothese gegenübergestellt. Erstere besagt, dass domänenspezifische Selbstkonzepte einen positiven Effekt auf entsprechende Leistungen haben, da ein hohes Selbstkonzept u.a. zu einer größeren Anstrengungsbereitschaft führt. Letztere nimmt an, dass Leistungsergebnisse zu sozialen Vergleichen führen und darüber vermittelt Effekte auf domänenspezifische Selbstkonzepte haben (vgl. Köller, 2004; Möller et al., 2009). Besonders leistungsstarke Schülerinnen und Schüler führen eher Abwärtsvergleiche durch, was ihr Selbstkonzept erhöht, leistungsschwache eher Aufwärtsvergleiche, die ihr Selbstkonzept senken. Mittlerweile wurden viele empirische Studien zusammengetragen,

welche die Skill-Development- und Self-Enhancement-Hypothese im Modell reziproker Zusammenhänge (REM) zusammenführen (vgl. auch Abbildung 1; Marsh & Craven, 2006; Marsh & O'Mara, 2008; Retelsdorf et al., 2014). In diesem Modell, das nur in längsschnittlichen Forschungsdesigns überprüfbar ist, werden zunächst hohe Stabilitäten des domänenspezifischen Selbstkonzepts und der entsprechenden Leistungen über die Zeit angenommen. Darüber hinaus wird ein Effekt des zeitlich vorgeordneten Selbstkonzepts auf die zeitlich spätere Leistung nach Kontrolle der Vorleistung postuliert, ebenso wie ein Effekt der zeitlich vorgeordneten Leistung auf das spätere Selbstkonzept, nachdem der Effekt des früheren Selbstkonzepts kontrolliert wurde. Schließlich werden innerhalb der Messzeitpunkte positive Korrelationen zwischen Leistung und Selbstkonzept angenommen.

Auch in Längsschnittstudien wurden Moderatoren des Zusammenhangs zwischen Selbstkonzepten und Leistungen untersucht. Dabei ergaben sich bei Untersuchungen verschiedener Altersgruppen uneinheitliche Ergebnisse. Skaalvik und Hagtvet (1990) untersuchten zwei Kohorten zu jeweils zwei Messzeitpunkten und fanden, dass bei Schülerinnen und Schülern in der zweiten bis zur dritten Klassenstufe der Effekt von der Leistung auf das Selbstkonzept stärker zu sein scheint, was die Skill-Development-Hypothese in dieser Altersgruppe unterstützt. In der zweiten Kohorte mit Schulkindern von der fünften bis zur sechsten Klassenstufe zeigten sich die Effekte in jeweils ähnlicher Höhe, so dass sich Leistungen und Selbstkonzept reziprok beeinflussten. Hingegen zeigten sich bei Guay, Marsh und Bovin (2003) in drei unterschiedlichen Kohorten mit erstem Messzeitpunkt in der zweiten, dritten bzw. vierten Klasse jeweils zu den ersten beiden von drei im einjährigen Abstand erfolgten Messzeitpunkten reziproke Effekte zwischen Selbst-

konzept und Leistungen. Zum jeweiligen dritten Messzeitpunkt stützten die Ergebnisse jedoch die Self-Enhancement-Hypothese, da nur ein Effekt des Selbstkonzepts auf die Leistung signifikant wurde.

Damit bleibt insgesamt festzuhalten, dass mittlerweile belastbare Befunde zu Schulleistungen und schulischen Selbstkonzepten von Schülerinnen und Schülern vorliegen, in wenigen Studien auch zu Schülerinnen und Schülern mit und ohne Migrationshintergrund (z. B. Roeber et al., 1998; Edele et al., 2013; Shajek et al., 2006). In diesen wenigen Arbeiten konnte gezeigt werden, dass Schülerinnen und Schüler mit Migrationshintergrund geringere sprachliche Leistungen aufweisen und sich diese Sprachdefizite auch auf Leistungen in anderen Fächern auswirken. Die wenigen längsschnittlichen Befunde weisen darauf hin, dass die migrationsbedingten Disparitäten in den Leistungen zumindest in der Sekundarstufe I eher stabil bleiben und nicht zunehmen (Walter, 2006).

Die verbalen schulischen Selbstkonzepte der Schülerinnen und Schüler mit Migrationshintergrund fallen in ihrer Höhe geringer aus als ihre mathematischen. Erklärt wird dies über das I/E-Modell (Marsh, 1984, 1986; Möller & Köller, 2004), demzufolge individuelle Vergleiche zu einem Absinken des verbalen zugunsten des mathematischen Selbstkonzepts bei Kindern und Jugendlichen mit Migrationshintergrund führen könnten (Shajek et al., 2006).

Bemerkenswert ist allerdings auch, dass die Leistungsabstände zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund üblicherweise deutlich größer sind als die Abstände im verbalen Selbstkonzept. Wir vermuten hier, dass dies eine Folge der relevanten Attribute der gewählten Partner für soziale Vergleiche ist. Berücksichtigen Schülerinnen und Schüler bei der Wahl der Vergleichsperson relevante Attribute (Goe-

thals & Darley, 1977), so ist zu erwarten, dass diejenigen mit Migrationshintergrund auch eher Klassenkameraden mit nicht-deutscher Herkunft für den sozialen Aufwärtsvergleich suchen. Da deren Leistungen relativ schwächer sind, sollte dieser Vergleich geringere Kosten haben als der Vergleich mit einer Person deutscher Herkunft. Dessen unbenommen finden aber soziale Vergleiche statt, so dass für Schülerinnen und Schüler mit Migrationshintergrund ebenso wie für solche ohne Migrationshintergrund steigende Leistungen und damit ein günstiger sozialer Vergleich auch zu einem hohen Selbstkonzept und ein hohes Selbstkonzept auch zu höheren Leistungen führen sollte.

## **2.2 Die vorliegende Untersuchung**

Unter Berücksichtigung der zuvor geschilderten Erwägungen zu Leistungen und domänenspezifischen Selbstkonzepten in Gruppen mit und ohne Migrationshintergrund wurden folgende Voraussagen getroffen:

- Schülerinnen und Schüler mit Migrationshintergrund zeigen geringere verbale Leistungen, auch wenn für die besuchte Schulform kontrolliert wird.
- Das verbale Selbstkonzept der Schülerinnen und Schüler mit Migrationshintergrund sollte niedriger ausfallen als das ihrer Klassenkameradinnen und Klassenkameraden ohne Migrationshintergrund.
- Diese Differenz im verbalen Selbstkonzept sollte allerdings deutlich geringer ausfallen als die in den Leistungen.
- Es sollte keine Zunahme der Differenzen im verbalen Selbstkonzept zwischen Schülerinnen und Schülern ohne Migrationshintergrund und denen mit Migrationshintergrund beobachtbar sein.

- Ebenso sollten keine Unterschiede zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund in der Veränderung der verbalen Leistungen nachweisbar sein.
- Zwischen dem verbalen Selbstkonzept und den sprachlichen Leistungen sollten reziproke Zusammenhänge in beiden Gruppen (mit vs. ohne Migrationshintergrund) beobachtbar sein: Ein höheres verbales Selbstkonzept sollte leistungsfördernd sein und höhere Leistungen sollten auch zu einem höheren verbalen Selbstkonzept führen.

### 2.3 Methode

Für die vorliegende Arbeit wurden Daten aus dem Projekt *Entwicklung und Implementierung eines neuen Konzeptes zur Eingliederung Jugendlicher in die Berufs- und Arbeitswelt in Schulen mit erhöhtem Förderbedarf* (EIKA) verwendet. Es handelt sich dabei um eine Initiative der Stadt Bremen, in der seit 2004 versucht wird, Schulen in schwierigen sozialen Lagen mit Maßnahmen der Schul- und Unterrichtsentwicklung zu unterstützen. Ziel ist die erfolgreiche Einfädung benachteiligter Jugendlicher in die berufliche Erstausbildung. Das Projekt wurde in einem Mehrkohorten-Längsschnittsdesign von 2004 bis 2008 wissenschaftlich begleitet. Eine genaue Beschreibung des Untersuchungsplans findet sich bei Eßel-Ullmann (2008) und Köller (2009). Die Daten dieser Studie eignen sich besonders gut für die Beantwortung unserer Forschungsfragen, da die Stichproben erhebliche Anteile von Schülerinnen und Schülern mit Migrationshintergrund enthalten.

Zur Beantwortung unserer Fragestellung wurden zwei Kohorten aus dem EIKA-Projekt berücksichtigt. Bei der Kohorte 1 handelt es sich um  $N = 1001$  Schülerinnen (50%) und Schüler aus 44 Klassen der 5. Jahrgangsstufe (T1), die zu Beginn des Schuljahres 2004/2005 untersucht wurden. 30.6 Prozent besuchten ein Gymnasium, die übrigen 69.4 Prozent Sekundar- oder Gesamtschulen. Insgesamt berichteten 47.1 Prozent der Schülerinnen und Schüler, einen Migrationshintergrund zu haben. In der großen Mehrzahl (86.3%) waren die Eltern im Ausland, die Kinder selbst in Deutschland geboren. Eine zweite Erhebung (T2) fand zu Beginn der 7. Jahrgangsstufe im Schuljahr 2006/2007 statt. Die zweite Kohorte besteht aus  $N = 855$  Schülerinnen (51.2%) und Schülern aus 38 Klassen der 5. Jahrgangsstufe, die parallel zur Kohorte 1 zu Beginn des Schuljahres 2006/2007 (T1) getestet wurden. 33.5 Prozent der Schülerinnen und Schüler besuchten ein Gymnasium, die übrigen eine Sekundar- oder Gesamtschule, 50.3 Prozent der Schülerinnen und Schüler kamen aus Familien mit Migrationshintergrund. Wiederum gaben die meisten Kinder an (86.2%), dass ihre Eltern im Ausland, sie selbst in Deutschland geboren wurden. Die zweite Datenerhebung (T2) in dieser Stichprobe fand zu Beginn der 7. Jahrgangsstufe im Schuljahr 2008/2009 statt.

Die Klassen beider Kohorten stammten aus denselben elf Schulen. In beiden Kohorten wurden standardisierte Lese- und Rechtschreibtests sowie Schüler- und Elternfragebögen eingesetzt. Die Testung erfolgte durch vorher geschulte Lehrkräfte an jeweils zwei Tagen in je zwei Schulstunden.

### **2.3.1 Instrumente**

Zur Erfassung der Lese- und Rechtschreibleistungen kamen in beiden Kohorten breit validierte Instrumente aus der Hamburger Lernausgangslagenuntersuchung LAU

(vgl. Lehmann & Peek, 1997; Lehmann, Peek, Gänsfuß & Husfeldt, 2002) zum Einsatz. Die Tests zum Leseverstehen beanspruchten zu beiden Messzeitpunkten eine Bearbeitungszeit von einer Schulstunde (45 Minuten) und enthielten Sachtexte ebenso wie literarische Texte. Pro Text wurden 4 bis 7 Fragen im Multiple-Choice-Antwortformat gestellt. Die Tests erreichten in beiden Kohorten zu beiden Messzeitpunkten Reliabilitäten (KR-20) über .80. Die Korrelationen mit der Deutschnote (Kohorte 1:  $r = -.58$  zu T1;  $r = -.54$  zu T2; Kohorte 2:  $r = -.52$  zu T1;  $r = -.46$  zu T2; alles  $ps < .001$ ) sprachen für eine zufriedenstellende Validität. Eine Teilmenge der Aufgaben (sogenannte Ankeritems) wurde in der 5. und der 7. Jahrgangsstufe eingesetzt. Aufgrund dieser Anker-Item-Designs (zwischen den Messzeitpunkten; vgl. Hambleton & Swaminathan, 1989) war es möglich, mit Hilfe von probabilistischen Testmodellen die Leistungen auf einer gemeinsamen Metrik abzutragen. Die Personenparameter (Weighted Maximum Likelihood Estimates) wurden an der ersten Kohorte standardisiert zu T1 ( $M = 100$ ,  $SD = 30$ ). Die gemeinsame Metrik zu T1 und T2 erlaubt eine Abschätzung der Leistungszuwächse über zwei Schuljahre.

Zur Testung der Rechtschreibleistungen wurde ebenfalls ein LAU-Test („Die Geheimnisse unserer Schule“) zu beiden Erhebungszeitpunkten eingesetzt. Es handelt sich dabei um einen fortlaufenden Text, bei dem unter jedem Wort ein kleines Kästchen ist, das anzukreuzen ist, sofern das darüber stehende Wort falsch geschrieben ist. Der Text besteht insgesamt aus 307 Wörtern, von denen 31 falsch geschrieben sind. Der Testwert ist definiert als

$$RS = \frac{(\text{Anzahl richtig angekreuzter Fehler})^2}{\text{Anzahl insgesamt angekreuzter Wörter}}$$

Durch diese Operationalisierung werden die richtig identifizierten Fehler im Text an der Zahl der insgesamt angekreuzten Wörter relativiert. Kreuzt also ein Schüler

bzw. eine Schülerin alle 31 falsch geschriebenen Wörter an und kein Weiteres, so erhält er bzw. sie den Wert 31 ( $312/31 = 31$ ). Kreuzt er oder sie viel mehr Wörter an, so sinkt der Wert entsprechend.

Anders als beispielsweise in einem Diktat müssen Schülerinnen und Schüler in diesem Test keine aktive Phonem-Graphem-Zuordnung vornehmen. Angesprochen werden vielmehr grammatische Kompetenzen (z. B. Groß- und Kleinschreibung) und das lexikalische Gedächtnis, in dem Wörter in ihrer Schreibweise abgespeichert sind. Lehmann et al. (1997) berichten zufriedenstellende Reliabilitäten ( $> .80$ ) für den Test. Die hinreichende Validität des Tests zeigte sich in unserer Untersuchung in substanziellen Korrelationen der Testleistungen mit den Deutschnoten der Schülerinnen und Schüler (Kohorte 1:  $r = -.55$ , zu T1;  $r = -.50$  zu T2; Kohorte 2:  $r = -.58$  zu T1,  $r = -.44$  zu T2; alle  $ps < .001$ ).

Das verbale Selbstkonzept wurde mit Hilfe von fünf Items untersucht (Beispielitem: *Deutsch würde ich lieber machen, wenn das Fach nicht so schwer wäre*, vierstufiges Antwortformat von 1 = *trifft voll und ganz zu* bis 4 = *trifft überhaupt nicht zu*). Die Skala hat sich in vielen empirischen Untersuchungen als hinreichend reliabel und valide erwiesen (vgl. z. B. Marsh & Köller, 2004; Möller & Köller, 2001). In der vorliegenden Untersuchung lag die Reliabilität (Cronbachs  $\alpha$ ) zu T1 bei .85, zu T2 bei .87. Die Skalenbildung erfolgte durch Aufsummierung der fünf Items.



### 2.3.2 Statistische Analysen

Die erhobenen Daten stammen aus geschachtelten Stichproben (Schülerinnen und Schüler innerhalb von Klassen). Wird diese Datenstruktur nicht bei den statistischen Analysen berücksichtigt, so können die Standardfehler statistischer Parameter unterschätzt werden. Diese Verzerrung steigt mit Zunahme der Intraklassenkorrelation (*ICC*), d.h. der Varianz zwischen den Klassen (zum Ausmaß der *ICCs*. Tabelle 3). Da zudem Prädiktoren auf unterschiedlichen Ebenen vorlagen (Schulform auf Klassenebene, die übrigen Variablen auf Individualebene), wurden in *Mplus* (Version 6.12, Muthén & Muthén, 2010) im Falle der Berücksichtigung der Schulformvariable Mehrebenenanalysen durchgeführt. Wurden allein Individualmerkmale analysiert, so wurden die Standardfehler in *Mplus* korrigiert (Befehl: Analysis: Type = Complex;). Die Beurteilung der Modellgüte erfolgte mit Hilfe des Root Mean Square Error of Approximation (RMSEA), des Comparative Fit Index (CFI) und des Tucker-Lewis-Index (TLI). Zusätzlich haben wir zur Beurteilung der Modellgüte den Quotienten aus der  $\chi^2$ -Statistik und der Zahl der Freiheitsgrade des jeweils getesteten Modells gebildet (vgl. Marsh, 2007 zu diesen verschiedenen Fit-Statistiken). Eine sehr gute Modellanpassung liegt vor, wenn der CFI und der TLI Werte über .95 erreichen und der RMSEA kleiner oder gleich .06 liegt. Der Quotient aus  $\chi^2$ -Statistik und der Zahl der Freiheitsgrade ist sehr stark stichprobenabhängig. Wir betrachten im Folgenden Quotienten unter vier als Hinweise auf einen guten Modellfit.

Da bei allen Variablen fehlende Werte auftraten (je nach Variable bis zu 25%), wurden in *Mplus* mittels multipler Imputation 50 vollständige Datensätze generiert (zur Behandlung von fehlenden Werten vgl. Lüdtke, Robitzsch, Trautwein & Köller, 2007). Die entsprechenden Analysen wurden dann für jeden der 50 Datensätze durchgeführt und

die Ergebnisse entsprechend dem Vorgehen bei Rubin (1987) automatisch in *Mplus* gemittelt.

## 2.4 Ergebnisse

Sofern man von Invarianz der Parameter in beiden Kohorten ausgehen kann, vereinfacht sich die Darstellung der Ergebnisse durch die gemeinsame Betrachtung beider Kohorten. In einem in *Mplus* analysierten Zweigruppenmodell wurden deshalb zunächst für alle Variablen die Mittelwerte, Streuungen und Kovarianzen auf Gleichheit in beiden berücksichtigten Kohorten getestet. Die gefundenen Fitstatistiken ( $\chi^2/df = 1.615$ ; RMSEA = .025; CFI = .999; TLI = .999) ergaben eine klare Unterstützung für die Annahme identischer Parameter in beiden Kohorten.

Weiterhin wurde geprüft, ob für die fünf Selbstkonzeptitems die Invarianz der Messmodelle (identische Faktorladungen und Schwellenparameter) in beiden Kohorten und über die Zeit gegeben war. Die Invarianz der Messmodelle über die Zeit ist eine notwendige Voraussetzung für die unten berichteten Analysen zum reziproken Zusammenhang von Selbstkonzept und Leistung. Wiederum sprachen die Fit-Statistiken für die Invarianzannahme ( $\chi^2/df = 3.570$ ; RMSEA = .053; CFI = .985; TLI = .988). Aufgrund der Ergebnisse wurde bei allen weiteren Analysen die Kohortenzugehörigkeit nicht mehr berücksichtigt.

### 2.4.1 Deskriptive Befunde

Tabelle 2.1 zeigt deskriptive Ergebnisse für die berücksichtigten Variablen aufgebrochen nach Schulform (andere vs. Gymnasium), Migrationshintergrund (ohne vs. mit) und Messzeitpunkt (T1 vs. T2). Alle Befunde sind gemittelt über 50 imputierte Da-

tensätze. Paarweise Mittelwertsdifferenzen wurden in Mplus auf Signifikanz getestet. Wegen der vielen Einzelvergleiche und der damit verbundenen Inflation von  $\alpha$ -Fehlern wurde ein Signifikanzniveau von  $p < .001$  festgelegt. Erwartungskonform liegen die Leistungen der Schülerinnen und Schüler mit Migrationshintergrund in beiden Schulformen niedriger als die der Schülerinnen und Schüler ohne Migrationshintergrund. Alle paarweise getesteten Mittelwertsdifferenzen sind im Einklang mit unseren Hypothesen signifikant ( $p < .001$ ), die Effektstärken schwanken zwischen  $d = 0.26$  (Rechtschreiben zu T1 in nichtgymnasialen Bildungsgängen) und  $d = 0.77$  (Lesen zu T1 im Gymnasium).

Die verbalen Selbstkonzepte der Schülerinnen und Schüler mit Migrationshintergrund liegen zwar durchgängig etwas niedriger, keine Mittelwertsdifferenz wird allerdings im Einklang mit der Literatur und unseren Hypothesen signifikant.

----- Tabelle 2.1 in etwa hier -----

Hinsichtlich der Schulformunterschiede in den Leistungen ergibt sich das erwartete Bild. Schülerinnen und Schüler an Gymnasien zeigen durchgängig signifikant höhere Leistungen (alle Mittelwertsdifferenzen  $p < .001$ ), die Effektstärken schwanken zwischen  $d = 0.94$  (T1 Rechtschreibleistungen bei Schülerinnen und Schülern mit Migrationshintergrund) und  $d = 1.10$  (T2 Rechtschreiben bei Schülerinnen und Schülern ohne Migrationshintergrund). Beim verbalen Selbstkonzept ergeben sich signifikante Schulformeffekte zu T1, d.h. Schülerinnen und Schüler an Gymnasien liegen generell signifikant höher als Schülerinnen und Schüler an Nicht-Gymnasien ( $d = 0.61$  ohne Migrationshintergrund;  $d = 0.52$  mit Migrationshintergrund). Zu T2 verschwinden diese Unterschiede, alle durchgeführten Mittelwertvergleiche verfehlen das Signifikanzniveau. Die konvergierenden Mittelwerte in den Schulformen sind Ausdruck des Fischteich-Effekts (vgl.

Marsh, 1987; Köller, 2004). Die verbalen Selbstkonzepte zu T1 speisen sich noch aus der Zeit der Grundschule, in der soziale Vergleiche der späteren Gymnasiastinnen und Gymnasiasten sehr günstig, die der übrigen Schülerinnen und Schüler vergleichsweise ungünstig ausfielen. Die neue Bezugsgruppe in der Sekundarstufe I sorgt dafür, dass das verbale Selbstkonzept am Gymnasium sinkt, im nichtgymnasialen Bereich dagegen steigt.

Schließlich belegt die Tabelle 2.1, dass die Leistungen im Rechtschreiben und im Lesen in allen Gruppen deutlich angestiegen sind (signifikante Zuwächse zwischen  $d = 0.68$ ; Lesen in nichtgymnasialen Bildungsgängen bei Schülerinnen und Schülern ohne Migrationshintergrund; und  $d = 1.23$ ; Rechtschreiben im Gymnasium bei Schülerinnen und Schülern mit Migrationshintergrund).

#### 2.4.2 Zusammenhänge zwischen Leistungen und Selbstkonzept

Um Zusammenhänge zwischen den Leistungs- und Selbstkonzeptmaßen festzustellen, wurden Interkorrelationen berechnet. Die Tabelle 2.2 zeigt die entsprechenden Befunde. Alle Koeffizienten sind hoch signifikant ( $p < .001$ ).

----- Tabelle 2.2 in etwa hier -----

Für die Leistungsmaße ergeben sich deutlich höhere Stabilitäten ( $r = .725$  im Lesen;  $r = .676$  im Rechtschreiben) als für das verbale Selbstkonzept ( $r = .289$ ), letzteres mag dem Umstand geschuldet sein, dass die neue Bezugsgruppe in der 5. Klasse zu Beginn der Sekundarstufe I noch zu erheblichen interindividuellen Differenzen in den Veränderungen des Selbstkonzepts führte.

Die Leistungsmaße korrelieren höher miteinander als mit dem Selbstkonzept, die weiteren Zusammenhänge fallen in einer Höhe aus, die kongruent zu früheren Befunden ist (vgl. u.a. Möller & Köller, 2004).

### 2.4.3 Veränderungsanalysen

Im nächsten Schritt wurden Mehrebenenanalysen durchgeführt, um Effekte des Migrationshintergrunds und der Schulformzugehörigkeit auf die Leistungs- und Selbstkonzeptveränderung zu identifizieren. Da vielfach Ausgangsunterschiede zwischen den Gruppen der Migranten und Nicht-Migranten bestanden, wurde ein kovarianzanalytisches Vorgehen bevorzugt, bei dem die T1-Maße als Kovariaten bei der Vorhersage der T2-Maße verwendet wurden. In Tabelle 3 sind standardisierte Regressionsgewichte aus entsprechenden Mehrebenenanalysen in *Mplus* 6.12 aufgeführt. Für beide Leistungsbereiche ergibt sich eine homogene Befundlage mit hohen Stabilitäten (Effekt der Leistung zu T1), substantziellen Effekten der Schulform und sehr schwachen und nur in einem Fall (Rechtschreiben) signifikanten Koeffizienten des Migrationsstatus. Die anfänglich bestehenden Leistungsunterschiede zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund bleiben im Lesen über die Zeit stabil und nehmen im Rechtschreiben leicht zu.

----- Tabelle 2.3 in etwa hier -----

Hinsichtlich der Effekte auf das Selbstkonzept zu T2 wird allein das Regressionsgewicht des verbalen Selbstkonzepts zu T1 signifikant, Schülerinnen und Schüler mit höherem Selbstkonzept zu T1 weisen dies auch eher zu T2 auf. Das Regressionsgewicht der Schulform wird zwar im Sinne des Fischteicheffekts negativ, im Gegensatz zu den paarweisen Tests der Mittelwertsdifferenzen (s.o.) wird der Effekt aber nicht signifikant. Keine Unterschiede in der Selbstkonzeptveränderung ergeben sich erwartungskonform zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund.

#### **2.4.4 Reziproke Effekte zwischen dem verbalen Selbstkonzept und sprachlichen Leistungen**

In den abschließenden Analysen gingen wir der Frage nach den längsschnittlichen Zusammenhängen zwischen dem verbalen Selbstkonzept und den Rechtschreib- und Leseleistungen nach. Auf der Basis der existierenden Literatur zum REM (z. B. Retelsdorf et al., 2014) erwarteten wir zunächst einmal für den Bereich Lesen wechselseitige Effekte, d.h. nach Kontrolle der Leistung zu T1 wurde ein Effekt des Selbstkonzepts zu T1 auf die Leistung zu T2 erwartet, ebenso wie ein Effekt der Leistung zu T1 auf das Selbstkonzept zu T2 nach Kontrolle des Selbstkonzepts zu T1. Obwohl zum Bereich der Orthografie bislang belastbare Befunde fehlen, erwarteten wir ein analoges Befundmuster. In Mehrgruppenanalysen gingen wir zudem der Frage nach, ob das Zusammenhangsmuster invariant in den unterschiedlichen Schulformen und in den beiden Gruppen mit unterschiedlichem Migrationsstatus (mit vs. ohne Migrationshintergrund) ist. In Analogie zur bisherigen Forschung zum REM wurden die Messmodelle der Selbstkonzeptvariablen berücksichtigt und ein Modell getestet, das in Abbildung 2.1 dargestellt ist. Durch die korrelierenden Fehler der Indikatoren wurde eine Überschätzung der Stabilität des Selbstkonzepts vermieden.

----- Abbildung 2.1 in etwa hier -----

Die Tabelle 2.4 zeigt die Fit-Indizes aus Mehrgruppen-Strukturgleichungsmodellen zur Überprüfung des REM für Lesen und Rechtschreiben, in denen die Mess- und Strukturmodelle invariant gesetzt wurden. Für alle vier getesteten Modelle weisen die Gütestatistiken aus, dass die Hypothese invarianter Modellparameter,

sowohl in den Schulformen als auch in Abhängigkeit des Migrationsstatus beibehalten werden kann.

----- Tabelle 2.4 in etwa hier -----

Die Abbildung 2.2 zeigt das Zusammenhangsmuster zwischen dem verbalen Selbstkonzept und den Leistungen, wie es im Modell mit invarianten Parametern bei Schülerinnen und Schülern mit und ohne Migrationshintergrund gefunden wurde. Vor dem Schrägstrich befinden sich die voll standardisierten Koeffizienten für die Rechtschreibleistungen, hinter dem Schrägstrich für die Leseleistungen. Das Messmodell ist aus Übersichtlichkeitsgründen weggelassen.

----- Abbildung 2.2 in etwa hier -----

Insgesamt ergibt sich klare Evidenz für reziproke Zusammenhänge in beiden Gruppen (Migranten vs. Nicht-Migranten), wenn auch die Koeffizienten vom Selbstkonzept auf die Leistung und vice versa eher klein ausfallen, dies gilt insbesondere für den Pfad von der Leseleistung zu T1 auf das verbale Selbstkonzept zu T2 (.03). Zu bedenken ist hier allerdings auch, dass das Selbstkonzeptmaß sich nicht explizit auf das Lesen bzw. Rechtschreiben bezieht.

## 2.5 Diskussion

In der vorliegenden Untersuchung gingen wir Fragen zum Zusammenspiel von sprachlichen Leistungen und verbalen Selbstkonzepten bei Schülerinnen und Schülern mit und ohne Migrationshintergrund nach. Neben der Beantwortung der Frage, wie sich das Selbstkonzept und die Leistungen in Abhängigkeit vom Migrationsstatus entwickeln, wurde untersucht, ob sich das Modell wechselseitiger Zusammenhänge zwischen Selbstkonzepten und Leistungen (REM) in der Gruppe mit Migrationshintergrund ebenso bele-

gen lässt wie in der Gruppe ohne Migrationshintergrund. Mit diesen Analysen kamen wir einem Desiderat der Metaanalyse von Valentine et al. (2004) nach. Insgesamt entsprechen die Befunde unseren Vorhersagen. So zeigten Schülerinnen und Schüler mit Migrationshintergrund geringere Leistungen (auch wenn die besuchte Schulform kontrolliert wurde), allerdings war die Leistungsentwicklung der Gruppe ohne Migrationshintergrund vergleichbar. Das verbale Selbstkonzept fiel in der Gruppe mit Migrationshintergrund kaum niedriger aus, auch die Entwicklung verlief nicht ungünstiger als in der Gruppe ohne Migrationshintergrund. Schließlich zeigten die Analysen zum REM die hohe Robustheit des Zusammenspiels von domänenspezifischen Selbstkonzepten und Leistungen. Die gefundenen Modellparameter variierten weder zwischen Migrationsgruppen noch zwischen Schulformen.

Im Folgenden sollen die Befunde vor dem Hintergrund der rezenten Literatur diskutiert werden, bevor wir auf Grenzen unserer Untersuchung und praktische Implikationen der Ergebnisse eingehen.

### **2.5.1 Sprachliche Leistungen in Abhängigkeit vom Migrationsstatus**

Unsere Befunde zu den Leistungen der Schülerinnen und Schüler mit Migrationshintergrund bestätigen frühere Befunde (z. B. Stanat et al., 2010), wonach diese im Vergleich zu Schülerinnen und Schülern ohne Migrationshintergrund schwächere sprachliche Leistungen erreichen. Indem wir in unseren Analysen die Schulform kontrollierten und zudem die Leistungsentwicklung über zwei Jahre in der Sekundarstufe I berücksichtigen konnten, geht unser Beitrag über die Befunde der meisten großen Large-Scale-Assessments hinaus und erlaubt ähnlich wie die Arbeit von Walter (2006) die Beantwortung der Frage, ob migrationsbedingte Disparitäten im Laufe der Schulzeit ansteigen.



Weder für die Lese- noch für die Rechtschreibkompetenzen ergaben sich hier in der Leistungsentwicklung substantielle Unterschiede zwischen den Schülergruppen mit und ohne Migrationshintergrund. Der signifikante Effekt im Rechtschreiben war sehr klein und praktisch unbedeutend. Die Befunde implizieren, dass sich die migrationsbedingten Disparitäten in der Sekundarstufe I nicht erhöhen, zumindest nicht in der Phase von der 5. bis zur 7. Jahrgangsstufe. Im Übrigen passen die Befunde sehr gut zu längsschnittlichen amerikanischen Studien, in denen wiederholt gezeigt wurde, dass Leistungsunterschiede zwischen benachteiligten und privilegierten Schülergruppen in Phasen der Beschulung gleich bleiben; wenn sie überhaupt zunehmen, dann in den Schulferien (vgl. Alexander & Entwisle, 1996; Entwisle & Alexander, 1992).

Allerdings nahmen die Disparitäten auch nicht ab, d.h. es gelang den beteiligten Schulen nicht, die Nachteile der Schülerinnen und Schüler mit Migrationshintergrund zu Beginn der Sekundarstufe I zu reduzieren, und zwar weder im gymnasialen noch im nichtgymnasialen Bereich. Der Abbau herkunftsbedingter Disparitäten scheint demnach weiterhin ein Desiderat der unterschiedlichen Bildungsgänge zu sein.

### **2.5.2 Verbale Selbstkonzepte und Migrationsstatus**

Erwartungskonform zeigten unsere Befunde, dass die verbalen Selbstkonzepte der Jugendlichen mit Migrationshintergrund nur unwesentlich und nicht statistisch signifikant unter denen der Schülerinnen und Schüler ohne Migrationshintergrund lagen. Die Befunde decken sich *cum grano salis* mit denen von Shajek et al. (2006) und implizieren, dass die geringeren Leistungen der Schülerinnen und Schüler mit Migrationshintergrund sich nicht entsprechend in den verbalen Selbstkonzepten zeigen. Wir hatten im Theorie- teil mit der Arbeit von Goethals und Darley (1977) argumentiert, wonach sich Personen

bei sozialen Vergleichen an relevanten Attributen der Vergleichspersonen orientieren. Es ist dementsprechend plausibel anzunehmen, dass die Jugendlichen aus der Gruppe mit Zuwanderungsgeschichte eher Mitschülerinnen und Mitschüler zu ihren Vergleichen auswählen, die ebenfalls einen Migrationshintergrund haben. In der Tat war der Anteil der Schülerinnen und Schüler mit Migrationshintergrund in der vorliegenden Untersuchung sehr groß (rund 50%), so dass entsprechende Vergleichsmöglichkeiten gegeben waren. Andernorts ist angemerkt worden, dass die wiederholten schulischen Misserfolge der Kinder und Jugendlichen aus Zuwanderungsfamilien Kosten für das Selbstkonzept haben sollten (vgl. Shajek et al., 2006). Unsere Befunde sprechen dafür, dass dem in Klassen mit relativ hohem Anteil von Schülerinnen und Schülern mit Migrationshintergrund nicht so sein muss. Gleichzeitig gilt hier aber auch zu bedenken, dass bei sehr hohen Anteilen von Kindern bzw. Jugendlichen mit Zuwanderungsgeschichte die Lernbedingungen ungünstiger zu sein scheinen (vgl. Stanat, 2006). Reflektiert man die praktischen Konsequenzen (s. auch unten), so wird sicherlich nicht die Lösung des Problems in der Steigerung des Migrantenteils in der Klasse liegen, sondern in zusätzlichen Maßnahmen seitens der Lehrkraft zur Unterstützung der Leistungsentwicklung der Schülerinnen und Schüler.

### **2.5.3 Modell der reziproken Zusammenhänge**

Analog zu früheren Befunden ließen sich im Sinne des REM wechselseitige Effekte des verbalen Selbstkonzepts und der sprachlichen Leistungen nachweisen, die eher klein ausfielen allerdings auch in dieser Höhe durchaus mit der einschlägigen Literatur (Möller et al., 2011) vereinbar waren. Hervorzuheben ist, dass diese Befunde nicht durch die Herkunft der Schülerinnen und Schüler moderiert wurden. Die vorliegende Arbeit

schließt damit eine Lücke, auf die Valentine et al. (2004) hingewiesen haben. Gleichzeitig reiht sie sich aber auch in vielen Arbeiten ein (u.a. Marsh & Hau, 2003, 2004), die implizieren, dass das Zusammenspiel von domänenspezifischen Selbstkonzepten und Schulleistungen ein universelles Phänomen darstellt, das mehr oder weniger kulturunabhängig gilt.

Weiterhin zählt unsere Arbeit zu den vergleichsweise wenigen, in denen das REM für den sprachlichen Bereich, hier Lesen und Rechtschreiben, repliziert werden konnte (vgl. hierzu auch Retelsdorf et al., 2014). Ehm et al. (2014) haben vor kurzem Effekte entsprechend dem I/E-Modell (Marsh, 1984, 1986) von Rechtschreib-, Lese- und Mathematikleistungen auf verbale Selbstkonzepte zeigen können, unsere Arbeit zeigt die Effekte auch in der entgegengesetzten Richtung.

#### **2.5.4 Limitationen**

Trotz der weitgehend hypothesenkonformen Befunde und der umfangreichen Datenbasis, mit der die Ergebnisse gewonnen wurden, bleiben auch Schwächen. Die von uns verwendete Skala zur Erfassung des verbalen Selbstkonzepts bezog sich auf das Fach Deutsch und nicht spezifisch auf die beiden Kompetenzbereiche Lesen und Rechtschreiben, wie dies beispielsweise in der Arbeit von Ehm et al. (2014) der Fall ist. Auf der anderen Seite stellt sich schon die Frage, wie spezifisch Selbstkonzepte erhoben werden sollten. Folgt man der Argumentation von Bong und Clark (1999), so stellen domänenspezifische Selbstkonzepte stärker aggregierte Repräsentationen eigener Fähigkeiten dar, während sich Selbstwirksamkeitsüberzeugungen auf vergleichsweise spezifische Tätigkeiten bzw. Handlungen beziehen. Schließt man sich dieser Argumentation nicht an und konzeptualisiert Selbstkonzepte für spezifische Kompetenzen innerhalb von Fächern, so

kann man auch die Frage stellen, warum das mathematische Selbstkonzept nicht in ein geometrisches, ein arithmetisches, ein algebraisches und ein stochastisches aufgeteilt wird. Wir folgen dieser feinkörnigeren Argumentation nicht, sondern denken, dass gerade im schulbezogenen Kontext das Fach und nicht die einzelnen Subkompetenzen im Fach bei der Erfassung des Selbstkonzepts im Vordergrund stehen sollte. Weitere Forschungsarbeiten zu den Vor- und Nachteilen einer besonders feinkörnigen Erfassung von Selbstkonzeptfacetten könnten hier aber sicher weitere empirische Argumente liefern.

Hinsichtlich der Messung der Leistungen kann ohne Frage kritisiert werden, dass der eingesetzte Rechtschreibtest nur passive orthografische Kompetenzen erfasste. Wir sehen diese Einschränkung, weisen allerdings auf die Befunde von Lehmann, Peek, Gänßfuß und Husfeldt (2002) hin, die zeigen konnten, dass passive und aktive Rechtschreibleistungen (ohne Messfehlerkorrektur) Korrelationen über .60 aufweisen. Für die Validität des von uns eingesetzten Tests sprechen sicher auch die substanziellen Korrelationen mit der Deutschnote (s. Methodenabschnitt).

Schließlich muss eingeräumt werden, dass wir in der vorliegenden Untersuchung nicht zwischen unterschiedlichen Zuwanderungsgruppen (nach Herkunftsland oder 1. bzw. 2. Generation) differenzieren konnten (s. hierzu beispielsweise Walter, 2006), zum einen weil die Schülerinnen und Schüler fast ausschließlich in Deutschland bereits geboren waren und zum anderen jenseits der türkischen Gruppe die übrigen Subgruppen mit Migrationshintergrund in der Stichprobe sehr gering besetzt waren. Hier können zukünftige Studien mit einer stärkeren Differenzierung der Zuwanderungsgruppe hilfreich sein.

### 2.5.5 Praktische Implikationen

Die vorliegende Untersuchung macht ergänzend zu den Analysen von Walter (2006) für sprachliche Kompetenzen deutlich, dass migrationsbedingte Disparitäten in den Schulleistungen im Laufe der Sekundarstufe I nicht weiter zunehmen. Diese Befunde implizieren, dass entsprechende Disparitäten, die sich in unserer Untersuchung in den Unterschieden in den Ausgangsleistungen manifestierten, bereits in der Grundschule und im vorschulischen Bereich entstehen. Die IGLU-Befunde (Bos et al., 2012) weisen in der Tat auf sehr große Unterschiede zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund am Ende der 4. Jahrgangsstufe hin. Die Ergebnisse für die Lesegeschwindigkeit und das Textverstehen bei Mücke und Schröder Lenzen (2008) belegen migrationsbedingte Disparitäten bereits am Ende der 1. Jahrgangsstufe. Internationale Studien (vgl. u.a. Hansen & Joshi, 2007) zeigen darüber hinaus, dass Disparitäten in der kognitiven, emotionalen und motorischen Entwicklung bereits lange vor der Einschulung in den ersten Lebensmonaten entstehen. Das Problem der Reduzierung migrationsbedingter Disparitäten scheint demnach nicht allein ein schulisches, sondern mindestens ebenso stark ein vorschulisches und familiäres zu sein.

Hinsichtlich des verbalen Selbstkonzepts weisen die Befunde darauf hin, dass Schülerinnen und Schüler mit Migrationshintergrund hier kaum geringere Werte aufweisen als Schülerinnen und Schüler ohne Migrationshintergrund, und dies, obwohl ihrer Leistungen geringer sind. Man mag hier argumentieren, dass sie ihre eigene Leistungsfähigkeit überschätzen. Traut man allerdings den Zusammenhängen mit den Leistungen, so schadet diese Überschätzung keineswegs, sondern ist sogar leistungsförderlich. Maßnah-

men zur Absenkung des überhöhten Selbstkonzepts wären hier vermutlich unzweckmäßig.

## 2.6 Literatur

- Alexander, K. L. & Entwisle, D. R. (1996). Schools and children at risk. In A. Booth & J. F. Dunn (Eds.), *Family-school links. How do they affect educational outcomes?* (pp. 67-88). Mahwah: Erlbaum.
- Böhme, K., Tiffin-Richards, S. P., Schipolowski, S. & Leucht, M. (2010). Migrationsbedingte Disparitäten bei sprachlichen Kompetenzen. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 203-225). Münster: Waxmann.
- Bong, M. & Clark, R.E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, 34, 139-153.  
doi:/10.1207/s15326985ep3403\_1
- Bos, W., Tarelli, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.) (2012). *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Wendt, H., Köller, O. & Selter, C. (Hrsg.) (2012). *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Calsyn, R., & Kenny, D. (1977). Self-concept of ability and perceived evaluations by others: Cause or effect of academic achievement? *Journal of Educational Psychology*, 69, 136-145. doi:/10.1037/0022-0663.69.2.136
- Edele, A., Stanat, P., Radmann, S. & Segeritz, M. (2013). Kulturelle Identität und Lesekompe-

tenz von Jugendlichen aus zugewanderten Familien. *Zeitschrift für Pädagogik*, 59. Beiheft, 84-131.

Ehm, J. H., Nagler, T., Lindberg, S & Hasselhorn, M. (2014). Dimensionale Vergleichseffekte zwischen Lesen, Rechtschreiben und Rechnen. Eine Erweiterung des I/E-Modells für die Grundschule. *Zeitschrift für Pädagogische Psychologie*, 28, 51-56.  
doi:/10.1024/1010-0652/a000117

Entwisle, D. R. & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review*, 57, 72-84. doi:/10.2307/2096145

Eßel-Ullmann, G. (2008). *Effekte eines Konzepts zur Entwicklung von Schule und Unterricht an Schulen in benachteiligter Lage. Eine Evaluationsstudie des Programms EIKA zur Eingliederung Jugendlicher in die Berufs- und Arbeitswelt*. Dissertation: Universität Erlangen-Nürnberg.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.  
doi:/10.1177/001872675400700202

Goethals, G. R. & Darley, J. M. (1977). Social comparison theory: An attributional approach. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 259-278). Washington, DC: Hemisphere.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: developmental perspectives on their causal ordering. *Journal of Educational Psychology*, 95, 124-136. doi:/10.1037/0022-0663.95.1.124



- Hansen, K. & Joshi, H. (2007). *Millenium Cohort Study. Second survey. A user's guide to initial findings*. London: Institute of Education, Centre for Longitudinal Studies.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W. & Stanat, P. (Hrsg.) (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Köller, O. (2004). *Konsequenzen von Leistungsgruppierungen*. Münster: Waxmann.
- Köller, O. (2009). Evaluation pädagogisch-psychologischer Maßnahmen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 333-351). Berlin: Springer.  
doi:/10.1007/978-3-540-88573-3\_14
- Köller, O., Knigge, M. & Leucht, M. (Hrsg.) (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Lehmann, R. H. & Peek, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klasse an Hamburger Schulen*. Hamburg: Hamburger Schulbehörde.
- Lehmann, R. H., Peek, R., Gänsfuß, R. & Husfeldt, V. (2002). *LAU 9: Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 9 – Ergebnisse einer längsschnittlichen Untersuchung in Hamburg*. Hamburger Schulbehörde.
- Lüdtke, O., Köller, O., Marsh, H. W. & Trautwein, U. (2005). Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology*, 30, 263-285.  
doi:/10.1016/j.cedpsych.2004.10.002
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rund-*

*schau*, 58, 103-117. doi:/10.1026/0033-3042.58.2.103

Marsh, H. W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education*, 28, 165-181.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129-149.

doi:/10.3102/00028312023001129

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79, 280-295. doi:/10.1037/0022-0663.79.3.280

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of sport psychology* (pp. 774-798). Hoboken, NJ: Wiley.

doi:/10.1080/10705510701303921

Marsh, H. W. & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1 (2), 133-163.

doi:/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Hau, K. T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, 58, 364-376. doi:/10.1037/0003-066X.58.5.364

Marsh, H. W., Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame

of reference predictions across 26 countries. *Journal of Educational Psychology*, 96, 56-67. doi:/10.1037/0022-0663.96.1.56

Marsh, H. W. & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: Reunification of East and West German school systems after the fall of the Berlin wall. *Contemporary Educational Psychology*, 29, 264-282. doi:/10.1016/S0361-476X(03)00034-1

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, 81, 59-77. doi:/10.1348/000709910X503501

Marsh, H. W. & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin*, 34 (4), 542-552. doi:/10.1177/0146167207312313

Möller, J. & Köller, O. (2001). Frame of reference effects following the announcement of exam results. *Contemporary Educational Psychology*, 26, 277-287. doi:/10.1006/ceps.2000.1055

Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau*, 55, 19-27. doi:/10.1026/0033-3042.55.1.19

Möller, J. & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, 120, 544-560. doi:/10.1037/a0032459

- Möller, J., Pohlmann, B., Köller, O. & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79, 1129-1176.  
doi:/10.3102/0034654309337522
- Möller, J., Retelsdorf, J., Köller, O. & Marsh, H. W. (2011). The reciprocal I/E model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 48, 1315-1346.  
doi:/10.3102/0002831211419649
- Mücke, S. & Schründer-Lenzen, A. (2008). Zur Parallelität der Schulleistungsentwicklung von Jungen und Mädchen im Verlauf der Grundschule. In B. Rendtorff & A. Prenzel (Hrsg.), *Kinder und ihr Geschlecht* (S. 135-146). Opladen: Budrich.
- Muthén, L. K. & Muthén, B. (2010). *Mplus. Statistical analysis with latent variables. User's guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.
- Pant, A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.) (2013). *IQB Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- PISA-Konsortium Deutschland (Hrsg.) (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (2013). *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.
- Retelsdorf, J., Köller, O., Möller, J. (2014). Reading achievement and reading self-concept –

Testing the reciprocal effects model. *Learning and Instruction*, 29, 21-30.

doi:/10.1016/j.learninstruc.2013.07.004

Roebers, C. M., Mecheril, A. & Schneider, W. (1998). Migrantenkinder in deutschen Schulen.

Eine Studie zur Persönlichkeitsentwicklung. *Zeitschrift für Pädagogik*, 44, 723-736.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

doi:/10.1002/9780470316696

Shajek, A., Lüdtke, O. & Stanat, P. (2006). Akademische Selbstkonzepte bei Jugendlichen mit

Migrationshintergrund. *Unterrichtswissenschaft*, 34, 125-145.

Skaalvik, E. M., & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analy-

sis of causal predominance in a developmental perspective. *Journal of Personality and*

*Social Psychology*, 58, 292-307. doi:/10.1037/0022-3514.58.2.292

Stanat, P. (2006). Schulleistungen von Jugendlichen mit Migrationshintergrund: Die Rolle der

Zusammensetzung der Schülerschaft. In J. Baumert, P. Stanat & R. Watermann

(Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle*

*Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (S. 189-219). Wiesbaden:

VS Verlag für Sozialwissenschaften. doi:/10.1007/978-3-531-90082-7\_5

Stanat, P., Rauch, D. & Segeritz, M. (2010). Schülerinnen und Schüler mit

Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel,

W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200-

230). Münster: Waxmann.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and

academic achievement: A meta-analytic review. *Educational Psychologist*, 39, 111-123. doi:/10.1207/s15326985ep3902\_3

Walter, O. (2006). Die Entwicklung der mathematischen und der naturwissenschaftlichen Kompetenzen von Jugendlichen mit Migrationshintergrund im Verlauf eines Schuljahres. In Deutsches PISA-Kosortium (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 249-275). Münster: Waxmann.

Walter, O. & Stanat, P. (2008). Der Zusammenhang des Migrantenanteils in Schulen mit der Lesekompetenz: Differenzierte Analysen der erweiterten Migrantenstichprobe von PISA 2003. *Zeitschrift für Erziehungswissenschaft*, 11, 84-105. doi:/10.1007/s11618-008-0005-7

Tabelle 2.1. Mittelwerte und (in Klammern) Standardabweichungen für die berücksichtigten Variablen

	Nicht-Gymnasium		Gymnasium	
	Ohne MH	mit MH	Ohne MH	mit MH
T1 Lesen	101.94 (26.15)	84.40 (23.74)	128.38 (24.43)	109.43 (24.79)
T2 Lesen	119.75 (25.31)	107.45 (23.97)	147.26 (27.14)	133.82 (28.15)
T1 Rechtschreiben	7.53 (4.43)	6.44 (4.12)	12.52 (5.41)	10.57 (5.03)
T2 Rechtschreiben	12.71 (5.82)	11.19 (5.57)	18.92 (5.39)	16.78 (5.86)
T1 Selbstkonzept	8.56 (4,57)	8.08 (4.46)	11.17 (3.75)	10.34 (4.09)
T2 Selbstkonzept	9.60 (4.03)	9.06 (4.15)	10.25 (3.70)	9.33 (3.84)

*Anmerkungen.* Analysen in *Mplus* 6.12 mit 50 imputierten Datensätzen; Werte wurden entsprechend den Vorgaben bei Rubin (1987) gemittelt und beziehen sich auf die Daten aus beiden Kohorten.

Tabelle 2.2. Korrelationen der berücksichtigten Variablen

	(1)	(2)	(3)	(4)	(5)	(6)
T1 Lesen (1)	1.00					
T2 Lesen (2)	.725	1.00				
T1 Rechtschreiben (3)	.494	.473	1.00			
T2 Rechtschreiben (4)	.512	.545	.676	1.00		
T1 Selbstkonzept (5)	.298	.281	.312	.270	1.00	
T2 Selbstkonzept (6)	.201	.216	.234	.290	.289	1.00

*Anmerkungen.* Analysen in *Mplus* 6.12 mit 50 imputierten Datensätzen; Werte wurden entsprechend den Vorgaben bei Rubin (1987) gemittelt und beziehen sich auf die Daten aus beiden Kohorten; alle Koeffizienten sind hoch signifikant ( $p < .001$ ).



Tabelle 2.3. Befunde aus Mehrebenenanalysen (standardisierte Koeffizienten und ihre Standardfehler aus *Mplus* 6.12) zur Vorhersage von Veränderungen im verbalen Selbstkonzept und den Lese- und Rechtschreibleistungen

	<i>Kriterium</i>		
	T2 Selbstkonzept ( <i>ICC</i> = .124)	T2 Lesen ( <i>ICC</i> = .288)	T2 Rechtschreiben ( <i>ICC</i> = .312)
<i>Prädiktoren</i>			
T1 Selbstkonzept	.291 (.034)**		
T1 Lesen		.696 (.020)**	
T1 Rechtschreiben			.639 (.021)**
igrationsstatus (1 = mit Migrationshintergrund)	-.057 (.031)	-.014 (.023)	-.056 (.025)*
Schulform (1 = Gymnasium)	-.115 (.223)	.861 (.078)**	.728 (.073)**
<i>R</i> <sup>2</sup>	.09/.03	.49/.74	.42/.53

*Anmerkungen.* Analysen in *Mplus* 6.12 mit 50 imputierten Datensätzen; Werte wurden entsprechend den Vorgaben bei Rubin (1987) gemittelt und beziehen sich auf die Daten aus beiden Kohorten; *ICC*: Intraklassenkorrelation; *R*<sup>2</sup>: Aufgeklärte Varianz; vor dem Schrägstrich auf Individualebene, hinter dem Schrägstrich auf Klassenebene; alle Modelle sind saturiert; \*  $p < .05$ ; \*\*  $p < .01$ .

Tabelle 2.4. Fit-Statistiken für verschiedene Modelle zum REM mit gruppeninvarianten Parametern im Mess- und Strukturmodell (Ergebnisse aus *Mplus*-Analysen)

	$\chi^2/df$	RMSEA	CFI	TLI
<i>Lesen</i>				
Invariantes Mess- und Strukturmodell in beiden Schulformen	2.47	.040	.991	.992
Invariantes Mess- und Strukturmodell für Schülerinnen und Schüler mit und ohne Migrationshintergrund	1.85	.030	.996	.996
<i>Rechtschreiben</i>				
Invariantes Mess- und Strukturmodell in beiden Schulformen	2.66	.042	.985	.987
Invariantes Mess- und Strukturmodell für Schülerinnen und Schüler mit und ohne Migrationshintergrund	2.88	.045	.986	.988

*Anmerkungen.*  $\chi^2/df$ : Quotient aus  $\chi^2$ -Statistik und der Zahl der Freiheitsgrade des Modells;

$df = 150$  in allen Modellen; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index.

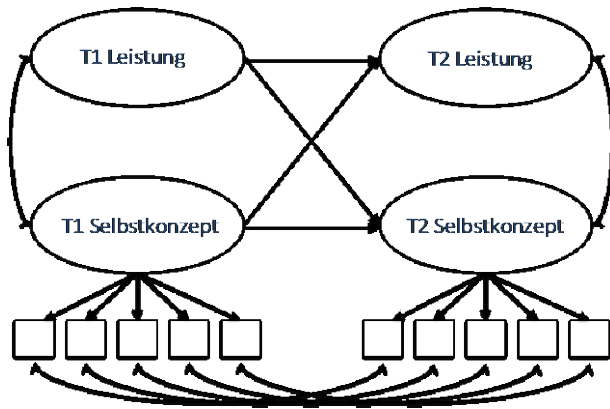


Abbildung 2.1. Spezifiziertes Modell zur Überprüfung reziproker Effekte des verbalen Selbstkonzepts und der sprachlichen Leistungen.

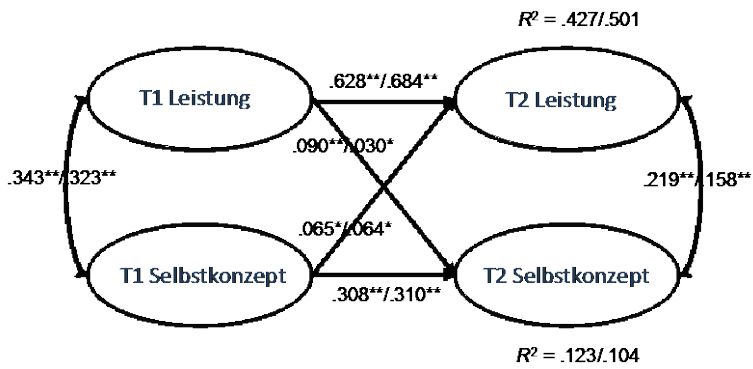


Abbildung 2.2. Befunde aus Strukturgleichungsanalysen (voll standardisierte Koeffizienten aus Mplus 6.12) zum Zusammenspiel von sprachlichen Leistungen und dem verbalen Selbstkonzept; Vor dem Schrägstrich sind die Befunde für das Rechtschreiben, nach dem Schrägstrich für die Leseleistungen aufgeführt; \*  $p < .05$ ; \*\*  $p < .01$ .

# 3

## **Studie 2: Reciprocal effects between self-efficacy and achievement in mathematics and reading**

Schöber, C., Schütte, K., Köller, O., McElvany, N., Gebauer, M. M. (Manuscript submitted for publication). Reciprocal effects between self-efficacy and achievement in mathematics and reading.

This is an Authors Original Manuscript of an article published with open access by Elsevier in *Learning and Individual Differences* on January 19<sup>th</sup>, 2018. Please do not copy without authors' permission. The final article is available online: doi 10.1016/j.lindif.2018.01.008

## Zusammenfassung

Schulische Selbstwirksamkeitsüberzeugungen sind weithin sowohl als Ursache als auch als Effekt von schulischen Leistungen anerkannt. Allerdings sind empirische Untersuchungen unter Verwendung von Längsschnittdaten auf einer domänenspezifischen Ebene selten und fehlt gerade in nicht-mathematischen Domänen offensichtlich völlig. Wir haben daher an einer Stichprobe von  $N = 1597$  Sekundarschulkindern in Deutschland mit zwei Messzeitpunkten innerhalb eines Schuljahres das reziproke Zusammenspiel zwischen Selbstwirksamkeitsüberzeugungen und Leistungen in den Domänen Lesen und Mathematik untersucht. Trotz hoher Stabilitäten von Leistungen und Selbstwirksamkeitsüberzeugungen konnten wir mit Strukturgleichungsmodellen Unterstützung für die Self-Enhancement-Hypothese im Bereich Mathematik und für die Skill-Development-Hypothese im Bereich Lesen finden. Die Berücksichtigung des Migrationshintergrundes der Schulkinder in Mehrgruppenmodellen lieferte zudem Evidenz für reziproke Effekte im Bereich Lesen. Diese Ergebnisse heben die Notwendigkeit frühzeitiger Förderung und domänenspezifischer Betrachtung der Selbstwirksamkeitsüberzeugungen hervor.

### 3.1 Introduction

While we are students, our job is to learn, and to acquire new skills and knowledge. But does it matter whether or not students are confident about dealing with the requirements made in school? Originally introduced by Albert Bandura as part of the social cognitive theory (Bandura, 1977, 1986, 1997), academic self-efficacy is defined as “personal judgments of one’s capabilities to organize and execute courses of action to attain designated types of educational performances” (Zimmerman, 1995, p. 203). It is hypothesized that academic self-efficacy influences students’ level of effort, and persistence, and their choice of activities (Bandura, 1977)—as a consequence, higher academic achievement can be expected from students with higher self-efficacy than from students with lower self-efficacy.

At the same time, enactive mastery experience is considered to be the strongest source of self-efficacy (Bandura, 1997), that is, high academic achievement strengthens students’ judgments of their capability to deal with future performance requirements. Academic self-efficacy is thus considered to be both the cause and an effect of academic achievement. However, there is still a lack of longitudinal studies investigating the reciprocal relationship between self-efficacy and achievement. Moreover, existing research often comprised investigations on a domain-general level or only in the mathematics domain. In addition, the relationship between academic self-efficacy and achievement appeared to be universal across different cultural groups; only a small number of investigations regarded the cultural universality of this relationship on a domain-specific level.

The goal of this article is to contribute empirical evidence to support the hypothesized reciprocal relationship between self-efficacy and achievement. To this end,

we conducted an extensive investigation of the reciprocal relationship between self-efficacy and achievement in the domains of mathematics and reading as assessed at the beginning and towards the end of a school year in early secondary school. We drew on a sample that comprised a large number of students with a migration background to test whether the relationship between self-efficacy and achievement in both domains showed measurement and structural invariance between students with and without a migration background. Therefore, we add to existing research by being, to the best of our knowledge, the first to investigate the relationship between self-efficacy and achievement in the domain of reading. By investigating this relationship separately in two domains, while using data from the same sample, our study makes it possible to contrast the effects found between the two domains.

In the remainder of this introduction we briefly introduce competing models regarding the relationship between self-beliefs and achievement. We then review empirical results on this relationship with subsections regarding domain-specific self-efficacy and achievement, before contemplating the cultural universality of this relationship.

### **3.1.1 Relationship Between Academic Self-Beliefs and Achievement**

Academic achievement has often been related to self-beliefs, with self-concepts and self-efficacy being the most important self-beliefs. Self-concepts include cognitive and affective appraisals of the self in contrast to self-efficacy, which only includes cognitive appraisals. Self-concepts are past-oriented and relatively stable, whereas self-efficacy is future-oriented and relatively malleable (see Bong & Clark, 1999; Bong & Skaalvik, 2003, for comparisons). Both should be assessed on a domain-specific level (Bong, Cho, Ahn, & Kim, 2012) and in a manner corresponding to the level of specificity



on which the achievement is to be measured (Marsh, 1992; Pajares, 1996b) in order to obtain valid results. In a meta-analysis, both constructs showed discriminant validity as predictors of achievement, but self-efficacy had a higher incremental validity than self-concept (Huang, 2011). Due to self-efficacy's malleability, educators might have a notable influence on the development of students' academic self-efficacy (see Pajares, 1996a; Schunk & Pajares, 2009, for overviews).

Three models that assume a causal relationship between self-beliefs and achievement have been discussed in the literature (Calsyn & Kenny, 1977). For domain-specific achievement and self-efficacy, the skill development model implies that students' self-efficacy in a specific domain will become higher, the higher their achievement in the same domain is. Self-efficacy has four sources: (a) enactive mastery experience, (b) vicarious experience, (c) verbal persuasion, and (d) physiological and affective states (Bandura, 1997; Usher & Pajares, 2008). Enactive mastery experience is considered to be the most powerful of these sources because one's own success or failure is assumed to have a longer-lasting influence than information from the other three sources and can even abrogate them (Bandura, 1997). For example, verbal persuasion may be useless if it is followed by contradictory enactive mastery experience. The primacy of enactive mastery experience supports the idea of a causal influence of achievement on future self-efficacy because successful behavior could lead to an improvement in the corresponding self-efficacy whereas failure could diminish it. Thus, the skill development model represents the influence of enactive mastery experience on self-efficacy. According to this model, self-efficacy does not have a causal effect on later achievement.

In contrast, the self-enhancement model signifies that students' high self-efficacy in a specific domain enhances their achievement in that domain, but that higher achievement levels do not influence the development of higher self-efficacy. Galla et al. (2014) confirmed Bandura's (1977) claim that academic self-efficacy has a positive influence on the effort students make and, therefore, on their achievement. They found that students' scores in standardized reading and mathematics tests were higher, the higher their self-reported academic self-efficacy was; this effect was mediated by the students' level of effortful engagement in academic situations, as rated by their teachers. Thus, academic self-efficacy is one important precursor of engagement; engaging in effortful learning behavior, in turn, leads to higher academic achievement. The importance of engagement for reading achievement has been shown by other studies (e.g., Guthrie & Wigfield, 2000; Lee, 2013).

The reciprocal effects model (REM) of causal relationships, which was promoted by Marsh (1990), integrates the causal relationships proposed by the aforementioned models; it hypothesizes that self-beliefs and achievement mutually reinforce each other. This view is consistent with Bandura's (1997) assertion that self-efficacy is both a cause and an effect of achievement.

### **3.1.2 Empirical Results Regarding the Relationship Between Self-Beliefs and Achievement**

The best investigated self-beliefs in terms of a reciprocal relationship with academic achievement are self-concepts. A reciprocal relationship between self-concepts and achievement has been demonstrated on a domain-general level (Guay, Marsh, & Boivin, 2003; Marsh, Byrne, & Yeung, 1999; Preckel, Niepel, Schneider, & Brunner,

2013), and in the mathematics domain (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Pinxten, Marsh, De Fraine, Van Den Noortgate, & Van Damme, 2014). Investigating two nationally representative samples of seventh-grade students in Germany, Marsh et al. (2005) found reciprocal effects between mathematics self-concept and mathematics achievement with path coefficients favoring the self-enhancement model. Regarding the domain of reading, Retelsdorf, Köller, and Möller (2014) found evidence to support the REM in students who had recently started secondary school in Germany (i.e., from Grade 5 to the end of Grade 6), but, in higher grades, there was only evidence for the skill development model. They attributed the vanishing support for the self-enhancement model after Grade 6 to the increased homogeneity of achievement in the classroom: in Germany, students start secondary school in Grade 5; depending on which secondary school track students attend, the self-concepts they formed during primary schooling may benefit or suffer from the new reference group they have to use for social comparisons (Ireson & Hallam, 2009). The inconsistent support for reciprocal effects in the domain of reading in contrast to the findings on either domain-general or mathematics-specific levels highlights the necessity for research in different domains.

In a nationally representative sample of U.S. children, Ganley and Lubienski (2016) found evidence for a reciprocal relationship between mathematics confidence and mathematics test performance from Grade 5 to Grade 8, but from Grade 3 to Grade 5, their data only supported the skill development model. This may be due to the age of the students; younger students often report overly optimistic views of the self and, therefore, the relationship between self-beliefs and achievement may have been flawed (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Wigfield & Karpathian, 1991). Moreover, Gan-

ley and Lubienski used four items to assess mathematics confidence at Grade 3 and Grade 5; for Grade 8, they exchanged some of their items, making the interpretation of stability and causality difficult.

Empirical evidence for reciprocal effects between reading motivation and text comprehension was provided by Kigel, McElvany, and Becker (2015) using a sample of primary school students in the German capital city of Berlin, where students make the transition to secondary schools after Grade 6. The students were assessed at the end of Grade 3, in the middle of Grade 4, and at the end of Grade 6. The results of Kigel et al. did not reveal reciprocal effects from one measurement time point to the other but supported the self-enhancement model in a medium-term interval from Grade 3 to Grade 4 and the skill development model in a long-term interval from Grade 3 to Grade 6.

In their meta-analysis on the relationship between self-efficacy and performance, Multon, Brown, and Lent (1991) found an unbiased effect size estimation of  $r = .32$  for pretreatment and correlational measures. Longitudinal correlational data were not yet available at that time and, thus, an investigation of cause and effect was impossible. In meta-analyses controlling for prior levels of self-beliefs (including self-esteem, self-concept, and self-efficacy) or prior achievement (including grades, standardized test scores, and attainment), evidence for a prospective relationship was shown (Valentine & Dubois, 2005; Valentine, Dubois, & Cooper, 2004). However, the path coefficient from achievement to later self-beliefs ( $\beta = .15$ ) was about twice as much as the path coefficient from self-beliefs to later achievement (Valentine & Dubois, 2005). In a recently published meta-analysis on the influence of academic self-efficacy on the academic achievement of university students, a moderate correlation of  $r = .33$  was revealed (Honicke &

Broadbent, 2016). Reciprocal effects could not be rigorously investigated due to a lack of appropriate studies and the authors concluded that further research was still needed to establish the causality in the relationship between academic self-efficacy and academic achievement.

Alongside the paucity of longitudinal studies investigating the REM, studies using domain-specific measures of self-efficacy are also scarce. Mittag, Kleine, and Jerusalem (2002) tracked the 4-year development of academic self-efficacy and achievement across domains in  $N = 434$  students in Germany from Grade 7 to Grade 10 with annual measurements. They did not find support for a reciprocal relationship between academic self-efficacy and academic achievement but they did find evidence to support the self-enhancement model (i.e., positive path coefficients from academic self-efficacy to later academic achievement) from Grade 7 to Grade 8 and from Grade 9 to Grade 10. Likewise, Hwang, Choi, Lee, Culver, and Hutchison (2016) tracked students in Korea from Grade 8 to Grade 12 and investigated their academic self-efficacy and achievement with 1-year intervals. Their path models revealed consistent reciprocal effects between achievement and later academic self-efficacy ( $\beta$ s from .36 to .44) and between academic self-efficacy and later achievement ( $\beta$ s from .13 to .14). Investigating reciprocal effects between self-efficacy and achievement with longer intervals, Caprara, Vecchione, Alessandri, Gerbino, and Barbaranelli (2011) assessed the mathematics and self-regulatory self-efficacy of students aged 13 and 16 and their achievement aged 14 and 19. They revealed reciprocal effects between self-efficacy and achievement ( $\beta$ s from .19 to .22) even when openness and conscientiousness were taken into account. However, in these studies, self-efficacy was investigated on a domain-general level and the achievement meas-

ure comprised several domains. Thus, the assessment was unfavorable (Bong, 2002; Bong et al., 2012) and the reported effects cannot be applied to specific domains.

### 3.1.2.1 Domain-Specific Relationship Between Self-Efficacy and Achievement

To test the domain-specific reciprocal relationship between mathematics self-efficacy and mathematics achievement, Williams and Williams (2010) based their structural equation model on the data of the 2003 survey of the *Programme for International Student Assessment* (PISA; Organisation for Economic Co-operation and Development, 2005). To deal with the cross-sectional nature of the data, they assumed the system between both constructs to have reached equilibrium, meaning that 15-year-old students were able to give realistic appraisals of their mathematics competence. With this approach, they found evidence for the reciprocal relationship of mathematics self-efficacy and mathematics achievement in 24 of the 33 countries included in their data set, and claimed that these findings could be replicated using longitudinal data.

Using nationally representative longitudinal data from Finland, Hannula, Bofah, Tuohilampi, and Metsämuuronen (2014) found evidence for reciprocal effects between mathematics self-efficacy and mathematics achievement, with the dominant direction being from achievement to self-efficacy. They used a sample of students who were surveyed at the beginning of Grade 3 and Grade 6 as well as at the end of Grade 9 and, thus, were surveyed over seven years with intervals of at least three years between the measurements. Comparing the two cross-lagged paths from self-efficacy to achievement, they revealed that the effect was stronger when students were older ( $\beta = .16$  between Grade 3 and Grade 6, and  $\beta = .26$  between Grade 6 and Grade 9), whereas the effect of achievement on self-efficacy did not change as much (from  $\beta = .34$  to  $\beta = .30$ ). They concluded

that the relationship between self-efficacy and achievement may transform from an achievement-dominated relationship in younger students to a reciprocal relationship when students get older. With regard to the mathematics domain, this conclusion is in line with the results of Ganley and Lubienski (2016), although they assessed mathematics confidence rather than mathematics self-efficacy. In contrast, in the results of Hwang et al. (2016), the dominance of achievement over self-efficacy was constant for domain-general self-efficacy, even in older students up to Grade 12. Thus, more research regarding the causal relationship between mathematics self-efficacy and mathematics achievement is required and, due to the absence of any studies on this topic, it is even more necessary in the domain of reading.

### 3.1.2.2 Cultural Universality of the Relationship Between Self-Efficacy and Achievement

As the studies mentioned in section 1.2.1 stem from different countries, it is worth contemplating the cultural universality of the relationship between self-efficacy and achievement. Investigations of self-efficacy that considered different cultures (i.e., individualistic cultures, such as the USA and Western Europe, vs. the collectivistic cultures in Asia and Eastern Europe) showed that academic self-efficacy was an equally good predictor of individual achievement for adolescents from both cultures, despite the lower mean levels of self-efficacy and higher mean levels of achievement in collectivistic cultures (Klassen, 2004). This means that the relationship between academic self-efficacy and achievement is equal in both cultural groups, although members of individualistic cultures report higher self-efficacy than members of collectivistic cultures compared to the respective achievement. Likewise, after reviewing existent research and making her own comparison of the self-efficacy and achievement of students in East and West Berlin, Moscow, and Los Angeles, Oettingen (1995) concluded that the effects of self-efficacy on achievement are universal across cultures.

In contrast, in a sample from Germany, students' migration background was positively related to the relative change in their intrinsic reading motivation, while both the initial level and the absolute change in their intrinsic reading motivation was unrelated to their migration background (Kigel et al., 2015). At the same time the initial level of these students' text comprehension was significantly lower. Likewise, when comparing groups of students stemming from European or other western countries (labeled as the ethnic majority of the Netherlands) with students stemming from non-western countries outside of Europe (labeled as the ethnic minority), using five waves of measurement and



multivariate latent growth curve analyses, Hornstra, van der Veen, Peetsma, and Volman (2013) found no significant differences in the slope of self-efficacy development between Grade 3 and Grade 6, but students of the ethnic minority had significantly higher intercepts than students of the ethnic majority. However, none of these studies investigated domain-specific self-efficacy or contrasted different domains.

The results of Hornstra et al. (2013) also indicate that self-efficacy may develop similarly in students of different achievement levels. There are similar findings on the development of verbal self-concepts and achievement in different secondary school tracks in Germany (Schöber, Retelsdorf, & Köller, 2015). Recent reviews regarding reading self-efficacy (Piercey, 2013) as well as meta-analytical findings (Valentine et al., 2004) corroborate these findings.

### 3.1.3 Objectives

In a nutshell, what has been empirically documented so far is that academic achievement can be predicted by self-efficacy (Bandura, 1997; Schunk & Meece, 2006). Domain-general self-efficacy develops similarly across different cultural groups despite there being different levels of self-efficacy within these groups (Hornstra et al., 2013). Although reciprocal effects between self-beliefs and achievement are well documented in several empirical studies, there is still a lack of research using longitudinal data to examine these effects with regard to domain-specific self-efficacy. Research on reciprocal effects between domain-specific self-concepts and achievement suggests that the effects are different, favoring the self-enhancement model in the domain of mathematics (Marsh et al., 2005) and the skill development model in the domain of reading (Retelsdorf et al., 2014). Therefore, the main objective of this study was to empirically corroborate the assumed reciprocal relationship between self-efficacy and achievement for both reading and mathematics on two measurement occasions.

This study extends previous research by applying the REM to domain-specific self-efficacy using longitudinal data. In addition, this study is one of the first empirical investigations of the cross-lagged paths of self-efficacy and achievement in the domain of reading. We posited the following four hypotheses, based on the postulation of reciprocal effects between personal and cognitive factors (Bandura, 1978; Bandura, 1986; Bandura, 1997), on empirical results on the effect of academic self-efficacy on later achievement (Valentine et al., 2004), on results indicating that achievement is a source of academic self-efficacy (Usher & Pajares, 2008), and on the empirical findings of a reciprocal relationship between self-efficacy and achievement (Caprara et al., 2011; Ganley

& Lubienski, 2016; Hannula et al., 2014; Hwang et al., 2016; Williams & Williams, 2010):

- Hypothesis 1a: Mathematics achievement has a positive effect on later mathematics self-efficacy.
- Hypothesis 1b: Mathematics self-efficacy has a positive effect on later mathematics achievement.
- Hypothesis 2a: Reading achievement has a positive effect on later reading self-efficacy.
- Hypothesis 2b: Reading self-efficacy has a positive effect on later reading achievement.

Hypothesis 1a and Hypothesis 2a represent the relationship postulated in the skill development model, whereas Hypothesis 1b and Hypothesis 2b represent the relationship postulated in the self-enhancement model. Concurrent support for Hypotheses 1a and 1b or for Hypotheses 2a and 2b would thus confirm the REM. We adopted a Type I error rate of  $\alpha = .05$  to evaluate the statistical significance of the path coefficients.

As a secondary objective, we assumed the hypothesized domain-specific relationship between self-efficacy and achievement to be culturally universal, as indicated by previous research on a domain-general level. In our study, we followed the approach of Hornstra et al. (2013) and compared students with and without a migration background within one country (i.e., Germany). To provide insights into the cultural universality of the domain-specific relationship between self-efficacy and achievement, we conducted additional two-group analyses considering students' migration background (migration background vs. no migration background) to test the consistency of the relationship men-

tioned in our hypotheses. In addition, we also conducted two-group analyses for students in different school tracks (academic track schools versus non-academic track schools) to ascertain whether the hypothesized relationships are universal across school tracks. In Germany, students with a migration background often attend non-academic school tracks (Autorengruppe Bildungsberichterstattung, 2016); groups of school tracks and migration background are confounded. Therefore, we also conducted four-group models, dividing students by migration background and school track.

## 3.2 Method

### 3.2.1 Sample

We investigated a stratified sample of  $N = 1597$  seventh-grade students (49.5% female) with a mean age of 12.15 years ( $SD = 0.63$ ) stemming from the study [name deleted to maintain the integrity of the review process]. The students attended three different school tracks in four federal states in Germany. In the vocational school track, students graduate after Grade 9 or 10 to start vocational training. Students attending the comprehensive school track have the choice after Grade 10 to leave school to start vocational education or to continue schooling at upper secondary level. Data from students attending either of these school tracks were included in one group (non-academic track,  $n = 897$ ). Finally, students who graduate from academic track schools are qualified to enter tertiary education ( $n = 700$ ). Students with a migration background were oversampled in this study, leading to  $n = 957$  students with a migration background, (among them many students with a migration background from Turkey or the successor states of the Former Soviet Union—the two largest immigrant groups in Germany) and  $n = 640$  stu-

dents without a migration background. Data were collected in October 2012 and June 2013.

Due to the fact that there are considerable differences in the intervals between the measurement occasions used in past research on reciprocal effects between self-beliefs and achievement, and that recommendations regarding those intervals range from several months (Honicke & Broadbent, 2016; Valentine et al., 2004) to more than one school year (Marsh et al., 1999), the optimal temporal distance between two assessments of self-efficacy and achievement is unclear. In addition to the time intervals between two measurement occasions, the amount of change between them may also be critical (Marsh et al., 2005), resulting in a trade-off between the malleability of self-efficacy and the presumably more stable achievement in that age group. In contrast, in the study of Kigel et al. (2015) reading motivation showed a medium-term relationship with later text comprehension from the end of Grade 3 to the middle of Grade 4 but text comprehension was only related to reading motivation in the long-term from the end of Grade 3 to the end of Grade 6. Therefore, assessing both with an interval of several months seemed to be a viable approach that could theoretically be supported by the mediating roles of level of effort, persistence, and choice of activities (Bandura, 1977) and, thus, we chose an interval of approximately one school year between both measurement occasions: investigating the students in October 2012 and June 2013.

### 3.2.2 Measures

#### 3.2.2.1 Students' achievement

Students' achievement was assessed using standardized tests with age appropriate items stemming from large-scale studies (e.g., Lehmann, Gänsfuß, & Peek, 1999). The tests were administered by trained research students and each test had a time limit of one school lesson (45 minutes) including the test instructions.

In the domain of mathematics, students' achievement in the content areas of quantity, space and shape, change and relationships, and data and chance was assessed. Our reading achievement test comprised several texts and questions that focused on students' skills in forming a broad and general understanding of the texts and retrieving information from them. The students' task was to read the texts and answer questions on their content. Most of the questions from both domains had to be answered in a multiple-choice format, but the reading achievement test also included a few short constructed-response items.

On each of the two measurement occasions, we administered two different test forms per domain to students in the academic track and two test forms to all other students, resulting in eight different test forms (2 measurement occasions  $\times$  2 domains  $\times$  2 school tracks). To create these test forms, we established an anchor-item test design (Kolen & Brennan, 2004), which allowed us to estimate person parameters on a common metric. Based on item response theory (IRT) models, we generated five plausible values (PVs; Wu, 2005) for each student using the software ConQuest 3.0 (Adams, Wu, & Wilson, 2012). The PVs were scaled to have a mean of 100 and a standard deviation of 30

for the complete sample of  $N = 1597$  on the first measurement occasion. All measures of students' achievement had good reliabilities (EAP/PV; see Table 3.1).

### 3.2.2.2 Self-efficacy

Self-efficacy was assessed with items from well-established scales (Jerusalem & Satow, 1999; Kunter et al., 2002), which were slightly adapted to a domain-specific level. There were 6 self-efficacy items with a 4-point response scale (0 = *not at all true*, 3 = *absolutely true*) for each of the domains of reading and mathematics. Example items were: "I'm certain that I can understand texts presented in lessons that are very difficult to read"; "Even if I were ill for a long time, I would still be capable to do well in mathematics". The reliabilities (Cronbach's alpha) were very good (see Table 3.1).

### 3.2.2.3 Migration background

Information from student and parent questionnaires was used to categorize students' migration background. Students were asked where they, their parents, and their grandparents were born, and which language they usually speak at home. The same information was requested from their parents. Data from the parent questionnaire were considered the most valid; in cases where they were missing, we resorted to data from the student questionnaire. Students were categorized as having a migration background when both of their grandparents, one of their parents, or the students themselves had been born outside Germany or when they usually speak a foreign language at home.

### 3.2.3 Missing Data

In our study, nonresponse rates on the indicators of self-efficacy ranged from 3.0% to 6.0% ( $M = 4.3$ ,  $SD = 0.9$ ) on the first and from 17.7% to 19.8% ( $M = 19.0$ ,  $SD = 0.6$ ) on the second measurement occasion. Using *Mplus* 7.11 (Muthén & Muthén, 1998-2012), we generated multiple imputations as this approach is considered to outperform more conventional methods of dealing with missing data (Enders, 2010). Plausible values are themselves imputed values and, as such, our achievement scores were complete data. We imputed the indicators of self-efficacy ten times per PV, leading to  $m = 50$  datasets used in our analyses. The results were averaged according to Rubin (1987).

### 3.2.4 Modeling Issues

We estimated structural equation models with auto-regressive paths using *Mplus* 7.11 (Muthén & Muthén, 1998-2012), following the guidelines presented by Marsh et al. (1999) to perform appropriate analyses using suitable data. In the structural part of the model, we estimated within-time correlations as well as stabilities, and cross-lagged relationships between the constructs. Because the self-efficacy items had a 4-point response format, we treated them as categorical. In our parametrization using thresholds, there was no identifiable residual variance (Byrne, 2011) and, therefore, correlated uniquenesses—as suggested in the guidelines of Marsh et al. (1999)—cannot be specified. We used the weighted least squares means and variances adjusted (WLSMV) estimation because, given our 4-point scales, using the WLSMV estimation was expected to lead to superior performance compared to using the maximum likelihood (ML) estimation (Beauducel & Herzberg, 2006). To adjust the clustered data structure (students were nested in classes),



we applied the type is complex option in our analyses. To reduce the covariance between the factors and thereby improve the convergence of our models, we centered the PVs by subtracting the mean and dividing them by the standard deviation of the first measurement occasion (i.e., [PV-100]/30 for all PVs). In total, we estimated four models for each domain: an overall model for all students (Model 1) and three multigroup models in which students were divided into groups according to migration background (Model 2), school track (Model 3), or both (Model 4).

### 3.3 Results

#### 3.3.1 Descriptive Statistics

The overall patterns of bivariate correlations were comparable between subsamples with and without a migration background (see Table 3.1). In both groups, there were small to moderate correlations between mathematics achievement and mathematics self-efficacy (from  $r = .20$  to  $r = .38$ ) as well as between reading achievement and reading self-efficacy (from  $r = .22$  to  $r = .33$ ). These correlations tended to be larger when measured concurrently rather than on different measurement occasions. The correlations between self-efficacy and achievement in different domains were low or even not statistically significant (from  $r = .05$  to  $r = .24$ ). The test-retest correlations of mathematics achievement ( $r = .87$ ) and reading achievement ( $r = .79$ ) were very high; mathematics self-efficacy and reading self-efficacy also correlated highly from one occasion to the other (from  $r = .53$  to  $r = .61$ ).

---Table 3.1 about here---

Regarding the means of the total sample, there was an increase in mathematics achievement (Wald's  $\chi^2(1) = 74.06, p < .001$ ) as well as in both mathematics self-efficacy (Wald's  $\chi^2(1) = 6.26, p = .01$ ) and reading self-efficacy (Wald's  $\chi^2(1) = 25.59, p < .001$ ), whereas the mean reading achievement was stable over time (Wald's  $\chi^2(1) = 0.52, p = .47$ ). Students with a migration background showed no statistically significant difference in mathematics self-efficacy compared to students without a migration background (T1: Wald's  $\chi^2(1) = 2.975, p = .09$ ; T2: Wald's  $\chi^2(1) = 0.200, p = .66$ ). Regarding reading self-efficacy and achievement in both domains, students with a migration background had significantly lower values on each measurement occasion than students without a migration background (all Wald's  $\chi^2(1) \geq 10.60, ps \leq .001$ ).

### 3.3.2 Measurement Invariance

Most of our estimated multigroup models with reciprocal relationships between achievement and self-efficacy showed at least acceptable  $\chi^2/df$ -ratios (from 1.70 to 3.24) and displayed good model fit indices (RMSEA  $\leq .05$ ; CFI and TLI  $\geq .97$ ; Schermelleh-Engel & Moosbrugger, 2003). Details of the model fit for all of the estimated models are depicted in Table 3.2. The  $\Delta CFI$  of the multigroup models (i.e., Model 2 to Model 4) was lower than .01 and, thus, the model fit was not meaningfully reduced as compared with the overall model (Cheung & Rensvold, 2002). This suggests a universality of the relationship between self-efficacy and achievement in the domains of mathematics and reading with respect to students in Germany with and without a migration background or attending different school tracks. However, the multigroup models in the domain of reading showed better model fit than the overall model ( $\Delta CFI > .01$ ). Therefore, in conjunction with students' migration background being a secondary objective of our study, we

complemented the results of Model 1 by reporting the results of Model 2 in which students were divided according to whether they had a migration background or not.

Because the model fit indices for the more parsimonious models with measurement and structural invariance were slightly better than those of the models with freely estimated structural parameters, we only report the results for the models with measurement and structural invariance in order to reduce the complexity.

---Table 3.2 about here---

### 3.3.3 Structural Models

The correlations and path coefficients of our overall structural model (Model 1) and the two-group models with students divided by migration background (Model 2) for the domains of mathematics and reading are depicted in Table 3.3. In the omnibus models, the correlations at T1 between self-efficacy and achievement were modest in both domains, whereas the residual correlations at T2 were not statistically significant. The stabilities for achievement and self-efficacy in each domain were high. Despite these high stabilities, mathematics self-efficacy at T1 had a small but statistically significant positive effect on mathematics achievement at T2, supporting the self-enhancement model and, thus, our Hypothesis 1b. In the domain of reading, reading achievement at T1 had a small but significant effect on reading self-efficacy at T2, supporting the skill development model and, thus, our Hypothesis 2a. However, both relationships were not statistically significant and, thus, our Hypotheses 1a and 2b could not be supported.

---Table 3.3 about here---

The stability coefficients for mathematics self-efficacy in the two-group model were  $\beta < .70$ , while the stabilities for mathematics achievement were  $\beta > .80$ . Despite

these high stabilities, in the domain of mathematics, the support for the hypothesis that mathematics self-efficacy has an effect on later mathematics achievement was also robust across both models as was the absence of a statistically significant effect of achievement on later self-efficacy.

The stability coefficients for reading self-efficacy were, overall, slightly lower than for mathematics. As in the omnibus model, reading achievement at T1 predicted reading self-efficacy at T2 in students with and without a migration background. Moreover, the two-group model revealed weak to small statistically significant cross-lagged effects from reading self-efficacy to later reading achievement, supporting the self-enhancement model and, thus, providing some evidence for reciprocal effects between reading self-efficacy and reading achievement.

### **3.4 Discussion**

The aim of our study was to find empirical support for the assertions that self-efficacy has positive effects on later achievement and achievement is a source of later self-efficacy (Bandura, 1977) in the domains of mathematics and reading, which would also support the assumption that there are reciprocal effects between domain-specific self-efficacy and achievement. However, in our sample, we did not find consistent empirical support for reciprocal effects between self-efficacy and achievement in either domain. Nonetheless, we did find consistent support for the self-enhancement model in the domain of mathematics (Hypothesis 1b) and for the skill development model in the domain of reading (Hypothesis 2a). In addition, when students were grouped together depending on whether or not they had a migration background, there were also significant

effects with regard to the self-enhancement model in the domain of reading (Hypothesis 2b) and, thus, there was some evidence for reciprocal effects.

Moreover, we established measurement and structural invariance by using multigroup models that considered students with and without a migration background, students attending different school tracks, and both constellations. Thus, for example, although students with a migration background showed lower achievement than their peers without a migration background, the effects of achievement on later self-efficacy in both domains were similar in these groups.

### **3.4.1 Relationship Between Mathematics Self-Efficacy and Mathematics Achievement**

In the overall model, the effect of mathematics self-efficacy on later mathematics achievement ( $\beta = .08$ ) was in line with the meta-analytical results of Valentine et al. (2004). This is remarkable because of the high stability of mathematics achievement in our models ( $\beta > .80$ ) compared to the meta-analysis, where stabilities of  $\beta = .62$  or above led to a mean effect size of  $\beta = .07$  (Valentine et al., 2004). Hence, although, in our results, there was not much variance in mathematics achievement left to be explained by prior mathematics self-efficacy, the robustness of the self-enhancement model is compelling.

However, finding support for the self-enhancement model without any support for the skill development model is peculiar, as evidence has been found for the consistent dominance of achievement over self-efficacy on a domain-general level until upper secondary school (Hwang et al., 2016) and over mathematics confidence from Grade 3 to Grade 8 (Ganley & Lubienski, 2016). Our results are also in contrast to the results of Hannula et al. (2014), who found reciprocal effects with a dominance of mathematics

achievement over mathematics self-efficacy (i.e., stronger relationship favoring the skill development model). They established intervals of three years between their three measurement occasions from Grade 3 to Grade 9 and found stabilities of mathematics achievement and mathematics self-efficacy that were  $\beta = .65$  or below. Moreover, both constructs showed large differences as their stabilities with mathematics achievement were higher than with mathematics self-efficacy. In their study, the dominance of achievement was negligible from the second to the third measurement occasion (i.e., from Grade 6 to Grade 9) when the difference between both constructs' stabilities was lower than that from the first to the second measurement occasion. Thus, the high difference in stabilities might explain the dominance of achievement over self-efficacy in the cross-lagged effects they found up until Grade 6.

On the other hand, our results are in line with the study of Marsh et al. (2005), which found support for reciprocal effects between mathematics self-concept and mathematics achievement, with more support for the self-enhancement model, in seventh graders in Germany. In a similar way, the initial support for the skill development model regarding mathematics achievement and confidence presented by Ganley and Lubienski (2016) up until Grade 5 changed to support for the REM up until Grade 8, indicating stronger support for the self-enhancement model with increasing age. Therefore, the support for the self-enhancement model in the domain of mathematics found in our study might be attributable to contextual variables, such as the school system in Germany, or might be specific to the age group, or may be due to both of these factors.

### 3.4.2 Relationship Between Reading Self-Efficacy and Reading Achievement

We did not find any previous studies on the reciprocal relationship between reading achievement and reading self-efficacy. However, our results that support the skill development model (Hypothesis 2a) correspond to the findings of Retelsdorf et al. (2014), who investigated the reciprocal relationship between reading achievement and reading self-concept in students in Germany of a similar age to the students in our sample.

Moreover, when dividing students into groups according to whether or not they had a migration background, our results supported the conviction that reading self-efficacy impacts later reading achievement. We revealed small reciprocal effects in the domain of reading over the course of Grade 7 in these groups. Because the effects of reading self-efficacy on later reading achievement were  $\beta = .06$ , the practical importance of the statistically significant effect in the two-group model seems to be low. The additional effect in relation to the self-enhancement model might be attributable to less stability in reading achievement than in the overall model, especially in the group of students with a migration background, and, therefore, to some additional variance that could be explained by previous reading self-efficacy.

Nevertheless, compared to mathematics, we did not find strong and consistent support for the self-enhancement model in the domain of reading. Our models favored different paths, depending on the domain under consideration. Had both domains been considered together, the results might have supported reciprocal effects on a domain-general basis, thus leading to false implications. Therefore, our results complement re-

search that emphasizes the necessity of domain-specific research regarding self-efficacy (Bong et al., 2012).

### **3.4.3 Limitations and Directions for Future Research**

Before coming to the practical implications of our study, we want to mention the shortcomings of our study, which also highlight possible directions that future research could encompass. The lack of consistent reciprocal effects between self-efficacy and achievement might be due to the interval between our measurement occasions. Short intervals often go along with high stability and, thus, little variance is left that could be explained by the cross-lagged effects; that is, the chances of detecting a cross-lagged effect are poor. Our results from the two-group model in the domain of reading suggest that reciprocal effects between reading achievement and reading self-efficacy might be detectable if the stabilities are below  $\beta = .80$ . As Marsh et al. (2005) stated, the interval between two measurement occasions may not be as critical as the amount of change that took place between them.

For reading achievement, where the basic decoding skills are taught and developed in primary school and more advanced reading comprehension skills might take longer to develop, it may thus be more fruitful to establish longer intervals between two measurement occasions in order to reveal an effect of reading self-efficacy when investigating secondary school students. In our study, the students did not show any improvement in their reading achievement, regardless of whether they had a migration background or not. The lack of improvement might be attributable to the fact that reading is only one part of language instruction, alongside listening, writing, and speaking, and, thus, less time is spent on developing reading comprehension skills. This stands in con-



trast to mathematics achievement where our tests reflected several different topics. Nevertheless, reading comprehension is an embedded part of other subjects taught at school and reading might also be part of students' leisure activities. Therefore, there are more opportunities to improve reading comprehension skills outside language instruction in school.

Studies reporting reciprocal effects between mathematics self-efficacy or mathematics confidence and mathematics achievement that used longitudinal data established intervals of more than one year between their measurement occasions (Ganley & Lubienski, 2016; Hannula et al., 2014). In contrast, Hwang et al. (2016) found evidence for reciprocal effects between academic self-efficacy and academic achievement with yearly assessments of secondary school students. Therefore, the assertion of Marsh et al. (2005) that the amount of change is more critical than mere intervals should be investigated in different educational systems, paying special attention to the curricula. However, this was not possible with our data and could be considered in applications of the REM in future investigations.

Investigating reciprocal effects between self-efficacy and achievement assumes that these constructs have mutual causal influences. However, as we stated at the beginning, the influence of self-efficacy on achievement is mediated by other variables, for example, engagement (Galla et al., 2014; see Honicke & Broadbent, 2016, for a discussion of other mediators of the relationship between academic self-efficacy and achievement). Likewise, achievement as a direct mastery experience is not the only source of later self-efficacy (Bandura, 1997; Usher & Pajares, 2008). With our data, we were not able to take mediating variables and different sources into account and, thus, the processes un-

derlying the causal link from self-efficacy to later achievement and vice versa could not be examined. Therefore, we are still uncertain about why mathematics self-efficacy had a significant effect on later mathematics achievement and why reading self-efficacy did not have an effect on later reading achievement in our overall models, although the stabilities of achievement were both around  $\beta = .85$ . Mediating variables that possibly reveal different processes in the domains of reading and mathematics could possibly provide further information to explain our results, which were in opposite directions in the two domains. Qualitative investigations in the domain of reading, like the work of Usher (2009) in the domain of mathematics, may provide useful insights into the sources of reading self-efficacy and perhaps allow for comparisons of the heuristics students use to develop their self-efficacy in different domains.

In the domain of mathematics, we found an effect of self-efficacy on later achievement in our structural models and an improvement in mathematics achievement in our descriptive results. In the domain of reading, there was neither an effect of reading self-efficacy on later reading achievement in the overall model nor an improvement of reading achievement. As engagement has been detected as a mediating variable of self-efficacy's positive effect on achievement (Galla et al., 2014) and engagement also showed an impact on reading achievement (Guthrie & Wigfield, 2000; Lee, 2013), the question of whether reading self-efficacy and mathematics self-efficacy exerted different influences on engagement and thereby on achievement remains open. Moreover, the comparison of the fit indices of our multigroup models supported the acceptance of structural invariance between students with or without a migration background and between students in different school tracks. Thus, the structural paths of our models could be con-

strained to be invariant. However, this does not provide any information about the underlying processes in the different groups or whether they can be seen as being equal. The interpretation of these findings is therefore somewhat limited. We can state, however, that the same causal relationships were found for students from different mean achievement levels (i.e., from different school tracks) and from different ethnic groups. The basic practical implications we consider below may therefore apply regardless of these affiliations. The question of whether interventions tailored specifically to these groups and concentrating on different processes in these groups could gain better results in terms of achievement and self-efficacy could be examined in future research. Investigations encompassing mediators and more than two measurement occasions seem to be a more viable approach to reveal the processes behind the reciprocal effects investigated in this study.

We concentrated on the reciprocal relationship between self-efficacy and achievement on a domain-specific level in an instructional setting. We also tested for moderating effects of students' migration background and the attended school track. This approach does not consider the possible influences of other moderating variables or third parties, such as family and peers. For example, vicarious learning from a sibling might be possible or successful verbal persuasion by peers or parents that boosts students' self-efficacy besides enactive mastery experience might take place. Especially in the domain of reading, the opportunity to read at home may be crucial in fostering both reading achievement and reading self-efficacy.

### 3.4.4 Practical Implications

The results of our study have some implications for educators. The consistent support we found for the self-enhancement model in the domain of mathematics could encourage teachers to foster their students' mathematics self-efficacy, for example, by creating a supportive classroom environment, by teaching effective life skills (Schunk & Meece, 2006) and, thereby, by improving their students' mathematics achievement. For the case of reading instruction, it might take more time for reading self-efficacy to exert an influence on reading achievement during secondary schooling. For this reason, it may be advisable to avoid fostering task-specific reading self-efficacy because it might be too short-lived to have a consistent influence on periods longer than one school year.

When dealing with the practical implications for educators, we want to emphasize the high stabilities of both self-efficacy and achievement in the two domains: Not only was achievement very stable within one school year; students' self-efficacy also appeared to be stable. Thus, we believe it is very important to accentuate the necessity of promoting achievement and even slightly optimistic self-efficacy beliefs right from the beginning of schooling, as it may no longer be possible to change either factors during secondary schooling. This is especially important with regard to the achievement gap between students with and without a migration background, which was constant across the school year. Students with a migration background were neither able to catch up with the mathematics achievement nor with the reading achievement of their peers who did not have a migration background. However, as our results suggest, there is still a chance for change despite the high stabilities, even within the course of one school year. Teachers could therefore be motivated to facilitate students' learning by using this chance. Con-

sidering the empirical evidence provided by previous research regarding the positive effect of engagement on reading achievement (Guthrie & Wigfield, 2000; Lee, 2013), and findings that self-efficacy has a positive influence on engagement (Galla et al., 2014), educators are advised to keep their students' reading self-efficacy in mind. This may be particularly true for students with a migration background because these students not only lagged behind their peers without a migration background in achievement but also in reading self-efficacy.

### 3.5 Conclusions

In sum, our study on the reciprocal effects between self-efficacy and achievement in the domains of mathematics and reading using a sample of seventh graders in Germany adds to the research on reciprocal effects between self-beliefs and achievement. Moreover, our results once again confirm that students with a migration background show lower mean levels of achievement than their peers without a migration background; this lower level remained stable over one school year. Thus, students with a migration background seem to have not tapped their full potential. Our main findings fit well to the results of self-concept research in reading (Retelsdorf et al., 2014) and mathematics (Marsh et al., 2005) using German samples with students of a similar age, thereby supporting the assumption that the self-concepts and self-efficacy in both domains appeared to make a difference in developing achievement. However, mathematics achievement benefitted consistently from mathematics self-beliefs, supporting the self-enhancement model, whereas in the domain of reading, we found some evidence for reciprocal effects between self-efficacy and achievement and consistent support for the skill development model. Thus, the significance of self-beliefs seems to vary between the domains at least

in seventh graders; this highlights the importance of a domain-specific approach in research and practice. In sum, our study supports the assertion that it does matter whether or not students are confident about dealing with the requirements made in school.

### 3.6 References

- Adams, R. J., Wu, M. L., & Wilson, M. R. (2012). ConQuest: ACER.
- Autorengruppe Bildungsberichterstattung. (2016). *Bildung in Deutschland 2016: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration* [Education in Germany 2016: An indicator-based report featuring an analysis of education and migration]. Bildung in Deutschland: Vol. 2016. Bielefeld: Bertelsmann. doi: 10.3278/6001820ew
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215. doi:10.1037/0033-295X.84.2.191
- Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist*, 33(4), 344–358.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall series in social learning theory. Englewood Cliffs, N.J: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(2), 186–203. doi:10.1207/s15328007sem1302\_2
- Bong, M. (2002). Predictive utility of subject-, task-, and problem-specific self-efficacy judgments for immediate and delayed academic performances. *The Journal of Experimental Education*, 70(2), 133–162. doi:10.1080/00220970209599503

- Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *The Journal of Educational Research*, *105*(5), 336–352. doi:10.1080/00220671.2011.627401
- Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, *34*(3), 139–153. doi:10.1207/s15326985ep3403\_1
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40. doi:10.1023/A:1021302408382
- Byrne, B. M. (2011). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge Academic.
- Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, *69*(2), 136–145. doi:10.1037/0022-0663.69.2.136
- Caprara, G. V., Vecchione, M., Alessandri, G., Gerbino, M., & Barbaranelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *British Journal of Educational Psychology*, *81*(1), 78–96. doi:10.1348/2044-8279.002004
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. doi:10.1207/S15328007SEM0902\_5



- Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, *64*(3), 830–847. doi:10.2307/1131221
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York, NY: Guilford Press.
- Galla, B. M., Wood, J. J., Tsukayama, E., Har, K., Chiu, A. W., & Langer, D. A. (2014). A longitudinal multilevel model analysis of the within-person and between-person effect of effortful engagement and academic self-efficacy on academic performance. *Journal of School Psychology*, *52*(3), 295–308. doi:10.1016/j.jsp.2014.04.001
- Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences*, *47*, 182–193. doi:10.1016/j.lindif.2016.01.002
- Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. doi:10.1037/0022-0663.95.1.124
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research, Vol. III* (pp. 403–422). New York; NY: Longman.
- Hannula, M. S., Bofah, E. A., Tuohilampi, L., & Metsämuuronen, J. (2014). A longitudinal analysis of the relationship between mathematics-related affect and achievement in Finland. In *Proceedings of the Joint Meeting of PME 38 and PME-NA 36* (pp. 249–256). Vancouver, Canada: PME. Retrieved from <https://www.unikas->

sel.de/fb10/fileadmin/datas/fb10/mathematik/didaktik/YERME\_Summer\_School/Exports/RR\_Bofah.pdf

- Honnicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review, 17*, 63–84. doi:10.1016/j.edurev.2015.11.002
- Hornstra, L., van der Veen, I., Peetsma, T., & Volman, M. (2013). Developments in motivation and achievement during primary school: A longitudinal study on group-specific differences. *Learning and Individual Differences, 23*, 195–204. doi:10.1016/j.lindif.2012.09.004
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*(5), 505–528. doi:10.1016/j.jsp.2011.07.001
- Hwang, M. H., Choi, H. C., Lee, A., Culver, J. D., & Hutchison, B. (2016). The relationship between self-efficacy and academic achievement: A 5-year panel analysis. *The Asia-Pacific Education Researcher, 25*(1), 89–98. doi:10.1007/s40299-015-0236-3
- Ireson, J., & Hallam, S. (2009). Academic self-concepts in adolescence: Relations with achievement and ability grouping in schools. *Learning and Instruction, 19*(3), 201–213. doi:10.1016/j.learninstruc.2008.04.001
- Jerusalem, M., & Satow, L. (1999). Schulbezogene Selbstwirksamkeitserwartung. In R. Schwarzer & M. Jerusalem (Eds.), *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen* (pp. 15–16). Berlin: R. Schwarzer.

- Kigel, R. M., McElvany, N., & Becker, M. (2015). Effects of immigrant background on text comprehension, vocabulary, and reading motivation: A longitudinal study. *Learning and Instruction, 35*, 73–84. doi:10.1016/j.learninstruc.2014.10.001
- Klassen, R. M. (2004). Optimism and realism: A review of self-efficacy from a cross-cultural perspective. *International Journal of Psychology, 39*(3), 205–230. doi:10.1080/00207590344000330
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., & Neubrand, M., . . . Weiß, M. (Eds.). (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente* [PISA 2000: Documentation of the data collection instruments]. Berlin, Germany: Max-Planck-Institut für Bildungsforschung.
- Lee, J.-S. (2013). The relationship between student engagement and academic performance: Is it a myth or reality? *The Journal of Educational Research, 107*(3), 177–185. doi:10.1080/00220671.2013.807491
- Lehmann, R. H., Gänsfuß, R., & Peek, R. (1999). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen — Klassenstufe 7. Bericht über die Untersuchung im September 1998* [Aspects of students' initial level and development at schools in Hamburg — Grade 7. Report on the study in September 1998]. Hamburg, Germany: Landesinstitut für Lehrerbildung und Schulentwicklung.
- Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82*(4), 646–656. doi:10.1037/0022-0663.82.4.646

- Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology*, *84*(1), 35–42.  
doi:10.1037/0022-0663.84.1.35
- Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, *34*(3), 155–167. doi:10.1207/s15326985ep3403\_2
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416. doi: 10.1037/0003-066X.33.4.344
- Mittag, W., Kleine, D., & Jerusalem, M. (2002). Evaluation der schulbezogenen Selbstwirksamkeit von Sekundarschülern. *Zeitschrift für Pädagogik, Beiheft*. (44), 145–173.
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30–38. doi:10.1037/0022-0167.38.1.30
- Muthén, L. K. & Muthén, B. O. (1998-2012). Mplus User's Guide. Los Angeles, CA: Muthén & Muthén. Retrieved from [www.statmodel.com](http://www.statmodel.com)
- Oettingen, G. (1995). Cross-cultural perspectives on self-efficacy. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 149–176). Cambridge; UK: Cambridge University Press. doi: 10.1017/CBO9780511527692.007
- Organisation for Economic Co-operation and Development. (2005). *PISA 2003 technical report*. Paris, France: OECD.

- Pajares, F. (1996a). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543–578. doi:10.3102/00346543066004543
- Pajares, F. (1996b, April). *Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence*. Annual meeting of the American Educational Research Association. 1996 AERA Annual Meeting, New York.
- Piercey, R. R. (2013). *Reading self-efficacy in early adolescence: Which measure works best?* (Dissertation). College of Education at the University of Kentucky, Lexington, Kentucky. Retrieved from [http://uknowledge.uky.edu/edp\\_etds/10/](http://uknowledge.uky.edu/edp_etds/10/)
- Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, 84(1), 152–174. doi:10.1111/bjep.12028
- Preckel, F., Niepel, C., Schneider, M., & Brunner, M. (2013). Self-concept in adolescence: a longitudinal study on reciprocal effects of self-perceptions in academic and social domains. *Journal of Adolescence*, 36(6), 1165–1175. doi:10.1016/j.adolescence.2013.09.001
- Retelsdorf, J., Köller, O., & Möller, J. (2014). Reading achievement and reading self-concept – Testing the reciprocal effects model. *Learning and Instruction*, 29, 21–30. doi:10.1016/j.learninstruc.2013.07.004
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys. Wiley series in probability and mathematical statistics. Applied probability and statistics*. New York, NY: John Wiley & Sons. doi: 10.1002/9780470316696

- Schermelleh-Engel, K., & Moosbrugger, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research online*, 8(2), 23–74. Retrieved from [http://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130\\_13.pdf](http://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf)
- Schöber, C., Retelsdorf, J., & Köller, O. (2015). Verbales schulisches Selbstkonzept und sprachliche Leistungen in Gruppen mit und ohne Migrationshintergrund [Verbal academic self-concept and achievement in groups with and without migration background]. *Psychologie in Erziehung und Unterricht*, 62(2), 89–105.  
doi:10.2378/peu2015.art10d
- Schunk, D. H., & Meece, J. L. (2006). Self-efficacy development in adolescences. In F. Pajares & T. Urdan (Eds.), *Adolescence and education: Vol. 5. Adolescence and education. Vol. 5: Self-efficacy and adolescence* (pp. 71–96). Greenwich, CT: Information Age Publishing.
- Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Educational psychology handbook series. Handbook of motivation at school* (pp. 35–54). New York, NY: Routledge.
- Usher, E. L. (2009). Sources of middle school students' self-efficacy in mathematics: A qualitative investigation. *American Educational Research Journal*, 46(1), 275–314.  
doi:10.3102/0002831208324517
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, 78(4), 751–796.  
doi:10.3102/0034654308321456

- Valentine, J. C., & Dubois, D. L. (2005). Effects of self-beliefs on academic achievement and vice versa: Separating the chicken from the egg. In H. W. Marsh, R. Craven, & D. M. McInerney (Eds.), *International advances in self-research: Vol. 2. New frontiers for self-research* (pp. 53–77). Greenwich, CT: Information Age Publishing.
- Valentine, J. C., Dubois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111–133. doi:10.1207/s15326985ep3902\_3
- Wigfield, A., & Karpathian, M. (1991). Who am I and what can I do?: Children's self-concepts and motivation in achievement situations. *Educational Psychologist, 26*(3-4), 233–261. doi:10.1080/00461520.1991.9653134
- Williams, T., & Williams, K. (2010). Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology, 102*(2), 453–466. doi:10.1037/a0017271
- Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202–231). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511527692.009

Table 3.1

Means, standard deviations, reliabilities (Cronbach's alpha for self-efficacy scales and EAP/PV for achievement plausible values), and manifest correlations of mathematics achievement, reading achievement, mathematics self-efficacy, and reading self-efficacy

Variable	T1 MAch	T2 MAch	T1 RAch	T2 RAch	T1 MSEff	T2 MSEff	T1 RSEff	T2 RSEff	
T1 MAch	-	.87***	.79***	.70***	.30***	.20***	.07	.12**	
T2 MAch	.87***	-	.74***	.69***	.32***	.25***	.12**	.15**	
T1 RAch	.79***	.74***	-	.79***	.13***	.05	.25***	.22***	
T2 RAch	.69***	.67***	.79***	-	.15***	.08	.24***	.23***	
T1 MSEff	.35***	.38***	.16***	.17***	-	.58***	.28***	.22***	
T2 MSEff	.25***	.30***	.10*	.11*	.61***	-	.09*	.28***	
T1 RSEff	.16***	.20***	.34***	.30***	.18***	.04	-	.53***	
T2 RSEff	.20***	.24***	.31***	.33***	.16**	.18***	.55***	-	
<i>M</i>	Total	100.00	104.52	100.00	99.55	1.95	2.01	2.08	2.17
	GER	113.09	116.86	115.34	114.73	2.00	2.02	2.16	2.25
	MIG	91.25	96.26	89.75	89.40	1.92	2.00	2.03	2.12
<i>SD</i>	Total	29.99	29.29	29.99	31.00	0.70	0.82	0.61	0.63
	GER	29.01	28.28	26.47	27.12	0.69	0.82	0.59	0.64
	MIG	27.33	26.96	27.75	29.24	0.71	0.81	0.62	0.62
Reliability		.96	.91	.94	.85	.90	.94	.88	.88

Note. MAch = mathematics achievement, RAch = reading achievement, MSEff = mathematics self-efficacy, RSEff = reading self-efficacy.

Correlations for students without a migration background (GER) below the diagonal and for students with a migration background (MIG) above the diagonal; reliabilities are reported for the complete sample.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .



Table 3.2

Model fit statistics of different reciprocal effects models with group-invariant parameters in measurement models or in measurement and structural models in mathematics and reading

	$\chi^2/df$		RMSEA		CFI		TLI	
	MI	SI	MI	SI	MI	SI	MI	SI
<i>Mathematics</i>								
Model 1		4.78		.049		.990		.991
Model 2	2.66	1.99	.046	.035	.992	.995	.993	.996
Model 3	2.52	2.62	.044	.045	.993	.992	.993	.993
Model 4	1.77	1.71	.044	.042	.993	.993	.994	.994
<i>Reading</i>								
Model 1		5.89		.055		.971		.972
Model 2	3.24	2.55	.053	.044	.977	.983	.979	.985
Model 3	2.96	2.49	.050	.043	.979	.984	.981	.986
Model 4	1.97	1.69	.049	.042	.980	.986	.983	.988

*Note.* Model 1: Omnibus model; Model 2: two-group models for groups of students with or without a migration background; Model 3: two-group models for groups of students at the academic school track and at other school tracks; Model 4: four-group models for groups of students at the academic school track and at other school tracks and with or without a migration background;  $\chi^2/df$  = quotient of  $\chi^2$ -statistic and degrees of freedom;  $df = 96$  for Model 1;  $df = 201$  (MI) or 207 (SI) for Model 2 and Model 3;  $df = 411$  (MI) or 429 (SI) for Model 4; RMSEA = Root Mean Square Error of Approximation; CF = Comparative Fit Index; TLI = Tucker-Lewis Index; MI = Invariant measurement model; SI = Invariant measurement and structural model.

← Formatierte Tabelle

Table 3.3  
Results of reciprocal effects models in mathematics and reading

Model	Group	SEff T1 ↔		SEff T2 ↔		SEff T1 →		Ach T1 →		Ach T1 →		SEff T1 →	
		Ach T1 <i>r</i>	<i>S.E.</i>	Ach T2 <i>r</i>	<i>S.E.</i>	SEff T2 $\beta$	<i>S.E.</i>	Ach T2 $\beta$	<i>S.E.</i>	SEff T2 $\beta$	<i>S.E.</i>	Ach T2 $\beta$	<i>S.E.</i>
Mathematics													
Model1		.33***	0.03	.07	0.05	.66***	0.02	.86***	0.01	.02	0.03	.08***	0.02
Model2	GER	.31***	0.03	.09	0.05	.69***	0.03	.84***	0.03	.03	0.03	.09***	0.02
	MIG	.34***	0.03	.07	0.05	.64***	0.03	.82***	0.02	.02	0.03	.09***	0.02
Reading													
Model1		.31***	0.03	.01	0.06	.59***	0.02	.84***	0.02	.14***	0.03	.03	0.02
Model2	GER	.29***	0.03	.08	0.06	.62***	0.03	.80***	0.03	.11***	0.03	.06*	0.03
	MIG	.30***	0.03	.06	0.05	.59***	0.02	.73***	0.04	.11***	0.03	.06*	0.03

*Note.* Standardized coefficients of structural equation models estimated using measurement and structural invariance across time and groups, model fit indices are displayed in Table 2;  
SEff = Self-efficacy beliefs, Ach = Achievement, GER = Students without a migration background, MIG = Students with a migration background. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

# **4**

## **Studie 3: Teachers' judgments of students' self-efficacy in reading: Accuracy and students' migrant origins**

Schöber, C., Köller, O., Schütte, K., Huelmann, T., Gebauer, M. M., Ferdinand, H. D. et al.

(Manuscript in preparation for publication). Teachers' judgments of students' self-efficacy in reading: Accuracy and students' migrant origins.

## Zusammenfassung

Um ihren Unterricht gestalten zu können, benötigen Lehrkräfte eine Bandbreite an Wissen über Eigenschaften ihrer Schulkinder. Das Wissen über die Selbstwirksamkeitsüberzeugungen der Schulkinder bilden sich Lehrkräfte durch eigene Urteile. Wir haben die Urteile in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen von Schulkindern mit einem Fokus auf Schulkinder mit Migrationshintergrund an einer Stichprobe von  $N = 1573$  Siebtklässlerinnen und Siebtklässlern und ihren 112 Deutschlehrkräften untersucht. Der mittlere Zusammenhang zwischen den Lehrkrafturteilen und den entsprechenden Schulkindselbsteinschätzungen war schwach ( $r = .23$ ), was eine geringe Urteilsgenauigkeit indiziert. Darüber hinaus zeigten Ergebnisse aus Zwei-Ebenen-Regressionsanalysen, dass die Lehrkräfte ihre Urteile am Migrationshintergrund der Schulkinder ausrichteten. Unter Kontrolle der Leseleistung erhielten Schulkinder ohne Migrationshintergrund und Schulkinder aus den Gebieten der ehemaligen Sowjetunion und aus Polen geringere lesespezifische Selbstwirksamkeitsüberzeugungen zugeschrieben als Schulkinder mit einem anderweitigen Migrationshintergrund. Für Schulkinder mit türkischem Migrationshintergrund konnten wir zusätzlich einen Interaktionseffekt mit dem Leseleistungsurteil der Lehrkräfte aufdecken. Dieser Befund weist auf einen geringeren Zusammenhang mit dem Niveau der Leseleistung bei den Schulkindern mit türkischem Migrationshintergrund im Vergleich zu den anderen Gruppen hin. Trotz der insgesamt geringen Urteilsgenauigkeit legen unsere Ergebnisse nahe, dass Lehrkräfte unterschiedliche Beziehungen zwischen Leseleistungen und lesespezifischer Selbstwirksamkeitsüberzeugung bei der Urteilsfindung einbeziehen.

## 4.1 Introduction

Teachers can foster their students' reading achievement and reading motivation by providing effective instruction (Guthrie et al., 2004). Effective instruction, however, is based on teachers' knowledge about their students' learning prerequisites (Givvin, Stipek, Salmon, & MacGyvers, 2001). One such prerequisite is students' domain-specific self-efficacy. Instructional methods incorporating modeling strategies, progress feedback, goal setting, and self-evaluations of progress can be used to enhance students' self-efficacy and achievement (Schunk, 2003). The higher students' self-efficacy the more self-regulative strategies to promote their academic achievement they use (Pajares & Schunk, 2003; see Schunk & Pajares, 2009, for an overview). In contrast, students lacking self-efficacy may refuse to work on a given task although they possibly have the skills to complete it. The more accurate teachers' knowledge about students' domain-specific self-efficacy the better they are equipped to adjust their instruction and thereby promote student achievement.

To obtain accurate knowledge on their students' self-efficacy teachers have to judge it using everyday lessons. Teachers' judgments may be influenced by individual teacher and student characteristics (Südkamp, Kaiser, & Möller, 2012) such as students' migration background. In the German student population of 15-year-olds about one quarter has a migration background of which the two largest immigrant groups stem either from Turkey or from the successor states of the Former Soviet Union or Poland (Gebhardt, Rauch, Mang, Sälzer, & Stanat, 2013). These two groups are most often descendants of a) so-called guest workers (i.e., people having migrated to Germany during the economic miracle in the 1950s to 1970s) stemming from Mediterranean countries like

Turkey or b) so-called ethnical German resettlers (i.e., descendants of Germans that settled in Eastern Europe in the 18th century) who migrated from the successor states of the Former Soviet Union or Poland after the second world war. Students with migration backgrounds are lagging behind their peers without migration background regarding their achievement (Gebhardt et al., 2013), especially within the language domain (Stanat & Christensen, 2006). At the same time they report similar or even higher levels of motivation (Schleicher, 2006). As teachers interpret actual achievement as an indicator of the level of motivation (Praetorius, Greb, Lipowsky, & Gollwitzer, 2010) the differential relations of self-reported motivation and achievement between students with different migration backgrounds may pose an additional challenge for teachers when judging their students' self-efficacy in reading—different levels of reading achievement might be related to the same level of self-efficacy in reading.

To investigate teachers' judgments of students' self-efficacy in reading we applied a powerful multilevel approach simultaneously analyzing the impact of predictors on the individual student level and the classroom level. Our findings thus advance the scientific understanding of teachers' judgments and may ultimately contribute to reducing the gap between students with and without a migration background regarding educational attainment.

#### **4.1.1 Self-Efficacy Beliefs**

Self-efficacy beliefs evolve in different areas of human functioning, for example, social, physical, health, and academics. Academic self-efficacy refers to “personal judgments of one's capabilities to organize and execute courses of action to attain designated types of educational performances” (Zimmerman, 1995, p. 203). Self-efficacy be-

liefs are hypothesized to influence level of effort, persistence, and choice of activities (Bandura, 1977). Galla, Wood, Tsukayama, Har, Chiu, and Langer (2014) partially confirmed Bandura's (1977) claim for academic self-efficacy by finding that students' scores in standardized reading and mathematics tests were higher the higher their self-reported academic self-efficacy was; teacher-rated students' effortful engagement in academic situations (including effort, persistence, attention, and class participation) mediated this effect. Thus, academic self-efficacy is one important precursor of engagement; engaging in effortful learning behavior, in turn, leads to higher academic achievement (see also Hughes, Luo, Kwok, & Loyd, 2008).

Within academics, self-efficacy beliefs can further be differentiated by domains (e.g., mathematics, science, language) or subjects (e.g., algebra, geometry, reading, writing). Valentine, DuBois, and Cooper (2004) compared the relations between self-beliefs and achievement associated with level of measurement in a meta-analytical review. Controlling for the type of self-belief measured (i.e., self-efficacy, self-concept, self-esteem) effect sizes were larger for measures of the self conducted at an academic level ( $\beta = .13$ ) compared to measures conducted at either a subject-specific ( $\beta = .06$ ) or a global ( $\beta = .07$ ) level (Valentine, DuBois, & Cooper, 2004). Comparing domain-specific with global measures Bong, Cho, Ahn, and Kim (2012) found evidence that domain-specific assessment of self-beliefs leads to higher relations with achievement than general self-esteem as a global measure. Piercey (2013) investigated the predictive validity of several scales for the assessment of self-efficacy in reading. Her results emphasized the importance of corresponding types of self-efficacy for reading (i.e., general reading self-efficacy, reading skills self-efficacy, self-efficacy for self-regulation in reading) and achievement out-

comes (i.e., language arts grades, scores in a standardized reading test, minutes read daily).

In the domain of reading, girls sometimes showed higher self-efficacy than boys (Coddington & Guthrie, 2009). In other studies, however, gender differences were minimal (Smith, Smith, Gilmore, & Jameson, 2012) or just found in one of two measurements (Wigfield & Guthrie, 1997). Piercey (2013) reviewed individual differences of students' reading self-efficacy and reading achievement with respect to gender, ethnicity, and age. She found inconsistent results in the literature. Investigating potential differences with her own sample of  $N = 364$  students from Grade 4 to Grade 6 from schools in the Southeastern United States with several reading-specific self-efficacy scales she only revealed significant mean differences in younger students showing higher levels of reading test self-efficacy than older students, but no significant differences regarding gender or differences between American students of European descent compared to American students of African descent. Likewise, Valentine et al. (2004) found no significant relations of gender or ethnicity to effect size in their meta-analysis.

#### **4.1.2 Teachers' Judgment Accuracy**

One facet of teacher professionalism affecting students' learning outcomes is their judgment accuracy (Artelt & Rausch, 2014); ultimately it thus affects educational participation and success. Everyday tasks of teachers include judging their students in several areas. The most obvious of these is student achievement. These achievement judgments are not only important for grading, but also for instructional decisions (e.g., selecting tasks of varying difficulty for individual students, organizing learning groups) and



can have immediate as well as long-term effects (e.g., by suggesting retention or remediation courses; Begeny, Eckert, Montarello, & Storie, 2008).

Because some instructional decisions are not only based on mere achievement teachers should also accurately judge other student characteristics, particularly motivational and affective characteristics. In Germany for example teachers' decisions regarding students' transition from primary to secondary schools is mainly based on school grades, but also on perceived parental support, school anxiety, persistence, and effort (Gräsel, Krolak-Schwerdt, Nölle, & Hörstermann, 2010; for fictive examples of teachers misjudging their students motivation in everyday instruction see Givvin et al., 2011). Note that teachers' ratings of students' effortful engagement mediated the relation between students' academic self-efficacy and students' scores in a standardized reading test (Galla et al., 2014).

Teachers' judgment accuracy is supposed to be affected by several individual and technical influences (Südkamp et al., 2012). Individual influences of teachers and students are characteristics such as age, gender, and ethnicity. For instance, Ready and Wright (2011) found that teachers systematically underestimated basic literacy skills among American children of African descent (effect size = -0.13) and of Spanish descent (effect size = -0.11) compared to their classmates of European descent. In an experimental setting in Germany, Glock and Krolak-Schwerdt (2013) showed that student teachers' judgments of students' competence were also affected by stereotypical beliefs. In their study they described fictive students as below-average or as best in class but varied their first name and the language spoken at home and during leisure time (Turkish vs. German). Below-average and thus stereotype-consistent students of Turkish descent

were judged as less competent in language comprehension, reading comprehension, and spelling competency than below-average but thereby stereotype-inconsistent students without migration background, whereas there was no significant difference in the competence judgments of stereotype-inconsistent (i.e., best in class) students of Turkish descent and stereotype-consistent (i.e., best in class) students without migration background. This finding supports the necessity to shed light on moderating effects of students' migrant origins on teachers' judgments.

Factors related to teachers' professionalism are possibly additional moderators. Especially job experience is assumed to enhance teachers' professional competence (McElvany et al., 2009), but empirical evidence for a moderating effect of job experience on judgment accuracy of students' achievement is inconsistent (Demaray & Elliott, 1998; McElvany et al., 2009; Ready & Wright, 2011). To our knowledge, a possible effect of job experience on teachers' judgments of students' self-efficacy beliefs has not yet been investigated.

The technical influences on teachers' judgment accuracy, namely test characteristics and judgment characteristics, are far better researched (see Südkamp et al., 2012, for an overview). One of the judgment characteristics is whether teacher judgments are informed or uninformed; informed judgments are distinguished from uninformed judgments by the availability of a standard of comparison. Informed teacher judgments (e.g., asking teachers to give an estimate of the number of solved items on a standardized achievement test after showing example items) showed higher correlations with students' academic achievement than uninformed teacher judgments (e.g., a judgment on a Likert-type rating scale) did (Südkamp et al., 2012). A second judgment characteristic, congru-

ence in domain-specificity of the achievement test and the teacher rating (e.g., comparing reading achievement with judgments of reading achievement as high and comparing reading achievement with judgments in language as low congruence in domain-specificity), led to larger effect sizes compared to studies with incongruence in domain-specificity (Südkamp et al., 2012). Therefore, informed teacher judgments with high congruence in domain-specificity should be assessed to avoid an underestimation of their accuracy.

#### 4.1.2.1 Components of judgment accuracy

Teachers' judgment accuracy is typically assessed by evaluation of three descriptive components: (1) the level component, (2) the differentiation component, and (3) the rank component (Schrader & Helmke, 1987). All three components contribute unique information and are therefore important to gain full insight into the accuracy of teachers' judgments. Teachers with high accuracy on the level component are well able to judge the absolute levels of their students on a certain construct. High accuracy regarding the differentiation component means that teachers accurately assess the construct's variance in their classrooms. The rank component signifies how accurately teachers rank order their students. The rank component has been stated to reflect the diagnostic sensitivity in the classical sense (Schrader & Helmke, 1987) and is most frequently reported when teachers' judgment accuracy is analyzed (Südkamp et al., 2012).

Advanced research on teachers' judgment accuracy applied multilevel models in addition to or instead of reporting the three descriptive components. Multilevel models are able to simultaneously model influences on the student (within) level and the teacher or classroom (cluster or between) level (Raudenbush & Bryk, 2002; see Ready & Wright, 2011, for an application of multilevel modeling in this research context). Student level

variables such as students' gender, age, or migrant origin and cluster level variables such as teachers' job experience can simultaneously be included in multilevel models to provide an insight into variables potentially affecting teachers' judgments.

#### 4.1.2.2 Empirical findings on teachers' judgment accuracy

Studies on teachers' judgment accuracy regarding student achievement typically find an overestimation of students' mean scores (Bates & Nettelbeck, 2001; Madelaine & Wheldall, 2005) and an underestimation of the variance in the classroom (Helmke, Hosenfeld, & Schrader, 2004; Spinath, 2005). Teachers especially showed a tendency to overestimate the reading achievement of students whose actual achievement was low (Bates & Nettelbeck, 2001; Begeny et al., 2008; Feinberg & Shapiro, 2009). Two meta-analyses demonstrated moderate teacher ability to rank order students' academic achievement (Hoge & Coladarci, 1989; Südkamp et al., 2012).

For fifth graders' reading achievement in German (i.e., the respective language of instruction), Karing, Matthäi, and Artelt (2011) found a mean rank component (Pearson correlation) of  $r = .34$  ( $SD = .30$ ) and Rjosk, McElvany, Anders, and Becker (2011) found a mean Spearman rank correlation of  $r = .60$  ( $SD = .37$ ) for the rank orders of seven randomly chosen sixth graders per classroom and their teacher-judged rank orders. One possible reason for the notable effect size difference may be differences in schooling between both samples. The sample for the study of Karing et al. (2011) was drawn in the first year of secondary school in Bavaria and Hesse, where students leave primary school after Grade 4. Rjosk et al. (2011) drew their sample in Berlin where sixth graders are in their last year of primary school. Thus, the difference in the mean rank components of these studies may be an effect of teachers in the study of Rjosk et al. (2011) having taught

their students for a longer period than those in the study of Karing et al. (2011). However, there were no differences in judgment accuracy between primary school teachers that had already taught their students for two years (4<sup>th</sup> Graders in Bavaria) compared to primary school teachers that had already taught their students for four years (4<sup>th</sup> Graders in Hesse; Karing, 2009). Rather, it may be possible that differences in teacher education between primary and secondary school teachers or effects of ability grouping and accordingly possible smaller variance in achievement in secondary schools led to these differences (see also Karing, 2009; Karing et al., 2011). Both explanations are in line with Karing (2009) who compared judgments from teachers at primary schools and academic track schools and reported higher rank components for primary school teachers (mean  $r = .61$ ) compared to teachers at academic track schools (mean  $r = .40$ ) in their judgments of students' text comprehension.

Teachers' judgment accuracy of students' self-beliefs has been far less investigated and the results of these studies are not consistent. Marsh and Craven (1991) found correlations between students' self-reports and teachers' judgments on reading, mathematics, and general school self-concept of  $r = .39$ ,  $r = .47$ , and  $r = .62$ , respectively. Spinath (2005) found a reverse pattern for judgments of students' academic self-concepts as compared with the common result for achievement: Teachers underestimated the level and overestimated the variance of their students' academic self-concepts. The rank component was modest with a median of  $r = .39$ , but ranged from  $r = -.39$  to  $r = .82$ . In effect, some teachers rank-ordered their students far less accurately than they would have had they been simply guessing. For domain-specific self-concepts Praetorius, Karst,

Dickhäuser, and Lipowsky (2011) found correlations of  $r = .25$ ,  $r = .52$ , and  $r = .55$  in writing, reading, and mathematics, respectively.

For other motivational constructs related to student achievement like student engagement teachers' judgment accuracy was equally low (Kaiser, Retelsdorf, Südkamp, & Möller, 2013). Moreover, in the Kaiser et al. study structural equation modeling revealed that teachers' judgments of engagement and achievement were each not only positively related to the respective student characteristic but also to the non-corresponding characteristic. These mutual relations of student achievement and self-reported engagement with teachers' judgments of achievement and engagement from a field study were replicated by two experimental studies in which student achievement and student engagement were uncorrelated (Kaiser et al., 2013).

Regarding self-efficacy, Huelmann et al. (2014) demonstrated that teachers significantly underestimated the level of their students' mathematics self-efficacy, but on average correctly judged the variance of the mathematics self-efficacy within their classroom. The rank component for general academic self-efficacy was fairly low; a statistically significant positive correlation coefficient was observed for only 7 out of the 37 teachers, suggesting that teachers mainly guess their students' academic self-efficacy. For reading self-efficacy of first graders Coddington and Guthrie (2009) found a correlation of  $r = .37$  between students' self-reports and teachers' judgments and no significant difference regarding the levels of self- and other-reported self-efficacy in reading.

Praetorius et al. (2010) found that students' mathematics achievement as measured by a standardized test was a stronger predictor than students' self-reported mathematics self-concept for primary school teachers' judgments of students' mathematics self-

concepts. This indicates that teachers use more obvious cues like achievement to judge not directly accessible constructs of student motivation. During daily instruction teachers may refer to the quality and intensity of educational engagement or to their achievements on quizzes and exams when judging students' motivational constructs.

#### **4.1.3 Realistic Accuracy Model**

Besides methodological issues due to unreliability of the assessments, a helpful approach to explain the less accurate judgments of students' motivational variables compared to achievement is the realistic accuracy model (RAM; Funder, 1995) which specifies prerequisites of accurate personality judgment. The RAM describes interpersonal (i.e., judge – target) and cognitive processes that lead to more or less accurate judgments depending on the success while completing four stages. The four stages of the RAM are based on the assumption that a trait produces a behavioral effect in some context. A behavior must be (1) relevant to that trait, (2) available to the judge, (3) detected by the judge, and (4) correctly utilized by the judge to produce an accurate judgment of the trait. The model posits that each stage makes an indispensable contribution to judgment accuracy; if either one is absent, correct judgments are impossible. Applying the four stages to teachers' judgments of student motivation, an example could be that a teacher recognizes that a student reads complicated texts during leisure time and the teacher interprets this behavior as indicating higher reading achievement levels than expected before, but does not consider this cue as also indicating high or fairly optimistic levels of self-efficacy in reading. In this case, the stage of utilization equals zero for self-efficacy leading to a zero increase of judgment accuracy of students' self-efficacy in reading, although the judgment accuracy of reading achievement may increase at the same time.

#### 4.1.4 The Present Investigation

To sum up existing research, the judgment of student achievement seems much easier for teachers than judging motivational characteristics. Looking at less easily observable motivational constructs, teachers are far less accurate. Little research exists for self-efficacy beliefs although academic self-efficacy is an important precursor of students' engagement and achievement in reading (Galla et al., 2014).

To the extent that teachers use student achievement as an indicator of their self-efficacy, comparable levels of self-evaluations despite different levels of achievement between (groups of) students pose an additional challenge for teachers' judgments of students' self-related beliefs. If migrant origin was a student characteristic moderating self-efficacy's relation to achievement in the respective domain and teachers took these differential relations into account when judging students' self-efficacy this should be reflected in a significant interaction term between students' migration background and teachers' judgments of their students' achievement. As teachers use obvious cues (see RAM) students' achievement and individual characteristics are expected to be the strongest predictors of teacher-reported student self-efficacy.

Therefore, the objectives of our study are two-fold. First, we report the traditional components of judgment accuracy (i.e., level, differentiation, and rank component) on a large sample of students to allow for comparisons with previous findings on different constructs. Given the results of previous research on teachers' judgment accuracy of students' motivation we assume that teachers underestimate the level and overestimate the variance of their students' self-efficacy in reading. The rank component is assumed to be positive but weak.



Second, we provide the first multilevel approach to analyzing teachers' judgments of students' self-efficacy in the domain of reading simultaneously considering predictors and moderators of teachers' judgments of their students' self-efficacy in reading. Specifically, we explore if students' migrant origin and teachers' judgments of students' reading achievement as obvious cues with different relations to self-efficacy beliefs interact on teachers' judgments of their students' self-efficacy in reading.

## 4.2 Method

### 4.2.1 Sample

The data stem from the study [name deleted to maintain the integrity of the review process]. We investigated a stratified sample of seventh-grade students from four federal states and three school tracks in Germany. The vocational school track is the lowest track in Germany and students leave school after Grade 9 or 10 to start vocational training. Students on the comprehensive school track take somewhat ambitious courses through Grade 10 and then have the choice to leave school to start vocational education or to continue schooling at upper secondary level. Finally, students who graduate from academic track schools after Grade 12 or Grade 13 are qualified to enter higher education. Sampling included one intact class and additional students with a migrant origin in Turkey or the Former Soviet Union or Poland from another class of each participating school. We discarded clusters with less than three students to avoid perfect correlation coefficients within clusters.

As a result, our final sample consisted of  $N = 1573$  seventh-grade students (49.5% female) with a mean age of 12.15 years ( $SD = 0.63$ ) from 112 clusters with  $n = 3$

to  $n = 30$  students (average cluster size  $n = 14.05$ ) and their  $N = 112$  teachers who teach German (82.3% female) with a mean age of 43.23 years ( $SD = 11.77$ ) and a mean teaching experience of 13.59 years ( $SD = 11.80$ ). The students attended the vocational school track ( $n = 432$  students from 40 clusters), the comprehensive school track ( $n = 454$  students from 30 clusters), or the academic school track ( $n = 687$  from 42 clusters).

Our sample comprised three groups of students with different migrant origins. Students of Turkish descent ( $n = 403$ ) or with migrant origin in the Former Soviet Union or Poland ( $n = 230$ ) represent the two largest demographic groups with migrant origins living in Germany. Students in the third group ( $n = 302$ ) have migrant origins in a variety of other countries. A fourth group of students ( $n = 638$ ) had no migration background. More detailed information on our sample regarding migrant origin and school track is depicted in Table 4.1 (see section *Descriptive Results*). Whereas the majority of students without a migration background attended academic track schools, students with a migration background were more likely to attend vocational track or comprehensive track schools. This difference in educational participation is typical for the German school system (Autorengruppe Bildungsberichterstattung, 2014).

#### 4.2.2 Procedure

Data were collected in November 2012. Trained research assistants administered tests and student questionnaires during regular school lessons in the classrooms. Teachers received information about the study and a questionnaire to be returned within two weeks' time. We also informed the parents about the study and asked them to consent to their child's participation. Parents also received a short questionnaire. Participation of students, parents, and teachers was voluntary. Only students with parental consent

participated in this study. Neither the teachers nor the students received personal incentives for their participation, but each school was rewarded 500 Euros.

### 4.2.3 Measures

#### 4.2.3.1 Self-efficacy in reading.

Self-efficacy in reading was assessed by slightly adapted well-established scales (e.g., Kunter et al., 2002). There were 6 self-efficacy items with a 4-point response scale (1 = *not at all true*, 4 = *exactly true*). An example item was: “I’m certain I can understand the material presented in readings that is very difficult to read”.

To establish high congruence in domain-specificity, teachers received similarly worded items assessing their students’ self-efficacy to be answered on the same response scale. The teachers were asked to rate these items for every student in their class who answered the student questionnaire. An example item was: “The student is certain that she/he can understand the material presented in readings that is very difficult to read”. The reliabilities of the instruments were satisfactory (Cronbach’s  $\alpha = .88$  for student questionnaire;  $\alpha = .94$  for teacher questionnaire).

#### 4.2.3.2 Reading achievement.

We assessed students’ reading achievement with a standardized reading test. The items stem from German large-scale studies on “Aspects of Students’ Initial Level and Development at Schools in Hamburg” (Lehmann, Gänßfuß, & Peek, 1999; Lehmann & Peek, 1997; Lehmann, Peek, Gänßfuß, & Husfeldt, 2002) and nationwide standardized tests for the whole population of eighth graders in Germany (Vergleichsarbeiten VERA-8

of the Institute for Educational Quality Improvement, IQB; e.g., Richter & Böhme, 2014). Our test comprised several texts and questions focusing on students' skills in forming a broad and general understanding of the texts and retrieving information from them. The students' task was to read the texts and answer questions on their content. Most of the questions had to be answered in a multiple-choice format, but the tests also included a few items with an open-ended response format. Detailed coding instructions were made available by the authors of those items.

We administered one of two test forms to students on the academic track and one of two test forms to all other students resulting in four different test forms. Depending on test form students had to answer 24 to 34 items (24 or 26 items for students on the vocational and comprehensive school tracks and 33 or 34 items for students on the academic track). Using the four test forms we were able to establish an anchor-item test design (Kolen & Brennan, 2004) which allowed us to estimate person parameters on a common metric. Based on IRT models we generated five plausible values (PVs; Wu, 2005) for each student using the software ConQuest 3.0. The reliability of the PVs was high (.93). The PVs were rescaled to have a mean of 100 and a standard deviation of 30 for the complete sample of  $N = 1597$ .

To obtain informed teacher judgments of the students' reading achievement with a high congruence in domain-specificity teachers were provided with sample items from the reading achievement test and the mean number of solved items by students regardless of school track. For each student in their class who took the reading achievement test they were then asked to estimate how many test items were solved.

#### 4.2.3.3 Migrant origin.

Multiple information was used to operationalize students' migrant origin. Students were asked where they, their parents, and their grandparents had been born. Furthermore, they were asked which language they usually speak at home (German vs. another). Parents likewise indicated where their child, they themselves, and their parents (i.e., the child's grandparents) had been born and which language they usually speak at home. A third data source was the tracking list from the school-coordinator (usually a teacher) which included the migrant origins of students and their parents as well as the language spoken at home of each participating student. Based on all available information we built a variable with the four categories (1) migrant origin in Turkey, (2) migrant origin in the former Soviet Union or in Poland, (3) migrant origin in another country, and (4) no migration background. Data from the parent questionnaire were considered the most valid; in cases where they were missing we resorted to data from the student questionnaire. The tracking list was only considered when no information was available from the other two sources. The students were assigned to one of the three categories with migrant origins if the language spoken at home was not only German or the student, at least one parent, or one pair of grandparents were born in a foreign country with assignment to categories 1 or 2 if language or country of birth fitted the focused countries.

#### 4.2.3.4 School grades.

School grades for the subject German were assessed by asking the school coordinator to make a list with grades in German from the last report cards, that is, from the end of the sixth grade. Grades in Germany vary between 1 (grade A) and 6 (grade F).

Only a few students received a grade worse than 5, therefore we recoded 6 into 5. Finally, grades were reverse coded so that higher grades indicate better achievement.

#### 4.2.3.5 Teachers' job experience.

Teachers' job experience was assessed by asking teachers in an open-ended format how many years they had worked as a teacher ( $M = 13.59$ ,  $SD = 11.80$ ). In addition, teachers indicated how many years they had been teaching students in the current sample following a multiple-choice format (1/4 years, 1 1/4 years, and 2 1/4 years). Because only seven of the teachers in our study knew their students for 1 1/4 years we dichotomized the variable in 0 = *less than two years* (51.0%) and 1 = *more than two years* (49.0%).

### 4.2.4 Statistical Analyses

#### 4.2.4.1 Missing data.

Missing data resulted primarily for the teacher questionnaire. The nonresponse rate was 41.1% for teachers' judgments of their students' self-efficacy in reading and 47.8% for judgments of reading achievement. Students had 3.3% missing responses for their self-efficacy in reading. On the cluster level teachers' answers on the background variables were missing from 38.6% (teachers' age and experience with class) to 39.4% (teachers' gender and total job experience).

Restricting the analyses to cases where teachers really judged their students (i.e., applying listwise deletion) possibly would provide biased results with reduced power (Newman, 2014). Thus, the application of modern missing data techniques is recom-

mended (e.g., Schafer & Graham, 2002). Multiple imputation to date is one of the best solutions to deal with missing data and, compared to maximum likelihood methods, offers the advantage to apply an imputation model that is more complex than the data analysis model (Schafer & Graham, 2002). Results of simulation studies support the application of multiple imputation when using such an inclusive strategy even for rates of missing information that are higher than those we had (e.g., Collins, Schafer, & Kam, 2001).

In line with the existing literature (Enders, 2010) we computed multiply imputed data using *Mplus* 7.11 (Muthén & Muthén, 1998-2012). The two-level imputation model incorporated the relevant variables (i.e., all predictors and the criterion variable) as well as interaction terms of the core variables used in later analyses and additional background variables. The imputation procedure was conducted 10 times for each of the 5 PVs. All subsequent analyses, also carried out using *Mplus* 7.11, were thus based on a total of  $m = 50$  imputed data sets. For all inferential statistics we adopted an a priori Type I error rate of 5%.

#### 4.2.4.2 Components of judgment accuracy.

The three components of teachers' judgment accuracy were calculated for both self-efficacy in reading and reading achievement. The level component was obtained by subtracting the cluster mean of the students' scores from the cluster mean of the teacher judgments. If both cluster means were the same the result would be exactly zero, indicating a perfect accuracy of the teacher's judgments regarding the absolute level of students' scores in his or her classroom. Positive values indicate that the teachers overestimate their students' mean level and negative values correspond to an underestimation of the mean level.

To gain insight into the accuracy of the teachers' judgments regarding the variance among students the differentiation component of teachers' judgment accuracy was built as a quotient of the standard deviations of the teachers' judgments and the students' scores. A perfect estimation of the variance among students' scores would lead to a differentiation component of 1, whereas values above 1 would indicate that it is overestimated.

The rank component provided a Pearson correlation of the teachers' judgments and the respective student scores. Thus, a correlation coefficient of 1 indicates that teachers rank their students perfectly on self-efficacy in reading or reading achievement; the more a correlation coefficient deviates from 1 the higher the degree of teachers' misjudgments. A rank component of -1 would indicate that teachers rank their students in exactly the opposite order with, in the case of reading achievement for example, the best student judged as being the worst and vice versa.

Due to the different number of items in the four test forms we took the relative frequency of solved items for students and teachers to calculate the three components of teachers' judgment accuracy of students' reading achievement. Since teachers and students answered similar items for students' self-efficacy in reading, the mean scores of both scales per student were used to calculate the three components of teachers' judgment accuracy of students' self-efficacy in reading. We computed weighted mean scores and weighted standard deviations across the 112 clusters for the level component to account for the different cluster sizes using the *Model Constraint* option in *Mplus* to create the new parameters. To test for equality to the perfect value (i.e., 0 for the level component



and 1 for the differentiation component) we computed Wald tests (Asparouhov & Muthén, 2010) using the *Model Test* option in *Mplus*.

Regarding the rank component we averaged the correlation coefficients obtained by *Mplus* using the Fisher- $z$  transformation weighted by the degrees of freedom as proposed by Hedges and Olkin (1985). We then corrected for bias by the adjustment proposed by Hotelling (1953) and report the coefficient transformed by the inverse Fisher- $z$  transformation as an estimate of the mean correlation.

#### 4.2.4.3 Multilevel analyses.

For analyzing predictors and moderators of teachers' judgments of students' self-efficacy in reading we had to take into account the multilevel structure of our data. Judgments on students' self-efficacy by the same teacher are not independent of each other; they may depend on teachers' rater biases. This is indicated by significant variance between clusters (i.e., teachers). The intra-class correlation (*ICC*) indicates the proportion of variance between clusters to the total variance (i.e., variance between clusters and variance between individuals). If the *ICC* coefficient exceeds zero, an ordinary least squares regression would underestimate the standard errors to this extent resulting in significance tests that are too liberal and confidence intervals that are too small (Cohen, Cohen, West, & Aiken, 2003). In our study, teachers' judgments of students' self-efficacy in reading were dependent on the teachers ( $ICC = .07$ ) making multilevel analyses indispensable. Please note, however, that the *ICC* is quite small and therefore most variance (93%) of teachers' judgments of their students' self-efficacy in reading was situated at the individual student level. An implication of this *ICC* is that there was only little variance between teachers when judging their students' self-efficacy in reading and therefore indi-

vidual teacher characteristics should not be expected to be powerful predictors of these judgments.

We conducted two-level regression models with random intercepts and both random slopes and fixed slopes using *Mplus*. The models were estimated using a sandwich estimator (Robust Maximum Likelihood; MLR). To facilitate interpretation we standardized ( $M = 0$ ,  $SD = 1$ ) the scores of the scales (i.e., self-efficacy ratings of students and teachers' judgments), students' test scores, and the German grade prior to estimating the two-level model. For each categorical variable we generated  $n_{\text{categories}}-1$  dummy variables coded 0 and 1.

## 4.3 Results

### 4.3.1 Descriptive Results

Students' reading achievement scores were  $M = 0.61$  ( $SD = 0.20$ ) in terms of relative frequency of items students solved. Teachers' mean ratings of students' reading achievement corresponded to a relative frequency of solved items of  $M = 0.60$  ( $SD = 0.18$ ). Students' self-efficacy in reading was  $M = 3.08$  ( $SD = 0.61$ ) and teachers judged it with  $M = 2.68$  ( $SD = 0.70$ ). More detailed information on the means and standard deviations of these four central variables broken down by school track and migrant origin is depicted in Table 4.1.

---Table 4.1 about here---

Looking at students' reading achievement it is obvious, however, that students with a migrant origin in Turkey performed worse in our reading achievement test than their peers without a migration background across all school tracks. The effect sizes (Co-

hen's  $d$  adjusted for different variances and group size) within each school track of the differences in reading achievement of each group of student with a migration background compared to the group of students without a migration background were from  $d = 0.42$  to  $d = 1.02$  with the exception ( $d = 0.15$ ) of students with a migrant origin in the Former Soviet Union visiting the comprehensive school track (overall  $M = 0.62$ ,  $SD = 0.28$ ). Still, all student groups showed comparable levels of self-efficacy in reading within their school tracks as effect sizes of the differences in self-efficacy in reading of these groups were between  $d = -0.12$  and  $d = 0.19$  and therefore can be considered as negligible. The same conclusion holds true for gender differences with effect sizes of  $d < |.20|$  (descriptive results broken down by gender are available as supplement).

#### 4.3.2 Components of Judgment Accuracy

To achieve our first objective on how accurately teachers judge their students' self-efficacy in reading we computed the three descriptive components of judgment accuracy. In addition, we report the results on teachers' judgment accuracy of their students' reading achievement to allow for comparisons. In line with our assumptions, the results indicate a significant underestimation of the level of self-efficacy in reading ( $M = -0.40$ ,  $SE = 0.03$ ; Wald's  $\chi^2(1) = 197.57$ ,  $p < .001$ ) and a significant overestimation of the variance (differentiation component) in the classroom ( $M = 1.32$ ,  $SE = 0.08$ ; Wald's  $\chi^2(1) = 17.29$ ,  $p < .001$ ). For teachers' judgment accuracy of their students' reading achievement the level component yields nonsignificant misjudgment of the level of reading achievement ( $M = -0.01$ ,  $SE = 0.01$ ; Wald's  $\chi^2(1) = 0.60$ ,  $p = .44$ ); the differentiation component, however, indicates that teachers underestimate the variance of reading achievement in the classroom ( $M = 0.78$ ,  $SE = 0.05$ ; Wald's  $\chi^2(1) = 18.06$ ,  $p < .001$ ). The rank component of

teachers' judgment accuracy regarding their students' self-efficacy in reading was weak ( $r = .23$ ;  $z = 7.92$ ;  $p < .001$ ; 95% CI [0.17, 0.28]) and modest for reading achievement ( $r = .39$ ,  $z = 13.73$ ,  $p < .001$ , 95% CI [0.34, 0.44]). The confidence intervals of both correlation coefficients did not overlap.

### 4.3.3 Application of Multilevel Models

The following results are based on multilevel analyses with teachers' judgments of students' self-efficacy in reading as the criterion. We standardized the criterion variable and the continuous predictor variables prior to estimating the model so that the results can be interpreted as standardized regression coefficients. To reduce computational effort we fixed the mean of the standardized criterion variable on the between level.

First of all, we estimated a two-level model with random intercept and students' self-reported self-efficacy in reading as the predictor. The model yielded a regression coefficient of  $B = 0.24$  ( $SE = 0.03$ ,  $p < .001$ ) which is similar to the rank component of teachers' judgment accuracy in the previous section. The estimated residual variance of this model was  $\hat{\sigma}_{r_{ij}} = .89$  ( $SE = .04$ ,  $p < .001$ ) indicating that there is much more variance on the individual student level to be explained. A second model that allowed random slopes in addition to random intercepts revealed an identical slope coefficient as compared with the first model, but the slope variance was not statistically significant ( $\hat{\sigma}_{u_{1j}} = .002$ ,  $SE = .01$ ,  $p = .80$ ). We therefore decided to estimate models with fixed slopes for our further analyses.

#### 4.3.4 Predictors of Teachers' Judgments and Interactions

Besides students' self-reported self-efficacy in reading we added three indicators of students' achievement, students' gender, and students' migrant origin (gender and migrant origin were dummy coded; reference = male and German, respectively) to the model. The three indicators of achievement were teachers' judgments of students' reading achievement on our test, and the German grade from the last report card as indicators supposed to be accessible by the teachers, and students' actual reading achievement as an indicator supposed not to be directly accessible by the teachers; all standardized.

To investigate whether there are differences in teachers' judgments of students' self-efficacy in reading depending on students' migrant origin we included interaction terms of the teachers' judgments of students' reading achievement with students' migrant origin on the individual level. On the cluster level we included school track (reference = comprehensive school track) and teachers' job experience with the class as dummy-coded predictors and teachers' total job experience as a continuous predictor. We estimated a two-level regression model with random intercepts, fixed slopes, eight predictors and three interaction terms on the student level, and four predictors on the cluster level. On the individual level the estimated residual variance of this model was reduced to  $\hat{\sigma}_{r_{ij}} = .63$  ( $SE = .04, p < .001$ ). The complete model results are depicted in Table 4.2.

---Table 4.2 about here ---

The German grade and teachers' judgments of their students' reading achievement—both are indicators of students' achievement that are supposed to be accessible by the teachers—significantly predicted teachers' judgments of their students' self-efficacy in reading. Students' self-reported self-efficacy in reading and students' actual reading

achievement added to the prediction of the criterion variable. After controlling for the other predictors in the model female students were judged less efficacious than their male classmates ( $B = -0.14$ ,  $SE = 0.06$ ,  $p = .01$ ). Students in the mixed migration group ( $B = 0.28$ ,  $SE = 0.07$ ,  $p < .001$ ) as well as students with a Turkish migrant origin ( $B = 0.15$ ,  $SE = 0.07$ ,  $p = .05$ ) were judged more efficacious than their classmates without a migration background. That is a male student without a migration background showing intermediate levels of reading achievement and of self-efficacy in reading and holding the same German grade is judged about 0.15 standard deviation units less self-efficacious in reading compared to the same student with a migrant origin in Turkey.

Furthermore, our analyses yielded a negative interaction effect of migrant origin in Turkey  $\times$  teachers' rating of students' reading achievement on students' self-efficacy in reading ( $B = -0.19$ ,  $SE = 0.07$ ,  $p = .01$ ), indicating that teachers' judgments of students' self-efficacy in reading for students with a Turkish migrant origin seems to change more gently with changing reading achievement than for students without a migration background. This means, if a teacher's judgments of the reading achievement of both students mentioned above change by one standard deviation unit the teacher's judgments of the students' self-efficacy in reading change by 0.56 standard deviation units for the student without a migration background and by  $0.56 - 0.19 = 0.37$  standard deviation units for the student with a migrant origin in Turkey.

Looking at the cluster level we only found a significant main effect of academic track schools where intermediate students are judged significantly less self-efficacious in reading than students at the comprehensive school track ( $B = -0.18$ ,  $SE = 0.07$ ,  $p = .01$ ), whereas the regression coefficient of the vocational school track was nonsignificant.

Teachers' total job experience and experience with their class also showed no significant main effects.

## **4.4 Discussion**

Our study was the first to investigate teachers' judgment accuracy of secondary school students' self-efficacy in reading as well predictors of those judgments on the student and the classroom level. To this end we computed the three descriptive components to get insights into how accurate teachers can judge their students' self-efficacy in reading as well as additional multilevel analyses with special regard to students' migrant origins.

### **4.4.1 Level and Discrimination Components**

Results for the level and discrimination components showed that teachers underestimate the level and overestimate the variance of students' self-efficacy in reading within their classrooms. These results met our assumptions and are consistent with findings on teachers' judgment accuracy of other motivational student characteristics (Huelmann et al., 2014; Praetorius et al., 2011; Spinath, 2005). In contrast to judgments of student achievement in terms of number of items solved the rating scales for assessing self-efficacy in reading and the corresponding teachers' judgments have no meaningful zero point and therefore are arbitrary. In addition, answering rating scales allows students to choose their answers with higher ratings than it would reflect their actual level of self-efficacy (faking good; e.g., Davier, 2010). Faking good is not possible for standardized achievement tests where better outcomes cannot be chosen by intention but just by chance. The arbitrariness of our self-efficacy scale should make it more difficult for

teachers to accurately judge the mean level of self-efficacy than judging the number of solved items on an achievement test with a given anchor of mean items correct for all students in a population and a meaningful point of zero reading achievement. The additional possibility of faking good on the students' side could lead to higher levels of self-evaluations with possible shrinkage of variance. On the teachers' side this could therefore lead to an underestimation of the mean level and an overestimation of the variance in a classroom. Therefore, our results may at least partly arise due to the reliance on self-reports but in the context of teachers' judgment accuracy of students' self-efficacy these assessments seem indispensable to us.

From a statistical point of view teachers underestimated their students' self-efficacy in reading by about two third of a standard deviation which is statistically meaningful. From a practical point of view this difference represents less than half of a point on an arbitrary scale. The practical relevance of teachers' underestimation of their students' self-efficacy in reading thus seems questionable to us.

#### **4.4.2 Rank Component**

Regarding the rank component teachers judged their students' reading achievement with a degree of accuracy for reading achievement in secondary schools ( $r = .39$ ) that corresponds well to previous findings in Germany ( $r = .34$ ; Karing et al., 2011). Teachers' judgments of students' self-efficacy in reading were weak but of comparable accuracy ( $r = .23$ ) to teachers judging the academic self-efficacy of their students in secondary schools ( $r = .21$ ; Huelmann et al., 2014). This result on the accuracy of teachers' judgments of their students self-efficacy in reading supports our hypothesis and is consistent with results of other motivational constructs (e.g., academic self-efficacy: Huel-



mann et al., 2014; learning goal orientation: Dicke, Lüdtke, Trautwein, Nagy, & Nagy, 2012; learning motivation: Spinath, 2005) and of self-concepts of secondary school students (Praetorius, Berner, Zeinz, Scheunpflug, & Dresel, 2013). Alternatively, findings regarding primary school students' self-concepts (Praetorius et al., 2010; Praetorius et al., 2011; Spinath, 2005) and reading self-efficacy (Coddington & Guthrie, 2009) reported moderate rank components.

There are two possible explanations for the weak correlation. First of all the weak correlation of teachers' judgments of their students' self-efficacy in reading might be the fact that student-reported self-efficacy in reading showed comparatively less variance ( $M = 3.08$ ,  $SD = 0.61$ ) than the reading achievement scores ( $M = 0.60$ ,  $SD = 0.20$ ). Thus, for teachers it should be more difficult to establish the rank order of their students' self-efficacy in reading which results in low correlations for judged self-efficacy in reading compared to the judged reading achievement. Likewise, the student-reported self-efficacy in reading in the study of Coddington and Guthrie (2009;  $M = 2.88$ ,  $SD = 0.76$ ) varied more strongly than that in our sample making it easier to judge the rank order, too.

Secondly, it may be explained at least partly by the RAM (Funder, 1995) because weak judgments may occur if the judge has a modicum of cues (e.g., by a lack of visibility) or uses irrelevant cues. This claim may particularly hold true for reading in secondary schools as teaching to read and fostering reading-related self-beliefs should be a major task during instruction on primary schools and therefore both could be less salient to teachers in secondary schools. In addition, there may be a lack of knowledge about self-efficacy possibly causing teachers to refer to students' achievement as a well accessible student characteristic when judging students' self-efficacy for the first time, particu-

larly in reading in secondary schools. However, students' ability to read in terms of relative frequency of items solved was not on a high level and varied considerably. Thus, teachers should be aware of their students' reading ability and reading-related constructs, even in secondary schools.

#### **4.4.3 Student Migration Background**

Teachers estimate the self-efficacy in reading of students with migrant origin in Turkey or the mixed migration group as higher compared to students without a migration background after controlling for all other predictor variables. Note that this effect did not occur for students with a migrant origin in the Former Soviet Union or Poland. In part this may be because these students are mainly resettlers or children of resettlers and therefore ethnical Germans with similar names, appearance, and maybe also attitudes, behavior, and overt forms of expressions. The positive regression coefficient for students with a migrant origin in Turkey or from the mixed migration group may at first indicate an overestimation of students' self-efficacy in reading. In our descriptive results students with a migrant origin showed weaker reading achievement with Cohens'  $d$  around 0.60 in the three school tracks, but there were no or only small effects of  $d < |0.20|$  regarding differences in the students' self-efficacy in reading. Thinking of the RAM, these levels of self-evaluations might be expressed by the students in a way that teachers are able to consider them for their judgments. Our findings suggest that teachers perceive these possible expressions and take this information into account when judging their students' self-efficacy in reading. This should be supported by further research where experimental settings are preferable using for example the Simulated Classroom setting (see Kaiser et al., 2013, for an example).

Moreover, the relationship between teachers' judgments of their students' reading achievement and teachers' judgments of their students' self-efficacy in reading is moderated by students' migrant origin in Turkey. If students with a migrant origin in Turkey were judged at the top of the scale and with a small variance this result could reflect a problem with the scale. But obviously our descriptive results do not support such an interpretation. Instead, compared to the other groups of students with and without migration backgrounds, teachers' judgments of students' self-efficacy in reading of students with a migrant origin in Turkey seem not as closely tied to teachers' judgments of students' reading achievement. As students' reading achievement should be a reasonable and obvious cue to judge students' self-efficacy in reading we prefer to assert that this result supports the idea that students with a migrant origin in Turkey compared to students without or with a different migration background display more tangible cues on which teachers can base their judgments of these students' self-efficacy in reading besides reading achievement.

#### **4.4.4 Student Gender**

An interesting auxiliary finding is that, controlling for the other predictors, teachers judged girls as less self-efficacious in reading than boys. Our descriptive results, however, showed no considerable differences favoring girls or boys, neither regarding their reading achievement nor their self-efficacy in reading. Therefore, the lower coefficient for girls' compared to boys' self-efficacy in reading after controls may result from boys and girls providing different behavioral cues to their teachers for judging self-efficacy in reading. It may also reflect gender stereotypes of teachers regarding students'

self-efficacy in reading leading to a possible gender bias in their judgments. Whether this bias really exists should be a topic in future research.

#### 4.4.5 Strengths and Limitations

Our analyses comprised the classical components of judgment accuracy as well as a sophisticated multilevel approach to further investigate teachers' judgments of their students' self-efficacy in reading. Besides the still open questions mentioned above, the present data limited the investigation to moderation on the individual level using interaction terms. Addressing possible moderators on the cluster level (i.e., teachers' individual characteristics) presumably requires a bigger and, regarding gender, a more diversified teacher sample. In our study we applied multilevel multiple imputation procedures to reduce possibly biased estimation due to teacher nonresponse. But although multiple imputation reflects the uncertainty in the estimates and is superior to listwise deletion of cases with missing data (e.g., Newman, 2014), it seems more favorable to motivate teachers to participate in such studies to increase statistical power and reduce uncertainty of the estimates. But this claim conflicts with the recommended assessment of the teachers' judgments. We used informed judgments with a high congruence in domain-specificity as they are considered the most appropriate for studying judgment accuracy (Südkamp et al., 2012). To establish this, the teachers participating in our study had to answer items to judge their students self-efficacy in reading that were similarly worded to the items for students' self-reported self-efficacy in reading. This resulted in excessive burden for participating teachers as they had to answer six items for each student to judge their students' self-efficacy. With an average cluster size of 14 students per class, teachers typically had to answer 84 items on students' self-efficacy in reading. To overcome this bur-

den it may be worth a try to adapt the assessment of self-efficacy to our assessment of reading achievement: showing an example item and giving an expected mean of the student responses. This possibly would lead to increased rates of teacher response.

Nevertheless, our study has certain strengths besides being the first to analyze teachers' judgment accuracy of their students' self-efficacy in reading on a large sample using multilevel analyses. The ancillary results on judgment accuracy regarding students' reading achievement allow for comparisons on how accurately these teachers judge different student characteristics. In our study, on average teachers were more accurate in judging their students' achievement ( $r = .39$ ) than their students' self-efficacy in reading ( $r = .23$ ).

The stratified sample containing a large number of students with a migration background stemming from the largest immigrant groups in Germany is an additional major strength of our study. This sample allowed us to regard students from these groups separately, thus differentiating effects of particular migrant origins on teachers' judgments of students' self-efficacy in reading. As self-efficacy in reading is comparably high regardless of students' migration background the validity of our findings on teachers' judgment accuracy should not be limited. With regard to students' reading achievement, however, such a claim might not be warranted. Teachers tend to overestimate students with low reading achievement (Bates & Nettelbeck, 2001; Begeny et al., 2008; Feinberg & Shapiro, 2009) and our sampling procedure presumably resulted in an overrepresentation of low achieving students.

#### 4.4.6 Implications for Theory and Practice

The topic is of particular interest for politicians, practitioners, and scientists as students with a migration background are lagging behind their peers without a migration background regarding achievement (Gebhardt et al., 2013; Stanat & Christensen, 2006) and students' self-efficacy in reading may be useful for teachers to promote students' reading achievement (Schunk, 2003). Our findings suggest that teachers heavily rely on students' reading achievement when judging their students self-efficacy in reading. As reading achievement and self-efficacy in reading are correlated (Bong et. al., 2012; Piercey, 2013) and students' achievement can be judged quite accurate by teachers (Südkamp et al., 2012) bridging the gap by relying on judgments of achievement seems to be easy for teachers. However, as the weak rank component shows, this approach tends to be insufficient. There is a lot more to self-efficacy in reading than students' reading achievement, with differences especially when considering individual students. Our results suggest that even on secondary schools teachers should put an effort to promote students' reading achievement and accurate knowledge on students' self-efficacy in reading can serve to accomplish this claim.

As discussed above, secondary school teachers may be baffled when asked to judge students' self-efficacy in reading and this could trigger vested interest leading to more accurate judgments later on. Longitudinal studies should be conducted to provide insights into that topic with a focus on teachers' vested interest in judging students' self-efficacy beliefs as a possible moderator of the development of their judgment accuracy. This could be extended by investigating the development of students' domain-specific self-efficacy and achievement.

Teacher interventions consisting of building advanced training courses to foster students' self-efficacy with evaluations according to these programs seem to be another promising field for researchers and practitioners. However, our results should motivate teachers to also engage in existing advanced training courses on topics of student' motivation, especially on domain-specific self-efficacy. The low rank component and the possible gender stereotypes reflected in teachers' judgments indicate that the instructional quality might benefit if teachers acquired more knowledge about self-efficacy.

#### **4.5 Conclusion**

In sum, our results correspond to previous findings and add to the field of research on teachers' judgments of students' motivation and achievement. Our findings support the importance of valid cues to judge at least students' self-efficacy in reading accurately.

The rank component was weak but according to results on teachers' judgment accuracy of students' self-beliefs on secondary schools and highlights the importance of teachers' knowledge about students' motivation. However, looking at individual student characteristics teachers' judgments may not be as weak as our results suggest because the different relations between self-efficacy and achievement of students with specific migrant origins may already be accounted for. Nevertheless, because teachers also base some of their decisions important for individual students' educational pathways on students' characteristics related to their self-efficacy our results strengthen the position that at least their knowledge about self-efficacy beliefs should be improved, guided by more fruitful research regarding this topic.

## 4.6 References

- Artelt, C., & Rausch, T. (2014). Accuracy of teacher judgments. In S. Glock & M. Böhmer (Eds.), *Teachers' professional development. Assessment, training, and learning* (pp. 27–43). Rotterdam, The Netherlands: SensePublishers. doi: 10.1007/978-94-6209-536-6\_3
- Asparouhov, T. & Muthén, B. (27.07.2010). *Chi-square statistics with multiple imputation*. Retrieved from [www.statmodel.com/download/MI7.pdf](http://www.statmodel.com/download/MI7.pdf) (13.07.2015)
- Autorengruppe Bildungsberichterstattung (Eds.). (2014). *Bildung in Deutschland 2014: Ein indikatorengestützter Bericht mit einer Analyse zur Bildung von Menschen mit Behinderungen* [Education in Germany 2014: An indicator-based report including an analysis of the situation of people with special educational needs and disabilities]. Bielefeld, Germany: Bertelsmann. Retrieved from: <http://www.bildungsbericht.de/zeigen.html?seite=11123>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- Bates, C., & Nettelbeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, 21(2), 177–187. doi:10.1080/01443410020043878
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43–55. doi:10.1037/1045-3830.23.1.43



- Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *The Journal of Educational Research, 105*(5), 336-352. doi:10.1080/00220671.2011.627401
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. doi:10.1023/A:1021302408382
- Coddington, C. S., & Guthrie, J. T. (2009). Teacher and students perceptions of boys' and girls' reading motivation. *Reading Psychology, 30*(3), 225-249. doi:10.1080/02702710802275371
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*, 330–351. doi:10.1037/1082-989X.6.4.330
- Davies, M. v. (2010). Why sum scores may not tell us all about test takers. *Newborn and Infant Nursing Reviews, 10*(1), 27–36. doi:10.1053/j.nainr.2009.12.011
- Demaray, M. K., & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly, 13*(1), 8–24. doi:10.1037/h0088969
- Dicke, A.-L., Lüdtke, O., Trautwein, U., Nagy, G., & Nagy, N. (2012). Judging students' achievement goal orientations: Are teacher ratings accurate? *Learning and Individual Differences, 22*(6), 844–849. doi:10.1016/j.lindif.2012.04.004

- Eccles, J. S., Wigfield, A., Flanagan, C. A., Miller, C., Reuman, D. A., & Yee, D. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality*, *57*(2), 283–310. doi:10.1111/j.1467-6494.1989.tb00484.x
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York, NY: Guilford Press.
- Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, *102*(6), 453–462. doi:10.3200/JOER.102.6.453-462
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652–670. doi:10.1037/0033-295X.102.4.652
- Galla, B. M., Wood, J. J., Tsukayama, E., Har, K., Chiu, A. W., & Langer, D. A. (2014). A longitudinal multilevel model analysis of the within-person and between-person effect of effortful engagement and academic self-efficacy on academic performance. *Journal of School Psychology*, *52*(3), 295–308. doi:10.1016/j.jsp.2014.04.001
- Gebhardt, M., Rauch, D., Mang, J., Sälzer, C., & Stanat, P. (2013). Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund [Competences in mathematics among students with a migration background]. In M. Prenzel, C. Sälzer, E. Klieme, & O. Köller (Eds.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (pp. 275–308). Münster, Germany: Waxmann.
- Givvin, K. B., Stipek, D. J., Salmon, J. M., & MacGyvers, V. L. (2001). In the eyes of the beholder: students' and teachers' judgments of students' motivation. *Teaching and Teacher Education*, *17*(3), 321–331. doi:10.1016/S0742-051X(00)00060-3

- Glock, S., & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Social Psychology of Education, 16*(1), 111–127. doi:10.1007/s11218-012-9197-z
- Gräsel, C., Krolak-Schwerdt, S., Nölle, I., & Hörstermann, T. (2010). Diagnostische Kompetenz von Grundschullehrkräften bei der Erstellung der Übergangsempfehlung: Eine Analyse aus der Perspektive der sozialen Urteilsbildung [Diagnostic competence of primary school teachers for the preparation of secondary school recommendations – a social judgment perspective]. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. 56. Beiheft zur Zeitschrift für Pädagogik* (pp. 286–295). Weinheim, Germany: Beltz.
- Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., . . . Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology, 96*(3), 403–423. doi:10.1037/0022-0663.96.3.403
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften [Comparative tests as an instrument for the improvement of teachers' diagnostic competence]. In R. Arnold & C. Griesse (Eds.), *Schulmanagement und Schulentwicklung* (pp. 119–144). Hohengehren, Germany: Schneider.

- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313.  
doi:10.3102/00346543059003297
- Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)*, 15(2), 193–232.  
doi:10.2307/2983768
- Huelmann, T., Ferdinand, H. D., Gebauer, M. M., McElvany, N., Bos, W., Köller, O., & Schöber, C. (2014). Urteilsgenauigkeit von Lehrkräften bezüglich der Selbstwirksamkeitsüberzeugungen von Schülerinnen und Schülern [Teachers' judgment accuracy regarding students' self-efficacy beliefs]. In H.-G. Holtappels, M. Pfeifer, A. S. Willems, W. Bos, & N. McElvany (Eds.), *Jahrbuch der Schulentwicklung. Band 18. Daten, Beispiele und Perspektiven* (1st ed., pp. 298–320). Weinheim, Germany: Beltz Juventa.
- Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher-Student Support, Effortful Engagement, and Achievement: A 3-Year Longitudinal Study. *Journal of Educational Psychology*, 100(1), 1–14. doi:10.1037/0022-0663.100.1.1
- Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, 28, 73–84. doi:10.1016/j.learninstruc.2013.06.001
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen [Diagnostic competence of elementary and secondary school teachers in the domains of competence and interests]. *Zeit-*

*schrift für Pädagogische Psychologie*, 23(3), 197–209. doi:10.1024/1010-0652.23.34.197

- Karing, C., Matthäi, J., & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? [Lower secondary school teacher judgment accuracy of students' reading competence – A matter of specificity?]. *Zeitschrift für Pädagogische Psychologie*, 25(3), 159–172. doi:10.1024/1010-0652/a000041
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). *Statistics for Social Science and Public Policy*. New York, NY: Springer. doi:10.1007/978-1-4757-4310-4
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., & Neubrand, M., . . . Weiß, M. (Eds.). (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente* [PISA 2000: Documentation of the data collection instruments]. Berlin, Germany: Max-Planck-Institut für Bildungsforschung.
- Lehmann, R. H., Gänsfuß, R., & Peek, R. (1999). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen — Klassenstufe 7. Bericht über die Untersuchung im September 1998* [Aspects of students' initial level and development at schools in Hamburg — Grade 7. Report on the study in September 1998]. Hamburg, Germany: Landesinstitut für Lehrerbildung und Schulentwicklung.
- Lehmann, R. H. & Peek, R. (1997). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen — Klassenstufe 5. Bericht über die Erhebung im September 1996* [Aspects of students' initial level and devel-

opment at schools in Hamburg — Grade 5. Report on the study in September 1996].

Hamburg, Germany: Landesinstitut für Lehrerbildung und Schulentwicklung.

Lehmann, R. H., Peek, R., Gänsfuß, R., & Husfeldt, V. (2002). *Aspekte der Lernausgangslage und der Lernentwicklung — Klassenstufe 9. Ergebnisse einer längsschnittlichen Untersuchung in Hamburg* [Aspects of students' initial level and development — Grade 9. Results of a longitudinal study in Hamburg]. Hamburg, Germany: Behörde für Bildung und Sport, Amt für Schule.

Madelaine, A., & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education*, 52(1), 33–42.

doi:10.1080/10349120500071886

Marsh, H. W., & Craven, R. G. (1991). Self-other agreement on multiple dimensions of pre-adolescent self-concept: Inferences by teachers, mothers, and fathers. *Journal of Educational Psychology*, 83(3), 393–404. doi:10.1037/0022-0663.83.3.393

McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., . . . Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern [Teachers' diagnostic skills to judge student performance and task difficulty when learning materials include instructional pictures]. *Zeitschrift für Pädagogische Psychologie*, 23(3), 223–235. doi:10.1024/1010-0652.23.34.223

Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus User's guide: Statistical analysis with latent variables*. (7th ed.). Los Angeles, CA: Muthén & Muthén.

- Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods, 17*, 372-411. doi: 10.1177/10944281145485909
- Pajares, F., & Schunk, D. H. (2003). Self-efficacy and self-concept beliefs: Jointly contributing to the quality of human life. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self research. International advances in self research: speaking to the future* (pp. 95–121). Greenwich, CT: Information Age Publishing.
- Piercey, R. R. (2013). *Reading self-efficacy in early adolescence: Which measure works best?* (Doctoral dissertation, University of Kentucky). Retrieved from [http://uknowledge.uky.edu/edp\\_etds/10/](http://uknowledge.uky.edu/edp_etds/10/)
- Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging self-concepts of students. *The Journal of Educational Research, 106*(1), 64–76.  
doi:10.1080/00220671.2012.667010
- Praetorius, A.-K., Greb, K., Lipowsky, F., & Gollwitzer, M. (2010). Lehrkräfte als Diagnostiker: Welche Rolle spielt die Schülerleistung bei der Einschätzung von mathematischen Selbstkonzepten? [Teachers' diagnostic perception – What role does students' grade performance play in the teacher's assessment of mathematical self-concepts?] *Journal for Educational Research Online, 2*(1), 121–144.
- Praetorius, A.-K., Karst, K., Dickhäuser, O., & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften [How teachers rate their students: On teachers' diagnostic compe-

tence regarding the academic self-concept]. *Psychologie in Erziehung und Unterricht*, 58, 81–91.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). *Advanced quantitative techniques in the social sciences: Vol. 1*. Thousand Oaks, CA: Sage.

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, 48(2), 335–360.

doi:10.3102/0002831210374874

Richter, D., & Böhme, K. (2014). Vergleichsarbeiten im Fokus: Welche Funktionen erfüllt der Test aus Sicht von Lehrkräften? [Focus on comparative tests: Which features do these tests fulfill from teachers' perspectives?]. *Schulmanagement*, 45(2), 12–14.

Rjosk, C., McElvany, N., Anders, Y., & Becker, M. (2011). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung der basalen Lesefähigkeit ihrer Schülerinnen und Schüler [Teachers' diagnostic skills in estimating students' basic reading capacities]. *Psychologie in Erziehung und Unterricht*, 58, 92–105.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. doi: 10.1037/1082-989X.7.2.147

Schleicher, A. (2006). Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003 1. *Intercultural Education*, 17(5), 507–516.

doi:10.1080/14675980601063900



- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen [Diagnostic competency of teachers: Components and effects]. *Empirische Pädagogik*, *1*(1), 27–52.
- Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading and Writing Quarterly*, *19*, 159–172.  
doi:10.1080/10573560308219
- Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Educational Psychology Handbook Series. Handbook of motivation at school* (pp. 35–54). New York, NY: Routledge.
- Smith, J. K., Smith, L. F., Gilmore, A., & Jameson, M. (2012). Students' self-perception of reading ability, enjoyment of reading and reading achievement. *Learning and Individual Differences*, *22*(2), 202–206. doi:10.1016/j.lindif.2011.04.010
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teacher judgments on student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie*, *19*(1/2), 85–95.
- Stanat, P., & Christensen, G. (2006). *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003*. Paris, France: OECD.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. doi:10.1037/a0027627

- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111–133. doi:10.1207/s15326985ep3902\_3
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*(3), 420–432. doi:10.1037/0022-0663.89.3.420
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31*(2-3), 114–128. doi:10.1016/j.stueduc.2005.05.005
- Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202–231). Cambridge, UK: Cambridge University

Table 4.1

Means and standard deviations (in parentheses) of students' reading achievement and self-efficacy in reading and the corresponding teachers' judgments, broken down by school track and migrant origin

School track	Migrant origin	<i>n</i>	Student scores				Teacher judgments			
			Reading achievement		Self-efficacy in reading		Reading achievement		Self-efficacy in reading	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
VT	GER	106	0.54	0.16	2.94	0.65	0.52	0.16	2.52	0.78
	TUR	143	0.41	0.15	3.01	0.57	0.46	0.17	2.55	0.68
	FSU	49	0.43	0.18	2.82	0.72	0.52	0.18	2.21	0.77
	MIX	134	0.45	0.17	2.99	0.66	0.48	0.16	2.62	0.70
CT	GER	139	0.62	0.18	3.04	0.66	0.64	0.16	2.76	0.68
	TUR	168	0.47	0.15	2.97	0.58	0.54	0.17	2.54	0.67
	FSU	68	0.60	0.16	2.97	0.61	0.64	0.16	2.67	0.61
	MIX	79	0.54	0.18	2.91	0.67	0.55	0.17	2.75	0.65
AT	GER	393	0.78	0.13	3.26	0.53	0.71	0.15	2.86	0.72
	TUR	92	0.65	0.15	3.18	0.57	0.62	0.15	2.72	0.67
	FSU	113	0.72	0.15	3.16	0.60	0.67	0.16	2.62	0.63
	MIX	89	0.72	0.16	3.16	0.62	0.66	0.14	2.84	0.65

*Note.* VT: Vocational school track; CT: Comprehensive school track; AT: Academic school track; GER: No migration background; TUR: Migrant origin in Turkey; FSU: Migrant origin in the Former Soviet Union or Poland; MIX: Group with mixed migrant origins.

Table 4.2

Results of a two-level regression model on teachers' judgments of their students' self-efficacy in reading

	<i>B</i>	<i>SE</i>
Intercept ( $\gamma_{00}$ )	0.00	0.00
<i>Predictors on student level</i>		
Teachers' judgment of students' reading achievement <sup>a</sup> ( $\gamma_{10}$ )	0.56***	0.05
German grade, reverse coded <sup>a</sup> ( $\gamma_{20}$ )	0.14***	0.03
Students' gender, female = 1 ( $\gamma_{30}$ )	-0.14*	0.06
Migrant origin in Turkey <sup>b</sup> ( $\gamma_{40}$ )	0.15*	0.07
Migrant origin in the Former Soviet Union or Poland <sup>b</sup> ( $\gamma_{50}$ )	-0.10	0.07
Migrant origin in another country <sup>b</sup> ( $\gamma_{60}$ )	0.28***	0.07
Interaction $\gamma_{10} \times \gamma_{40}$ ( $\gamma_{70}$ )	-0.19**	0.07
Interaction $\gamma_{10} \times \gamma_{50}$ ( $\gamma_{80}$ )	-0.04	0.07
Interaction $\gamma_{10} \times \gamma_{60}$ ( $\gamma_{90}$ )	-0.12	0.07
Self-efficacy in reading <sup>a</sup> ( $\gamma_{100}$ )	0.10***	0.03
Reading achievement <sup>a</sup> ( $\gamma_{110}$ )	0.12**	0.05
<i>Predictors on cluster level</i>		
Vocational track school <sup>c</sup> ( $\gamma_{01}$ )	0.17	0.09
Academic track school <sup>c</sup> ( $\gamma_{02}$ )	-0.18*	0.07
Teachers' job experience in years <sup>a</sup> ( $\gamma_{03}$ )	0.00	0.03
Teachers' experience with classroom, 1 = more than 2 years ( $\gamma_{04}$ )	-0.02	0.06

*Note.* The predictors  $\gamma_{10}$  to  $\gamma_{60}$  are expected to be accessible by the teachers, whereas the interactions and the predictors  $\gamma_{100}$  and  $\gamma_{110}$  are not;

<sup>a</sup>Measure is standardized ( $M = 0$ ;  $SD = 1$ ).

<sup>b</sup>Compared to students without migration background.

<sup>c</sup>Compared to the comprehensive school track.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

## Supplementary Table

Means and standard deviations (in parentheses) of students' reading achievement and self-efficacy in reading and the corresponding teachers' judgments, broken down by school track and student gender

School track	Student gender	<i>n</i>	Student scores				Teacher judgments			
			Reading achievement		Self-efficacy in reading		Reading achievement		Self-efficacy in reading	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
VT	Male	238	0.46	0.17	2.92	0.63	0.49	0.17	2.54	0.75
	Female	194	0.45	0.17	3.02	0.65	0.48	0.17	2.51	0.70
CT	Male	225	0.54	0.18	2.97	0.62	0.58	0.16	2.63	0.65
	Female	229	0.55	0.18	2.99	0.63	0.60	0.18	2.69	0.69
AT	Male	331	0.75	0.14	3.21	0.58	0.68	0.15	2.84	0.69
	Female	356	0.74	0.15	3.23	0.54	0.68	0.15	2.75	0.69

*Note.* VT: Vocational school track; CT: Comprehensive school track; AT: Academic school track.



# **5**

## **Gesamtdiskussion**

## 5.1 Zusammenfassung

The influence of people's self-beliefs on their achievement does not end with their schooling. Consequently, the aim of education must transcend the development of academic competence. Schools have the added responsibility of preparing self-assured and fully functioning individuals capable of pursuing their hopes and their ambitions. (Pajares & Schunk, 2001, S. 258)

Das hier aufgeführte Zitat verdeutlicht einerseits die zentrale Funktion selbstbezogener Fähigkeitskognitionen für die Leistungsfähigkeit auch über die Schulzeit hinaus und andererseits die Aufgabe der Schule, sich in ihrem Einfluss auf die Schülerinnen und Schüler nicht nur auf die Vermittlung von Wissen zu beschränken. In Theorie wie Empirie wird aktuell ein positiver reziproker Effekt zwischen selbstbezogenen Fähigkeitskognitionen und Leistungen angenommen (Bandura, 1993; Marsh & Craven, 2006). Insofern können Lehrkräfte die selbstbezogenen Fähigkeitskognitionen der Schulkinder mittelbar durch die Verbesserung von Leistungen fördern. Gleichzeitig sollte sich dadurch ein weiterer positiver Effekt auf die Leistungsentwicklung einstellen. Ob diese angenommenen positiven Zusammenhänge jedoch in ethnisch gemischten Gruppen gleichermaßen gültig sind, wurde bisher nicht empirisch untersucht. Zugleich wurde nur wenig Forschung zu den Zusammenhängen im sprachlichen Bereich betrieben, der durch die unterschiedliche Erstsprache bei Kindern und Jugendlichen mit und ohne Migrationshintergrund besonders interessant für die Untersuchung kulturell differenzieller Zusammenhänge erscheint.

Neben der indirekten Förderung der Leistung können Lehrkräfte auch direkt auf die Entwicklung der selbstbezogenen Fähigkeitskognitionen Einfluss nehmen, insbeson-



dere auf deren Selbstwirksamkeitsüberzeugungen (Bandura, 1997; Möller & Köller, 2004; Pajares, 1996; vgl. Abschnitt 1.2.4). Für eine zielführende Förderung von Leistungen und Selbstwirksamkeitsüberzeugungen ist es allerdings unerlässlich, dass Lehrkräfte beides bei ihren Schülerinnen und Schülern möglichst akkurat einschätzen können (Anders, Kunter, Brunner, Krauss & Baumert, 2010; Huelmann et al., 2014; Schrader, 2009). Obwohl die Güte von Lehrkrafturteilen in Bezug auf Schulkindleistungen bereits häufig untersucht wurde, fand kaum Forschung zur Urteilsgenauigkeit in Bezug auf Selbstwirksamkeitsüberzeugungen von Schulkindern statt. Hinzu kommt, dass ein Einfluss auf die Urteilsgenauigkeit durch persönliche Schulkindmerkmale angenommen wird. Dennoch wurde der Migrationshintergrund von Schulkindern bisher überwiegend in experimentellen Studien betrachtet (Glock & Krolak-Schwerdt, 2013; Kaiser, Südkamp & Möller, 2016) und noch nicht um entsprechende Feldstudien ergänzt, was die ökologische Validität der Erkenntnisse einschränkt (Schrader, 2009).

In der vorliegenden Arbeit wurden diese Fragen aufgegriffen und mit zwei empirischen Studien der Versuch unternommen, reziproke Effekte zwischen Leistungen und selbstbezogenen Fähigkeitskognitionen nachzuweisen, wobei auf den Bereich des Lesens und die Berücksichtigung des Migrationshintergrunds der Schulkinder fokussiert wurde. Eine dritte Studie diente der empirischen Klärung, wie genau Lehrkräfte die lesespezifischen Selbstwirksamkeitsüberzeugungen ihrer Schulkinder beurteilen können und welche Rolle der Migrationshintergrund der Schulkinder für die Lehrkrafturteile spielt.

Die Ergebnisse dieser drei Studien werden im Folgenden kurz zusammengefasst, bevor eine Gesamtbetrachtung der Ergebnisse erfolgt. Im Anschluss werden diese

Ergebnisse hinsichtlich ihrer Limitationen und Implikationen für Forschung und Praxis betrachtet.

### 5.1.1 Studie 1

Zunächst wurde im Rahmen der ersten Studie die kausale Ordnung zwischen dem verbalen Fähigkeitsselbstkonzept und sprachlichen Leistungen im Lesen und Rechtschreiben über den Verlauf der ersten beiden Jahre in der Sekundarstufe I untersucht. Erstmals wurde dabei der Migrationshintergrund der Schülerinnen und Schülern (mit bzw. ohne Migrationshintergrund) berücksichtigt.

Um die Frage der kausalen Ordnung zwischen verbalem Fähigkeitsselbstkonzept und Lese- beziehungsweise Rechtschreibleistungen zu beantworten wurde den methodischen Vorgaben von Marsh, Byrne & Yeung (1999) gefolgt und Strukturgleichungsmodelle berechnet. Eine mögliche Moderation der Zusammenhänge durch den Migrationshintergrund der Schulkinder oder durch die besuchte Schulform (Gymnasium vs. Nicht-Gymnasium) wurde durch die Modellpassungsmaße invarianter Mess- und Strukturmodelle evaluiert.

Insgesamt bestätigen die Ergebnisse bestehende Befunde aus großen Bildungsvergleichsstudien, die einen deutlichen Leistungsrückstand von Kindern und Jugendlichen mit Migrationshintergrund im Vergleich zu ihren Mitschülerinnen und Mitschülern ohne Migrationshintergrund aufzeigten (Schwippert, Wendt & Tarelli, 2012; Stanat, Rauch & Segeritz, 2010) auch für Jugendliche in den ersten beiden Jahren nach dem Übertritt in die Sekundarstufe und unter Kontrolle der Schulform. Erwartungsgemäß wiesen die Jugendlichen mit Migrationshintergrund ein Niveau des verbalen Selbstkonzeptes

auf, das kaum geringer war als bei ihren Mitschülerinnen und Mitschülern ohne Migrationshintergrund (Roebbers, Mecheril & Schneider, 1998; Shajek, Lüdtke & Stanat, 2006).

Die Entwicklung der sprachlichen Leistungen sowie des verbalen Selbstkonzeptes verlief zwischen den Gruppen mit und ohne Migrationshintergrund vergleichbar. Das ist einerseits eine gute Nachricht, da Kinder mit geringeren Ausgangsleistungen keine weiteren Leistungsnachteile im Vergleich zu ihren leistungsstärkeren Mitschülerinnen und Mitschülern hinnehmen mussten. Damit wurden mit dieser Arbeit Befunde bestätigt, die gegen das Vorliegen eines sogenannten *Matthäuseffekts* sprechen, demzufolge Schulkinder mit stärkeren Ausgangsleistungen auch stärkere Leistungszugewinne erzielen sollten (Baumert, Nagy & Lehmann, 2012).

Bezüglich der zentralen Frage der kausalen Ordnung zeigten sich für Jugendliche mit und ohne Migrationshintergrund wechselseitige Effekte des verbalen Selbstkonzeptes und der sprachlichen Leistungen. Sowohl für die Leseleistungen als auch für die Rechtschreibleistungen waren diese Effekte eher gering. Allerdings wurde in Studie 1 ein Messintervall von zwei Jahren verwendet, mit dem ersten Messzeitpunkt kurz nach dem Übergang von der Grundschule auf die Sekundarstufe. Während damit zum ersten Messzeitpunkt das verbale Fähigkeitsselbstkonzept noch maßgeblich durch die Erfahrungen aus der Grundschulzeit beeinflusst sein sollte, können zum zweiten Messzeitpunkt bereits Anpassungen aufgrund der neuen sozialen Bezugsgruppe und dem unterschiedlichen Fächerkanon angenommen werden (vgl. Abschnitt 1.2.3). In einer ähnlichen Stichprobe fanden sich bei Retelsdorf, Köller & Möller (2014) nur reziproke Effekte zwischen Leseleistungen und lesespezifischem Fähigkeitsselbstkonzept nach dem Übergang auf die Sekundarstufe und Effekte im Sinne der Skill-Development-Hypothese im weiteren Ver-

lauf. Hierbei wurden jedoch kürzere Messintervalle etabliert und mit der Erfassung des lesespezifischen Fähigkeitsselbstkonzeptes eine bessere Passung mit entsprechend stärkeren Zusammenhängen gewählt (vgl. Abschnitt 1.2.2). Insofern sind die in Studie 1 aufgedeckten reziproken Effekte über Schulformen und den Migrationshintergrund hinweg durchaus bemerkenswert.

Die Robustheit der Effekte über den Migrationshintergrund der Schulkinder und die besuchte Schulform ergänzt Befunde früherer Studien, die reziproke Effekte in Hongkong im Vergleich zu englischsprachigen westlichen Staaten (Marsh, Hau & Kong, 2002) sowie bei Schulkindern aus unterschiedlichen Schulsystemen (Marsh & Köller, 2004) nachweisen konnten. Damit leistet diese Studie einen Beitrag zur Füllung der von Valentine, Dubois & Cooper (2004) identifizierten Forschungslücke der kulturellen Übertragbarkeit des Zusammenspiels zwischen Leistungen und selbstbezogenen Fähigkeitskognitionen.

Ausgehend von diesen Befunden wurde in Studie 2 erstmals versucht der kausalen Ordnung zwischen Leseleistungen und lesespezifischen Selbstwirksamkeitsüberzeugungen nachzugehen, sowie die bisher dünne Befundlage zur kausalen Ordnung zwischen Mathematikleistungen und mathematikspezifischen Selbstwirksamkeitsüberzeugungen zu stärken.

### **5.1.2 Studie 2**

Hauptziel der zweiten Studie war, die in Studie 1 gefundenen reziproken Effekte zwischen sprachlichen Leistungen und dem verbalen Fähigkeitsselbstkonzept für Leseleistungen und lesespezifische Selbstwirksamkeitsüberzeugungen innerhalb des siebten Schuljahres nachzuweisen und eine mögliche Moderation durch den Migrationshinter-

grund der Jugendlichen zu prüfen. Zugleich sollte das Vorliegen reziproker Effekte zwischen Mathematikleistungen und mathematikspezifischen Selbstwirksamkeitsüberzeugungen überprüft werden, um so Vergleiche zwischen den beiden Fachbereichen anhand derselben Stichprobe zu erlauben.

Methodisch wurde analog zur Studie 1 den Vorgaben von Marsh et al. (1999) gefolgt und wiederum Strukturgleichungsmodelle berechnet. Aufgrund der sehr unterschiedlichen Bildungsbeteiligung von Jugendlichen mit und ohne Migrationshintergrund und der damit einhergehenden Konfundierung von Migrationshintergrund und Schulform, wurde zusätzlich zu den beiden Zweigruppenmodellen (Migrationshintergrund vs. kein Migrationshintergrund bzw. Gymnasium vs. Nicht-Gymnasium) die Passung eines aus Migrationshintergrund und Schulform kombinierten Viergruppenmodells überprüft.

Die Ergebnisse wiesen in der Gesamtgruppe auf eine Unterstützung der Self-Enhancement-Hypothese im Bereich Mathematik hin; höhere mathematikspezifische Selbstwirksamkeitsüberzeugungen wirkten sich innerhalb des siebten Schuljahres positiv auf die zeitlich nachgeordnete Mathematikleistung aus. Für den Bereich Lesen zeigte sich hingegen eine Unterstützung der Skill-Development-Hypothese, so dass eine hohe Leseleistung zu Beginn des Schuljahres die Entwicklung der lesespezifischen Selbstwirksamkeitsüberzeugungen förderte. Die Modellpassungsmaße ergaben keine Hinweise auf eine mögliche Moderation dieser Beziehungen durch den Migrationshintergrund oder die Schulform. Die Inspektion der Pfadkoeffizienten des Zweigruppenmodells bei Aufteilung der Schülerinnen und Schüler in Gruppen mit und ohne Migrationshintergrund ergab jedoch Evidenz für das Vorliegen reziproker Effekte zwischen der Leseleistung und lese-

spezifischen Selbstwirksamkeitsüberzeugungen innerhalb des siebten Schuljahres in beiden Gruppen.

Mit dieser Studie wurde erstmals das Vorliegen eines reziproken Zusammenhangs zwischen Selbstwirksamkeitsüberzeugungen und Leistungen im Bereich Lesen nachgewiesen, wenn auch erst nach Aufteilung der Schülerinnen und Schüler nach dem Migrationshintergrund. Ohne diese Aufteilung fand sich Unterstützung der Skill-Development-Hypothese in der siebten Schulklasse. Dieser Befund schließt an die Ergebnisse von Retelsdorf et al. (2014) an, die für das Zusammenspiel zwischen Leseleistung und dem lesespezifischen Fähigkeitsselbstkonzept in der siebten Klasse ohne Berücksichtigung des Migrationshintergrunds ebenfalls nur Effekte in diese Richtung nachweisen konnten. Die Modellpassungsmaße aus der zweiten Studie bestätigten ergänzend zur Studie 1 die kulturelle Übertragbarkeit des Zusammenspiels zwischen Selbstwirksamkeitsüberzeugungen und Schulleistungen im Lesen und in Mathematik.

### **5.1.3 Studie 3**

In der dritten Studie wurde aufgegriffen, dass reziproke Effekte zwischen lesespezifischen Selbstwirksamkeitsüberzeugungen und Leseleistungen in Gruppen mit und ohne Migrationshintergrund innerhalb eines Schuljahres nachweisbar waren. Durch diesen Befund empfiehlt sich eine Förderung von beidem im Unterricht (Pajares & Schunk, 2001; Schunk & Pajares, 2009). Für die optimale Förderung von Selbstwirksamkeitsüberzeugungen und Leistungen sollte jedoch beides von den Lehrkräften möglichst akkurat beurteilt werden können (Huelmann et al., 2014; Schrader, 2009). Entsprechend wurde in Studie 3 überprüft, wie akkurat Lehrkräfte die Leseleistungen und die lesespezifischen Selbstwirksamkeitsüberzeugungen ihrer Schülerinnen und Schüler zu Beginn der

siebten Klasse einschätzen können. Zudem wurde explorativ untersucht, ob der spezifische Migrationshintergrund der Schulkinder eine Rolle bei der Urteilsbildung in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen spielt.

Die Urteilsgenauigkeit der Lehrkräfte wurde dabei über zwei unterschiedliche Herangehensweisen bewertet. Zunächst wurden die klassischen Komponenten der Urteilsgenauigkeit (Niveauelemente, Differenzierungskomponente und Rangkomponente; siehe Schrader & Helmke, 1987) berechnet. Damit sollte ein Vergleich sowohl der lesespezifischen Selbstwirksamkeitsüberzeugungen als auch der Leseleistung mit bestehenden Befunden zur Urteilsgenauigkeit von Lehrkräften gelingen. Im zweiten Schritt wurde ein mehrebenenanalytisches Verfahren zur Regression des Lehrkrafturteils auf die lesespezifischen Selbstwirksamkeitsüberzeugungen der Schülerinnen und Schüler gewählt. Durch dieses Vorgehen konnte einerseits der mehrebenenanalytisch berechnete Regressionskoeffizient des Lehrkrafturteils mit der klassischen Rangkomponente verglichen werden. Andererseits ermöglichte das mehrebenenanalytische Verfahren in einer erweiterten Analyse simultan verschiedene Prädiktoren auf der Ebene der einzelnen Schulkinder, z. B. deren Leseleistung, Geschlecht und Migrationshintergrund, wie auch auf der Klassenebene, z. B. die Schulform und die Berufserfahrung der Lehrkräfte, in das Modell aufzunehmen. Dadurch konnten die Auswirkungen der Prädiktoren auf die Lehrkrafturteile evaluiert werden. Außerdem war es möglich, eine Regression des Urteils der lesespezifischen Selbstwirksamkeitsüberzeugungen auf eine Interaktion des Migrationshintergrunds der Schulkinder mit dem Leseleistungsurteil der Lehrkräfte auf das Urteil zu überprüfen.

Insgesamt konnte nur eine unzureichende Urteilsgenauigkeit der Lehrkräfte in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen ihrer Schulkinder festgestellt werden, die im Wesentlichen den Befunden zur Rangkomponente der Urteilsgenauigkeit in Bezug auf die fachunspezifischen Selbstwirksamkeitsüberzeugungen bei Huelmann et al. (2014) entsprach. Das mehrebenenanalytische Analogon zur Rangkomponente bestätigte dieses Ergebnis erwartungsgemäß. Zudem unterschätzten die Lehrkräfte im Mittel die Schulkindselbsteinschätzungen der lesespezifischen Selbstwirksamkeitsüberzeugungen und überschätzten deren Streuung. Auch die Unterschätzung des Niveaus steht im Einklang mit der Literatur zur Urteilsgenauigkeit in Bezug auf selbstbezogene Fähigkeitskognitionen (Praetorius, Karst, Dickhäuser & Lipowsky, 2011; Spinath, 2005). Hingegen zeigten sich vorliegende Befunde in Bezug auf die Differenzierungskomponente uneinheitlich. Einige Ergebnisse belegten eine Überschätzung (Spinath, 2005), andere eine genaue Einschätzung der Streuung (Huelmann et al., 2014; Praetorius et al., 2011).

Die Betrachtung der Prädiktoren auf Schulkindebene ergab unter anderem, dass Schulkinder mit einem türkischen Migrationshintergrund signifikant höher in den lesespezifischen Selbstwirksamkeitsüberzeugungen eingeschätzt wurden, als ein vergleichbares Schulkind ohne Migrationshintergrund. Zudem zeigte sich ein negativer Interaktionsterm zwischen dem türkischen Migrationshintergrund und den Lehrkrafturteilen der Schulkindleseleistung. Dieser lässt sich wie folgt interpretieren: Von zwei Schulkindern, eines mit türkischem Migrationshintergrund und eines ohne Migrationshintergrund, die in der Leseleistung um eine Standardabweichung besser eingeschätzt werden als ein durchschnittlich eingeschätztes Schulkind, wäre bei dem Schulkind mit türkischem Migrationshintergrund der Anstieg in der Lehrkräfteeinschätzung der lesespezifischen Selbst-



wirksamkeitsüberzeugung geringer als bei dem Schulkind ohne Migrationshintergrund. Bei einer um eine Standardabweichung geringer eingeschätzten Leseleistung würde das Lehrkrafturteil der lesespezifischen Selbstwirksamkeitsüberzeugung für ein Schulkind mit türkischem Migrationshintergrund entsprechend geringer abfallen. Insgesamt wäre damit das Lehrkrafturteil in Bezug auf lesespezifische Selbstwirksamkeitsüberzeugungen von Schulkindern mit einem türkischen Migrationshintergrund im Vergleich zu den anderen Schulkindern relativ konstant und auf hohem Niveau. Dieser Befund kann als Hinweis darauf gedeutet werden, dass Lehrkräfte das hohe Niveau der Selbstwirksamkeitsüberzeugungen bei Jugendlichen mit Migrationshintergrund bei dennoch geringeren Leistungen als deren Mitschülerinnen und Mitschüler ohne Migrationshintergrund erkennen.

## 5.2 Diskussion

Die Ergebnisse der drei Studien sind sowohl theoretisch für die empirische Unterstützung der in der Theorie formulierten reziproken Effekte (Bandura, 1977, 1978) zwischen Selbstwirksamkeitsüberzeugungen und Leistungen sowie praktisch für Erkenntnisse in die Urteilsgenauigkeit von Lehrkräften interessant. Mit Blick auf die aufgeworfenen Forschungsfragen kann zusammenfassend konstatiert werden, dass die kausale Ordnung zwischen Schulleistungen und selbstbezogenen Fähigkeitskognitionen kulturübergreifend vergleichbar betrachtet werden kann. Einschränkend ist zu erwähnen, dass die gefundenen Effekte stets sehr gering waren, wenn auch angesichts der hohen Stabilitäten und der bereits über mehrere Schuljahre fortgeschrittenen Entwicklung durchaus bemerkenswert. Dennoch stellen selbstbezogene Fähigkeitskognitionen keine Panazee zur Beseitigung von Leistungsdisparitäten dar.

In Bezug auf die Urteilsgenauigkeit der Lehrkräfte bleibt festzuhalten, dass die in dieser Arbeit untersuchten Lehrkräfte bezüglich der Leistungsurteile übliche Ergebnisse erreicht haben. Die Urteile in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen fielen hingegen äußerst schwach aus. Besonders ist der Befund hervorzuheben, dass Jugendlichen mit türkischem Migrationshintergrund offensichtlich hinsichtlich des Niveaus ihrer lesespezifischen Selbstwirksamkeitsüberzeugungen entsprechend auf höherem Niveau eingeschätzt werden, als ihre Mitschülerinnen und Mitschüler mit vergleichbaren Leseleistungen. Die bei Jugendlichen mit türkischem Migrationshintergrund möglicherweise verzerrte Kalibrierung der lesespezifischen Selbstwirksamkeitsüberzeugungen mit den Leseleistungen scheint daher von den Lehrkräften bei der Urteilsbildung berücksichtigt zu werden.

Nachfolgend werden zunächst die Einschränkungen der in dieser Arbeit berichteten empirischen Studien zusammenfassend diskutiert, um jeweils einen Forschungsausblick zu präsentieren. Im Anschluss folgen Ausführungen zu den praktischen Implikationen der vorgestellten Arbeiten.

### **5.2.1 Kausale Ordnung**

Die Ergebnisse im sprachlichen Bereich waren sich sowohl über die Bereiche (Lesen und Rechtschreiben) als auch über die selbstbezogenen Fähigkeitskognitionen (verbales Fähigkeitsselbstkonzept und lesespezifische Selbstwirksamkeitsüberzeugungen) hinweg sehr ähnlich. Leider verhindern gleich mehrere Unterschiede in der Operationalisierung eine direkte Vergleichbarkeit der gewonnenen Erkenntnisse. Zum einen wurden in den ersten beiden Studien die jeweiligen selbstbezogenen Kognitionen auf unterschiedlichen Ebenen der Spezifität gemessen. Obwohl die Erfassung des verbalen Fähigkeits-

selbstkonzeptes mit der Struktur des Marsh/Shavelson-Modells (Marsh & Shavelson, 1985) übereinstimmt, wäre für einen Vergleich mit den lesespezifischen Selbstwirksamkeitsüberzeugungen die Erfassung auf einer entsprechenden lesespezifischen Ebene notwendig, wie sie etwa von Retelsdorf et al. (2014) vorgenommen wurde.

Letztendlich ist jedoch noch unklar, inwieweit auch die Selbstwirksamkeitsüberzeugungen in einzelnen Schulfächern generalisiert werden (Schunk & Pajares, 2009). Lesekompetenzen sind bereits ein Aggregat aus unterschiedlichen Kompetenzen der Leserin beziehungsweise des Lesers, wie beispielsweise dem Wortschatz und dem Vorwissen (Artelt et al. 2007). Hinzu kommen persönliche Merkmale des Lesers oder der Leserin, wie der adaptive Einsatz von Lesestrategien oder die Selbstregulation (Artelt et al. 2007). Auch lesespezifische Selbstwirksamkeitsüberzeugungen sollten somit bereits über unterschiedliche Teilkompetenzen aggregierte Fähigkeitsbeschreibungen darstellen. Im gleichen Zuge werden mathematikspezifische Selbstwirksamkeitsüberzeugungen als hinreichend spezifisch angenommen, obwohl auch hier unterschiedliche Teilleistungsbereiche wie Geometrie und Arithmetik durch spezifische Selbstwirksamkeitsüberzeugungen repräsentiert sein sollten. Im Bereich der Arithmetik könnten wiederum noch spezifischere Erfassungen beispielsweise der Selbstwirksamkeitsüberzeugungen für Divisionsaufgaben und noch spezifischer in Selbstwirksamkeitsüberzeugungen für Multiplikation und Subtraktion vorgenommen werden. Der optimale Grad an Spezifität erscheint somit völlig unklar und die offene Frage, inwieweit selbstbezogene Fähigkeitskognitionen aggregierte feinkörnige Operationalisierungen darstellen oder eben nicht, erschwert die Interpretation vorliegender Befunde.

Weitere Einschränkung der Vergleichbarkeit bestehen in den zwar ähnlichen aber nicht identischen Leseleistungstests, den unterschiedlichen Stichproben und den unterschiedlichen Messintervallen. Insofern kann zwar auf Basis der in den ersten beiden Studien gewonnenen Erkenntnisse die Schlussfolgerung gezogen werden, dass reziproke Effekte im Bereich des Lesens innerhalb der Sekundarstufe I nachweisbar sind. Ein Vorteil für eine der beiden selbstbezogenen Kognitionen lässt sich jedoch aufgrund der Operationalisierung nicht ausmachen. Ein solcher Unterschied könnte beispielsweise in der unterschiedlichen Stabilität der Konstrukte angenommen werden (vgl. Abschnitt 1.2.5). So könnten durch die fragileren Selbstwirksamkeitsüberzeugungen kurzfristig wirksame Interventionen initiiert werden und eine Vielzahl unterschiedlicher Quellen adressiert werden (vgl. Abschnitt 1.2.4). Veränderungen des Selbstkonzeptes erscheinen hingegen schwieriger, sind jedoch beispielsweise bei ungünstigen sozialen Bezugsgruppen durch eine Betonung der individuellen Leistungsentwicklung beim Feedback möglich (Fischbach, Brunner, Krauss & Baumert, 2015; Lüdtke & Köller, 2002; Lüdtke, Köller, Marsh & Trautwein, 2005). Aufgrund der skizzierten Unterschiedlichkeit der in dieser Dissertation durchgeführten Studien können hier jedoch keine neuen Empfehlungen an Lehrkräfte ausgesprochen werden, eine der beiden selbstbezogenen Kognitionen als besonders förderungswürdig herauszustellen, sondern es sollten die hier belegten Vorzüge von beiden genutzt werden.

Mit der Frage der Generalisierung der Konstrukte ist auch die Frage der empirischen Trennbarkeit verknüpft. Die im Rahmen dieser Dissertation durchgeführten empirischen Arbeiten bieten keine weiteren Hinweise darauf, inwieweit die beiden Konstrukte empirisch unterscheidbar sind. In bisherigen Studien fand die Untersuchung der empiri-

schen Trennbarkeit und der Vergleich, welche Variablen sich besser durch Fähigkeits-selbstkonzepte oder Selbstwirksamkeitsüberzeugungen vorhersagen lassen, auf Basis von unterschiedlich spezifischen Operationalisierungen der Konstrukte statt (Ferla, Valcke & Cai, 2009; Jansen, Scherer & Schroeders, 2015; Parker, Marsh, Ciarrochi, Marshall & Abduljabbar, 2014). In der vorliegenden Dissertation wurden leicht adaptierte Versionen von in der Forschung etablierten Items zur Erfassung der Selbstwirksamkeitsüberzeugungen verwendet. Diese verwendeten abstrakte Barrieren anstelle konkreter Aufgaben. Die abstrakten Barrieren dürften jedoch eine gewisse Unsicherheit bezüglich der konkreten Anforderungen hervorrufen. Wenn die Aufgabenanforderungen unbekannt sind, sollten zur Einschätzung der eigenen Erfolgszuversicht vermehrt soziale Vergleiche genutzt werden (Bandura, 1997; Bouffard-Bouchard, 1990; vgl. Abschnitt 1.2.3). Insofern könnte die empirische Trennbarkeit der Konstrukte eingeschränkt sein. Da die hier verwendeten Stichproben aus unterschiedlichen Studien stammten und in beiden Stichproben nur jeweils eines der Konstrukte auf fach- beziehungsweise domänenspezifischer Ebene erfasst wurde, konnte die empirische Trennbarkeit der hier verwendeten Operationalisierungen nicht untersucht werden und bleibt ebenso wie die Beantwortung der Frage nach der Generalisierung eine Aufgabe zukünftiger Forschung.

Um die Vergleichbarkeit von Ergebnissen zu verbessern soll an dieser Stelle die Empfehlung ausgesprochen werden, eine neue Skala zur Erfassung von Selbstwirksamkeitsüberzeugungen und Fähigkeits-selbstkonzepten zu entwickeln. Das in Studie 1 erfasste verbale Fähigkeits-selbstkonzept war nicht auf die Leseleistungen abgestimmt. Insofern liegt hier eher eine Unterschätzung der möglichen Effekte vor. Bei den Selbstwirksamkeitsüberzeugungen wurde auf die in der pädagogischen-psychologischen Forschung üb-

liche domänenspezifische Spezifitätsebene zurückgegriffen. Dies gilt umso mehr aufgrund der längsschnittlichen Erfassung. Um die kausale Ordnung selbstbezogener Fähigkeitenkognitionen und Leistungen bestimmen zu können ist die Nutzung von Längsschnittdaten erforderlich (Marsh et al., 1999). Um dabei Veränderungen abschätzen zu können, ist es ratsam, dass die verwendeten Items möglichst unverändert zu allen Messzeitpunkten vorgegeben werden (Ployhart & Vandenberg, 2010; Vandenberg & Lance, 2000). Die mehrfache Verwendung identischer Items erscheint jedoch gerade im Kontext von Schule durch die sich schnell verändernde Leistung ungeeignet bei Erfassung aufgabenspezifischer Selbstwirksamkeitsüberzeugungen über einen längeren Zeitraum. Abhilfe könnte hier die Entwicklung einer neuen Skala bieten, mit der Selbstwirksamkeitsüberzeugungen in unterschiedlichen Schulfächern mit an die jeweilige Altersgruppe angepassten aufgabenbezogenen Items und der Möglichkeit zur Verankerung von Items erhoben werden. Dabei könnten sogar international anwendbare Skalen konstruiert werden. Dass querschnittliche aufgabenspezifische Erfassung von Selbstwirksamkeitsüberzeugungen zumindest für Mathematik und Naturwissenschaften international möglich ist, zeigen die in den PISA-Studien verwendeten Items (Jansen et al., 2015; Parker et al., 2014; Williams & Williams, 2010). Dieser Erfolg sollte Motivation sein, den Ansatz zu erweitern und auch im Längsschnitt anwendbare aufgabenspezifische Items zu konstruieren. Eine aufgabenspezifische und längsschnittlich anwendbare Operationalisierung von Selbstwirksamkeitsüberzeugungen könnte nicht nur bei Lösung der bereits aufgeworfenen Frage der Generalisierung unterstützen. Zusätzlich böte sich die Möglichkeit auch die noch unbewältigte Herausforderung der Kalibrierung von Selbstwirksamkeitsüberzeugungen und Leistungen (Pajares, 1996; Schunk & Pajares, 2009) mit neuen Erkenntnissen empi-

risch zu unterfüttern, da nur bei aufgabenspezifischer Erfassung von Selbstwirksamkeitsüberzeugungen eine Einschätzung zur Kalibrierung sinnvoll erscheint (Schunk & Pajares, 2009).

Eine angemessene längsschnittliche Operationalisierung auf Aufgabenebene erfordert aufgrund der nötigen Veränderungen der Items die Anwendung der *Item-Response-Theorie* (IRT; Embretson & Reise, 2000; Hambleton & Swaminathan, 1985; van der Linden & Hambleton, 1997). Diese ermöglicht das Abtragen von Itemparametern und Personparametern auf einer gemeinsamen Skala. Damit ist es möglich, die Personenmerkmale, wie beispielsweise die Ausprägung der Selbstwirksamkeitsüberzeugungen, unabhängig von den konkret beantworteten Items einzuschätzen. Damit können Schulkinder in jedem Schuljahr auf das Curriculum abgestimmte aufgabenspezifische Selbstwirksamkeitsitems beantworten. Gleichzeitig können die Items in zeitlich nachgelagerten Befragungen so gestaltet sein, dass diese auf die aktuellen Anforderungssituationen in der Schule abgestimmt sind. Gleichzeitig wird durch die Verwendung der gemeinsamen Skala und ein Anker-Item-Design die längsschnittliche Entwicklung trotz veränderter Itemformulierung messbar.

Das sogenannte Anker-Item-Design ermöglicht es, dass Items, die im Verlauf längsschnittlicher Untersuchungen eine zu hohe Zustimmungswahrscheinlichkeit erhalten (also zu „leicht“ sind), durch Items mit einer geringeren Zustimmungswahrscheinlichkeit (also „schwereren“) ersetzt werden (Kolen & Brennan, 2004). Durch diese Austauschmöglichkeit kann Deckeneffekten vorgebeugt werden. Um die Interpretation jedoch nicht zu gefährden, ist es notwendig, dass der Test den strengen Anforderungen des *Rasch-Modells* genügt (Kubinger & Draxler, 2007). Ein gewichtiger Nachteil besteht in der

aufwändigen Skalenkonstruktion. Es werden initial sehr viele Items benötigt, von denen sich viele zum Teil mangels Passung an die strengen Modellvorgaben als ungeeignet erweisen werden. Für die Kalibrierung der Items ist zudem eine sehr große Anzahl an Personen notwendig (Maier, 2014). Die Gestaltung des Fragebogens könnte dann angelehnt an das Prinzip der revidierten Fassung des Kognitiven Fähigkeitstests für 4. bis 12. Klassen (KFT 4–12+R; Heller & Perleth, 2000) erfolgen, indem je nach eingesetzter Klassenstufe einige leichtere Items, die für vorhergehende Klassenstufen vorgesehen sind, zu Beginn des Fragebogens eingesetzt werden und mit zunehmend schwierigeren Items, die also eine geringere Zustimmungswahrscheinlichkeit haben, fortgefahren wird. Für die nachfolgenden Klassenstufen bilden diese schwierigeren Items wiederum den leichteren Einstieg und werden durch nachfolgende schwerere Items ergänzt und so weiter.

Mit einer solchen Skala könnte die Herausforderung der Kalibrierung angegangen werden. Um mögliche Differenzen im Zusammenspiel zwischen selbstbezogenen Fähigkeitskognitionen und Schulleistungen über die Lebensspanne zu evaluieren bieten sich Mehrkohorten-Längsschnittstudien an (siehe z. B. Marsh & Craven, 2006, S. 152). Diese sollten zudem mehr als zwei Messzeitpunkte umfassen, um Trends abbilden zu können. Dabei bietet sich zudem die Anwendung von Wachstumskurvenmodellen an (Bollen & Curran, 2006; Urban, 2004). Im Gegensatz zu den in den ersten beiden Studien dieser Arbeit verwendeten autoregressiven Pfadmodellen, welche die Rangstabilität unter Kontrolle der initialen Merkmalsausprägungen abbilden, bieten Wachstumskurvenmodelle auch eine Evaluation der absoluten Veränderung von Selbstwirksamkeitsüberzeugungen (Hornstra, van der Veen, Peetsma & Volman, 2013; Kigel, McElvany & Becker,



2015). Auf einer aufgabenspezifischen Ebene können dadurch Hinweise zur Generalisierung und Kalibrierung von Selbstwirksamkeitsüberzeugungen erhalten werden.

Durch die Verwendung von mehr als zwei Messzeitpunkten wird es ermöglicht einen Einblick in die Prozesse hinter den in dieser Dissertation untersuchten Zusammenhängen zu gewinnen. Obwohl in zwei Teilstudien dieser Dissertation Evidenz für die kulturelle Übertragbarkeit des Zusammenspiels zwischen selbstbezogenen Fähigkeitskognitionen und Schulleistungen gefunden wurde, bedeuten diese Befunde lediglich, dass die Pfadkoeffizienten in den Strukturgleichungsmodellen nicht zwischen Schulkindgruppen mit und ohne Migrationshintergrund variieren, zumindest nicht in Deutschland. Die positiven Effekte von Leistungen und selbstbezogenen Kognitionen auf die Entwicklung zukünftiger Leistungen und selbstbezogener Kognitionen könnte jedoch auf unterschiedlichen Prozessen basieren. Diese konnten jedoch in dieser Arbeit nicht modelliert werden, hätte dieser Umweg im Längsschnitt im Idealfall noch mindestens einen dritten Messzeitpunkt erfordert, um nicht nur die Rangstabilität sondern auch absolute Veränderungen der Prozessvariablen verfolgen zu können (McElvany, Kortenbruck & Becker, 2008). Insofern bleibt weiter unklar, ob unter Umständen in einer der beiden Gruppen die höhere Ausprägung selbstbezogener Kognitionen eher die Auswahl herausfordernder Aufgaben auslöste, oder die Anstrengung oder die Ausdauer bei der Aufgabenbearbeitung erhöhte, oder vielleicht auch die aktive Teilnahme am Unterricht oder die Bearbeitung von Hausaufgaben förderte und damit die Entwicklung der Leistungen positiv beeinflusste.

Zwischen kulturell unterschiedlichen Staaten wurde die Übertragbarkeit von Bezugsrahmenmodellen gezeigt (Marsh & Hau, 2003; Marsh & Hau, 2004). In Studie 1

wurden nun invariante Zusammenhänge im Sinne des REM zwischen Jugendlichen mit und ohne Migrationshintergrund innerhalb eines Staates gefunden. In der Forschung zu Fähigkeitsselbstkonzepten wurde bereits das REM mit dem Bezugsrahmenmodell (I/E-Modell) zum RI/E-Modell vereint (Möller, Retelsdorf, Köller & Marsh, 2011). Auch vor dem Hintergrund des Vergleichs von Selbstwirksamkeitsüberzeugungen und Fähigkeitsselbstkonzepten erscheint eine Übertragung des RI/E-Modells auf Selbstwirksamkeitsüberzeugungen interessant. Da Selbstwirksamkeitsüberzeugungen weniger stark auf soziale Vergleiche rekurren als Fähigkeitsselbstkonzepte und dimensionale Vergleiche als nahezu wirkungslos angenommen werden (vgl. Abschnitt 1.2.3), sollten sich keine Kontrasteffekte zwischen einzelnen Fächern beziehungsweise Domänen im Sinne des I/E-Modells zeigen und damit die reziproken Effekte unbeeinflusst von den Selbstwirksamkeitsüberzeugungen und Leistungen in den anderen erfassten Bereichen sein. Die Überprüfung dieser theoretisch relevanten Annahme bleibt jedoch eine Aufgabe zukünftiger Arbeiten in diesem Bereich.

Das RI/E-Modell könnte gerade im sprachlichen Bereich unter Berücksichtigung des Migrationshintergrundes beziehungsweise der zu Hause gesprochenen Sprache mit der Untersuchung einer gemeinsamen Fremdsprache, also z. B. des lesespezifischen Fähigkeitsselbstkonzepts in Englisch, zusätzlich zum Konstrukt in der Unterrichtssprache, überprüft werden. In den berichteten Studien wurden jeweils selbstbezogene Fähigkeitskognitionen in der Unterrichtssprache erfasst, unabhängig davon, ob dies der zu Hause gesprochenen Sprache entspricht, oder nicht. Da jedoch dabei im Ergebnis keine Hinweise gegen die kulturelle Übertragbarkeit der Pfadkoeffizienten zu finden waren, sollte sich auch die kulturelle Übertragbarkeit der Pfadkoeffizienten bei der Untersu-

chung von rein fremdsprachlichen selbstbezogenen Fähigkeitskognitionen ergeben. Allerdings bietet die Erfassung eines gemeinsamen (fremdsprachlichen) Konstruktes den Vorteil, dass beide Konstrukte gleichermaßen fremdsprachlich sind und deshalb, bei Untersuchung der Prozesse zwischen den Pfaden wie der Mediation zwischen Fähigkeitskognition und Leistung über die mit der Aufgabendurchführung verbundene Ausdauer im Falle von Schwierigkeiten, unter diesen Voraussetzungen validere Schlussfolgerungen gezogen werden könnten, als wenn unterschiedliche Variablen, also erstsprachliche und zweitsprachliche Leistungen und Fähigkeitskognitionen, erfasst und miteinander verglichen werden. Dies gilt selbst dann, wenn die Pfadkoeffizienten kulturell übertragbar sein mögen.

### **5.2.2 Urteilsgenauigkeit**

Eine Schwäche der dritten Studie ist, dass die Klassenkomposition keine Berücksichtigung finden konnte. Experimentelle Befunde zeigten, dass Schulkinder als Mitglieder einer Minorität zumindest in Bezug auf ihre Mathematikleistungen akkurater eingeschätzt wurden, als Schulkinder der Majorität (Kaiser et al., 2016). Auch bezüglich der Einschätzung anderer Schulkindmerkmale als der Mathematikleistung könnte die Genauigkeit des Urteils durch die Zusammensetzung der Klasse moderiert sein. Insofern könnte es sinnvoll sein, dass die Zusammensetzung der Klasse berücksichtigt wird. Die ist in Feldstudien leider nur eingeschränkt möglich, da oftmals nicht die komplette Klasse an der Studie teilnimmt. So sind auch in der vorliegenden Arbeit, insbesondere durch fehlende Elterngenehmigungen, einige Schulkinder nicht in der Stichprobe enthalten gewesen. Diese gehören jedoch zur kompletten Klassenzusammensetzung und sollten damit auch einen Einfluss auf die Urteilsgenauigkeit von Lehrkräften haben können. Eine er-

gänzende Untersuchung, wie sich nicht zu bewertende Schulkinder auf die Urteilsgenauigkeit von Lehrkräften auswirken, könnte eine Idee für zukünftige experimentelle Forschung darstellen.

Experimentelle Ansätze unter Verwendung des Simulierten Klassenraums (Kaiser et al., 2016) zeigten, dass die Zugehörigkeit zur Minorität zu akkurateren Einschätzungen der Leistungen durch Lehrkräfte führt. In Verknüpfung mit dem Prozessmodell von Funder (1995), welches den Weg zu einem möglichst akkuraten Urteil über vier Stufen beschreibt, könnte dieser Befund durch die verringerte Wahrnehmung relevanter Hinweisreize durch die Lehrkraft und durch unterschiedlich verfügbare Hinweisreize zur Leistungseinschätzung durch die Schulkinder zustande kommen. Neben der unterschiedlichen Salienz wird die Einschätzung von Schulkindern mit Migrationshintergrund auch dadurch schwieriger, dass diese schlechtere Leistungen erbringen (siehe Abschnitt 1.3.1), Lehrkräfte Schulkinder mit stärkeren Leistungen aber genauer einschätzen können (Begeny, Eckert, Montarello & Storie, 2008; Feinberg & Shapiro, 2009). Da die schlechte Leistung und der ethnische Hintergrund somit konfundiert sind, ist es für Feldstudien schwierig, hier eine Differenzierung vorzunehmen. Hier könnten Arbeiten unter Anwendung von *Propensity-Score-Matching* (Rubin & Thomas, 2000) für weitere Klärung sorgen.

Die Ergebnisse von Kaiser et al. (2016) können auch als Hinweis darauf gedeutet werden, dass die Salienz, verbunden mit einer besseren Wahrnehmbarkeit, ein entscheidender Faktor für die Genauigkeit von Lehrkrafturteilen in Bezug auf die Leistungsstärke von Schulkindern ist. Die Wahrnehmbarkeit von Merkmalen kann abseits der vorab diskutierten technischen und persönlichen Faktoren einen hilfreichen Ansatz bieten,

um Unterschiede in der Urteilsgenauigkeit zu erklären. Im Theorieteil dieser Arbeit wurde mit dem RAM (Funder, 1995; vgl. Abschnitt 1.4.3) ein Prozessmodell vorgestellt, das verschiedene Stufen annimmt, an denen es zu einem Scheitern des Urteils kommen kann. Bisher ist noch unklar, inwieweit die Beurteilung der Leistungsstärke von Schulkindern durch Unterbrechungen in diesem Urteilsprozess ungenauer wird. Allerdings scheint sich zu bestätigen, dass Schulkindmerkmale, die weniger salient als Leistungen sind, zunehmend ungenau beurteilt werden können. Zur Klärung der lehrkraftseitigen Prozesse könnten möglicherweise Videostudien beitragen. So könnten beispielsweise zunächst die Lehrkräfte im Anschluss an eine Unterrichtseinheit befragt werden, wie viele Leistungshinweise sie bei bestimmten Schulkindern wahrgenommen hätten (Stufe 3 im RAM). Anschließend können Dritte anhand des Videomaterials eine komparable Einschätzung abgeben. Ein Vergleich dieser Antworten könnte einen Hinweis darauf geben, wie abgelenkt eine Lehrkraft durch den Unterricht von der Wahrnehmung der urteilsrelevanten Leistungshinweise der Schulkinder ist. Ob ein solcher Ansatz in der Folge zur weiteren Klärung der schwachen Urteilsgenauigkeit bei leistungsrelevanten Merkmalen wie selbstbezogenen Fähigkeitskognitionen fruchtbar wird, bleibt abzuwarten.

Eine weitere Schwäche der Studie 3 ist die zum Teil sehr geringe Gruppengröße in den Klassen. Hierdurch war es unmöglich die drei klassischen Komponenten der Urteilsgenauigkeit für die vier nach dem Migrationshintergrund aufgeteilten Gruppen einzeln zu berechnen. Obwohl die Ergebnisse der Mehrebenenanalysen für eine Anpassung der Urteile der Lehrkräfte an die übermäßig ausgeprägten lesespezifischen Selbstwirksamkeitsüberzeugungen ihrer Schulkinder sprechen, konnte dies nicht durch eine separate Urteilsgenauigkeit in Bezug auf die mittleren Urteile des Niveaus, der Streuung und der

Rangfolge innerhalb der einzelnen Gruppen überprüft werden und muss, obwohl solche Stichproben schwierig zu finden sein werden, zukünftiger Forschung überlassen bleiben.

In Studie 3 zeigte sich nicht nur ein höheres Niveau der Urteile in Bezug auf die lesespezifische Selbstwirksamkeitsüberzeugung von Kindern mit türkischem oder sonstigem Migrationshintergrund bei einer durchschnittlichen Leseleistung. Zudem ergab sich eine Interaktion mit dem Leistungsurteil. Inwieweit dieser Interaktionseffekt zeitlich überdauernd ist, sollten längsschnittliche Betrachtungen der Migrationshintergrund-Leistungs-Interaktion zeigen. Längsschnittliche Befunde können auch dabei helfen die Stabilität der Urteilsgenauigkeit abzuschätzen. Mit Studie 3 wurde bewusst der erste Messzeitpunkt aus der längsschnittlich angelegten Studie gewählt, um eine erste Abschätzung der Urteilsgenauigkeit in Bezug auf die lesespezifischen Selbstwirksamkeitsüberzeugungen der Schulkinder zu erhalten. Der Gedanke dahinter war, dass hier eine unverfälschte Erfassung dieses Konstruktes erfolgen kann, während zum zweiten Messzeitpunkt bereits differenzielles Interesse der Lehrkraft an dem Thema zu einer unterschiedlichen Auseinandersetzung mit dem Thema Selbstwirksamkeitsüberzeugung und damit einhergehenden systematischen Verzerrungen hätte führen können. Bei einer längsschnittlichen Betrachtung sollte also möglichst auch das Interesse an Selbstwirksamkeitsüberzeugungen als Moderator möglicher Urteilsveränderungen betrachtet werden.

### **5.3 Implikationen**

Mit Blick auf die Ergebnisse der empirischen Arbeiten bleiben zwei grundsätzliche Implikationen für die Praxis festzuhalten: Die Kalibrierung von selbstbezogenen Fähigkeitskognitionen und Leistungen sowie die Sensibilisierung der Lehrkräfte für die

dazu notwendigen akkuraten Urteile in Bezug auf die Selbstwirksamkeitsüberzeugungen ihrer Schulkinder.

Die Ergebnisse der ersten beiden Studien weisen einerseits darauf hin, dass höhere selbstbezogene Fähigkeitskognitionen positiv mit der Leistungsentwicklung zusammenhängen. Dennoch konnten trotz relativ hoher selbstbezogener Fähigkeitskognitionen im Vergleich zu den Leistungen die Jugendlichen mit Migrationshintergrund den Leistungsrückstand zu ihren Mitschülerinnen und Mitschülern nicht aufholen. Die Kalibrierung der Selbstwirksamkeitsüberzeugungen und Fähigkeitsselbstkonzepte könnte ein Weg zur Lösung dieses persönlichen wie gesellschaftlichen Problems sein, denn es scheint bei Schulkindern mit Migrationshintergrund tendenziell eine übermäßige Zuversicht vorzuliegen. Obwohl eine optimistische Erfolgszuversicht durchaus leistungsfördernd wirken kann, da entsprechend Aufgaben leicht über dem eigentlichen Kompetenzniveau ausgewählt werden, die mit Anstrengung und Ausdauer jedoch lösbar sind und positive Leistungsentwicklungen mit sich bringen (Bandura, 1997). Hingegen führt Überoptimismus zu einer geringen Beschäftigung mit den Anforderungen der Aufgabe und geht oft mit einem Scheitern einher (Schunk & Pajares, 2009). Der Überoptimismus resultiert oft aus der Unklarheit über die tatsächlichen Anforderungen der Aufgabe und die zu ihrer Bewältigung erforderlichen Kompetenzen (Schunk & Pajares, 2009). Die unzureichende Kalibrierung kann jedoch auch durch die Lehrkräfte ausgelöst werden, wenn sie mit verbalen Bekräftigungen die Selbstwirksamkeitsüberzeugungen ihrer Schulkinder immer weiter erhöhen („Ich weiß, dass Du das schaffst!“; siehe Abschnitt 1.2.4), ohne dabei darauf zu achten, dass die notwendigen Kompetenzen erlernt werden, um die Aufgaben tatsächlich zu bewältigen. Dadurch können die Schulkinder sehr hohe Selbstwirk-

samkeitsüberzeugungen erlangen und dennoch sehr geringe Leistungen zeigen (Schunk & Pajares, 2009). Eine solche Situation mag bei den Schulkindern mit Migrationshintergrund eingetreten sein. Zu beachten ist auch, dass diese hohen Selbstwirksamkeitsüberzeugungen bei geringer Leistungsfähigkeit auch durch andere Quellen oder andere einflussreiche Personen, beispielsweise Eltern oder Gleichaltrige hervorgebracht worden sein können, ebenso wie Modelllernen durch Medien wie das Fernsehen (Usher & Pajares, 2008). Wenn in solchen Fällen von den Lehrkräften die Förderung der Selbstwirksamkeitsüberzeugungen der Schulkinder verlangt wird, dann sollte dies also nicht in einer weiteren Steigerung der Selbstwirksamkeitsüberzeugungen münden. Stattdessen sollten die Lehrkräfte zunächst das Niveau der Selbstwirksamkeitsüberzeugungen mit dem korrespondierenden Leistungsniveau abgleichen.

Bei der Arbeit an der Kalibrierung sollten die Lehrkräfte möglichst darauf achten, dass, sofern stark überhöhte Selbstwirksamkeitsüberzeugungen wie bei den Kindern mit Migrationshintergrund vorliegen, weder der Optimismus noch die Zuversicht der Schulkinder unter der Anpassung leiden (Pajares, 1996). Um die Kalibrierung der Selbstwirksamkeitsüberzeugungen erfassen zu können, erscheint eine aufgabenspezifische Operationalisierung unumgänglich. Eine domänen- oder fachspezifische Formulierung könnte kaum angemessen hinsichtlich der Passung der Stärke der Selbstwirksamkeitsüberzeugungen zur individuellen Leistungsfähigkeit genutzt werden, ohne auf soziale Vergleiche zurückzugreifen, wie es bei Vergleichen zwischen Schulkindern mit und ohne Migrationshintergrund im Theorieteil nötig war (vgl. Abschnitt 1.3.2). Oft müssen die Lehrkräfte jedoch auf informelle Urteilsbildung zurückgreifen, wie sie in einem Teil der vorliegenden Arbeit untersucht wurde. Die dort gefundenen Ergebnisse sollten Lehr-



kräfte motivieren, sich mit dem Konstrukt der Selbstwirksamkeitsüberzeugungen zu befassen, um die Vorteile für die Effektivität ihres Unterrichts nutzbar zu machen.

#### **5.4 Schlussfolgerung**

Mit den Erkenntnissen dieser Arbeit lässt sich einerseits zusammenfassen, dass die Frage nach der kausalen Ordnung, wie sie einleitend gestellt wurde, weiterhin nicht abschließend geklärt werden konnte. Dennoch liegen durch zwei Studien wichtige Erkenntnisse vor, mit denen die kulturelle Übertragbarkeit reziproker Effekte zwischen Leistungen und selbstbezogenen Fähigkeitskognitionen gestützt werden kann. Für die Aufgabe der Schule, ihre Schulkinder mit entsprechenden Fähigkeiten und selbstbezogenen Fähigkeitskognitionen für den weiteren Lebensweg auszustatten, wie im Eingangszitat zur Gesamtdiskussion angeführt, erscheint jedoch eine Sensibilisierung der Lehrkräfte für Selbstwirksamkeitsüberzeugungen von Schulkindern und deren Kalibrierung sinnvoll.

## 5.5 Literatur

- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57, 175-193.
- Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J. et al. (2007). *Förderung von Lesekompetenz - Expertise* (Bundesministerium für Bildung und Forschung (BMBF), Hrsg.) (Bildungsforschung Nr. 17), Bonn.
- Bandura, A. (1977). Self-efficacy. Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist*, 33, 344-358.
- Bandura, A. (1993). Perceived self-efficacy in cognitive functioning and development. *Educational Psychologist*, 28, 117-148.
- Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York, NY: W. H. Freeman.
- Baumert, J., Nagy, G. & Lehmann, R. (2012). Cumulative Advantages and the Emergence of Social and Ethnic Inequality: Matthew Effects in Reading and Mathematics Development Within Elementary Schools? *Child Development*, 83, 1347-1367.
- Begeny, J. C., Eckert, T. L., Montarello, S. A. & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23, 43-55.

- Bollen, K. A. & Curran, P. J. (2006). *Latent curve models. A structural equation perspective*. Hoboken: Wiley-Interscience.
- Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *Journal of Social Psychology, 130*, 353-363.
- Embretson, S. E. & Reise, S. (2000). *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum Associates, Publishers.
- Feinberg, A. B. & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *Journal of Educational Research, 102*, 453-462.
- Ferla, J., Valcke, M. & Cai, Y. (2009). Academic self-efficacy and academic self-concept. Reconsidering structural relationships. *Learning and Individual Differences, 19*, 499-505.
- Fischbach, A., Brunner, M., Krauss, S. & Baumert, J. (2015). Die Bezugsnormorientierung von Mathematiklehrkräften am Ende der Sekundarstufe I. Konvergenz verschiedener Messverfahren und Wirkung auf motivational-affektive Aspekte des Mathematiklernens und Leistung. *Journal for Educational Research Online, 7*(3), 3-27.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652-670.
- Glock, S. & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Social Psychology of Education, 16*, 111-127.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston: Kluwer Nijhoff Publishing.

- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision* (3. Aufl.). Göttingen: Beltz.
- Hornstra, L., van der Veen, I., Peetsma, T. & Volman, M. (2013). Developments in motivation and achievement during primary school: A longitudinal study on group-specific differences. *Learning and Individual Differences*, 23, 195-204.
- Huelmann, T., Ferdinand, H. D., Gebauer, M. M., McElvany, N., Bos, W., Köller, O. et al. (2014). Urteilsgenauigkeit von Lehrkräften bezüglich der Selbstwirksamkeitsüberzeugungen von Schülerinnen und Schülern. In H.-G. Holtappels, M. Pfeifer, A. S. Willems, W. Bos & N. McElvany (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Veröffentlichungen des Instituts für Schulentwicklungsforschung, Bd. 18, 1. Aufl., S. 298-320). Weinheim, Bergstr: Beltz Juventa.
- Jansen, M., Scherer, R. & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences. Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, 41, 13-24.
- Kaiser, J., Südkamp, A. & Möller, J. (2016). The effects of student characteristics on teachers' judgment accuracy. Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology* (Advance Online Publication).
- Kigel, R. M., McElvany, N. & Becker, M. (2015). Effects of immigrant background on text comprehension, vocabulary, and reading motivation. A longitudinal study. *Learning and Instruction*, 35, 73-84.
- Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking. Methods and practices* (Statistics for Social Science and Public Policy, 2. Aufl.). New York NY: Springer New York.

- Kubinger, K. D. & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem *Rasch-Modell*. *Diagnostica*, 53, 131-143.
- Lüdtke, O. & Köller, O. (2002). Individuelle Bezugsnormorientierung und soziale Vergleiche im Mathematikunterricht. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 34, 156-166.
- Lüdtke, O., Köller, O., Marsh, H. W. & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary Educational Psychology*, 30, 263-285.
- Maier, U. (2014). *Lerndiagnostik und Leistungsbeurteilung* (UTB M (Medium-Format), Bd. 4178, 1. Aufl.). Bad Heilbrunn: UTB; Klinkhardt, Julius.
- Marsh, H. W., Byrne, B. M. & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement. Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, 34, 155-167.
- Marsh, H. W. & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1, 133-163.
- Marsh, H. W. & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, 58, 364-376.
- Marsh, H. W. & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, 96, 56-67.

- Marsh, H. W., Hau, K.-T. & Kong, C.-K. (2002). Multilevel causal ordering of academic self-concept and achievement. Influence of Language of Instruction (English Compared With Chinese) for Hong Kong Students. *American Educational Research Journal*, 39, 727-763.
- Marsh, H. W. & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations. Reunification of east and west German school systems after the fall of the Berlin wall. *Contemporary Educational Psychology*, 29, 264-282.
- Marsh, H. W. & Shavelson, R. (1985). Self-concept. Its multifaceted, hierarchical structure. *Educational Psychologist*, 20 (3), 107-123.
- McElvany, N., Kortenbruck, M. & Becker, M. (2008). Lesekompetenz und Lesemotivation. *Zeitschrift für Pädagogische Psychologie*, 22, 207-219.
- Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte. Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau*, 55, 19-27.
- Möller, J., Retelsdorf, J., Köller, O. & Marsh, H. W. (2011). The Reciprocal Internal/External Frame of Reference Model: An Integration of Models of Relations Between Academic Achievement and Self-Concept. *American Educational Research Journal*, 48, 1315-1346.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66, 543-578.
- Pajares, F. & Schunk, D. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. J. Riding & S. Rayner (Hrsg.), *International perspectives on individual differences* (S. 239-266). Westport, CT: Ablex.

- Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S. & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology, 34*, 29-48.
- Ployhart, R. E. & Vandenberg, R. J. (2010). Longitudinal Research. The Theory, Design, and Analysis of Change. *Journal of Management, 36*, 94-120.
- Praetorius, A.-K., Karst, K., Dickhäuser, O. & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht, 58*, 81-91.
- Retelsdorf, J., Köller, O. & Möller, J. (2014). Reading achievement and reading self-concept – Testing the reciprocal effects model. *Learning and Instruction, 29*, 21-30.
- Roebers, C. M., Mecheril, A. & Schneider, W. (1998). Migrantenkinder in deutschen Schulen. Eine Studie zur Persönlichkeitsentwicklung. *Zeitschrift für Pädagogik, 44*, 723-736. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:0111-opus-68373>
- Rubin, D. B. & Thomas, N. (2000). Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates. *Journal of the American Statistical Association, 95*(450), 573-585.
- Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie, 23*, 237-245.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik, 1*, 27-52.
- Schunk, D. H. & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (Educational Psychology Handbook Series, pp. 35-54). New York, NY: Routledge.

- Schwippert, K., Wendt, H. & Tarelli, I. (2012). Lesekompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 191-208). Münster: Waxmann.
- Shajek, A., Lüdtke, O. & Stanat, P. (2006). Akademische Selbstkonzepte bei Jugendlichen mit Migrationshintergrund. *Unterrichtswissenschaft, 34*, 125-145.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie, 19*, 85-95.
- Stanat, P., Rauch, D. & 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). Münster: Waxmann.
- Urban, D. (2004). *Neue Methoden der Längsschnittanalyse. Zur Anwendung von latenten Wachstumskurvenmodellen in Einstellungs- und Sozialisationsforschung*. Münster: Lit.
- Usher, E. L. & Pajares, F. (2008). Sources of self-efficacy in school. Critical review of the literature and future directions. *Review of Educational Research, 78*, 751-796.
- Valentine, J. C., Dubois, D. L. & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*, 111-133.
- Van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.



Vandenberg, R. J. & Lance, C. E. (2000). A Review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-70.

Williams, T. & Williams, K. (2010). Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology*, 102, 453-466.



## Lebenslauf

### Christian Schöber

Geburtstag und –ort: 07.01.1975, Hameln

Staatsangehörigkeit: deutsch

### Werdegang

---

2017	Promotion in Psychologie, CAU, Kiel
Seit 2011	Wissenschaftlicher Mitarbeiter mit Zielsetzung Promotion in Psychologie am IPN, Kiel
2002-2011	Studium Psychologie, CAU, Kiel Abschluss: Diplompsychologe
1999-2002	Studium Rechtswissenschaften, CAU, Kiel
1981-1994	Schulbildung, zuletzt: Fachgymnasium Wirtschaft, Hameln Abschluss: Abitur