

Herkunftsassoziierte Lehrkrafturteile und -erwartungen:
Soziale Kognitionen und Urteilsbildungsprozesse im Kontext einer
ethnisch und sozial heterogenen Schülerschaft

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der

Philosophisch-Sozialwissenschaftlichen Fakultät

der

Universität Augsburg

vorgelegt von
Anita Tobisch
(2017)

Erstgutachter: Prof. Dr. Markus Dresel

Lehrstuhl für Psychologie

Zweitgutachter: Prof. Dr. Andreas Hartinger

Lehrstuhl für Grundschulpädagogik und Grundschuldidaktik

Tag der mündlichen Prüfung: 20.11.2017

Zusammenfassung

Basierend auf empirischen Befunden zu Disparitäten im Bildungsverlauf und -erfolg von Schüler(inne)n unterschiedlicher ethnischer und sozialer Herkunft (z. B. Braun & Mehringer, 2010; Stubbe, Bos & Euen, 2012) stellt sich die Frage nach den Ursachen für das schlechtere Abschneiden von Schüler(inne)n mit (v. a. türkischem) Migrationshintergrund und/oder niedrigem sozioökonomischem Status (vgl. Mehringer, 2013; Müller & Ehmke, 2013). Diese herkunftsassoziierten Disparitäten können nicht ausschließlich durch geringere Fähigkeiten oder Leistungen erklärt werden. So erhalten Schüler(innen) mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status im Vergleich zu Schüler(inne)n ohne Migrationshintergrund und mit hohem sozioökonomischem Status, auch unter Kontrolle der tatsächlichen Leistungen und Fähigkeiten, schlechtere Zensuren sowie seltener Empfehlungen für das Gymnasium (z. B. Arnold, Bos, Richert & Stubbe, 2007; Merkens & Wessel, 2002). Angenommen werden multikausale Einflussfaktoren (z. B. Stubbe & Bos, 2008), wobei insbesondere die Grundschule (u. a. durch ihre Selektionsfunktion) und auch die Lehrkraft eine besondere Verantwortung für Bildungsverläufe von Schüler(inne)n trägt (vgl. Ditton, 2016).

Theoretische Modelle betonen v. a. die Bedeutung von akkuraten Lehrkrafturteilen und herkunftsunabhängigen Erwartungen für einen adaptiven Unterricht (z. B. Helmke, 2009; Schrader & Helmke, 2001). Jedoch konnten empirische Studien meist nur eine moderat ausgeprägte diagnostische Kompetenz von Lehrkräften feststellen (z. B. Südkamp, Kaiser & Möller, 2012) sowie herkunftsassoziierte Unterschiede in Lehrkrafturteilen und -erwartungen aufdecken (z. B. Dunkake & Schuchart, 2015; Glock, 2016). Bislang ist jedoch nicht geklärt, ob diese herkunftsassoziierten Urteilsunterschiede Ausdruck negativ verzerrter Lehrkrafturteile bei Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status sind oder positiv verzerrte Urteile bei Schüler(inne)n ohne Migrationshintergrund und mit hohem sozioökonomischem Status darstellen. Die ersten Befunde für den deutschsprachigen Raum, die sich mit dieser Frage befassen, stammen von Kaiser, Südkamp und Möller (2016) und deuten auf eine akkurate Einschätzung von Schüler(inne)n mit Migrationshintergrund hin.

Darauf aufbauend stellt sich die Frage nach den zugrundeliegenden kognitiven Prozessen, die herkunftsassoziierte Urteile und Erwartungen von Lehrkräften erklären könnten. Anzunehmen ist, dass der soziale Informationsverarbeitungsprozess durch bereits bestehende mentale Strukturen, wie z. B. Einstellungen und Stereotype, beeinflusst wird (z. B. Fiske & Neuberg, 1990; Hamilton, Sherman & Ruvolo, 1990). Dabei weisen Lehrkräfte und Lehramtsstudierende negativere Einstellungen gegenüber Kindern mit Migrationshintergrund im Vergleich zu Kindern ohne Migrationshintergrund sowie negativere Einstellungen gegenüber Kindern mit niedrigem sozioökonomischem Status im Vergleich zu Kindern mit hohem sozioökonomischem Status auf (z. B. Glock & Karbach, 2015; Glock, Krolak-Schwerdt & Hörstermann, 2016). Basierend auf dem Kontinuum-Modell der Eindrucksbildung (Fiske & Neuberg, 1990) stellt sich die Frage, ob bei der Beurteilung von Schüler(inne)n unterschiedlicher Herkunft diese mentalen Strukturen wirksam sind und sich für manche Schülergruppen eine eher kategoriebasierte oder eher individuumsbasierte Informations-

verarbeitung zeigt. Die vorliegende Arbeit setzt an diesen Fragen an und geht ihnen in zwei empirischen Untersuchungen nach.

In einer experimentellen Online-Studie mit $N = 237$ erfahrenen Grundschullehrkräften konnten unter dem Einsatz von Fallvignetten (männlicher Schüler; Manipulation der ethnischen und sozialen Herkunft) zunächst herkunftsassoziierte Urteils- und Erwartungsunterschiede von Lehrkräften für Leistungen bestätigt und für leistungsrelevante Schülermerkmale aufgedeckt werden. Dabei deuten die Befunde auf eine akkurate Einschätzung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status sowie auf eine positive Urteilsverzerrung bei Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status hin. In einer zweiten experimentellen Studie mit $N = 45$ Lehramtsstudierenden wurden ebenfalls Fallvignetten eingesetzt und die ethnische und soziale Herkunft von männlichen Schülern manipuliert. Durch die Erfassung okulomotorischer Daten mit einem Eye-Tracker konnten Rückschlüsse auf die Informationsverarbeitung gezogen werden. Die Ergebnisse verweisen auf eine höhere kognitive Aktivierung bei der Urteilsbildung über Schüler mit Migrationshintergrund und niedrigem sozioökonomischem Status und eine geringere kognitive Aktivierung bei der Bildung von Urteilen über Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status. Entsprechend kann eine eher kontrollierte Informationsverarbeitung bei Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status und eine eher automatisierte Informationsverarbeitung bei Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status angenommen werden. Unter Berücksichtigung der impliziten ethnischen Einstellungen von Lehramtsstudierenden zeigte sich zudem, dass eine eher automatisierte Informationsverarbeitung bei der Beurteilung von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status, v. a. bei Studierenden mit starken positiven Assoziationen gegenüber Personen ohne Migrationshintergrund auftritt.

Inhaltsverzeichnis

Zusammenfassung	i
1. Heterogenität als Herausforderung für Lehrkräfte	1
2. Herkunftsassoziierte Diversität im schulischen Kontext	5
2.1 Ethnische und soziale Herkunft von Schüler(inne)n in Deutschland	5
2.2 Herkunftsassoziierte Disparitäten im Bildungsverlauf und -erfolg	9
2.2.1 Schulische Leistungen nach ethnischer und sozialer Herkunft von Schüler(inne)n	11
2.2.2 Bildungsbeteiligung nach ethnischer und sozialer Herkunft von Schüler(inne)n	14
2.2.3 Bildungserfolg nach ethnischer und sozialer Herkunft von Schüler(inne)n	17
2.3 Erklärungsansätze für ethnische und soziale Disparitäten im Bildungssystem ..	19
3. Lehrkrafturteile und -erwartungen im Kontext einer heterogenen Schülerschaft	29
3.1 Diagnostische Kompetenz von Lehrkräften.....	32
3.1.1 Empirische Befunde zur Urteilsgenauigkeit von Lehrkräften	35
3.1.2 Einflussfaktoren auf die Akkuratheit von Lehrkrafturteilen.....	37
3.2 Lehrkrafterwartungen und Erwartungseffekte	42
3.3 Herkunftsassoziierte Lehrkrafturteile und -erwartungen	48
4. Einstellungen und Stereotype von Lehrkräften im Kontext einer heterogenen Schülerschaft	55
4.1 Einstellungen und Stereotype gegenüber ethnischen und sozialen Gruppen	61
4.2 Herkunftsspezifische Einstellungen und Stereotype von Lehrkräften	65
5. Lehrkrafturteilen und -erwartungen unterliegende soziale Kognitionen	68
5.1 Der Prozess sozialer Informationsverarbeitung	71
5.2 Soziale Kategorien im Kontext der Informationsverarbeitung von Lehrkräften	74
5.3 Top-down und bottom-up Prozesse bei der Informationsverarbeitung von Lehrkräften.....	80
6. Zusammenfassung und Herleitung der Forschungsfragen	89
6.1 Zusammenfassung des Forschungsstands zu herkunftsassoziierten Lehrkrafturteilen und -erwartungen und Forschungsdefizite.....	89
6.2 Forschungsfragen und Hypothesen der vorliegenden Arbeit.....	92

7. Überblick über das empirische Vorgehen in Studie 1 und Studie 2	94
7.1 Stimulusmaterial in Studie 1 und Studie 2.....	98
7.1.1 Konstruktion der Fallvignetten	98
7.1.2 Namensauswahl zur Aktivierung der ethnischen und sozialen Schülerherkunft	100
7.2 Messung von Lehrkrafturteilen über und -erwartungen an Leistungen und leistungsrelevante Merkmale in Studie 1 und Studie 2.....	103
8. Studie 1: Herkunftsassoziierte Verzerrungen von Lehrkrafturteilen und -erwartungen	106
8.1 Vorstudie zum Einsatz von Fallvignetten und Testung der Erhebungsinstrumente.....	106
8.2 Spezifische Fragestellungen und Hypothesen von Studie 1	112
8.3 Untersuchungsdesign und Stichprobe der ersten Studie	113
8.4 Erhebungsinstrumente in Studie 1	115
8.5 Analysemethoden der ersten Studie.....	116
8.6 Ergebnisse von Studie 1	117
8.7 Diskussion von Studie 1.....	123
9. Studie 2: Herkunftsabhängige Informationsverarbeitungsprozesse von Lehramtsstudierenden.....	127
9.1 Eye-Tracking als Methode zur Erfassung kognitiver Prozesse	127
9.1.1 Physiologische Grundlagen der Okulomotorik.....	129
9.1.2 Psychologische Grundlagen der Okulomotorik	132
9.2 Vorstudie zum Einsatz eines Eye-Trackers bei der Schülerbeurteilung.....	136
9.3 Spezifische Fragestellungen und Hypothesen von Studie 2	138
9.4 Untersuchungsdesign und Stichprobe der zweiten Studie	140
9.5 Erhebungsinstrumente in Studie 2	146
9.6 Analysemethoden und Aufbereitung der okulomotorischen Daten der zweiten Studie	150
9.7 Ergebnisse von Studie 2.....	153
9.8 Diskussion von Studie 2.....	167
10. Zusammenfassung und Gesamtdiskussion der empirischen Untersuchungen dieser Arbeit.....	172
Literaturverzeichnis.....	187
Appendix	220
Appendix A Noten und Textbausteine der Fallvignetten.....	221
Appendix B Ziel- und Attributionskategorien im impliziten Assoziationstest.....	225
Appendix C Dokumentation der Fragebogenskalen der Studien 1 & 2.....	226
Appendix D Ergänzende Tabelle: Studie 1	236
Appendix E Ergänzende Tabellen: Studie 2	237

1. Heterogenität als Herausforderung für Lehrkräfte

Seit geraumer Zeit existiert auf bildungspolitischer und wissenschaftlicher Ebene ein Diskurs über interkulturelle Bildung und herkunftsunabhängige Bildungsgerechtigkeit (z.B. Kultusministerkonferenz [KMK], 1996; Göbel & Hesse, 2004; Neumann & Reuter, 2004; Stojanov, 2011). Dieser führt zwar zu einem wachsenden Problembewusstsein, jedoch bestehen weiterhin große Herausforderungen im Umgang mit ethnisch-kultureller Diversität in der schulischen Praxis (Allemann-Ghionda, 2006). Dabei umfasst der Diskurs nicht nur ethnisch-kulturelle Herausforderungen, sondern weitere Heterogenitätsmerkmale der Lernenden – so weisen Schüler(innen) u.a. heterogene Vorwissensbestände und Vorstellungen, Fähigkeiten, Sprachkenntnisse sowie heterogene familiäre und soziale Hintergründe auf. Zudem ist insbesondere die ethnische Herkunft der Schüler(innen) meist mit weiteren sozialen Merkmalen, z.B. dem sozioökonomischen Status, verknüpft (z.B. Herwartz-Emden, Schurt & Waburg, 2010), wodurch sich ein komplexes Bündel heterogener Schülermerkmale ergibt.

In Bezug auf einzelne Dimensionen der Heterogenität, wie individuelle Lern- und Leistungsfähigkeit, wird eine Homogenisierung durch Selektionsmechanismen im Bildungssystem angestrebt. Diese Bemühungen stoßen jedoch an ihre Grenzen, da sich auch in unterschiedlichen Schulformen der Sekundarstufe I Lernende mit unterschiedlichen Fähigkeiten und Kompetenzen befinden (z.B. Hohn, Schiepe-Tiska, Sälzer & Artelt, 2013; Schiepe-Tiska, Schöps, Rönnebeck, Köller & Prenzel, 2013). Ein professioneller und reflektierter Umgang mit Heterogenität ist somit für Lehrkräfte aller Schularten und Jahrgangsstufen eine bedeutsame Herausforderung. Dabei hat jedoch insbesondere die Grundschule durch ihre Selektionsfunktion eine besondere Bedeutung für den Bildungsverlauf von Schüler(inne)n. Um eine bestmögliche Unterstützung für jedes Kind und jeden Jugendlichen zu ermöglichen, müssen individuelle Stärken, Schwächen und Ressourcen durch die Lehrkraft wahrgenommen werden; ohne durch leistungsirrelevante Schülermerkmale verzerrt zu werden.

Terhart (2015) argumentiert im Sinne der kommunikationstheoretischen Annahme Watzlawicks (z.B. Watzlawik, Bavelas & Jackson, 2011): „[I]n irgendeiner Weise reagiert Schule, reagieren Lehrer immer auf Heterogenität – auch dann, wenn sie ignoriert wird. Man kann nicht nicht darauf reagieren“ (S. 25). Ein reflektierter und sensibilisierter Umgang mit jeder Art von Diversität ist demnach gefordert. Neben strukturellen Determinanten des Bildungssystems gelten auch Überzeugungen, Einstellungen und Kompetenzen der Lehrkräfte als maßgeblich für einen professionellen Umgang mit Heterogenität (Baumert & Kunter, 2006; Terhart, 2015). Zwei wesentliche Herausforderungen im Umgang mit vielfältigen Schülermerkmalen und der spezifischen Rolle der Lehrkraft können dabei herausgestellt werden: (1) Trotz großer Bemühungen seitens zahlreicher Lehrkräfte, dieser Aufgabe gerecht zu werden, fehlt es oftmals an personellen Ressourcen (z.B. Lehrkräftemangel; Rakhkotchikine & Kotthoff, 2015), um Kinder und Jugendliche in ihrer individuellen Entwicklung optimal zu unterstützen. (2) Es liegt in der Natur sozialer Informationsverarbeitungsprozesse, dass in Anbetracht begrenzter (z.B. zeitlicher oder kognitiver) Ressourcen auch Urteilsverzerrungen (z.B. Bless, Fiedler & Strack, 2004) zum Tragen kommen können. Dies erscheint insbesondere im Hinblick

auf herkunftsassoziierte Disparitäten in den schulischen Leistungen als ein bedeutsamer Faktor, den es zu berücksichtigen gilt.

Neben internationalen Schulleistungsstudien, wie z. B. der *Internationalen Grundschul-Lese-Untersuchung* (IGLU), *Trends in International Mathematics and Science Study* (TIMSS) oder *Programme for International Student Assessment* (PISA) – die regelmäßig auf die herkunftsassoziierten Disparitäten im deutschen Bildungssystem verweisen – liefern auch weitere empirische Befunde zum Bildungsverlauf und -erfolg ein überwiegend konsistentes Bild über das schlechtere Abschneiden von Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status im Vergleich zu Schüler(inne)n ohne Migrationshintergrund und/oder hohem sozioökonomischem Status (z. B. Dahl & Lochner, 2012). Differenzen in den Leistungen und Kompetenzen zwischen Kindern und Jugendlichen der Mehrheitsgesellschaft und ethnischen sowie sozialen Minoritäten sind nicht nur in Deutschland, sondern auch in zahlreichen weiteren Ländern aufzufinden (z. B. in den USA; Lee, 2002; Lesaux & Harris, 2013).

Die ethnische und soziale Herkunft scheint in Deutschland jedoch einen besonders hohen Einfluss auf den Bildungserfolg zu haben. So zeigen sich in internationalen Studien für Deutschland, im Vergleich zu anderen Ländern, besonders hohe Differenzen zwischen den Kompetenzen von Schüler(inne)n mit Migrationshintergrund und ohne Migrationshintergrund sowie von Kindern und Jugendlichen mit hohem und niedrigem sozioökonomischem Status (z. B. Gebhardt, Rauch, Mang, Sälzer & Stanat, 2013; Schwippert, Hornberg, Freiberg & Stubbe, 2007; Stubbe, Schwippert & Wendt, 2016). Schüler(innen) mit Migrationshintergrund sind im Vergleich zu Schüler(inne)n ohne Migrationshintergrund zudem überrepräsentiert an Haupt-/Mittelschulen und deutlich unterrepräsentiert an Gymnasien (z. B. Diefenbach, 2010; Statistisches Bundesamt, 2014), sie erreichen entsprechend häufiger nur niedrige Bildungsabschlüsse (z. B. Organisation for Economic Co-operation and Development [OECD], 2016a).

Oftmals liegt jedoch nicht nur ein Risikofaktor für schlechtere schulische Leistungen vor, sondern mehrere Risikofaktoren gleichzeitig. Personen mit Migrationshintergrund verfügen z. B. durchschnittlich über ein geringeres Haushaltseinkommen und sind häufiger von Armut betroffen als Personen ohne Migrationshintergrund (z. B. Cortés Núñez & Küçük, 2016; Fuhr, 2012). Kinder und Jugendliche mit spezifischen ethnischen und sozialen Hintergründen sind in ihrem Bildungsverlauf somit nachweislich mit zahlreichen zusätzlichen Herausforderungen konfrontiert. Anzunehmen sind infolgedessen multikausale Ursachen für herkunftsassoziierte Disparitäten, wobei auch Lehrkräfte einen bedeutsamen Einfluss haben.

Neben anderen Faktoren sind insbesondere Defizite in der Vorbereitung von Lehrkräften auf heterogenitätsspezifische Herausforderungen und Chancen zu vermerken. Die Qualifikation von Lehrkräften hat einen bedeutsamen Einfluss auf die Unterrichtsqualität, die Lern- und Leistungsentwicklung von Schüler(inne)n (vgl. Weinert, 2001) sowie auf den Umgang der Lehrkraft mit Heterogenität im Unterrichtsalltag. Trotz der bereits seit Beginn der 1980er Jahre in diesem Zusammenhang thematisierten Bedeutung der Lehrkraftausbildung wurden die Ausbildungsinhalte nur bedingt angepasst (Seitz, 2006). Wobei sich zunehmend ein Wandel durch Maßnahmen, wie z. B. die Qualitätsoffensive Lehrerbildung (Bundesministerium für Bildung und Forschung, 2016) und u. a. dem in diesem Zusammenhang geförderten Projekt

Lehrerprofessionalität im Umgang mit Heterogenität (LeHet; Universität Augsburg, 2016) abzeichnet. Um die Aus- und Weiterbildung von Lehrkräften jedoch adäquat verbessern zu können, bedarf es theoretischer Modelle und empirischer Befunde zu herkunftsassoziierten Lehrkrafturteilen und -erwartungen und den zugrundeliegenden psychischen Mechanismen.

Anzunehmen ist zwar, dass einige der Einflussfaktoren auf Disparitäten im Bildungsverlauf und -erfolg strukturellen und nicht primär psychologischen Ursprungs sind, jedoch sind psychologische Erklärungsansätze intrapersoneller (z.B. kognitiver und motivationaler Prozesse) und interpersoneller (z.B. Prozesse sozialer Informationsverarbeitung) Faktoren der Lehrkraft nicht zu unterschätzen (vgl. American Psychological Association [APA], 2012), die nicht erst seit den Befunden der Meta-Analyse von Hattie (2015) diskutiert werden. So gibt es bereits seit den 1960er Jahren einige empirische Hinweise darauf, dass z.B. bei der Notenvergabe leistungsirrelevante Informationen zu einer Urteilsverzerrung führen können (z.B. Weiss, 1965). Die Schülerbeurteilung durch Lehrkräfte unterliegt offenbar (wie alle sozialen Informationsverarbeitungsprozesse) der Gefahr von automatisierten Abläufen, die zu Urteilsfehlern führen können. Unklar ist jedoch, ob diese Urteilsverzerrungen eine Benachteiligung von Schüler(inne)n ethnischer Minoritäten und/oder mit niedrigem sozioökonomischem Status oder eine Bevorzugung von Schüler(inne)n ohne Migrationshintergrund und/oder mit hohem sozioökonomischem Status darstellen. Bislang ist zudem nicht geklärt, ob sich die zugrundeliegenden psychischen Mechanismen der Urteilsbildung von Lehrkräften in Bezug auf die ethnische und soziale Herkunft von Schüler(inne)n unterscheiden.

An dieser Stelle setzt die vorliegende Arbeit an und fokussiert Lehrkrafturteile und -erwartungen im Kontext einer ethnisch und sozial heterogenen Schülerschaft. In zwei experimentellen Studien wurden zum einen herkunftsassoziierte Unterschiede in Lehrkrafturteilen und -erwartungen untersucht und zum anderen die zugrundeliegenden psychischen Mechanismen der sozialen Informationsverarbeitung fokussiert. Damit befindet sich die vorliegende Arbeit an der Schnittstelle zwischen pädagogisch-psychologischen Ansätzen (u. a. durch die Berücksichtigung der diagnostischen Kompetenz von Lehrkräften), sozialpsychologischen Theorien (u. a. durch Einbezug der Urteils- und Erwartungsbildung) sowie kognitionspsychologischen Grundlagen (u. a. durch die Analyse der Informationsverarbeitung).

Ziel dieser Arbeit ist es, Unterschiede in Lehrkrafturteilen über und -erwartungen an Leistungen und leistungsrelevante Schülermerkmale in Abhängigkeit der ethnischen und sozialen Herkunft der Lernenden zu analysieren. Zudem soll geklärt werden, ob bestimmte Schülergruppen durch ihre Herkunft von Lehrkräften benachteiligt oder bevorzugt werden. Basierend auf dem Kontinuum-Modell der Eindrucksbildung (Fiske & Neuberg, 1990) wird angenommen, dass sowohl eine Bevorzugung als auch eine Benachteiligung bestimmter Schüler(innen) wegen ihrer ethnischen und/oder sozialen Herkunft durch entsprechend positive oder negative Assoziationen mit diesen sozialen Kategorien und einer eher kategoriebasierten Informationsverarbeitung entstehen kann. Rückschlüsse auf die Informationsverarbeitung wurden durch die Erfassung der *Okulomotorik* (Augenbewegungen) bei der Urteils- und Erwartungsbildung von angehenden Lehrkräften ermöglicht. Hierdurch kann die vorliegende Arbeit einen wesentlichen Beitrag zur Analyse herkunftsassoziiierter Urteils- und Erwartungsunterschiede in Bezug auf schulische Leistungen und leistungsrelevante Merkmale liefern,

mögliche Benachteiligungen oder Bevorzugungen von Schülergruppen identifizieren und zudem die zugrundeliegenden Informationsverarbeitungsprozesse aufdecken.

Um Fragen nach herkunftsassoziierten Unterschieden von Lehrkrafturteilen und deren kognitiver Verarbeitung zu beantworten, werden in dieser Arbeit zunächst die grundlegenden Befunde zu Disparitäten im schulischen Kontext in Abhängigkeit der ethnischen und sozialen Herkunft von Kindern und Jugendlichen dargestellt (Kapitel 2). In diesem Zusammenhang erfolgt eine Auseinandersetzung mit allgemeinen Einflussfaktoren auf den Bildungsverlauf und -erfolg sowie eine Beantwortung der Frage, welche Schüler(innen) unter den Kategorien *Schüler(innen) mit Migrationshintergrund* und/oder *niedrigem sozioökonomischem Status* zu fassen sind. Aufbauend auf einer Übersicht über mögliche Einflussfaktoren auf Disparitäten im Bildungsverlauf und -erfolg, wird die Rolle der Lehrkraft v.a. durch ihre Urteile über und Erwartungen an Leistungen und leistungsrelevante Merkmale von Schüler(inne)n fokussiert (Kapitel 3). In diesem Zusammenhang werden theoretische Modelle und empirische Befunde zur Akkuratheit von Lehrkrafturteilen vorgestellt und im Kontext einer heterogenen Schülerschaft diskutiert. Anschließend erfolgt die Darstellung sozial-kognitiver Erklärungsansätze für das Zustandekommen herkunftsassoziierter Lehrkrafturteile. Dabei werden herkunftsassozierte Einstellungen und Stereotype thematisiert (Kapitel 4) und der Prozess sozialer Informationsverarbeitung betrachtet (Kapitel 5), da diese sozialpsychologischen Theorien die Grundlage für das theoretische Verständnis und die empirische Erfassung von herkunftsassozierten Lehrkrafturteilen und -erwartungen sowie den zugrundeliegenden psychischen Mechanismen bilden. Nach einer Zusammenfassung des Forschungsstands sowie der Darstellung der daraus resultierenden Fragestellungen (Kapitel 6) erfolgt ein Überblick über die beiden experimentellen Studien dieser Arbeit (Kapitel 7).

Der Frage, inwiefern sich Lehrkrafturteile und -erwartungen nach ethnischer und/oder sozialer Herkunft unterscheiden, wurde in einer experimentellen Online-Studie (Studie 1) mit erfahrenen Grundschullehrkräften unter dem Einsatz von Fallvignetten nachgegangen (Kapitel 8). Anschließend nähert sich eine experimentelle Laboruntersuchung (Studie 2) mit Hilfe der Erfassung verhaltensnaher Daten (Okulomotorik) den zugrundeliegenden kognitiven Prozessen bei der Urteilsbildung im Kontext einer heterogenen Schülerschaft an (Kapitel 9). Abschließend werden die Ergebnisse beider Studien zusammenfassend kritisch diskutiert, Limitationen reflektiert und praktische Implikationen abgeleitet sowie weitere Forschungsdesiderata formuliert (Kapitel 10).

2. Herkunftsassoziierte Diversität im schulischen Kontext

„Zu den wichtigsten bildungspolitischen Zielen demokratischer Gesellschaften gehört es, allen Heranwachsenden gleich gute Bildungschancen zu geben, sie individuell optimal zu fördern und gleichzeitig soziale, ethnische und kulturelle Disparitäten der Bildungsbeteiligung und des Bildungserfolgs auszugleichen“ (Baumert & Schümer, 2001, S. 323). Trotz zahlreicher Forderungen nach herkunftsunabhängigen Bildungschancen (vgl. z.B. Baumert & Schümer, 2001) verweisen empirische Befunde weiterhin auf herkunftsassoziierte Disparitäten im Bildungsverlauf und -erfolg von Schüler(inne)n. Durch das Zusammenspiel unterschiedlicher Herkunftsmerkmale, wie z.B. einem Migrationshintergrund, einer anderen Muttersprache als Deutsch, einem bildungsfernen Elternhaus oder einem niedrigen sozioökonomischen Status, steigt die Wahrscheinlichkeit, schlechtere schulische Leistungen zu erzielen (OECD, 2016a). Beschlüsse der Kultusministerkonferenzen (z.B. KMK, 1971, 2002, 2007, 2015) verweisen dabei regelmäßig auf die Bedeutung der Eingliederung von Personen mit Migrationshintergrund in das deutsche Schulsystem und auf die Entwicklung geeigneter Maßnahmen in Form von z.B. individueller Beratung oder Sprachfördermaßnahmen. Auch die Qualifizierung der Lehrkräfte sowie die Notwendigkeit der Personalaufstockung werden in diesem Zusammenhang häufig diskutiert. Da bedingt durch begrenzte finanzielle Ressourcen nur in seltenen Fällen eine Erhöhung des Personals möglich ist und auch die Qualifizierung von Lehrkräften nur über einen langfristigen Prozess mit zahlreichen Einflussfaktoren erfolgen kann, erweist sich der Umgang mit Heterogenität im schulischen Alltag häufig als große Herausforderung für Lehrende.

Für einen adäquaten Umgang mit einer ethnisch und sozial heterogenen Schülerschaft sowie der empirischen Analyse möglicher herkunftsassoziierter Urteils- und Erwartungsunterschiede ist es unabdingbar, zunächst einen differenzierten Blick auf diese Schülergruppen zu werfen. Welche Schüler(innen) werden unter der Kategorie des Migrationshintergrunds zusammengefasst und wie hoch ist der Anteil dieser Schüler(innen) im deutschen Schulsystem? Wie stellt sich die sozioökonomische Lage von Schüler(inne)n dar und wie hängt diese mit der ethnischen Herkunft zusammen? Basierend auf der Beantwortung dieser Fragen wird in den folgenden Abschnitten ebenfalls ein differenzierter Blick auf die herkunftsassozierten Disparitäten im deutschen Schulsystem geworfen. Hierzu wird zunächst geklärt, wie sich ein erfolgreicher Bildungsverlauf beschreiben lässt und welche Einflussfaktoren identifiziert werden können. Abschließend werden theoretische Annahmen und empirische Befunde als Erklärungsansätze für herkunftsassoziierte Disparitäten diskutiert.

2.1 Ethnische und soziale Herkunft von Schüler(inne)n in Deutschland

Migrationsbewegungen sind aus historischer Perspektive kein neues Phänomen (z.B. Bade, 2000; Marsella & Ring, 2003), dennoch stellen sie weiterhin große Herausforderungen an gesellschaftliche, bildungspolitische sowie schulische Entscheidungen und Handlungen. In Deutschland erfolgten erstmals in den 1970er Jahren im Rahmen der sogenannten *Ausländerpädagogik* – bedingt durch die Zuwanderung von angeworbenen Gastarbeitern in den 1960er

Jahren – pädagogische und psychologische Diskussionen über den Umgang mit Schüler(inne)n mit Migrationshintergrund und die Konsequenzen der Zuwanderung für den schulischen Kontext (vgl. u.a. Seitz, 2006). Die anfänglich noch stark defizitorientierte Perspektive fokussierte dabei meist auf Unzulänglichkeiten der Schüler(innen) im Beherrschen der deutschen Sprache und deren Förderung in z.B. Vorbereitungsklassen (Auernheimer, 1996; Diehm & Radtke, 1999; Seitz, 2006).

Seit 1950 wurden in Deutschland u. a. 4.5 Millionen (Spät-)Aussiedler(innen) aufgenommen und insgesamt mehr als 5.2 Millionen Personen eingebürgert (Brückner, 2016). Durch anhaltende Immigrationsprozesse und Nachwuchs der immigrierten Personen nahm auch die Zahl der Menschen mit einer Migrationsgeschichte in Deutschland in den folgenden Jahren kontinuierlich zu. Laut Statistischem Bundesamt (Brückner, 2016) liegt ein Migrationshintergrund vor, wenn man selbst oder mindestens ein Elternteil ohne deutsche Staatsbürgerschaft geboren wurde. Unberücksichtigt bleiben dabei z. B. Schüler(innen), deren Großeltern aus sogenannten Gastarbeiterstaaten nach Deutschland kamen. Häufig haben diese Kinder und ihre Eltern einen deutschen Pass, sind kulturell und sprachlich jedoch z. T. noch mit der Herkunftskultur der Großeltern verbunden.

Die häufig uneinheitliche Operationalisierung des Migrationshintergrunds, aber auch die teilweise Vernachlässigung einiger relevanter Merkmale wie z. B. der Sprache, Aufenthaltsdauer und -status oder der ethnisch-kulturellen Identität erschweren eine Vergleichbarkeit unterschiedlicher Untersuchungen und die Berücksichtigung relevanter Aspekte für den schulischen Kontext und die gesellschaftliche Integration (vgl. u. a. Kemper, 2010). Zur Berücksichtigung zumindest einiger dieser Aspekte und den Einbezug des Migrationshintergrunds der Kinder in empirischen Untersuchungen, erweist sich das Geburtsland der Eltern (oder eines Elternteils) als besserer Indikator für einen Migrationshintergrund als die Staatsangehörigkeit. Die Berücksichtigung von Kindern, bei denen nur ein Elternteil im Ausland geboren wurde, ist dabei ebenfalls von Bedeutung, da sich auch hier je nach Herkunftsregion bedeutsame Differenzen in z. B. dem Umgang mit Mitschüler(inne)n aber auch dem schulischen Problemverhalten zeigen (Karssen, van der Veen & Volman, 2017).

Laut Datenreport des Statistischen Bundesamtes lebten 2014 ca. 16.4 Millionen Menschen mit Migrationshintergrund in Deutschland (dies entspricht ca. 20.3 % der Gesamtbevölkerung Deutschlands), davon sind 21.8 % der Personen mit Migrationshintergrund unter 15 Jahren (Brückner, 2016; Statistisches Bundesamt, 2015). Insgesamt haben ein Drittel der unter 15-Jährigen und 35 % der unter 10-Jährigen in Deutschland einen Migrationshintergrund (Maaz et al., 2016; Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration, 2014) und stellen somit einen bedeutsamen Anteil der Schülerschaft im deutschen Schulsystem. Zu beachten ist jedoch, dass die Gruppe der als mit Migrationshintergrund bezeichneten Personen in sich deutlich heterogener ist als die Außenwahrnehmung durch vereinheitlichende Begriffe suggeriert. Unterschieden werden müssen dabei nicht nur unterschiedliche Herkunftsländer und -regionen, sondern u. a. auch Gründe für die Migration (z. B. Arbeitsmigration, Flucht, akademischer Austausch), die Aufenthaltsdauer und Bleibeaussicht, eigene Migrationserfahrungen, Zugang zu Sozialleistungen oder Einbürgerungsmöglichkeiten (z. B. Cortés Núñez & Küçük, 2016; Herwartz-Emden, 2003). Je nach spezifischer Konstellation dieser

Merkmale ist mit besonderen Herausforderungen und Chancen der Personen mit Migrationshintergrund zu rechnen.

Mit Blick auf die unterschiedlichen Herkunftsländer und -regionen sind Personen aus ehemaligen Gastarbeiter-Anwerbeländern gefolgt von (Spät-)Aussiedler(inne)n die größten Herkunftsgruppen (Statistisches Bundesamt, 2017). In Bayerns Grundschulen sind alleine bezogen auf die Staatsbürgerschaft im Schuljahr 2008/09 Kinder mit türkischen (gefolgt von serbischen, italienischen, griechischen, albanischen, kroatischen und bosnischen) Wurzeln die größte Gruppe der ausländischen Kinder (Kemper, 2010). Dabei sind jedoch Kinder mit ausländischen Wurzeln und deutscher Staatsbürgerschaft nicht in dieser Statistik aufgeführt, wodurch mit einem deutlich höheren Anteil an Kindern mit Migrationshintergrund im deutschen Schulsystem zu rechnen ist. Da z. B. 38 % der in Deutschland lebenden Personen aus ehemaligen Gastarbeiterstaaten über die deutsche Staatsbürgerschaft verfügen (Statistisches Bundesamt, 2017), erschweren einige amtliche Statistiken und empirische Studien den Überblick über die tatsächliche Anzahl von Kindern und Jugendlichen mit Migrationshintergrund. Dennoch bleibt die Gruppe der Personen mit türkischem Migrationshintergrund auch nach differenzierter Betrachtung die zahlenmäßig bedeutsamste Migrantengruppe. Für die gesamte Bundesrepublik stellen Schüler(innen) mit türkischen Wurzeln mit 42.5 % die größte Gruppe mit Migrationshintergrund in allgemeinbildenden Schulen dar (vgl. Diefenbach, 2010).

Die meist unumstritten hohe Konfundierung der ethnischen und sozialen Herkunft (z. B. Herwartz-Emden, 2003; Herwartz-Emden et al., 2010; Mehringer, 2013) erschwert dabei die kausale Zuordnung der bestehenden Differenzen im Bildungsverlauf und -erfolg. Die enge Verschränkung von Armutsgefährdung und Migrationshintergrund zeigt sich in zahlreichen Ländern, wie z. B. in den USA. Dort sind 12 % der Personen bis 17 Jahren mit europäischen Wurzeln und 12 % mit asiatischem Hintergrund von Armut betroffen, während dies 33 % der lateinamerikanischen und 38 % der afroamerikanischen Bevölkerung der entsprechenden Altersgruppe betrifft (Kena et al., 2015). Deutlich wird in den Analysen von Kena et al. (2015), dass zwischen unterschiedlichen Herkunftsregionen differenziert werden muss. Personen asiatischer Abstammung haben ein vergleichbares Armutsrisiko wie Personen der amerikanischen Majorität. Afro- und lateinamerikanische Personen dagegen scheinen mit anderen sozio-ökonomischen Herausforderungen konfrontiert zu sein. In Deutschland ist das Armutsrisiko für Personen mit Migrationshintergrund im Vergleich zu Personen ohne eine Migrationsgeschichte ebenfalls mehr als doppelt so hoch (Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration, 2014; Cortés Núñez & Küçük, 2016; Statistisches Bundesamt, 2015).¹

Auch ein Vergleich des durchschnittlichen monatlichen Nettoeinkommens von Personengruppen unterschiedlicher Herkunft verdeutlicht, dass die ethnische Herkunft und soziale Schicht in Deutschland eng miteinander verknüpft sind. Bei Erwerbstätigen mit einer Vollzeitbeschäftigung liegt das monatliche Nettoeinkommen für Personen ohne Migrationshintergrund bei durchschnittlich 2.235 €, während Menschen mit Wurzeln in den ehemaligen Anwerbeländern für Gastarbeiter mit 1.946 € und mit Wurzeln in den neuen EU-

¹ Ein Armutsrisiko liegt nach behördlichen Definitionen vor, wenn das Einkommen weniger als 60 % des durchschnittlichen Einkommens in Deutschland beträgt (Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration, 2014; Statistisches Bundesamt, 2015).

Mitgliedsstaaten (z.B. Slowenien, Ungarn, Rumänien; Bundesamt für Migration und Flüchtlinge [BAMF], 2014) mit 1.789 € rechnen können (Brückner, 2016). Dementsprechend lässt sich auch erklären, warum Schüler(innen) mit Migrationshintergrund in internationalen Klassifizierungen des beruflichen Status (z.B. *International Socio-Economic Index*, ISEI; Ganzeboom, de Graaf & Treiman, 1992) einen deutlich geringeren sozioökonomischen Status erreichen. Dies betrifft dabei insbesondere Schüler(innen) mit türkischem Migrationshintergrund (u. a. Mehringer, 2013).

Die Koppelung des Migrationshintergrunds und der sozialen Schicht zeigt sich jedoch nicht nur in Einkommensdifferenzen, sondern auch in den unterschiedlichen Bildungsabschlüssen, die einen Einfluss auf den beruflichen Status und entsprechendes Einkommen haben können. Die Eltern der unter 20-jährigen Kinder und Jugendlichen mit Migrationshintergrund weisen dabei deutlich häufiger keinen Schul- oder Berufsabschluss auf und haben seltener einen Hochschulabschluss als Eltern ohne Migrationshintergrund (Maaz et al., 2016). Die geringeren Bildungsabschlüsse sowie die höhere Arbeitslosenquote unter Personen mit Migrationshintergrund (u. a. Herwartz-Emden, 2003) sind dabei als Risikofaktoren für den Bildungsverlauf ihrer Kinder anzusehen. Im Jahr 2014 wiesen 55 % der Kinder im Alter zwischen sechs und zehn Jahren der ersten Migrationsgeneration und 42 % der zweiten Generation mindestens einen Risikofaktor (Erwerbslosigkeit, geringes Einkommen oder niedrige Qualifikation der Eltern) auf. Im Gegensatz hierzu beträgt der Anteil an Kindern ohne Migrationshintergrund, die einen der Risikofaktoren aufweisen, lediglich 20 % (Maaz et al., 2016). Der Anteil an Kindern zwischen sechs und zehn Jahren, die alle drei Risikofaktoren vereinen, beträgt bei Kindern ohne Migrationshintergrund 2 %, in der zweiten Generation 6 % und in der ersten Generation 10 % (Maaz et al., 2016). Groos und Jehles (2015) berichten von 40 % der Kinder mit Migrationshintergrund und 13 % ohne Migrationshintergrund, die von Armut betroffen sind.

Kinder, deren Eltern einen hohen Bildungsabschluss aufweisen, sind zudem mit lediglich 8 % auf Sozialleistungen angewiesen, bei Kindern mit Eltern, die einen niedrigeren Bildungsabschluss aufweisen, sind dies knapp 65 % (Groos & Jehles, 2015). Insgesamt leiden in Deutschland 19 % aller unter 18-jährigen Kinder und Jugendlichen unter Armut (von zur Gathen & Liebert, 2016). Nach von zur Gathen und Liebert (2016) werden im Bericht zur Armutsentwicklung in Deutschland unter Armut neben der Einkommensarmut (Ressourcenansatz) auch die Teilhabechancen (Lebenslagenansatz) sowie die Chancen zur realen Verwirklichung von Potenzialen (Verwirklichungsansatz) verstanden. Fokussiert wird dabei jedoch meist auf die Einkommensarmut, die bei weniger als 60 % des monatlichen Medianeinkommens vorliegt (vgl. auch Bundesministerium für Arbeit und Soziales, 2013). Eine Analyse der Kinder in Haushalten mit Arbeitslosengeld-II-Bezügen (Hartz-IV) weist dabei jedoch starke regionale Schwankungen auf. In Bayern liegt der Anteil bei 7.2 %, in Berlin dagegen betrifft dies jedes dritte Kind unter 15 Jahren (von zur Gathen & Liebert, 2016).

Dabei sind die Auswirkungen von Armut deutlich umfangreicher als ausschließlich ökonomische Einschränkungen. Zusammenhänge der familiären sozioökonomischen Situation zeigen sich mit einer Vielzahl weiterer Faktoren, die Einfluss auf die allgemeine kindliche Entwicklung und den schulischen Bildungsverlauf haben können. Brooks-Gunn und Duncan (1997) verweisen auf zahlreiche Befunde großangelegter Untersuchungen, die verdeutlichen, dass von Armut betroffene Kinder und Jugendliche einen schlechteren gesundheitlichen

Zustand aufweisen sowie häufiger von emotionalen Problemen oder Verhaltensauffälligkeiten betroffen sind. Der Entwicklungsverlauf dieser Kinder und Jugendlichen ist somit durch zahlreiche Risiken gefährdet, die sich auch auf die schulischen Fähigkeiten und Leistungen auswirken können.

Durch die Analyse der amtlichen Statistiken und empirischen Befunde zur ethnischen und sozialen Herkunft von Schüler(inne)n im deutschen Schulsystem wird deutlich, dass ein nicht zu vernachlässigender Anteil an Kindern und Jugendlichen eine Migrationsgeschichte aufweist und/oder aus sozioökonomisch niedrigeren Schichten kommt. Der hohe Zusammenhang zwischen ethnischer Herkunft und sozialem Status von Schüler(inne)n in Deutschland erfordert zudem eine Berücksichtigung beider Faktoren in empirischen Analysen zum Bildungsverlauf und -erfolg sowie in Untersuchungen zu herkunftsassoziierten Lehrkraturteilen und -erwartungen. Nur durch die Berücksichtigung sowohl der ethnischen und als auch der sozialen Herkunft kann der Einfluss des jeweiligen Merkmals auf Disparitäten im Bildungsverlauf aufgeklärt werden. Um jedoch die Bedeutung und das Ausmaß herkunftsassoziiierter Disparitäten zu verstehen, wird im folgenden Abschnitt geklärt, in welchen Bereichen und in welchem Ausmaß herkunftsabhängige Unterschiede im Bildungsverlauf und -erfolg vorliegen.

2.2 Herkunftsassoziierte Disparitäten im Bildungsverlauf und -erfolg

Trotz eines leichten Rückgangs herkunftsbezogener Disparitäten im Schulerfolg sind weiterhin signifikante Unterschiede zwischen Schüler(inne)n unterschiedlicher ethnischer und sozialer Herkunft zu verzeichnen (Baumert, Watermann & Schümer, 2003; Geißler, 1994; Marks, 2004; Schimpl-Neimanns, 2000). Zunächst stellt sich dabei jedoch einerseits die Frage, welche schulischen Meilensteine im Bildungsverlauf berücksichtigt werden sollten und andererseits, inwiefern sich Bildungserfolg definieren lässt. Um die spezifische Situation sozialer Gruppen im deutschen Schulsystem zu analysieren und um einschätzen zu können, inwiefern diese Gruppen das Schulsystem mehr oder weniger erfolgreich durchlaufen, müssen somit zunächst Indikatoren definiert werden, welche die wesentlichen strukturellen Merkmale des Bildungsverlaufs und -erfolgs operationalisieren. Der schulische Werdegang lässt sich nach Diefenbach (2010) in drei zentrale Bereiche untergliedern: Schulleistung, Bildungsbeteiligung und Bildungserfolg.

Die Erfassung der *Schulleistungen* spezifischer Schülergruppen erfolgt dabei u. a. über die Zensuren sowie über erreichte Testwerte der Schülergruppen. Des Weiteren wird der Anteil an Übergangsempfehlungen für unterschiedliche weiterführende Schulen für diese Schülergruppen ermittelt. Da die Übergangsempfehlung der Lehrkraft maßgeblich für die Wahl der weiterführenden Schulform ist, wäre auch eine Kategorisierung im Unterpunkt Bildungsbeteiligung denkbar. Zu beachten ist dabei, dass durch die föderale Struktur des deutschen Bildungssystems bundeslandspezifische Regelungen zur Verbindlichkeit der Übergangsempfehlungen vorliegen und Eltern z. B. in Bayern stärker an die Lehrkräftempfehlung gebunden sind als Eltern in z. B. Hessen (vgl. z. B. Kleine, Birnbaum, Zielonka, Doll & Blossfeld, 2010). Aber auch in Bayern wurde der Elternwille seit dem Schuljahr 2009/2010 deutlich gestärkt (Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, o. J., 2009).

Die Übergangsempfehlung ist zwar v.a. durch die Leistungen der Schüler(innen) beeinflusst, jedoch nicht ausschließlich durch leistungsrelevante Schülermerkmale erklärbar. Ditton und Krüsken (2006a) konnten zeigen, dass schulische Vorleistungen und Kompetenztests lediglich 56% der Varianz in den Übergangentscheidungen erklären. Zu erwähnen ist dabei, dass bereits die schulischen Leistungen in Form von Noten durch Verzerrungen im Urteilsbildungsprozess der Lehrkräfte beeinflusst sein können und nicht unbedingt die tatsächlichen Fähigkeiten und Leistungen der Schüler(innen) widerspiegeln müssen (vgl. Kapitel 3). Neben leistungsrelevanten Faktoren, wie z. B. motivationalen Ausprägungen, erweist sich u. a. auch der soziale Status als leistungsirrelevanter Einflussfaktor (Stubbe & Bos, 2008). Dollmann (2011) konnte zeigen, dass eine bindende Übergangsempfehlung einen positiven Effekt für Kinder aus sozial niedrigeren Schichten haben kann. Die Übergangsempfehlung der Lehrkraft hat somit einen bedeutsamen Einfluss auf den Bildungsverlauf. Auch wenn Lehrkräftempfehlungen für die weiterführenden Schulen nicht in jedem Bundesland bindend sind, haben sie dennoch einen bedeutsamen Einfluss auf die elterliche Wahl der besuchten weiterführenden Schulform (Stubbe & Bos, 2008) und bedingen die Bildungsbeteiligung in der Sekundarstufe.

Da sie als Ergebnis bisheriger Leistungen und Leistungsbewertungen anzusehen sind, werden Übergangsempfehlungen in dieser Arbeit im Gegensatz zur Gliederung nach Diefenbach (2010) nicht dem Bereich der Leistungen, sondern der *Bildungsbeteiligung* zugeordnet. Über den Vergleich des Anteils spezifischer Schülergruppen an verschiedenen Schulformen werden Informationen über die Bildungsbeteiligung gesammelt, wobei sich insbesondere die Frage nach dem Anteil an Kindern mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status an spezifischen weiterführenden Schulformen stellt. Zudem ist ein relevanter Indikator für die Bildungsbeteiligung die Über- oder Unterrepräsentation bestimmter Schülergruppen im Vergleich zur Gesamtschülerschaft an den Schulformen der Sekundarstufe I.

Der *Bildungserfolg* stellt den dritten Indikator für die schulische Situation spezifischer Schülergruppen dar. Dieser wird in Form von erreichten formalisierten Bildungsabschlüssen, wie z. B. dem Abitur und der entsprechenden Abschlussnote erfasst. Dabei stehen Schulleistungen, Bildungsbeteiligung und Bildungserfolg in engem Zusammenhang zueinander. Die Bildungsbeteiligung an weiterführenden Schulen ergibt sich in hohem Maße aus den schulischen Leistungen in der Primarstufe. Die schulischen Leistungen sowie die Bildungsbeteiligung in der Sekundarstufe wiederum bedingen den Bildungserfolg. Neben diesen von Diefenbach (2010) benannten Indikatoren, veranschaulicht Tabelle 1 weitere Aspekte, die Hinweise für einen mehr oder weniger erfolgreichen Bildungsverlauf von Schüler(inne)n liefern:

Tabelle 1

Indikatoren zur Erfassung des Bildungsverlaufs und -erfolgs spezifischer Schülergruppen (eigene Erweiterung nach Diefenbach, 2010)

Schulleistungen	Bildungsbeteiligung	Bildungserfolg
Erreichte Punktzahlen/ Kompetenzstufen in Leistungstests	Vorschulische Bildungs- beteiligung	Erreichte Bildungs- abschlüsse und Bildungszertifikate
Noten in einzelnen Schulfächern	Übergangsempfehlung für weiterführende Schulen	Anteil an Schüler(inne)n ohne einen Schulabschluss
	Anteil sowie Über-/ Unterrepräsentation an verschiedenen Schulformen	Durchschnittsnote in den Abschlussprüfungen
	Anteil an spezifischen Förderprogrammen	

Der Bildungserfolg kann zwar über die erreichten Qualifikationen der Bildungsabschlüsse erfasst werden, jedoch ist es hierbei ebenfalls von Bedeutung, Schüler(innen) zu berücksichtigen, die bereits vor dem Antritt der Abschlussprüfung die Schullaufbahn abgebrochen haben. Dabei sollte entsprechend der Fokussierung auf bestimmte ethnische und soziale Schülergruppen der Anteil an Schulabbrecher(inne)n betrachtet werden. Für den Bildungsverlauf ebenfalls relevant erweist sich nicht nur die Frage nach der Bildungsbeteiligung an unterschiedlichen Schulformen, sondern auch der Anteil an Schüler(inne)n in besonderen Förderprogrammen, wie z. B. Begabtenförderung, die einen wesentlichen Einfluss auf den weiteren Bildungsverlauf und -erfolg von Schüler(inne)n haben können. Um insbesondere den Bildungsverlauf nachvollziehen zu können, ist es zudem von wesentlicher Bedeutung, die Bildungsbeteiligung nicht erst zum Schuleintritt zu betrachten, sondern auch die Nutzung vorschulischer pädagogischer Einrichtungen zu berücksichtigen, da hier bereits zentrale Weichen für den weiteren schulischen Werdegang gestellt werden.

Um Disparitäten im deutschen Bildungssystem nach ethnischer und sozialer Herkunft der Schüler(innen) analysieren zu können, werden im Folgenden empirische Befunde zu den schulischen Leistungen, der vorschulischen und schulischen Bildungsbeteiligung, sowie dem Bildungserfolg von Schüler(inne)n unterschiedlicher ethnischer und sozialer Herkunft systematisch nach den hier dargestellten Kriterien zusammengefasst.

2.2.1 Schulische Leistungen nach ethnischer und sozialer Herkunft von Schüler(inne)n

In den Zensuren der Schüler(innen) zeigen sich z. T. große Unterschiede je nach ethnischer und/oder sozialer Herkunft, wobei v. a. Schüler(innen) mit türkischem Migrationshintergrund schlechtere Noten erhalten. In der vierten Jahrgangsstufe zeigte sich z. B. im Fach Deutsch eine durchschnittliche Abweichung von einer Notenstufe für türkischstämmige Schüler(innen) im Vergleich zu deren Mitschüler(inne)n ohne Migrationshintergrund (vgl. u. a. Mehringer, 2013). Die schulischen Noten unterscheiden sich aber nicht nur in Bezug auf die

ethnische Herkunft. In einer Untersuchung von Merkens und Wessel (2002) zeigte sich, dass sowohl in leistungsstarken als auch leistungsschwachen Klassen (ermittelt über einen allgemeinen Schulleistungstest) Schüler(innen) mit hohem sozioökonomischem Status (auch unter Kontrolle der tatsächlichen Leistungen) durchgängig bessere Noten erhalten als Schüler(innen) aus niedrigen sozialen Schichten. Dabei können zunächst vielfältige Ursachen für die Disparitäten in den Zensuren angenommen werden. Einerseits könnten tatsächliche Unterschiede in den schulischen Fähigkeiten vorliegen, andererseits könnten die Noten bereits Ausdruck von verzerrten Lehrkrafturteilen sein. Befunde, die zwar keine Unterschiede in standardisierten Tests, jedoch unterschiedliche Zensuren auffinden (z.B. Merkens & Wessel, 2002), verdeutlichen, dass schlechtere Zensuren nicht durch die tatsächlichen Leistungen erklärt werden können und Urteilsfehler der Lehrkräfte vorliegen könnten (vgl. Kapitel 3).

Differenzielle herkunftsassoziierte Unterschiede finden sich jedoch nicht nur in Zensuren, auch in Kompetenztests zeigt sich ein vergleichbares Muster. Ditton und Krüsken (2006a) konnten z.B. zeigen, dass unter Kontrolle der Vorleistungen die Leistungsentwicklung v. a. in Mathematik von Herkunftsmerkmalen (v. a. dem Bildungsabschluss der Eltern und der Muttersprache) abhängt. Sowohl für die Primar- als auch Sekundarstufe I bestehen weiterhin große Unterschiede in den erreichten Kompetenzniveaus unterschiedlicher Schulleistungstests. Besonders Schüler(innen) türkischer Herkunft weisen bereits in der ersten Jahrgangstufe eine deutlich geringere Mathematik-, Lese- und Rechtschreibkompetenz im Vergleich zu Schüler(inne)n ohne Migrationshintergrund auf (Mehringer, 2013). Erwartungsgemäß zeigt sich zwar eine Verbesserung der Kompetenzen der Kinder mit Migrationshintergrund in den vier Grundschuljahren, jedoch bleiben sie dauerhaft unter den Kompetenzen der Schüler(innen) ohne Migrationshintergrund (Mehringer, 2013). Zudem haben Schüler(innen) türkischer Herkunft bereits im ersten Schuljahr schlechtere Mathematik-, Lese- und Rechtschreibkompetenzen als Schüler(innen) anderer ethnischer Herkunft, wie z. B. aus GUS-Staaten (*Gemeinschaft unabhängiger Staaten*; Zusammenschluss einiger Nachfolgestaaten der ehemaligen Sowjetunion) oder Osteuropa (Mehringer, 2013).

In den IGLU und TIMSS Untersuchungen bestätigen sich die substantiellen Unterschiede nach ethnischer und sozialer Herkunft der Grundschüler(innen) in den Lesekompetenzen und in Kompetenzen im mathematisch-naturwissenschaftlichen Bereich (Bonsen, Frey & Bos, 2008; Bonsen, Kummer & Bos, 2008; Bos, Schwippert & Stubbe, 2007; Gebhardt et al., 2013; Müller & Ehmke, 2013; Schwippert, Bos & Lankes, 2003; Schwippert et al., 2007; Schwippert, Wendt & Tarelli, 2012; Stubbe et al., 2016; Stubbe, Tarelli & Wendt, 2012; Tarelli, Schwippert & Stubbe, 2012; Wendt, Schwippert & Stubbe, 2016; Wendt, Stubbe & Schwippert, 2012) (vgl. Abbildung 1 & Abbildung 2): In diesen Untersuchungen schnitten Schüler(innen) aus hohen Sozialschichten signifikant besser ab als Schüler(innen) aus niedrigen sozialen Schichten. Zudem erreichten Schüler(innen) ohne Migrationshintergrund höhere Kompetenzstufen als Schüler(innen) mit Migrationshintergrund. Dabei schnitten Kinder mit beiden im Ausland geborenen Elternteilen schlechter ab als Kinder mit einem im Ausland geborenen Elternteil.

Trotz gewisser positiver Veränderungen in den vergangenen Jahren – z. B. Kompetenzzuwachs im Lesen für Kinder mit beiden im Ausland geborenen Elternteilen (vgl. Abbildung 1) sowie insgesamt für Schüler(innen) mit Migrationshintergrund in den mathematischen Kompe-

tenzen (vgl. Abbildung 2) – bestehen weiterhin signifikante Unterschiede nach ethnischer und sozialer Herkunft der Schüler(innen).

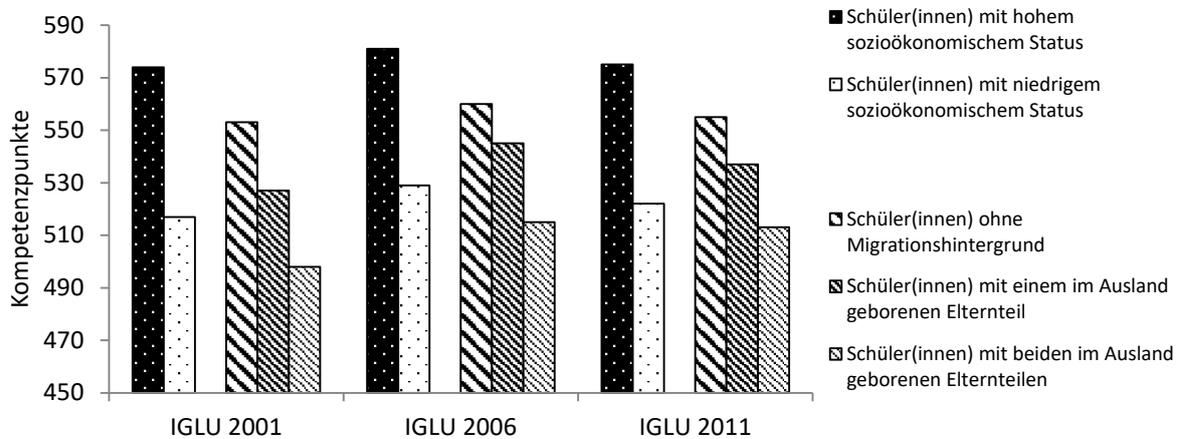


Abbildung 1. Mittelwerte der Lesekompetenz nach Schülerherkunft (IGLU, 2001, 2006, 2011)²

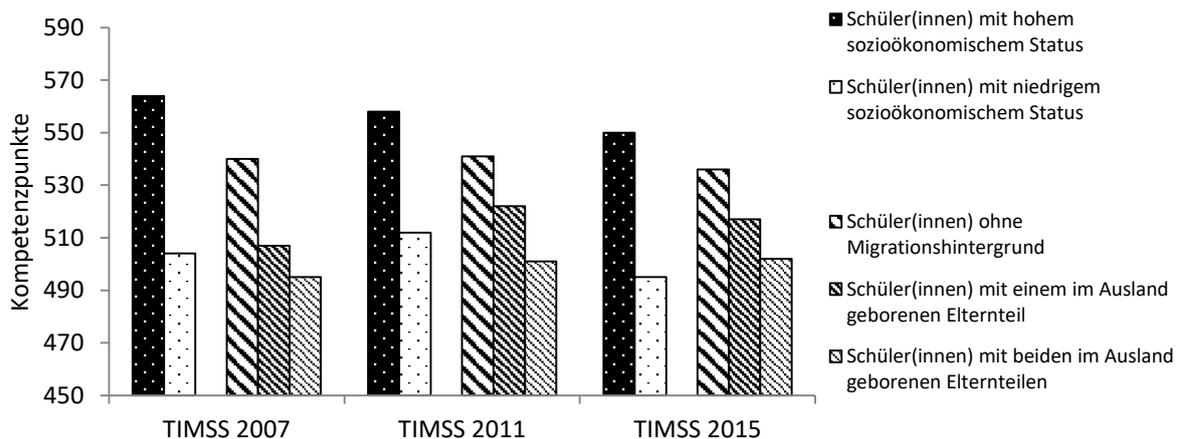


Abbildung 2. Mittelwerte der Mathematikkompetenz nach Schülerherkunft (TIMSS, 2007, 2011, 2015)³

Ähnliche Befundmuster der Leistungsdifferenzen nach ethnischer und sozialer Herkunft zeigen sich auch in den PISA-Daten für die Sekundarstufe I (u.a. Ehmke, Hohensee, Heidemeier & Prenzel, 2004; Ramm, Prenzel, Heidemeier & Walter, 2004). Dabei konnten in Deutschland 21.1 % der Varianz in den mathematischen Kompetenzen durch soziokulturelle Faktoren erklärt werden, wobei hier insbesondere der sozioökonomische Status der Schüler(innen) einen bedeutsamen Einfluss hatte (Prenzel, Heidemeier, Ramm, Hohensee & Ehmke, 2004). Die Abhängigkeit der erreichten Kompetenzstufen von der sozialen Herkunft konnten

² Eigene Darstellung nach Daten der IGLU Erhebungen (Schwippert et al., 2012; Wendt et al., 2012).

³ Eigene Darstellung nach Daten der TIMSS Erhebungen (Stubbe et al., 2012; Tarelli et al., 2012; Stubbe et al., 2016; Wendt et al., 2016).

Baumert et al. (2003) ebenfalls in den PISA-Daten durch die Berechnung des sozialen Gradienten aufzeigen.⁴ Dabei wies Deutschland im internationalen Vergleich den steilsten sozialen Gradienten auf – bei der Veränderung der Sozialschicht um eine Standardabweichung verändert sich die Lesekompetenz um 45 Kompetenzpunkte – im Vergleich dazu lag z. B. in Finnland lediglich eine Veränderung um 20 Punkte vor.⁵ Jedoch ist der sozioökonomische Status kein alleiniger Erklärungsansatz für die unterschiedlichen Kompetenzen von Kindern und Jugendlichen mit und ohne Migrationshintergrund. In den Daten der PISA Studie 2015 zeigten sich für den naturwissenschaftlichen Bereich auch nach Kontrolle des sozioökonomischen Hintergrunds signifikant geringere Kompetenzwerte bei Schüler(inne)n mit Migrationshintergrund (OECD, 2016b). Die unterschiedlichen schulischen Zensuren und erreichten Kompetenzen wirken sich dabei natürlich auch auf die Entscheidung für spezifische Schulformen aus und führen somit u. a. zu einer entsprechenden Bildungsbeteiligung.

2.2.2 Bildungsbeteiligung nach ethnischer und sozialer Herkunft von Schüler(inne)n

Ungleiche Bildungsverläufe beginnen bereits vor der Grundschulzeit, zum einen durch unterschiedliche lernförderliche familiäre Umgebungen (z. B. García, 2015), zum anderen durch differenzielle frühkindliche Förderung (z. B. Groos & Jehles, 2015) sowie einer selektiven Nutzung vorschulischer Einrichtungen (z. B. Becker, 2010b; Kreyenfeld, 2004). Auch wenn sich in den vergangenen Jahren eine Annäherung der Betreuungsquote von Kindern mit und ohne Migrationshintergrund abzeichnet, besuchen Kinder mit Migrationshintergrund im Vergleich zu Kindern ohne Migrationshintergrund seltener Kinderkrippen und Kindergärten (Groos & Jehles, 2015; Kreyenfeld, 2004; Maaz et al., 2016; Schwippert et al., 2012). Zudem verweisen empirische Arbeiten darauf, dass Kinder aus einkommensschwachen Familien sowie Kinder, deren Eltern über ein niedriges Bildungsniveau verfügen, ebenfalls deutlich seltener vorschulische Einrichtungen nutzen (Groos & Jehles, 2015; Kreyenfeld, 2004). So sind Eltern zudem durch geringe ökonomische Ressourcen und/oder lange Arbeitszeiten in ihrer Wahl der Betreuungseinrichtung eingeschränkt und entscheiden sich hierdurch z. T. für alternative Betreuungsmöglichkeiten oder berücksichtigen nur bedingt die pädagogische Qualität einer Einrichtung bei ihrer Entscheidung (Becker, 2010b).

Die geringeren sprachlichen Kompetenzen (z. B. Wortschatz und Grammatik) bei Kindern im Vorschulalter mit (insbesondere türkischem) Migrationshintergrund sowie geringerem Bildungsstand der Eltern (Maaz et al., 2016) können somit nicht immer durch eine optimale Förderung in pädagogischen Einrichtungen ausgeglichen werden. Dabei erweist sich ein Kindergartenbesuch – unter Berücksichtigung der pädagogischen Qualität der Einrichtung – in vielerlei Hinsicht als förderlich, z. B. für den schulischen Werdegang (z. B. Garces, Thomas & Currie, 2002). Insbesondere durch eine Förderung kognitiver, sprachlicher und sozialer Kompetenzen (z. B. Becker & Biedinger, 2006; Biedinger & Becker, 2006, 2010), aber auch

⁴ Sozialer Gradient: Steigung und Senkung der Regressionsgeraden bei Berechnung einer linearen Regression.

⁵ Jedoch zeigt sich für Deutschland im Bundesländervergleich ein heterogenes Bild. Während die Steigung des sozialen Gradienten in Nordrhein-Westfalen knapp 53 Punkte oder in Bayern 43 Punkte betrug, lag er in Brandenburg bei ca. 35 Punkten (Baumert et al., 2003). Ein differenzierter Blick auf die Unterschiede in den Bundesländern erscheint somit lohnenswert.

durch eine Vorbereitung auf den Schulalltag (vgl. Diefenbach, 2010) erweist sich der Kindergartenbesuch als besonders relevant und nützlich. Vor allem Kinder mit Migrationshintergrund und/oder mit niedrigem sozioökonomischem Status profitieren von einer vorschulischen Betreuung (Becker & Lauterbach, 2008; Büchel, Spieß & Wagner, 1997). Schöler, Guggenmos, Hasselbach und Iseke (2005) konnten entsprechend enge Zusammenhänge zwischen dem Besuch einer Kindertagesstätte und den Deutschkenntnissen zu Beginn der Schulzeit aufdecken. Zudem zeigte sich, dass die Deutschkenntnisse umso besser waren, je länger eine vorschulische Einrichtung besucht wurde (Schöler et al., 2005). Diese Befunde deuten bereits vor Schuleintritt auf eine ungünstige Bildungsbeteiligung von Kindern mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status hin.

Für die ersten Schuljahre ist ein differenzierter Blick auf eine herkunftsassoziierte Bildungsbeteiligung durch die eingliedrige Primarstufe kaum möglich. Neben privaten Trägern mit unterschiedlichen pädagogischen Konzepten (z. B. Montessori- oder Waldorfschulen) besteht eine fähigkeitsbezogene Differenzierung der Schülerschaft lediglich im Hinblick auf besonderen Förderbedarf. Dabei zeigt sich ebenfalls ein deutlich höherer Anteil an Kindern mit Migrationshintergrund und/oder mit niedrigem sozioökonomischem Status an Förderschulen als dies für Kinder ohne Migrationshintergrund und/oder mit hohem sozioökonomischem Status der Fall ist (Hasselhorn et al., 2014). In der Primarstufe zeichnen sich somit trotz eingliedriger Struktur herkunftsassoziierte Differenzen in der Bildungsbeteiligung, zumindest in Bezug auf den besonderen Förderbedarf, ab.

Zum Ende der Grundschulzeit werden die Weichen für eine selektive Bildungsbeteiligung durch die Vergabe von Übergangsempfehlungen gestellt, wobei sich ebenfalls herkunftsassoziierte Zusammenhänge feststellen lassen (z. B. Braun & Mehringer, 2010). Böhmer, Hörstermann, Gräsel, Krolak-Schwerdt und Glock (2015) verweisen darauf, dass neben den Zensuren und dem Arbeitsverhalten der Schüler(innen) v. a. die elterliche Unterstützung sowie der elterliche Schulwunsch als relevante Informationen für Lehrkräfte bei der Übergangsempfehlung dienen. Somit sind neben den tatsächlichen Fähigkeiten und Leistungen der Schüler(innen) ebenso familiäre Aspekte bedeutsam für die Entscheidung der Lehrkraft. Insbesondere nach Kontrolle der kognitiven Fähigkeiten sowie der Kompetenzen in den Bereichen Lesen, Mathematik und Naturwissenschaften haben Schüler(innen) mit höherem sozioökonomischem Status deutlich höhere Chancen das Gymnasium zu besuchen als Schüler(innen) mit niedrigem sozioökonomischem Status (Arnold, Bos, Richert & Stubbe, 2007; Baumert et al., 2003; Stubbe, Bos & Euen, 2012).

Dies korrespondiert mit Befunden von Mehringer (2013), wonach migrationsspezifische Differenzen in der Lehrkräftempfehlung für eine weiterführende Schule bestehen. Kinder ohne Migrationshintergrund haben eine fünfmal höhere Chance das Gymnasium zu besuchen als ihre Mitschüler(innen) mit Migrationshintergrund. Durch die Differenzierung nach Staatsbürgerschaft wurde deutlich, dass v. a. Kinder ohne deutsche Staatsbürgerschaft die höchste Wahrscheinlichkeit für einen Haupt-/Mittelschulbesuch aufweisen und zudem auch eine geringere Chance für den Realschulbesuch haben. Anzumerken ist zudem, dass im Vergleich zu Kindern ohne Migrationshintergrund auch Schüler(innen), die als eingebürgerte Migrant(inn)en gelten, aber auch Kinder mit einem (Spät-)Aussiedler-Hintergrund eine signifikant geringere Wahrscheinlichkeit haben, ein Gymnasium zu besuchen (Mehringer, 2013). Zudem konnten Unter-

schiede zwischen Schüler(inne)n mit einem im Ausland geborenen Elternteil und beiden im Ausland geborenen Elternteilen aufgezeigt werden (Mehringer, 2013): Kinder, deren beide Elternteile im Ausland geboren wurden, besuchen mit einer höheren Wahrscheinlichkeit eine Haupt-/Mittelschule als Kinder, bei denen ein Elternteil im Ausland geboren wurde. Angenommen wird, dass sich dieser Unterschied u. a. durch eine stärkere soziale Integration und bessere sprachliche Fähigkeiten der Kinder mit einem Elternteil ohne Migrationshintergrund erklären lässt (Mehringer, 2013). In den Analysen von Kristen und Dollmann (2009, 2010) zeigte sich jedoch, dass sich die Nachteile türkischstämmiger Schüler(innen) vollständig durch deren Leistungen und soziale Herkunft erklären lassen. Zudem ergab sich durch eine sehr hohe Bildungsmotivation der türkischstämmigen Familien sogar eine höhere Wahrscheinlichkeit auf eine höhere Schulform in der Sekundarstufe I zu wechseln. Dabei verbleibt jedoch weiterhin eine Benachteiligung durch den sozioökonomischen Status. Die Befunde einer höheren Wahrscheinlichkeit für einen Gymnasialbesuch können jedoch nicht einheitlich in anderen empirischen Studien bestätigt werden.

Für Schüler(innen) mit Migrationshintergrund zeigte sich ebenfalls unter Kontrolle der schulischen Zensuren und der kognitiven Fähigkeiten eine geringere Wahrscheinlichkeit für eine Gymnasialempfehlung durch die Lehrkraft (Arnold et al., 2007; Kristen, 2002; Mehringer, 2013; Stubbe et al., 2012). Zwar können die kognitiven Fähigkeiten und v. a. die Lesekompetenz die geringere Wahrscheinlichkeit auf einen Gymnasialbesuch z. T. erklären, dennoch verbleibt auch nach Kontrolle der Fähigkeiten und Leistungen in den IGLU 2006 Daten ein Effekt des Migrationshintergrunds (Arnold et al., 2007). Unter Berücksichtigung des sozioökonomischen Status verringert sich zwar der Effekt der ethnischen Herkunft auf die Übergangsempfehlung, bleibt jedoch (außer bei Kindern aus den GUS-Staaten) signifikant (Mehringer, 2013). Nach Kontrolle der Lesekompetenz zeigte sich in den IGLU 2011 Daten kein signifikanter Effekt des Migrationshintergrunds (Stubbe et al., 2012).

Betrachtet man die regionale Herkunft der Schüler(innen) mit Migrationshintergrund, erhalten v. a. Schüler(innen) mit türkischen Wurzeln am häufigsten eine Empfehlung für die Haupt-/Mittelschule und am seltensten für das Gymnasium (Mehringer, 2013). Unter Berücksichtigung des kulturellen Kapitals (z. B. Anzahl der Bücher im Haushalt) lassen sich zwar für Kinder mit Wurzeln in den ehemaligen GUS-Staaten oder Osteuropa keine signifikanten Unterschiede mehr in den Übergangsempfehlungen finden, für Schüler(innen) türkischer Herkunft bleiben die Differenzen in den Lehrkräftempfehlungen jedoch bestehen (Mehringer, 2013). Der sozioökonomische Hintergrund der Familie (berechnet durch den *HISEI*, *highest international socio-economic index of occupational status*) erklärt jedoch 20 % der Varianz bei der Vergabe der Übertrittsempfehlung (Mehringer, 2013).

Entsprechend der Befunde zu den Übergangsempfehlungen zeigen sich auch in der Verteilung auf die Schulformen der Sekundarstufe I ethnizitäts- und sozialschichtspezifische Unterschiede. Schüler(innen) mit Migrationshintergrund sowie Schüler(innen) mit niedrigem sozioökonomischem Status besuchen seltener das Gymnasium und sind dafür an Haupt-/Mittelschulen deutlich häufiger vertreten (z. B. Müller & Ehmke, 2013; Ramm et al., 2004). Dabei sind insbesondere Schüler(innen) mit türkischem Migrationshintergrund überproportional häufig an Haupt-/Mittelschulen zu finden, auch im Vergleich zu Schüler(inne)n mit einem anderen Migrationshintergrund (u. a. Diefenbach, 2010).

Merkens und Wessel (2002) konnten zeigen, dass Schüler(innen) mit einkommensstarken Eltern sowie einem hohem kulturellen Kapital (nach Bourdieu, 1982; vgl. auch Merkens & Wessel, 2002) deutlich häufiger das Gymnasium besuchen als Kinder mit geringem kulturellem Kapital und aus einkommensschwachen Elternhäusern. Herkunftsassoziierte Disparitäten in der Bildungsbeteiligung zeigen sich somit über den gesamten schulischen Bildungsverlauf, begonnen bei der vorschulischen Bildung, bis hin zum Studium. Auch hier ist der Anteil an Studierenden aus sozial schwachen Familien geringer, zudem müssen diese Studierenden auch häufiger als finanziell besser gestellte Studierende durch studienbegleitende Erwerbstätigkeit ihr Studium finanzieren, wodurch zusätzlich zum Studium eine höhere Belastung entsteht (Bargel & Bargel, 2010).

Aber auch der Anteil an Kindern und Jugendlichen in besonderen Förderprogrammen weist große Unterschiede zwischen Lernenden unterschiedlicher Herkunft auf. Auch wenn Schüler(innen) mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status ebenso Spitzenleistungen erbringen (z.B. Stamm, 2009), sind sie in Begabtenförderprogrammen deutlich unterrepräsentiert (z.B. Imhasly, 2004; Stamm, 2007, 2009). Nicholson-Crotty, Grissom und Nicholson-Crotty (2011) zeigten für den US-amerikanischen Raum, dass afroamerikanische Schüler(innen) eine 54 % geringere Wahrscheinlichkeit haben für ein Begabtenprogramm vorgeschlagen zu werden. Jedoch stieg die Wahrscheinlichkeit, vorgeschlagen zu werden, wenn ihre Lehrkraft ebenfalls afroamerikanischer Herkunft war, was auf den bedeutsamen Einfluss der Lehrkraft hindeutet. Dabei ist anzunehmen, dass die eigene Ethnizität und damit auch ethnizitätsspezifische Wahrnehmungen und Einstellungen der Lehrkräfte von Bedeutung sind, die im weiteren Verlauf der Arbeit vertieft werden (Kapitel 4). Somit zeigt sich auch für Schüler(innen) des oberen Leistungsspektrums eine erheblich höhere Hürde, wenn sie einer ethnischen Minorität angehören oder einen niedrigen sozioökonomischen Status aufweisen.

2.2.3 Bildungserfolg nach ethnischer und sozialer Herkunft von Schüler(inne)n

Wie sich aufgrund der Befunde zu Disparitäten in den schulischen Leistungen sowie der Bildungsbeteiligung erwarten lässt, verweisen empirische Arbeiten ebenfalls auf entsprechende Differenzen im Bildungserfolg von Schüler(inne)n unterschiedlicher ethnischer und sozialer Herkunft. In den Daten des Mikrozensus 2014 (vgl. Brückner, 2016; Statistisches Bundesamt, 2015) zeigte sich zwar, dass sowohl ca. 30 % der Personen mit als auch ohne Migrationshintergrund in Deutschland einen hohen Schulabschluss aufweisen. Einen mittleren Abschluss können dagegen nur noch 23 % der Personen mit und 36 % der Personen ohne Migrationshintergrund aufweisen. Ein relativ ausgeglichenes Bild zeigt sich für einen niedrigen Schulabschluss (28 % ohne Migrationshintergrund; 31 % mit Migrationshintergrund). Jedoch haben 12 % der in Deutschland lebenden Menschen mit Migrationshintergrund keinen Schulabschluss, wobei dies bei Personen ohne Migrationshintergrund lediglich 2 % betrifft.

Dabei sind starke Schwankungen je nach Herkunftsland, Aufenthaltsdauer und -status zu attestieren (Brückner, 2016). So haben nur 4 % der Spätaussiedler(innen) und ihrer Nachkommen und 21 % der Personen mit Wurzeln in Gastarbeiterländern keinen Schulabschluss. 55 % der Personen mit Wurzeln in EU-Mitgliedsstaaten haben dagegen einen hohen

Bildungsabschluss. Auch durch den höheren Anteil an Kindern mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status an Förderschulen sowie der hohen Quote an Schüler(inne)n, die die Förderschule ohne Abschluss verlassen (Hasselhorn et al., 2014), sind ethnizitäts- und sozialspezifische Häufungen unter Schüler(inne)n ohne Schulabschluss zu erwarten. An US-amerikanischen Schulen ist ebenfalls der Anteil an Schüler(inne)n, welche die Schule ohne Abschluss verlassen, unter lateinamerikanischen Schüler(innen) im Vergleich zur europäischstämmigen Mehrheit viermal so hoch und afroamerikanische Schüler(innen) werden dreimal so häufig suspendiert oder von der Schule verwiesen wie ihre europäischstämmigen Mitschüler(innen) (APA, 2012; Aud et al., 2010; Fenning & Rose, 2007; Gregory & Weinstein, 2008). Dabei kann dies zu einer sozialen Marginalisierung (Solga, 2003) und damit zur Reproduktion sozialer Ungleichheiten führen, die an Merkmale der ethnischen und sozialen Herkunft von Personen geknüpft ist.

Kristen und Granato (2007) differenzieren in ihren Analysen der Mikrozensusdaten deutlich stärker nach unterschiedlichen Herkunftsländern von Schüler(inne)n der zweiten Generation. Dabei zeigt sich v. a. bei türkisch- und italienischstämmigen Personen ein deutlich geringerer Anteil an Abiturient(inn)en als z. B. bei Personen mit griechischem Migrationshintergrund oder keinem Migrationshintergrund. Auch ein akademischer Abschluss ist bei Personen mit türkischen Wurzeln mit 8 % im Vergleich zu 24 % bei Personen ohne Migrationshintergrund, aber auch im Vergleich zu anderen Migrantengruppen (15-30 %) am geringsten (Tucci, 2016). Über den zeitlichen Verlauf (1991-2004) ließ sich zwar in allen ethnischen Gruppen ein Anstieg der Abiturrate verzeichnen, bei Schüler(inne)n türkischer und italienischer Herkunft fiel dieser Anstieg jedoch deutlich schwächer aus als bei Schüler(inne)n ohne Migrationshintergrund. Nach Kontrolle des sozioökonomischen Status waren jedoch keine differenziellen Unterschiede zwischen Schüler(inne)n mit türkischem und ohne türkischen Migrationshintergrund, in der Chance das Abitur zu erreichen, zu beobachten. Lediglich für Schüler(innen) italienscher Herkunft blieb auch unter Kontrolle des sozioökonomischen Status der Eltern eine geringere Wahrscheinlichkeit, das Abitur abzulegen. Als Konsequenz der geringeren beruflichen Abschlüsse sind auch entsprechende Unterschiede in der beruflichen Beschäftigung zu finden. Dabei zeigt sich, dass türkischstämmige Migrant(inn)en eine geringere Beschäftigungsrate aufweisen als Personen ohne Migrationshintergrund, die sich jedoch überwiegend durch individuelle Merkmale (z. B. Alter) erklären lässt (Euwals, Dagevos, Gijsberts & Roodenburg, 2007). Dennoch weisen die Berufe der türkischstämmigen Personen, entsprechend den geringeren Schulabschlüssen, ein geringes berufliches Prestige auf (Euwals et al., 2007).

Resümierend zeigt die Analyse der schulischen Leistungen, der Bildungsbeteiligung sowie des Bildungserfolgs in allen drei Bereichen und Unterkategorien deutliche Differenzen zwischen Schüler(inne)n mit und ohne Migrationshintergrund, wobei v. a. türkischstämmige Schüler(innen) deutlich schlechter abschneiden als ihre Mitschüler(innen). Der unterschiedliche Bildungsverlauf und -erfolg von Schüler(inne)n verschiedener ethnischer und sozialer Herkunft ist kein spezifisch deutsches Phänomen, sondern ebenfalls in zahlreichen anderen Staaten auffindbar. So erreichen in den USA Schüler(innen) mit afro-, lateinamerikanischen oder indigenen Wurzeln deutlich geringere schulische Leistungen, verlassen die Schule häufiger ohne Abschluss, besuchen seltener Hochschulen und sind ebenfalls in Programmen

zur Förderung besonders begabter Schüler(innen) unterrepräsentiert (u. a. Aud et al., 2010). Auch in Luxemburg zeigt sich, dass Schüler(innen) mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status trotz gleicher Leistungen eine höhere Wahrscheinlichkeit haben eine Klassenstufe zu wiederholen (Klapproth & Schaltz, 2015).

Nicht nur im Hinblick auf die ethnische Zugehörigkeit zeigen sich Unterschiede im Bildungsverlauf und -erfolg, sondern auch im Vergleich der sozialen Herkunft der Schüler(innen) werden große Diskrepanzen deutlich. Schüler(innen) mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status erreichen deutlich seltener Bildungserfolge und beginnen ihren Bildungsverlauf bereits vor Schuleintritt mit oftmals schlechteren Eingangsvoraussetzungen. Diese allein können die teilweise sehr großen Leistungsunterschiede jedoch nicht erklären. Als weitere Einflussfaktoren zur Erklärung der bestehenden herkunftsassoziierten Disparitäten im deutschen Bildungssystem werden dabei eine Vielzahl unterschiedlicher Ursachen diskutiert (z. B. Allemann-Ghionda, 2006; Billmann-Mahecha & Tiedemann, 2006; Diefenbach, 2008, 2010; Seitz, 2006), die im Folgenden näher betrachtet werden.

2.3 Erklärungsansätze für ethnische und soziale Disparitäten im Bildungssystem

Als Erklärungsansätze für schulische Leistungen und Bildungserfolg lassen sich zahlreiche, unterschiedliche Einflussfaktoren differenzieren, wobei diese Faktoren bei bestimmten Schülergruppen z. T. kumulativ auftreten können. Stubbe und Bos (2008) sprechen in diesem Zusammenhang von einer mehrfachen Benachteiligung sozial schwacher Kinder, da sie geringere schulische Leistungen erreichen, bei gleichen Leistungen trotzdem schlechter bewertet werden, ihre Eltern eine geringere Bildungsaspiration aufweisen und somit bedingt durch vielerlei Gründe eine geringere Wahrscheinlichkeit für einen Gymnasialbesuch haben. Nach Helmkes (2009) *Angebots-Nutzungs-Modell der Wirkungsweise des Unterrichts* sowie dem *Makromodell der Bedingungsfaktoren schulischer Leistungen* lassen sich eine Vielzahl unterschiedlicher Einflussfaktoren auf das Lernen der Schüler(innen) sowie auf ihre Leistungen differenzieren. Dabei wird u. a. die Persönlichkeit der Lehrkraft, ihre Erwartungen und diagnostischen Kompetenzen als bedeutsame Faktoren thematisiert, aber auch soziokulturelle Merkmale der Familie (Helmke, 2009).

Nach Boudon (1974) lassen sich die Ursachen für herkunftsassoziierte Disparitäten unter Fokussierung des familiären Hintergrunds in primäre und sekundäre Herkunftseffekte untergliedern. Primäre Herkunftseffekte entstehen dabei durch den direkten Einfluss der familiären Bedingungen auf die kindliche Entwicklung, z. B. geringere Kompetenzen durch eine weniger anregungsreiche Lernumgebung in der Familie. Sekundäre Herkunftseffekte beschreiben hingegen Disparitäten in der Bildungsbeteiligung, die sich nicht durch unterschiedliche Leistungen erklären lassen, sondern z. B. durch herkunftsassoziierte Bildungsentscheidungen (Baumert et al., 2010; Dumont, Maaz, Neumann & Becker, 2014). In dem *Modell zur Identifikation primärer und sekundärer Herkunftseffekte* von Baumert et al. (2010) hat die soziale Herkunft eines Kindes dabei direkte Auswirkungen auf seine Leistungen, die sich in Noten niederschlagen und damit zu einer Empfehlung für eine weiterführende Schule führen, die dann in einem entsprechenden Übergang resultiert (vgl. Abbildung 3). Somit sind primäre

Herkunftseffekte z.B. unterschiedliche Leistungen, Effekte die direkt durch entsprechende familiäre Bedingungen entstehen.

Abbildung 3 stellt zudem die sekundären Herkunftseffekte dar. Dabei zeigt sich, dass sich die familiäre Herkunft der Lernenden auch bei gleichen Leistungen auf die Noten, aber auch auf die Übergangsempfehlung sowie die tatsächliche Entscheidung für eine weiterführende Schule auswirken kann. Dabei konnten sowohl für primäre als auch sekundäre Herkunftseffekte zahlreiche empirische Evidenzen für Deutschland erbracht werden (Überblick bei Dumont et al., 2014). Hierbei stellt sich die Frage, auf welche Weise der ethnische und/oder soziale Hintergrund der Familie überhaupt Einfluss auf die schulischen Leistungen der Kinder und Jugendlichen haben sollte.

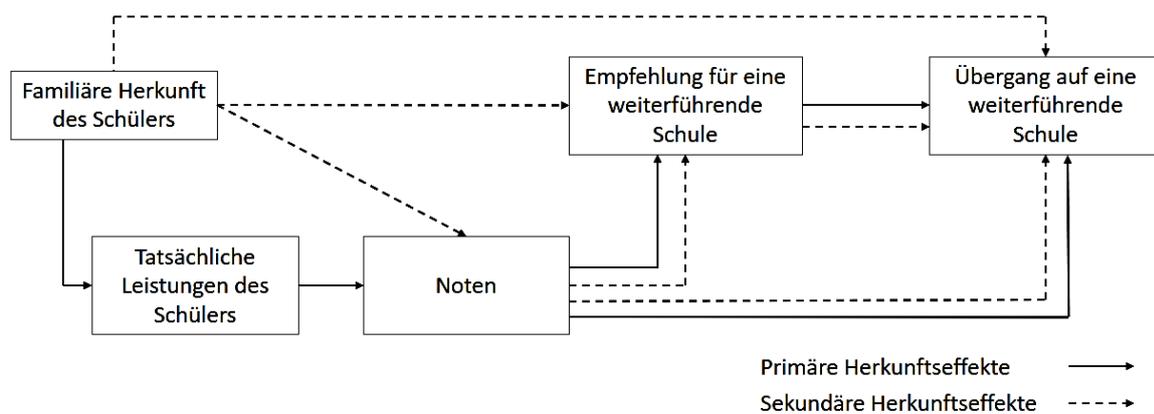


Abbildung 3. Primäre und sekundäre Herkunftseffekte auf den Übergang auf eine weiterführende Schule (nach Baumert et al., 2010, S. 8)

In diesem Zusammenhang verweisen theoretische Modelle zur Erklärung schulischer Leistungen oder leistungsrelevanter Entscheidungen wie z.B. das *Erwartungs-Wert-Modell* (Eccles, 1994; Wigfield & Eccles, 2000) bereits auf den Einfluss des familiären Hintergrunds. So können sich spezifische ethnisch-kulturelle Einstellungen und Erwartungen der Sozialisationsinstanzen u. a. auf die Interpretation vorheriger schulischer Leistungen und damit zusammenhängend auch auf die Kausalattribution der Schüler(innen) bei Erfolg oder Misserfolg auswirken. Die entsprechende Kausalattribution kann wiederum die affektiven Leistungserinnerungen beeinflussen und damit auch den Wert schulischer Aufgaben. Hierzu können empirische Befunde zunächst unabhängig von der soziokulturellen Herkunft den Einfluss elterlicher Einstellungen und Verhaltensweisen auf kindliche Überzeugungen nachweisen. Grolnick, Ryan und Deci (1991) zeigten z.B., dass leistungsrelevante kindliche Faktoren (z.B. die wahrgenommene Kompetenz) von der elterlichen Unterstützung der Autonomie sowie der elterlichen Einbindung in schulisch relevante Aspekte (z. B. Interesse und Zeit) abhängen. Dabei werden die elterlichen Einstellungen über spezifisches Erziehungsverhalten wirksam und beeinflussen unterschiedliche Aspekte der kindlichen Entwicklung (z.B. Borelli et al., 2010; Sigel, 1994). Die Vermittlung kultur- und ethnizitätsgeprägter elterlicher Einstellungen zeigt sich auch in herkunftsspezifischem

Erziehungsverhalten und beeinflusst hierdurch ebenfalls die Entwicklung kindlicher Einstellungen (Jäkel & Leyendecker, 2009; Liebwein, 2008).

Im Sinne des Erwartungs-Wert-Modells (Eccles, 1994; Wigfield & Eccles, 2000) sind zunächst die Eltern als Sozialisationsinstanzen zu verstehen, deren Herkunft Einfluss auf schulisch relevante Einstellungen haben können. Auch bei Boudon (1974) wird zunächst der Einfluss der familiären Entscheidungen im Hinblick auf sekundäre Disparitäten thematisiert (vgl. auch Dumont et al., 2014). Jedoch sind auch Lehrkräfte als Sozialisationsinstanzen zu sehen, deren spezifischer soziokultureller Hintergrund mit spezifischen Einstellungen verknüpft ist, die sich ebenfalls auf die kindliche Wahrnehmung und Interpretation auswirken können und damit einen Einfluss auf die schulischen Leistungen und den Bildungsweg der Schüler(innen) haben können. Zudem können auch herkunftsassoziierte Einstellungen und Erwartungen der Lehrkräfte einen Einfluss auf Bildungsentscheidungen im Sinne sekundärer Disparitäten haben. Lehrkräfte sind dabei jedoch nicht nur durch ihren Einfluss als Sozialisationsinstanzen relevant, sondern in erster Linie durch ihre Rolle als Wissensvermittler, Lernbegleiter und Leistungsbewertende.

Hierbei wird bereits deutlich, dass Einflussfaktoren auf den Schulerfolg im Allgemeinen, aber auch unter der Berücksichtigung einer ethnisch sowie sozial heterogenen Schülerschaft, auf unterschiedlichen Ebenen zu lokalisieren sind und dennoch in komplexen Wirkmechanismen zusammenhängen. Während Diefenbach (2010) die schulische bzw. institutionelle Ebene von der individuellen Ebene – wobei hierunter neben Schülermerkmalen auch Merkmale des Elternhauses zu fassen sind – unterscheidet, trennt Rüesch (1998) die individuelle Ebene in zwei separate Bereiche auf und unterscheidet zwischen individuellen und familiären Einflussfaktoren. Zudem werden neben schulischen Merkmalen auch gesellschaftliche Rahmenbedingungen betrachtet (Rüesch, 1998).

Dabei ist davon auszugehen, dass individuelle Schülermerkmale und familiäre Bedingungen zwar eng zusammenhängen und sich gegenseitig beeinflussen können, eine systematische Trennung individueller Schülermerkmale und familiärer Faktoren jedoch durchaus sinnvoll erscheint. Auch sollte die institutionelle Ebene stärker differenziert werden, da sicherlich grundlegende gesellschaftliche Rahmenbedingungen sowie schulische bzw. konkrete institutionelle Merkmale des Bildungssystems eine Rolle spielen. Dennoch wird in beiden Gliederungs- bzw. Differenzierungsversuchen der individuellen Bedeutung der Lehrkraftmerkmale nicht ausreichend Rechnung getragen. Zwar werden sie bei Diefenbach (2010) unter „Kontextbedingungen des Schulerfolgs“ (S. 90) subsumiert, dennoch erscheinen insbesondere aus pädagogisch-psychologischer sowie sozialpsychologischer Perspektive die individuellen Lehrkraftmerkmale als sehr bedeutsam, weshalb ihnen eine stärkere Positionierung in entsprechenden Modellen gerechter werden würde. Aus diesem Grund wird in dieser Arbeit eine zusätzliche Spezifizierung für lehrkraftspezifische Einflussfaktoren vorgenommen, die nicht nur durch schulische oder gesellschaftliche Rahmenbedingungen beeinflusst werden, sondern auch die individuellen Merkmale der Lehrperson selbst berücksichtigt. Wie Abbildung 4 zu entnehmen ist, lassen sich die Einflussfaktoren auf den Bildungsverlauf und -erfolg von Kindern unterschiedlicher ethnischer und sozialer Herkunft in unterschiedliche Bereiche gliedern, die jedoch bedingt durch die gesellschaftlichen und systemischen Zusammenhänge interdependent verbunden sind.

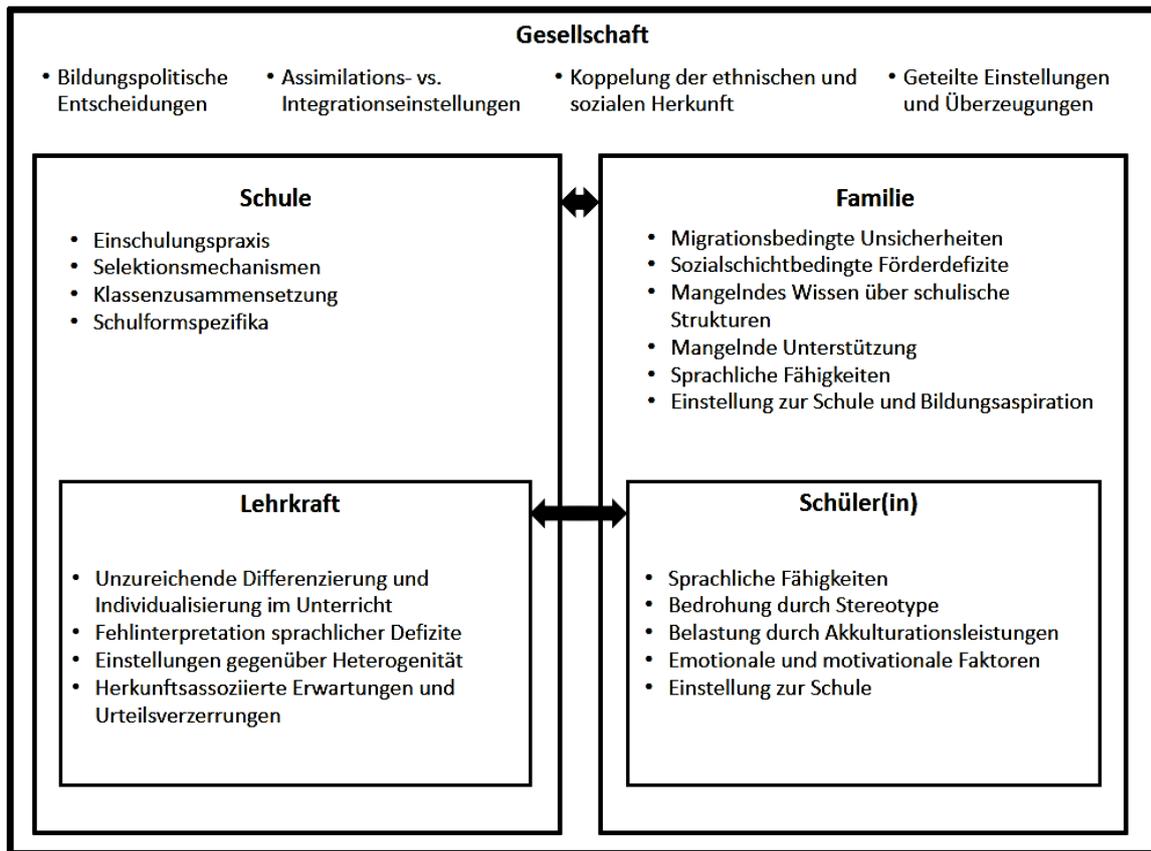


Abbildung 4. Überblick über zentrale Einflussfaktoren herkunftsspezifischer Disparitäten im Bildungsverlauf

Zu berücksichtigen ist, dass wesentliche Einflussfaktoren, die strenggenommen eigenständige Subsysteme darstellen und damit auch eigenständige Einflussfaktoren auf den Bildungsverlauf haben können, hier in übergeordnete Systeme integriert wurden (vgl. Abbildung 4). So befindet sich das Subsystem der Schulklasse auf der schulischen Ebene. Aber auch im familiären Umfeld sind Einflüsse, z. B. der Nachbarschaft, subsumiert, die als soziales Umfeld der Familie ein eigenständiges und zudem sicherlich einflussreiches Subsystem darstellen. Da die vorliegende Arbeit v. a. die Bedeutung der Lehrkraft durch ihre Urteils- und Erwartungsbildung fokussiert, wird im Folgenden lediglich ein kurzer Überblick über die weiteren Erklärungsansätze für herkunftsassoziierte Bildungsdisparitäten auf anderen Ebenen gegeben.

Durch bildungspolitische Entscheidungen, gesellschaftlich geteilte Überzeugungen und die Konfundierung der ethnischen und sozialen Herkunft werden auf *gesellschaftlicher Ebene* die grundlegenden Strukturen für herkunftsassoziierte Disparitäten gebildet (u. a. Neumann, 2009). Die soziale und rechtliche Benachteiligung bestimmter Migrantengruppen, aber auch die mangelnde Verständigung zwischen den Kulturen und damit eines wertschätzenden Umgangs mit Minoritätskulturen (Seitz, 2006), Diskriminierungsprozesse (z. B. Auernheimer, 2001; Cortés Núñez & Küçük, 2016; Diehl & Fick, 2016; Gomolla & Radtke, 2007), sowie unterschiedliche (integrative vs. separierende) strukturelle Ausrichtungen, werden als gesellschaftliche Ursachen für Schwierigkeiten und Probleme der erfolgreichen Integration,

und damit eines erfolgreichen Durchlaufens des deutschen Bildungssystems diskutiert (z. B. Fincke, 2009).

Gesellschaftliche Benachteiligung verläuft dabei häufig nicht subtil und versteckt ab, v. a. Kinder mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status berichten bereits im Grundschulalter von empfundener bzw. selbst erlebter Benachteiligung (Andresen, Fegter & Hurrelmann, 2013). Die gesellschaftliche Ausrichtung auf eine eher assimilationsorientierte Eingliederung kann ebenfalls eine große Herausforderung darstellen. Zander und Hannover (2013) konnten zeigen, dass ein negativer Zusammenhang zwischen der Identifikation mit der Herkunftskultur von Schüler(inne)n und der Identifikation mit der Aufnahmekultur besteht. Dabei ist davon auszugehen, dass es für Jugendliche mit Migrationshintergrund in Deutschland eine Herausforderung darstellt, sich mit beiden Kulturen zu identifizieren, was u. a. zu einem Assimilationsdruck führen kann (Zander & Hannover, 2013). Aber auch die z. T. gesellschaftlich bedingte hohe Konfundierung der ethnischen und sozialen Herkunft (vgl. Abschnitt 2.1) bildet die Grundlage für reproduktive Mechanismen der Chancenungleichheit.

Neben Einflussfaktoren auf gesellschaftlicher Ebene spielt auch eine Vielzahl an *Bedingungen in den Familien* eine bedeutsame Rolle für Disparitäten im Bildungsverlauf und -erfolg von Kindern und Jugendlichen unterschiedlicher ethnischer und sozialer Herkunft. Baumert et al. (2003) gehen in ihrem Modell des Zusammenhangs zwischen Struktur- und Prozessmerkmalen der familiären Lebensverhältnisse und Bildungsbeteiligung bzw. Kompetenzerwerb davon aus, dass die familiären Strukturmerkmale (sozioökonomischer Status, Bildungsniveau und Migrationsstatus) Einfluss auf die Wohlstandsgüter der Familie, aber auch auf die kulturelle, kommunikative und soziale Praxis innerhalb der Familie haben und darüber auf die Bildungsbeteiligung und den Kompetenzerwerb wirken, was im Sinne Boudons (1974) als primärer Herkunftseffekt zu verstehen wäre. Unterschiede in den kognitiven Leistungen zwischen der ethnischen Majorität und Minoritäten lassen sich zu einem hohen Anteil durch die sozioökonomische Lage der Familie und das mütterliche Erziehungsverhalten erklären (Brooks-Gunn, Klebanov & Duncan, 1996).

Neben den Effekten der Strukturmerkmale wird in dem Modell von Baumert et al. (2003) ebenfalls davon ausgegangen, dass der Migrationshintergrund und der sozioökonomische Status zusätzlich einen direkten Effekt auf die Bildungsbeteiligung und den Kompetenzerwerb haben. Ergänzend gehen Baumert et al. (2003) darauf ein, dass Struktur- und Prozessmerkmale der Familie ebenfalls Einfluss auf psychologische Merkmale wie die Fähigkeiten und Motivation haben können, die wiederum die Kompetenzen und Bildungsbeteiligung direkt beeinflussen. Je nach sozioökonomischem Status werden Bildungsentscheidungen der Eltern zudem durch unterschiedliche Kosten-Nutzen-Rechnungen getroffen. Der Statuserhalt in sozial niedrigeren Schichten muss z. B. nicht durch einen Gymnasialbesuch erreicht werden (vgl. Baumert et al., 2003). Dies wäre nach Boudon (1974) unter den sekundären Herkunftseffekten zu fassen. Dabei gehen Baumert et al. (2003) davon aus, dass Ungleichheiten der Übergangsentscheidungen sowohl durch primäre als auch sekundäre Herkunftseffekte entstehen. Leistungsunterschiede, die durch primäre Herkunftseffekte bedingt sind, haben so z. B. auch Auswirkungen auf die Übergangsentscheidung, die wiederum durch sekundäre Effekte verstärkt werden können. Die genauen (interdependenten) Wirkmechanismen der

einzelnen Faktoren, sowie die jeweilige Einflussgröße sind empirisch jedoch noch nicht vollständig geklärt.

Angenommen wird zudem, dass ein Mangel an ökonomischen Mitteln zu geringeren Ressourcen in unterschiedlichsten Bereichen führen kann, was möglicherweise auch eine weniger entwicklungsfördernde Umwelt bedingt. Dabei ist davon auszugehen, dass diese Gefahr in von Armut betroffenen Familien in zweierlei Hinsicht deutlich höher ist. Durch geringere ökonomische Ressourcen könnte die Ausstattung an z. B. Büchern oder lernanregenden Materialien geringer sein, aber auch die direkte Nachbarschaft ist oftmals durch einen geringen sozioökonomischen Status und damit einhergehend durch zahlreiche soziale Probleme geprägt (u. a. Brooks-Gunn & Duncan, 1997). Hierdurch erleben insbesondere Personen mit niedrigem sozioökonomischem Status Einschränkungen, die ihnen eine optimale Förderung des Bildungsverlaufs ihrer Kinder erschweren (Kristen, 2003). In Familien mit niedrigem sozioökonomischem Status kommt es auch seltener zu einer vielseitigen Gestaltung der kindlichen Freizeit oder auch Mitgliedschaft in Vereinen, wie z. B. Musik- oder Sportgruppen (Jänsch & Schneekloth, 2013). Eltern mit höherem sozioökonomischem Status geben z. B. das Lesen deutlich häufiger als wichtige häusliche Aktivität an als Eltern mit niedrigem sozioökonomischem Status (Schwippert et al., 2003; Stubbe, Buddenberg, Hornberg & McElvany, 2007). Kinder aus sozial schwächeren Schichten haben ebenfalls eine größere Gefahr bei Eltern aufzuwachsen, deren physischer und psychischer Gesundheitszustand beeinträchtigt ist, was ebenfalls zu einer geringeren Förderung der kindlichen Entwicklung führen kann (Brooks-Gunn & Duncan, 1997).

Zudem zeigen sich Unterschiede hinsichtlich der Bildungsaspiration von Eltern unterschiedlicher Herkunft. Eltern mit eigenem hohem Schulabschluss, mit hohem kulturellem Kapital sowie hohem sozioökonomischem Status haben durchschnittlich eine höhere Bildungsaspiration (Merkens & Wessel, 2002). Während sich Eltern mit hohem sozioökonomischem Status mit über 80 % einen Gymnasialabschluss für ihr Kind wünschen, sind es bei Eltern mit niedrigem sozioökonomischem Status etwas weniger als halb so viele, die einen Gymnasialwunsch äußern, wobei sich dieses Muster auch in den Bildungsaspirationen der Kinder wieder spiegelt (Merkens & Wessel, 2002). Eltern mit geringerem sozioökonomischem Status versprechen sich z. B. auch einen geringeren Vorteil von einem höheren Bildungsabschluss, als Eltern mit hohem sozioökonomischem Status (Erikson & Jonsson, 1996). Die hohe Bildungsaspiration v. a. in Familien mit hohem sozioökonomischem Status wird auch häufiger umgesetzt. Ein gewünschter Gymnasialbesuch wurde bei 55 % der Kinder aus niedrigen und bei 84 % aus hohen Sozialschichten umgesetzt (Merkens & Wessel, 2002). Bei beiden im Ausland geborenen Elternteilen zeigt sich hingegen eine besonders hohe Bildungsaspiration für ihre Kinder (Ditton, Krüsken & Schauenberg, 2005). Für diese soziale Gruppe ist dies somit kein relevanter Einflussfaktor, der die Disparitäten im Bildungsverlauf und -erfolg erklären kann.

Angenommen wird ebenfalls, dass Eltern mit Migrationshintergrund über ein geringeres Wissen über das deutsche Bildungssystem und weniger Zugang zu Informationen hierüber verfügen (Kristen, 2003). Kristen (2005) konnte zeigen, dass bei türkischstämmigen Migrant(inn)en mit zunehmenden sprachlichen Fähigkeiten auch die Wahrscheinlichkeit für den Zugang zu grundlegenden und relevanten Informationen steigt. Sprachliche Defizite der Schüler(innen) und Familien werden dabei in vielerlei Hinsicht als ein wesentlicher Einflussfaktor

diskutiert (z. B. Stanat, 2006a). In Familien aus niedrigen sozialen Schichten wird zudem eher in der Herkunftssprache gesprochen (Andresen et al., 2013), was erneut zu einem kumulierten Risiko für den Bildungsverlauf dieser Schülergruppe führen kann. Dennoch gibt es auch einige Kinder aus spezifischen Herkunftsländern, die das deutsche Schulsystem trotz ihres Migrationshintergrunds erfolgreich durchlaufen. Diskutiert wird in diesem Zusammenhang, dass womöglich spezielle ethnisch-kulturelle Faktoren (z. B. Selbstorganisation) spezifischer ethnischer Gruppen (z. B. mit griechischen Wurzeln) zum Ausbau besserer Unterstützungssysteme führen könnten oder aber auch, dass diese erfolgreichen Migrantengruppen stärkere Assimilationstendenzen aufzeigen (Stanat, 2006a). Jedoch bedarf es weiterer empirischer Studien, die die jeweiligen ethnizitäts- und kulturspezifischen Bedingungen berücksichtigen.

Familiäre Einflussfaktoren haben immer auch Auswirkungen auf *individuelle Merkmale der Kinder*, die entsprechend für Disparitäten im Bildungsverlauf verantwortlich sein können. Mehringer (2013) konnte in den Daten einer Grundschulstudie zeigen, dass 80 % der Kinder mit Migrationshintergrund (im Vergleich zu 30 % der Kinder ohne Migrationshintergrund) nur über unzureichende Sprachkenntnisse vor Schuleintritt verfügen. Dabei betrifft dies in der Gruppe der Schüler(innen) mit türkischem Migrationshintergrund sogar 90 % der Kinder (Mehringer, 2013). Problematisch ist in diesem Zusammenhang, dass die geringere Bildungsbeteiligung in vorschulischen Einrichtungen eine sprachliche Förderung vor Schuleintritt kaum ermöglicht. Befunde des Bildungsberichts 2016 verweisen darauf, dass 63 % der vier- und fünfjährigen Kinder mit Migrationshintergrund zu Hause überwiegend eine andere Sprache als Deutsch sprechen (Maaz et al., 2016). Anzumerken ist allerdings, dass Bilingualität nicht per se als negativer Einflussfaktor identifiziert werden kann, da einige empirische Befunde auf gleiche oder teilweise sogar höhere kognitive Fähigkeiten hindeuten (z. B. Bialystok, Craik & Luk, 2012). Sprachliche Schwierigkeiten zum Schuleintritt sind jedoch nicht auf die Gruppe der Kinder mit Migrationshintergrund beschränkt, sondern treten auch im Zusammenhang mit einer niedrigen sozialen Schicht auf (Schöler et al., 2005). Kinder, deren Familien auf Sozialleistungen angewiesen sind, zeigen in der Schuleingangsuntersuchung ebenfalls geringere Deutschfähigkeiten, aber auch größere Probleme in der Körperkoordination, der Visuomotorik, der selektiven Aufmerksamkeit sowie beim Zählen (Groos & Jehles, 2015).

Auch in einigen emotionalen Bereichen zeigen sich negative Ausprägungen bei Kindern mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status. Zwar konnten in einer Studie mit Grundschüler(inne)n positive motivationale Ausrichtungen beobachtet werden, jedoch haben Schüler(innen) mit Migrationshintergrund und/oder niedrigem sozialem Status auch nach Kontrolle der Vorleistungen signifikant höhere Ausprägungen von Leistungsangst und Hilflosigkeitserleben (Tobisch, Kopp, Martschinke, Kröner & Dresel, 2016). Diese stärker ausgeprägten negativen Emotionen können, z. T. vermittelt über das Selbstkonzept, Auswirkungen auf schulische Leistungen haben (Faber, 2007; König, 2009). Als weiterer Einflussfaktor ist zudem die Bedrohung durch Stereotype (*stereotype threat*) zu erwähnen, wobei Mitglieder stigmatisierter sozialer Kategorien durch die Aktivierung des entsprechenden Stereotyps in negativ stereotypisierten Domänen eine Bedrohung wahrnehmen, die sich u. a. auf Arbeitsgedächtniskapazitäten, emotionale und motivationale Aspekte und darüber vermittelt auf tatsächliche Leistungen auswirken kann (z. B. Martiny, Götz & Keller, 2013; Steele, 1997; Steele & Aronson, 1995). Jedoch sind stereotype threat Effekte für den deutsch-

sprachigen Raum und die hier lebenden Migrantengruppen und stereotypisierten Domänen noch nicht ausreichend erforscht.

Neben gesellschaftlichen, familiären und individuellen Merkmalen der Schüler(innen) sind ebenfalls *Einflussfaktoren auf institutioneller Ebene* auszumachen, die eine Chancengleichheit im deutschen Bildungssystem erschweren können. Dabei ist zunächst die strukturelle Diskriminierung zu nennen, „die aus der Beschaffenheit der vorhandenen Institutionen bzw. ihrer Funktionsweise selbst resultieren und daher gewöhnlich unbeabsichtigt und unbewußt [sic!] erfolgen“ (Diefenbach, 2010, S. 89). Da Schulen nicht nur die Aufgabe der Wissensvermittlung tragen, sondern auch eine Allokations- und Selektionsfunktion erfüllen müssen (u. a. Stanat, 2006a), sind die grundlegenden Strukturen bereits auf eine statuszuweisende und -generierende Funktion ausgelegt. Die Schule als Institution birgt somit bereits strukturelle Merkmale in sich, die einen ungleichen Bildungsverlauf unterstützen.

Aber auch auf Klassenebene können Erklärungsansätze identifiziert werden, so z. B. durch den Einfluss der Zusammensetzung der Schülerschaft (z. B. Rüesch, 1998; Stanat, 2006b). So wirkt sich die Zusammensetzung der Schülerschaft einer Klasse in Bezug auf ihre soziale Herkunft auf den Lernerfolg aus (Ditton & Krüskens, 2006b). Kristen (2002) konnte zeigen, dass auch bei einer Zunahme des Migrantenteils in der Klasse die Wahrscheinlichkeit für einen Übertritt auf die Realschule oder das Gymnasium für Kinder mit Migrationshintergrund abnahm. Stanat (2006b) konnte zwar ebenfalls ein geringeres Leistungsniveau in Klassen mit einem hohen Anteil an Schüler(inne)n mit Migrationshintergrund zeigen, jedoch brachten detailliertere Analysen ebenfalls eine hohe Konfundierung zahlreicher Benachteiligungsfaktoren zu Tage (z. B. geringerer sozioökonomischer Status, schlechtere Eingangsvoraussetzungen). Dabei ist zudem kritisch zu sehen, dass es einige Schulen in Deutschland gibt, die einen besonders hohen Anteil an Schüler(inne)n mit Migrationshintergrund haben und es hier womöglich zu einer Kumulierung kritischer Einflussgrößen kommt. Diese „ethnische Schulsegregation“ (Kristen, 2003, S. 32) führt dabei zu einer geringeren Wahrscheinlichkeit für einzelne Schüler(innen), nach dem Übertritt auf eine höhere weiterführende Schulform zu wechseln. Anzunehmen ist, dass multikausale Erklärungen notwendig sind, um herkunftsassoziierte Disparitäten zu analysieren, jedoch bedarf es hierbei weiterer vertiefter Untersuchungen. Neben gesellschaftlichen Strukturen, familiären und schulischen Einflüssen sowie den Schüler(inne)n selbst spielt insbesondere die Lehrkraft eine entscheidende Rolle, da sie maßgeblich den Lernverlauf der Schüler(innen) durch u. a. die Aufgabenstellung oder die Lernstandsdiagnostik beeinflusst.

Die *Bedeutung der Lehrkraft* wird im Speziellen durch die Folgen der direkten Interaktion mit den Schüler(inne)n, des Einflusses ihrer interkulturellen Kompetenz, aber auch durch verzerrte Urteile und Erwartungen im Kontext heterogener Schülermerkmale diskutiert (z. B. Dee, 2005; Schofield & Alexander, 2012; Strasser, 2016; Tobisch & Dresel, 2016; Weir, 2016). Dabei wird Lehrkräften u. a. durch ihre Kompetenzen eine bedeutsame Rolle für das Lernen der Schüler zugeschrieben (Lipowsky, 2006). Lehrkräfte haben durch ihren täglichen Kontakt mit Schüler(inne)n einen großen Einfluss auf motivationale und emotionale Bedingungen der Schüler(inne)n, sie können ihre Lernprozesse aktiv gestalten und durch ihre Instruktionen, Feedback und individualisierte Aufgabenstellungen dafür sorgen, dass jeder Schüler und jede Schülerin sich entsprechend seiner bzw. ihrer individuellen Stärken entwickeln kann. Die

Bedeutung der Lehrkraft wird u. a. durch die großen Unterschiede von herkunftsassoziierten Disparitäten zwischen Schulklassen deutlich (Dresel, Martschinke, Kopp, Tobisch & Kröner, 2017). Dies zeigt, dass Lehrkräfte einen bedeutsamen Einfluss auf die Stärke der Leistungsdisparitäten haben. In der Meta-Analyse von Hattie (2015) wird dabei zudem deutlich, dass die Lehrperson u. a. durch die Beziehung zu ihren Schüler(inne)n, durch das Nicht-Etikettieren der Lernenden, aber auch durch Erwartungen, einen bedeutsamen Einfluss auf den Lernerfolg von Schüler(inne)n hat. Die Rolle der Lehrkraft für schulische Leistungen wird auch in Helmkes (2009) Modell der Bedingungsfaktoren schulischer Leistungen deutlich. Dabei wird davon ausgegangen, dass u. a. Persönlichkeit, Erwartungen und diagnostische Fähigkeiten der Lehrkraft sowohl Einfluss auf Prozessmerkmale des Unterrichts als auch auf kindliche Kompetenzen haben können.

Insbesondere eine hohe diagnostische Kompetenz von Lehrkräften kann dabei als Voraussetzung für einen adaptiven Unterricht angesehen werden, da sie die zentrale Grundlage für schulische Fördermaßnahmen darstellt. Jedoch zeigt sich oftmals, dass die hierfür erforderliche diagnostische Kompetenz von Lehrkräften nicht optimal ausgeprägt ist (z. B. Südkamp, Kaiser & Möller, 2012; vgl. Abschnitt 3.1) und zudem teilweise von leistungsirrelevanten Herkunftsmerkmalen beeinflusst wird (z. B. Dunkake, Kiechle, Klein & Rosar, 2012; vgl. Abschnitt 3.1.2). So gibt es Indizien dafür, dass bei einer übereinstimmenden ethnischen Herkunft von Lehrkräften und Schüler(inne)n, die Schüler(innen) bessere Leistungen erzielen (Dee, 2004). Dabei wird zudem Diskriminierung als mögliche Ursache für den geringeren Bildungserfolg von Schüler(inne)n mit Migrationshintergrund diskutiert (Cortés Núñez & Küçük, 2016; Gomolla & Radtke, 2007; Kristen & Granato, 2007), da sich häufig trotz der Kontrolle leistungsrelevanter Merkmale Disparitäten im Bildungsverlauf und -erfolg von Kindern unterschiedlicher Herkunft zeigen (z. B. Farkas, 2003). Trotz gleicher Leistungen, Erfahrungen oder anderer relevanter Faktoren werden somit unterschiedliche Ergebnisse basierend auf leistungsirrelevanten Gründen erreicht (Kristen, 2005). Auch wenn davon auszugehen ist, dass direkte negative Diskriminierung von Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozialem Status eher die Ausnahme bildet, sind auch hierzu Fälle bekannt. So ist aus den Daten der Antidiskriminierungsstelle des Bundes (2013) zu entnehmen, dass die häufigsten Beratungsanfragen ethnische Diskriminierung beinhalten.

Ditton (2016) vermutet dabei u. a., dass bedingt durch eine eher Mittel-/Oberschichtgeprägte Wertevorstellung der Schule und der Lehrkräfte, Kinder aus ähnlich gelagerten Schichten durch gleiche Werte und Verhaltensmuster Vorteile in der Wahrnehmung der Lehrkraft genießen. Dabei können auch schichtspezifische Sprachmuster ein Problem darstellen (Ditton, 2016). So können mangelnde sprachliche Fähigkeiten womöglich auch als mangelnde kognitive Fähigkeiten fehlinterpretiert werden. Entsprechend zeigt sich z. B., dass Lehrkräfte die Kompetenzen mehrsprachiger Schüler(innen) häufig unterschätzen (Fürstenau & Gomolla, 2012). Dies hängt auch eng mit allgemeinen Einstellungen der Lehrkraft gegenüber Heterogenität in ihren zahlreichen Facetten zusammen, sowie mit herkunftsspezifischen Assoziationen bestimmter Merkmale, die als Stereotype in der Urteilsbildung von Lehrkräften wirksam werden können (vgl. Kapitel 4 & Kapitel 5).

Da die Bildung von Urteilen über und Erwartungen an Schüler(innen) durch Lehrkräfte eine essentielle Bedeutung für die tägliche Unterrichtspraxis haben kann und maßgeblich für sekundäre Herkunftseffekte verantwortlich sein könnte, existieren hierzu bereits einige empirische Untersuchungen. Welche Rolle Lehrkrafturteile und -erwartungen im Kontext einer heterogenen Schülerschaft jedoch spielen, wurde im deutschsprachigen Raum bislang noch unzureichend untersucht. Dabei wurde vor allem die Konfundierung der ethnischen und sozialen Herkunft kaum berücksichtigt und meist ausschließlich Leistungsurteile und -erwartungen fokussiert. Theoretische Annahmen und bisherige empirische Befunde zu möglichen herkunftsassoziierten Unterschieden in Urteilen und Erwartungen von Lehrkräften bestätigen allerdings die besondere Bedeutung der diagnostischen Kompetenz, die im folgenden Kapitel neben Urteilen und Erwartungen näher beleuchtet wird.

3. Lehrkrafturteile und -erwartungen im Kontext einer heterogenen Schülerschaft

Wie bereits in Abschnitt 2.3 skizziert wurde, sind Lehrkräfte mit all ihren Überzeugungen, Einstellungen, Erwartungen und Kompetenzen maßgeblich für den Bildungsverlauf und -erfolg von Lernenden mitverantwortlich (z. B. Hattie, 2015; Seitz, 2006). Insbesondere die Beurteilung von Leistungen und leistungsrelevanten Schülermerkmalen stellt hierbei eine zentrale und bedeutsame Aufgabe des Lehrberufs dar (z. B. Helmke, 2009; Jürgens & Lissmann, 2015; Schrader & Helmke, 2001). So hat die Einschätzung der Leistungen und leistungsrelevanten Merkmale durch Grundschullehrkräfte einen beträchtlichen Einfluss auf ihre Übergangsempfehlung für weiterführende Schulen der Sekundarstufe I (z. B. Richert, 2012) und ist ebenfalls von besonderer Bedeutung, um im Unterricht individuelle Schülermerkmale berücksichtigen zu können.

Die fachspezifische Leistungsbeurteilung findet oftmals Ausdruck in Zensuren und wird in regelmäßigen Abständen in Form von Zeugnissen an die Schüler(innen) und ihre Eltern ausgegeben. Noten und Zeugnisse erfüllen u. a. die Funktion einer Rückmeldung über die Lern- und Leistungsentwicklung an die Eltern (Bartnitzky, 1999), beinhalten jedoch häufig auch weitere u. a. motivationale Komponenten oder dienen z. T. auch als lernförderliches Feedback für die Schüler(innen) (Bartnitzky, 1999; Weiss, 1965). Kinder im Grundschulalter erhalten in Zeugnissen neben Ziffernbewertungen in den einzelnen Schulfächern zudem eine verbale Beurteilung über ihr Sozial-, Lern- und Arbeitsverhalten (§ 15, Grundschulordnung [GrSO]; Bayerische Staatskanzlei, 2008a).⁶

Somit sind neben akkuraten Einschätzungen der fachspezifischen Leistungen auch die Beurteilungen weiterer Schülermerkmale durch die Lehrkraft erforderlich. Lehrkrafturteile zu weiteren Schülermerkmalen sind jedoch nicht nur für die formale Bewertung in Form von Zeugnissen bedeutsam, sondern sind v. a. im regulären Unterrichtsalltag relevant. Nur durch eine akkurate Einschätzung der Schüler(innen) ist es der Lehrkraft möglich, den Unterricht und schulische Aufgaben adäquat an die spezifischen Voraussetzungen der Schüler(innen) anzupassen. Die genaue Beurteilung ist somit eine Grundlage für individualisierten, adaptiven und differenzierten Unterricht. Dabei können unterschiedliche Variablen identifiziert werden, die einen bedeutsamen Einfluss auf das Lernen und die Leistungen von Schüler(inne)n haben und entsprechend akkurat von Lehrkräften wahrgenommen und beurteilt werden sollten. Einige zentrale Prädiktoren des Lern- und Leistungserfolgs von Schüler(inne)n werden im Folgenden in einem kurzen Überblick dargestellt.

Prädiktoren schulischen Lern- und Leistungserfolgs

Wigfield und Eccles (2000) gehen davon aus, dass die subjektiv eingeschätzte *Erwartung einen Erfolg zu erzielen* und der *subjektive Wert einer Handlung* die Motivation von

⁶ Anzumerken ist, dass eine Ziffernbewertung erst ab dem Jahreszeugnis der zweiten Jahrgangsstufe eingeführt wird, zuvor erhalten Schüler(innen) einen verbalen Bericht, in dem auch der Leistungsstand beschrieben wird (§ 15GrSO; Bayerische Staatskanzlei, 2008a).

Schüler(inne)n bedingen. Dabei kann der subjektive Wert einer Aufgabe oder einer spezifischen Domäne sich auch auf die Leistungen auswirken (Steinmayr & Spinath, 2009) und entsteht nach Wigfield und Eccles (2000) durch die individuell zugeschriebene Nützlichkeit oder Wichtigkeit einer Aufgabe. Bei Grundschüler(inne)n zeigt sich ein negativer Zusammenhang des subjektiven Werts im Fach Deutsch mit negativen Lernemotionen und ein positiver Zusammenhang mit einer Lernzielorientierung (Tobisch et al., 2016).⁷ Der subjektive Wert hängt entsprechend mit weiteren relevanten Einflussfaktoren auf den Lernerfolg von Schüler(inne)n zusammen. Nach dem Erwartungs-Wert-Modell (z.B. Eccles, 1994) ist der subjektive Wert zwar eine wichtige, jedoch nicht ausreichende Voraussetzung für die Motivation. Auch die Erfolgserwartung der Schüler(innen) sollte für eine optimale Lernmotivation möglichst hoch ausgeprägt sein.

Anzunehmen ist, dass die Erfolgserwartung u.a. von dem Selbstkonzept der Schüler(innen) abhängt. Das hierarchisch strukturierte Selbstkonzept lässt sich zunächst in ein schulisches Selbstkonzept und ein nicht akademisches Selbstkonzept unterteilen (Marsh & Shavelson, 1985). Das schulische Selbstkonzept, als „die Gesamtheit der kognitiven Repräsentationen eigener Fähigkeiten“ (Pekrun, 1983, S. 79) – entsprechend auch *Fähigkeitsselbstkonzept* genannt – ist für Lernprozesse und damit für den Lernerfolg von essentieller Bedeutung (z.B. Dickhäuser, 2006; Schöne, Dickhäuser, Spinath & Stiensmeier-Pelster 2003, 2012). Das Fähigkeitsselbstkonzept repräsentiert dabei die subjektive Einschätzung der Höhe eigener Fähigkeiten, in Abgrenzung zur Repräsentation der Veränderbarkeit eigener Fähigkeiten durch implizite Fähigkeitstheorien (Schloz & Dresel, 2011). Für Grundschüler(innen) zeigt sich ein negativer Zusammenhang des Fähigkeitsselbstkonzepts mit negativen Lernemotionen, wie dem Hilflosigkeitserleben sowie ein positiver Zusammenhang mit einem handlungsadaptiven Umgang mit Fehlern (z.B. Schloz & Dresel, 2011).⁸

Aber nicht nur die kognitive Repräsentation der eigenen Fähigkeiten ist relevant für den Lernerfolg, sondern auch die tatsächlichen *schulischen Fähigkeiten* der Lernenden. Dabei werden fachspezifische Fähigkeiten meist über die fachlichen Leistungen erfasst, z.B. durch entsprechende Tests. Allgemeine schulische Fähigkeiten gehen jedoch über fachspezifische Anteile hinaus und entsprechen einer generellen Begabung (Meyer, 1984) im Sinne einer „individuelle[n] Befähigung (d.h. Potenzial) für bestimmte Leistungen“ (Wirtz, 2013, S. 248). Die akkurate Einschätzung der Fähigkeiten wird z.B. durch die *Flow-Theorie* (Csikszentmihalyi, 1975) deutlich. Demnach ist es von besonderer Bedeutung, Anforderungen an die Lernenden entsprechend ihrer Fähigkeiten auszuwählen, um eine Über- oder Unterforderung zu vermeiden. Die Bedeutsamkeit der Fähigkeiten wird auch im Kontext der Attributionstheorie (Weiner, 1986) ersichtlich. Die Ursache von schulischem Erfolg oder Misserfolg kann nicht nur durch die Schüler(innen) selbst, sondern auch durch die Lehrkräfte,

⁷ Zielorientierungen können dabei allgemein als motivationale Ausrichtungen zur Erreichung bestimmter Ziele beschrieben werden, wobei eine Lernzielorientierung sich im Wesentlichen durch das Bestreben der eigenen Kompetenzerweiterung definieren lässt (z.B. Dweck, 1986; Elliott & Dweck, 1988).

⁸ Das Hilflosigkeitserleben beschreibt dabei das Gefühl, keinen Einfluss auf die Konsequenzen der eigenen Handlungen zu haben (z.B. König, 2009). Ein adaptiver Umgang mit Fehlern kann sich auf behavioraler Ebene z.B. in der „Aufrechterhaltung von Anstrengung und eine[r] angemessene[n] Schwierigkeitswahl bei zukünftigen Aufgaben“ (Steuer, 2014, S. 33) zeigen.

auf unterschiedlichen Ebenen lokalisiert werden. Neben externalen Faktoren kann Erfolg und Misserfolg auf internaler Ebene entweder auf die Fähigkeiten oder die Anstrengungen attribuiert werden (Schunk, Pintrich & Meece, 2008; Weiner, 1986).⁹

Die *Anstrengungsbereitschaft* – als verhaltensbezogene motivationale Ausprägung, die schulische Anstrengung bedingt (Pekrun, 1983) – ist von besonderer Bedeutung, da sie u. a. im Zusammenhang mit der Motivationsregulation, der Selbstwirksamkeitserwartung und der Leistung steht (z. B. Komarraju & Nadler, 2013; Wolters, 1999; Yeo & Neal, 2004).¹⁰ Die Anstrengung, die Kinder und Jugendliche beim Lernen aufbringen, kann als Indikator für ihre Motivation angesehen werden (Schunk et al., 2008). Bei einem höheren Fähigkeitsniveau wird jedoch entsprechend weniger Anstrengung benötigt (Schunk et al., 2008). Somit ist die Anstrengung von Schüler(inne)n bei hohen Fähigkeiten nicht zwingend erforderlich, um Lernerfolge zu erzielen. Die entsprechenden wechselseitigen Beziehungen einiger Einflussfaktoren auf den Lernerfolg von Schüler(inne)n verdeutlichen dabei, dass Lehrkräfte alle Prädiktoren schulischer Leistungen im Blick haben sollten, um bei ihren Schüler(inne)n erkennen zu können, an welchen Stellschrauben sie im individuellen Fall ansetzen können oder müssen.

Als Prädiktoren für schulische Leistungen sind somit zahlreiche Faktoren, wie z. B. schulische Fähigkeiten und Aspekte der Lern- und Leistungsmotivation (u. a. Harackiewicz, Barron, Tauer & Elliot, 2002; Steinmayr & Spinath, 2009), aber auch die eigene Einschätzung der Schüler(innen) von hoher Bedeutung. Die akkurate Einschätzung von tatsächlichen Leistungen und weiteren Schülermerkmalen ist somit die Voraussetzung für realistische Lehrkraft-erwartungen und akkurate Lehrkrafturteile. Dabei fließen häufig jedoch leistungsirrelevante Informationen in die Urteils- und Erwartungsbildung mit ein (z. B. Böhmer, Gräsel, Hörstermann & Krolak-Schwerdt, 2012), so ist dem diagnostischen Urteil von Lehrenden auch stets eine subjektive Interpretation der Beobachtung von Schülerleistungen und Verhaltensweisen sowie eigene Erwartungen vorgeschaltet (Schrader & Helmke, 2001). Lehrkraft-erwartungen beeinflussen das Lehrkraftverhalten und damit die Gestaltung von Unterrichts- und Leistungssituationen, jedoch auch die Wahrnehmung der Schülerleistungen, was sich anschließend im Urteil der Lehrkraft niederschlagen kann (Schrader & Helmke, 2001).

Da anzunehmen ist, dass Kognitionen der Lehrkräfte (z. B. Urteile über und Erwartungen an Schülerleistungen und leistungsrelevante Schülermerkmale) auch Auswirkungen auf ihre Handlungen (z. B. spezifisches Lehrkraftverhalten gegenüber leistungsschwachen oder -starken Schüler(inne)n) haben (Clark & Peterson, 1986), lässt sich ebenfalls die besondere Bedeutung von Lehrkrafturteilen und -erwartungen vermuten. Ergänzend zum Modell von Schrader und Helmke (2001) ist jedoch hervorzuheben, dass Erwartungen nicht nur das Lehrkrafturteil beeinflussen, sondern dieses Urteil auch entsprechend wieder einen Einfluss auf die Erwar-

⁹ Dabei kann attributionales Feedback der Lehrkraft sich u. a. auf das Fähigkeitsselbstkonzept der Schüler(innen) auswirken (Dresel & Ziegler, 2006).

¹⁰ Motivationsregulation kann dabei als zielgerichtete Aktivität zur Initiation, Aufrechterhaltung oder Steigerung der Motivation beschrieben werden (Wolters, 2003). Selbstwirksamkeitserwartung beschreibt die Erwartung von Personen, durch eigene Kompetenzen zukünftige Handlungen zu beeinflussen (Bandura, 1977, 1982), was ebenfalls einen bedeutsamen Einfluss auf die Lernmotivation von Schüler(inne)n hat (z. B. Zimmerman, 2000).

tungen der Lehrkraft haben kann (vgl. Abbildung 5). Folglich ist eine reziproke Beziehung zwischen Urteilen und Erwartungen anzunehmen.

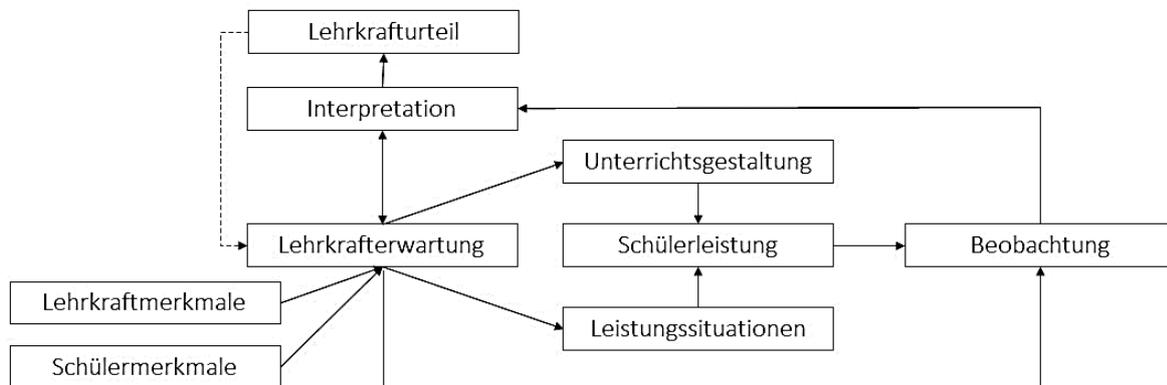


Abbildung 5. Modell zu Einflussfaktoren auf Lehrkrafterwartungen und -urteile (adaptiert nach Schrader & Helmke, 2001, S. 47)

Leistungsbeurteilungen durch Lehrkräfte sind somit nicht nur durch die tatsächlichen Leistungen der Schüler(innen) beeinflusst. Dabei können geringere Übereinstimmungen zwischen tatsächlichen Leistungen und der Lehrkraftbewertung z. T. beabsichtigt sein, um Schüler(innen) zu motivieren, sie können jedoch auch unbeabsichtigt durch u. a. systematische Urteilverzerrungen entstehen. Lehrkrafturteile sind demnach auch der Gefahr von z. B. Beobachtungsfehlern, Urteilstendenzen oder Artefakten durch Einstellungen und Erwartungen unterworfen (Tent & Birkel, 2010). Dabei ist die akkurate Beurteilung von Schülermerkmalen essentieller Bestandteil des Lehrberufs und der diagnostischen Kompetenz von Lehrkräften (Artelt & Gräsel, 2009; Lorenz & Artelt, 2009).

3.1 Diagnostische Kompetenz von Lehrkräften

Der diagnostischen Kompetenz wird eine Schlüsselrolle in den Lehrkompetenzen zugeschrieben (z. B. Artelt & Gräsel, 2009). Sie lässt sich in dem Modell professioneller Handlungskompetenzen (Baumert & Kunter, 2006) in dem Bereich allgemeiner pädagogischer Kompetenzen verorten. Nach Artelt und Gräsel (2009) ist die diagnostische Kompetenz als Übereinstimmung zwischen der Lehrkräfteeinschätzung eines Merkmals (z. B. Schülerleistung oder Aufgabenschwierigkeit) und der tatsächlichen Merkmalsausprägung zu definieren. In diesem Sinn ist die diagnostische Kompetenz u. a. als Urteilsgenauigkeit oder Akkuratheit von Urteilen zu verstehen (Helmke, Hosenfeld & Schrader, 2004; Schrader, 2009). Nur durch eine genaue und richtige Einschätzung von Schüler- und Aufgabenmerkmalen kann eine adäquate Unterrichtsplanung, -durchführung sowie eine gezielte individuelle Förderung (z. B. Auswahl passenden Materials und adäquater Aufgabenschwierigkeiten) erfolgen und angemessene instruktionale Entscheidungen getroffen werden (Shavelson & Stern, 1981; Helmke et al., 2004). Dies ermöglicht einen adaptiven Unterricht – im Sinne einer optimalen Passung auf die individuellen Lernvoraussetzungen der Schüler(innen) (Schrader & Helmke, 2001) – und

ist eine Voraussetzung für die frühzeitige Erfassung von Lernschwierigkeiten, die adäquate Beratung der Schüler(innen) und ihrer Eltern sowie für einen gelungenen Umgang mit Lern- und Leistungsheterogenität (z. B. Bailey & Drummond, 2006; Teisl, Mazzocco & Myers, 2001). Zudem können gute diagnostische Fähigkeiten helfen, Urteilsverzerrung durch Faktoren der ethnischen und sozialen Herkunft zu vermeiden oder zumindest abzumildern (Bos & Hovenga, 2010).

In dem Beschluss der Kultusministerkonferenz vom 16.12.2004 wurde u. a. auf die besondere Verantwortung von Lehrkräften bei ihrer Aufgabe der Diagnose individueller Lernprozesse verwiesen (KMK, 2004). Diagnosen lassen sich dabei von alltäglichen Personenbeschreibungen dadurch abgrenzen, dass diagnostische Urteile „anhand vorgegebener Kategorien, Begriffe und Konzepte“ (Helmke, 2009, S. 122) erstellt werden. Hierbei müssen jedoch die formale Diagnose – die theoriegeleitet ist und sich auf systematisch erhobene Daten stützt – und die informelle Diagnose – die eher auf impliziten Theorien und unsystematisch erhobenen Daten basiert – unterschieden werden (Helmke, 2009; Schrader & Helmke, 2001; Schrader, 2010). Im Gegensatz zu expliziten Diagnosen (z. B. durch Tests oder mündliche Prüfungen), bei denen die Aufmerksamkeit der Lehrkraft bewusst auf die diagnostische Situation gerichtet ist, verlaufen implizite Diagnosen (z. B. durch die intuitive Einschätzung beim Aufrufen eines Schülers oder einer Schülerin) eher unbewusst und meist ohne Formulierung eines direkten Urteils (Schrader & Helmke, 2001). Beide Urteilsformen sind jedoch für die Interaktion mit Schüler(inne)n sowie für die Erwartungsbildung und weitere Beurteilungen von Bedeutung. Zentral für die Bestimmung der diagnostischen Kompetenz ist jedoch – unabhängig davon, ob explizite oder implizite Urteile gefällt werden – inwieweit die Lehrkrafturteile mit den tatsächlichen Merkmalsausprägungen übereinstimmen.

Basierend auf sozialpsychologischen Vorarbeiten zur Übereinstimmung des Urteils einer beurteilenden Person mit den tatsächlichen Angaben der beurteilten Person zu einem spezifischen Merkmal identifizierte Cronbach (1955) Komponenten, die sich zu einem globalen Maß der Urteilsakkuratheit zusammensetzen. In Anlehnung an Cronbach (1955) adaptierten Schrader und Helmke (1987) die Überlegungen zur Berechnung der Urteilsgenauigkeit auf den schulischen Kontext und unterscheiden drei Komponenten der Urteilsgenauigkeit von Lehrkräften (vgl. Helmke et al., 2004):

- (1) Die *Niveauekomponente* (NK) beschreibt die Übereinstimmung zwischen dem Mittelwert der tatsächlichen Schülermerkmale einer Klasse (\bar{x}) und der Einschätzung des Klassenmittelwerts des gleichen Merkmals durch die Lehrkraft (\bar{y}). Dabei lässt sich die Niveauekomponente durch die Subtraktion des tatsächlichen Klassenmittelwerts von dem durch die Lehrkraft eingeschätzten Klassenmittelwert beschreiben ($NK = \bar{y} - \bar{x}$). Liegt das Ergebnis der Subtraktion bei 0, ist von einer optimalen Passung zwischen Lehrkrafturteil und tatsächlicher Merkmalsausprägung in Bezug auf den Klassenmittelwert auszugehen. Ist die Differenz > 0 , ist von einer Überschätzung des Klassenmittelwerts und bei einem Wert < 0 von einer Unterschätzung des Klassenmittelwerts auszugehen.¹¹

¹¹ Nach Schrader und Helmke (1987) beinhaltet die Niveauekomponente neben der adäquaten Einschätzung des durchschnittlichen Leistungsniveaus auch die pädagogische Haltung der Lehrkraft.

- (2) Die akkurate Einschätzung der Leistungsstreuung innerhalb einer Klasse wird durch die *Differenzierungskomponente* (DK) beschrieben. Dabei wird die Streuung des Lehrkrafturteils (s_y) durch die tatsächliche Streuung des Merkmals innerhalb der Klasse (s_x) dividiert ($DK = \frac{s_y}{s_x}$). Somit indiziert $DK = 1$ eine akkurate Einschätzung der Streuung durch die Lehrkraft. Ein Wert < 1 kann demnach als Unterschätzung der Streuung und ein Wert > 1 als Überschätzung der Streuung interpretiert werden.¹²
- (3) Die Vergleichskomponente/Rangkomponente (RK) beschreibt die akkurate Lehrkräftein-schätzung individueller Unterschiede in den Merkmalsausprägungen einer Klasse. Schrader und Helmke (1987) bezeichnen sie „als Indikator für die diagnostische Kompetenz im eigentlichen Sinne“ (S. 35), da Lehrkräfte hierbei die Rangfolge der Ausprägungen eines Merkmals innerhalb einer Klasse beurteilen. Sie berechnet sich als Produktmomentkorrelation zwischen den tatsächlichen individuellen Merkmalsausprägungen der Schüler(innen) (x) und der Einschätzung der individuellen Ausprägungen durch die Lehrkraft (y): $RK = r_{xy}$.¹³

Schrader und Helmke (1987) postulieren, dass diese drei Komponenten der Urteils-genauigkeit unterschiedliche Auswirkungen auf den Lernerfolg von Schüler(inne)n haben können. In der entsprechenden empirischen Überprüfung dieser Annahmen erwies sich jedoch weder die Niveauelemente noch die Rangkomponente als bedeutsam für den Lernerfolg der Schüler(innen) (Schrader & Helmke, 1987). Für die Differenzierungskomponente zeigte sich hingegen, dass bei einer Überschätzung der Streuung geringere Leistungszuwächse bei leistungsschwachen Schüler(inne)n zu verzeichnen waren als bei einer Unterschätzung oder nur geringen Überschätzung.

Auch wenn der Begriff der diagnostischen Kompetenz den Eindruck einer allgemeinen Fähigkeit von Lehrkräften erweckt, konnte Spinath (2005) durch die Analyse der Korrelationen zwischen den Komponenten der Urteils-genauigkeit lediglich vereinzelte Zusammenhänge feststellen. Somit ist nicht von einer allgemeinen Kompetenz auszugehen, was auch in empirischen Analysen eine differenzierte Betrachtung der einzelnen Komponenten erfordert (vgl. McElvany et al., 2009; Spinath, 2005). Dabei ist zu beachten, dass nicht in allen empirischen Untersuchungen, die sich mit der Urteils-genauigkeit von Lehrkräften befassen, explizit auf die einzelnen Komponenten eingegangen wird und zudem oftmals nur einzelne Aspekte der diagnostischen Kompetenz adressiert werden. In zahlreichen Studien zeigt sich jedoch, dass die Ausprägungen der Komponenten der diagnostischen Kompetenz von Lehrkräften im Durchschnitt häufig nicht zufriedenstellend ist (z. B. Südkamp et al., 2012) und auch leistungsirrelevante Faktoren in die Beurteilung miteinfließen können (z. B. Dunkake et al., 2012).

¹² Schrader und Helmke (1987) nehmen an, dass Lehrkräfte mit einer stark ausgeprägten Spannweite der Leistungsstreuung besonders auf Leistungsunterschiede innerhalb einer Klasse achten würden.

¹³ Die von Schrader und Helmke (1987) eingeführte Vergleichskomponente wird in aktuelleren Untersuchungen zur diagnostischen Kompetenz (z. B. Stang & Urhahne, 2016a) als Rangkomponente (z. T. auch Korrelationskomponente, Helmke et al., 2004) bezeichnet. In der vorliegenden Arbeit wird der aktuell geläufigere Begriff der Rangkomponente verwendet.

3.1.1 Empirische Befunde zur Urteilsgenauigkeit von Lehrkräften

Untersuchungen zur Niveauebene berichten häufig von einer Überschätzung der Leistungen durch die Lehrkräfte (Doherty & Conolly, 1985; Kaiser, Möller, Helm & Kunter, 2015; Schrader & Helmke, 1987; Stang & Urhahne, 2016b; Zhu & Urhahne, 2015). Jedoch verweisen einzelne Untersuchungen auch auf gegenteilige Befunde. So konnten z. B. McElvany et al. (2009) eine Unterschätzung der Schüler(innen) durch die Lehrkraft feststellen. Die meisten empirischen Untersuchungen zur diagnostischen Kompetenz und im Speziellen zur Niveauebene fokussieren fachliche Leistungen der Schüler(innen). Weitere relevante Schülermerkmale sind hingegen seltener im Fokus empirischer Studien. Ausnahmen hierzu finden sich u. a. bei Spinath (2005), die in einer Untersuchung mit Grundschullehrkräften weitere leistungsrelevante Merkmale (Intelligenz, Fähigkeitsselbstwahrnehmung, Lernmotivation und Leistungsängstlichkeit) berücksichtigte. Dabei schätzen Lehrkräfte die durchschnittliche Intelligenz akkurat ein, die Fähigkeitsselbstwahrnehmung und die Lernmotivation wurden hingegen eher unterschätzt und die Leistungsängstlichkeit eher überschätzt (Spinath, 2005). Praetorius, Karst, Dickhäuser und Lipowsky (2011) konnten hingegen eine durchschnittliche Übereinstimmung des Lehrkrafturteils und des Fähigkeitsselbstkonzepts feststellen. Itskowitz, Navon und Strauss (1988) zeigten in Bezug auf die Akkuratheit der Lehrkräfteeinschätzungen des Selbstkonzepts, dass Lehrkräfte mit hoher emotionaler Nähe zu ihren Schüler(inne)n diese eher überschätzten. Im Gegensatz dazu wurden Schüler(innen), denen gegenüber die Lehrkraft eine geringe Nähe verspürte, in ihrem Selbstkonzept eher unterschätzt (Itskowitz et al., 1988). Hosenfeld, Helmke und Schrader (2002) konnten in Bezug auf die Lehrkrafturteile zum Interesse, der Aufmerksamkeit und dem Verständnis der Schüler(innen) eine deutliche Unterschätzung feststellen. Dies lässt darauf schließen, dass nicht von einer generellen Tendenz der Lehrkräfte zur Überschätzung ausgegangen werden kann, sondern dass das jeweilige einzuschätzende Merkmal der Schüler(innen) einen entsprechenden Einfluss auf die Genauigkeit der Lehrkrafturteile haben kann und je nach Merkmal eher eine Über- oder Unterschätzung auftreten kann.

Stang und Urhahne (2016b) untersuchten die Einschätzung der Mathematikleistung, der Konzentration sowie des Arbeits- und Sozialverhaltens von Schüler(inne)n der fünften Jahrgangsstufe durch Mathematiklehrkräfte, wobei die Befunde auf eine akkurate Einschätzung der Streuung der Schülerleistungen hindeuten. Südkamp und Möller (2009) berichten hingegen von einer Unterschätzung der Streuung der Schülerleistungen durch die Lehrkräfte (vgl. auch Südkamp, Möller & Pohlmann, 2008). Auch Kaiser et al. (2015) konnten in ihrer Studie eine Unterschätzung der Leistungsstreuung nachweisen. Spinath (2005) berichtet in Bezug auf die Intelligenz und die Leistungsängstlichkeit eher eine Unterschätzung der Streuung. Die Streuung der Lernmotivation und des Fähigkeitsselbstkonzepts wurden hingegen eher überschätzt (Spinath, 2005). Praetorius et al. (2011) berichten allerdings eine durchschnittlich eher akkurat eingeschätzte Streuung des Fähigkeitsselbstkonzepts.

Studien in denen die Rangkomponente der diagnostischen Kompetenz berücksichtigt wurde, konnten häufig einen überwiegend moderaten Zusammenhang zwischen dem Lehrkrafturteil und den tatsächlichen Schülerleistungen ermitteln (Demaray & Elliot, 1998; Feinberg & Shapiro, 2003; Kaiser, Retelsdorf, Südkamp & Möller, 2013; Zhu & Urhahne, 2015). Südkamp

et al. (2012) fanden in ihrer Meta-Analyse aus 75 Studien zur Urteilsübereinstimmung von Lehrkräften mit den tatsächlichen Schülerleistungen eine relative hohe Übereinstimmung ($r = .63$), jedoch darf dabei nicht vernachlässigt werden, dass mindestens ein Drittel der Varianz nicht durch die tatsächlichen Schülerleistungen erklärt werden konnte. Hoge und Coladarci (1989) berichten in ihrer Meta-Analyse ebenfalls eine hohe Korrelation von $r = .66$ zwischen Lehrkrafturteilen und tatsächlichen Schülerleistungen. Jedoch setzte sich dies aus einer Spannweite der Korrelationen von $r = .28$ bis $r = .92$ in den einzelnen Untersuchungen zusammen (Hoge & Coladarci, 1989). Auch Stang und Urhahne (2016b) fanden lediglich eine geringe Überstimmung zwischen der tatsächlichen Leistung der Beurteilung durch die Lehrkräfte.

Neben den Befunden zur Übereinstimmung der Leistung und des Lehrkrafturteils, untersuchte Spinath (2005) die Überstimmung zwischen weiteren Schülermerkmalen und den entsprechenden Urteilen der Lehrkräfte. Die Intelligenz und das Fähigkeitsselbstkonzept der Schüler(innen) hingen moderat ($r_{\text{Intelligenz}} = .40$; $r_{\text{Fähigkeitsselbstkonzept}} = .39$) mit den Lehrkrafturteilen zusammen (Spinath, 2005). Die Befunde zur Einschätzung des Fähigkeitsselbstkonzepts stehen im Gegensatz zu den Befunden von Praetorius et al. (2011), die den Lehrkräften eher Schwierigkeiten bei der Bestimmung der Rangfolge des Fähigkeitsselbstkonzepts attestierten. Im Einklang mit den Befunden von Urhahne et al. (2010) zeigten sich bei der Leistungsängstlichkeit und Lernmotivation ($r_{\text{Leistungsängstlichkeit}} = .15$; $r_{\text{Lernmotivation}} = .20$) jedoch geringe Zusammenhänge (Spinath, 2005). Lohbeck, Nitkowski, Petermann & Petermann (2014) konnten für das Sozial- und Lernverhalten kleine bis moderate Zusammenhänge zwischen den Merkmalsausprägungen der Schüler(innen) und den Lehrkrafturteilen ermitteln (vgl. auch Stang & Urhahne, 2016b).

Da die diagnostische Kompetenz nicht nur die akkurate Einschätzung von Schülerleistungen, sondern auch von Aufgabenmerkmalen beinhaltet, sei an dieser Stelle auch auf eine Untersuchung von Hoffmann und Böhme (2013) mit Deutsch- und Mathematiklehrkräften der Primarstufe verwiesen. Dabei zeigte sich eine moderate Korrelation für die Einschätzung der Aufgabenschwierigkeit, die vergleichbar mit den Befunden zur Einschätzung von Schülermerkmalen durch eine hohe Varianz zwischen den Lehrkräften gekennzeichnet war. In Bezug auf die Über- und Unterschätzung der Aufgabenschwierigkeit verweisen die Befunde von Hoffmann und Böhme (2013) auf eine Überschätzung der Aufgabenschwierigkeit von leichten Aufgaben sowie eine Unterschätzung der Schwierigkeit bei schweren Aufgaben. Insgesamt zeigte sich eine Unterschätzung der Streuung bei den Aufgabenschwierigkeiten.

Zusammenfassend lässt sich festhalten, dass Lehrkräfte dazu tendieren, die durchschnittlichen Schülerleistungen (Niveauelemente) zu überschätzen (vgl. u. a. Kaiser et al., 2015; Hosenfeld et al., 2002; Stang & Urhahne, 2016a, 2016b; Südkamp & Möller, 2009; Südkamp et al., 2008). Die Streuung der Schülerleistungen (Differenzierungskomponente) wurde in einigen Studien akkurat eingeschätzt (z. B. Stang & Urhahne, 2016b). In anderen Studien zeigte sich hingegen eine Unterschätzung der Streuung (z. B. Kaiser et al., 2015; Südkamp & Möller, 2009; Südkamp et al., 2008). Die akkurate Zuordnung des Rangs der Schülerleistungen (Rangkomponente) scheint Lehrkräften leichter zu gelingen, Untersuchungen für diese Komponente fanden häufig eine moderate bis gute Übereinstimmung (z. B. Karing, Pfohl & Artelt, 2011; Kaiser et al., 2015; Stang & Urhahne, 2016b; Südkamp et al., 2008). Die akkurate Beurteilung leistungsrelevanter Schülermerkmale, die nur in vereinzelt Studien berücksichtigt wurden

(z. B. Lohbeck et al., 2014; Spinath, 2005; Stang & Urhahne, 2016b), scheint den Lehrkräften dabei häufig größere Probleme zu bereiten.

Durchschnittlich moderate Ausprägungen der Rangkomponente zeigten sich jedoch eher bei der Beurteilung von Leistungen und z. T. noch für Verhaltensmerkmale, intraindividuelle Merkmale wie z. B. motivationale oder emotionale Aspekte hingegen waren für Lehrkräfte deutlich schwerer akkurat einzuschätzen. Da diese nicht direkt beobachtbar sind, ließe sich die geringere Akkuratheit für Lernmotivation und -emotionen hierdurch möglicherweise erklären. Kritisch ist zu beachten, dass keine konsistenten Kriterien zur Erfassung der Akkuratheit von Lehrkrafturteilen und damit auch keine einheitlichen Messmethoden des Lehrkrafturteils und der tatsächlichen Schülerleistung vorliegen (Südkamp et al., 2012). So werden z. T. unterschiedliche Komponenten der diagnostischen Kompetenz erfasst und zudem unterschiedliche Grundlagen für die Informationen und Urteile der Lehrkräfte geliefert (z. B. direkte vs. indirekte Urteile; vgl. Hoge & Coladarci, 1989; Südkamp et al., 2012).

Bereits 1965 lieferte Weiss Befunde aus einer experimentellen Untersuchung mit österreichischen Lehrkräften, die große Unterschiede zwischen Lehrkräften bei der Notenvergabe für zwei Aufsätze der vierten Jahrgangsstufe sowie für eine Rechenarbeit der vierten und fünften Jahrgangsstufe nachwies. Dabei lagen die vergebenen Noten der Deutscharbeiten zwischen eins und vier und die Zensuren der Mathematikarbeiten zwischen eins und fünf.¹⁴ Somit zeigen sich deutliche interindividuelle Unterschiede zwischen Lehrkräften in der Beurteilung von schulischen Leistungen, was im Durchschnitt bereits auf eine eher geringe diagnostische Kompetenz hindeutet. Trotz der wachsenden Bewusstheit über die Bedeutung einer akkuraten Beurteilung durch Lehrkräfte berichten auch aktuellere Untersuchungen häufig eine große Streuung der Lehrkrafturteile bei der Bewertung von Schülerarbeiten (z. B. Brimi, 2011). So konnten auch in Untersuchungen von Birkel und Birkel (2002) für Deutschsaufsätze sowie für Mathematikarbeiten von Birkel (2005) die Befunde von Weiss (1965) repliziert werden.

Diese Befunde verdeutlichen die großen interindividuellen Unterschiede in der diagnostischen Kompetenz von Lehrkräften. Zudem zeigt sich, dass Lehrkräfte ihre Schüler(innen) nicht immer akkurat beurteilen. Dies deutet darauf hin, dass eine der wichtigsten Aufgaben im Lehrberuf gleichzeitig eine der schwierigsten Aufgaben darstellt. Basierend auf den in diesem Abschnitt dargestellten Ergebnissen empirischer Untersuchungen liegt die Annahme nahe, dass Lehrkrafturteile neben tatsächlichen Leistungs- und Merkmalsausprägungen von Schüler(inne)n durch weitere Faktoren beeinflusst werden. Mögliche Einflussfaktoren auf die Urteilsakkuratheit werden im folgenden Abschnitt beleuchtet.

3.1.2 Einflussfaktoren auf die Akkuratheit von Lehrkrafturteilen

Als Ursachen für inakkurate Lehrkrafturteile werden eine Vielzahl an möglichen Einflussfaktoren diskutiert, z. B. logische Fehler oder der Halo-Effekt (vgl. u. a. Kaiser et al., 2015; Stang & Urhahne, 2016b). Funder (1995) beschreibt in seinem *realistic accuracy model* als Einflussfaktoren für die akkurate Einschätzung von Persönlichkeitsmerkmalen sowohl Merk-

¹⁴ Anzumerken ist, dass in Österreich ein fünfstufiges Notensystem, von 1 (*sehr gut*) bis 5 (*nicht genügend*) eingesetzt wird (§ 14 Leistungsbeurteilungsverordnung; Bundeskanzleramt, 2012).

male des Urteilenden (z. B. Wahrnehmung), des Beurteilten (z. B. Konsistenz), des Merkmals (z. B. Sichtbarkeit) als auch der vorliegenden Informationen (z. B. Qualität). Die Vielzahl von Einflussfaktoren auf die Urteilsgenauigkeit lassen sich auch auf den pädagogischen Kontext und die Einschätzung von Schüler- und Aufgabenmerkmalen übertragen. Südkamp et al. (2012) benennen in dem *Modell der Akkuratheit von Lehrkrafturteilen* (vgl. Abbildung 6) ebenfalls vier Einflussfaktoren, die vergleichbar mit den Merkmalen im Modell von Funder (1995) sind.

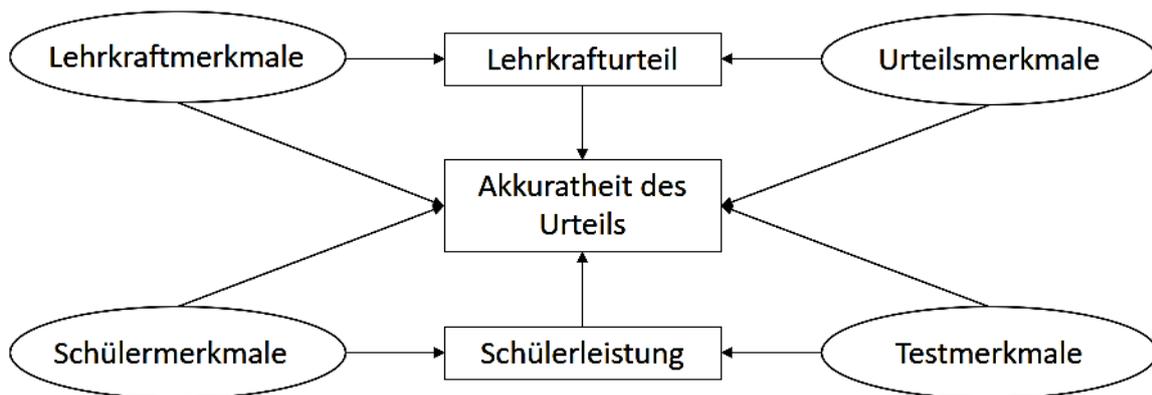


Abbildung 6. Modell der Akkuratheit von Lehrkrafturteilen nach Südkamp et al. (2012, S. 756)

Südkamp et al. (2012) verweisen in ihrem Modell der Akkuratheit von Lehrkrafturteilen auf den Einfluss von Urteils- und Testmerkmalen sowie auf Merkmale der Schüler(innen) und Lehrkräfte. Empirische Befunde bekräftigen dabei die Annahme des Einflusses der vier Bereiche auf die Urteilsakkuratheit, jedoch blieben Analysen zum Zusammenspiel aller Merkmale und der Bedeutungsgröße der einzelnen Einflussgrößen bislang aus. Dennoch gibt es zu den jeweiligen Merkmalen bereits einige empirische Untersuchungen, die im Folgenden exemplarisch skizziert werden.

Einfluss von Urteil- und Testmerkmalen auf die Akkuratheit von Lehrkrafturteilen

Karing und Artelt (2013) verweisen auf die Bedeutung der Spezifität des Urteils. Sekundarschullehrkräfte in den Fächern Mathematik und Deutsch weisen eine höhere Übereinstimmung bei globalen, also allgemeineren, Urteilen als bei aufgabenspezifischen Urteilen auf. Karing und Artelt (2013) vermuten, dass bei globalen Urteilen Wissen über die Fähigkeiten des Schülers oder der Schülerin notwendig sind, bei spezifischen Urteilen jedoch zusätzlich Wissen über die Aufgabe vorhanden sein muss. Bei der Erfassung der diagnostischen Kompetenz müssen nach Südkamp et al. (2012) zudem auch die Domänenspezifität des Urteils sowie die Informiertheit der Lehrkräfte über die Tests oder Vergleichsmaßstäbe (in Anlehnung an die Unterscheidung in direkte und indirekte Lehrkrafturteile; Hoge & Coladarci, 1989) berücksichtigt werden. Eine informierte Lehrkraft schätzt dabei z. B. die Schülerleistung in einem bestimmten Leistungstest ein, eine uninformierte Lehrkraft würde hingegen die Leistung in einer bestimmten Domäne angeben, ohne den spezifischen Test zur Erfassung der tatsächlichen Schülerleistung zu kennen. Informierte Lehrkräfte weisen erwartungsgemäß eine höhere

Urteilsübereinstimmung mit den Schülerleistungen auf als uninformierte Lehrkräfte (Feinberg & Shapiro, 2003; Demaray & Elliot, 1998; Südkamp et al., 2012). Dies erscheint auch plausibel, da Lehrkräfte durch die Information über den konkreten Test, den Schüler(innen) bearbeiten sollen, weitaus mehr Hinweise über konkrete Stärken und Schwächen der Schüler(innen) auf die Aufgabenstellung übertragen können. Bei der Einschätzung des Schwierigkeitsniveaus von Aufgaben zeigt sich eine höhere Differenzierung der Lehrkrafturteile bei leichten Aufgaben im Vergleich zu schweren Aufgaben (McElvany et al., 2009).

Einfluss von Schülermerkmalen auf die Akkuratheit von Lehrkrafturteilen

Insbesondere Merkmale der Schüler(innen) werden als bedeutsame Einflussfaktoren auf die Urteilsbildung von Lehrkräften diskutiert und auch in dem Modell von Südkamp et al. (2012) thematisiert. Weiss (1965) konnte in einer Untersuchung zur Bewertung von Deutschaufsätzen zeigen, dass neben den tatsächlichen Leistungen weitere Schülerinformationen bei der Urteilsbildung der Lehrkräfte eine Rolle spielen. Neben allgemeinen positiven (sprachlich begabt) und negativen (liest gerne Schundhefte) Informationen wurden den Lehrkräften dabei ebenfalls Angaben zur Berufstätigkeit der Eltern mitgeteilt, über die Rückschlüsse zum familiären Hintergrund und Sozialstatus möglich waren (beide Elternteile berufstätig vs. Vater ist Redakteur einer großen Tageszeitung). Unter der positiven Manipulation bewerteten die Lehrkräfte sowohl Deutsch- als auch Mathematikarbeiten der vierten Jahrgangsstufe signifikant besser als unter der negativen Schülerbeschreibung. Durch die Konfundierung von Schülerinformationen und familiären Informationen sind, basierend auf der vorliegenden Publikation, jedoch keine differenzierten Rückschlüsse über den Einfluss der sozialen Herkunft der Schüler auf die Urteilsbildung möglich. Neben positiven oder negativen Informationen über Schüler(innen) kann z. B. auch das physische Aussehen einen bedeutsamen Effekt auf Lehrkrafturteile haben. Ritts, Patterson und Tubbs (1992) berichten, dass Lehrkräfte physisch attraktivere Schüler(innen) durchschnittlich besser beurteilten (z. B. Intelligenz, soziale Fähigkeiten, akademisches Potential) als weniger attraktivere Schüler(innen) (vgl. auch Dunkake et al., 2012).

Bereits 1984 konnten Hoge und Butcher ebenfalls den Einfluss der Schülerintelligenz auf die Urteilsgenauigkeit nachweisen. Lehrkräfte überschätzten die Testleistung von Schüler(inne)n mit hohen kognitiven Fähigkeiten und unterschätzten die Leistung von Schüler(inne)n mit geringeren kognitiven Fähigkeiten (Hoge & Butcher, 1984; vgl. auch Kaiser et al., 2015). Auch weitere Studien verweisen auf den Einfluss der Intelligenz der Lernenden bei der Leistungsbeurteilung (Kaiser et al., 2015; Schrader & Helmke, 1990). Fehlerhafte Lehrkräfteeinschätzungen zeigten sich auch bei der Einschätzung des Fähigkeitsselbstkonzepts, das ebenfalls durch die Intelligenz als auch die Leistung der Schüler(innen) verzerrt wird (Praetorius, Greb, Lipowsky & Gollwitzer, 2010; Schrader & Helmke, 1990).

Kaiser et al. (2013) konnten in einer Feld- und zwei Experimentaluntersuchungen den Einfluss der Schülerleistung sowie der Unterrichtsbeteiligung der Schüler(innen) auf die Akkuratheit von Lehrkrafturteilen nachweisen. Einige empirische Befunde deuten darauf hin, dass Lehrkräfte eine höhere Urteilsgenauigkeit bei Schüler(inne)n mit hohen Leistungen aufweisen, als bei Schüler(inne)n mit durchschnittlichen oder schlechten Leistungen (z. B. Begeny, Eckert, Montarello & Storie, 2008; Demaray & Elliot, 1998; Südkamp & Möller, 2009). Das Geschlecht hingegen konnte bis auf einige Ausnahmen (z. B. Doherty & Conolly, 1985; Kaiser

et al., (2015) i. d. R. nicht als bedeutsamer Einflussfaktor identifiziert werden (z. B. Dicke, Lüdtke, Trautwein, Nagy & Nagy, 2012; Hoge & Butcher, 1984; Glock, Krolak-Schwerdt, Klapproth & Böhmer, 2012).

Die Befunde zum Einfluss von Schülermerkmalen auf die Urteilsgenauigkeit lassen den Schluss zu, dass sich die Urteilsakkuratheit von Lehrkräften bei bestimmten Schülergruppen unterscheidet und z. T. Rückschlüsse auf die Leistung durch implizite Persönlichkeitstheorien gezogen werden. Die Ergebnisse zum Einfluss des physischen Erscheinungsbildes der Schüler(innen) lassen zudem systematische Urteilsverzerrungen aufgrund leistungsirrelevanter Schülerinformationen vermuten. Bislang ist jedoch weitestgehend unklar, ob auch die ethnische und/oder soziale Schülerherkunft einen Einfluss auf die Urteilsakkuratheit von Lehrkräften hat.

Einfluss von Lehrkraftmerkmalen auf die Akkuratheit von Lehrkrafturteilen

Kaiser et al. (2015) nehmen an, dass inakkurate Notenbewertungen dadurch entstehen, dass Lehrkräfte das Potential der Schüler(innen) in ihre Bewertung einfließen lassen, um dieses zu würdigen oder die Schüler(innen) zu motivieren. Dies würde eine Überschätzung der Schülerleistungen teilweise erklären, wenn die Urteilsakkuratheit über vergebene Zensuren erfasst wird. Jedoch liefert diese Annahme weder eine Erklärung für Befunde, in denen um die Einschätzung der Schülerleistung in einem spezifischen Test gebeten wurde, noch für eine verzerrte Beurteilung der Leistungsstreuung innerhalb einer Klasse oder eine fehlerhafte Rangordnung der Schüler(innen). Schrader und Helmke (1990) vermuten, dass Lehrkräfte Leistung und Intelligenz wegen des hohen empirischen Zusammenhangs nur schwer unterscheiden können. Hiermit ließe sich auch erklären, dass die Deutschleistung einen Einfluss auf die Beurteilung der Mathematikleistung hat (Kaiser et al., 2015). So könnte angenommen werden, dass Lehrkräfte die Einschätzung der kognitiven Fähigkeiten von Schüler(inne)n als Indikator für fachspezifische Leistungen heranziehen oder fachspezifische Leistungen umgekehrt als Indikator für die Intelligenz interpretieren und dies in die Beurteilung anderer Domänen einfließt.

Diskutiert werden auch spezifische Lehrkraftmerkmale, die einen Einfluss auf die Urteilsgenauigkeit haben könnten, wie z. B. die Berufserfahrung, das Alter oder das Geschlecht der Lehrkraft (z. B. Baumert & Kunter, 2006; Stang & Urhahne, 2016b). Jedoch konnten empirische Untersuchungen i. d. R. keinen oder nur einen geringen Einfluss der Berufserfahrung sowie keinen Einfluss des Alters oder des Geschlechts auf die diagnostische Kompetenz nachweisen (z. B. Dicke et al., 2012; Hoffmann & Böhme, 2013; McElvany et al., 2009; Praetorius et al., 2011; Stang & Urhahne, 2016b; Zhu & Urhahne, 2015). Die Berufserfahrung der Lehrkraft könnte demnach auch dazu führen, dass Lehrkräfte im Unterricht auf Handlungsroutrinen zurückgreifen und damit einige Urteile im Unterricht eher automatisiert ablaufen (vgl. Schrader & Helmke, 2001).

Empirische Untersuchungen konnten zudem Hinweise auf den Einfluss der Urteilssicherheit (Praetorius, Berner, Zeinz, Scheunpflug & Dresel, 2013) finden, in den überwiegenden Fällen lies sich die Vermutung des Einflusses der Kontaktdauer mit den Schüler(inne)n (z. B. Hoffmann & Böhme, 2014; Stang & Urhahne, 2016a) jedoch nicht bestätigen (vgl. z. B. Ambady & Rosenthal, 1992, 1993; Oerke, McElvany, Ohle, Ullrich & Horz, 2016; Praetorius et al., 2014). Karing (2009) konnte allerdings eine höhere Urteilsgenauigkeit bei Grundschul-

lehrkräften im Vergleich zu Gymnasiallehrkräften feststellen. Zu vermuten wäre, dass u. a. die zeitlich intensivere Arbeit von Grundschullehrkräften in ihren Klassen (im Vergleich zu Gymnasiallehrkräften, die meist als Fachlehrkräfte weniger Zeit in den Klassen verbringen) eine Rolle spielen könnte. Dies deutet darauf hin, dass womöglich die Intensität des Kontakts, jedoch nicht die reine Quantität, die Urteilsgenauigkeit beeinflusst. Hierzu fehlt es jedoch bislang an ausreichender empirischer Evidenz.

Auch die Annahme, dass Lehrkräfte, die sich zusätzlich weitere Informationen zu Schüler(inne)n (z. B. durch eine enge Kooperation mit anderen Lehrkräften oder Eltern) suchen, akkuratere Urteile bilden, konnte in der empirischen Prüfung nicht bestätigt werden (Karing, 2009). Dabei ist auch in Frage zu stellen, welche Informationen Lehrkräfte durch diese weitere Informationssuche erlangen würden. Angenommen werden könnte, dass durch die weitere Informationssuche auch zahlreiche leistungsirrelevante Informationen über die Schüler(innen) erhalten, die sich in entsprechenden Erwartungen und ungenauen Lehrkrafturteilen niederschlagen könnten.

Bisherige Studien konnten somit den Einfluss z. B. der Berufserfahrung oder der Kontaktdauer mit den Schüler(inne)n weitestgehend widerlegen und im Gegenzug Hinweise für die Bedeutung der Urteilssicherheit von Lehrkräften auf die Akkuratheit ihrer Urteile finden. Jedoch zeigen sich in diesem Forschungsfeld noch zahlreiche offene Fragen zum Einfluss unterschiedlicher Lehrkraftmerkmale auf die Beurteilung von Schüler(inne)n. In diesem Zusammenhang erscheint neben der Akkuratheit der Lehrkrafturteile auch die Frage nach der Bedeutung von Lehrkrafterwartungen für die Urteilsbildung bedeutsam. Schrader und Helmke (2001) verweisen auf den Einfluss von Erwartungen, die sich einerseits aus Schülermerkmalen (z. B. auch der Zusammensetzung der Klasse) aber auch aus Lehrkraftmerkmalen, wie z. B. Einstellungen oder diagnostischen Fähigkeiten zusammensetzen. Diese Faktoren spielen eine entscheidende Rolle für die Gestaltung von Lern- und Leistungssituationen, aber auch für die Beobachtung von Leistungen und deren Interpretation. Somit ist die Urteilsbildung und entsprechend auch die Akkuratheit der Urteile stets dem Einfluss der subjektiven Erwartungen unterworfen, die im folgenden Abschnitt näher beschrieben werden.

3.2 Lehrkrafterwartungen und Erwartungseffekte

„Erwartungen sind realitätsbezogene Vorstellungen, die auf die Zukunft gerichtet sind und von denen der Erwartende überzeugt ist, daß [sic!] sie durch die nachfolgende Ereignisentwicklung bestätigt werden“ (Ludwig, 1991, S. 32). Dabei kann ein Individuum entweder davon ausgehen, dass seine Erwartungen sich in Zukunft mit Sicherheit (*deterministisch*) oder mit einer relativ hohen Wahrscheinlichkeit (*probabilistisch*) erfüllen werden, wobei Erwartungen nach Ludwig (1991) meist einen deterministischen Charakter aufweisen, was sie von Vermutungen oder Hoffnungen mit eher probabilistischem Charakter abgrenzt. Menschen sind demnach eher dazu geneigt, davon auszugehen, dass ihre Erwartungen eintreten werden. Personen können dabei Erwartungen an unterschiedliche Objekte richten, sie können sich z. B. auf Dinge oder Situationen beziehen, aber auch als intra- oder interpersonelle Erwartungen auf die eigene Person oder andere Individuen gerichtet sein (Ludwig, 1991).

Lehrkrafterwartungen können demnach als auf die Zukunft gerichtete interindividuelle Überzeugungen angesehen werden. Good und Brophy (2003) beschreiben Lehrkrafterwartungen als Schlüsse, die Lehrkräfte über zukünftiges Verhalten oder Leistungen ihrer Schüler(innen) ziehen. Dabei gehen sie davon aus, dass diese Erwartungen auf dem aktuellen Wissen über die Schüler(innen) basieren (Good & Brophy, 2003). Es ist jedoch davon auszugehen, dass Erwartungen nicht nur auf der Basis des Wissens über spezifische Schüler(innen) gebildet werden, sondern auch bedingt durch Erfahrungen entstehen, die im Verlauf des Lebens gesammelt werden. Hierdurch ist es Menschen möglich sich auf ähnliche zukünftige Situationen durch „antizipatorische Überzeugungen“ (Ludwig, 1991, S. 31) einzustellen. Entsprechend sind Erwartungen von hoher Bedeutung, um v. a. in komplexen sozialen Situationen erfahrungsbasierte Urteile fällen zu können. Dies kann im schulischen Kontext jedoch bedeuten, dass Lehrkräfte Erwartungen an spezifische Schüler(innen) richten, die auf Grund von Erfahrungen mit anderen Schüler(inne)n entstanden sind. Zum einen kann dies im Unterrichtsalltag Handlungsroutinen schaffen, die es der Lehrkraft ermöglichen, mit spezifischen Herausforderungen schnell umgehen zu können. Zum anderen kann dies jedoch auch dazu führen, dass hohe oder niedrige Erwartungen an einzelne Schüler(innen) gerichtet werden, obwohl diese Erwartungen nicht dem tatsächlichen Potenzial der Schüler(innen) entsprechen.

Die Grundlage zur Erwartungsbildung kann dabei auch auf tatsächlichen Beobachtungen der Schüler(innen) beruhen, die jedoch stets durch die subjektive Wahrnehmung und Interpretation eingefärbt sind (Darley & Fazio, 1980). Grundsätzlich ist davon auszugehen, dass Wahrnehmung ein konstruktiver und interpretativer Prozess ist (Darley & Fazio, 1980), der das individuelle Verhalten entsprechend der subjektiven Interpretationen der sozialen Umwelt beeinflusst. Lehrkräfte interpretieren das Schülerverhalten und deren Leistungen somit immer aus ihrer subjektiven Perspektive. Demnach ist auch die Erwartungsbildung durch subjektive Annahmen beeinflusst. Entsprechend lassen sich hierdurch große interindividuelle Erwartungsunterschiede zwischen Lehrkräfte (vgl. Hoge, 1984) erklären.

Lehrkrafterwartungen können allgemein als das Produkt aus bisherigen eigenen Erfahrungen mit einzelnen Lernenden, bestimmten Schülergruppen oder ganzen Klassen sowie Informationen Dritter (z. B. Aussagen im Lehrerkollegium über Verhaltensweisen oder Leistungen einer Klasse) angesehen werden. Im schulischen Kontext können Erwartungen durch eine Viel-

zahl an Informationen beeinflusst werden, so z. B. durch Vorerfahrungen mit Geschwisterkindern. Seaver (1973) konnte in einer quasiexperimentellen Untersuchung u. a. zeigen, dass Lehrkrafterwartungen für Schüler(innen), deren ältere Geschwister bereits positiv aufgefallen waren, besser ausfielen. Erwartungen können aber auch durch gesellschaftlich geteilte Überzeugungen und eigene Einstellungen sowie Stereotype beeinflusst werden und stehen somit in einem engen Zusammenhang mit sozialen Kategorien und entsprechenden Assoziationen und Überzeugungen (Rosenberg & Jones, 1972; vgl. Kapitel 4).

Nach Miller und Turnbull (1986) können sich Erwartungshaltungen auch aus Überzeugungen gegenüber demographischen Charakteristika, wie z. B. der ethnischen oder sozialen Herkunft, entwickeln. Erwartungen sind als wesentliche Bestandteile des täglichen Lebens unverzichtbar und nicht per se eine Bedrohung (Jussim, 1990). Sind Erwartungen akkurat – repräsentieren sie also die Realität – sind sie zwar womöglich ein Indikator für bestehende Differenzen zwischen sozialen Gruppen, erzeugen diese jedoch nicht (Jussim, 1990). Entsprechen Erwartungen jedoch nicht den tatsächlichen potentiellen Leistungen oder Verhaltensweisen, können sie sich in Erwartungseffekten niederschlagen (Jussim, 1990). Lehrkrafterwartungseffekte können dabei allgemein als Effekte auf Schülerleistungen und -verhalten verstanden werden, die durch Lehrkraftverhalten, das auf Erwartungen beruht, entstehen (Good & Brophy, 2003). Erwartungen können im Wesentlichen zwei bedeutsame Konsequenzen haben (Darley & Fazio, 1980; Darley & Oleson, 1993; Jussim, 1990; Jussim, Eccles & Madon, 1996):

- (1) Erwartungen können die Wahrnehmung beeinflussen und somit dazu führen, dass Urteile über eine Person durch erwartetes Verhalten eingefärbt sind bzw. sich der Wahrnehmungsfokus verschiebt (*perceptual confirmation effect*). Miller und Turnbull (1986) gehen davon aus, dass Erwartungen den Informationsverarbeitungsprozess dahingehend beeinflussen, dass eher erwartungskonforme Informationen wahrgenommen werden, aber auch Informationen entsprechend der eigenen Erwartungen interpretiert werden. Nach Good und Brophy (2003) kann dieser Effekt auch als *sustaining expectation effect* wirksam werden. Dieser beschreibt, dass die Einordnung von Schüler(inne)n auf einem bestimmten Leistungsniveau zu entsprechenden Erwartungen an die zukünftigen Leistungen führen und die dafür sorgen, dass Leistungsveränderungen unbemerkt bleiben (Good & Brophy, 2003).
- (2) Erwartungen können aber auch zu tatsächlichen Verhaltensänderungen führen (*behavioral confirmation effect*). Der in diesem Zusammenhang bekanntere Begriff der *self-fulfilling prophecy* (selbsterfüllende Prophezeiung) wurde von dem US-amerikanischen Soziologen Merton (1948, 1968) geprägt, der selbsterfüllende Prophezeiungen als „in the beginning, a false definition of the situation evoking a new behavior which makes the originally false conception become true“ (Merton, 1968, S. 477) beschreibt. Verhaltensändernde Erwartungseffekte können einerseits positive (*Pygmalion/Galatea*) und andererseits negative (*Golem*) Konsequenzen hervorrufen (vgl. u. a. Babad, Inbar & Rosenthal, 1982): Pygmalion-Effekte gehen im schulischen Kontext somit von einer überhöhten Erwartungshaltung der Lehrkraft aus und können entsprechend zu einer Verbesserung z. B. der Schülerleistungen führen. Golem-Effekte hingegen entstehen

durch zu niedrige Lehrkrafterwartungen, die in der Folge z.B. zu einer Leistungsminderung führen können.¹⁵

Grundsätzliche Ausgangslage einer selbsterfüllenden Prophezeiung ist in beiden Fällen die falsche Einschätzung der ursprünglichen Situation und daraus resultierenden falschen Erwartungen (vgl. z. B. Jussim, 1990; Madon, Jussim & Eccles, 1997; Madon, Willard, Guyll & Scherr, 2011). So können „Lehrererwartungen ... nur dann einen eigenständigen Erklärungsbeitrag für die Leistungsentwicklung liefern, wenn sie nicht perfekt mit dem tatsächlichen Leistungsstand der Schüler korrelieren“ (Schrader & Helmke, 1987, S. 29). Fehlerhafte Lehrkrafterwartungen an Schülerleistungen weichen dabei systematisch von den tatsächlichen Vorleistungen ab, während inakkurate Lehrkrafturteile allgemein auch zufällige Abweichungen beinhalten können (Timmermans, Kuyper & van der Werf, 2015). Fehlerhafte Lehrkrafterwartungen sind somit immer inakkurat, inakkurate Erwartungen sind dagegen nicht automatisch fehlerhaft (vgl. Ready & Wright, 2011). Dementsprechend können herkunftsabhängige Erwartungen, die für den einzelnen Lernenden der entsprechenden Herkunft jedoch unzutreffend sind, als systematische Abweichungen und damit als fehlerhafte Lehrkrafterwartungen angesehen werden. Fehlerhafte Erwartungen können zwar durch die Lehrkraft erkannt und hierdurch unwirksam werden (Jussim et al., 1996). Problematisch ist jedoch, dass Erwartungen „mit einem hohen Grad subjektiver Gewissheit verbunden [sind], was i. d. R. die Handlungssicherheit erhöht“ (Schrader & Helmke, 2001), jedoch auch die Reflexion und das in Frage stellen der eigenen Erwartungen reduziert.

Dabei hat die Wirkungsweise einer falschen Erwartung einen bestimmten Ablauf um als selbsterfüllende Prophezeiungen definiert zu werden (vgl. Jussim & Fleming, 1996). Nach Good und Brophy (2003) sowie Darley und Fazio (1980) ist das Verhalten einer Person durch die vorher gebildeten (falschen) Erwartungen beeinflusst. Dieses Verhalten wird durch das Erwartungsobjekt wahrgenommen und ebenfalls interpretiert, was Auswirkungen auf die Wahrnehmung der eigenen Person haben kann und sich entsprechend im eigenen Selbstkonzept niederschlagen kann (vgl. Fazio, Effrein & Falender, 1981). Wenn die Ursachen des ihm entgegengebrachten Verhaltens auf sich selbst bezogen werden, kann sich dies entsprechend auch in erwartungskonformem Verhalten zeigen (Brophy & Good, 1974; Darley & Fazio, 1980).

Auswirkungen von Erwartungen wurden für den pädagogischen Kontext dabei v. a. durch Rosenthal und Jacobson (1968, 1992) relevant.^{16, 17} Ihre Studie zu Erwartungseffekten an einer Grundschule in den USA stellt immer noch eine bedeutsame Basis für weitere Forschungen zu

¹⁵ Anzumerken ist zudem, dass Erwartungen auch zu einer entgegengesetzten Reaktion führen können (*self-disconfirming prophecies*). Durch eine erwartungswidrige Veränderung des Verhaltens des Erwartenden kommt es zu einer der Erwartung entgegengesetzten Reaktion des Erwartungsobjekts (Miller & Turnbull, 1986).

¹⁶ Die wissenschaftliche Auseinandersetzung mit diesem Phänomen erfolgte jedoch bereits in den 1950er Jahren im Zusammenhang mit der psychologischen Forschungspraxis, unter der Fragestellung, welchen Einfluss der Versuchsleiter auf das Ergebnis eines Experiments (*Versuchsleitereffekt*) hat (vgl. z. B. Ludwig, 1991; Rosenthal, 1994). Dabei präsentierten Rosenthal und Fode (1963) ebenfalls Befunde zu Erwartungseffekten in Experimenten mit Laborratten, die einen Weg durch ein Labyrinth finden sollten.

¹⁷ Im medizinischen Bereich werden vergleichbare Effekte bereits seit dem 18. Jahrhundert thematisiert (vgl. Cullen, 1797). Neben dem positiven Placebo-Effekt existiert zudem der negative Nocebo-Effekt (z. B. Enck, Benedetti & Schedlowksi, 2008).

Erwartungseffekten im schulischen Kontext dar. In dieser Untersuchung zeigte sich, dass in den ersten beiden Jahrgangsstufen Schüler(innen), die zuvor zufällig ausgewählt wurden und für die den Lehrkräften ein Zuwachs der intellektuellen Fähigkeiten prognostiziert wurde, tatsächlich einen signifikanten Anstieg der kognitiven Fähigkeiten im Vergleich zur Kontrollgruppe aufwiesen (Rosenthal & Jacobson, 1968, 1992). Jedoch ist anzumerken, dass diese Studie zu äußerst kritischen Kontroversen führte (vgl. z.B. Elashoff & Snow, 1972; Spitz, 1999; Thorndike, 1968) und Replikationsversuche die Befunde zum Effekt selbsterfüllender Prophezeiungen nicht konsistent bestätigen konnten (z.B. Claiborn, 1969), wobei jedoch teilweise größere methodische Abweichungen vorlagen. Andere Forschergruppen konnten hingegen die Befunde zu Erwartungseffekten in den ersten beiden Jahrgangsstufen bestätigen (Meichenbaum, Bowers & Ross, 1969; Raudenbush, 1984; Rist, 1970; Zanna, Sheras, Cooper & Shaw, 1975). Uneinheitliche Definitionen und Operationalisierungen des Erwartungskonstrukts oder methodische Zugänge können die heterogene Befundlage teilweise erklären (vgl. Cooper, Findley & Good, 1982). Trotz einer teilweisen inkonsistenten Befundlage und eher kleinen Effekten (vgl. Hoge, 1984; Jussim & Harber, 2005) kann dennoch davon ausgegangen werden, dass Lehrkrafterwartungen einen nicht zu vernachlässigenden Einfluss auf das Lernen und damit auch den Lernerfolg der Schüler(innen) haben können (z. B. Hattie, 2015).

Verzerrungen von Lehrkrafterwartungen können dabei allgemein oder spezifisch sein, d. h. eine Lehrkraft hat entweder eine generell zu hohe oder niedrige Erwartung an die eigenen Schüler(innen) oder sie hat nur für bestimmte Schülergruppen oder einzelne Schüler(innen) zu hohe oder zu niedrige Erwartungen (Timmermans et al., 2015). Doherty und Conolly (1985) berichten, dass (unter Kontrolle der tatsächlichen Leistungen) an Jungen, die als besonders ordentlich wahrgenommen wurden, höhere Leistungserwartungen gerichtet wurden als an Jungen, die als weniger ordentlich wahrgenommen wurden. Bei Mädchen ließ sich hingegen kein Unterschied in den Erwartungen in Abhängigkeit der wahrgenommenen Ordentlichkeit feststellen (Doherty & Conolly, 1985). Auch wenn Erwartungen an sich nicht problematisch sein müssen, können sie durch die Verschränkung mit sozialen Kategorien und damit assoziierten Stereotypen zu herkunftsassoziierten Disparitäten führen (Jussim, 1990).

Jussim und Harber (2005) gehen davon aus, dass Erwartungen häufig akkurat sind und daher Schülerleistungen meist akkurat voraussagen (vgl. auch Jussim et al., 1996). Sie stellen jedoch auch fest, dass es – trotz eher kleiner Effektstärken – zu selbsterfüllenden Prophezeiungen im schulischen Kontext kommt. Es bleibt jedoch unklar, ob sie eher einen positiven oder negativen Effekt haben (Jussim & Harber, 2005). Dennoch kommen sie zu dem Schluss, dass insbesondere bei stigmatisierten Schülergruppen stärkere Effekte selbsterfüllender Prophezeiungen zum Tragen kommen können (vgl. z.B. Jussim et al., 1996; Jussim & Harber, 2005; Sorhagen, 2013; McKown & Weinstein, 2002). Insbesondere bei männlichen Schülern, die einer ethnischen Minorität angehören, traten die negativsten Erwartungseffekte bei zu niedrigen Erwartungen der Lehrkraft auf, aber ebenso der höchste Leistungszuwachs bei einer Überschätzung durch die Lehrkraft (Hinnant, O'Brien & Ghazarian, 2009). Erwartungseffekte sind dabei nicht nur bei stigmatisierten sozialen Gruppen besonders wirksam, sondern auch bei Schüler(innen) mit niedrigen Leistungen (Madon et al., 1997).

Somit kann von einem additiven Effekt bestimmter Schülermerkmale ausgegangen werden (Jussim et al., 1996). Da Schüler(innen) mit Migrationshintergrund und/oder niedrigem

sozioökonomischem Status häufig niedrigere schulische Leistungen aufweisen (vgl. Kapitel 2), ist anzunehmen, dass diese Schüler(innen) demnach besonders empfänglich für den Einfluss von Lehrkrafterwartungen sind. McKown und Weinstein (2003) zeigten, dass Kinder bereits im Grundschulalter – insbesondere aus stigmatisierten Gruppen – Stereotype anderer Personen wahrnehmen und in diagnostischen Situationen schlechter abschneiden als Kinder, die weniger Bewusstheit über die Stereotype anderer haben (vgl. auch *stereotype threat effect*; z.B. Steele, 1997). Bedenklich sind zudem die Befunde von Crano und Mellon (1978) die ebenfalls darauf hindeuten, dass Lehrkräfteeinstellungen und daraus resultierende Erwartungen einen größeren Einfluss auf Schülerleistungen haben, als Schülerleistungen auf Lehrkräfteeinstellungen.

Theoretische Ansätze zur Klärung, wie Lehrkrafterwartungen sich auf Schülerleistungen und -verhalten auswirken können, gehen dabei zunächst von der entsprechenden Erwartungsbildung der Lehrkräfte aus, die Transmission der Erwartung durch entsprechendes Verhalten, die Absorption und Interpretation des Verhaltens durch den Schüler bzw. die Schülerin und abschließend die Entstehung des entsprechenden Schülerverhaltens. (vgl. u. a. Babad, 1992; Cooper, 1979). Das Verhalten der Lehrkraft wird somit maßgeblich durch ihre Erwartungen (z. B. Meichenbaum et al., 1969) und Überzeugungen (z. B. Swann & Snyder, 1980) beeinflusst. Nach der Vier-Faktoren-Theorie der Kommunikation von Erwartungseffekten (*four-factor-theory of the communication of expectancy effects*) von Rosenthal (u. a. 1993, 1994), sind (1) Klima, (2) Feedback, (3) Input und (4) Output maßgeblich für das Zustandekommen von Erwartungseffekten im schulischen Kontext verantwortlich. Rosenthal (1974, 1993) nimmt dabei an, dass Lehrkräfte gegenüber Schüler(inne)n denen sie hohe Erwartungen entgegenbringen...:

- (1) ... ein wärmeres sozioemotionales Klima schaffen,
- (2) ... ein deutlich differenzierteres Feedback geben,
- (3) ... ebenso quantitativ wie qualitativ mehr Inhalte vermitteln und
- (4) ... mehr Zeit und Möglichkeiten für die Beantwortung von Fragen bereits stellen.

In empirischen Untersuchungen sind eine Vielzahl von Variablen identifiziert worden, die Lehrkrafterwartungen und Schülerleistung und -verhalten mediiieren (vgl. Babad, 1992; Brophy, 1983; Brophy & Good, 1974; Harris & Rosenthal, 1985). Das Klima und die konkrete Interaktion durch Lob und Feedback der Lehrkraft spielen eine entscheidende Rolle, jedoch auch emotionale Aspekte und nonverbale Verhaltensweisen sind von hoher Relevanz (vgl. Babad, 1992; Harris & Rosenthal, 1985; Rosenthal, 1974; Rubovits & Maehr, 1971; Trouilloud, Sarrazin, Bressoux & Bois, 2006; Weinstein, 2002). Lehrkräfte bauen für Schüler(innen), an die sie eher geringere Erwartungen haben, zwar weniger Druck auf, diese Schüler(innen) erhalten jedoch auch weniger emotionale Unterstützung als Schüler(innen), für die Lehrkräfte höhere Erwartungen haben (Babad, 1990). Dieses unterschiedliche Lehrkraftverhalten wird dabei auch von den Lernenden wahrgenommen (Weinstein, 2002). Lehrkräfte fordern bessere Leistungen und stellen anspruchsvollere Fragen bei Schüler(inne)n, für die sie hohe Erwartungen haben und loben diese bei entsprechend guten Leistungen mehr (Brophy & Good, 1970; Rubie-Davis, 2007). Bei Schüler(inne)n, denen gegenüber sie eher niedrigere Erwartungen

haben, werden auch geringere Leistungen schneller akzeptiert, zudem erfolgt weniger Lob bei guten Leistungen (Brophy & Good, 1970).

Neben mediiierenden Mechanismen ist zu vermuten, dass auch moderierende Faktoren die Stärke von Erwartungseffekten beeinflussen. Nach Madon et al. (1997) spielt dabei u. a. auch das Selbstkonzept der Schüler(innen) sowie ihre Vorleistungen eine Rolle, aber auch die Über- oder Unterschätzung durch die Lehrkraft. Schüler(innen), deren Leistungen durch die Lehrkraft unterschätzt werden, weisen ein geringeres Selbstkonzept und zudem stärkere negative emotionale und motivationale Ausprägungen auf (Zhu & Urhahne, 2015). Aber auch die Interpretation der Lehrkraftinteraktion kann altersabhängig unterschiedlich erfolgen (Lord, Umezaki & Darley, 1990). Die Bedeutung von direkten Erwartungseffekten nimmt mit zunehmendem Alter der Kinder ab, ein indirekter Effekt – vermittelt über die intraindividuelle Erwartung der Schüler(innen) – zeigt sich dagegen in höheren Klassenstufen der Grundschule (Kuklinski & Weinstein, 2001).

Verstärkt werden können Erwartungseffekte somit auch dadurch, dass die Interaktion zwischen Lernendem und Lehrenden nicht getrennt vom Klassenkontext und damit von den Mitschüler(inne)n betrachtet werden kann (vgl. Darley & Oleson, 1993). Nicht nur das betroffene Kind selbst, sondern auch seine Mitschüler(innen) können das erwartungskongruente Lehrkraftverhalten beobachten und ziehen hieraus Schlüsse über die Fähigkeiten des Kindes. Dadurch könnten sich die Einschätzungen und Erwartungen der Lehrkraft gegenüber den jeweiligen Schüler(inne)n auch auf Mitschüler(innen) übertragen. Ein Kind, dem hohe Lehrkrafterwartungen entgegengebracht werden, wird vor der Klasse vermutlich häufiger gelobt. Dies nehmen auch die Mitschüler(innen) wahr und schließen daraus hohe Fähigkeiten. Bei schulischen Problemen kann dieses Kind wegen der Erwartungen der Mitschüler(innen) zur Anlaufstelle für Fragen werden. Dies kann sich positiv auf das Fähigkeitsselbstkonzept des Kindes auswirken und Erwartungseffekte verstärken. Dieser Prozess ist allerdings nicht nur bei positiven Erwartungen wirksam.

Der Einfluss von Lehrkrafterwartungen auf Schülerleistungen ist dabei jedoch nicht unumstritten. Goldenberg (1992) verweist basierend auf einer qualitativen Felduntersuchung z. B. auf einen paradoxen Erwartungseffekt (hohe Lehrkrafterwartungen gingen mit schlechteren Leistungen, niedrigere Erwartungen mit besseren Leistungen einher). Anzumerken ist jedoch, dass der methodische Ansatz – der Vergleich zweier Einzelfälle – keine Generalisierung der Ergebnisse erlaubt. Hattie (2015) verweist in seiner Meta-Analyse hingegen auf eine mittlere Effektstärke ($d = 0.43$) der Lehrkrafterwartungen auf das Lernen der Schüler(innen). Alvidrez und Weinstein (1999) konnten zudem ebenso wie Babad et al. (1982) aufzeigen, dass v. a. negative Erwartungen stärkere Erwartungseffekte hervorrufen als positive Erwartungen.

Zusammenfassend lässt sich festhalten, dass Lehrkrafturteile und -erwartungen anfällig für Verzerrungen sind (vgl. u. a. Heller & Hany, 2001). Unter bestimmten Bedingungen können falsche Erwartungen zu positiven oder negativen Erwartungseffekten führen. Dünnebier, Gräsel und Krolak-Schwerdt (2009) konnten z. B. Ankereffekte bei der Beurteilung von Schülerleistungen nachweisen. Ebenso wie diagnostische Urteile werden auch Erwartungen an Leistungen und Verhaltensweisen der Schüler(innen) von weiteren Faktoren beeinflusst, die nicht immer in direktem Zusammenhang mit den tatsächlichen Fähigkeiten stehen. Ausgehend von

der Annahme, dass Erwartungen Auswirkungen auf die Wahrnehmung, das Gedächtnis sowie die Urteilsbildung haben (vgl. z. B. Schwarz, Pfister & Büchel, 2016), ist anzunehmen, dass sie ebenfalls im schulischen Alltag im Kontext einer heterogenen Schülerschaft von Bedeutung sind. Lehrkrafterwartungen an Schüler(innen) können dabei auch in Anbetracht unterschiedlicher sozialer Kategorien differieren, z. B. in Bezug auf das Geschlecht von Kindern und Jugendlichen (vgl. Ludwig & Ludwig, 2007; Timmermans et al., 2015). Timmermans, de Boer und van der Werf (2016) verweisen jedoch darauf, dass sich unterschiedliche Lehrkrafterwartungen zumindest zum Teil durch differenzielle Verhaltensweisen der Mädchen und Jungen erklären lassen. Diese Schülermerkmale (z. B. Selbstvertrauen und Arbeitsverhalten) könnten auch durch stereotype Zuschreibungen zu sozialen Gruppen eine Bedeutung bei der Erwartungsbildung haben. Somit ist auch denkbar, dass sich Lehrkrafturteile und -erwartungen auch in Abhängigkeit der ethnischen und sozialen Herkunft von Schüler(inne)n unterscheiden. Da Erwartungseffekte insbesondere bei stigmatisierten Gruppen wirksam sind, könnte dies ein weiterer Erklärungsansatz für herkunftsassoziierte Disparitäten im Bildungssystem sein. Da Erwartungen in einem engen Zusammenhang mit Lehrkrafturteilen stehen, ist ebenfalls davon auszugehen, dass sich auch Lehrkrafturteile in Abhängigkeit der ethnischen und sozialen Herkunft von Schüler(innen) unterscheiden. Im folgenden Abschnitt werden daher empirische Befunde zu herkunftsassoziierten Urteilen und Erwartungen von Lehrkräften zusammengefasst.

3.3 Herkunftsassoziierte Lehrkrafturteile und -erwartungen

Wie in den beiden vorhergehenden Abschnitten gezeigt werden konnte, weisen theoretische Annahmen und empirische Befunde darauf hin, dass auch leistungsirrelevante Informationen einen Einfluss auf Urteile und Erwartungen von Lehrkräften haben können. In einer Studie von Böhmer et al. (2015) zur Übergangentscheidung von Lehrkräften, gaben diese an, dass die elterlichen Berufe (6. Rang) und der Migrationshintergrund (10. Rang) von Schüler(inne)n für sie eher unwichtige Informationen bei der Entscheidung für eine weiterführende Schulform darstellen. Die Lehrkräfte berichteten jedoch, dass neben dem Arbeitsverhalten der Schüler(innen) v. a. die elterliche Unterstützung und der Schulwunsch der Eltern bedeutsame Informationen für sie darstellen (Böhmer et al., 2015). Diese Studie zeigte zwar eine nachrangige Bedeutung der ethnischen und sozialen Herkunft als konkrete Information für die Übergangentscheidung, dennoch spielten Informationen über die Eltern eine wichtige Rolle. Die Schülerherkunft könnte jedoch einen indirekten Einfluss darauf haben wie Lehrkräfte z. B. die Unterstützung der Familie wahrnehmen oder interpretieren.

Empirische Studien konnten allerdings nicht nur Befunde erbringen, die einen indirekten Einfluss der Schülerherkunft auf Urteile und Erwartungen von Lehrkräften vermuten lassen, sondern auch direkte Herkunftseffekte feststellen. Den Einfluss der sozialen Herkunft konnten Darley und Gross (1983) in einer experimentellen Untersuchung verdeutlichen. Dabei wurde die Leistung eines Mädchens, unter der experimentellen Manipulation eines niedrigen sozioökonomischen Status, schlechter eingeschätzt als bei der Manipulation eines hohen sozioökonomischen Status (Darley & Gross, 1983). Tom, Cooper und McGraw (1984) konnten ebenfalls in einer experimentellen Untersuchung zeigen, dass Lehrkräfte höhere Erwartungen an Schüler(innen) der Mittelschicht als an Schüler(innen) der Unterschicht hatten. Insbesondere die Bil-

dungsnähe der Eltern als soziales Merkmal erweist sich als besonders relevant. Schüler(innen), deren Eltern einen höheren Bildungsabschluss aufweisen, erhielten in einer niederländischen Untersuchung ebenfalls deutlich häufiger die Empfehlung für eine höhere weiterführende Schule als Schüler(innen), deren Eltern einen mittleren Bildungsabschluss hatten (Timmermans et al., 2015).

Aber nicht nur für schulische Leistungen zeigen sich Urteils- und Erwartungsunterschiede in Bezug auf die soziale Herkunft von Schüler(inne)n. Auch die Intelligenz von Schüler(inne)n wird umso höher eingeschätzt, je höher der sozioökonomische Status der Familien ist (Alvidrez & Weinstein, 1999). Dunkake und Schuchart (2015) berichten, dass Lehramtsstudierende unterschiedliche Erwartungen an Verhaltensweisen der Schüler(innen) in Abhängigkeit ihrer sozialen Herkunft haben. So wurden von Schüler(inne)n mit niedrigem sozioökonomischem Status negativere Verhaltensweisen erwartet als von Schüler(inne)n mit mittlerem sozioökonomischem Status. Informationen zum sozialen Hintergrund wurden jedoch v. a. bei inkonsistenten Schülerinformationen zur Urteilsbildung herangezogen. Als schwierig erweist sich ein Vergleich der Ergebnisse zur sozialen Herkunft von Schüler(inne)n, da diese in verschiedenen Studien unterschiedlich operationalisiert wurden. Kaiser et al. (2015) konnten unter Berücksichtigung des kulturellen Kapitals (vgl. Bourdieu, 1982; operationalisiert durch die Anzahl der Bücher im familiären Haushalt) keinen Effekt auf die Urteilsgenauigkeit von Referendaren feststellen. Aber auch in einer Untersuchung von Karing, Matthäi und Artelt (2011) zeigte sich nach der Operationalisierung der sozialen Herkunft durch den HISEI kein signifikanter Einfluss der sozialen Herkunft auf die Lehrkrafturteile.

Auch in Bezug auf die soziale Herkunft erweist sich die Befundlage als heterogen. Der Migrationshintergrund ist in einigen Untersuchungen ein bedeutsamer Einflussfaktor auf Lehrkrafturteile und -erwartungen (z. B. Überblick bei Kaiser et al., 2016; Riley & Ungerleider, 2012; van den Bergh, Denessen, Hornstra, Voeten & Holland, 2010). Andere Studien finden wiederum keine Hinweise auf einen Einfluss der Ethnizität auf Urteile und Erwartungen. In einer experimentellen Studie, in der angehende Lehrkräfte zukünftige schulische und berufliche Erfolge eines Schülers anhand eines Aufsatzes einschätzen sollten, konnten Chen und Weseley (2011) keine ethnizitätsspezifischen Erwartungsunterschiede feststellen. Methodisch ist dabei jedoch anzumerken, dass das Stimulusmaterial in dem die Ethnizität des Schülers manipuliert wurde sich u. a. auch aktiv mit dem Thema ethnisch-kulturelle Diversität auseinandersetzte. Dabei könnte es auch zu einem Priming-Effekt gekommen sein, der die Informationsverarbeitung bei der Bearbeitung beeinflusst haben könnte. Chen und Weseley (2011) konnten zwar keine ethnizitätsabhängigen Erwartungsunterschiede finden, jedoch zeigten sich signifikant höhere Urteile in der Einschätzung der Intelligenz für asiatische Schüler(innen) im Vergleich zur Kontrollgruppe, jedoch nicht im Vergleich zu afroamerikanischen oder europäischstämmigen Schüler(inne)n. Für den US-amerikanischen Raum weisen zahlreiche Befunde auf höhere Erwartungen für europäischstämmige Schüler(innen) hin, die nur noch von Schüler(inne)n asiatischer Abstammung übertroffen werden.

Zudem konnten in zahlreichen Studien negativere Lehrkrafturteile gegenüber latein- und afroamerikanischen Minoritäten im Vergleich zur europäischstämmigen Majorität berichtet werden (z. B. Diamond, Randolph & Spillane, 2004; Rubovits & Maehr, 1973; Sbarra & Pianta, 2001; Tom et al., 1984). Tenenbaum und Ruck (2007) konnten dies in einer Meta-Analyse be-

stätigen. Dabei hatten Lehrkräfte die höchsten Erwartungen an Schüler(innen) asiatischer Abstammung und die geringsten Erwartungen an Schüler(innen) mit latein- oder afroamerikanischen Wurzeln (Tenenbaum & Ruck, 2007). Rubie-Davies, Hattie und Hamilton (2006) konnten bei neuseeländischen Lehrkräften, auch nach Kontrolle der tatsächlichen Leseleistungen, geringere Erwartungen an Schüler(innen) mit einem Maori-Hintergrund im Vergleich zu Schüler(inne)n der ethnischen Majorität und anderen ethnischen Gruppen aufdecken (vgl. auch Meissel, Meyer, Yao & Rubie-Davis, 2017). In dieser Untersuchung zeigte sich ferner, dass Schüler(innen) mit einem Maori-Hintergrund auch die geringsten Leistungszuwächse im Verlauf eines Schuljahres zeigten, was als möglicher Effekt der Lehrkrafterwartungen interpretiert werden kann (vgl. Rubie-Davis et al., 2006).

Für den deutschsprachigen Raum liegt ebenfalls eine heterogene Befundlage zu ethnizitätsabhängigen Urteils- und Erwartungsunterschieden vor. In einer Untersuchung von Sprietsma (2013) konnte der Einfluss der ethnischen Herkunft auf die Urteilsbildung von Lehrkräften aufgedeckt werden. Lehrkräfte beurteilten Aufsätze (deren Vorname systematisch nach Schülerherkunft variiert wurde) von Kindern mit Migrationshintergrund schlechter. Zudem erhielten Schüler mit türkischem Migrationshintergrund seltener eine Empfehlung für das Gymnasium (Sprietsma, 2013). Bonefeld, Dickhäuser, Janke, Praetorius und Dresel (2017) zeigten für Schüler(innen) mit Migrationshintergrund unter Kontrolle der tatsächlichen Schülerleistungen im Fach Mathematik ebenfalls schlechtere Noten als für ihre Mitschüler(innen) ohne Migrationshintergrund. Wobei diese Effekte auch über den zeitlichen Verlauf von zwei Schuljahren stabil blieben (Bonefeld et al., 2017). Kaiser et al. (2016) konnten hingegen keine Unterschiede der Lehrkrafturteile in Bezug auf die Ethnizität der Schüler(innen) feststellen. Jedoch verweisen die Befunde auf einen bedeutsamen Einfluss des Minoritätenstatus.

In einer experimentellen Untersuchung mit erfahrenen Lehrkräften konnte Glock (2016) ebenfalls Urteilsunterschiede in Bezug auf die Schülerethnizität feststellen. Jedoch hatte die (experimentell manipulierte) hohe oder niedrige Leistung des männlichen Schülers in den Fallvignetten einen bedeutsamen Einfluss. Lehrkräfte beurteilten das Engagement des Schülers mit Migrationshintergrund im Vergleich zu dem Schüler ohne Migrationshintergrund nur höher, wenn beide Schüler überdurchschnittliche Leistungen (im Vergleich zu unterdurchschnittlichen Leistungen) hatten (Glock, 2016). Für die Beurteilung der Mathematikleistung zeigten sich in dieser Untersuchung keine ethnizitätsabhängigen Unterschiede in den Lehrkrafturteilen. In Bezug auf die Sprachkenntnisse wurde der Schüler mit Migrationshintergrund ausschließlich in der experimentellen Bedingung unterdurchschnittlicher Leistungen signifikant schlechter beurteilt als der Schüler ohne Migrationshintergrund. In einer experimentellen Untersuchung mit Lehramtsstudierenden konnte eine negativere Beurteilung von Schüler(inne)n mit türkischem Migrationshintergrund unter der stereotypenkonformen Bedingung (unterdurchschnittliche Leistungen) bestätigt werden (Glock & Krolak-Schwerdt, 2013).

Jedoch lassen sich nicht nur bei der Beurteilung von Leistungen ethnizitätsabhängige Unterschiede finden. Auch bei der Ursachensuche für Schülerleistungen sind herkunftsassoziierte Unterschiede aufzufinden. So schreiben Lehrkräfte schlechtere aber auch bessere schulische Leistungen von afroamerikanischen Kindern eher externalen Faktoren zu (Wiley & Eskilson, 1978). Downey und Pribesh (2004) konnten für Kindergartenkinder sowie Sekundarstufenschüler(innen) aufzeigen, dass Vorschullehrkräfte und Sekundarschullehrkräfte afro-

amerikanische Vorschul- und Schulkinder in ihrem Verhalten negativer einschätzten. Unter Berücksichtigung der Ethnizität der Lehrkräfte zeigten sich hingegen keine signifikanten Unterschiede mehr (Downey & Pribesh, 2004). Es ist somit davon auszugehen, dass Schüler(innen), die nicht zur selben ethnischen Gruppe der Lehrkräfte gehören, auf eine andere Art beurteilt werden. Harber (1998) berichtete, dass europäischstämmige Lehrkräfte bei schlechten Aufsätzen Schüler(inne)n der gleichen Ethnizität kritischeres Feedback geben als afroamerikanischen Schüler(inne)n. Dies deutet zumindest auf ethnizitätsabhängige Verhaltensweisen der Lehrkräfte hin. Gershenson, Holt und Papageorge (2015) zeigten, dass europäischstämmige Lehrkräfte im Vergleich zu afroamerikanischen Lehrkräften von afroamerikanischen Schüler(inne)n 12 % weniger häufig erwarten, dass sie einen College Abschluss erreichen. Dee (2005) berichtet, dass Lehrkräfte Schüler(innen) die eine andere Ethnie als sie selbst hatten, als signifikant unaufmerksamer und störender wahrnahmen. Diese Befunde aus dem US-amerikanischen Raum verweisen zwar auf unterschiedliche Urteile und Erwartungen von Lehrkräften, deuten jedoch darauf hin, dass in diesem Zusammenhang auch die Ethnizität der Lehrkraft relevant ist. Ein möglicher Erklärungsansatz hierzu lässt sich in sozialpsychologischen Theorien finden, die von einer Bevorzugung der Eigengruppe (*ingroup*) ausgehen.

Tom et al. (1984) berücksichtigten neben der ethnischen Herkunft, das Geschlecht der Schüler(innen) und konnten z.B. zeigen, dass Lehrkräfte höhere Erwartungen an Jungen asiatischer Abstammung sowie an Mädchen weißer Herkunft hatten. Zwar wurde in dieser Studie nicht der sozioökonomische Hintergrund der Lernenden berücksichtigt, jedoch verdeutlichen die Ergebnisse die bedeutsame Verschränkung sozialer Kategorien. Die Verschränkung von Ethnizität und Geschlecht, aber auch von Ethnizität und Attraktivität bei der Urteilsbildung konnte in einer experimentellen Studie von Parks und Kennedy (2007) bestätigt werden. So schrieben Lehrkräfte unattraktiven, afroamerikanischen Jungen die geringste Kompetenz zu (Parks & Kennedy, 2007).

Da insbesondere Effekte des Migrationshintergrunds auch durch die soziale Schicht erklärt werden könnten, sollten die Befunde zu ethnizitätsspezifischen Urteils- und Erwartungsunterschieden von Lehrkräften nur mit Vorsicht interpretiert werden. Eine gemeinsame Berücksichtigung beider Merkmale sollte entsprechend erfolgen. Für den US-amerikanischen Raum konnten bereits einige empirische Befunde zu herkunftsassoziierten Erwartungsunterschieden von Lehrkräften und Lehramtsstudierenden, unter der Berücksichtigung der ethnischen und sozialen Herkunft, erbracht werden (z. B. Cooper, Baron & Lowe, 1975). Meist lässt sich ein konsistentes Bild der höheren Erwartungen für Schüler(innen) der ethnischen Majorität und/oder hohem sozioökonomischem Status bzw. geringere Erwartungen für Schüler(innen) mit afroamerikanischen oder lateinamerikanischen Wurzeln und/oder niedrigem sozioökonomischem Status finden. In einer Studie von Harvey und Slatin (1975) wurde zwar der sozioökonomische Hintergrund und die Ethnizität der Schüler(innen) berücksichtigt, jedoch wurden beide Merkmale in den Analysen separat betrachtet. Dabei lassen sich zwar höhere Erwartungen für Kinder ohne Migrationshintergrund und für Kinder mit hohem sozioökonomischem Status finden, eine Verschränkung beider Merkmale blieb jedoch aus. Befunde einer experimentellen Untersuchung von McCombs und Gay (1988) zeigten unter Berücksichtigung beider Merkmale negativere Lehrkrafturteile für lateinamerikanische Schüler(innen) sowie für Schüler(innen) mit niedrigem sozioökonomischem Status. Nachdem die Lehrkräfte

jedoch zusätzlich Informationen über die kognitiven Fähigkeiten der Kinder erhielten, blieb nur der Effekt der ethnischen Herkunft bestehen.

De Boer, Bosker und van der Werf (2010) konnten für niederländische Grundschulen eine negative Verzerrung der Lehrkrafterwartung bei niedrigem sozioökonomischem Status der Schüler(innen) aufdecken, die ethnische Herkunft hingegen hatte keinen signifikanten Einfluss. Eine Untersuchung aus den Niederlanden zu Übergangsempfehlungen zeigte, dass an Schüler(innen) aus niedrigen sozialen Schichten höhere Erwartungen gerichtet wurden, wenn sie einen türkischen oder marokkanischen Hintergrund besaßen als wenn sie niederländische Wurzeln hatten (Timmermans et al., 2015). Die Verschränkung von Ethnizität und sozialer Herkunft muss somit nicht zwangsweise zu einer kumulierten Benachteiligung führen. Unklar ist bislang jedoch, welche Ursachen zu höheren Erwartungen an Schüler(innen) mit niedrigem sozioökonomischem Status und Migrationshintergrund im Vergleich zu Schüler(inne)n ohne Migrationshintergrund führen könnten.

Für Luxemburg konnten Klapproth, Glock, Böhmer, Krolak-Schwerdt und Martin (2012) nach Kontrolle der Leistungen keinen Effekt des sozioökonomischen Status auf das Lehrkrafturteil zum Übergang auf eine weiterführende Schule ermitteln. Jedoch zeigte sich in dieser Studie ein bedeutsamer Einfluss der ethnischen Herkunft von Schüler(inne)n auf das Lehrkrafturteil (vgl. auch Klapproth, Glock, Krolak-Schwerdt, Martin & Böhmer, 2013). Glock et al. (2012) berücksichtigten ebenfalls sowohl die ethnische als auch soziale Herkunft von Schüler(inne)n in zwei experimentellen Studien mit luxemburgischen Grundschullehrkräften. Die Ethnizität der Schüler(innen) hatte lediglich in der ersten Studie und unter der experimentellen Bedingung einer geringen Verantwortung für die Entscheidung, einen Einfluss auf die Übergangentscheidung. Die soziale Herkunft hatte hingegen in keiner der beiden Studien einen Einfluss auf die Übergangentscheidung (Glock et al., 2012).

Für den deutschen Kontext ist dabei noch keine ausreichende empirische Grundlage vorhanden, um den Einfluss der Ethnizität und der sozialen Herkunft von Schüler(inne)n auf die Urteils- und Erwartungsbildung eindeutig zu klären. Im vorschulischen Kontext untersuchte Kratzmann (2013) die Fähigkeitseinschätzung von Erzieherinnen bei Kindergartenkindern unterschiedlicher Herkunft. Unter Kontrolle der sozialen Herkunft und der tatsächlichen Fähigkeiten konnten keine durchgängigen bedeutsamen Differenzen identifiziert werden. Jedoch lässt sich die Einschätzung nicht ohne Einschränkungen auf den schulischen Kontext übertragen. So ist sowohl die inhaltliche Ausgestaltung als auch die zeitliche Intensität der Interaktion zwischen Erzieher(inne)n und Vorschulkindern nicht direkt mit Lehrkräften und Schüler(inne)n vergleichbar. In Grundschulen konnten Lorenz, Gentrup, Kristen, Stanat und Kogan (2016) jedoch ebenfalls herkunftsabhängige Unterschiede feststellen. Unter Kontrolle der tatsächlichen Schülerleistungen (erfasst durch einen Leistungstest) zeigten sich im Fach Deutsch negativere Erwartungen der Lehrkräfte an Schüler(innen) türkischer Herkunft sowie an Kinder mit niedrigem sozioökonomischem Status (Lorenz et al., 2016). Im Fach Mathematik hatten Lehrkräfte negativere Erwartungen an Schüler(innen) mit niedrigem sozioökonomischem Status, in Bezug auf den türkischen Migrationshintergrund konnten hingegen keine Erwartungsunterschiede festgestellt werden (Lorenz et al., 2016). Auch Hoenig und Wenz (2013) konnten in einer experimentellen Studie, in der die ethnische und soziale Herkunft durch Vornamen manipuliert wurde, Effekte der ethnischen und sozialen Herkunft finden.

Jedoch bleibt weiterhin unklar, ob es sich bei den herkunftsassoziierten Unterschieden um eine Benachteiligung von Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status handelt oder ob Schüler(innen) ohne Migrationshintergrund und mit einem hohen sozioökonomischem Status einen Erwartungsvorsprung genießen und hier eine positive Urteilsverzerrung zum Tragen kommt. Bisher fokussierten nur einzelne Untersuchungen auf die Urteilsgenauigkeit im Kontext einer ethnisch und sozial heterogenen Schülerschaft. Für US-amerikanische Vorschulkinder zeigte sich in diesem Zusammenhang jedoch eine signifikante Unterschätzung der sprachlichen Fähigkeiten von Kindern mit niedrigem sozioökonomischem Status (Ready & Wright, 2011). Dieser Befund konnte von Ready und Chu (2015) ebenfalls bestätigt werden. Zudem wurden Vorschulkinder mit hohem sozioökonomischem Status überschätzt (Ready & Chu, 2015). Pit-ten Cate, Krolak-Schwerdt und Glock (2016) berichten für luxemburgische Lehrkräfte von einer geringeren Urteilsakkuratheit bei Schüler(inne)n der ethnischen Minorität im Vergleich zu Schüler(inne)n der Majorität.

In Deutschland sind bislang nur wenige Untersuchungen bekannt, die sich explizit mit der Frage der diagnostischen Kompetenz von Lehrkräften im Zusammenhang mit der ethnischen und sozialen Herkunft von Schüler(inne)n befassen. Hachfeld, Anders, Schroeder, Stanat & Kunter (2010) berichten dabei allgemein von einer Überschätzung der Schüler(inne)n durch die untersuchten Mathematiklehrkräfte. Bei sprachlich komplexen Aufgabenstellungen wurden die bilingualen Schüler(innen) gefolgt von monolingualen Schüler(inne)n mit Migrationshintergrund am meisten überschätzt. Die niedrigste Überschätzung zeigte sich bei Schüler(inne)n ohne Migrationshintergrund (Hachfeld et al., 2010). Diese Befunde lassen sich jedoch nur mit Vorsicht interpretieren. So könnte einerseits die Bilingualität als sprachliche Kompetenz angesehen werden und entsprechend zu einer Überschätzung bei sprachlich komplexen Aufgaben geführt haben (vgl. Hachfeld et al., 2010). Kaiser et al. (2016) berichten, dass Lehrkrafturteile bei ethnischen Minoritäten akkurater urteilen als bei der Majorität. Kaiser et al. (2016) vermuten, dass sich dies durch den Minoritätenstatus und die hierdurch erhöhte Salienz, jedoch nicht durch die spezifische Ethnizität erklären ließe. Beide Studien deuten darauf hin, dass der Migrationshintergrund selbst keine ausreichende Erklärung für Unterschiede in der Urteilsgenauigkeit von Lehrkräften liefert und hierbei weitere Faktoren einen bedeutsamen Einfluss haben. Jedoch wurde dabei die soziale Herkunft der Schüler(innen) vernachlässigt. Zudem zeigte sich, dass die Urteilsgenauigkeit stark von der individuellen Lehrkraft abhängt, da einige u. a. einer starken negativen, andere einer stark positiven Verzerrung unterliegen (z. B. Timmermans et al., 2015).

Einige Untersuchungen verweisen auf bedeutsame Einflussfaktoren, die den Zusammenhang zwischen Schülerherkunft und Lehrkrafturteil moderieren könnten. Glock und Krolak-Schwerdt (2013) konnten in einer experimentellen Studie unter dem Einsatz von Fallvignetten zeigen, dass sich die Urteile der angehenden Lehrkräfte bei besonders guten Schülern nicht in Abhängigkeit der Schülerherkunft unterschieden. Beschrieb die Fallvignette jedoch einen unterdurchschnittlichen Schüler, erwarteten Lehramtsstudierende von einem türkischstämmigen Schüler signifikant schlechtere Leistungen als von einem Schüler ohne Migrationshintergrund. Als mögliche Gründe hierfür wurden eher automatisierte Informationsverarbeitungsprozesse bei erwartungskonsistenten, und kontrollierte Prozesse bei inkonsistenten Informationen vermutet. Zudem konnten Glock, Krolak-Schwerdt und Hörstermann (2016) aufzeigen, dass posi-

tive Einstellungen von Lehramtsstudierenden gegenüber ethnischen Minoritäten auch zu positiveren Urteilen führen können. Glock und Krolak-Schwerdt (2014) konnten in zwei experimentellen Studien zu herkunftsassoziierten Lehrkrafturteilen keine Unterschiede in Bezug auf die ethnische oder soziale Herkunft von Schüler(inne)n finden. Die Befunde verweisen jedoch auf den Einfluss von Stereotypen auf die Wahrnehmung und den Gedächtnisabruf (Glock & Krolak-Schwerdt, 2014).

Zusammenfassend lässt sich festhalten, dass insgesamt eine heterogene Befundlage zu herkunftsassoziierten Lehrkrafturteilen und -erwartungen aufzufinden ist. So lassen sich in einigen Untersuchungen Unterschiede in den Lehrkrafturteilen und -erwartungen sowohl für die ethnische (z.B. Bonefeld et al., 2017; Rubie-Davis et al., 2006; Sprietsma, 2013; Tenenbaum & Ruck, 2007) als auch die soziale Herkunft (z.B. Darley & Gross, 1983; Dunkake & Schuchart, 2015; Tom et al., 1984) von Schüler(inne)n berichten. Dennoch konnten einige andere Studien keine herkunftsabhängigen Urteile und Erwartungen von Lehrkräften finden (z.B. Kaiser et al., 2015; Karing et al., 2011). Bisherige Untersuchungen im Kontext der diagnostischen Kompetenz konnten zudem nicht klären, ob Unterschiede in Lehrkrafturteilen und -erwartungen als Benachteiligung von Schüler(inne)n mit Migrationshintergrund und niedrigem sozioökonomischem Status zu interpretieren sind, oder ob es sich wie in der Untersuchung von Ready und Chu (2015) eher um eine Bevorzugung von Kindern ohne Migrationshintergrund und mit hohem sozioökonomischem Status handelt.

Da Einstellungen und Stereotype einen bedeutsamen Einfluss auf die Informationsverarbeitung sowie einen Einfluss auf das Verhalten haben können, ist denkbar, dass auch herkunftsassoziierte Einstellungen einen Einfluss auf die Urteile und Erwartungen von Lehrkräften in Bezug auf Schüler(innen) spezifischer ethnischer und/oder sozialer Herkunft haben können. Die Befunde von Glock et al. (2016) können hierfür auch als erste empirische Indizien im deutschsprachigen Raum angesehen werden. Diverse Untersuchungen zeigen zudem, dass soziale Kategorien, wie z.B. die ethnische oder soziale Herkunft, bei der Erwartungsbildung keinen zu unterschätzenden Einfluss haben (vgl. z.B. Word, Zanna & Cooper, 1974; Überblick bei Jussim & Fleming, 1996). Dies lässt sich dadurch erklären, dass Stereotype eine bedeutsame Quelle für Erwartungen darstellen (Hamilton et al., 1990; Jussim & Fleming, 1996). Entsprechend ist anzunehmen, dass die Schülerherkunft und damit zusammenhängende Einstellungen und Stereotype einen entscheidenden Einfluss auf die Erwartungsbildung der Lehrkräfte haben kann (Cooper, 1979). Im Folgenden wird daher ein detaillierter Blick auf herkunftsassoziierte Einstellungen und Stereotype geworfen.

4. Einstellungen und Stereotype von Lehrkräften im Kontext einer heterogenen Schülerschaft

Einstellungen und Stereotype können einen bedeutsamen Einfluss auf die Erwartungs- und Urteilsbildung von Lehrkräften haben (vgl. z.B. Miller & Turnbull, 1986; Schrader & Helmke, 2001). Insbesondere im alltagssprachlichen Gebrauch, aber auch im wissenschaftlichen Diskurs ist jedoch nicht immer eine klare Trennung zwischen Einstellungen und Stereotypen ersichtlich. Besonders die Verwendung des Begriffs *Vorurteil* erfolgt häufig ohne klare Abgrenzung zu den anderen beiden Begriffen. Zwar sind inhaltliche Überlappungen der Konstrukte zu konstatieren, dennoch ist eine theoretische Trennung möglich und notwendig.

Abgrenzung von Einstellungen, Vorurteilen und Stereotypen

Eagly und Chaiken (1993) definieren Einstellungen als psychologische Tendenz (interner Zustand), die Ausdruck in einer positiven oder negativen Bewertung (Reaktion) gegenüber einem Einstellungsobjekt findet.¹⁸ Die Bewertungen (z.B. etwas mögen oder nicht mögen) können z.B. als Zustimmung oder Ablehnung, als Annäherung oder Vermeidung ausgedrückt – und erfasst – werden. Allgemein können Einstellungen als kognitive Strukturen zur Organisation von bereits vorhandenem Wissen angesehen werden (Eagly & Chaiken, 1993; Fiske & Linville, 1980). Olson und Fazio (2009) beschreiben Einstellungen als mentale Assoziationen zwischen Bewertungen und Einstellungsobjekten, d.h. Einstellungen sind als Verknüpfungen des Einstellungsobjekts mit einem positiven oder negativen Urteil zu verstehen (vgl. auch Albarracín & Vargas, 2010; Crano & Prislín, 2006; Fiske, 2014). Fazio (2007) bezeichnet Einstellungen entsprechend als summative Evaluationen.

Einige theoretische Ausführungen setzen Einstellungen und Vorurteile weitestgehend gleich, wobei beide Konstrukte die Bewertung eines Objekts beinhalten (z.B. Ajzen & Fishbein, 1980; Eagly & Chaiken, 1993; Fazio, 2007; Tajfel 1969; Thurstone, 1931; Überblick bei Banaji & Heiphetz, 2010). Vorurteile beziehen sich jedoch stets auf soziale Kategorien; Einstellungen sind hingegen nicht auf spezifische Objekte oder Gruppen beschränkt (Eagly & Chaiken, 1993). Da es sich bei Schüler(inne)n unterschiedlicher ethnischer und sozialer Herkunft – die in der vorliegenden Arbeit fokussiert werden – um soziale Gruppen bzw. Kategorien handelt, werden dieser Abgrenzung folgend Einstellungen in der vorliegenden Arbeit als Vorurteile verstanden.

¹⁸ Die wesentlichen Merkmale erster Annäherungen an das Einstellungskonstrukt fasst Allport (1935) unter folgender Einstellungsdefinition zusammen: “An attitude is a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual’s response to all objects and situations with which it is related” (S. 810). Nach Eagly und Chaiken (1993) ist durch diese Definition jedoch kaum eine Abgrenzung zu anderen Konstrukten wie z.B. Gewohnheiten oder Persönlichkeitseigenschaften möglich.

Stereotype haben mit Vorurteilen den Bezug zu sozialen Kategorien gemeinsam. Tajfel (1969) definiert Stereotype als Attribution allgemeiner psychologischer Merkmale zu sozialen Gruppen.¹⁹ Stereotype können als kognitive Assoziationen dieser Merkmale mit sozialen Kategorien aufgefasst werden, ohne zwingend eine positive oder negative Valenz zu beinhalten (Kunda & Thagard, 1996).²⁰ Entsprechend unterscheiden sich Stereotype in diesem wesentlichen Punkt von Vorurteilen und Einstellungen. Greenwald und Banaji (1995) unterscheiden Stereotype von Einstellungen folgendermaßen: „A stereotype is a socially shared set of beliefs about traits that are characteristic of members of a social category. Whereas an attitude implies a consistent evaluative response to its object, a stereotype may encompass beliefs with widely diverging evaluative implications“ (S. 14).

Die assoziierten Merkmale werden als typisch für die soziale Kategorie angesehen. Lehrkräfte können z. B. mit bestimmten Migrantengruppen eine hohe Religiosität verknüpfen, weil sie dies als charakteristisches Merkmal dieser sozialen Kategorie ansehen. Da hohe Religiosität nicht per se eine Bewertung enthält, kann diese Merkmalszuschreibung zu einer bestimmten ethnischen Gruppe als Stereotyp angesehen werden. Zeigen Lehrkräfte jedoch eine positive oder negative Bewertungstendenz gegenüber einer ethnischen Schülergruppe, wird dies in der vorliegenden Arbeit als Einstellung oder Vorurteil angesehen. Stereotype als kognitive Konstrukte hängen aber auch stark mit emotionalen Aspekten zusammen.²¹

Kontrovers wird z. T. diskutiert, ob Vorurteile eine ausschließlich negative Komponente aufweisen oder auch in positiver Form auftreten können (vgl. z. B. Dovidio, Brigham, Johnson & Gaertner, 1996; Zick, 1997). Werden Vorurteile, wie in der vorliegenden Arbeit, nach Eagly und Chaiken (1993) als Einstellungen gegenüber sozialen Kategorien verstanden, kann sowohl eine negative als auch eine positive Tendenz angenommen werden (vgl. auch Zick, 1997). Lehrkräfte können somit sowohl negative als auch positive Vorurteile gegenüber spezifischen ethnischen und sozialen Schülergruppen haben.

Unterschieden werden muss zwischen offenen und verdeckten Vorurteilen bzw. Einstellungen: Offene Vorurteile werden von Pettigrew und Meertens (1995) als „hot, close and direct“ (S. 58) und verdeckte Vorurteile eher als „cool, distant, and indirect“ (S. 58) beschrieben. Offene Vorurteile treten, als Zu- oder Abneigung zu einer bestimmten Gruppe, eher in direkten Äußerungen oder Verhaltensweisen auf. Verdeckte Vorurteile zeigen sich hingegen eher in-

¹⁹ Der Begriff Stereotyp wurde erstmals zum Beginn des 20. Jahrhunderts als Stereotypie für wiederholte und gleichförmige Verhaltensweisen im Bereich der Psychiatrie von Bedeutung (Überblick bei Ashmore & Del Boca, 1981), für die Sozialwissenschaften wurde er von Lippmann in seinem Buch *Public Opinion* (1922) eingeführt. Ursprünglich wurde der Begriff des *Stereotyps* Ende des 18. Jahrhunderts für einen Buchdruckvorgang eingeführt (Überblick bei Peterson & Six, 2008). Dabei hat das Wort seinen Ursprung in den griechischen Wörtern *stereos* (fest oder starr) und *typos* (feste Norm, Entwurf oder charakteristisch Geprägtes) (Peterson & Six, 2008).

²⁰ Tajfel (1969) postuliert in seinem Beitrag *The Cognitive Aspects of Prejudice*, dass die kognitive Komponente von Vorurteilen i. d. R. als Stereotyp bezeichnet wird und Vorurteile hingegen eine positive oder negative Valenz aufweisen. Jedoch sind andere theoretische Strömungen von dieser Trennung nicht überzeugt und sehen Vorurteile als unausweichliche Konsequenz von Stereotypen (die automatisiert Mitgliedern sozialer Gruppen zugewiesen werden) (Überblick bei Devine, 1989).

²¹ Cikara und Fiske (2012) konnten aufbauend auf das *stereotype content model* (Fiske, Cuddy, Glick & Xu, 2002) zeigen, dass Personen höhere Schadenfreude gegenüber Outgroup-Mitgliedern empfinden, wenn diese einen höheren Status haben oder zu konkurrierenden Gruppen gehören.

direkt, z. B. durch Aufwertung der Eigengruppe (vgl. Pettigrew & Meertens, 1995). Verdeckte negative Vorurteile einer Lehrkraft gegenüber z. B. Schüler(inne)n einer ethnischen Minorität zeigen sich somit eher durch die Bevorzugung oder Aufwertung der ethnischen Majorität. Offene Vorurteile von Lehrkräften können hingegen in direkten auf- oder abwertenden Aussagen der ethnischen Minorität resultieren.

Vorurteile bzw. Einstellungen können als hypothetische Konstrukte nicht direkt beobachtet werden, sondern lediglich durch Rückschlüsse aus der Erfassung des Stimulus bzw. Einstellungsobjekts und den (verbalen, behavioralen oder physischen) Reaktionen auf dieses Objekt abgeleitet werden (Eagly & Chaiken, 1993; vgl. auch Rosenberg & Hovland, 1969). Die entweder offenen oder verdeckten Reaktionen auf ein Einstellungsobjekt können sich kognitiv, affektiv oder behavioral ausdrücken (Eagly & Chaiken, 1993; vgl. Abbildung 7).

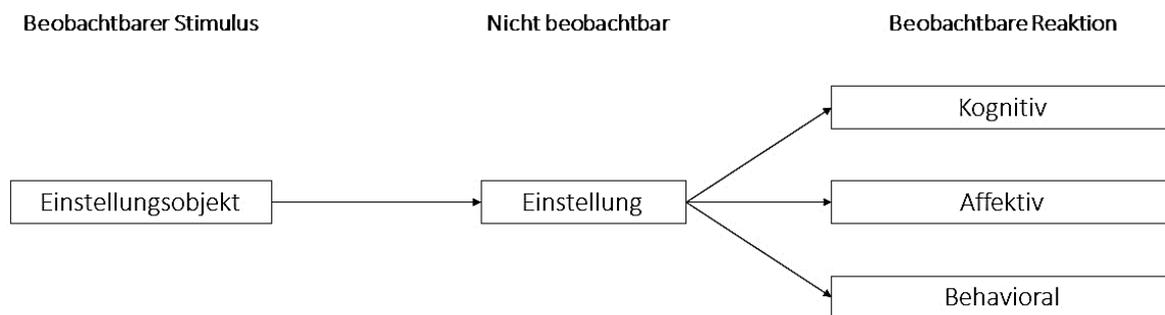


Abbildung 7. Schematische Konzeption von Einstellungen (Eagly & Chaiken, 1993, S. 10)

Einstellungsausprägungen können als Punkte auf einem Kontinuum angesehen werden, die sich in Valenz und Intensität unterscheiden (Banaji & Heiphetz, 2010; Eagly & Chaiken, 1993; Maio & Haddock, 2010; Olson & Fazio, 2009). Sie können somit mehr oder weniger positiv oder negativ sein. Diese bewertenden Tendenzen gegenüber einem sozialen Objekt entwickeln sich aus wiederholten positiven oder negativen Bewertungen des Einstellungsobjekts, die im Gedächtnis als mentale Repräsentation abgespeichert werden (Eagly & Chaiken, 1993).²² Nach Allport (1935) sind sie ebenso wie Stereotype als Produkt aller eigenen sowie beobachteten Erfahrungen (z. B. der Bezugspersonen oder anderer sozialer Quellen) mit dem Einstellungsobjekt zu verstehen (vgl. Ashmore & Del Boca, 1981; Lippmann, 1922).²³ Hierdurch entsteht die psychische Tendenz, bei einer zukünftigen Begegnung mit dem Einstellungsobjekt mit der gespeicherten Bewertung zu reagieren (Eagly & Chaiken, 1993). Unterschiedliche Aspekte können für die inhaltliche Ausprägung und Intensität von ethnischen Stereotypen und Einstellungen bedeutsam sein. Nach Bar-Tal (1997) sind dies: (1) Hintergrundvariablen, wie z. B. das politisch-soziale Klima oder Verhaltensweisen der ethnischen Gruppe; (2) vermit-

²² Inwieweit Einstellungen erlernt oder durch genetische Faktoren bedingt sind, wird im wissenschaftlichen Diskurs z. T. kontrovers diskutiert (vgl. Albarracín & Vargas, 2010; Allport, 1935; Eagly & Chaiken, 1993; Lykken, 1982; Thomas & Znaniecki, 1918; Waller, Kojetin, Bouchard, Lykken & Tellegen, 1990).

²³ Einstellungen Dritter sind nach Allport (1935) nicht nur individuelle Einstellungen von spezifischen Personen, sondern auch gemeinsame Überzeugungen der Gesellschaft, die sich durch gemeinsame ethnische, kulturelle und soziale Einflüsse und Bedingungen ergeben.

telnde Variablen, wie z. B. erzieherische Mechanismen oder der direkte Kontakt zu ethnischen Gruppen; (3) intrapersonelle Variablen, wie z. B. motivationale Aspekte oder persönliche Wertvorstellungen.

Anzumerken ist, dass einige Definitionen Einstellungen als Dispositionen beschreiben. Schwarz und Bohner (2001) sind dagegen der Ansicht, dass Einstellungen sich auch erst bei Bedarf entwickeln können (vgl. auch Banaji & Heiphetz, 2010; Schwarz, 2007). Eagly und Chaiken (1993) sowie weitere Einstellungsforscher(innen) (z. B. Fazio, 2007; Schwarz & Bohner, 2001) gehen davon aus, dass nicht alle Einstellungen über einen langen Zeitraum stabil bleiben, sondern manche Einstellungen sich unter bestimmten Umständen verändern.²⁴ Einstellungen von Lehrkräften gegenüber ethnischen oder sozialen Gruppen sind somit als relativ stabil aber veränderbar anzusehen. Sie entstehen zum einen durch Einstellungen ihres Umfelds und zum anderen durch eigene Erfahrungen mit diesen Schülergruppen und äußern sich in positiven oder negativen Gedanken, Gefühlen oder Verhaltensweisen.

Der Einfluss von Einstellungen und Stereotypen auf das Verhalten

Einstellungen und Stereotype können das Verhalten von Personen unbewusst beeinflussen. So zeigen Personen nach der Aktivierung eines Stereotyps häufig stereotypkonsistentes Verhalten (vgl. Bargh, Chen & Burrows, 1996). Zu Beginn der Einstellungsforschung wurde Einstellungen ein verhaltensdeterminierender Charakter zugeschrieben (vgl. z. B. Allport, 1935; Thomas & Znaniecki, 1918), mittlerweile besteht jedoch weitestgehend Einigkeit, dass Einstellungen gegenüber einem Einstellungsobjekt nicht zwangsläufig zu einem entsprechenden Verhalten führen müssen (vgl. z. B. Ajzen & Fishbein, 1980; Schwarz & Bohner, 2001). Eine Lehrkraft kann z. B. eine negative Einstellung gegenüber ethnischen Minoritäten haben, sich im individuellen Fall jedoch gegen negative Verhaltensweisen gegenüber den entsprechenden Schüler(inne)n entscheiden. Werden Einstellungen nicht im Verhalten gezeigt, kann dies unterschiedliche Ursachen haben, z. B. soziale Normen oder die Zugänglichkeit der Einstellung im Gedächtnis (vgl. Ajzen & Fishbein, 1980; Fazio, Powell & Williams, 1989). Nach dem MODE-Modell (*motivation and opportunity as determinants model*) werden insbesondere die Motivation zur Verarbeitung und die Gelegenheit, aber auch die Zugänglichkeit der Einstellung als wesentliche Einflussfaktoren auf den Zusammenhang zwischen Einstellungen und entsprechendem Verhalten angesehen (Fazio, 1990; Olson & Fazio, 2009; Schuette & Fazio, 2016). Einstellungen können somit nicht mit dem betreffenden Verhalten in einem individuellen Fall gleichgesetzt werden (Ajzen & Fishbein, 1980). Angenommen wird, dass Einstellungen einen handlungsmotivierenden, jedoch keinen determinierenden Charakter aufweisen (Eagly & Chaiken, 1993; Überblick bei Fazio, 1986).

Fazio (1986) geht in dem *attitude accessibility model* davon aus, dass Einstellungen zunächst aktiviert sein müssen, um das Verhalten beeinflussen zu können. Anzunehmen ist,

²⁴ Allport (1935) und Lippmann (1922) beschreiben Einstellungen nicht als unveränderbar, jedoch als sehr starr und unflexibel. Da Einstellungen zwar für eine kurze Zeit stabil sind, sich aber auch verändern können, präferieren Eagly und Chaiken (1993) die Beschreibung von Einstellungen als Tendenzen im Gegensatz zu Dispositionen, da dies einen zeitlich stabilen Charakter zuschreiben würde (z. B. Ajzen, 2005).

dass die Einstellungsaktivierung dabei ohne bewusste Aufmerksamkeit verläuft (Bargh, Chaiken, Govender & Pratto, 1992; Greenwald & Banaji, 1995). Einstellungen können durch das Einstellungsobjekt selbst oder aber durch Objekte, die mit dem Einstellungsobjekt assoziiert sind, aktiviert werden (Eagly & Chaiken, 1993). Nach Fazio (1986) hängt die Aktivierung von der Assoziationsstärke zwischen Einstellungsobjekt und einer entsprechenden Bewertung ab, da die Assoziationsstärke die Zugänglichkeit der Einstellung beeinflusst (Bargh et al., 1992). In Replikationsstudien der Untersuchungen von Fazio, Sanbonmatsu, Powell & Kardes (1986) konnten Bargh et al. (1992) feststellen, dass eine automatische Aktivierung auch bei Einstellungen mit geringerer Assoziationsstärke auftreten kann. Die Abrufleichtigkeit von Gedächtnisinhalten kann einen bedeutsamen Einfluss auf die Stereotypaktivierung haben. Dies hängt u. a. mit der Häufigkeit ihrer Nutzung zusammen, kann jedoch durch aktuelle oder kürzlich aktivierte Eindrücke (*priming*) erhöht werden (vgl. Klauer, 2008; Lepore & Brown, 1997; van Twuyver & van Knippenberg, 1995; Wasel & Gollwitzer, 1997). Basierend auf den Ergebnissen zweier experimenteller Studien von Gilbert und Hixon (1991) kann davon ausgegangen werden, dass kognitive Beanspruchung die Aktivierung von Stereotypen eher hemmt. Jedoch zeigte sich auch, dass trotz kognitiver Beanspruchung aktivierte Stereotype wirksam werden (Gilbert & Hixon, 1991).

Der Einfluss von Einstellungen und Stereotypen auf die Informationsverarbeitung

Die Aktivierung von Einstellungen und Stereotypen kann nicht nur die sichtbaren Verhaltensweisen, sondern auch die Wahrnehmung und die Interpretation des Einstellungsobjekts beeinflussen (Bargh et al., 1992; vgl. Kapitel 5). Aktivierte Einstellungen oder Stereotype müssen sich aber nicht zwangsläufig in einem expliziten Urteil niederschlagen. So können z. B. die soziale Erwünschtheit oder konträre Überzeugungen, die entgegen den Einstellungen oder stereotypbasierten Urteilen stehen, verhindern, dass ein einstellungs- bzw. stereotypenkonformes Urteil erfolgt (vgl. Klauer, 2008).

Kognitive Verknüpfungen zwischen sozialen Kategorien und Merkmalen sowie Reaktionstendenzen ermöglichen es, Einstellungen und Stereotype in einer komplexen sozialen Umwelt handlungsfähig zu sein und schnelle Urteile fällen zu können (z. B. Bodenhausen, 2005). Jedoch besteht v. a. bei Stereotypen die Gefahr, individuelle Merkmale bei der Beurteilung von Personen zu vernachlässigen (Bodenhausen & Macrae, 1996). Stereotype können dabei v. a. in heuristischen Informationsverarbeitungsprozessen relevant sein (Bodenhausen & Wyer, 1985; vgl. Kapitel 5). McCauley, Stitt und Segal (1980) gehen davon aus, dass Stereotype als subjektive Auftretenswahrscheinlichkeiten angesehen werden können, wobei diese nicht mit den tatsächlichen Auftretenswahrscheinlichkeiten von Merkmalen in sozialen Kategorien übereinstimmen müssen und hierdurch z. T. fragwürdige diagnostische Aussagen getroffen werden (Bodenhausen & Lichtenstein, 1987). Stereotype müssen jedoch nicht unzutreffend sein (Jussim et al., 1996). Jussim et al. (1996) nennen zwei zentrale Gründe, warum Stereotype nicht automatisch zu fehlerhaften Urteilen führen müssen: (1) Kontrollierende oder regulierende Faktoren können die Nutzung von Stereotypen unterbinden. (2) Stereotype repräsentieren gültige Merkmalsausprägungen und führen dadurch zu akkuraten Urteilen. Stereotype können für einige Mitglieder sozialer Gruppen zutreffen, müssen dies jedoch nicht für jedes Individuum der sozialen Gruppe.

Einstellungen können dem Individuum direkt zugänglich sein oder aber implizit vorliegen (Albarracín & Vargas, 2010). Greenwald und Banaji (1995) definieren implizite Einstellungen dabei als „introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action towards social objects” (S. 8). Vergleichbar mit der Struktur von Einstellungen können Stereotype neben expliziten auch implizite Ausprägungen haben (Greenwald & Banaji, 1995). Dabei stellt eine eher automatisierte Verarbeitung von Stereotypen die Grundlage für implizite Stereotypisierungsprozesse dar (Greenwald & Banaji, 1995).²⁵ Entsprechend besteht auch die Gefahr, dass v.a. implizite Einstellungen Lehrkrafturteile beeinflussen. So könnten Lehrkräfte, die z. B. besonders positive implizite Einstellungen gegenüber bestimmten Schüler(inne)n haben, diese Schüler(innen) unbewusst tendenziell besser benoten oder ihr Verhalten positiver bewerten.

In Studien, in denen der *implizite Assoziationstest* (IAT; Greenwald, McGhee & Schwarz, 1998) zur Erfassung impliziter Einstellungen eingesetzt wurde, zeigten sich meist nur geringe oder keine Korrelationen mit expliziten Einstellungsmaßen (vgl. z. B. die Meta-Analyse von Hofmann, Gawronski, Gschwendner, Le & Schmitt, 2005).²⁶ Eine mögliche Erklärung wäre, dass es sich bei impliziten und expliziten Einstellungen um zwei unterschiedliche Facetten handelt (vgl. Albarracín & Vargas, 2010). Die Differenzierung von impliziten und expliziten Einstellungen ließ sich auch durch neurowissenschaftliche Befunde bestätigen, die auf die Aktivität unterschiedlicher Gehirnregionen verweisen (Überblick bei Albarracín & Vargas, 2010). Hofmann et al. (2005) konnten insbesondere die Spontanität der expliziten Selbstberichtsdaten als Moderator identifizieren, was Annahmen zur Automatisierung impliziter Einstellungen bestätigt.²⁷ Zudem wird vermutet, dass die geringen Korrelationen an einer niedrigen Übereinstimmung der Konzeptualisierung impliziter und expliziter Maße (z. B. Fokussierung auf kognitive oder affektive Aspekte) liegt (Hofmann et al., 2005).

Dasgupta und Greenwald (2001) verweisen zudem auf mögliche Priming-Effekte, die die Stärke von impliziten Einstellungen verändern können. Versuchspersonen wurden vor der Bearbeitung eines IAT – zur Erfassung impliziter Einstellungen gegenüber afroamerikanischen und europäischstämmigen Personen – je nach Experimentalgruppe Bilder von berühmten und beliebten afroamerikanischen Personen (z. B. Schauspieler oder Sportler) und berühmten aber eher unbeliebten europäischstämmigen Personen (für die andere Experimentalgruppe umgekehrt) gezeigt (Dasgupta & Greenwald, 2001). Wenn vorher unbeliebte Personen präsentiert wurden, zeigten sich geringere Präferenzen für die jeweilige ethnische Gruppe. Stärkere Präferenzen für die ethnische Gruppe zeigten sich hingegen, wenn zuvor beliebte Personen der gleichen Ethnie gezeigt wurden (Dasgupta & Greenwald, 2001). Positive oder negative Einstellungen können auch durch vorher präsentierte, angenehme oder unangenehme Stimuli

²⁵ Die Aktivierung von Stereotypen ist dabei u. a. von persönlichen Motiven und Zielen sowie situationalen Faktoren abhängig (Übersicht bei Blair, 2002).

²⁶ Eine der wenigen Ausnahmen, die Korrelationen zwischen impliziten und expliziten Einstellungen finden konnten, ist die Studie von Carlsson und Björklund (2010). Sie konnten mit zwei IATs und den Attribut-Kombinationen warm/kalt und kompetent/inkompetent (in Anlehnung an das *stereotype content model*; Fiske et al., 2002) stereotype Zuschreibungen zu Berufsgruppen aufzeigen und Korrelationen des Kompetenz-IATs mit expliziten Maßen finden (Carlsson & Björklund, 2010).

²⁷ Automatisiertheit zeichnet sich dabei u. a. durch folgende Charakteristika aus: unbewusst, unkontrolliert, unintentional, Stimulus geleitet, effizient und schnell (Moors & de Houwer, 2006).

induziert werden (*subliminal attitude conditioning*), wenn das Einstellungsobjekt vorher neutral besetzt ist (Greenwald & Banaji, 1995). Es ist davon auszugehen, dass die Einstellung gegenüber dem angenehmen oder unangenehmen Stimulus fälschlicherweise auf das neutrale Einstellungsobjekt attribuiert wird (Greenwald & Banaji, 1995). Es ist aber auch davon auszugehen, dass Kontexteffekte zumindest den Abruf von Einstellungen beeinflussen (Greenwald & Banaji, 1995).²⁸

Die Auswirkungen von Einstellungen und Stereotypen auf alle Bereiche des sozialen Lebens sind weitestgehend unumstritten (vgl. z. B. Albarracín und Vargas, 2010; Tajfel, 1969) und damit auch für den schulischen Kontext von essentieller Bedeutung. Insbesondere implizite Einstellungen können einen Einfluss auf die Urteilsbildung haben. Zudem sind Einstellungen nur schwer veränderbar. Somit stellt sich die Frage, wie Einstellungen und Stereotype von Lehrkräften gegenüber ethnischen Minoritäten und sozialen Gruppen ausgeprägt sind, um mögliche Verzerrungen in Urteilen und Erwartungen hierdurch erklären zu können. Herkunftsassoziierte Einstellungen und Stereotype im schulischen Kontext sind zudem eng mit allgemeinen gesellschaftlichen Einstellungen und Stereotypen gegenüber ethnischen und sozialen Gruppen verknüpft. Da individuelle Einstellungen auch durch Einstellungen Dritter beeinflusst werden können, fokussiert der folgende Abschnitt zunächst allgemeine Einstellungen und Stereotype gegenüber ethnischen und sozialen Gruppen.

4.1 Einstellungen und Stereotype gegenüber ethnischen und sozialen Gruppen

Bereits in den 1920er Jahren untersuchte Bogardus (1925b, 1925c, 1926) die soziale Distanz – d.h. das Ausmaß an Mitgefühl, Verständnis, und Intimität – zwischen ethnischen Gruppen (nach Park, 1924). Bogardus (1926) ging der Annahme nach, dass trotz einer räumlichen Nähe der unterschiedlichen ethnischen Gruppen in Städten eine soziale Distanz besteht. In einer Befragung von gebürtigen US-amerikanischen Personen aus urbanen Gegenden konnte er die soziale Distanz zu anderen ethnischen und sprachlichen Gruppen als der eigenen aufzeigen (Bogardus, 1926).²⁹ Für bestimmte ethnische Gruppen, z.B. Personen mit afroamerikanischen oder türkischen Wurzeln, erwies sich die soziale Distanz ausgeprägter als bei anderen ethnischen Gruppen, wie z.B. Personen mit Wurzeln in England oder Norwegen (Bogardus, 1926). Die geringste soziale Distanz wurde gegenüber Personen berichtet, die der eigenen ethnischen Herkunft entsprachen (Bogardus, 1926).³⁰ Unter Berücksichtigung der sozialen Schicht zeigten sich zudem negativere Einstellungen von Personen aus höheren Schichten gegenüber anderen Ethnien (Bogardus, 1926). Bogardus (1926) nahm an, dass diese höhere

²⁸ Schwarz und Clore (1983) konnten z. B. zeigen, dass die aktuelle Stimmung der Versuchspersonen oder auch die Wetterlage bei der Befragung Auswirkungen auf Urteile haben können.

²⁹ Mit sieben Items (u. a. der Bereitschaft, Personen anderer Ethnien als Nachbarn zu haben) erfasste Bogardus (1926) die soziale Distanz auf einer Ratingskala.

³⁰ Bogardus (1925a, 1926) erbrachte erste Hinweise, dass ethnische Einstellungen veränderbar sind. Personen mit eher neutraler Einstellung gegenüber spezifischen ethnischen Gruppen entwickelten durch einzelne positive Kontakte mit Gruppenmitgliedern der entsprechenden Ethnie eine positivere Einstellung, während Personen mit eher negativen Einstellungen für eine positivere Einstellungsänderung deutlich mehr positive Erfahrungen mit Mitgliedern der entsprechenden ethnischen Gruppe benötigten (Bogardus, 1925a, 1926).

soziale Distanz ein Ergebnis des Wunsches nach eigenem Statuserhalt ist, bzw. Ausdruck der Befürchtung, durch aufstrebende andere soziale Gruppen im eigenen Status gefährdet zu sein.

In einer weiteren Untersuchung von Bogardus (1925c) sollten Proband(inn)en 39 ethnische und nationale Gruppen in die Kategorien *Freundlichkeit*, *neutral* und *Abneigung* einteilen. Dabei drückten die Versuchspersonen ihre größte Abneigung gegenüber afroamerikanischen und türkischstämmigen Personen aus, denen gegenüber sie auch die höchste soziale Distanz hatten (Bogardus, 1925c). Diese ersten Befunde zu positiven Einstellungen gegenüber der eigenen ethnischen Gruppe und negativen Einstellung gegenüber anderen ethnischen Gruppen (v. a. gegenüber afroamerikanischen und türkischstämmigen Personen) konnten für den US-amerikanischen Raum der 1920er und 1930er ebenfalls durch Forschungsarbeiten von Thurstone (1928, 1931) bestätigt werden (Überblick bei Katz & Braly, 1933).

Diese Befunde wurden von Katz und Braly (1933) um empirische Evidenz erweitert. Sie legten Studierenden eine Liste mit 84 sowohl positiven (z. B. *intelligent* oder *freundlich*) als auch negativen (z. B. *faul* oder *grausam*) Adjektiven vor, die zehn unterschiedlichen ethnischen Gruppen als typische Merkmale zugeordnet werden sollten (*adjective checklist*).³¹ In den prozentualen Nennungshäufigkeiten der Eigenschaften nach Ethnizität zeigten sich die positivsten Zuschreibungen für die Eigengruppe (Amerikaner(innen) – *fleißig*: 48.5 % und *intelligent*: 47.5 %), aber auch für andere Ethnien (Engländer(innen) – *sportlich*: 53.5 %; *intelligent*: 46.5 %; Deutsche – *wissenschaftlich interessiert*: 78.0 % und *fleißig*: 65.0 %). Afroamerikaner(innen) (*abergläubisch*: 84.0 %; *faul*: 75.0 %) und Türk(inne)en (*grausam*: 54.0 %; *sehr religiös*: 29.9 %) wurden hingegen deutlich negativere Eigenschaften zugeschrieben (Katz & Braly, 1933). Katz und Braly (1933) vermuteten in diesem Zusammenhang, dass die hohe Übereinstimmung der Studierenden in ihren Zuschreibungen bedingt durch die Bekanntheit der Gruppen, also dem verfügbaren bzw. abrufbaren gespeicherten gesellschaftlich geteilten Wissen oder auch über die persönlich verfügbaren Kontakte zu diesen Gruppen, war.

Neben diesen Meilensteinen der frühen Vorurteilsforschung im US-amerikanischen Raum verweisen jedoch Untersuchungen im Verlauf der letzten Jahrzehnte immer wieder auf negativere Einstellungen und Stereotype gegenüber ethnischen Minoritäten im Vergleich zur Majorität (z. B. Devine & Elliot, 1995; Maykovich, 1972; Schaller, Park & Mueller, 2003; Wittenbrink, Judd & Park, 1997). Negative Einstellungen, Stereotype und Vorurteile und gegenüber Minoritäten konnten dabei auch in zahlreichen europäischen Ländern festgestellt werden (z. B. Hagendoorn & Hraba, 1989; Kirchler & Zani, 1995; Zick, Pettigrew & Wagner, 2008). Insgesamt kann i. d. R. eine Präferenz für die Eigengruppe (*ingroup*) im Vergleich zur Fremdgruppe (*outgroup*) festgestellt werden. Dunham und Emory (2014) konnten in einer Untersuchung mit Kindern zeigen, dass die Präferenz für die Eigengruppe bei Kindern mit drei Jahren im Gegensatz zu Kindern mit sechs Jahren jedoch noch nicht ausgeprägt ist, was auf eine entwicklungsabhängige Komponente bei der Entstehung der Ingroup-Präferenz hindeutet.

³¹ Katz und Braly (1933) untersuchten dabei die charakteristischen Eigenschaftszuordnungen der Studierenden zu folgenden ethnischen Gruppen: Personen aus China, Deutschland, England, Irland, Italien, Japan und der Türkei sowie gegenüber Amerikaner(inne)n mit europäischer Herkunft, Personen mit jüdischen und afroamerikanischen Wurzeln. Dabei sollten die Studierenden alle Wörter aus der Liste auswählen, die sie für diese Ethnie als typisch erachteten und konnten diese Liste auch um eigene Adjektive erweitern, falls diese nicht in der *adjective checklist* aufgeführt waren (Katz & Braly, 1933).

Auch neurologische Befunde deuten auf unterschiedliche neuronale Prozesse bei der Bewertung der Eigengruppe und der Fremdgruppe hin. In einer Untersuchung unter Einsatz des IAT und bildgebender Verfahren (fMRT) konnte aufgezeigt werden, dass bei der Betrachtung von ethnizitätsspezifischem Stimulusmaterial unterschiedliche Regionen im Gehirn aktiv sind (Beer et al., 2008). Bei negativen Assoziationen zur ethnischen Fremdgruppe zeigten sich Aktivitäten im *nucleus caudatus* sowie im *linken lateralen orbitofrontalen cortex*, der u. a. für die Steuerung und Kontrolle emotionaler Prozesse zuständig ist (Beer et al., 2008). Positive Assoziationen mit der eigenen ethnischen Gruppe führten hingegen zu einer Aktivität der *amygdala*, dem *medialen orbitofrontalen cortex* und dem *rechten lateralen orbitofrontalen cortex*, der eine zentrale Rolle bei der Emotions- und Impulskontrolle spielt (Beer et al., 2008).

In Deutschland vertreten 53 % der Personen die Einstellung, dass ethnische, kulturelle und religiöse Vielfalt eine Bereicherung für die Gesellschaft darstellt (Ward & Masgoret, 2008). Im internationalen Vergleich liegt Deutschland damit hinter Ländern wie Neuseeland, den Niederlanden oder Italien (Ward & Masgoret, 2008). Faulbaum und Stein (2000) stellten für Deutschland fest, dass Personen sich in ihren Einstellungen gegenüber ethnischen Minderheiten dahingehend unterscheiden, ob sie eine eher integrative oder antiintegrative Einstellung aufweisen.³² Anzunehmen ist, dass negative ethnische Einstellungen von Personen sich nicht nur auf eine ethnische Gruppe beschränken. Alba und Johnson (2000) verweisen darauf, dass eine hohe positive Korrelation zwischen Einstellungen gegenüber Personen mit ausländischen und jüdischen Wurzeln in Deutschland besteht.

Empirische Studien berichten zudem konkrete positive bzw. negative Assoziationen gegenüber spezifischen ethnischen Gruppen in Deutschland. Asbrock (2010) konnte in Anlehnung an das *stereotype content model* (Fiske et al., 2002) in einer Untersuchung mit Studierenden zeigen, dass die Gruppe der türkischen Migrant(inn)en in das Cluster *geringe Wärme* und *geringe Kompetenz*, Personen ohne Migrationshintergrund hingegen in das Cluster *warm* und *kompetent* eingestuft wurden (Asbrock, 2010). In einer offenen Befragung von Studierenden wurde zwar eine vergleichbare Anzahl negativer Stereotype für Personen ohne Migrationshintergrund und in Deutschland lebende türkischstämmige Personen berichtet, jedoch wurden Personen mit türkischem Migrationshintergrund deutlich weniger positive Stereotype zugeschrieben als Personen ohne Migrationshintergrund (Kahraman & Knoblich, 2000).³³

Die Befunde zur annähernd gleichen Verteilung negativer Attribute zur Eigen- und Fremdgruppe im Vergleich zu den Ergebnissen von Katz und Braly (1933) könnten an zwei wesentlichen Unterschieden liegen. Zum einen sind kultur- und gesellschaftsspezifische sowie historische Besonderheiten im Umgang mit Minoritäten zwischen dem US-amerikanischen

³² Dabei verweisen Faulbaum und Stein (2000) in ihren Analysen darauf, dass v. a. jüngere Personen mit einer höheren Bildung und mehr Kontakten zu anderen Ethnien eine integrative Einstellung aufweisen. Im Gegensatz dazu zeigt sich, dass Personen mit einer starken antiintegrativen Einstellung im Durchschnitt eher älter sind, geringere Bildungsabschlüsse aufweisen und weniger Kontakt zu anderen Ethnien haben (Faulbaum & Stein, 2000).

³³ Sowohl die genannten Stereotype gegenüber türkischstämmigen Migrant(inn)en (z.B. traditionell und männliche Dominanz) als auch gegenüber Personen ohne Migrationshintergrund (z.B. intolerant und leistungsorientiert) wiesen eine hohe Übereinstimmung zwischen den Versuchspersonen auf (Kahraman & Knoblich, 2000).

Raum und Deutschland zu erwähnen, die einen Einfluss auf entsprechende Ausprägungen ethnischer Einstellungen haben können und die Ergebnisse somit nicht ohne Einschränkungen vergleichbar machen. Durante et al. (2012) weisen z.B. darauf hin, dass sich herkunftsassoziierte Einstellungen je nach Einkommensgerechtigkeit in verschiedenen Ländern unterscheiden. Zum anderen erscheint insbesondere die Beachtung der fast 70 Jahre zwischen den Publikationen von besonderer Bedeutung. Während der 1930er Jahre war die offene Kommunikation negativer Einstellungen und ethnizitätsspezifischer Stereotype deutlich weniger verpönt, als dies im Jahr 2000 der Fall war. Während eine offene Äußerung negativer Stereotype gesellschaftlich weniger toleriert wird, erfolgt hier womöglich eine Kontrolle sozialer Äußerungen und ein eher sozial erwünschtes Antwortverhalten. Anders verhält es sich jedoch möglicherweise mit der Vergabe von weniger positiven Eigenschaften für andere Ethnien, da dies nur indirekt als negative Zuschreibung gewertet wird und somit sozial verträglicher erscheinen könnte.

Der Einfluss der sozialen Erwünschtheit wird auch deutlich, wenn eine Bogus-Pipeline-Messung erfolgt.³⁴ Mummendey, Bolten und Isermann-Gerke (1982) konnten in einer Studie keine Unterschiede in der Einschätzung von türkischstämmigen Personen und Personen ohne Migrationshintergrund, wenn eine Fragebogenerhebung ohne Bogus-Pipeline-Messungen eingesetzt wurde, feststellen. Unter der Bogus-Pipeline-Bedingung zeigten sich hingegen signifikant negativere Einschätzungen türkischstämmiger Personen. Neben dem Einfluss der sozialen Erwünschtheit ist auch der Kontext bedeutsam, in dem Versuchspersonen zu ethnischen Stereotypen befragt werden bzw. in dem Stimulusmaterial präsentiert wird. Mit zwei experimentellen Untersuchungen konnten Wittenbrink, Judd und Park (2001) diese Annahme bestätigen. In ihren Studien wurden afroamerikanische Personen in einem positiven Kontext (z. B. Familie) positivere Einstellungen entgegengebracht als in einem negativen Kontext (z. B. Straßenecke) (Wittenbrink et al., 2001).

Für die Gruppe der türkischen Migrant(inn)en in Deutschland kann neben ethnizitätsspezifischen Einstellungen, Vorurteilen und Stereotypen eine religiös begründete, vorurteilsbehaftete Einstellung gegenüber dem Islam vermutet werden (vgl. Benz, 2012). Aber auch durch die hohe Konfundierung der ethnischen und sozialen Herkunft könnten türkische Migrant(inn)en ebenfalls mit Stereotypen gegenüber niedrigen sozialen Schichten (z. B. Rössel & Pape, 2010) konfrontiert sein. Einzelne empirische Hinweise deuten auf negativere Einstellungen gegenüber Personen mit einem niedrigen sozioökonomischen Status (z. B. Sielschott, 2010). Insgesamt ist jedoch ein Mangel an empirischen Studien zu verzeichnen, die sich mit Einstellungen, Vorurteilen und Stereotypen gegenüber unterschiedlichen sozialen Schichten im deutschsprachigen Raum befassen und zudem die Überlappung sozialer Kategorien berücksichtigen.

Herkunftsassoziierte Einstellungen, Vorurteile und Stereotype können auf unterschiedliche Weise bedeutsam werden. Sie können die Wahrnehmung von ethnischen Gruppen (z. B.

³⁴ Das Bogus-Pipeline-Paradigma besagt, dass Personen unter der Annahme, ihre körperlichen Reaktionen würden über eine Messapparatur aufgezeichnet und ließen Rückschlüsse über ihre tatsächlichen Einstellungen zu, ehrlicher antworten und damit der Einfluss der sozialen Erwünschtheit gemindert werden kann (vgl. z. B. Jones & Sigall, 1971; Sigall & Page, 1971).

Wilson, Hugenberg & Rule, 2017), aber auch das Verhalten gegenüber ethnischen Gruppen beeinflussen (z. B. Baltés & Rudolph, 2010; Bertrand & Mullainathan, 2004; Jowell & Prescott-Clark, 1970; Klink & Wagner, 1999). Auch die Erwartungen an Verhaltensweisen der ethnischen Gruppe können entsprechend der Einstellungen variieren sowie die Interpretation des Verhaltens beeinflussen. Gawronski, Geschke und Banse (2003) berichten, dass v. a. Personen mit starken negativen Assoziationen gegenüber türkischstämmigen Personen eher negativere Verhaltensweisen von ihnen erwarten, als von Personen ohne Migrationshintergrund. Die Befunde einer experimentellen Untersuchung von Sagar und Schofield (1980) deuten darauf hin, dass afroamerikanische Personen bei der Durchführung der gleichen Handlung als bedrohlicher und aggressiver wahrgenommen werden (vgl. auch Duncan, 1976). Personen mit lateinamerikanischen Wurzeln wurden in empirischen Untersuchungen ebenfalls als aggressiver eingeschätzt (vgl. Marín, 1984).³⁵ Eine Erklärung der APA (2012) für Einstellungsunterschiede gegenüber ethnischen Gruppen sind unterschiedliche Einwanderungsgründe. Während ein großer Anteil der lateinamerikanischen Einwanderer(innen) zur Familienwiedervereinigung in die USA einreist, sind die meisten asiatischen Einwanderer(innen) aus beruflichen Gründen immigriert und weisen einen entsprechend höheren Bildungsgrad auf. Somit lässt sich vermuten, dass primär ethnisch-anmutende Einstellungen auch auf assoziierte sozioökonomische Ressourcen und die Bildungsnähe zurückzuführen sind.

Zusammenfassend lässt sich festhalten, dass sowohl internationale als auch nationale Studien auf negativere Einstellungen, Vorurteile und Stereotype gegenüber ethnischen Minoritäten verweisen. In Deutschland ist dabei von negativeren Einstellungen gegenüber türkischen Migrant(inn)en auszugehen, wobei zu vermuten ist, dass hierbei auch der sozioökonomische Status dieser Gruppe einen nicht zu vernachlässigenden Einfluss hat. Dennoch fehlt es weiterhin an Untersuchungen, die den sozialen Status ethnischer Gruppen stärker berücksichtigen.

4.2 Herkunftsspezifische Einstellungen und Stereotype von Lehrkräften

Die Befunde zu herkunftsassoziierten Urteilen und Erwartungen von Lehrkräften (z. B. Darley & Gross, 1983; Sprietsma, 2013; Tenenbaum & Ruck, 2007; vgl. Abschnitt 3.3) lassen bereits darauf schließen, dass sowohl ethnische als auch soziale Schülergruppen mit spezifischen herkunftsassoziierten Stereotypen und Einstellungen durch Lehrkräfte besetzt sind. Dabei erfassen diese Studien die Konsequenzen, die sich möglicherweise durch herkunftsspezifische Einstellungen und Stereotype von Lehrkräften ergeben. Sie berichten jedoch nicht, welche konkreten Assoziationen mit den jeweiligen Schülergruppen verknüpft sind. Abschnitt 4.1 konnte einige empirische Belege für herkunftsabhängige Einstellungen und Stereotype in der allgemeinen Bevölkerung darlegen, jedoch verbleibt die Frage, ob sich negativere Einstellungen gegenüber ethnischen Minoritäten oder sozialen Gruppen auch bei Lehrkräften

³⁵ Duncan (1976) berichtet in diesem Zusammenhang, dass für afroamerikanische Personen eher dispositionelle Ursachen, für europäischstämmige Personen hingegen eher situationale Ursachen für aggressives Verhalten attribuiert werden. Auch werden je nach ethnischer Herkunft andere Ursachen für den geringeren Bildungserfolg betont. Bei Migrant(inn)en lateinamerikanischer Abstammung werden sprachliche Defizite fokussiert, afroamerikanische Schüler(innen) werden stärker mit Disziplinproben verknüpft (APA, 2012).

und Lehramtsstudierenden zeigen und ob sich auch hier eine positivere Einstellung für die eigene ethnische und soziale Gruppe feststellen lässt.

Gebauer und McElvany (2017) berichten durchschnittlich eher positive Einstellungen von Lehramtsstudierenden in Deutschland gegenüber einer kulturell heterogenen Schülerschaft. Hachfeld, Hahn, Schroeder, Anders und Kunter (2015) verweisen darauf, dass Lehrkräfte, die eine multikulturelle Einstellung haben, auch von einer höheren Selbstwirksamkeitserwartung und mehr Enthusiasmus berichteten, als Lehrkräfte, die versuchen Unterschiede zwischen Ethnien und Kulturen zu ignorieren (*colorblind belief*). Jedoch bezieht sich dies noch nicht auf die Einstellungen von Lehrkräften gegenüber spezifischen ethnischen oder sozialen Gruppen, sondern deutet eine eher positive Einstellung gegenüber einer heterogenen Zusammensetzung der Schülerschaft an.

Leacock (1969) erfasste für die USA der 1960er Jahre die Einstellungen von Lehrkräften zur ethnischen (afroamerikanischer vs. europäischer Hintergrund) und sozialen (niedriger vs. mittlerer sozioökonomischer Status) Herkunft von Schüler(inne)n. Die Lehrkräfte trafen die negativsten Äußerungen bei afroamerikanischen Schüler(inne)n mit niedrigem sozioökonomischem Status. Europäischstämmige Schüler(innen) mit mittlerem sozioökonomischem Status wurden hingegen am wenigsten negativ bewertet (Leacock, 1969; siehe auch Rosenthal, 1974). In Bezug auf die Konfundierung der ethnischen und sozialen Schülerherkunft im schulischen Kontext ist die Befundlage jedoch äußerst gering.

Vervaeet, D'hondt, van Houtte und Stevens (2016) bestätigten negative Einstellungen von Lehrkräften gegenüber ethnischen Minoritäten auch für Belgien. Bei einem hohen Bildungsgrad der Lehrkräfte, sowie einem hohen Anteil an Schüler(inne)n ethnischer Minoritäten an ihrer Schule, hatten sie aber die geringsten negativen Einstellungen gegenüber ethnischen Minoritäten (Vervaeet et al., 2016). Van den Bergh et al. (2010) konnten durch die Erfassung expliziter Einstellungen gegenüber ethnischen Minoritäten (erfasst durch eine Adaption der *modern racism scale* nach McConahay, 1986) und impliziter Einstellungen durch den IAT (vgl. Greenwald et al., 1998) eher geringe explizite negative Einstellungen feststellen, jedoch auch nicht zu vernachlässigende negative implizite Assoziationen feststellen (van den Bergh et al., 2010). Dabei zeigte sich zudem, dass Lehrkräfte mit negativen Einstellungen gegenüber ethnischen Minoritäten eher dazu neigen, diese Schüler(innen) schlechter einzuschätzen und diesen eine weniger aussichtreiche Schullaufbahn zu prognostizieren (van den Bergh et al., 2010). Dies weist auf den moderierenden Einfluss von Einstellungen bei der Urteilsbildung von Lehrkräften hin (vgl. auch Glock et al., 2016).

Glock und Karbach (2015) konnten in einer Untersuchung mit Lehramtsstudierenden unter Einsatz von drei unterschiedlichen impliziten Messverfahren zur Erfassung impliziter Einstellungen deutliche Unterschiede in den Einstellungen gegenüber ethnischen Gruppen aufzeigen.³⁶ Im Vergleich zu Schüler(inne)n ohne Migrationshintergrund erbrachten die Analysen negativere implizite Einstellungen gegenüber Schüler(inne)n mit Migrationshintergrund. Allerdings wurde auch deutlich, dass dies zwar im Vergleich mit Schüler(inne)n ohne Migra-

³⁶ Eingesetzt wurde der implizite Assoziationstest (IAT, *implicit association test*), die affektive Priming-Aufgabe (APT, *affective priming task*) und das affektive Missattributions-Prozedere (AMP, *misattribution procedure*) (Glock & Karbach, 2015).

tionshintergrund negativere Ausprägungen waren, jedoch v. a. davon auszugehen ist, dass die Versuchsteilnehmer deutlich positivere Einstellungen gegenüber den Schüler(inne)n ohne Migrationshintergrund hatten (Glock & Karbach, 2015). Entsprechend ist eher von positiveren Einstellungen gegenüber Schüler(inne)n ohne Migrationshintergrund auszugehen ist (Glock & Karbach, 2015). Ebenfalls unter Einsatz impliziter Messerverfahren zeigte sich, dass angehende Lehrkräfte positivere implizite und explizite Einstellungen für Schüler(innen) mit hohem sozioökonomischem Status aufweisen (Glock et al., 2016). Diese Befunde deuten darauf hin, dass Lehrkräfte entsprechend positive Einstellungen gegenüber Schüler(inne)n ohne Migrationshintergrund und mit einem hohen sozioökonomischen Status haben.

Zwar erweist sich die bisherigere Befundlage zu impliziten und expliziten Einstellungen und Stereotypen von Lehrkräften als nicht ausreichend, jedoch deuten auch die Befunde zu Lehrkrafturteilen und -erwartungen auf herkunftsassoziierte Einstellungen und Stereotype hin. Vergleichbar mit den Befunden zu allgemeinen Einstellungen und Stereotypen zeichnet sich auch für Lehrkräfte und Lehramtsstudierende ein eher negatives Bild für Minoritäten und ein eher positives Bild für die eigene Ethnie ab. Festzuhalten ist an dieser Stelle auch, dass Einstellungen und Stereotype für das menschliche Handeln in sozialen Situationen unverzichtbar sind. Einerseits bergen sie zwar die Gefahr, Individuen einer sozialen Gruppe basierend auf assoziierten Merkmalen (ohne Berücksichtigung der individuellen Merkmale) der entsprechenden sozialen Gruppe zu beurteilen. Andererseits ermöglichen sie es erst in einer komplexen sozialen Welt mit unzähligen Reizen handlungsfähig zu sein (vgl. z.B. Tajfel, 1969). Einstellungen, Vorurteile und Stereotype nehmen somit einen bedeutsamen Einfluss auf die soziale Informationsverarbeitung um schnelle Urteile zu ermöglichen. Da dies auch für die Urteils- und Erwartungsbildung von Lehrkräften der Fall ist, wird im folgenden Kapitel ein detaillierter Blick auf die zugrundeliegenden kognitiven Prozesse geworfen.

5. Lehrkrafturteilen und -erwartungen unterliegende soziale Kognitionen

Da die Bildung von Lehrkrafturteilen und -erwartungen als soziale Informationsverarbeitungsprozesse anzusehen sind, eignen sich insbesondere sozial-kognitive Theorien zur Erklärung dieser Prozesse.³⁷ Sozial-kognitive Ansätze orientieren sich dabei zunächst stark an kognitionspsychologischen Annahmen (vgl. Fiedler & Bless, 2003). Kognitionen bezeichnen dabei alle mentalen Prozesse, die den Erwerb, die Organisation und die Nutzung von Wissen umfassen (Neisser, 1976). Aus sozialpsychologischer Perspektive sind Kognitionen im Zusammenhang mit sozialen Kontexten als *soziale Kognitionen* zu bezeichnen (Fiske & Taylor, 2013). Nach Brewer (1988) beschreiben soziale Kognitionen die Interaktion zwischen internalen Wissensstrukturen (Schemata) mit neuen Informationen über soziale Objekte oder Ereignisse. Der Aspekt der Interaktion verdeutlicht, dass bereits vorhandene kognitive Strukturen durch neue soziale Informationen verändert werden können. Aber auch bestehende mentale Strukturen können die kognitive Verarbeitung sozialer Informationen beeinflussen (vgl. auch Neisser, 1976).³⁸ Bereits die Wahrnehmung, als erster Schritt der sozialen Informationsverarbeitung, ist abhängig von existierenden kognitiven Strukturen (vgl. z. B. Feuß, 2013; Neisser, 1976). Diese vorhandenen Schemata bilden zudem das Vorwissen, das nötig ist um soziale Stimuli enkodieren zu können (Bless et al., 2004). Entsprechend verknüpfen auch Lehrkräfte neue Informationen über Schüler(innen) mit vorhandenen Schemata, die sie bereits über Lernende haben. Gleichzeitig wird die Wahrnehmung von Schüler(inne)n durch bestehende kognitive Strukturen der Lehrkräfte beeinflusst. Somit können auch Einstellungen und Stereotype von Lehrkräften, die ebenfalls mentale Strukturen darstellen, die Wahrnehmung von Schüler(inne)n unterschiedlicher ethnischer und sozialer Herkunft beeinflussen. Schemata sind jedoch eine Voraussetzung für effektive kognitive Prozesse (Crick & Dodge, 1994). Entsprechend sind auch Einstellungen und Stereotype von Lehrkräften notwendige Voraussetzungen für eine schnelle Urteilsbildung.

In Abgrenzung zur Kognitionspsychologie sind zwei zentrale Merkmale zu nennen, mit denen sich sozial-kognitive Theorien und Studien befassen: (1) *soziale Stimuli* und (2) *distale Merkmale*. Die spezifischen Eigenschaften sozialer Stimuli (z. B. Dynamik sozialer Objekte) und distaler Merkmale (z. B. keine direkte Erfassbarkeit abstrakter Eigenschaften) führen dazu, dass der soziale Informationsverarbeitungsprozess durchaus komplex sein kann (vgl. u. a. Bless et al., 2004; Brewer, 1988; Fiedler & Bless, 2003; Fiske & Taylor, 2013). Übertragen auf die Urteils- und Erwartungsbildung von Lehrkräften sind Schüler(innen) als soziale Stimuli anzusehen, wobei Lehrkräfte sich Urteile über z. B. Fähigkeiten oder motivationale Ausprägungen von Schüler(inne)n (die nur über proximale Hinweise erschlossen werden können) bilden. Da

³⁷ Im Gegensatz zu behavioristischen Ansätzen (z. B. Skinner, 1938) erklären sozial-kognitive Theorien soziales Verhalten nicht nur durch externe Reize, sondern auch durch mentale Repräsentationen sozialer Situationen und Objekte und berücksichtigen dabei bereits bestehende kognitive Schemata, die soziales Verhalten und soziale Kognitionen beeinflussen können (vgl. Bless et al., 2004; Fiedler & Bless, 2003).

³⁸ Da individuelle Schemata keine statischen kognitiven Strukturen darstellen (Neisser, 1976), sind Einstellungen, Vorurteile und Stereotype somit ebenfalls potentiell veränderbar.

Personen sich äußerst schnell Urteile über andere Menschen bilden, ist insbesondere der erste Eindruck, den Schüler(innen) bei ihren Lehrkräften hinterlassen, von großer Bedeutung.

Der erste Eindruck

Eine der ersten Arbeiten zur Eindrucksbildung wurde von Asch im Jahr 1946 veröffentlicht, in denen bereits auf die Schnelligkeit der Eindrucksbildung und die geringe Anzahl hierfür notwendiger Informationen verwiesen wird: „We look at a person and immediately a certain impression of his character forms itself in us. A glance, a few spoken words are sufficient to tell us a story about a highly complex matter.“ (S. 258). In einer Reihe experimenteller Untersuchungen konnte Asch (1946) zeigen, dass der erste Eindruck über Personen nicht ausschließlich durch die Summe der einzelnen wahrgenommenen Eigenschaften gebildet wird, sondern darüber hinaus dynamische Verknüpfungen zwischen den Eigenschaften bestehen und weitere Assoziationen über tatsächlich bekannte Merkmale hinaus die Eindrucksbildung bestimmen.

Dabei können *implizite Persönlichkeitstheorien* wirksam werden. Menschen nehmen an, dass verschiedene Attribute zusammenhängen und gehen entsprechend davon aus, dass wenn eines dieser Merkmale auftritt auch die damit verknüpften Attribute zutreffen (z. B. Bruner & Tagiuri, 1954; Schneider, 1973). Da diese Persönlichkeitstheorien implizit sind, sind sie dem Individuum selbst oft nicht zugänglich oder bewusst. Lehrkräfte könnten z. B. bei besonders störenden Schüler(inne)n auch den Schluss ziehen, diese Schüler(innen) wären z. B. auch faul. Im Zusammenhang mit der ethnischen und sozialen Herkunft von Schüler(inne)n ist davon auszugehen, dass auch beim ersten Eindruck, den sich Lehrkräfte von ihren Schüler(inne)n bilden, Merkmale, die mit der jeweiligen Schülerherkunft kognitiv verknüpft sind (Stereotype; vgl. Abschnitt 4.1) aktiviert werden und die Eindrucksbildung beeinflussen. Diese stereotypen Annahmen über spezifische assoziierte Merkmale können dabei durch weitere implizite Persönlichkeitstheorien verstärkt werden. Eine Lehrkraft könnte z. B. das Merkmal *gebildet* mit einem hohen sozioökonomischen Status assoziieren. Mit der stereotypen Annahme, Schüler(innen) mit hohem sozioökonomischem Status seien gebildeter, könnten Lehrkräfte zusätzlich verknüpfen, dass gebildete Schüler(innen) z. B. intelligenter seien.

Asch (1946) erklärt dies dadurch, dass Personen sich stets einen Gesamteindruck einer Person bilden und aus diesem Grund weitere Eigenschaften aus den bekannten Merkmalen ableiten.³⁹ Zudem betont Asch (1946), dass manche Personenmerkmale einen größeren Einfluss auf die Eindrucksbildung haben als andere.⁴⁰ Lehrkräfte können bei der Eindrucksbildung, je

³⁹ Nach Anderson (1965) werden alle einzelnen Informationen (Kategoriezugehörigkeiten sowie individuelle Informationen) separat beurteilt und anschließend zu einem zusammenfassenden Eindruck vereint. Im Gegensatz dazu steht z. B. die Theorie von Asch (1946; Asch & Zukier, 1984). Diese besagt, dass Personen alle Informationen zu einem kohärenten Bild vereinen.

⁴⁰ In diesem Zusammenhang wurde zudem deutlich, dass auch die Reihenfolge der präsentierten Eigenschaften einen Einfluss auf die Eindrucksbildung hat (Asch, 1946). Basierend auf diesen Befunden konnte Anderson (1965) ebenfalls den Effekt der Reihenfolge personenbeschreibender Adjektive bestätigen. Dabei zeigte sich insbesondere ein Effekt der ersten Adjektive (*primacy-effect*) auf die Eindrucksbildung. Der Einfluss der folgenden Adjektive sank kontinuierlich und linear, je später sie in der Adjektivliste erwähnt wurden. Neben dem *primacy-effect* existiert auch der *recency-effect*. Dieser besagt, dass insbesondere die letzten Informationen besser erinnert werden, was u. a. daran liegen könnte, dass diese durch die zeitliche Nähe besser aus dem Gedächtnis abgerufen werden können (z. B. Jones & Goethals, 1971).

nach z.B. subjektiven Werten und Erwartungen, unterschiedliche Schülermerkmale für ihren ersten Eindruck als besonders bedeutsam erachten und entsprechend eher für die Eindrucksbildung heranziehen. Angenommen werden könnte, dass manche Lehrkräfte z.B. besonderen Wert auf ein ordentliches Erscheinungsbild legen und daher die Kleidung der Schüler(innen) für sie eine größere Bedeutung bei der Bildung des ersten Eindrucks hat als andere Schülermerkmale. Nach Asch (1946) werden aus diesen Merkmalen weitere Schlüsse gezogen. So könnte die Kleidung mit einem entsprechenden sozialen Status assoziiert sein (vgl. Dunkake & Schuchart, 2015), was wiederum sozialschichtspezifische Stereotype aktivieren könnte und damit die Eindrucksbildung beeinflusst.

Dabei verweisen empirische Befunde auf die Bedeutsamkeit des ersten Eindrucks, da dieser folgende Urteile maßgeblich beeinflusst. Gawronski et al. (2002) berichten aus einer experimentellen Untersuchung, dass Proband(inn)en, die in einem Experiment zunächst eine Zielperson als ruhig und gelassen erlebt haben, die anschließend in einer Situation als sehr nervös gezeigt wurde, die Zielperson in dieser Situation als weniger nervös einschätzten, als Versuchspersonen, die diese Zielperson bereits zuvor nervös gesehen hatten. Gawronski et al. (2002) vermuten, dass ein Assimilationseffekt vorliegt, der entweder dazu führt, dass sich die Aufmerksamkeit größtenteils auf mit dem ersten Eindruck konsistenten Verhaltensweisen richtet. Oder, dass während des Gedächtnisabrufs überwiegend oder ausschließlich Informationen des ersten Eindrucks erinnert werden (Gawronski et al., 2002). Eine weitere mögliche Erklärung könnte sein, dass neue Verhaltensweisen entsprechend dem ersten Eindruck interpretiert werden (vgl. Gawronski et al., 2002; Jones & Goethals, 1971). Festhalten lässt sich hieraus, dass der erste Eindruck einen bedeutsamen Einfluss auf folgende Urteile hat. Somit kann auch die Eindrucksbildung von Lehrkräften deren folgende Urteile über Schüler(innen) beeinflussen. Problematisch ist, dass sich der erste Eindruck in sehr kurzer Zeit bildet und sehr stabil ist. Willis und Todorov (2006) konnten in einer Serie experimenteller Untersuchungen zeigen, dass Personen sich bereits nach 100 ms einen ersten Eindruck von einer Zielperson bilden. Zudem verdeutlichen die Befunde, dass der erste Eindruck äußerst robust ist und sich auch nach längerer Betrachtung der Personen kaum verändert (Willis & Todorov, 2006).

Für den schulischen Kontext bedeutet dies, dass auch Lehrkräfte sich schnell überwiegend robuste Eindrücke über Schüler(innen) bilden. Dabei sind insbesondere zu Beginn eines neuen Schuljahres in einer neuen Klasse kaum Informationen über Schüler(innen) vorhanden. Somit wird auch hier der erste Eindruck zunächst über visuelle oder akustische Reize gebildet. Dabei erfolgen zunächst Kategorisierungen nach Geschlecht und anderen visuell wahrnehmbaren Merkmalen. Über spezifische morphologische Merkmale können dabei teilweise Rückschlüsse auf die Ethnie sowie über z.B. die Kleidung auf die soziale Herkunft der Schüler(innen) gezogen werden. Die erste Wahrnehmung dieser sozialen Kategorien kann dabei zu stereotypen Assoziationen bei Lehrkräften führen, die die weitere Wahrnehmung oder Interpretation bestimmter Verhaltensweisen der Schüler(innen) beeinflussen könnte. Die stereotypen Assoziationen können dabei zu stereotypkonformen Erwartungen führen, die für die einzelnen Schüler(innen) unzutreffend sein können. Diese falschen Erwartungen können zu selbsterfüllenden Prophezeiungen oder sustaining expectation effects führen (vgl. Good & Brophy, 2003; Merton, 1948; vgl. Abschnitt 3.2). Um den Einfluss der Schülerherkunft auf die

Eindrucks- und Urteilsbildung von Lehrkräften nachvollziehen zu können, ist zunächst eine genauere Betrachtung der kognitiven Verarbeitung sozialer Informationen notwendig.

5.1 Der Prozess sozialer Informationsverarbeitung

Crick und Dodge (1994) gehen davon aus, dass die erste Phase der Informationsverarbeitung die Enkodierung darstellt. Hierunter fassen sie v.a. Aufmerksamkeits- und Wahrnehmungsprozesse.⁴¹ Anschließend erfolgt eine Integration in bestehende Wissensstrukturen und darauf aufbauend eine Interpretation der Informationen (Crick & Dodge, 1994). Somit erfolgt nach der Wahrnehmung die Zuordnung von Personen zu Kategorien und ein entsprechender Abgleich mit bereits vorhandenen kognitiven Strukturen, die dann zu einem Urteil oder einer Entscheidung führen (vgl. Abbildung 8).

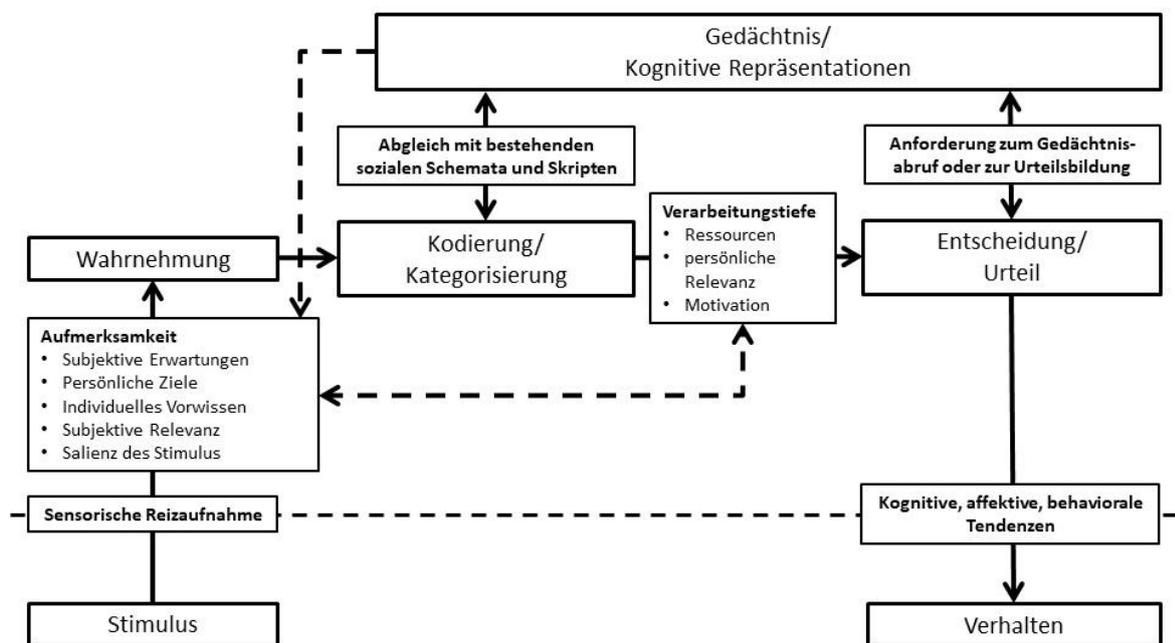


Abbildung 8. Kognitive Stufen der sozialen Informationsverarbeitung (eigene Erweiterung nach Fiedler & Bless, 2003, S. 133)

Bedingt durch begrenzte kognitive Ressourcen sowie die Komplexität und hohe Anzahl sozialer Informationen ist es Individuen nicht möglich, alle vorliegenden Informationen bzw. Umweltreize zu verarbeiten. Nach Miller (1956) ist die Wahrnehmung, Verarbeitung und der Gedächtnisabruf von Informationen auf eine geringe Anzahl an Informationseinheiten

⁴¹ Nach Feuß (2013) kann Aufmerksamkeit als „Selektionsmechanismus bei der Aufnahme von Umgebungsreizen“ (S. 27) und Wahrnehmung als „kognitiver Verarbeitungsprozess aufgenommener Reize“ (S. 26) verstanden werden.

begrenzt.⁴² Bereits nach der sensorischen Reizaufnahme wird durch die Aufmerksamkeit einer Person bestimmt, welche Informationen wahrgenommen werden (Bless et al., 2004; Neisser, 1976; Zebrowitz-McArthur, 1981).

Aber auch die Salienz eines Stimulus kann die Aufmerksamkeit beeinflussen (vgl. Taylor & Fiske, 1978). Ein Reiz ist nicht per se salient, sondern nur durch seine spezifischen Merkmale, die sich von anderen Umweltreizen abheben. So kann z. B. die Ethnizität von Schüler(inne)n (sofern sie auf den ersten Blick erkennbar ist) für die Lehrkraft salient sein, wenn die anderen Schüler(innen) nicht der gleichen Ethnizität angehören. Eine Lehrkraft wird so z. B. die türkische Herkunft eines Schülers oder einer Schülerin besonders wahrnehmen, wenn die anderen Kinder im Klassenverband nicht ebenfalls türkischer Herkunft sind. Kaiser et al. (2016) gehen davon aus, dass die Salienz einer ethnischen Minorität in einer Klasse zu einer akkurateren Urteilsbildung führt, da die Aufmerksamkeit und Wahrnehmung hierdurch auf diese Schüler(innen) gerichtet ist. Aber auch empirische Befunde, die darauf verweisen, dass afroamerikanische Schüler(innen) im Unterricht störender wahrgenommen werden als Schüler(innen) mit europäischen Wurzeln (z. B. Dee, 2005), könnten – sofern die afroamerikanischen Schüler(innen) die ethnische Minorität in der Klasse darstellen – durch die Salienz der Herkunft und einer entsprechend höheren Aufmerksamkeit erklärt werden.

Zudem sorgt auch die individuelle Bedeutsamkeit eines Stimulus dafür, ob die selektive Aufmerksamkeit einer Person auf diesen Stimulus gelenkt wird (vgl. Cocktailparty-Effekt; Pollack, 1958). Somit ist nicht nur die tatsächliche Zielperson von Bedeutung für die Wahrnehmung, sondern auch die subjektive Relevanz und bereits bestehende mentale Strukturen (Bless et al., 2004; Neisser, 1976). Dabei wird die Salienz eines sozialen Objekts und die Bereitschaft dieses wahrzunehmen von internalisierten Werten und Bedürfnissen bestimmt, die wahrnehmende Person ist also nicht passiv den Reizen der Umwelt ausgesetzt (Bruner, 1950; Bruner & Goodman, 1947).⁴³ Die Aufmerksamkeit einer Person wird zudem von persönlichen Zielen (z. B. Bless et al., 2004; Gollwitzer & Moskowitz, 1996; Wyer & Srull, 1989), aber auch von Erwartungen oder impliziten Persönlichkeitstheorien einer Person beeinflusst (Bless et al., 2004; Bruner & Postman, 1949).⁴⁴ Entsprechend ist davon auszugehen, dass auch Ziele oder Erwartungen von Lehrkräften die Aufmerksamkeit bei der Schülerwahrnehmung beeinflussen.

Anzunehmen ist, dass erwartungsinkongruente Ereignisse als besonders salient erscheinen und somit zu einer erhöhten Aufmerksamkeit führen. Zebrowitz-McArthur (1981) geht hin-

⁴² Werden einzelne Informationseinheiten jedoch in *chunks* (Bündelung einzelner Informationen) organisiert, können diese wiederum zu einer höheren Anzahl verarbeitbarer Informationen führen (vgl. Miller, 1956).

⁴³ Bruner & Goodman (1947) konnten in einem Experiment mit Kindern zeigen, dass auch die Interpretation des Wahrgenommenen von der subjektiven und sozialen Wertigkeit abhängt. So schätzten die Kinder die Größe von Münzen im Vergleich zu einfachen Scheiben größer ein als sie tatsächlich waren, insbesondere wenn die Wertigkeit der Münzen stieg. Aber es zeigte sich zudem auch ein Unterschied in der Wahrnehmung zwischen Kindern aus einkommensstarken und -schwachen Haushalten, wobei letztere die Münzen weitaus größer einschätzten. Bruner und Goodman (1947) interpretierten dies als Hinweis auf die Bedeutung der subjektiven Bedeutsamkeit bzw. des Bedürfnisses (*need*) des wahrgenommenen Objekts.

⁴⁴ Erwartungen beeinflussen dabei alle Bereiche des menschlichen Lebens. Bruner und Postman (1949) konnten in einer experimentellen Untersuchung mit Spielkarten zeigen, dass bekannte Objekte schneller erkannt werden und die Erwartung, dass bestimmte Objekte präsentiert werden, zu einer schnelleren Identifikation der Objekte führt.

gegen davon aus, dass die Aufmerksamkeit eher auf erwartungskongruente Ereignisse gelenkt wird, da die Erwartung selbst zu einem erhöhten Erregungslevel führt und entsprechend die Aufmerksamkeit lenkt. Beide Mechanismen wären theoretisch erklärbar: Zum einen kann eine spezifische Erwartung einer Person die Aufmerksamkeit darauf richten, ob dieses Ereignis tatsächlich eintritt. Zum anderen kann jedoch die Auffälligkeit von erwartungsinkongruenten Ereignissen ebenfalls zu einer erhöhten Aufmerksamkeit führen. Ist ein Schüler oder eine Schülerin üblicherweise eher ruhig und zurückhaltend, führt z. B. ein aufbrausendes Verhalten zu mehr Aufmerksamkeit, als bei Schüler(inne)n von denen man derartiges Verhalten gewohnt ist.

Gibson (1969) geht davon aus, dass auch das Erfahrungslevel die Wahrnehmung sowie die weitere Informationsverarbeitung beeinflusst. Erfahrene Personen nehmen bestimmte Informationen wahr, die unerfahrene Personen womöglich übersehen oder ignorieren. Zudem ist anzunehmen, dass erfahrene Personen mehr Informationen aufnehmen und mit bereits bekannten Schemata verknüpfen können (Neisser, 1976). Für den schulischen Kontext könnte dies bedeuten, dass erfahrene Lehrkräfte als Expert(inne)en, im Vergleich zu Lehramtsstudierenden/Novizen, aus dem Lesen eines Zeugnisses mehr Informationen generieren können oder aber andere Textinformationen fokussieren. Lehrkräfte könnten durch ihre Erfahrung bereits kognitive Schemata entwickelt haben, die ihnen die Zuordnung relevanter Informationen erleichtern, da diese mit ihrem bestehenden Vorwissen verknüpft werden können. Novizen hingegen verfügen über kein ausgeprägtes Vorwissen und richten ihre Aufmerksamkeit somit eventuell auf andere Aspekte. Wahrnehmung ist somit ein interaktiver und konstruktiver Prozess, der durch Vorwissen determiniert ist und Antizipationen generiert (Neisser, 1976).

Basierend auf diesen Annahmen ist davon auszugehen, dass kognitive Strukturen auf die soziale Wahrnehmung von Individuen wirken. Wenn angenommen wird, dass Salienz abhängig von individuellen Erwartungen und Zielen ist und sich dies auf die Wahrnehmung auswirkt, ist dies bereits dem Prozess der Kategorisierung vorgeschaltet. Somit ist bereits für die (selektive) Aufmerksamkeit ein Einfluss der gespeicherten Wissensstrukturen bedeutsam (vgl. Abbildung 8). Hamilton et al. (1990) gehen davon aus, dass Stereotype dabei auch bei der Enkodierung von Informationen sowie beim Gedächtnisabruf eine Rolle spielen. So kann die Interpretation neuer Informationen durch bereits vorhandene Stereotype beeinflusst werden. Aber auch Informationen, die zu einem vorhandenen Stereotyp passen, sind leichter aus dem Gedächtnis abrufbar (Hamilton et al., 1990). Insbesondere bei der Bildung eines ersten Eindrucks über Schüler(innen) besteht demnach die Gefahr, diese in stereotypisierte Kategorien einzuordnen.

5.2 Soziale Kategorien im Kontext der Informationsverarbeitung von Lehrkräften

Es liegt in der Natur kognitiver Prozesse, neue Informationen mit bestehenden Wissensinhalten zu verknüpfen sowie Gemeinsamkeiten und Unterschiede zu bereits bekannten Kategorien zu finden (vgl. Bruner, 1957a).⁴⁵ Einzelne Informationen eines wahrgenommenen sozialen Objekts oder aber auch die Summe verschiedener sozialer Informationen werden mit bestehenden mentalen Strukturen (Schemata bzw. Kategorien) abgeglichen (Neisser, 1976). Wird ein spezifisches Muster erkannt (*pattern recognition*), wird das Beobachtungsobjekt bei einer gewissen Übereinstimmungsrate einer bestimmten Kategorie zugewiesen (Neisser, 1976).⁴⁶ Personen werden dabei sozialen Kategorien entweder durch visuell wahrnehmbare Merkmale (z. B. Hautfarbe), geteilte Überzeugungen (z. B. Religionszugehörigkeit) oder Ähnlichkeiten zu einem Typus (z. B. Akademiker) zugeordnet (Klauer, 2008). Lehrkräfte nehmen z. B. die Kleidung von Schüler(inne)n wahr und ordnen diese Schüler(innen) entsprechend einer sozialen Schicht zu (z. B. Dunkake & Schuchart, 2015). Aber auch weitere wahrnehmbare Informationen von Schüler(inne)n, wie z. B. der Vorname, können die Zuordnung zu einer sozialen Kategorie, z. B. einer ethnischen oder sozialen Gruppe, bedingen. Schüler(innen) gehören jedoch nie nur einer einzigen Kategorie an (vgl. Klauer, 2008). So können Schüler(innen) gleichzeitig z. B. einem Geschlecht, einer ethnischen und einer sozialen Gruppe angehören. Die Zugehörigkeit zu einer ethnischen Gruppe kann durch eine entsprechende regionale Herkunft spezifiziert und als Subkategorie aufgefasst werden. Ein Schüler oder eine Schülerin könnte z. B. deutscher Herkunft, im speziellen aus Bayern, sein und zudem aus einem christlich-konservativen Elternhaus stammen.

Dabei werden soziale Kategorien meist mit spezifischen Merkmalen und Eigenschaften, die auch einen stereotypen Charakter aufweisen können, verknüpft (vgl. z. B. Taylor, 1981; vgl. Kapitel 4). Die zugeordneten Kategorien bei der Personenwahrnehmung können insbesondere bei wenigen weiteren verfügbaren Informationen über eine Person zur Aktivierung stereotyper Zuschreibungen führen, die zwar einerseits eine Vereinfachung der sozialen Umwelt darstellt, gleichzeitig jedoch auch zu Urteilsfehlern führen kann (z. B. Klauer, 2008). Diese Simplifizierung der komplexen und unzähligen sozialen Umweltreize ist dabei bedingt durch begrenzte kognitive Ressourcen notwendig (Bless et al., 2004; Miller, 1956). So weist bereits Lippmann (1922) auf die Notwendigkeit der kognitiven Vereinfachung der sozialen Umwelt hin: „For the real environment is altogether too big, too complex, and too fleeting we have to reconstruct it on a simpler model before we can manage with it“ (S. 16).

Kategorisierungsprozesse sind damit zentrale Bestandteile der sozialen Informationsverarbeitung (Cantor & Mischel, 1979). Hierzu ist es notwendig, die sozialen Informationen zu kodieren bzw. zu strukturieren und sie durch bestimmte Muster in Schemata zu ordnen (Bruner,

⁴⁵ „Suppose a person to be shown any straight line and to be directed to imagine another one prolonging it. He will at once be able to comply, if the second line is directed to be perpendicular to, intersecting with, or in any other prescribed manner related to the first one“ (Spearman, 1923, S. 94). In Anlehnung an Spearman (1923) beschreibt Bruner (1957a) hierzu die Charakteristik des Mentalen als die Eigenschaft, dass Personen konstant über die gegebenen Informationen hinausgehen.

⁴⁶ Es ist davon auszugehen, dass die Bildung bestimmter Kategorien bereits in der frühen Kindheit beginnt (z. B. Geschlechtskategorien) und sich auch im gesamten Lebensverlauf neue Kategorien bilden können (vgl. Bar-Tal, 1996).

1957a; Klauer, 2008). Zum einen können dabei Informationen einer Person mit der Ähnlichkeit zu bestehenden Kategorien abgeglichen werden (Bruner, 1957a, Brewer, 1988; Fiske & Neuberg, 1990). Zum anderen werden bestehende Gemeinsamkeiten und Unterschiede durch die Bildung von Kategorien zunächst geordnet (Oakes, 1987). Dabei werden insbesondere die Zugänglichkeit (*accessibility*) und die Passung (*fit*) als bedeutsame Faktoren für die Kategorisierung angesehen (Bruner, 1957b; Oakes, 1987). Wenn Lehrkräfte Schüler(innen) kategorisieren, ist es somit notwendig, dass die entsprechende soziale Kategorie auch im Gedächtnis zugänglich bzw. verfügbar ist. Des Weiteren muss die Kategorie als passend für das spezifische Kind oder den Jugendlichen wahrgenommen werden. Die Kategorisierung ermöglicht es der Lehrkraft, die hohe Anzahl an sozialen Informationen zu systematisieren und Erwartungen zu generieren. Hierdurch kann sie auf neue soziale Informationen schneller reagieren. Problematisch ist jedoch, dass der Kategorisierungsprozess ein höchst automatisierter Prozess ist, der kaum durchbrochen werden kann. Bruner (1957a) beschreibt dies auch als „*problem of creativity*“ (S. 50). Durch die verinnerlichte menschliche Tendenz Muster zu erkennen, ist es eine große Herausforderung, aus diesen kognitiven Strukturen zumindest teilweise ausbrechen. Es ist davon auszugehen, dass vorherige Erfahrungen mit Personen einer sozialen Kategorie als Vergleichsbasis für Urteile über andere Mitglieder der sozialen Kategorie dienen, aber auch die Erwartung erzeugen, dass Individuen der sozialen Kategorie den bereits bekannten Individuen dieser Gruppe ähneln (Manis, Biernat & Nelson, 1991; Manis & Paskewitz, 1984). Haben Lehrkräfte mit bestimmten ethnischen und sozialen Gruppen in ihrer bisherigen Berufslaufbahn z.B. überwiegend positive Erfahrungen gesammelt und diese Gruppe als leistungsstarke Schüler(innen) kategorisiert, kann dies dazu führen, dass neue Schüler(innen), die diesen sozialen Kategorien angehören, entsprechend hohe Leistungserwartungen entgegengebracht werden.

Die von Personen als angenehm empfundene und angestrebte Konsistenz kognitiver Strukturen (Festinger, 1957) zeigt sich auch in der automatisierten Kategorisierung von Personen. Hugenberg und Bodenhausen (2004) konnten in einer experimentellen Untersuchung zeigen, dass bei unklaren ethnischen Merkmalen einer Person weitere inhaltlich unabhängige Informationen (wie ein glücklicher oder feindseliger Gesichtsausdruck) zur Urteilsbildung herangezogen werden. Dabei zeigen v. a. Personen mit einem hohen Maß negativer impliziter Assoziationen gegenüber ethnische Minderheiten eine Zuordnung zu dieser Gruppe, wenn ein negativer Gesichtsausdruck vorliegt (Hugenberg & Bodenhausen, 2004). Die Zuordnung zu einer sozialen Kategorie hängt somit auch davon ab, ob stereotypkonsistente Informationen wahrgenommen werden. Bei Abweichungen von dem typischen Muster der Kategorie werden entsprechend Subkategorien gebildet (Klauer, 2008).

Eine automatische Beurteilung von Personen, die auf Stereotypen basiert, findet nach Locksley, Hepburn und Ortiz (1982a, 1982b) jedoch nur statt, wenn keine bzw. nicht ausreichend individualisierte Informationen zur Verfügung stehen und nur auf stereotype Annahmen zurückgegriffen werden kann (vgl. auch Klauer, 2008). Ugwuegbu (1979) konnte zeigen, dass unter bestimmten Bedingungen bzw. für bestimmte soziale Gruppen trotz vieler Informationen eine kategoriebasierte Informationsverarbeitung stattfindet. So beurteilten europäischstämmige und afroamerikanische Personen in einer experimentellen Studie bei wenigen Beweisen in einem Strafprozess einen Verdächtigen der anderen Ethnizität als der eigenen

deutlich häufiger als schuldig (Ugwuegbu, 1979). Afroamerikanische Versuchsteilnehmer(innen) beurteilten Verdächtige der eigenen Ethnie bei einer Vielzahl an Beweismitteln jedoch weiterhin als weniger schuldig (Ugwuegbu, 1979). Trotz ausreichender Informationen erfolgte bei einem Teil der Versuchspersonen dennoch eine kategoriebasierte Informationsverarbeitung, die in diesem Fall eine Eigengruppenpräferenz darstellt. Somit kann davon ausgegangen werden, dass zahlreiche individuumsbasierte Informationen in der Konsequenz nicht zwangsläufig zu einem objektiveren Urteil führen müssen.

Nach Casper, Rothermund und Wentura (2010) ist die Kategorisierung von Personen und Aktivierung sozialer Stereotype an den Kontext gebunden. Somit sind soziale Kognitionen nur kontextualisiert sinnvoll zu verstehen (Bless et al., 2004; Neisser, 1976).⁴⁷ Je nach Situation kann eine soziale Handlung z. B. als adäquat oder als unangebracht wahrgenommen werden. Aber auch Personen können in unterschiedlichen Umgebungen oder Situationen anders wahrgenommen und beurteilt werden (z. B. Bierhoff, 1986). Dabei ist zentral, dass auch die Wahrnehmung des Kontexts, d. h. die Relation des Stimulus zur Umgebung, auch von den subjektiven Erfahrungen der wahrnehmenden Person abhängig ist. Zu beachten ist, dass für jede soziale Gruppe eine Vielzahl und z. T. unterschiedliche Attribute je nach spezifischem Kontext vorliegen. Beispielsweise wird der Gruppe der türkischen Migrant(inn)en im Kontext einer Gebetsumgebung eher das Stereotyp *religiös* zugeschrieben bzw. aktiviert (Casper et al., 2010). Zur Aktivierung des Stereotyps ist entsprechend auch der Kontext relevant, in dem der stigmatisierten Gruppe bestimmte Attribute zugeschrieben werden. Hechtman und Rosenthal (1991) zeigten zudem, dass das Lehrkraftverhalten durch stereotypassoziierten Kontext beeinflusst wird. So verhielten sich Lehrkräfte gegenüber Mädchen und Jungen in entsprechend geschlechtsstereotypen Domänen positiver als in stereotypinkonsistenten Domänen (vgl. Hechtman & Rosenthal, 1991). Im Fall der vorliegenden Arbeit ist dies durch den schulischen Leistungskontext gegeben. Da auch Lehrkräften die Ergebnisse der repräsentativen Schulleistungsstudien, wie z. B. PISA, TIMSS oder IGLU bekannt sind, in denen geringere Kompetenzen von Schüler(inne)n mit (türkischem) Migrationshintergrund und/oder geringem sozioökonomischem Status berichtet werden, ist naheliegend, dass Lehrkräfte für den schulischen Kontext die Kategorie Herkunft mit dem Attribut (geringere vs. höhere) Leistungsfähigkeit bzw. Kompetenz verknüpfen.

Zusammenfassend lässt sich festhalten, dass Wahrnehmungsprozesse von bestehenden Stereotypen beeinflusst werden (vgl. z. B. Schmid Mast & Krings, 2008). Dabei ist davon auszugehen, dass sie auf Gedächtnis- als auch Interpretationsprozesse der sozialen Urteilsbildung wirken (Peterson & Six, 2008). Personen sind jedoch häufig mit Informationen konfrontiert, die mit ihren bestehenden Stereotypen oder Kategorien konfliktieren oder sogar völlig inkompatibel sind. Dies kann z. B. selektive Aufmerksamkeit bei der Wahrnehmung zur Folge haben (Neisser, 1976). Dabei unterliegen Personen bei der sozialen Informationsverarbeitung nicht nur potenziellen Verzerrungen in der Wahrnehmung, sondern auch in anderen Informationsverarbeitungsabschnitten wie z. B. dem Gedächtnisabruf (z. B. Loftus & Greene,

⁴⁷ „We pursue the main idea that stereotypes are organized as multiple context-specific schemas that are typically activated only by a combination of category and context information“ (Casper et al., 2010, S. 131).

1980). Diese können zu entsprechenden Urteilsverzerrungen oder -fehlern führen, denen auch Lehrkräfte bei der Beurteilung von Schüler(inne)n unterliegen können.

Bereits durch subjektive und selektive Aufmerksamkeitsprozesse besteht die Gefahr von Wahrnehmungsfehlern. Simons (o. J.; 2010) *Gorilla-Studie* veranschaulicht, wie selektiv die menschliche Wahrnehmung verläuft und dass Personen, die auf das Erscheinen eines Gorillas warten, andere unerwartete Aspekte nicht wahrnehmen (vgl. auch Chabris & Simons, 2013). Diese Unaufmerksamkeitsblindheit (*inattentional blindness*) konnte in weiteren empirischen Studien nachgewiesen werden (z. B. Stothart, Wright, Simons & Boot, 2017). Personen nehmen Objekte und soziale Stimuli somit stets eingefärbt durch den eigenen Aufmerksamkeitsfokus war. Dies kann dazu führen, dass gravierende Veränderungen nicht bemerkt werden (*change blindness*; Simons, o. J.). Auch Lehrkräfte können, wenn sie ihre Aufmerksamkeit auf bestimmte Aspekte im Unterricht fokussieren, andere soziale Situationen übersehen. Auch unterschiedliche emotionale Zustände können den Prozess sozialer Urteilsbildung beeinflussen. Bodenhausen, Sheppard und Kramer (1994) konnten z. B. zeigen, dass ein ärgerlicher emotionaler Zustand eher zu stereotypen Urteilen führt als z. B. ein trauriger emotionaler Zustand. Auch bei Lehrkräften besteht die Gefahr, dass die eigenen affektiven Zustände Auswirkungen auf die Wahrnehmung und Beurteilung von Schüler(inne)n bestimmter ethnischer und sozialer Gruppen haben können. Neben diesen Aspekten sind weitere relevante Urteilsverzerrungen und -fehler zu nennen, die auch im schulischen Kontext, v. a. im Zusammenhang mit einer heterogenen Schülerschaft, bedeutsam sein können.

Dabei ist v. a. der *halo-effect* zu nennen, der beschreibt, dass ein bekanntes Merkmal einer Person die Einschätzung weiterer Merkmale beeinflusst bzw. überstrahlt (vgl. Thorndike, 1920). So konnten z. B. Nisbett und Wilson (1977) in einer experimentellen Untersuchung zeigen, dass bestimmte Eigenschaften einer Person entweder als positiv oder negativ bewertet werden, je nachdem ob der erste Gesamteindruck positiv oder negativ ausfällt. Greenwald und Banaji (1995) gehen davon aus, dass sich insbesondere durch den halo-effect implizite Einstellungen und implizite Persönlichkeitstheorien auf die Urteilsbildung auswirken können. Da wir uns einen ersten Eindruck meist zunächst über visuell verfügbare Informationen bilden, kann die äußere Erscheinung z. B. auch Auswirkungen auf die Einschätzung der Intelligenz und Kompetenz haben (Hödl-Nossberger, 1964; Jackson, Hunter & Hodge, 1995). Zudem existieren empirische Hinweise darauf, dass die wahrgenommene physische Attraktivität sich ebenfalls auf die Notengebung durch Lehrkräfte auswirkt – attraktivere Schüler(innen) erhielten in einer Untersuchung von Dunkake et al. (2012) entsprechend auch die besseren Noten. Einige empirische Untersuchungen konnten klären, dass die Attraktivität von Personen nichts darüber aussagt, ob sie tatsächlich in einem bestimmten Bereich besser oder intelligenter sind (Jackson et al., 1995). Dennoch verweisen Studien darauf, dass als attraktiv wahrgenommene Personen auch kompetenter eingeschätzt werden (z. B. Eagly, Ashmore, Makhijani & Longo, 1991; Jackson et al., 1995).⁴⁸

⁴⁸ Jackson et al. (1995) konnten in einer Meta-Analyse von 113 Studien zeigen, dass bei Erwachsenen kein Zusammenhang zwischen wahrgenommener Attraktivität und tatsächlicher Intelligenz vorlag. Für Kinder zeigte sich jedoch ein kleiner Zusammenhang, der von den Autoren durch die womöglich stärkeren Auswirkungen von Erwartungseffekten bei jüngeren Kindern als bei Erwachsenen erklärt wird.

Als Erklärungsansatz für den subjektiv wahrgenommenen Zusammenhang physischer Attraktivität und Intelligenz oder Kompetenz wird auch die *status generalization theory* (z. B. Webster & Foschi, 1988) herangezogen. Diese beschreibt, dass Personen den Status einer Person generalisieren und hieraus Schlüsse auf andere Merkmale ableiten. Dies ist vergleichbar mit *impliziten Persönlichkeitstheorien*, bei denen aus einem Merkmal einer Person Schlüsse über weitere Attribute gezogen werden. Im Einklang mit der *status generalization theory* zeigten Untersuchungen, dass Effekte der Generalisierung deutlich höher sind, wenn keine tatsächlichen Informationen zu den einzuschätzenden Bereichen wie z. B. Intelligenz vorliegen (z. B. Jackson et al., 1995). Der *mere exposure effect* kann als Effekt impliziter Einstellungen angesehen werden (Greenwald & Banaji, 1995). Eine häufige wiederholte Wahrnehmung eines Einstellungsobjekts kann zu einer positiveren Bewertung führen (Zajonc, 1968). Greenwald und Banaji (1995) gehen davon aus, dass der *mere exposure effect* eine Fehlinterpretation des Mögens des Einstellungsobjekts ist, die durch die Leichtigkeit der Identifikation (als Folge der häufigen und wiederholten Darbietung) des Einstellungsobjekts erfolgt. Somit kann auch bei Lehrkräften der häufigere Kontakt zu bestimmten ethnischen und sozialen Gruppen dazu führen, dass diese positiver bewertet werden, ohne dass es hierfür einen objektiven Grund gibt. Der *incremental threat effect* besagt hingegen, dass wiederholter Kontakt auch zu eher negativen Urteilen führen kann, wenn Individuen sich bedroht fühlen (Crisp, Hutter & Young, 2009).

Bedingt durch die unzähligen Informationen, mit denen ein Individuum in sozialen Situationen konfrontiert ist, sowie durch komplexe soziale Situationen oder unter Zeitdruck, sind Personen oftmals nicht bereit oder in der Lage, diese vertieft zu verarbeiten. Sie orientieren sich stattdessen an mentalen Abkürzungen (vgl. Bless et al., 2004). In diesem Zusammenhang sind Heuristiken von Bedeutung, die als kognitive Instrumente zur Urteilsbildung mit Hilfe von Daumenregeln zu verstehen sind und die unter geringem Aufwand häufig valide Ergebnisse liefern (Bless et al., 2004). In diesem Zusammenhang führen Fiske und Taylor (2013) auch den Begriff des „cognitive miser“ (S. 37) („kognitiver Geizhals“) ein. Häufig sind diese mentalen Abkürzungen effizient und führen zu zutreffenden Urteilen. Jedoch ist dieser Prozess durch die Übersimplifizierung auch fehleranfällig und kann unter bestimmten Bedingungen zu systematischen Verzerrungen führen (vgl. Bless et al., 2004). Kahneman und Tversky (1973) beschreiben dies folgendermaßen: „In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors“ (S. 237).

Nach Kahneman und Tversky (1972; 1973; Tversky & Kahneman, 1973, 1974) lassen sich grundlegend drei heuristische Informationsverarbeitungsprozesse unterscheiden: die *Repräsentativitäts-*, die *Verfügbarkeits-* und die *Ankerheuristik*.⁴⁹ Wird ein Individuum z. B. durch äußere Merkmale als typisch für eine bestimmte soziale Gruppe wahrgenommen, halten Personen es für subjektiv wahrscheinlicher, dass die Person dieser Gruppe auch angehört (Repräsentativitätsheuristik; Kahneman & Tversky, 1972). Dabei werden zur Urteilsbildung eher subjektiv wahrgenommene Ähnlichkeiten als Beurteilungsgrundlage herangezogen, als die Wahrscheinlichkeit, die sich aus der realen Auftretenshäufigkeit (*base rate*) ergibt (Kahneman

⁴⁹ Überblick über weitere Urteilsverzerrungen bei Jungermann, Pfister und Fischer (1998).

& Tversky, 1973).⁵⁰ Kahneman und Tversky (1973) gehen davon aus, dass drei Informationen berücksichtigt werden müssen, um eine statistische Vorhersage zu treffen: (1) Hintergrundinformationen (base rate), (2) Merkmale des Individuums und (3) die angenommene Akkuratheit der Vorhersage. Je geringer die angenommene Akkuratheit der Vorhersage ist, desto wahrscheinlicher ist ein tatsächliches Urteil in Anlehnung an die base rate. Personen tendieren jedoch dazu, die base rate zu ignorieren und sich auf Ähnlichkeiten zu verlassen, was zu einer fehlerhaften Urteilsbildung führen kann. Neben der Repräsentativitätsheuristik formulierten Tversky und Kahneman (1973) ebenfalls die Verfügbarkeitsheuristik als Möglichkeit der Urteilsbildung. Angenommen wird, dass Urteile über Häufigkeiten und Wahrscheinlichkeiten sich nicht immer an den tatsächlichen Wahrscheinlichkeiten orientieren, sondern durch die Abruflichkeit aus dem Gedächtnis verfälscht werden. Informationen, die durch häufige Wiederholungen wahrgenommen werden sind leichter aus dem Gedächtnis abrufbar und werden somit auch als wahrscheinlicher angesehen (Tversky & Kahneman, 1973). Higgins, King und Mavin (1982) konnten in einer experimentellen Untersuchung ebenfalls demonstrieren, dass Personeneigenschaften, die häufig genannt werden, auch bei einem späteren Gedächtnisabruf leichter erinnert werden als Eigenschaften, die selten bei einer Personenbeschreibung genannt wurden. Denkbar ist, dass Lehrkräfte, die häufig mit negativen Informationen über Schüler(innen) mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status konfrontiert sind, diese Informationen auch leichter aus dem Gedächtnis abrufen können.

Aber auch vorhandene Vorinformationen oder bereits bekannte Werte könnten die Schätzung einer Person in Richtung des vorher gesetzten Wertes beeinflussen (Ankerheuristik, Tversky & Kahneman, 1974). So können z. B. Informationen zu schulischen Leistungen durch eine vorherige Lehrkraft von Schüler(inne)n als Anker dienen (vgl. z. B. Dünnebier et al., 2009), wenn Erwartungen an zukünftige Leistungen gebildet werden. Hatte ein Schüler oder eine Schülerin im vorhergehenden Schuljahr eher schlechtere Zensuren in einem bestimmten Fach, kann dies die neue Lehrkraft insofern beeinflussen, als dass eher geringere Erwartungen an diesen Schüler bzw. diese Schülerin gerichtet werden. Diese Urteilsverzerrungen und -fehler laufen dabei meist automatisiert ab und unterliegen nicht der direkten Kontrolle von Personen. Dabei lassen sich sozial-kognitive Prozesse nicht nur als automatisierte Prozesse beschreiben, sondern einige Prozesse verlaufen auch kontrolliert. Der folgende Abschnitt stellt entsprechend dar, welche Arten der Informationsverarbeitung unterscheidbar sind und v. a. welche Bedeutung dies im Zusammenhang der Urteilsbildung von Lehrkräften hat.⁵¹

⁵⁰ Anzumerken ist, dass die Festlegung spezifischer Urteilsfehler durch den Widerspruch zu statistischen Annahmen bzw. Wahrscheinlichkeitsaussagen durch u. a. Gigerenzer (1991) in Frage gestellt wurde.

⁵¹ Dabei sind noch zahlreiche weitere potenzielle Verzerrungen und Fehler bei der Urteilsbildung möglich. Einen Überblick dazu gibt z. B. Funder (1987).

5.3 Top-down und bottom-up Prozesse bei der Informationsverarbeitung von Lehrkräften

„Dry psychologists who champion dual-process models are usually not stuck on two. Few would come undone if their models were recast in terms of three processes, or four, or even five. Indeed, the only number they would not happily accept is one“ (Gilbert, 1999, S. 4). Während zu Beginn der psychologischen Theoriebildung und empirischen Erforschung zur sozialen Eindrucks- und Urteilsbildung noch kaum auf unterschiedliche kognitive Verarbeitungsprozesse eingegangen wurde (z. B. Asch, 1946), werden mittlerweile in sozialpsychologischen Theorien üblicherweise Zwei-Prozess-Modelle zur Erklärung sozial-kognitiver Informationsverarbeitungsprozesse herangezogen (Übersicht bei Chaiken & Trope, 1999). Nach Gilbert (1999) ist jedoch nicht immer klar ob – u. a. bedingt durch die Komplexität menschlichen Erlebens und Verhaltens – sich dies auf lediglich zwei Prozesse reduzieren lässt.

Grundlegend besteht überwiegend Einigkeit über die theoretische Annahme, dass sich soziale Informationen auf unterschiedliche Arten verarbeiten lassen. Dabei lässt sich zum einen eine überwiegend unintentionale Informationsverarbeitung, die keiner bewussten Regulation unterliegt und somit eher *automatisiert* abläuft, von einer bewusst regulierten Informationsverarbeitung, die entsprechend eher *kontrolliert* verläuft, abgrenzen (Bless et al., 2004).⁵² Es ist davon auszugehen, dass der überwiegende Anteil menschlicher Kognitionen eher automatisiert und auf unbewusster Ebene verlaufen (vgl. Bargh & Chartrand, 1999). Zum anderen kann unterschieden werden, ob die Informationsverarbeitung eines sozialen Stimulus überwiegend durch bestehende kognitive Strukturen einer Person geleitet ist. Dies wird je nach theoretischem Ansatz als *konzeptgeleiteter*, *kategoriebasierter*, *automatisierter* oder allgemein als *top-down* Prozess bezeichnet (vgl. u. a. Bless et al., 2004). Dem gegenüber stehen sogenannte *datengeleitete*, *merkmalsbasierte*, *kontrollierte* oder allgemein *bottom-up* Prozesse, die überwiegend Informationen des spezifischen Stimulus in den Verarbeitungsprozess einbeziehen (vgl. u. a. Bless et al., 2004). Übertragen auf die Urteils- und Erwartungsbildung von Lehrkräften bedeutet dies, dass bei einem top-down Prozess Informationen überwiegend basierend auf der Kategoriezugehörigkeit von Schüler(inne)n verarbeitet werden. Im Gegensatz dazu werden bei einer bottom-up Verarbeitung individuelle Schülermerkmale einbezogen.

Ob ein simultaner oder zeitlich gestufter Ablauf von top-down und bottom-up Prozessen stattfindet, wird in verschiedenen theoretischen Ansätzen unterschiedlich beantwortet. So gehen z. B. Kunda und Thagard (1996) davon aus, dass individuelle und kategoriale Informationen simultan verarbeitet werden. Meist wird jedoch ein zweistufiger Prozess postuliert, in dem zunächst top-down Prozesse ablaufen, auf die anschließend bottom-up Prozesse folgen (z. B. Brewer, 1988; Devine, 1989). Neben dem Zwei-Stufen-Modell von Devine (1989), in dem automatisierte Prozesse von kontrollierten Prozessen unterschieden werden und

⁵² Während Bless et al. (2004) postulieren, dass eine automatische Informationsverarbeitung dem Individuum selbst nicht bewusst ist, relativieren Tzelgov, Porat und Henik (1997) dies, indem sie davon ausgehen, dass ein automatisierter Vorgang nicht zwangsläufig unbewusst ablaufen muss. Der Ablauf eines Prozesses kann zwar verinnerlicht sein und ohne hohe kognitive Beanspruchung in automatisierter Weise ablaufen, ohne dass der Prozess bewusst eingeleitet werden muss, auf einer zweiten Ebene kann der Prozessablauf dem Individuum jedoch durchaus bewusst sein.

davon ausgegangen wird, dass beide Prozesse auch unabhängig voneinander ablaufen können, ist v. a. das Zwei-Prozess-Modell von Brewer (1988) an dieser Stelle erwähnenswert.⁵³ In diesem Modell erweiterte Brewer (1988) zunächst die bis dahin bestehende Theorien (z. B. Asch, 1946; Anderson, 1965) um Aspekte der Begrenzung kognitiver Kapazitäten und möglicher Informationsüberlastung. Zudem greift sie im Gegensatz zu Asch (1946) in diesem Modell die Ausweitung einer merkmalsgeleiteten Informationsverarbeitung durch die Bedeutung sozialer Kategorien bei der Eindrucksbildung auf, wofür z. B. durch Pryor und Ostrom (1981) bereits empirische Belege vorlagen. Nach Brewer (1988) beginnt jede Eindrucksbildung mit einer automatischen, unbewussten und unintentionalen *Identifikation* einer Person, wobei die Zuordnung zu bereits vorhandenen Kategorien erfolgt. Dieser Schritt ist vergleichbar mit der *primitiven Kategorisierung* nach Bruner (1957b) und ist als automatisierter Prozess zu verstehen.⁵⁴ Anschließend werden individuelle Ziele integriert und entschieden, ob die wahrgenommene Person von Relevanz ist. Ist dies der Fall, erfolgt nach Brewer (1988) die *Kategorisierung*. Dabei werden allgemeine Muster im Sinne sozialer Kategorien erkannt und Personen diesen Kategorien zugeordnet.⁵⁵ Entspricht der Stimulus nicht dem Prototyp der sozialen Kategorie, wird eher eine passende Subkategorie gesucht als eine neue bzw. andere Kategorie.⁵⁶ Sind diese beiden Varianten jedoch nicht möglich, wird nach Brewer (1988) der Prozess der *Individuation* initialisiert. Dieser ist immer noch stark durch die zunächst zugeschriebene Kategorie geprägt, ermöglicht es jedoch bei der Eindrucksbildung auch individuelle Merkmale zu berücksichtigen. Bei persönlicher Bedeutsamkeit des sozialen Stimulus für die urteilende Person kommt es nicht zur Individuation, sondern zur *Personalisation*, wobei individuumbezogene Informationen maßgeblich für die Eindrucksbildung sind (Brewer, 1988). Neben diesem und anderen Zwei-Prozess-Modellen der Eindrucksbildung hat insbesondere das Kontinuum-Modell der Eindrucksbildung von Fiske und Neuberg (1990) eine hohe Bedeutung für die Erklärung der Urteilsbildung von Lehrkräften.

⁵³ Devine (1989) geht zudem davon aus, dass automatisierte Prozesse nicht bewusst sein müssen und trotz bewusster Vermeidungsbemühungen auftreten. Im Gegensatz dazu sind kontrollierte Prozesse als intentional anzusehen und erfordern die Aufmerksamkeit des Individuums. Zudem sind sie flexibler.

⁵⁴ Brewer (1988) geht davon aus, dass nur eine begrenzte Anzahl sozialer Kategorien simultan in diesem ersten automatisierten Prozess verarbeitet werden kann.

⁵⁵ Ist eine soziale Kategorie aktiviert, senkt dies nach Brewer (1988) den Schwellwert für die Zuordnung des Stimulus zu dieser Kategorie.

⁵⁶ Sind weitere Informationen vorhanden, die nicht in Übereinstimmung mit der zugeordneten Kategorie sind, können die inkonsistenten Informationen ignoriert oder abgewertet werden oder sie können – sofern sie bei mehreren Vertreter(inne)n der Kategorie auffindbar sind – dazu führen, dass eine Subgruppe der Kategorie gebildet wird (vgl. Brewer, 1988).

Kontinuum-Modell der Eindrucksbildung (Fiske & Neuberg, 1990)

Im Gegensatz zum Zwei-Prozess-Modell von Brewer (1988) wird der Prozess der Informationsverarbeitung von Fiske und Neuberg (1990) nicht in automatisierte und kontrollierte Anteile dichotomisiert, sondern von der Annahme geleitet, dass dieser Prozess als Kontinuum anzusehen ist.⁵⁷ Fiske und Neuberg (1990) beschreiben dieses Kontinuum dabei als stufenweise Abfolge unterschiedlicher Verarbeitungstiefen (vgl. Abbildung 9). Dabei erfolgt direkt nach der ersten Wahrnehmung einer Person eine *primäre Kategorisierung*, d.h. sensorisch wahrnehmbare Merkmale eines Menschen werden mit bestehenden Wissensstrukturen verknüpft und die Person wird aktivierten Kategorien zugeordnet. So nimmt eine Lehrkraft in einem Bruchteil von Sekunden z. B. anhand der Kleidung, Frisur, Stimme etc. das Geschlecht eines Schülers bzw. einer Schülerin wahr. Nach dem Modell von Fiske und Neuberg (1990) kann dies als ein Pol des Kontinuums angesehen werden, der sich zunächst nur auf Informationen zur Zugehörigkeit des Zielobjekts zu einer sozialen Kategorie stützt und keine besondere Aufmerksamkeit erfordert.

Hat die Zielperson keine Bedeutung für die/den Beurteilende(n) oder weckt kein Interesse, kann der Informationsverarbeitungsprozess an dieser Stelle bereits abgebrochen werden. In Folge dessen basieren die affektiven, kognitiven oder behavioralen Tendenzen der Person lediglich auf der Zuschreibung der Zielperson zu einer sozialen Kategorie. Da davon auszugehen ist, dass Lehrkräfte bei der Beurteilung von Schüler(inne)n ein Interesse an den Schüler(inne)n haben und ihnen ebenfalls die Relevanz bewusst ist, kann angenommen werden, dass die Eindrucksbildung von Lehrkräften in der Regel nicht auf diesem Ende des Kontinuums stehen bleibt, sondern dass sie den Schüler(inne)n in einem nächsten Schritt Aufmerksamkeit schenken, die zu einer *konfirmatorischen Kategorisierung* führt. In diesem Schritt werden Informationen berücksichtigt, die über die primäre Kategorisierung hinausgehen und durch die erhöhte Aufmerksamkeit wahrgenommen wurden. Dabei wird geprüft, ob diese zusätzlichen Informationen konsistent mit den Annahmen über die zugehörige Kategorie sind. So könnte bei einem Schüler oder einer Schülerin, der/die der Kategorie *Kind mit Migrationshintergrund* zugeordnet wurde, die Aufmerksamkeit der Lehrkraft auf z. B. spezifische sprachliche Merkmale gelenkt werden und geprüft werden, ob dies konsistent mit der Lehrkraftannahme über sprachliche Fähigkeiten von Kindern mit Migrationshintergrund ist. Ist dies der Fall, besteht nach Fiske und Neuberg (1990) kein Bedarf für eine vertiefte Verarbeitung und der Informationsverarbeitungsprozess wird an dieser Stelle beendet, was zu einer kategoriebasierten Urteilsbildung führt.

Wurden hingegen Informationen wahrgenommen, die nicht mit der sozialen Kategorie vereinbar sind, führt dies zur *Rekategorisierung*. Dabei werden neue Kategorien in den bereits vorhandenen mentalen Strukturen gesucht zu denen das Zielobjekt besser passen könnte, oder

⁵⁷ Des Weiteren unterscheiden sich beide Modelle darin, dass das Zwei-Prozess-Modell (Brewer, 1988) sich auf die mentale Repräsentation des Eindrucks fokussiert, wobei die Urteilsbildung im Fokus der Theorie von Fiske und Neuberg (1990) steht.

neue Kategorien oder Subkategorien werden gebildet.⁵⁸ Eine besondere eloquente Ausdrucksweise von Schüler(inne)n mit niedrigem sozioökonomischem Status könnte z. B. gegen die stereotypen Assoziationen mit dieser sozialen Kategorie stehen. Entsprechend könnten Lehrkräfte eine Subkategorie für diese Schülergruppe bilden. Nach Fiske und Neuberg (1990) kann auch die Ähnlichkeit der Zielperson mit anderen Personen, die bereits bekannt sind, verglichen werden. Eine Lehrkraft könnte sich z. B. an Schüler(innen) erinnern, die sie bereits kennt, die trotz eines niedrigen sozioökonomischen Status sprachlich äußerst gewandt waren. Ist dieser Prozess erfolgreich, erfolgt die Zuschreibung der Merkmale der neuen Kategorie. Ist auch dieser Schritt nicht erfolgreich, kommt es zu einer schrittweisen *Integration der individuellen Merkmale*. Dies bedeutet z. B., dass die Lehrkraft mehr Aufmerksamkeit auf den einzelnen Schüler bzw. die einzelne Schülerin legt und die individuellen Fähigkeiten genau betrachtet. Je weiter die Informationsverarbeitung auf diesem Kontinuum fortschreitet, desto mehr individuelle Informationen werden berücksichtigt.

Da soziale Kategorien häufig mit spezifischen Merkmalen assoziiert sind und diese bei einer Generalisierung auf einzelne Mitglieder der Kategorie als Stereotype wirksam sein können, zeigt sich die hohe Bedeutung einer individuumsbasierten Informationsverarbeitung von Lehrkräften v. a. im Kontext einer ethnisch und sozial heterogenen Schülerschaft. Im Hinblick auf positive Stereotype gegenüber Schüler(inne)n ohne Migrationshintergrund und mit einem hohen sozioökonomischen Status (vgl. Abschnitt 4.2), könnte eine kategoriebasierte Informationsverarbeitung zu positiv verzerrten Urteilen führen. Die eher negativen Stereotype gegenüber Schüler(inne)n mit Migrationshintergrund und Schüler(inne)n mit niedrigem sozioökonomischen Status (vgl. Abschnitt 4.2) könnten bei einer eher automatisierten Informationsverarbeitung hingegen eher zu negativ verzerrten Urteilen führen.

Die Befunde einer experimentellen Studie von Glock und Krolak-Schwerdt (2014) deuten dabei ebenfalls auf eine eher kontrollierte Informationsverarbeitung von Lehrkräften bei der Beurteilung von Schüler(inne)n mit türkischem Migrationshintergrund hin. So erinnerten sich die Versuchspersonen bei Schüler(inne)n türkischer Herkunft an mehr Informationen der Fallvignette als bei Schüler(inne)n ohne Migrationshintergrund (Glock & Krolak-Schwerdt, 2014). Auch die Ergebnisse von Kaiser et al. (2016) zur akkuraten Einschätzung von Schüler(inne)n mit Migrationshintergrund lassen eher auf eine individuumsbasierte Informationsverarbeitung schließen, da die akkurate Einschätzung die genaue Betrachtung individueller Informationen voraussetzt. Die Befunde von Ready und Chu (2015) und Ready und Wright (2011) zu einer Überschätzung von Vorschulkindern mit hohem sozioökonomischem Status und einer Unterschätzung von Kindern mit niedrigem sozioökonomischem Status, sprechen hingegen für eine eher automatisierte bzw. kategoriegeleitete Informationsverarbeitung. In den aktuellen Recherchen konnten jedoch keine empirischen Studien aufgefunden werden, die unterschiedliche Informationsverarbeitungsprozesse für unterschiedliche Schülergruppen erfasst haben. Entsprechend bleibt bislang lediglich die Möglichkeit, durch Rückschlüsse von herkunftsassoziierten Lehrkrafturteilen auf die Art der Informationsverarbeitung zu schließen.

⁵⁸ Personen weisen eine geringere Tendenz auf, Personen der eigenen sozialen Kategorien in Subgruppen für soziale Kategorien einzuteilen, denen sie selbst nicht angehören. Erklärt wird dies auch durch die geringere Wahrnehmung von Varianz und unterschiedlichen Ausprägungen in anderen Kategorien als der eigenen (Fremdgruppenhomogenitätseffekt) (vgl. Machunsky, 2008).

Anzunehmen ist, dass nicht nur die Schülerherkunft maßgeblich dafür verantwortlich ist, ob Informationen eher kategoriebasiert oder individuumsbasiert verarbeitet werden.

Bedingungen für kategoriebasierte bzw. individuumsbasierte Informationsverarbeitungsprozesse von Lehrkräften

Für eine individuumsbasierte Informationsverarbeitung setzen Fiske und Neuberg (1990) *zeitliche und kognitive Ressourcen* voraus.⁵⁹ Angenommen wird, dass eine eher automatisierte bzw. kategoriegeleitete Informationsverarbeitung geringere kognitive Ressourcen beansprucht und aus diesem Grund v. a. bei geringen zur Verfügungen stehenden zeitlichen oder kognitiven Kapazitäten zum Einsatz kommt (vgl. z.B. Bless et al., 2004; Klauer, 2008; Oakes, 1996). Rothbart, Fulero, Jensen, Howard und Birrell (1978) konnten empirische Evidenz hierfür erbringen und berichten, dass unter hoher kognitiver Belastung eher automatisiert bzw. kategoriebasiert verarbeitet wird. Olson und Fazio (2009) gehen ebenfalls davon aus, dass der Einfluss impliziter Einstellungen auf die Urteilsbildung größer ist, wenn zeitliche und kognitive Kapazitäten begrenzt sind.

⁵⁹ Bodenhausen (1990) konnte in zwei experimentellen Untersuchungen je nach Chronotyp der Versuchspersonen tagesrhythmische Unterschiede in der Stereotypaktivierung nachweisen. Fand die Untersuchung zu einer entgegen der dem individuellen Chronotyp entsprechenden Tageszeit statt, unterlagen die Urteile der Proband(inn)en häufiger Fehler im Sinne der Repräsentativitätsheuristik. Dies bestätigt sich auch in einer zweiten Untersuchung von Bodenhausen (1990), in der auch ethnische Stereotypen gegenüber Personen mit mittel-/südamerikanischen und afroamerikanischen Wurzeln berücksichtigt wurden. Dies bestätigt die Annahme, dass bei geringen kognitiven Kapazitäten eher automatisierte Verarbeitungsstrategien angewendet werden (Bodenhausen, 1990).

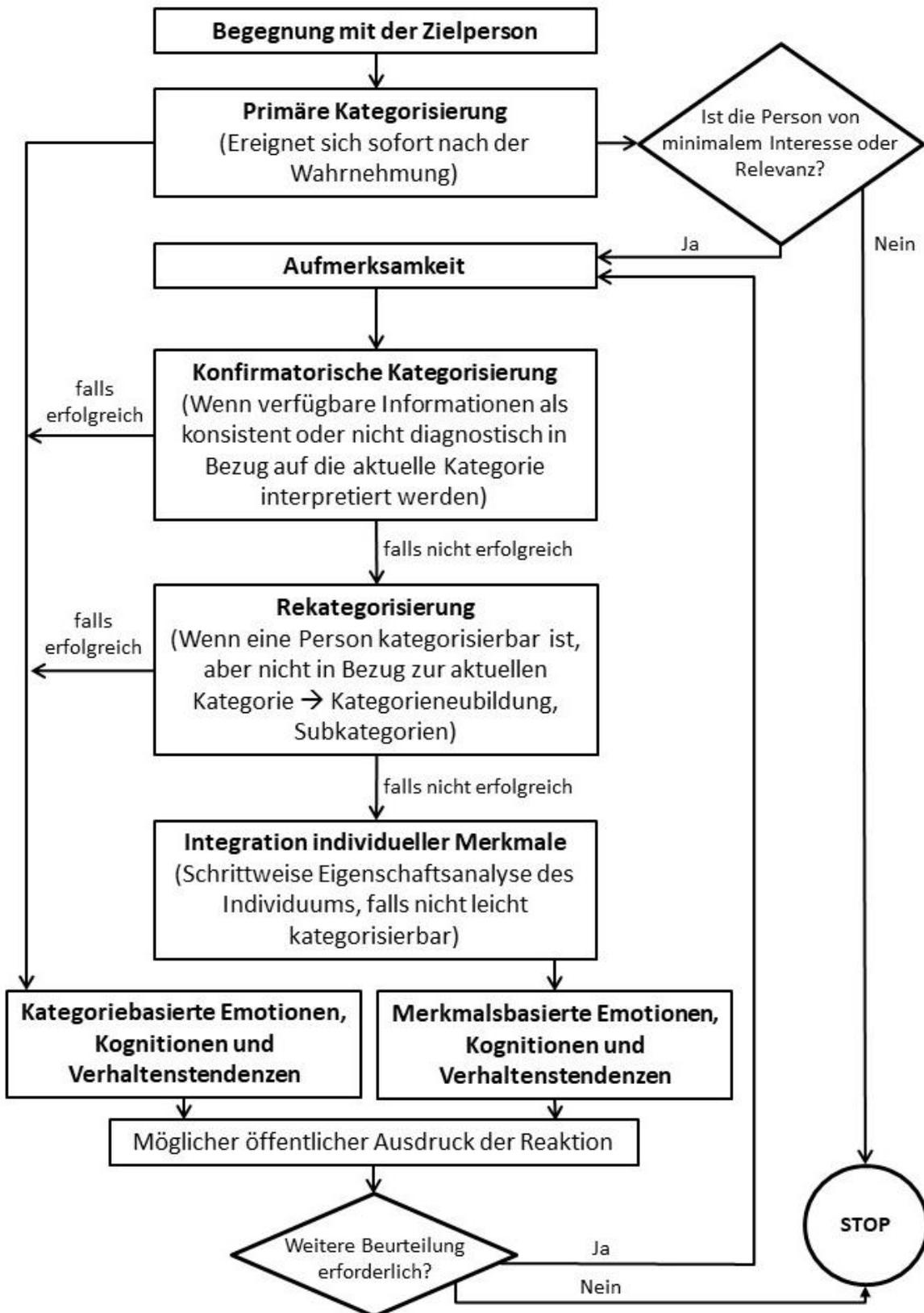


Abbildung 9. Kontinuum-Modell der Eindrucksbildung (nach Fiske & Neuberg, 1990, S. 5; Fiske, Lin & Neuberg, 1999, S. 233)

Lehrkräfte berichten häufig über Zeitmangel (Hüfner 2003; Schaarschmidt & Kieschke, 2007), wobei Zeitdruck oftmals mit kognitiver Belastung einhergeht und die Anwendung kategoriebasierter Informationsverarbeitung somit begünstigt (z.B. Dijker & Koomen, 1996; Kruglanski & Freund, 1983; Macrae & Bodenhausen, 2000). Zudem kann angenommen werden, dass auch Erfahrungen und Routinen dazu führen können, dass zunächst eher kontrollierte Prozesse zu eher automatisierten Prozessen werden. Für den schulischen Kontext ist dies z.B. bei der Beurteilung von Klassenarbeiten denkbar. Während Lehramtsanwärter hierfür womöglich noch sehr viel Zeit und kognitive Ressourcen benötigen, verläuft dies bei Lehrkräften durch langjährige Erfahrungen mit dem Unterrichtsstoff, dem verinnerlichten Erwartungshorizont und Korrekturroutinen womöglich eher automatisiert.

Neben Fiske und Neuberg (1990) bestätigen auch andere Forscher(innen), dass neben zeitlichen und kognitiven Ressourcen auch *Motivation zur vertieften Verarbeitung* vorhanden sein muss (z.B. Bodenhausen, 1990; Brewer, 1996; Devine, Plant, Amodio, Harmon-Jones & Vance, 2002; Schaal, 1997). Pendry und Macrae (1996) konnten zeigen, dass Proband(innen), die ein bedeutsames Ziel verfolgen, Personen differenzierter beurteilen. Empirische Untersuchungen zu kategoriebasierten und individuumsbasierten Informationsverarbeitungsprozessen von Lehrkräften verweisen darauf, dass auch hier die *Zielsetzung* der Beurteilung (Eindrucksbildung vs. Prognose) einen Einfluss hat. Sollen sich Lehrkräfte einen Eindruck von Schüler(inne)n bilden, greifen sie eher auf soziale Kategorien zurück (Dünnebier et al., 2009; Krolak-Schwerdt, Böhmer & Gräsel, 2009, 2012; Glock, Kneer & Krolak-Schwerdt, 2011). Unter der Vorgabe, eine Prognose zu erstellen, gingen sie hingegen eher individuumsbasiert vor (Dünnebier et al., 2009; Krolak-Schwerdt et al., 2009, 2012). Dabei ist anzunehmen, dass die Motivation das eigene Ziel zu erreichen, dazu führt, dass mehr individuumsbasierte Informationen aufgenommen und verarbeitet werden. Somit ist davon auszugehen, dass Lehrkräfte bei wichtigen Urteilen, wie z.B. verbindlichen Entscheidungen über den weiteren Bildungsvorlauf von Schüler(inne)n, bemüht sind, diese auf Basis individueller Informationen zu treffen. Jedoch konnte in diesem Abschnitt bereits verdeutlicht werden, dass der erste Eindruck weitere Urteile beeinflusst und relativ stabil ist. Somit sind auch vermeintlich unwichtige Urteile von hoher Bedeutung und unterliegen der Gefahr einer kategoriebasierten Informationsverarbeitung.

In Bezug auf die Expertise von Personen, deuten empirische Befunde darauf hin, dass Expert(inn)en und Laien sich in den Informationsverarbeitungsprozessen unterscheiden (z.B. Baron, 1988; Chi, Feltovich & Glaser, 1981). Böhmer, Gräsel, Krolak-Schwerdt, Hörstermann & Glock (2017) greifen die theoretischen Grundlagen des Kontinuum-Modells der Eindrucksbildung (Fiske & Neuberg, 1990) in ihrem Modell der adaptiven diagnostischen Kompetenz (*adaptive diagnostic competency model*; ADCM) auf und berücksichtigen dabei unterschiedliche Informationsverarbeitungsprozesse von Expert(inne)nen und Novizen (vgl. Abbildung 10). Dabei gehen Böhmer et al. (2017) davon aus, dass Novizen alle ihnen zur Verfügung stehenden Informationen in den Urteilsbildungsprozess integrieren und hier nicht zwischen relevanten und irrelevanten Informationen differenzieren. Expert(inne)nen hingegen können je nach situationaler Anforderung zwischen einer eher heuristischen (kategoriebasierten) und einer regelbasierten (individuumsbasierten) Informationsverarbeitung wechseln. Durch empirische Studien von Böhmer (2011) und Dünnebier et al. (2009) konnten diese Annahmen bestätigt

werden. Auch van Ophuysen (2006) konnte in einer Untersuchung zur Vergabe von Schullaufbahneempfehlungen zeigen, dass erfahrene Grundschullehrkräfte im Vergleich zu Lehramtsstudierenden die vorliegenden Schülerinformationen differenzierter betrachten und auch weniger anfällig für einen *confirmation bias* sind, d. h. bei der Hinzunahme inkonsistenter Informationen waren sie eher bereit, ihr erstes Urteil zu ändern.^{60, 61}

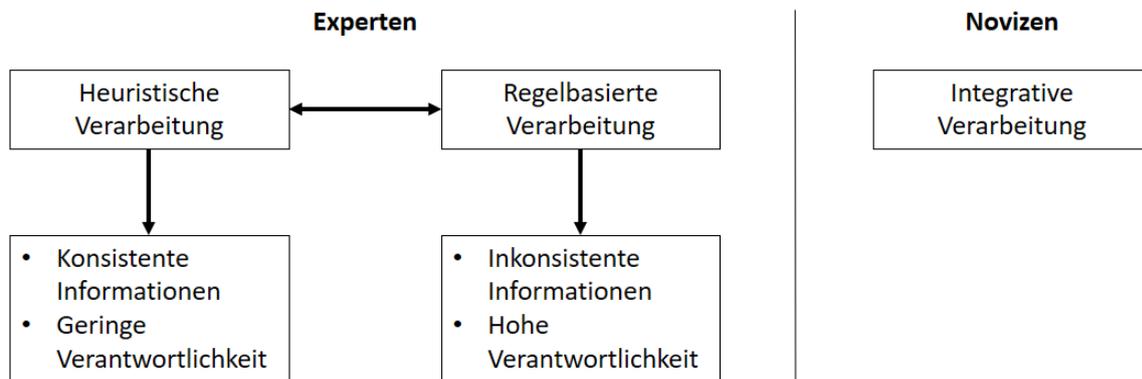


Abbildung 10. Modell der adaptiven diagnostischen Kompetenz (nach Böhmer et al., 2017)

Das ADCM greift dabei neben der Expertise und der Informationskonsistenz auch das Gefühl der Verantwortung für die Beurteilungsaufgabe auf (vgl. Abbildung 10). Krolak-Schwerdt, Böhmer und Gräsel (2013) konnten zeigen, dass Lehrkräfte unter der Bedingung geringer Verantwortlichkeit, Informationen eher kategoriengeleitet verarbeiteten, unter der experimentellen Bedingung hoher Verantwortlichkeit hingegen eher individuumsbasiert. Für luxemburgische Lehrkräfte berichten Glock et al. (2012), dass das Ausmaß an Verantwortung, das sie für eine Entscheidung (in diesem Fall eine Übergangsempfehlung) tragen, zu einer Veränderung der Urteilsgenauigkeit führt. Dabei zeigten sich bei einer höheren Verantwortlichkeit eine höhere Genauigkeit und weniger Urteilsfehler. Zudem konnten Glock et al. (2012) zeigen, dass auch die ethnische Herkunft von Schüler(inne)n bei der Urteilsbildung unter einer geringen Verantwortung eine Rolle spielt, während unter der Bedingung einer hohen Verantwortlichkeit kein Einfluss der Herkunft aufzufinden ist. Stattdessen werden in dieser Bedingung das Lern- und Arbeitsverhalten der Schüler(innen) berücksichtigt. Diese Befunde bestätigen die Annahme von Böhmer et al. (2017), dass bei Expert(inn)en, die eine hohe Verantwortung für ein Urteil haben, eine individuumsbasiere Informationsverarbeitung abläuft. Dennoch lassen sich herkunftsassoziierte Disparitäten in Lehrkrafturteilen (auch nach Kontrolle der tatsächlichen Leistungen) auch für wichtige Entscheidungen wie die Übergangsempfehlung auffinden (vgl. z. B. Sprietsma, 2013; vgl. Abschnitt 2.2).

⁶⁰ Der *confirmation bias* beschreibt die Tendenz von Personen, an ihren eigenen Überzeugungen und Hypothesen festzuhalten und entsprechend Informationen zu berücksichtigen, die die eigenen Annahmen bestätigen (vgl. z. B. Klayman, 1995).

⁶¹ Im Vergleich zu erfahrenden Lehrkräften scheinen Lehramtsstudierende auch über einen größeren „pädagogische[n] Optimismus“ (van Ophuysen, 2006, S. 159) zu verfügen, da sie häufiger eine Gymnasialempfehlung geben würden.

Neuberg und Fiske (1987) verweisen darauf, dass auch die Konsistenz des dargebotenen Stimulusmaterials bedeutsam für den Informationsverarbeitungsprozess ist, da diese u. a. die Aufmerksamkeit einer Person beeinflussen kann. Fiske, Kinder und Larter (1983) berichten, dass Experten mehr inkonsistente Informationen berücksichtigen, während diese von Novizen eher ausgeblendet werden. Angenommen wird, dass eine erhöhte Aufmerksamkeit – die durch inkonsistente Informationen entstehen kann – eine Voraussetzung für die vertiefte Informationsverarbeitung darstellt (Neuberg & Fiske, 1987). Allen, Sherman, Conrey und Stroessner (2009) konnten in zwei experimentellen Untersuchungen zeigen, dass die kognitiven Ressourcen einen bedeutsamen Einfluss auf den Aufmerksamkeitsfokus (stereotypkonsistente vs. inkonsistente Informationen) im Zusammenhang mit der Stärke impliziter Stereotype haben. Bei Personen, die ihre vollen kognitiven Kapazitäten nutzen, zeigte sich ein Zusammenhang zwischen der Stärke der Stereotype und zunehmender Aufmerksamkeit gegenüber stereotypkonsistenten Informationen. Dagegen stand das Maß impliziter Stereotype, unter der Bedingung begrenzter kognitiver Ressourcen, in Zusammenhang mit einer stärkeren Aufmerksamkeit gegenüber stereotypinkonsistenten Informationen. Böhmer et al. (2015) konnten in einer experimentellen Untersuchung mit erfahrenen Grundschullehrkräften zur Informationssuche bei Übergangsempfehlungen zeigen, dass bei inkonsistenten Schülerinformationen signifikant mehr weitere Informationen gesucht werden, was ein Indikator für eine eher kontrollierte Informationsverarbeitung sein könnte. Bodenhausen und Lichtenstein (1987) nehmen jedoch an, dass der Einfluss von Stereotypen bei inkonsistenten Informationen zunimmt, wenn nicht ausreichend eindeutige Informationen zur Urteilsbildung vorliegen.

Glock und Krolak-Schwerdt (2013) berichten im Rahmen einer experimentellen Untersuchung mit Lehramtsstudierenden, dass auch die Konsistenz der Informationen einen bedeutsamen Einfluss auf die Informationsverarbeitung bei der Beurteilung von Schülern mit türkischem Migrationshintergrund hat. Lag eine stereotypinkonsistente Fallbeschreibung (türkischer Schüler mit sehr guten Leistungen) vor, erfolgte keine herkunftsassoziierte Urteilsverzerrung. Im Gegensatz dazu zeigten sich Urteilsverzerrungen bei stereotypkonsistenten Fallvignetten (Glock & Krolak-Schwerdt, 2013; vgl. auch Glock et al., 2011). Glock und Krolak-Schwerdt (2013) verdeutlichen durch ihre Befunde, dass auch Novizen unter bestimmten Umständen – in diesem Fall durch Variation der Konsistenz von Informationen – Schülerinformationen unterschiedlich verarbeiten. Dies widerspricht dabei der Annahme von Böhmer et al. (2017) und deutet darauf hin, dass zumindest unter bestimmten Voraussetzungen auch bei Novizen unterschiedliche Informationsverarbeitungsprozesse aufzufinden sind. Wobei Glock et al. (2011) darauf hindeuten, dass Expert(inn)en zwar im Gegensatz zu Laien bei inkonsistenten Informationen eine heuristische Informationsverarbeitung anwenden können, aber Laien nicht zwischen automatisierten und kontrollierten Prozessen wechseln können. Dies könnte u. a. daran liegen, dass je nach Aufgabenstellung bestimmte Handlungsskripte noch nicht verinnerlicht sind und entsprechend noch nicht automatisch aktiviert werden können.

6. Zusammenfassung und Herleitung der Forschungsfragen

Basierend auf den Befunden zu herkunftsassoziierten Disparitäten im deutschen Bildungssystem (z. B. Arnold et al., 2007; Mehringer, 2013; OECD, 2016a, 2016b) und der Relevanz einer akkuraten und herkunftsunabhängigen Urteils- und Erwartungsbildung von Lehrkräften (z. B. Artelt & Gräsel, 2009; Bailey & Drummond, 2006; Teisl et al., 2001) wurden bereits einige empirische Untersuchungen durchgeführt, die das Ziel hatten den Zusammenhang zwischen der ethnischen und sozialen Schülerherkunft mit Lehrkrafturteilen und -erwartungen aufzuklären. Zudem existieren bereits erste Studien zu herkunftsassoziierten Lehrkrafteinstellungen und -stereotypen. Im folgenden Abschnitt werden daher die zentralen Befunde, die im Verlauf der vorliegenden Arbeit bereits ausführlich vorgestellt wurden, zusammengefasst. Zudem werden die bestehenden Forschungslücken aufgezeigt und darauf aufbauend im nachfolgenden Abschnitt die zentralen Forschungsfragen und Hypothesen der vorliegenden Arbeit dargestellt.

6.1 Zusammenfassung des Forschungsstands zu herkunftsassoziierten Lehrkrafturteilen und -erwartungen und Forschungsdefizite

Insbesondere aus dem Forschungsbereich zur diagnostischen Kompetenz konnten zahlreiche Studien bestätigen, dass große interindividuelle Unterschiede zwischen Lehrkräften in der Urteilsakkuratheit vorliegen (z. B. Schrader & Helmke, 1987; Stang & Urhahne, 2016b; Südkamp et al., 2012). Trotz der hohen Relevanz der diagnostischen Kompetenz (z. B. Artelt & Gräsel, 2009) verfügen somit nicht alle Lehrkräfte über ausreichende Fähigkeiten in der akkuraten Beurteilung von Schülerleistungen und weiteren leistungsrelevanten Merkmalen, wie z. B. motivationale Ausprägungen oder allgemeine schulische Fähigkeiten. Durchschnittlich kann meist nur von einer moderat ausgeprägten diagnostischen Kompetenz ausgegangen werden (z. B. McElvany et al., 2009; Südkamp et al., 2012), wobei den Lehrkräften v. a. die akkurate Einschätzung leistungsrelevanter Schülermerkmale Schwierigkeiten bereitet (z. B. Spinath, 2005; Urhahne et al., 2010). Die Ursachen interindividueller Differenzen in der Ausprägung der diagnostischen Kompetenz zwischen Lehrkräften konnten dabei bislang nur unzureichend identifiziert werden. Südkamp et al. (2012) verweisen in ihrem Modell der Akkuratheit von Lehrkrafturteilen auf unterschiedliche Einflussfaktoren, jedoch konnten bislang insbesondere für intraindividuelle Lehrkraftmerkmale keine ausreichenden empirischen Befunde erbracht werden. Dabei wäre zu vermuten, dass spezifische Einstellungen von Lehrkräften – die einen Einfluss auf die Informationsverarbeitung haben können (vgl. Kapitel 4 & Kapitel 5) – bei der Beurteilung von Schüler(inne)n relevant sind.

Des Weiteren verweisen einige empirische Befunde darauf, dass Lehrkrafturteile und -erwartungen durch leistungsirrelevante Informationen z. T. systematisch verzerrt sein können (z. B. Doherty & Conolly, 1985; Dunkake et al., 2012; Weiss, 1965). Dies wirft im Hinblick auf herkunftsassoziierte Disparitäten im deutschen Bildungssystem die Frage auf, ob auch ethnische und soziale Merkmale der Schüler(innen) zu verzerrten Lehrkrafturteilen und -erwartungen führen. Zum Einfluss der ethnischen und sozialen Schülerherkunft auf die Urteile

und Erwartungen von Lehrkräften und Lehramtsstudierenden ist jedoch eine eher heterogene Befundlage zu konstatieren. In einigen experimentellen Studien konnte festgestellt werden, dass Lehrkräfte und Lehramtsstudierende höhere Erwartungen an die Leistungen von Schüler(inne)n mit hohem sozioökonomischen Status haben als an die Leistungen von Schüler(inne)n mit niedrigem sozioökonomischem Status (z. B. Darley & Gross, 1983; Tom et al., 1984). Zudem zeigte sich, dass Lehrkräfte Schüler(innen) mit hohem sozioökonomischem Status eine höhere Intelligenz zuschreiben und ihr Verhalten positiver beurteilen (Alvidrez & Weinstein, 1999; Dunkake & Schuchart, 2015). Jedoch konnten nicht alle Untersuchungen Unterschiede in Lehrkrafturteilen und -erwartungen in Abhängigkeit der sozialen Herkunft feststellen (z. B. Kaiser et al., 2015; Karing et al., 2011).

Für ethnizitätsabhängige Urteils- und Erwartungsunterschiede zeigt sich ebenfalls kein einheitliches Bild. Teilweise konnten sowohl in experimentellen Studien als auch in Felduntersuchungen Unterschiede in Lehrkrafturteilen und -erwartungen ermittelt werden (z. B. Glock, 2016; Sprietsma, 2013; Tenenbaum & Ruck, 2007), die z. T. auch nach Kontrolle der tatsächlichen Leistungen bestehen blieben (z. B. Bonefeld et al., 2017). Jedoch konnten andere Forscher(innen) keine verzerrten Lehrkrafturteile in Abhängigkeit der Schülerethnizität feststellen (z. B. Chen & Weseley, 2011; Kaiser et al., 2016). Untersuchungen, die herkunftsassoziierte Unterschiede in Lehrkrafturteilen aufzeigen, liefern ein überwiegend einheitliches Bild der negativeren Urteile und Erwartungen für Schüler(innen) ethnischer Minoritäten und/oder mit niedrigem sozioökonomischem Status im Vergleich zu Schüler(inne)n der ethnischen Majorität oder hohem sozioökonomischem Status. Jedoch können diese Befunde bislang nicht klären, wie diese Disparitäten zustande kommen. Es bleibt unklar ob die unterschiedlichen Einschätzungen der Lehrkräfte daher kommen, dass sie Schüler(innen) mit Migrationshintergrund oder niedrigem sozioökonomischem Status unterschätzen, oder ob sie Schüler(innen) der ethnischen Majorität oder mit hohem sozioökonomischem Status überschätzen. Für US-amerikanische Vorschulkinder finden sich Hinweise für eine Überschätzung von Kindern mit hohem sozioökonomischem Status und eine Unterschätzung von Kindern mit niedrigem sozioökonomischem Status (Ready & Chu, 2015; Ready & Wright, 2011). Kaiser et al. (2016) berichten für Deutschland hingegen von einer eher akkuraten Einschätzung von Kindern mit Migrationshintergrund, vermuten jedoch, dass dies nicht per se durch die ethnische Herkunft, sondern die Salienz als Minorität zu erklären ist. Dabei bleibt eine Verschränkung der ethnischen und sozialen Herkunft in empirischen Studien hierzu jedoch bislang aus.

Erschwert wird die Interpretation zahlreicher Studien durch die hohe Konfundierung der ethnischen und sozialen Herkunft (vgl. z. B. Herwartz-Emden et al., 2010). Vermeintliche Effekte des Migrationshintergrunds könnten sich durch die enge Verknüpfung mit einem niedrigen sozioökonomischen Status erklären lassen. Die bislang wenigen empirischen Untersuchungen, die sowohl die ethnische als auch die soziale Herkunft der Schüler(innen) berücksichtigen (z. B. Hoenig & Wenz, 2013; Lorenz et al., 2016; McCombs & Gay, 1988) deuten dabei auf den Einfluss beider Herkunftsmerkmale hin. Jedoch liegen v. a. im deutschsprachigen Raum zu wenige empirische Studien vor, die beide Herkunftsmerkmale berücksichtigen.

Im Rahmen der Forschungen zur diagnostischen Kompetenz wurde zudem gefordert, den Fokus von der reinen Notengebung auf weitere Schülermerkmale zu erweitern (vgl. Artelt & Gräsel, 2009). Insbesondere unter Berücksichtigung der Schülerherkunft blieb dies bislang

jedoch weitestgehend aus. Der überwiegende Anteil empirischer Studien fokussierte lediglich herkunftsassoziierte Leistungsbeurteilungen und -erwartungen (z.B. Klapproth et al., 2012), zum Teil wurde jedoch bereits die Übergangsempfehlung für eine weiterführende Schule der Sekundarstufe I berücksichtigt (z.B. Glock et al., 2012; Hoenig & Wenz, 2013; Sprietsma, 2013). Da jedoch zahlreiche weitere Schülermerkmale bedeutsamen Einfluss auf den Lernerfolg haben können und entsprechend akkurat durch die Lehrkräfte eingeschätzt werden müssen, bedarf es hierzu empirischen Forschungsarbeiten die klären, ob sich herkunftsassoziierte Lehrkrafturteile und -erwartungen auch bei weiteren relevanten Faktoren zeigen. Dabei ist die akkurate Einschätzung leistungsrelevanter Schülermerkmale v. a. in der Primarstufe von hoher Bedeutung, da Lehrkräfte auch diese Informationen bei der Übergangsempfehlung berücksichtigen (Böhmer et al., 2012).

Kratzmann (2013) erfasste in den Analysen zu herkunftsassoziierten Urteilsunterschieden zwar die Einschätzung der Fähigkeiten der Kinder, jedoch ist diese Studie durch die Untersuchung von Vorschulkindern nicht direkt auf den Schulkontext übertragbar. Für Lehrkräfte existieren hierzu bislang nur Befunde aus dem US-amerikanischen Raum (Downey & Pribesh, 2004), wobei durch die unterschiedlichen Migrantengruppen und gesellschaftsspezifische Situationen keine direkte Übertragung auf Lehrkräfte in Deutschland möglich ist. Dunkake und Schuchart (2015) konnten für Deutschland zwar zeigen, dass Lehramtsstudierende negativere Verhaltensweisen von Schüler(inne)n mit niedrigem sozioökonomischen Status erwarten, jedoch müssten diese Studien mit erfahrenen Lehrkräften repliziert werden.

Im Vergleich zu zumindest einigen Befunden zu expliziten herkunftsassoziierten Urteilen und Erwartungen, steht die Erforschung der zugrundeliegenden kognitiven Prozesse noch am Anfang. In Anlehnung an das Kontinuum-Modell der Eindrucksbildung von Fiske und Neuberg (1990) verweisen einige Befunde darauf, dass Faktoren wie z.B. die Verantwortung für das Urteil die Expertise beeinflussen, ob die Schülerherkunft einen bedeutsamen Einfluss auf die Urteilsbildung hat (z.B. Böhmer et al., 2017; Glock et al., 2011; Glock et al., 2012). Dabei wird vermutet, dass unter z.B. geringer Verantwortung eine eher automatisierte Informationsverarbeitung abläuft, die auf Stereotypen fußt. Dennoch haben Untersuchungen zu herkunftsassoziierten Lehrkrafturteilen und -erwartungen den zugrundeliegenden Informationsverarbeitungsprozess bislang nur unzureichend analysiert. Meist erfolgte lediglich der Rückschluss auf automatisierte oder kontrollierte Informationsverarbeitungsprozesse durch die entsprechenden Verzerrungen in den tatsächlichen Urteilen und Erwartungen.

Dabei lässt sich vermuten, dass herkunftsassoziierte Einstellungen der Lehrkräfte und Lehramtsstudierenden einen bedeutsamen Einfluss auf den Prozess der Informationsverarbeitung haben. Vergleichbar mit Einstellungen der allgemeinen Bevölkerung zu ethnischen und sozialen Gruppen in Deutschland (z.B. Asbrock, 2010; Kahraman & Knoblich, 2000; Sielschott, 2010) weisen auch Lehrkräfte und Lehramtsstudierende spezifische Einstellungen gegenüber ethnischen und sozialen Gruppen auf. Insgesamt zeigt sich zwar eine eher positive explizite Einstellung von Lehrkräften zu einer heterogenen Schülerschaft (Gebauer & McElvany, 2017), jedoch erbrachten implizite Messverfahren eine Präferenz für Schüler(innen) ohne Migrationshintergrund sowie für Schüler(innen) mit einem hohen sozioökonomischen Status (Glock & Karbach, 2015; Glock et al., 2016). Inwieweit unterschiedliche Einstellungen jedoch die Informationsverarbeitung beeinflussen, wurde bisher nicht berücksichtigt.

6.2 Forschungsfragen und Hypothesen der vorliegenden Arbeit

Der Forschungsstand zu herkunftsassoziierten Lehrkrafturteilen und -erwartungen weist bereits auf einige Forschungslücken. Offen blieb bisher, ob sich herkunftsassoziierte Unterschiede durch den Migrationshintergrund und/oder die soziale Herkunft von Schüler(inne)n erklären lassen. Aus diesem Grund hat die vorliegende Arbeit das Ziel, beide Herkunftsmerkmale in den Untersuchungen zu herkunftsassoziierten Lehrkräfteeinschätzungen zu berücksichtigen. Bisherige Studien konnten entsprechend bislang nicht klären, ob sich herkunftsassoziierte Urteils- und Erwartungsunterschieden von Lehrkräften sowohl in Bezug auf die ethnische als auch soziale Herkunft von Schüler(inne)n zeigen. Zudem existieren kaum Befunde, die neben den Urteilsunterschieden für Leistungen auch weitere relevante Schülermerkmale, wie z.B. motivationale Ausprägungen oder allgemeine schulische Fähigkeiten berücksichtigten. Entsprechend geht die vorliegende Arbeit der Frage nach, ob sich herkunftsassoziierte Unterschiede in Urteilen und Erwartungen von Lehrkräften für Schülerleistungen und leistungsrelevante Merkmale zeigen (Fragestellung 1.1). Basierend auf den Ergebnissen bisheriger empirischer Studien und theoretischer Annahmen wird in dieser Arbeit angenommen, dass negativere Urteile über und Erwartungen an Schüler(innen) mit Migrationshintergrund im Vergleich zu Schüler(inne)n ohne Migrationshintergrund sowie negativere Urteile über und Erwartungen an Schüler(innen) mit niedrigem sozioökonomischem Status im Vergleich zu Schüler(inne)n mit hohem sozioökonomischen Status vorliegen.

Eine grundlegende Frage in diesem Zusammenhang – wie diese Urteilsunterschiede zustande kommen – konnte bislang nicht beantwortet werden. So ist nicht klar, ob es sich bei herkunftsabhängigen Differenzen in Lehrkrafturteilen und -erwartungen tatsächlich um eine Benachteiligung von Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status handelt, oder ob eher von einer Bevorzugung von Schüler(inne)n ohne Migrationshintergrund und mit hohem sozioökonomischem Status auszugehen ist (Fragestellung 1.2). Dabei geht die vorliegende Arbeit zum einen davon aus, dass Lehrkraft-erwartungen und -aspirationen für Schüler(innen) mit Migrationshintergrund und/oder mit niedrigem sozioökonomischem Status negativ verzerrt sind und zum anderen Lehrkraft-erwartungen und -aspirationen für Schüler(innen) ohne Migrationshintergrund und mit hohem sozioökonomischem Status positiv verzerrt sind. Diese Annahmen resultieren dabei aus den vorliegenden Befunden zu eher positiven Einstellungen und Stereotypen von Lehrkräften gegenüber Schüler(inne)n ohne Migrationshintergrund und/oder mit hohem sozioökonomischem Status und entsprechend eher negativen Einstellungen und Stereotypen gegenüber Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status. Zudem kann in Anlehnung an das Kontinuum-Modell der Eindrucksbildung (Fiske & Neuberg, 1990) angenommen werden, dass Einstellungen und Stereotype unter bestimmten Bedingungen die Informationsverarbeitung und damit auch die Urteilsbildung beeinflussen.

Jedoch zeigt der bisherige Forschungsstand v. a. noch große Lücken bei der Analyse der zugrundeliegenden kognitiven Mechanismen auf. Entsprechend verfolgt die vorliegende Arbeit das Ziel zu klären ob sich die Informationsverarbeitung bei der Beurteilung von Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status im Vergleich zu Schüler(inne)n ohne Migrationshintergrund und hohem sozioökonomischem Status unter-

scheidet (Fragestellung 2.1). Angenommen wird, dass sich, basierend auf ersten empirischen Hinweisen für eine Überschätzung von Kindern mit hohem sozioökonomischem Status, bei Schüler(inne)n ohne Migrationshintergrund und mit hohem sozioökonomischem Status eine eher automatisierte Informationsverarbeitung feststellen lässt. Zudem ist basierend auf den Ergebnissen einer akkurateren Einschätzung von Schüler(inne)n mit Migrationshintergrund davon auszugehen, dass sich bei dieser Schülergruppe eine eher kontrollierte Informationsverarbeitung zeigt. Zudem wird vermutet, dass die kognitiven Prozesse bei der Beurteilung von Schüler(inne)n unterschiedlicher Herkunft durch herkunftsassoziierte Einstellungen moderiert werden (Fragestellung 2.2).

Um diesen Fragen nachzugehen, wurden zwei experimentelle Untersuchungen mit Lehrkräften und Lehramtsstudierenden durchgeführt. Dabei wurde auf die Beurteilung von Kindern im Grundschulalter fokussiert, da v. a. zum Ende der Grundschulzeit die Weichen für den weiteren Bildungsweg und damit häufig auch für den Bildungserfolg gestellt werden und sich hier bereits zahlreiche herkunftsassoziierte Disparitäten auffinden lassen (vgl. Kapitel 2).

7. Überblick über das empirische Vorgehen in Studie 1 und Studie 2

Die vorliegende Arbeit verfolgt das Ziel, durch zwei experimentelle Untersuchungen einen Beitrag zur Schließung bestehender Forschungslücken im Kontext herkunftsassoziierter Lehrkrafturteile und -erwartungen zu liefern. Dabei hat Studie 1 das zentrale Ziel, bisherige Befunde zu herkunftsassozierten Urteils- und Erwartungsunterschieden von Lehrkräften zu replizieren und herkunftsassozierte Unterschiede in der Einschätzung leistungsrelevanter Schülermerkmale zu untersuchen. Des Weiteren soll eine mögliche Benachteiligung durch negative Urteilsverzerrungen oder Bevorzugung durch positive Urteilsverzerrungen bestimmter ethnischer und sozialer Schülergruppen aufgedeckt werden. Studie 2 fokussiert die Analyse der zugrundeliegenden kognitiven Prozesse, die herkunftsassozierte Urteile und Erwartungen erklären könnten. In dieser Untersuchung wird analysiert ob eine spezifische Ethnizität oder der soziale Status von Schüler(inne)n zu einer eher automatisierten oder eher kontrollierten Informationsverarbeitung führt. Zudem werden herkunftsassozierte Einstellungen in den Analysen berücksichtigt.

Bedingt durch methodische Überschneidungen werden in diesem Kapitel die zentralen abhängigen und unabhängigen Variablen beider Studien sowie die Auswahl des für beide Studien überwiegend identischen Stimulusmaterials beschrieben. Eingesetzt wurden Fallvignetten, die fiktive männliche Grundschüler der vierten Jahrgangsstufe beschreiben (vgl. Abschnitt 7.1.1).⁶² In beiden Studien wurde durch Vornamen die ethnische und soziale Schülerherkunft manipuliert (vgl. Abschnitt 7.1.2). Die Versuchspersonen wurden den experimentellen Gruppen in beiden Studien randomisiert zugewiesen. Durch die Konstanthaltung der weiteren Informationen ermöglicht dieses Vorgehen eine kausale Zuordnung der Effekte auf die unabhängige Variable und erhöhte die interne Validität der Untersuchungen. Auf Grund von standardisierten Erhebungsinstrumenten, standardisierten Erhebungsdurchführungen, Instruktionen sowie Auswertungen kann ebenfalls von einer hohen Durchführungs- und Auswertungsobjektivität der Untersuchungen ausgegangen werden. Tabelle 2 enthält eine Übersicht zu den wesentlichen Merkmalen beider Studien, die in den beiden folgenden Kapiteln dieser Arbeit vertieft beschrieben werden.

⁶² Da in den empirischen Untersuchungen der vorliegenden Arbeit ausschließlich das männliche Geschlecht von Schülern berücksichtigt wurde, wird in den folgenden Abschnitten und Kapiteln zu den beiden Studien dieser Arbeit bewusst nur die männliche Form *Schüler* verwendet (zur Begründung der Fokussierung auf das männliche Geschlecht, siehe Abschnitt 7.1).

Tabelle 2

Übersicht der Fragestellungen, Hypothesen, Studiendesigns, Erhebungs- und Analysemethoden sowie Versuchspersonen und Analysetechnik beider empirischer Studien der vorliegenden Arbeit

	Studie 1	Studie 2
Zentrale Fragestellungen	<p>1.1 Zeigen sich herkunftsassoziierte Unterschiede in den Erwartungen und Urteilen von Lehrkräften für Schülerleistungen und leistungsrelevante Merkmale?</p> <p>1.2 Kann von einer Benachteiligung von Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status und/oder einer Bevorzugung von Schüler(inne)n ohne Migrationshintergrund und hohem sozioökonomischem Status ausgegangen werden?</p>	<p>2.1 Unterscheidet sich die Tiefe der Informationsverarbeitung bei der Urteils- und Erwartungsbildung von Lehramtsstudierenden bei Schüler(inne)n unterschiedlicher ethnischer und sozialer Herkunft?</p> <p>2.2 Moderieren implizite Einstellungen die Tiefe der Informationsverarbeitung von Lehramtsstudierenden bei der Beurteilung von Schüler(inne)n unterschiedlicher ethnischer und sozialer Herkunft?</p>
Hypothesen	<p>H_{1.1}: Lehrkrafturteile über und -erwartungen an Leistungen sowie leistungsrelevante Merkmale fallen für Schüler mit niedrigem sozioökonomischem Status schlechter aus als für Schüler mit hohem sozioökonomischem Status.</p>	<p>H_{2.1}: Bei stereotypinkonsistenten Informationen zeigt sich eine höhere visuelle Aufmerksamkeit (Fixationsdauer) als bei stereotypkonsistenten Informationen.</p>

Tabelle 2 (Fortsetzung)

Übersicht der Fragestellungen, Hypothesen, Studiendesigns, Erhebungs- und Analysemethoden sowie Versuchspersonen und Analysestichprobe beider empirischer Studien der vorliegenden Arbeit

	Studie 1	Studie 2
Hypothesen	<p>H_{1.2}: Lehrkrafterteile über und -erwartungen an Leistungen so- wie leistungsrelevante Merkmale fallen für Schüler mit Migrationshintergrund schlechter aus als für Schüler ohne Migrationshintergrund.</p> <p>H_{1.3}: Lehrkrafterwartungen und -aspirationen sind für Schüler mit Migrationshintergrund und/oder mit niedrigem sozio- ökonomischem Status negativ verzerrt.</p> <p>H_{1.4}: Lehrkrafterwartungen und -aspirationen sind für Schüler ohne Migrationshintergrund und mit hohem sozioökono- mischem Status positiv verzerrt.</p>	<p>H_{2.2}: Beim Lesen der Fallvignette eines Schülers mit Migrati- onshintergrund und/oder mit niedrigem sozioökonomi- schem Status zeigt sich eine höhere kognitive Beanspru- chung (Pupillendilatation und erhöhte Lidschlagrate) als beim Lesen der Fallvignette eines Schülers ohne Migrati- onshintergrund und hohem sozioökonomischem Status.</p> <p>H_{2.3}: Personen mit hohen impliziten Einstellungen gegenüber Personen ohne Migrationshintergrund verarbeiten Infor- mationen von Schülern ohne Migrationshintergrund auto- matisierter als Personen mit weniger positiven impliziten Stereotypen.</p>
Studiendesign	<p>Experimentelle Online-Studie im between-subject-Design</p> <p>Experimentelle Manipulation der ethnischen und sozialen Schülerherkunft in drei Fallvignetten</p>	<p>Experimentelle Labor-Studie im within-subject-Design</p> <p>Experimentelle Manipulation der ethnischen und sozialen Schülerherkunft in neun Fallvignetten</p>

Tabelle 2 (Fortsetzung)
 Übersicht der Fragestellungen, Hypothesen, Studiendesigns, Erhebungs- und Analysemethoden sowie Versuchspersonen und Analyses Stichprobe beider empirischer Studien der vorliegenden Arbeit

	Studie 1	Studie 2
Erhebungsmethoden	<ul style="list-style-type: none"> - Fragebogenerhebung 	<ul style="list-style-type: none"> - Fragebogenerhebung - Erfassung der Okulomotorik durch Eye-Tracking - Impliziter Assoziationstest (IAT)
Unabhängige Variable	Ethnische und soziale Schülerherkunft	Ethnische und soziale Schülerherkunft
Abhängige Variablen	<ul style="list-style-type: none"> - Leistungserwartung und -aspiration - Einschätzung der Gymnasialeignung - Einschätzung der Anstrengungsbereitschaft - Einschätzung der Fähigkeiten 	<ul style="list-style-type: none"> - Fixationsdauer - Pupillendiameter - Lidschlagrate
Zentrale Analysemethoden	<ul style="list-style-type: none"> - Multi- und univariate einfaktorielle Varianzanalysen - Orthogonale a priori Helmert-Kontrastanalysen 	<ul style="list-style-type: none"> - Einfaktorielle Varianzanalysen mit Messwiederholung und drei Faktorstufen - Orthogonale a priori Helmert-Kontrastanalysen - Kovarianzanalysen mit Messwiederholung
Versuchspersonen	Erfahrene Grundschullehrkräfte	Lehramtsstudierende aller Schularten (ab 3. Fachsemester)
Analyses Stichprobe	N = 237	N = 45

7.1 Stimulusmaterial in Studie 1 und Studie 2

Da die durchgeführten Untersuchungen experimenteller Art sind, hatte die Konzeption des Stimulusmaterials besondere Bedeutung. Fallvignetten über Grundschüler der vierten Jahrgangsstufe bildeten in beiden Studien die Basis zur Aktivierung der sozialen Informationsverarbeitung und sollten daher ausreichende Beschreibungen enthalten, um möglichst realitätsnahe kognitive Repräsentationen der Lernenden zu erzeugen. Dabei konnten Krolak-Schwerdt, Hörstermann, Glock und Böhmer (2017) durch zwei Studien mit erfahrenen Lehrkräften zeigen, dass Fallvignetten ökologisch valide Instrumente darstellen um Lehrkrafturteile zu erfassen. Zudem war die Voraussetzung für eine erfolgreiche Manipulation der ethnischen und sozialen Herkunft, in den Fallvignetten Vornamen zu wählen, die von den Versuchspersonen möglichst automatisiert mit der entsprechenden Herkunft assoziiert werden. In den folgenden beiden Abschnitten werden daher die Kriterien zur Erstellung der Fallvignetten und Namensauswahl näher beschrieben.

7.1.1 Konstruktion der Fallvignetten

Die Auswahl von Verbal- und Ziffernbewertungen in Form von schulischen Zeugnissen als Stimulusmaterial erfolgte aus drei wesentlichen Gründen:

- (1) Mit dieser Form der Personenbeschreibung sind Lehrkräfte in ihrem Berufsalltag regelmäßig konfrontiert. Somit zählen das Lesen und Schreiben von Zeugnissen zu ihren gewohnten Aufgaben. Hierdurch können sich Lehrkräfte leichter in den Prozess der Urteilsbildung einfügen, ohne weitere Ressourcen wegen zu komplexen oder unbekanntem Stimulus-Formaten zu verlieren.
- (2) Die Entscheidung für die Verbalbewertung eines Zeugnisses zur Beschreibung leistungsrelevanter Schülermerkmale wurde gewählt, da diese im Gegensatz zu einzelnen getrennten Textfragmenten (z. B. in Form einer Auflistung von Personenmerkmalen) durch die sprachliche Struktur und syntaktische Verknüpfungen eher den vertrauten und gängigen Personenbeschreibungen im Alltag ähnelt (vgl. Krolak-Schwerdt & Wintermantel, 2004).
- (3) Zudem kann, auf Grund der üblicherweise häufigen Nennung des Schülernamens in Zeugnissen, die ethnische und soziale Herkunft manipuliert werden, ohne explizite Informationen hierüber nennen zu müssen. Dies erscheint insbesondere durch die Gefahr einer zu offensichtlichen Präsentation der Schülerherkunft von Bedeutung, da dies möglicherweise bereits einen Effekt auf die Informationsverarbeitung haben könnte. Um dies zu vermeiden, müsste eine umfangreichere Schülerbeschreibung erstellt werden, in der neben den Herkunftsinformationen weitere irrelevante Informationen zur Textfüllung eingefügt werden. Da dies jedoch insbesondere wegen des within-subject-Designs in Studie 2 und die hierfür notwendige Nutzung mehrerer Fallvignetten die Konstanzhaltung aller weiteren Informationen gefährden könnte, wurde die Aktivierung der Schülerherkunft ausschließlich durch den Vornamen vorgenommen und entsprechend auf zusätzliche Schülerbeschreibungen verzichtet.

Voraussetzung für differentielle Urteilsbildungen ist eine mögliche Zuordnung von Informationen zu unterschiedlichen kognitiven Schemata (Neisser, 1976). Aus diesem Grund sollte die Auswahl der Textbausteine inhaltliche Ambiguität erzeugen, da die Möglichkeit, die Informationen auf unterschiedliche Weise wahrzunehmen und zu interpretieren, gegeben sein sollte. Dabei sollten die Zeugnisse stereotypenkonsistente und -inkonsistente Informationen für alle drei Herkunftskombinationen enthalten. Hierdurch können bei den Versuchspersonen unterschiedliche Schemata aktiviert werden, die die Urteils- und Erwartungsbildung gegenüber den jeweiligen Schülern beeinflussen können.

Die Textbausteine wurden überwiegend aus den Formulierungshilfen für Grundschulzeugnisse der vierten Jahrgangsstufe des Staatsinstituts für Schulqualität und Bildungsforschung (2008) erstellt und um einzelne Textpassagen mit Textbausteinen zur Schülerbeurteilung (Birnstein, o. J.) ergänzt. Neben der Verbalbeurteilung enthielten die Grundschulzeugnisse ebenfalls eine Übersicht zu Schulnoten in allen Fächern, sowie eine Zusammenfassung der Durchschnittsnoten aller Fächer sowie der drei Hauptfächer. Für Studie 1 wurde eine Variante eines Halbjahreszeugnisses der vierten Jahrgangsstufe generiert. Durch die Manipulation der drei Herkunftskombinationen ergaben sich drei Fallvignetten. In Studie 2 wurde ebenfalls die identische Variante des Grundschulzeugnisses aus Studie 1 eingesetzt. Da in Studie 2 jede Versuchsperson alle Zeugnisse mit den entsprechenden drei Herkunftskombinationen bearbeiten sollte, mussten hierfür zwei weitere Zeugnisvarianten erstellt werden. Zeugnis 2 und Zeugnis 3 waren ebenfalls Halbjahreszeugnisse der vierten Jahrgangsstufe und enthielten ähnliche Informationen wie Zeugnis 1. Durch entsprechende Notenäquivalente der Textbausteine war es möglich, die drei Zeugnisvarianten so zu konzipieren, dass sie im Aufbau, den Informationen und der Verteilung positiver und negativer Aussagen zum Arbeits-, Lern- und Sozialverhalten sowie den schulischen Zensuren vergleichbar waren.

Dabei wurden zu Beginn der Verbalbewertung zunächst allgemeine Informationen zur Ordnung (z. B.: „Der Schüler gibt sich Mühe, mit seinen Schulbüchern und anderen Unterrichtsmaterialien gewissenhaft umzugehen.“) und zum Schulbesuch (z. B.: „Nie fehlt er unentschuldig und erscheint zum Unterricht jederzeit pünktlich.“) gegeben. Anschließend folgten positive und negative Informationen zum Sozialverhalten (positiv z. B.: „Der Schüler hilft Schwächeren stets unaufgefordert und eigenverantwortlich“; negativ z. B.: „Über vereinbarte Regeln setzt er sich oft hinweg.“), zum Arbeitsverhalten (positiv z. B.: „Der Schüler erledigte die Hausaufgaben zuverlässig.“; negativ z. B.: „Er ließ sich durch Störungen leicht von seiner Arbeit ablenken.“) und Lernverhalten (positiv z. B.: „Der Schüler war für neue Lerninhalte zu begeistern.“; negativ z. B.: „Durch Probleme ließ er sich leicht entmutigen“). Abschließend wurden Informationen zum Verhalten gegenüber der Lehrkraft geschildert (z. B.: „Er begegnet seinen Lehrern respektvoll und aufgeschlossen.“).

Im Anschluss an die Verbalbewertung wurde den Versuchspersonen die Benotung des Schülers für alle Fächer präsentiert sowie der gesamte Notendurchschnitt aller Fächer und der Durchschnitt in den Schulfächern Deutsch, Mathematik sowie Heimat- und Sachunterricht angegeben. Dabei wurde der Notendurchschnitt in den drei Hauptfächern für alle Fallvignetten so gewählt, dass er mit 2.33 die Grenze für Lehrkräftempfehlungen auf das Gymnasium markiert (§ 6, GrSO; Bayerische Staatskanzlei, 2008b). Anzumerken ist, dass diese Notengrenze – bedingt durch die föderale Struktur des Bildungssystems in Deutschland – nicht in allen

Bundesländern identisch ist (Überblick bei Hartinger, Graumann & Grittner, 2004), sondern spezifisch die bayerische Übertrittsregelung betrifft. Auch die Bedeutung der Lehrkrafturteile zum Übertritt variiert zwischen den Bundesländern. Auch wenn in den meisten Bundesländern der Elternwille von großer Bedeutung ist, haben die Empfehlungen der Lehrkräfte zumindest eine beratende Funktion (vgl. Hartinger et al., 2004) und sind somit trotzdem von hoher Relevanz. Um jedoch auszuschließen, dass die regionale Herkunft der Lehrkräfte in Studie 1 zu einer Verzerrung führt, wurde das entsprechende Bundesland erfasst, in dem die Lehrkräfte unterrichten. Da in Studie 2 ausschließlich Lehramtsstudierende einer bayerischen Universität rekrutiert wurden, musste dies in dieser Untersuchung nicht berücksichtigt werden.

7.1.2 Namensauswahl zur Aktivierung der ethnischen und sozialen Schülerherkunft

Bereits in zahlreichen empirischen Untersuchungen (z.B. Bertrand & Mullainathan, 2004; Bodenhausen & Lichtenstein, 1987; Glock & Krolak-Schwerdt, 2013; Hoenig & Wenz, 2013; Sprietsma, 2013; Schneider, Yemane & Weinmann, 2014) erwies sich die Manipulation der ethnischen und sozialen Herkunft durch Vornamen als erfolgreich. Entsprechend wurden in beiden Studien der vorliegenden Arbeit Vornamen zur Aktivierung der Herkunftswahrnehmung bei den Versuchspersonen eingesetzt. Somit lag besondere Bedeutung in der Auswahl der entsprechenden Namen, um sowohl die ethnische als auch die soziale Herkunft möglichst automatisiert zu aktivieren. Da Vornamen allerdings nicht nur diese beiden sozialen Kategorien implizieren können, sondern in erster Linie auch Informationen zum Geschlecht der Schüler(innen) beinhalten, musste dies ebenfalls in der Namensauswahl berücksichtigt werden. Da männliche Schüler mit größeren Herausforderungen im Bildungsverlauf konfrontiert zu sein scheinen, wurde in der vorliegenden Arbeit das männliche Geschlecht berücksichtigt.⁶³

Die Ergebnisse einer qualitativen Interview- und Beobachtungsstudie von Weber (2003) zeigen, dass einige Lehrkräfte der gymnasialen Oberstufe in der Fremdzuschreibung von Eigenschaften bei türkischen Migrant(inn)en deutliche Geschlechterdifferenzen zeigen. Zwar gibt es übereinstimmende ethnische Zuschreibungen (z. B. geringe Deutschkenntnisse, niedrige soziale Schicht der Herkunftsfamilien, starke familiäre Bindung), jedoch werden einige Charakteristika hauptsächlich Mädchen, andere überwiegend Jungen zugeschrieben. Die von Weber (2003) befragten Lehrkräfte heben bei Mädchen dabei eher ein angepasstes Verhalten hervor, bei Jungen beschreiben sie z.B. häufiger einen problematischen Umgang mit sozialen Regeln. Dabei wurde bei Mädchen oftmals weiter in eine introvertierte und eine eher verhaltensauffällige bzw. provokante Gruppe differenziert (Weber, 2003). Für Jungen beschreibt Weber (2003) dagegen eine eher eindimensionale Eigenschaftszuschreibung durch die Lehrkräfte, hier werden patriarchale Strukturen und Probleme im Sozialverhalten betont. Im US-amerikanischen Raum zeigt sich, dass afroamerikanische Schüler – trotz objektiv geringer Unterschiede in Verhaltensproblemen – häufiger wegen disziplinärer Auffälligkeiten sanktioniert werden als afroamerikanische Schülerinnen (APA, 2012). In Deutschland verlassen männliche Schüler mit Migrationshintergrund im Vergleich zu Schülerinnen mit Migrationshintergrund die Schule häufiger ohne Abschluss und erreichen durchschnittlich

⁶³ Die Entscheidung für ein Geschlecht erleichterte zudem einen möglichst ökonomischen Versuchsaufbau.

geringere Bildungsabschlüsse (Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration, 2014).

Essentielle Grundlage bei der Wahl der Namen zur Aktivierung spezifischer Stereotype, war die eindeutige Zuordnung des Vornamens zu einer ethnischen und sozialen Herkunft. Für Kinder ohne Migrationshintergrund lassen sich einige Namen auch bestimmten sozialen Schichten zuordnen. Utech (2011) analysierte die Vornamensvergabe der Geburtsjahrgänge 2004/2005. Hierbei zeigte sich, dass einige Vornamen, wie z. B. *Maximilian* und *Julius*, häufiger in höheren sozialen Schichten vergeben werden. Andere Namen, wie *Justin* und *Kevin*, waren häufiger in niedrigen sozialen Schichten vertreten (Utech, 2011). Zudem können Vornamen weitaus mehr Assoziationen als die ethnische und soziale Herkunft hervorrufen. Aus diesem Grund wurde – soweit empirische Daten vorlagen – ebenfalls berücksichtigt, ob die Vornamen als besonders attraktiv oder unattraktiv wahrgenommen wurden. Da Attraktivität oftmals mit Kompetenz oder Intelligenz assoziiert wird (z. B. Jackson et al., 1995; Webster & Driskell, 1983), wurden insbesondere Namensassoziationen mit Attraktivität und Intelligenz berücksichtigt. Namen die als besonders attraktiv (z. B. *Tim* oder *Lukas*) oder intelligent (z. B. *Christoph* oder *Maximilian*) wahrgenommen werden, wurden für die Studien dieser Arbeit genauso ausgeschlossen wie Namen, die als eher unattraktiv (z. B. *Roland* oder *Ralf*) oder weniger intelligent (z. B. *Helmut* oder *Werner*) wahrgenommen werden (Kröner & Dickhäuser, 2009; Rudolph, Böhm & Lummer, 2007; Rudolph & Spörrle, 1999).

Da diese Namensnormen jedoch nicht für alle Vornamen vorlagen, die nach Utech (2011) eindeutig einer sozialen Schicht zugeordnet werden konnten, konnten nur einige Vornamen durch bekannte Attraktivitäts- oder Intelligenzassoziationen ausgeschlossen werden. So wie z. B. *Maximilian*, der zwar einer hohen sozialen Schicht zugeordnet werden konnte, jedoch bereits allein durch den Namen als überdurchschnittlich intelligent und attraktiv wahrgenommen wird. Bedingt durch die lückenhafte Datenlage zur Wahrnehmung von Vornamen und Assoziationen mit Intelligenz und Attraktivität konnten diese Informationen bei der Namensauswahl nur teilweise berücksichtigt werden. Jedoch lagen zusätzliche Informationen zu assoziierten Verhaltensauffälligkeiten vor, die ebenfalls berücksichtigt wurden. In einer Befragung von $N = 500$ Grundschullehrkräften gaben 50% der Befragten an, dass sie den Namen *Kevin* mit Verhaltensauffälligkeiten assoziierten (Kaiser, 2009), somit wurde dieser Name ebenfalls ausgeschlossen.

Da im Rahmen der Literaturrecherche für diese Arbeit keine entsprechenden Untersuchungen aufgefunden werden konnten, konnten keine Befunde zu Assoziationen für türkische Vornamen mit Intelligenz oder Attraktivität berücksichtigt werden. Für türkischstämmige Schüler(innen) erweist sich eine Zuordnung zu hohen sozialen Schichten zudem als deutlich schwieriger. Lehrkräfte assoziieren mit Familien mit türkischen Wurzeln meist einen niedrigen sozioökonomischen Status und sehen eine hohe Sozialschicht eher als Ausnahme an (Weber, 2003). Somit ist davon auszugehen, dass mit einem türkischen Vornamen auch gleichzeitig ein niedriger sozioökonomischer Status assoziiert wird. Die Manipulation einer hohen sozialen Schicht bei Fallvignetten mit türkischem Vornamen wäre entsprechend nur über weitere Informationen möglich. Zusätzliche Informationen über die Familien, Wohnsituation oder Berufe der Eltern in den Fallvignetten könnten jedoch möglicherweise dazu führen, dass die Herkunft der Schüler zu offensichtlich im Wahrnehmungsfokus der Versuchspersonen

wäre, was evtl. zu einer generell eher kontrollierten Informationsverarbeitung oder stärkerem sozial erwünschtem Antwortverhalten führen könnte.

Um die Aktivierung eines türkischen Migrationshintergrunds durch den Vornamen zu erreichen, musste ein Name gewählt werden, der dieser ethnischen Herkunft möglichst eindeutig zugeordnet wird. Hierzu konnte auf Vorarbeiten zur Auswahl von Vornamen, die mit einem türkischen Migrationshintergrund assoziiert werden zurückgegriffen werden (Tobisch, 2013). Dabei wurden alle nicht eindeutig deutschen, männlichen Vornamen aus Daten einiger Standesämter der Geburtsjahrgänge 2002 bis 2009 (u. a. Gesellschaft für Deutsche Sprache, 2012; Standesamt Nürnberg, o. J.; Standesamt München, o. J.) ausgewählt. Anschließend gaben $N = 41$ Lehramtsstudierende an, welche dieser Namen bei ihnen die Assoziation mit einem türkischen Migrationshintergrund auslösen (Tobisch, 2013). Die Auswahl von typischen Vornamen, erscheint im Zusammenhang mit den Befunden von Macrae, Mitchell und Pendry (2002) besonders relevant, da in diesen Untersuchungen gezeigt werden konnte, dass bekannte Vornamen kategoriebasierte Informationsverarbeitung auslösen können und auch entsprechend assoziierte Merkmale aktiviert werden können, die u. a. zu stereotypen Bewertungen führen. Die Vornamen, die mit der höchsten Übereinstimmung mit einem türkischen Migrationshintergrund assoziiert wurden, kamen zunächst in eine engere Auswahl. Diese Namen wurden anschließend mit den deutschen Vornamen, die entsprechenden sozialen Schichten zugeordnet werden konnten, in der Anzahl der Buchstaben abgeglichen, um einen vergleichbaren Leseaufwand und einen symmetrischen Textaufbau zu ermöglichen. Aus diesen Voruntersuchungen und bereits vorhandenen empirischen Befunden wurden für alle Studien der vorliegenden Arbeit folgende drei Namen zur Aktivierung ethnischer und sozialer Stereotype gewählt (Tabelle 3):

Tabelle 2

Vornamen in den Fallvignetten und entsprechende Zuordnung zu den Kategorien Migrationshintergrund und sozioökonomischer Status

Vorname	Herkunftskombinationen in den Fallvignetten
Julius	Hoher sozioökonomischer Status; kein Migrationshintergrund
Justin	Niedriger sozioökonomischer Status; kein Migrationshintergrund
Murat	Niedriger sozioökonomischer Status; Migrationshintergrund

Da bei der Auswahl dieser Namen lediglich eine durchschnittliche Einschätzung der ethnischen und sozialen Herkunftsassoziationen berücksichtigt werden konnte, wurden in beiden Untersuchungen ein Manipulations-Check eingefügt, um die ethnische und soziale Herkunftswahrnehmung der jeweiligen Versuchspersonen zu prüfen. Nur durch eine Wahrnehmung der geplanten Herkunftskombinationen (vgl. Tabelle 3) durch die Versuchspersonen kann von einer erfolgreichen Manipulation der unabhängigen Variable ausgegangen werden.

7.2 Messung von Lehrkrafturteilen über und -erwartungen an Leistungen und leistungsrelevante Merkmale in Studie 1 und Studie 2⁶⁴

Da sowohl die akkurate und herkunftsunabhängige Lehrkräfteeinschätzung von Leistungen als auch von leistungsrelevanten Schülermerkmalen einen bedeutsamen Einfluss auf die Unterrichtsplanung und -durchführung sowie auf den Bildungsverlauf und -erfolg von Schüler(inne)n haben, fokussiert die vorliegende Arbeit auf entsprechende Lehrkrafturteile und -erwartungen als abhängige Variablen. Die Bedeutung von Lehrkrafterwartungen für die Lehrer-Schüler-Interaktion und den Lernerfolg von Schüler(inne)n wurde bereits in Abschnitt 3.2 erörtert. Dabei zeigt sich insbesondere die hohe Relevanz einer realistischen Erwartung sowie die möglichen Auswirkungen von (zu) hohen oder niedrigen Erwartungen auf lernförderliche oder -hemmende Interaktionsprozesse. Da Erwartungseffekte v. a. bei stigmatisierten Gruppen wirksam werden (z. B. Jussim & Harber, 2005), war die Erfassung von Lehrkrafterwartungen ein zentraler Aspekt dieser Arbeit. Durch die reziproke Beziehung zwischen Erwartungen und Urteilen, wurden neben auf die Zukunft ausgerichteten Einschätzungen auch Lehrkrafturteile über den aktuellen Stand der Schüler erfasst.

Da Lehrkräfte Schüler(innen) i. d. R. nicht global als lediglich gute oder schlechte Schüler(innen) beurteilen – sondern stattdessen z. T. verschiedene Aspekte wie z. B. die Persönlichkeit, Leistungen und soziale Aspekte berücksichtigen (Hoge, 1983), ist es notwendig bei der Erfassung von Lehrkrafturteilen auch spezifische Dimensionen im Gegensatz zu einem globalen Urteil abzudecken. In allen Untersuchungen dieser Arbeit wurden daher zentrale Faktoren für den Lern- und Leistungserfolg von Schüler(inne)n erfasst. Dabei wurden mit zwei Skalen die Leistungen fokussiert und zudem drei leistungsrelevante Aspekte (Gymnasialeignung, Anstrengungsbereitschaft und schulische Fähigkeiten) in die Untersuchung mit einbezogen (vgl. Kapitel 3).

Erfassung von Lehrkrafturteilen und -erwartungen an Schülerleistungen

Leistungserwartungen können auf vielfältige Weise erfasst werden, z. B. durch die Einschätzung des schulischen Potentials der Schüler(innen) oder die Schätzung zukünftiger schulischer Leistungen (Hoge, 1984). Die Einschätzung des schulischen Potentials ist aus theoretischer Perspektive jedoch nur bedingt als Erwartung anzusehen (vgl. Hoge, 1984), da Erwartungen auf die Zukunft gerichtete Einschätzungen darstellen (vgl. Abschnitt 3.2) und einen eher deterministischen Charakter aufweisen (vgl. Ludwig, 1991). Die Einschätzung des Potentials beinhaltet zwar implizit fähigkeitsbezogene zukünftige Möglichkeiten, könnte jedoch auch als Einschätzung aktueller Fähigkeiten verstanden werden und somit womöglich eine ausführlichere Instruktion benötigen. Aus diesem Grund, wurden Lehrkrafterwartungen für schulische Leistungen in der vorliegenden Arbeit explizit als Schätzung zukünftiger Leistungen operationalisiert. Diese wurden jeweils für die drei Hauptfächer Deutsch, Mathematik und Heimat- und Sachunterricht mit jeweils einem Item auf einer sechsstufigen Notenskala von 1 (*sehr gut*) bis 6 (*ungenügend*) erfasst („Welche Note würden Sie von dem Schüler in der

⁶⁴ In diesem Abschnitt werden lediglich die Instrumente beschrieben, die in beiden Studien zum Einsatz kamen. Ergänzende Erhebungsinstrumente in Studie 2 werden im entsprechenden Abschnitt 9.5 bzw. Abschnitt 9.1 für spezifische Eye-Tracking-Informationen beschrieben.

nächsten Probe in folgenden Fächern erwarten?“). Die Fokussierung auf diese drei Fächer erfolgte, da diese Fächer in der vierten Jahrgangsstufe der Grundschule eine bedeutsame Rolle für die Wahl einer weiterführenden Schulform darstellen (vgl. z. B. Ditton & Krüsken, 2006a).

Neben den konkreten Erwartungen an zukünftige Leistungen wurde zudem die *Leistungsaspiration* der Lehrkräfte erfasst. Frank (1935) beschreibt die Leistungsaspiration als „an objective estimate of a future level of performance on the basis of a past level of performance“ (S. 285) und kann nach Starbuck (1963) als subjektives Leistungsziel verstanden werden. In der vorliegenden Arbeit wird dabei in Anlehnung an Haller (1968) eine definitorische Trennung von Aspiration und Erwartung vorgenommen. Haller (1968) geht davon aus, dass beiden Konstrukten Zielstrukturen zugrunde liegen, die auf die Zukunft ausgerichtet sind. Dabei können idealistische von realistischen Aspirationen unterschieden werden (vgl. Becker, 2010a; Haller, 1968). Die Erfassung der Leistungserwartung fokussiert bereits eine möglichst realistische Einschätzung der zukünftigen Leistungsentwicklung der Schüler. Aspirationen fokussieren hingegen eher das Anstreben des Leistungsziels (vgl. z. B. Hill & Wang, 2015; Zander und Wolter, 2016). Durch die Messung der Leistungsaspiration werden hingegen weniger konkrete Erwartungen erfasst und in Anlehnung an Dresel, Stöger und Ziegler (2006) als „schlechteste, noch zufriedenstellende Zensur“ (S. 49) operationalisiert (vgl. auch Finsterwald, 2006). Hierzu wurde die Aspiration (ebenfalls separat für die drei Hauptfächer) mit jeweils einem Item („Mit welcher Note wären Sie bei diesem Schüler in der nächsten Probe in folgenden Fächern gerade noch zufrieden?“; adaptiert nach Dresel et al., 2006; Finsterwald, 2006) erfasst. Die Leistungsaspiration wurde, vergleichbar mit der Leistungserwartung, ebenfalls auf einer sechsstufigen Notenskala von 1 (*sehr gut*) bis 6 (*ungenügend*) gemessen. Beide leistungsbezogenen Maße wurden in den Studien dieser Arbeit für weitere Analysen rekodiert, um eine leichtere Interpretation der Ergebnisse zu ermöglichen. Entsprechend sind hohe Werte als hohe Leistungserwartungen und -aspirationen der Versuchspersonen zu interpretieren und niedrigere Werte als geringe Leistungserwartungen und -aspirationen.

Erfassung von Lehrkrafturteilen und -erwartungen an leistungsrelevante Schülermerkmale

Wie in Kapitel 3 verdeutlicht wurde, lassen sich zahlreiche Prädiktoren für schulische Leistungen identifizieren, die für einen adaptiven Unterricht entsprechend akkurat durch die Lehrkraft eingeschätzt werden müssen. Anstrengung und Fähigkeiten von Schüler(inne)n haben einen bedeutsamen Einfluss auf schulische Leistungen (vgl. Harackiewicz et al., 2002; Hochweber, Hosenfeld & Klieme, 2014). In dem Makromodell der Bedingungsfaktoren schulischer Leistungen werden dabei u. a. Fähigkeiten und motivationale Aspekte der Schüler(innen) genannt (Helmke, 2009). Die akkurate Einschätzung von Fähigkeiten ist auch bedeutsam um eine Über-/Unterforderung der Schüler(innen) zu vermeiden (vgl. Csikszentmihalyi, 1975). Nach der Attributionstheorie (Weiner, 1986) kann Erfolg und Misserfolg auf internaler Ebene entweder auf Anstrengung oder auf Fähigkeiten attribuiert werden, insbesondere im Hinblick darauf, dass attributionales Feedback der Lehrkraft einen Einfluss auf das Fähigkeitsselbstkonzept von Schüler(inne)n haben kann (Dresel & Ziegler, 2006). Auch Brühwiler, Helmke und Schrader (2017) verweisen im Rahmen des Angebots-Nutzungs-Modells auf den bedeutsamen Einfluss der Anstrengungsbereitschaft der Schüler(innen).

Entsprechend wurde in Anlehnung an Ramm et al. (2006) die Einschätzung der *Anstrengungsbereitschaft* mit fünf Items (z. B. „Wenn der Schüler eine Probearbeit schreibt, wird er sich sehr anstrengen.“) auf einer fünfstufigen Skala von 1 (*stimmt gar nicht*) bis 5 (*stimmt voll und ganz*) erhoben. Die Skala basiert auf einem Erhebungsinstrument der PISA 2003 Erhebungen. Die schulischen Fähigkeiten wurden dabei durch eine an die Lehrkraftperspektive adaptierte Version der *Subskala schulisches Selbstkonzept absolut* der *Skalen zur Erfassung des schulischen Selbstkonzepts* (SESSKO; vgl. Dickhäuser, Schöne, Spinath & Stiensmeier-Pelster, 2002) erfasst. Die bipolare fünfstufige Skala von 1 (*niedrig*) bis 5 (*hoch*) setzte sich aus vier Items (z. B. „Ich halte die Begabung des Schülers für...“) zusammen.

Ein wesentlicher Faktor für den Bildungsweg stellt auch die Einschätzung der *Eignung für das Gymnasium* durch Lehrkräfte dar. Da Lehrkräfte nicht nur die Leistungen der Schüler(innen) bei der Übergangsempfehlung berücksichtigen, sondern auch weitere Informationen in diese Entscheidung miteinfließen (vgl. Böhmer et al., 2015), wurde in der vorliegenden Arbeit explizit das Lehrkrafturteil über die aktuelle Eignung des Schülers für den Übertritt in eine höhere Schulform der Sekundarstufe I erfasst. Dies wurde mit einem fünfstufigen Item von 1 (*überhaupt nicht geeignet*) bis 5 (*voll und ganz geeignet*) erfasst („Ist der Schüler Ihrer Meinung nach für den Gymnasialbesuch geeignet?“).

8. Studie 1: Herkunftsassoziierte Verzerrungen von Lehrkrafturteilen und -erwartungen⁶⁵

Aufbauend auf bisherigen empirischen Befunden zu herkunftsassoziierten Lehrkrafturteilen und -erwartungen (vgl. Abschnitt 6.1), war das Ziel dieser Studie die differenzierte Analyse des Einflusses der ethnischen und sozialen Schülerherkunft auf Lehrkrafturteile und -erwartungen in Bezug auf Leistungen und leistungsrelevante Merkmale. Basierend auf dem Kontinuummodell der Eindrucksbildung von Fiske und Neuberg (1990) wurde in dieser Studie angenommen, dass Lehrkräfte Informationen bestimmter Schülergruppen (durch die Aktivierung von herkunftsassoziierten Stereotypen) eher automatisiert bzw. kontrolliert verarbeiten und sich dies in herkunftsassoziierten Urteilen und Erwartungen an Leistungen und leistungsrelevante Schülermerkmale zeigt (Fragestellung 1.1). Zudem sollte geklärt werden, ob es sich bei herkunftsabhängigen Unterschieden in Urteilen und Erwartungen von Lehrkräften um eine negative Urteilsverzerrung bei Schülern mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status handelt, oder ob unterschiedliche herkunftsassoziierte Lehrkrafturteile und -erwartungen das Resultat positiver verzerrter Urteile bei Schülern ohne Migrationshintergrund und hohem sozioökonomischem Status sind (Fragestellung 1.2).

Hierzu wurde eine experimentelle Online-Studie durchgeführt, in der erfahrene Lehrkräfte fiktive Schüler unterschiedlicher Herkunft, die in Fallvignetten präsentiert wurden, beurteilten. In den eingesetzten Fallvignetten wurde sowohl die ethnische als auch die soziale Herkunft der Schüler(innen) experimentell manipuliert (vgl. Abschnitt 7.1). So kann ausgeschlossen werden, dass Effekte des Migrationshintergrunds durch einen assoziierten niedrigen sozioökonomischen Status zustande kommen. Dabei wurden nach dem Lesen der Fallvignetten mit standardisierten Online-Fragebogen neben leistungsbezogenen Einschätzungen der Lehrkräfte für die Schulfächer Deutsch, Mathematik und Heimat- und Sachunterricht auch Urteile zu leistungsrelevanten Schülermerkmalen (Gymnasialeignung, Anstrengungsbereitschaft und Fähigkeiten) erfasst (vgl. Abschnitt 7.3). Zudem wurde untersucht, ob es sich bei herkunftsassoziierten Lehrkrafturteilen und -erwartungen um eine negative Urteilsverzerrung bei Schülern mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status handelt oder ob es sich um eine positive Verzerrung der Lehrkrafturteile bei Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status handelt. Um die Fallvignette sowie die Erhebungsinstrumente zu testen, wurde eine Vorstudie durchgeführt, die im Folgenden in einem Überblick dargestellt wird.

8.1 Vorstudie zum Einsatz von Fallvignetten und Testung der Erhebungsinstrumente

Die Vorstudie verfolgte zwei wesentliche Ziele: (1) Die Erhebungsinstrumente zur Erfassung von Urteilen über und Erwartungen an Leistungen und leistungsrelevante Merkmale sollten getestet werden. (2) Die konstruierten Fallvignetten in Form eines Halbjahreszeugnisses

⁶⁵ Teilergebnisse dieser Studie wurden bereits im Rahmen eines Zeitschriftenartikels veröffentlicht: Tobisch & Dresel (2017).

sollten erprobt werden. Zu diesem Zweck wurde eine experimentelle Studie im between-subject-Design konzipiert und mit Hilfe der Software EFS-Survey (Unipark, 2016) online durchgeführt. Über spezifische Online-Foren für Lehrkräfte wurden hierfür im November und Dezember 2013 Proband(inn)en für die Studie rekrutiert. Für die Teilnahme an der Studie erhielten die Versuchspersonen einen Gutschein in Höhe von 3,- € für einen Online Versandhändler. Insgesamt nahmen bundesweit 130 Lehrkräfte an der Untersuchung teil. Jedoch mussten 23 Versuchspersonen aus den Analysen ausgeschlossen werden, da entweder spezifische Antwortmuster oder eine zu geringe Bearbeitungszeit des Fragebogens (unter 100 Sekunden) eine ersthafte Bearbeitung der Aufgaben in Frage stellten. Die Analysestichprobe bestand somit aus $N = 107$ Lehrkräften mit einer durchschnittlichen Berufserfahrung von $M = 10.63$ Jahren ($SD = 9.67$), davon waren 61.5 % weiblich. Das durchschnittliche Alter der Lehrkräfte lag bei $M = 37.52$ Jahren ($SD = 8.81$). Mit 39.6 % war der überwiegende Teil der Versuchspersonen als Lehrkraft an einer Realschule beschäftigt, gefolgt von 32.1 % Gymnasiallehrkräften, 8.5 % Grundschullehrkräften, 5.7 % Haupt-/Mittelschullehrkräften und 3.8 % der Lehrkräfte waren an einer Sonder-/Förderschule tätig (6.6 % der befragten Lehrkräfte gaben an, an einer anderen Schulform zu unterrichten, z.B. Gesamtschule; während für 3.8 % der Teilnehmer(inne)n keine Angaben hierzu vorlagen).

Die hierfür entwickelten Fallvignetten bestanden aus Textfragmenten zum Arbeits- und Sozialverhalten, Informationen zur Regelmäßigkeit des Schulbesuchs, dem Ordnungsverhalten sowie dem Verhalten des Schülers gegenüber der Lehrkraft (vgl. Abschnitt 7.1). Die Fallvignetten waren für alle Versuchspersonen identisch, bis auf die randomisierte Zuweisung einer spezifischen ethnisch-sozialen Herkunftskombination, die durch die Vornamen der Schüler impliziert wurde. Die Versuchspersonen konnten dabei ohne zeitliche Einschränkung zunächst die Verbalbeurteilung und anschließend die Ziffernbewertung des Schülers lesen. Im Anschluss daran erhielten alle Lehrkräfte einen Fragebogen, in dem sie ihre Einschätzungen über zukünftige Leistungen und leistungsrelevante Merkmale des jeweiligen Schülers angeben sollten.

Wie in Abschnitt 7.1.2 beschrieben wurde, sollte der Vorname *Julius* einen Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status, der Vorname *Justin* einen Schüler ohne Migrationshintergrund und mit niedrigem sozioökonomischem Status und der Vorname *Murat* einen Schüler mit türkischem Migrationshintergrund und niedrigem sozioökonomischem Status beschreiben (vgl. Tabelle 4).

Tabelle 3

Übersicht über die Verteilung der Versuchspersonen zu den drei Fallvignetten nach Herkunftskombination der Schüler (Vorstudie 1)

Herkunftskombinationen in den Zeugnisvarianten	Vorname	<i>n</i>
Hoher sozioökonomischer Status; kein Migrationshintergrund	Julius	34
Niedriger sozioökonomischer Status; kein Migrationshintergrund	Justin	35
Niedriger sozioökonomischer Status; Migrationshintergrund	Murat	38

Anmerkung. $N = 107$.

Hierdurch ergaben sich drei Fallvignetten, die aus der Verbalbeurteilung und Ziffernbewertung eines Halbjahreszeugnisses der vierten Jahrgangsstufe mit identischem Aufbau und Inhalten bestand, in denen lediglich der Vorname der Schüler variiert wurde. Durch die randomisierte Zuweisung der Versuchspersonen zu einer der drei Zeugnisvarianten ergab sich dabei die in Tabelle 4 einsehbare Verteilung der Versuchspersonen zu einer der drei Fallvignetten nach Herkunftskombination.

Erhebungsinstrumente

Um herkunftsspezifische Lehrkrafturteile und -erwartungen an *Schülerleistungen* zu erfassen, wurden alle Versuchspersonen um die Angabe der konkreten Leistungserwartungen an Noten in zukünftigen Proben in den Hauptfächern Deutsch, Mathematik sowie Heimat- und Sachunterricht gebeten („Welche Note würden Sie von dem Schüler in der nächsten Probe in folgenden Fächern erwarten?“).⁶⁶ Dabei wurden die Leistungserwartungen für alle drei Fächer in den gängigen Notenstufen von 1 (*sehr gut*) bis 6 (*ungenügend*) erhoben. Insgesamt ergab sich eine durchschnittliche Leistungserwartung (rekodiert) von $M = 4.69$ ($SD = 0.49$) in den drei Hauptfächern. Die interne Konsistenz der Skala Leistungserwartung war mit $\alpha = .42$ nicht zufriedenstellend, was sich jedoch inhaltlich durch die domänenspezifischen Unterschiede erklären lässt. Neben der Leistungserwartung wurden ebenfalls die Leistungsaspiration („Mit welcher Note wären Sie bei diesem Schüler in der nächsten Probe in folgenden Fächern gerade noch zufrieden?“; $\alpha = .83$) in Anlehnung an Dresel et al. (2006) und Finsterwald (2006) in den Notenstufen 1 (*sehr gut*) bis 6 (*ungenügend*) erhoben. Die Leistungsaspiration (rekodiert) der Lehrkräfte lag durchschnittlich bei $M = 4.37$ ($SD = 0.82$). Da sich für die Leistungserwartung keine akzeptable interne Konsistenz ergab, wurde in den folgenden Analysen der Vorstudie eine fachspezifische Trennung vorgenommen.

Da auch eine akkurate und herkunftsunabhängige Beurteilung *leistungsrelevanter Merkmale* wie der Eignung für das Gymnasium, motivationaler Aspekte sowie allgemeiner schulischer Fähigkeiten von Bedeutung für adaptiven Unterricht sind (vgl. Kapitel 3), wurden die Lehrkräfte auch zu diesen Aspekten um ihre Einschätzung gebeten. Zur Validierung der adaptierten Skalen, die in Abschnitt 7.2 beschrieben wurden, wurden für jedes Konstrukt zunächst zwei ähnliche Konstrukte erfasst. So sollte inhaltlich durch entsprechende Korrelationen mit weiteren Variablen geklärt werden, ob das jeweilige Messinstrument geeignet ist. Da insbesondere Single-Item Messungen in ihrer Güte eingeschränkt sind, wurden diese zunächst kritisch betrachtet.

Die Beurteilung der Eignung für den Besuch eines Gymnasiums wurde auf zweifache Weise erfasst. Zum einen sollten die Lehrkräfte angeben, ob sie dem Schüler eine Empfehlung für den Gymnasialbesuch aussprechen würden (dichotom: 0 [*nein*] und 1 [*ja*]). Über alle Fallvignetten hinweg, gaben 84.1 % der Lehrkräfte an, dass sie eine Empfehlung für das Gymnasium aussprechen würden. Zum anderen wurde mit einem fünfstufigen, von 1 (*überhaupt nicht geeignet*) bis 5 (*voll und ganz geeignet*), Item („Ist der Schüler Ihrer Meinung nach für

⁶⁶ Zur leichteren Interpretation wurden die Noten der Leistungserwartung und -aspiration rekodiert. Hohe Werte entsprechend somit einer hohen Erwartung und Aspiration.

den Gymnasialbesuch geeignet?“) erfasst, ob die Versuchspersonen den Schüler geeignet für den Gymnasialbesuch halten. Dabei wurde der Schüler mit $M = 3.73$ ($SD = 0.82$) im Durchschnitt als eher geeignet für den Gymnasialbesuch angesehen.

Die Erhebung der Lehrkrafturteile und -erwartungen für motivationale Aspekte erfolgte ebenfalls durch zwei unterschiedliche Vorgehensweisen. Mit einem Item („Wie schätzen Sie die schulische Motivation des Schülers ein?“) wurde auf einer fünfstufigen Skala von 1 (*sehr gering*) bis 5 (*sehr hoch*) die Einschätzung der allgemeinen schulischen Motivation der Schüler erfragt. Die durchschnittliche Motivation der Schüler wurde mit $M = 3.58$ ($SD = 0.69$) als eher hoch eingeschätzt. Neben der allgemeinen Erfassung der schulischen Motivation wurde auch die Anstrengungsbereitschaft der Schüler als wichtige motivationale Komponente in Anlehnung an Ramm et al. (2006) mit fünf Items (z. B. „Wenn der Schüler eine Probearbeit schreibt, wird er sich sehr anstrengen.“) auf einer fünfstufigen Skala von 1 (*stimmt gar nicht*) bis 5 (*stimmt voll und ganz*) erhoben ($M = 3.85$, $SD = 0.57$). Die Skala erwies sich mit $\alpha = .72$ als zufriedenstellend intern konsistent.

Auch die Beurteilung der schulischen Fähigkeiten wurde durch zwei Aspekte erfasst. Zum einen wurde auf die kognitiven Fähigkeiten fokussiert und mit einem Item nach der Einschätzung der Intelligenz („Wie schätzen Sie die allgemeine kognitive Leistungsfähigkeit [im Sinne von Intelligenz] des Schülers ein?“) gebeten. Dabei lagen fünf Antwortoptionen von 1 (*deutlich unterdurchschnittlich*) bis 5 (*deutlich überdurchschnittlich*) vor. Die Lehrkräfte schätzten die durchschnittlichen kognitiven Fähigkeiten als leicht überdurchschnittlich ein ($M = 3.78$, $SD = 0.59$). Zum anderen wurden die allgemeinen schulischen Fähigkeiten auf einer bipolaren fünfstufigen Skala von 1 (*niedrig*) bis 5 (*hoch*) (z. B. „Ich halte die Begabung des Schülers für...“; $\alpha = .79$) erfasst (adaptiert nach Dickhäuser et al., 2002). Die Versuchspersonen schätzten die schulischen Fähigkeiten der Schüler dabei durchschnittlich eher hoch ein ($M = 3.94$; $SD = 0.65$). Da alle Items im Online-Fragebogen als Pflichtitems deklariert wurden, gab es keine fehlenden Werte für die vorliegenden Daten.

Neben den Informationen zu den internen Konsistenzen der Skalen wurden ebenfalls korrelative Befundmuster berücksichtigt um entsprechende Skalen für die Hauptstudie auszuwählen. Tabelle 5 sind die bivariaten Korrelationen der Lehrkrafteinschätzungen zu Leistungen und leistungsrelevante Schülermerkmale zu entnehmen. Dabei zeigten sich überwiegend erwartungsgemäß mittlere bis hohe Korrelationen zwischen der Einschätzung der Leistungen und leistungsrelevanten Merkmalen. Jedoch erbrachten die Analysen der Leistungserwartungen in den Fächern Deutsch sowie Heimat- und Sachunterricht keine Korrelationen mit den Leistungsaspirationen der entsprechenden Fächer. Dieser unerwartete Befund könnte dabei im Zusammenhang mit der geringen internen Konsistenz der Skala Leistungserwartung stehen. Sowohl die Reliabilitätsanalyse als auch die Korrelationsmuster verweisen auf Probleme bei der Messung. Entsprechend kann die Skala zur Leistungserwartung in der Hauptstudie nur unter Vorbehalt eingesetzt werden. Auch die Ergebnisse der Korrelationen der Motivation mit anderen leistungsrelevanten Merkmalen zeigten lediglich geringe signifikante Zusammenhänge mit der Gymnasialempfehlung und der Einschätzung der Gymnasialeignung. Jedoch hing die Einschätzung der allgemeinen Motivation moderat mit der Einschätzung der Anstrengungsbereitschaft zusammen.

Tabelle 5
Bivariate Korrelationen der Einschätzungen von Leistungen und leistungsrelevanten Merkmalen (Studie 1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Leistungsaspiration (rekodiert)											
(1) Deutsch	-										
(2) Mathematik	.68**	-									
(3) Heimat- und Sachunter- richt	.61**	.59**	-								
Leistungserwartung (rekodiert)											
(4) Deutsch	.13	-.01	.23*	-							
(5) Mathematik	.51**	.44**	.31**	.12	-						
(6) Heimat- und Sachunter- richt	.12	-.08	.07	.34**	.17	-					
Leistungsrelevante Merkmale											
(7) Gymnasialempfehlung	.44**	.56**	.40**	.05	.26**	.02	-				
(8) Gymnasialeignung	.59**	.60**	.54**	.13	.50**	.17	.70**	-			
(9) Motivation	.17	.21*	.01	-.06	.09	.13	.29**	.22*	-		
(10) Anstrengungsbereitschaft	.48**	.56**	.45**	.07	.46**	.01	.45**	.51**	.43**	-	
(11) Kognitive Fähigkeiten	.58**	.62**	.57**	.07	.57**	.04	.53**	.68**	.26**	.70**	-
(12) Schulische Fähigkeiten	.34**	.55**	.37**	.02	.43**	-.04	.61**	.55**	.36**	.69**	.67**

Anmerkungen. Kodierung der Gymnasialempfehlung 0 (nein), 1(ja). ** $p < .01$. * $p < .05$, $N = 107$.

Basierend auf den vorliegenden Befunden zu den einzelnen Messinstrumenten erfolgte für die leistungsrelevanten Schülermerkmale eine Reduktion auf jeweils ein Messinstrument zur Erfassung des jeweiligen Konstrukts in der Hauptstudie. Mittels explorativer Faktorenanalysen sollte zusätzlich geklärt werden ob die weiteren Variablen, die zur Validierung gedacht waren, womöglich doch einen zusätzlichen Erklärungswert beinhalten könnten und entsprechend auch in Studie 1 beibehalten werden sollten.

Die Lehrkräfteeinschätzungen der Gymnasialempfehlung und der Gymnasialeignung korrelierten mit $r = .70$ ($p < .01$) hoch miteinander.⁶⁷ Durch den höheren Informationsgehalt der kontinuierlichen Variable und die entsprechend inhaltlich vergleichbare Aussagen beider Variablen, wurde für Studie 1 die Gymnasialeignung der Schüler erfasst und auf die konkrete dichotome Empfehlung durch die Lehrkräfte verzichtet. Die beiden motivationalen Aspekte korrelierten moderat miteinander ($r = .43$, $p < .01$). Entgegen der Erwartung hing das Motivationsitem nicht mit den Leistungen oder der Gymnasialeignung zusammen. Die Anstrengungsbereitschaft erbrachte hingegen die erwarteten korrelativen Befundmuster mit den Leistungen und leistungsrelevanten Schülermerkmalen. Die Befunde der Faktorenanalyse ($MSA = .75$) zeigten, dass auch das einzelne Motivationsitem mit einer Ladung von 0.59 auf dem gleichen Faktor wie die Items der Anstrengungsbereitschaft lud. Basierend auf diesem Befund ist von keinem zusätzlichen Erklärungsgehalt durch das einzelne motivationale Item auszugehen. Die korrelativen Befunde standen zudem entgegen der theoretisch anzunehmenden Erwartungen, somit wurde für Studie 1 die Skala zur Anstrengungsbereitschaft als motivationale Komponente ausgewählt und das Item zur allgemeinen Motivation ausgeschlossen.

Mit $r = .67$ ($p < .01$) wies die Einschätzung der kognitiven Fähigkeiten einen moderaten Zusammenhang zu den schulischen Fähigkeiten auf. Beide Messinstrumente korrelierten erwartungsgemäß mit den Leistungen und weiteren leistungsrelevanten Schülermerkmalen. Die explorative Faktorenanalyse ($MSA = .75$) erbrachte eine Ladung des Einzelitems zu kognitiven Fähigkeiten von 0.81 auf den gleichen Faktor wie die Items der Skala zur Einschätzung der allgemeinen schulischen Fähigkeiten. Durch dieses Befundmuster konnten keine zusätzlichen inhaltlichen Informationen erwartet werden. Dabei wurde die Skala zur Erfassung der schulischen Fähigkeiten als geeignetes Instrument für Studie 1 ausgewählt, da angenommen wird, dass Lehrkräfte kognitive Fähigkeiten und Leistungen schwer voneinander unterscheiden können (vgl. Schrader & Helmke, 1990). Hierfür könnten auch die höheren Korrelationen der Einschätzung der kognitiven Fähigkeiten im Vergleich zu den Korrelationen der Einschätzung schulischer Fähigkeiten mit den Leistungsurteilen sprechen. Zudem liefert die Verbalbeurteilung der Schüler im Stimulusmaterial der vorliegenden Arbeit mehr Informationen über das schulische Verhalten, über welches eher Rückschlüsse auf die allgemeinen schulischen Fähigkeiten im Vergleich zu kognitiven Fähigkeiten möglich sind.

⁶⁷ Der Kaiser-Meyer-Olkin Test erbrachte für beide Variablen einen nicht geeigneten Wert (*measure of sampling adequacy*; $MSA = .50$) zur Berechnung einer Faktorenanalyse (vgl. Backhaus, Erichson, Plinke & Weiber, 2000).

Fallvignetten

Neben der Prüfung der eingesetzten Erhebungsinstrumente verfolgte diese Vorstudie ebenfalls das Ziel, erste Erfahrungen mit dem Stimulusmaterial zu sammeln. Die Fallvignetten wurden von den Lehrkräften als Grundschulzeugnis der vierten Jahrgangsstufe akzeptiert. Die Lehrkräfte hatten im Fragebogen die Möglichkeit in einem offenen Antwortformat Kommentare zur Untersuchung anzugeben. Zwar wurde von einigen Lehrkräften angemerkt, dass der persönliche Kontakt zu Schülern mehr Informationen liefert, die Aussagen über zukünftiges Verhalten erleichtern würden. Die Fallvignetten wurden insgesamt jedoch gut angenommen. Um zu vermeiden, dass Versuchspersonen sich ihr Urteil nur durch positive oder negative Informationen bilden, die in Abschnitten der Vignette gehäuft vorkommen, wurden im Anschluss an die Voruntersuchung positive und negative Textfragmente im Verlauf der Fallvignette gleichmäßig verteilt. Die Inhalte und Formulierungen der Textfragmente bleiben identisch, bis auf einzelne kleinere Anpassungen (z. B. Grammatik), um eine leichtere Lesbarkeit, infolge der Umstrukturierung der Textabschnitte, zu ermöglichen.

In dieser Voruntersuchung bleibt jedoch unklar, ob die Lehrkräfte die Vornamen auch tatsächlich mit den entsprechenden Herkunftskombinationen assoziierten, was jedoch nicht Ziel der Vorstudie war. Dementsprechend ist ein Manipulations-Check für Studie 1 noch ausstehend. Eine erste inhaltliche Annäherung war durch diese Voruntersuchung somit nicht möglich. Problematisch erscheint die heterogene Zusammensetzung der Stichprobe in Bezug auf Schulformen, an denen die Lehrkräfte unterrichten. Da es sich beim Stimulusmaterial um Grundschulzeugnisse handelt, das Curriculum und die Anforderungen der Grundschule überwiegend Lehrkräften der entsprechenden Schulform bekannt sind sowie die Einschätzung der Gymnasialeignung ebenfalls eine grundschulspezifische Aufgabe darstellt, wurde für Studie 1 eine entsprechende Auswahl der Stichprobe intendiert.

8.2 Spezifische Fragestellungen und Hypothesen von Studie 1

Aufbauend auf den Vorarbeiten der Vorstudie insbesondere zum Design, dem erstellten Stimulusmaterial und den Messinstrumenten, verfolgte die Hauptstudie das Ziel, herkunftsassoziierte Urteile und Erwartungen von erfahrenen Grundschullehrkräften zu untersuchen. Bisherige empirische Befunde zu herkunftsabhängigen Unterschieden in Lehrkrafturteilen und -erwartungen fokussierten jedoch meist nur die Schülerleistungen (z. B. Bonefeld et al., 2017; Darley & Gross, 1983; Sprietsma, 2013; Tom et al., 1984). Es zeigt sich v. a. für den Einfluss des Migrationshintergrunds eine eher heterogene Befundlage (vgl. z. B. Chen & Weseley, 2011; Überblick bei Kaiser et al., 2016). Nur vereinzelt wurden dabei auch leistungsrelevante Schülermerkmale berücksichtigt (z. B. kognitive Fähigkeiten: Alvidrez & Weinstein, 1999; Verhalten, Downey & Pribesh, 2004; Gymnasialeignung: Sprietsma, 2013). Eine Analyse herkunftsassoziierter Urteile und Erwartungen bezogen auf motivationale Aspekte und schulische Fähigkeiten blieb insbesondere im deutschsprachigen Raum bislang aus.

Aus diesem Grund fokussierte die erste Forschungsfrage dieser Arbeit darauf, ob sich bei erfahrenen Grundschullehrkräften herkunftsabhängige Urteils- und Erwartungsunterschiede sowohl für schulische Leistungen als auch für leistungsrelevante Schülermerkmale identifi-

zieren lassen. Basierend auf den theoretischen Annahmen einer möglichen stereotypgeleiteten Informationsverarbeitung (vgl. Kapitel 5) sowie empirischen Befunden zu negativen Einstellungen von Lehrkräften gegenüber Schüler(inne)n mit Migrationshintergrund und positiven Einstellungen gegenüber Schüler(inne)n mit hohem sozioökonomischem Status (z. B. Glock & Karbach, 2015; Glock et al., 2016), wurden folgende gerichtete Hypothesen formuliert:

H_{1.1}: Lehrkrafturteile über und -erwartungen an Leistungen sowie leistungsrelevante Merkmale fallen für Schüler mit niedrigem sozioökonomischem Status schlechter aus als für Schüler mit hohem sozioökonomischem Status.

H_{1.2}: Lehrkrafturteile über und -erwartungen an Leistungen sowie leistungsrelevante Merkmale fallen für Schüler mit Migrationshintergrund schlechter aus als für Schüler ohne Migrationshintergrund.

Geringere Erwartungen an und negativere Urteile über Schüler mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status müssen jedoch nicht Ausdruck einer direkten Benachteiligung sein. So konnten Ready und Chu (2015) für den US-amerikanischen Raum u. a. positiv verzerrte Lehrkrafturteile für Vorschulkinder mit hohem sozioökonomischem Status berichten. Unklar ist bisher jedoch, ob sich positive Urteilsverzerrungen auch im deutschsprachigen Raum für Schüler(innen) ohne Migrationshintergrund und mit hohem sozioökonomischem Status auffinden lassen. Bedingt durch negative Einstellungen gegenüber Schüler(inne)n mit Migrationshintergrund und/oder mit niedrigem sozioökonomischem Status ist allerdings auch eine negative Verzerrung bei der Urteils- und Erwartungsbildung von Lehrkräften zu vermuten. Da davon auszugehen ist, dass sowohl für Schüler mit Migrationshintergrund, als auch für Schüler mit niedrigem sozioökonomischem Status negative Einstellungen und Stereotype vorherrschen, wurden für die folgenden Hypothesen beide Herkunftsmerkmale verschränkt. Zudem kann davon ausgegangen werden, dass insbesondere die Kombination keines Migrationshintergrunds und eines hohen sozioökonomischen Status zu einer Assoziation von positiven Merkmalen und damit auch zu einer positiven Urteilsverzerrung führen kann. Entsprechend wurden die beiden folgenden gerichteten Hypothesen formuliert:

H_{1.3}: Lehrkrafteerwartungen und -aspirationen sind für Schüler mit Migrationshintergrund und/oder mit niedrigem sozioökonomischem Status negativ verzerrt.

H_{1.4}: Lehrkrafteerwartungen und -aspirationen sind für Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status positiv verzerrt.

Zur Prüfung dieser Hypothesen wurde eine Studie mit erfahrenden Grundschullehrkräften durchgeführt, deren Design und Stichprobe im folgenden Abschnitt näher beschrieben werden.

8.3 Untersuchungsdesign und Stichprobe der ersten Studie

In einer experimentellen Online-Studie im between-subject-Design wurden Grundschullehrkräfte um das Lesen einer Fallvignette in Form eines Halbjahreszeugnis der vierten Jahrgangsstufe (Verbal- und Ziffernbewertung; vgl. Abschnitt 7.1.1) gebeten. Da es sich bei den fiktiven Schülern in der Fallvignette um Grundschüler handelte, wurden gezielt Grundschul-

Lehrkräfte in spezifischen Online-Foren (die zum Austausch von Lehrkräften gedacht sind) angeworben. Für ihre Teilnahme erhielten die Lehrkräfte einen Gutschein in Höhe von 3,- € für einen Online-Versandhändler. Die Erhebung erfolgte dabei im Januar und Februar 2014 mittels Online-Fragebogen über die Software EFS-Survey (Unipark, 2016). Insgesamt wurden $N = 237$ Grundschullehrkräfte mit einer durchschnittlichen Berufserfahrung von $M = 15.34$ Jahren ($SD = 8.37$) befragt. Die Lehrkräfte waren dabei im Durchschnitt $M = 41.62$ ($SD = 8.49$) Jahre alt und wurden in allen Bundesländern deutschlandweit rekrutiert, wobei der größte Anteil aus Nordrhein-Westfalen (19.1 %) und Bayern (12.3 %) kam. Der Frauenanteil in der Stichprobe betrug 48.9 %.

Die Versuchspersonen konnten die Fallvignette in ihrem eigenen Tempo und ohne Zeitdruck lesen. Anschließend erhielten die Teilnehmer(innen) einen standardisierten Fragebogen, in dem ihre Erwartungen und Urteile über Leistungen und leistungsrelevante Merkmale des jeweiligen Schülers der Fallvignette erfragt wurden. In der Fallvignette wurde über den Vornamen der Schüler eine spezifische Herkunftskombination manipuliert (vgl. Abschnitt 7.1.2). Der Vorname *Julius* sollte keinen Migrationshintergrund und einen hohen sozioökonomischen Status beim Lesen aktivieren, der Vorname *Justin* sollte beim Bearbeiten der Fallvignette ebenfalls keinen Migrationshintergrund jedoch einen niedrigen sozialen Status aktivieren und der Vorname *Murat* sollte von den Versuchspersonen als Schüler mit türkischem Migrationshintergrund und niedrigem sozioökonomischem Status wahrgenommen werden. Jede Versuchsperson wurde randomisiert einer Fallvignette mit einer der drei Herkunftskombinationen zugewiesen (vgl. Tabelle 6).

Tabelle 4

Übersicht über die Verteilung der Versuchspersonen auf die drei Fallvignetten nach Herkunftskombination der Schüler (Studie 1)

Herkunftskombination in den Zeugnisvarianten	Vorname	<i>n</i>
Hoher sozioökonomischer Status; kein Migrationshintergrund	Julius	84
Niedriger sozioökonomischer Status; kein Migrationshintergrund	Justin	77
Niedriger sozioökonomischer Status; Migrationshintergrund	Murat	76

Anmerkung. $N = 237$.

Nach Bearbeitung der inhaltlichen Fragen zur Fallvignetten wurden die Lehrkräfte um Einschätzung der ethnischen und sozialen Herkunft der Schüler gebeten. Mit einem dichotomen Item – 1 (*Deutschland*) und 2 (*anderes Land*) – wurde nach dem Herkunftsland der Familie des Schülers gefragt („Was denken Sie, aus welchem Land der Schüler bzw. seine Familie kommt?“). Wurde ein anderes Land als Deutschland angegeben, wurden die Lehrkräfte darum gebeten in einem offenen Antwortfeld das von ihnen vermutete Herkunftsland der Schüler bzw. ihrer Familie einzutragen. Die anschließenden Analysen bestätigten eine erfolgreiche Manipulation der ethnischen ($\chi^2(2, N = 237) = 228.06, p < .001, \text{Cramer's } V = .981$) Herkunft. Dabei

wurde *Murat* auch im Abgleich mit den offenen Antwortoptionen als Schüler mit türkischem Migrationshintergrund wahrgenommen, während *Julius* und *Justin* als Schüler ohne Migrationshintergrund wahrgenommen wurden.

Mit einem fünfstufigen Item von 1 (*Unterschicht*) bis 5 (*Oberschicht*) wurde zudem um die Einschätzung der sozialen Schicht der Schüler gebeten („Was denken Sie, welcher sozialen Schicht der Schüler bzw. seine Familie zugehörig ist?“). Auch die Analyse der Angaben zum sozioökonomischem Status der Schüler zeigte eine erfolgreiche Manipulation durch die Vornamen ($F(2,234) = 28.024, p < .001, \eta^2 = .19$). Ein Vergleich der Mittelwerte (vgl. Abbildung 11) verdeutlicht dabei, dass erwartungsgemäß dem Schüler mit dem Vornamen *Julius* die höchste soziale Schicht zugeordnet wurde. Die beiden anderen Schüler wurden eher durchschnittlich bis unterdurchschnittlich eingestuft.

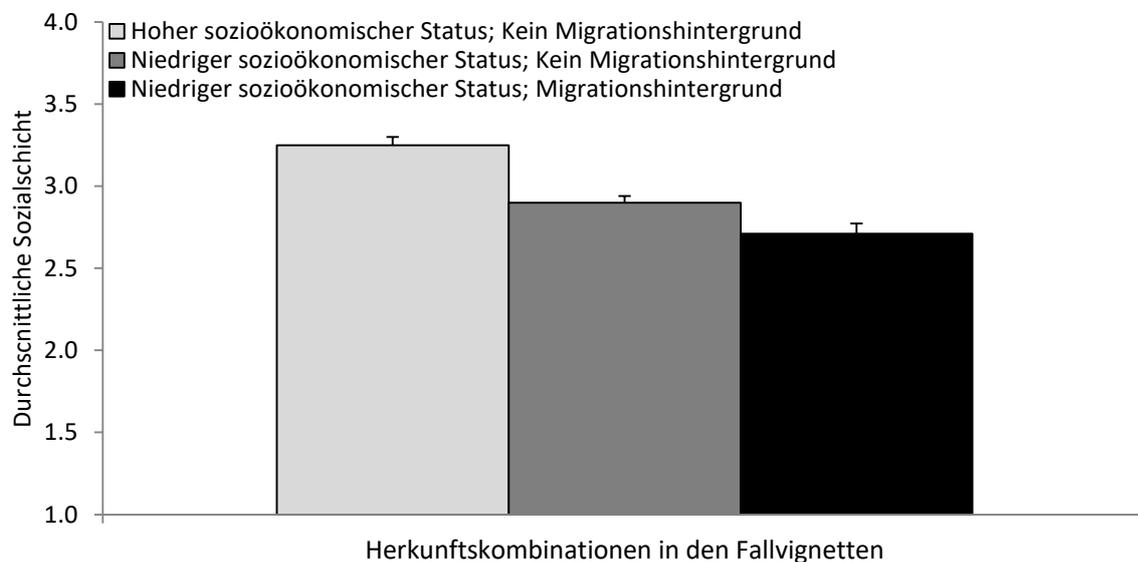


Abbildung 11. Mittelwerte und Standardfehler der eingeschätzten sozialen Herkunft der Schüler nach Herkunftskombinationen in den Fallvignetten

8.4 Erhebungsinstrumente in Studie 1

Eingesetzt wurden zwei Skalen zu Erfassung der leistungsbezogenen Lehrkrafturteile und -erwartungen sowie drei Skalen zur Erfassung der Lehrkrifteinschätzung leistungsrelevanter Schülermerkmale, die in der Vorstudie (Abschnitt 8.1) validiert wurden und in Abschnitt 7.2 beschrieben wurden. Zunächst wurde die Leistungsaspiration in Anlehnung an Dresel et al. (2006) und Finsterwald (2006) auf einer sechsstufigen Notenskala von 1 (*sehr gut*) bis 6 (*ungenügend*) für die drei Hauptfächer Deutsch, Mathematik sowie Heimat- und Sachunterricht erfasst („Mit welcher Note wären Sie bei diesem Schüler in der nächsten Probe gerade noch zufrieden?“; $\alpha = .79$). Neben der Leistungsaspiration wurde auch die Leistungserwartung für die Fächer Deutsch, Mathematik und Heimat- und Sachunterricht in den Notenstufen von 1 (*sehr gut*) bis 6 (*ungenügend*) („Welche Note würden Sie in der nächsten Probe von diesem Schüler erwarten?“; $\alpha = .77$) erhoben. Beide Skalen zur Leistungseinschätzung zeigten zufriede-

denstellende interne Konsistenzen. Indikatoren zu leistungsabhängigen Einschätzungen der Lehrkräfte wurden für die Analysen rekodiert um die Interpretation hoher Werte mit hohen Leistungserwartungen und -aspirationen zu erleichtern. Als leistungsrelevante Merkmale wurden die Lehrkräfteeinschätzungen der Gymnasialeignung, der Anstrengungsbereitschaft sowie der Fähigkeiten erfasst (vgl. Abschnitt 7.2). Mit einem Item („Ist der Schüler Ihrer Meinung nach für das Gymnasium geeignet?“) wurde die Einschätzung der Lehrkraft zur Eignung der Schüler für den Gymnasialbesuch auf einer fünfstufigen Skala von 1 (*überhaupt nicht geeignet*) bis 5 (*voll und ganz geeignet*) erhoben ($M = 3.59$, $SD = 0.71$). Die Lehrkräfte schätzten die Anstrengungsbereitschaft der Schüler mit fünf Items auf einer fünfstufigen Skala (adaptiert nach Ramm et al., 2006) von 1 (*stimmt gar nicht*) bis 5 (*stimmt voll und ganz*) ein (Beispielitem: „Wenn der Schüler eine Probearbeit schreibt, wird er sich sehr anstrengen.“, $\alpha = .87$). Die allgemeinen schulischen Fähigkeiten der Schüler wurden von den Lehrkräften mit vier Items auf einer fünfstufigen bipolaren Skala (adaptiert nach Dickhäuser et al., 2002) von 1 (*niedrig/schwer*) bis 5 (*hoch/leicht*) eingeschätzt (Beispielitem: „Neues zu lernen fällt dem Schüler...“, $\alpha = .88$).

Außer für die Einschätzung der Fähigkeiten der Schüler (18.6% fehlende Werte), lagen alle Angaben vollständig vor. Die fehlenden Werte wurden durch den Expectation-Maximization-Algorithmus (EM-Algorithmus; Graham, 2009; Lüdtke, Robitzsch, Trautwein & Köller, 2007) geschätzt und ersetzt.

8.5 Analysemethoden der ersten Studie

Neben deskriptiven Analysen wurden zur Hypothesenprüfungen zunächst multivariate Varianzanalysen für die Schülerleistungen sowie die leistungsrelevanten Merkmale durchgeführt und anschließend univariate Varianzanalysen für die einzelnen Konstrukte berechnet. Nach der Analyse der allgemeinen Herkunftseffekte durch Varianzanalysen wurden anschließend orthogonale a priori Kontrastanalysen berechnet (z. B. Abdi & Williams, 2010; Field, 2009; vgl. Tabelle 7).

Tabelle 5

Orthogonale a priori Kontraste nach ethnischer und sozialer Schülerherkunft

	Kontrast 1	Kontrast 2	Produkt <small>(Kontrast 1 x Kontrast 2)</small>
Hoher sozioökonomischer Status; kein Migrationshintergrund	1.0	0.0	0.00
Niedriger sozioökonomischer Status; kein Migrationshintergrund	-0.5	0.5	-0.25
Niedriger sozioökonomischer Status; Migrationshintergrund	-0.5	-0.5	0.25
Summe	0.0	0.0	0.00

Anmerkungen. Kontrast 1 = Hoher vs. niedriger sozioökonomischer Status. Kontrast 2 = Kein Migrationshintergrund vs. Migrationshintergrund.

Dies erfolgte um den Einfluss der ethnischen und sozialen Herkunft getrennt betrachten zu können. Dabei wurden die Schüler einmal entsprechend ihrer sozialen Herkunft und einmal entsprechend ihrer ethnischen Herkunft gewichtet. So wurde im ersten Kontrast der Schüler mit hohem sozioökonomischem Status mit den beiden Schülern mit niedrigem sozioökonomischem Status (unabhängig von der ethnischen Herkunft) verglichen (*Julius vs. Justin und Murat*). Im zweiten Kontrast wurden die beiden Schüler mit niedrigem sozioökonomischem Status je nach ethnischer Herkunft gegenübergestellt (*Justin vs. Murat*).

Um der Frage nach einer positiven oder negativen Urteilsverzerrung nachzugehen, wurden der Notendurchschnitt der Schüler aus den Fallvignetten ($M = 4.67$, rekodiert) als Vergleichsmaßstab genutzt. Die Lehrkraftaspiration sowie die Lehrkrafterwartung an zukünftige Noten wurden von dem tatsächlichen Notendurchschnitt der Schüler in den Hauptfächern abgezogen. Die daraus resultierende Differenz zeigt dabei ob eine positive oder negative Urteilsverzerrung vorliegt. Positive Werte entsprechen einer positiven Urteilsverzerrung und negative Werte einer negativen Urteilsverzerrung. Um zu berechnen ob die Abweichung vom tatsächlichen Notendurchschnitt statistisch bedeutsam ist, wurden Einstichproben t-Tests berechnet.

8.6 Ergebnisse von Studie 1

Deskriptive Befunde

Die Ergebnisse der Mittelwerte (vgl. Tabelle 8) weisen insgesamt auf eher positive Lehrkrafturteile über und -erwartungen für alle Schüler, unabhängig von ihrer Herkunft, hin.

Tabelle 6

Mittelwerte (Standardabweichungen) und bivariate Korrelationen der Einschätzungen der Leistungen und leistungsrelevanten Merkmale (Studie 1)

	<i>M</i> (<i>SD</i>)	(1)	(2)	(3)	(4)
(1) Leistungsaspiration (rekodiert)	5.00 (0.66)	-			
(2) Leistungserwartung (rekodiert)	4.98 (0.56)	.76**	-		
(3) Gymnasialeignung	3.59 (0.71)	.64**	.56**	-	
(4) Anstrengungsbereitschaft	3.86 (0.66)	.56**	.60**	.49**	-
(5) Fähigkeiten	3.71 (0.47)	.68**	.61**	.59**	.76**

Anmerkungen. ** $p < .01$, $N = 237$

Auch für die leistungsrelevanten Merkmale zeigten sich insgesamt positive Lehrkrafturteile. Die Schüler wurden dabei sowohl in ihren Fähigkeiten als auch in ihrer Anstrengungsbereitschaft über dem mittleren Skalenwert eingeschätzt und zudem auch als eher geeignet für einen Gymnasialbesuch angesehen. Der Tabelle 8 sind zudem die bivariaten Korrelationen zwischen den Leistungen und leistungsrelevanten Merkmalen zu entnehmen. Erwartungsgemäß zeigten sich hohe Korrelationen zwischen der Leistungsaspiration und der Leistungserwartung. Beide Leistungsurteile der Lehrkräfte hingen zudem moderat mit der Einschätzung zusammen, ob die Schüler für das Gymnasium geeignet sind sowie mit den Lehrkrafturteilen über die Anstrengungsbereitschaft und die Fähigkeiten der Schüler. Auch die Einschätzung der Anstrengungsbereitschaft korrelierte hoch mit den Fähigkeitsurteilen.

Hypothesenprüfende Analysen⁶⁸

Hypothese 1.1 und Hypothese 1.2

Die grundlegende Fragestellung der vorliegenden Arbeit basiert auf der Annahme, dass sich Lehrkrafturteile und -erwartungen in Abhängigkeit der ethnischen und sozialen Herkunft von Schülern unterscheiden. Dabei wurde davon ausgegangen, dass Lehrkrafturteile und -erwartungen über Leistungen und leistungsrelevante Merkmale für Schüler mit niedrigem sozioökonomischem Status im Vergleich zu Schülern mit hohem sozioökonomischem Status ($H_{1.1}$) und für Schüler mit Migrationshintergrund im Vergleich zu Schülern ohne Migrationshintergrund ($H_{1.2}$) schlechter ausfallen.

Die Mittelwertvergleiche nach Herkunftskombination der Schüler in den Fallvignetten (vgl. Abbildung 12) deuten bereits auf herkunftsassozierte Unterschiede in den Lehrkrafturteilen und -erwartungen für Leistungen und leistungsrelevante Schülermerkmale hin. Zur Prüfung der Hypothese 1.1 und Hypothese 1.2 wurden zunächst multivariate Varianzanalysen berechnet. Hierbei zeigten sich signifikante Haupteffekte der Schülerherkunft auf die Leistungsbeurteilung (Wilks $\Lambda = .820$, $F(12,458) = 3.980$; $p < .001$, $\eta^2 = .09$) sowie für die Einschätzung der leistungsrelevanten Merkmale (Wilks $\Lambda = .859$, $F(6,464) = 6.120$; $p < .001$, $\eta^2 = .07$). Anschließende univariate Analysen erbrachten dabei sowohl für die Leistungsaspiration ($F(2,234) = 14.692$, $p < .001$, $\eta^2 = .11$) als auch für die Leistungserwartung ($F(2,234) = 7.973$, $p < .001$, $\eta^2 = .06$) einen signifikanten Einfluss der Schülerherkunft auf die Lehrkrafturteile. Lehrkräfte schätzen somit die zukünftigen Leistungen der Schüler in Abhängigkeit der Schülerherkunft unterschiedlich ein.

Auch für die leistungsrelevanten Merkmale ergaben die univariaten Varianzanalysen signifikante Unterschiede in allen drei Aspekten: Gymnasialeignung ($F(2,234) = 10.550$, $p < .001$, $\eta^2 = .08$), Anstrengungsbereitschaft ($F(2,234) = 7.540$, $p < .001$, $\eta^2 = .06$) und Fähigkeiten ($F(2,234) = 16.432$, $p < .001$, $\eta^2 = .12$). Diese Befunde zeigen demnach, dass sich Urteile und Erwartungen erfahrener Grundschullehrkräfte für Leistungen und leistungsrelevante Merkmale ebenfalls nach Schülerherkunft in den Fallvignetten unterscheiden.

⁶⁸ Durch die gerichteten Hypothesen werden einseitige Signifikanzniveaus berichtet.

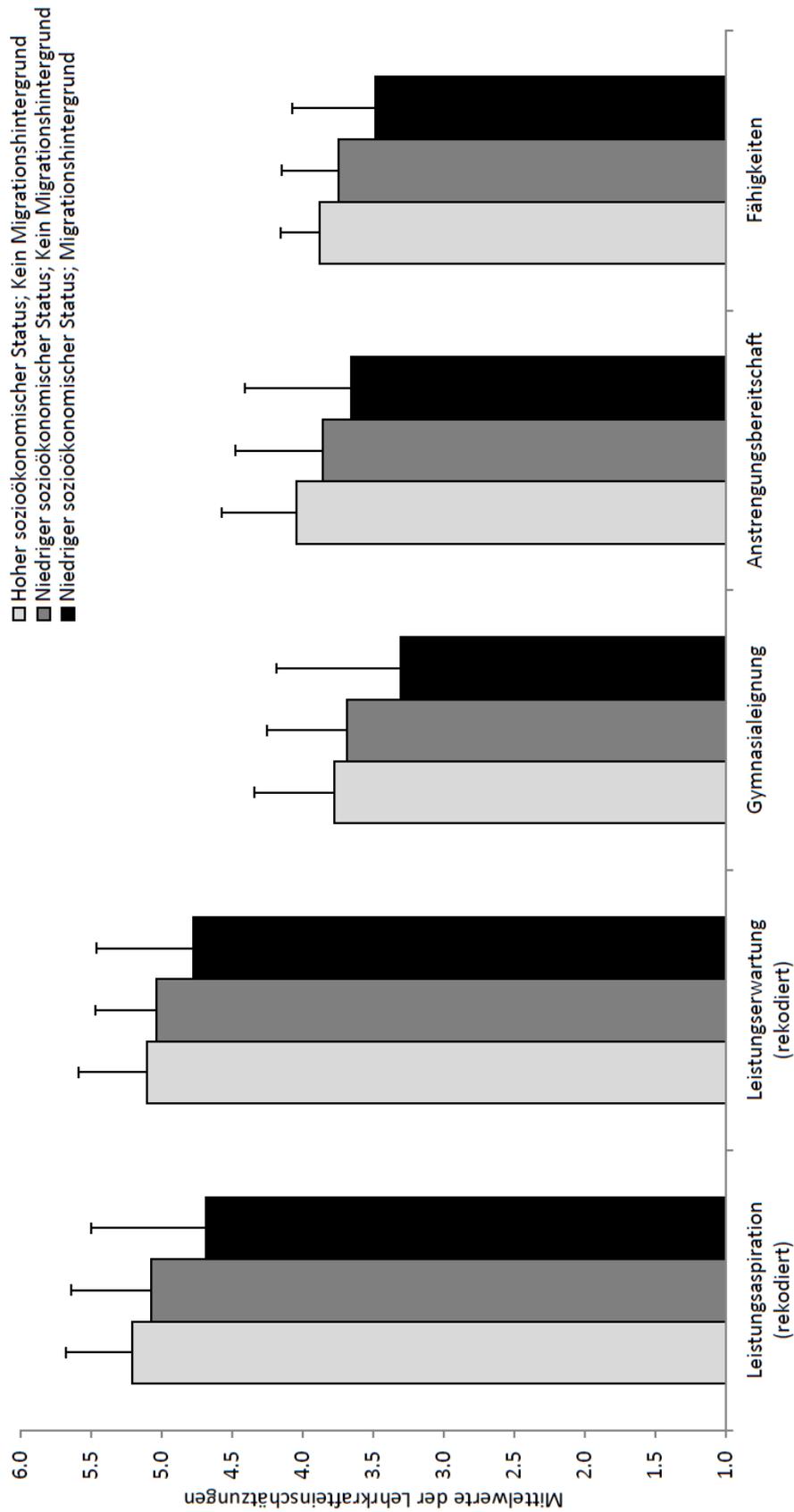


Abbildung 12. Mittelwerte und Standardfehler der Leistungen und leistungsrelevanten Schülermerkmale nach Herkunftscombination in den Fallvignetten

Um die spezifischen Hypothesen getrennt nach ethnischer und sozialer Herkunft zu prüfen, wurden orthogonale a priori Helmert-Kontrastanalysen berechnet. Hierbei zeigte sich sowohl für die Leistungsaspiration als auch für die Leistungserwartung ein signifikanter Unterschied in den Lehrkrafturteilen nach sozialer Herkunft der Schüler (vgl. Tabelle 9). Lehrkräfte hatten somit höhere Erwartungen an den fiktiven Schüler mit hohem sozioökonomischem Status als an die beiden Schüler der Fallvignetten mit niedrigem sozioökonomischem Status. Die Analysen erbrachten ebenfalls einen vergleichbaren Effekt der ethnischen Herkunft auf die Lehrkrafturteile und -erwartungen in Bezug auf die Leistungen der Schüler (vgl. Tabelle 9). Dabei zeigten sich bei vergleichbarem niedrigem sozioökonomischem Status der Schüler in den Fallvignetten für den Schüler ohne Migrationshintergrund höhere Leistungserwartungen und -aspirationen der Lehrkräfte als für den Schüler mit Migrationshintergrund.

Tabelle 7

Ergebnisse der Kontrastanalysen zur Leistungsaspiration und -erwartung (Studie 1)

Kontrast	Kontrastwert	<i>t</i>	<i>r</i> _{Kontrast}
Leistungsaspiration			
Hoher vs. niedriger sozioökonomischer Status	0.327	-3.841***	.24
Migrationshintergrund vs. kein Migrationshintergrund	0.195	-3.840***	.24
Leistungserwartung			
Hoher vs. niedriger sozioökonomischer Status	0.197	-2.692**	.17
Migrationshintergrund vs. kein Migrationshintergrund	0.129	-2.960**	.19

Anmerkungen. *df* = 234, *** *p* < .001. ** *p* < .01, *N* = 237.

Die Kontrastanalysen der leistungsrelevanten Schülermerkmale bestätigten dieses Befundmuster weitestgehend (vgl. Tabelle 10) und erbrachten ebenfalls für alle drei leistungsrelevanten Merkmale signifikante Effekte beider Herkunftsmerkmale. Die Lehrkrafturteile über die Gymnasialeignung, die Anstrengungsbereitschaft sowie die Fähigkeiten fielen für den Schüler mit hohem sozioökonomischem Status besser aus als für die Schüler mit niedrigem sozioökonomischem Status. Aber auch der Migrationshintergrund spielte eine nicht zu vernachlässigende Rolle. So waren die Einschätzungen der Lehrkräfte für die leistungsrelevanten Schülermerkmale durchweg positiver für den Schüler ohne Migrationshintergrund und mit niedrigem sozioökonomischem Status als für den Schüler mit Migrationshintergrund und mit niedrigem sozioökonomischem Status. Die Annahme negativerer Urteile und Erwartungen von Lehrkräften für Schüler mit niedrigem sozioökonomischem Status im Vergleich zu Schülern mit hohem sozioökonomischem Status konnte somit bestätigt werden (*H*_{1.1}). Auch in Bezug auf die ethnische Herkunft der Schüler zeigten sich bei vergleichbarem sozioökonomischem Status erwartungsgemäß negativere Lehrkräfteeinschätzungen der Schüler mit Migrationshintergrund im Vergleich zu Schülern ohne Migrationshintergrund (*H*_{1.2}).

Tabelle 8

Ergebnisse der Kontrastanalysen zu leistungsrelevanten Schülermerkmalen (Studie 1)

Kontrast	Kontrastwert	<i>t</i>	<i>r</i> _{Kontrast}
Gymnasialeignung			
Hoher vs. niedriger sozioökonomischer Status	0.278	3.000**	.19
Migrationshintergrund vs. kein Migrationshintergrund	0.193	3.491***	.22
Anstrengungsbereitschaft			
Hoher vs. niedriger sozioökonomischer Status	0.293	3.377***	.22
Migrationshintergrund vs. kein Migrationshintergrund	0.100	1.931*	.13
Fähigkeiten			
Hoher vs. niedriger sozioökonomischer Status	0.266	4.461***	.28
Migrationshintergrund vs. kein Migrationshintergrund	0.128	3.618***	.23

Anmerkungen. *df* = 234, *** *p* < .001, ** *p* < .01, * *p* < .05, *N* = 237.

Hypothese 1.3 und Hypothese 1.4

Zur Analyse der positiven und negativen Urteilsverzerrung nach Schülerherkunft wurden die Abweichungen der Leistungsaspiration und der Leistungserwartung vom tatsächlichen Notendurchschnitt ($M = 4.67$; rekodiert) berechnet (vgl. Abbildung 13). Für keinen der Schüler war eine negative Abweichung von den tatsächlichen Noten der Fallvignetten zu erkennen. Für den Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status verwiesen die Daten auf eine signifikant positive Verzerrung der Leistungsaspiration ($t(83) = 10.456$, $p < .001$) sowie der Leistungserwartung ($t(83) = 8.294$, $p < .001$). Auch für den Schüler ohne Migrationshintergrund und mit niedrigem sozioökonomischem Status zeigte sich eine positive Verzerrung der Leistungsaspiration ($t(76) = 6.315$, $p < .001$) und der Leistungserwartung ($t(76) = 7.545$, $p < .001$). Lediglich für den Schüler türkischer Herkunft und mit niedrigem sozioökonomischen Status konnte keine signifikante Abweichung der Leistungsaspiration ($t(75) = 0.201$, $p = .84$) und der Leistungserwartung ($t(75) = 1.417$, $p = .16$) vom tatsächlichen Notendurchschnitt der Fallvignetten festgestellt werden. Die Annahme einer negativen Urteilsverzerrung bei Schülern mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status muss somit verworfen werden ($H_{1.3}$). Es zeigt sich im Gegenteil eine starke Orientierung der Lehrkräfte am Notendurchschnitt der Fallvignetten. Für Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status konnte hingegen eine positive Verzerrung der Lehrkrafturteile bestätigt werden ($H_{1.4}$).

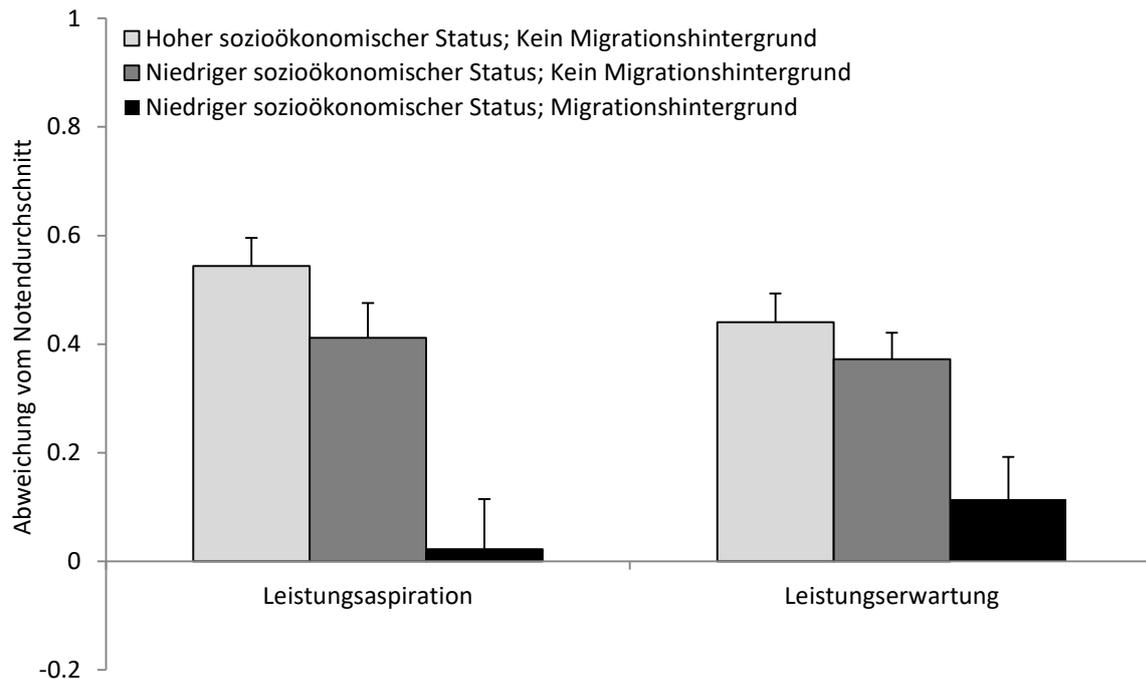


Abbildung 12. Durchschnittliche Abweichung und Standardfehler der Leistungsaspiration und Leistungserwartung vom tatsächlichen Notendurchschnitt nach Herkunftscombination der Fallvignetten

Ergänzende Analysen⁶⁹

Angenommen wird, dass auch Merkmale der Lehrkräfte einen Einfluss auf die Beurteilung von Schüler(inne)n haben, was die große Streuung von Lehrkrafturteilen (z. B. Brimi, 2011) erklären würde. So wurde im Kontext der diagnostischen Kompetenz u. a. der Einfluss der Berufserfahrung diskutiert (z. B. Baumert & Kunter, 2006; Stang & Urhahne, 2016b), jedoch konnten empirische Studien hierzu meist keinen bedeutsamen Einfluss ermitteln (vgl. z. B. Praetorius et al., 2011). Dies wurde bislang jedoch nicht im Zusammenhang mit der Urteilsbildung bei Schüler(inne)n unterschiedlicher Herkunft untersucht. Da u. a. im ADCM (Böhmer et al., 2017) angenommen wird, dass sich die Informationsverarbeitung von Novizen und Expert(inne)en unterscheiden, könnte angenommen werden, dass Lehrkräfte, die bislang wenig Berufserfahrung haben, durch eine integrierende Informationsverarbeitung auch stärker irrelevante Informationen, wie die ethnische und soziale Herkunft von Schüler(inne)n, für ihre Urteile berücksichtigen.

Zunächst wurden multivariate Varianzanalysen berechnet, in denen alle Urteile der Lehrkräfte zu Leistungen und leistungsrelevanten Schülermerkmalen als abhängige Variablen berücksichtigt wurden und die Schülerherkunft sowie die Berufserfahrung der Lehrkräfte (erfasst durch die Anzahl an bisherigen Dienstjahren) als abhängige Variablen hinzugefügt wurden. Da für zwei Versuchspersonen keine Angaben zur Berufserfahrung vorlagen, wurden diese in den ergänzenden Analysen ausgeschlossen. Dabei erbrachten die Analysen weiterhin

⁶⁹ Da keine gerichteten Hypothesen vorlagen, werden zweiseitige Signifikanzniveaus berichtet.

einen signifikanten Haupteffekt der Herkunft auf die Schülerbeurteilung (Wilks $\Lambda = .791$, $F(10,292) = 3.633$; $p < .001$, $\eta^2 = .11$) sowie einen signifikanten Haupteffekt der Berufserfahrung auf die Einschätzung der Schüler (Wilks $\Lambda = .263$, $F(185,729) = 1.218$; $p < .05$, $\eta^2 = .23$). Jedoch erbrachten die Analysen keine signifikante Interaktion der Schülerherkunft und der Berufserfahrung bei der Beurteilung der Schüler in den Fallvignetten (Wilks $\Lambda = .265$, $F(225,731) = 0.996$; $p = .51$, $\eta^2 = .23$). Nachfolgende univariate Varianzanalysen bestätigten dabei, dass bei keinem der Lehrkrafturteile eine signifikante Interaktion zwischen der Schülerherkunft und der Berufserfahrung vorlag. Lediglich bei der Einschätzung der Fähigkeiten zeigte sich ein marginal signifikanter Interaktionseffekt ($F(45,150) = 1.412$, $p = .06$, $\eta^2 = .30$). Dieser deutet in der Tendenz darauf hin, dass unerfahrenere Lehrkräfte den Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status geringere Fähigkeiten zuschreiben, als erfahrenere Lehrkräfte. Hingegen erbachten die univariaten Analysen Effekte der Dienstjahre auf die Beurteilung der Schüler (unabhängig von der Herkunftskombination in den Fallvignetten) in allen Bereichen, bis auf bei der Leistungserwartung ($F(37,150) = 1.208$, $p = .22$, $\eta^2 = .23$). Insgesamt zeigen die Unterschiede nach Dienstjahren, dass Lehrkräfte mit einer längeren Berufserfahrung im Vergleich zu Lehrkräften mit einer geringeren Berufserfahrung durchschnittlich eine höhere Leistungsaspiration haben, die Schüler als geeigneter für den Gymnasialbesuch betrachten sowie ihnen eine höhere Anstrengungsbereitschaft und Fähigkeiten zuschrieben

8.7 Diskussion von Studie 1

Studie 1 ging den Fragen nach ob sich Lehrkrafturteile und -erwartungen für Leistungen sowie leistungsrelevante Merkmale nach unterschiedlicher ethnischer und sozialer Schülerherkunft unterscheiden. Zudem sollte untersucht werden, ob Disparitäten in den Lehrkrafturteilen als negative Urteilsverzerrungen für Schüler mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status zu interpretieren sind und/oder ob eine positive Urteilsverzerrung für Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status identifizierbar ist. Hierzu wurden unter dem Einsatz von Fallvignetten (Halbjahreszeugnisses der vierten Jahrgangsstufe) mit experimenteller Variation der ethnischen und sozialen Herkunft der Schüler im Rahmen einer Online-Studie $N = 237$ Grundschullehrkräften befragt.

Durch die experimentelle Variation der ethnischen und sozialen Herkunft in dieser Studie ist zwar nur eine eingeschränkte externe Validität gegeben, es ist jedoch möglich kausale Rückschlüsse der Herkunftsmerkmale auf die Lehrkrafturteile und -erwartungen zu ziehen. Die Ergebnisse bestätigten dabei bisherige Befunde zu herkunftsassoziierten Disparitäten in Lehrkrafturteilen und zeigen negativere Lehrkrafturteile und -erwartungen an Schüler mit Migrationshintergrund im Vergleich zu Schülern ohne Migrationshintergrund und negativere Urteile und Erwartungen an Schüler mit niedrigem sozioökonomischem Status im Vergleich zu Schülern mit hohem sozioökonomischem Status (z. B. Glock & Krolak-Schwerdt, 2013; Ready & Chu, 2015; Rubie-Davis et al., 2006; Sprietsma, 2013).

Dabei konnten im Unterschied zu bisherigen empirischen Befunden nicht nur herkunftsassoziierte Lehrkrafturteile und -erwartungen in Bezug auf schulische Leistungen und die Gymnasialempfehlung bestätigt werden, sondern zusätzlich herkunftsabhängige Unterschiede bei

Lehrkrafturteilen bzgl. motivationaler Aspekte (Anstrengungsbereitschaft) sowie der Fähigkeiten der Schüler aufgedeckt werden. Auch hier zeigte sich eine negativere Einschätzung der Schüler mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status im Vergleich zu Schülern ohne Migrationshintergrund und hohem sozioökonomischem Status. Hypothese 1.1 und Hypothese 1.2 konnten durch die Analysen entsprechend bestätigt werden. Zudem erbrachte die Berücksichtigung sowohl der ethnischen als auch sozialen Herkunft durch die a priori Helmert-Kontrastanalysen, dass die herkunftsassoziierten Unterschiede in Lehrkrafturteilen und -erwartungen sowohl für Leistungen als auch für leistungsrelevante Schülermerkmale nicht alleine durch die soziale Herkunft von Schülern erklärt werden können. Der Migrationshintergrund der Schüler erwies sich im Vergleich mit Schülern des gleichen niedrigen sozioökonomischen Status weiterhin als bedeutsamer Prädiktor für die Einschätzungen der Lehrkräfte. Dabei zeigte sich in den ergänzenden Analysen zwar ein Unterschied in der generellen Beurteilung der Schüler nach Berufserfahrung der Lehrkräfte, jedoch konnte kein Einfluss der Berufserfahrung auf eine herkunftsassoziierte Beurteilung aufgedeckt werden.

Eine v. a. für den deutschsprachigen Raum neue und bedeutsame Erkenntnis erbrachte der Vergleich der Lehrkraftaspiration und -erwartung mit den tatsächlichen Noten der Fallvignetten. Die Analyse der Leistungserwartung und -aspiration zeigte unter Berücksichtigung der Vorleistungen der Schüler keine direkte Benachteiligung des Schülers mit Migrationshintergrund. Die Erwartungen an zukünftige Leistungen orientierten sich dabei an den tatsächlichen Leistungen des Schülers. Die Befunde dieser Untersuchung konnten keine negative Verzerrung der Lehrkrafturteile für Schüler mit Migrationshintergrund und niedrigem sozioökonomischem Status feststellen. So konnte Hypothese 1.3 zu negativen Urteilsverzerrungen bei Schülern mit Migrationshintergrund nicht bestätigt werden. Die Hypothese 1.4 zur positiven Urteilsverzerrung für Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status bestätigte sich jedoch, was in Einklang mit der Untersuchung von Ready und Chu (2015) steht. Damit lassen sich möglicherweise herkunftsassoziierte Urteils- und Erwartungsunterschiede nicht durch eine direkte Benachteiligung von Schülern mit Migrationshintergrund, sondern eher durch eine Bevorzugung von Kindern ohne Migrationshintergrund und mit hohem sozialem Status erklären. Da sich die akkurate Einschätzung bzw. die Überschätzung aus der Differenz zwischen der Lehrkrafteinschätzung und den tatsächlichen Leistungen der fiktiven Schüler der Fallvignetten handelt, können diese Befunde auch im Sinne der Niveauelemente der diagnostischen Kompetenz interpretiert werden.

Da eher von einer relativen Stabilität der Schülerleistungen auszugehen ist als von einem schnellen Anstieg der Leistungen (vgl. Hülür, Wilhelm & Robitzsch, 2011), ist die Erwartung von Lehrkräften, Schüler würden ihre durchschnittliche Leistung in den Hauptfächern bis zur nächsten Probe signifikant verbessern, eher unrealistisch und spricht für eine Überschätzung der Schüler. Im Gegensatz dazu wäre eine realistische Erwartung, dass sich in diesem kurzen Zeitraum keine substantiellen Leistungsveränderungen abzeichnen und könnte entsprechend als eine akkurate Einschätzung der Lehrkräfte interpretiert werden. Somit sind die insbesondere hohen Erwartungen und Aspirationen für Schüler ohne Migrationshintergrund als Überschätzung zu interpretieren. Die Orientierung der Lehrkräfte an den bisherigen Noten bei Schülern mit Migrationshintergrund und niedrigem Status kann hingegen als akkurate Einschätzung gedeutet werden. Die Befunde stehen dabei im Gegensatz zu den Ergebnissen von Pit-ten Cate

et al. (2016) die eine geringere Akkuratheit bei Schülern der ethnischen Minorität im Vergleich zu Schülern der luxemburgischen Majorität stehen. Jedoch sind die vorliegenden Befunde der Studie konform mit den Ergebnissen von Kaiser et al. (2016), die ebenfalls eine akkuratere Beurteilung der ethnischen Minorität aufzeigen konnten, jedoch interpretieren sie ihre Ergebnisse als Effekte der Salienz durch den Minoritätenstatus der Schüler mit Migrationshintergrund in ihren Experimenten. Dies kann für die vorliegende Studie nicht bestätigt werden, da durch das between-subject-Design, in dem jede Lehrkraft ausschließlich eine Fallvignette mit einer der drei Herkunftskombinationen las, kein Minoritätenstatus einer der ethnischen und sozialen Herkunftskombination entstand. Da auch der Schüler ohne Migrationshintergrund und mit niedrigem sozioökonomischem Status leicht überschätzt wurde, können die vorliegenden Befunde als Effekte des Migrationshintergrunds interpretiert werden.

In Anlehnung an das Kontinuummodell von Fiske und Neuberg (1990), könnten die positiven Urteilsverzerrungen bei Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status ein Indikator für eine eher automatisierte Informationsverarbeitung sein. Erklären ließe sich dies durch positive Einstellungen und Stereotypen dieser Schülergruppe gegenüber, die automatisiert zu positiveren Urteilen führen könnten. Bei Schülern mit Migrationshintergrund deuten die eher akkuraten Urteile stärker auf einen kontrollierten Informationsverarbeitungsprozess hin. Lehrkräften sind sich oftmals der bestehenden herkunftsassoziierten Disparitäten im Bildungsverlauf und -erfolg bewusst und sind hierdurch womöglich motiviert, sich nicht durch negative Assoziationen leiten zu lassen. Entsprechend würden sie Informationen kontrollierter verarbeiten und sich damit auch in ihren Urteilen stärker an den tatsächlichen Leistungen orientieren. Dies müsste jedoch durch vertiefte Analysen der kognitiven Prozesse bei der Erwartungs- und Urteilsbildung von Lehrkräften systematisch untersucht werden und wird in Studie 2 analysiert.

Zwar ist die Aussagekraft durch die Beschränkung auf männliche Grundschüler und einen ausschließlich türkischen Migrationshintergrund auf diese Merkmale begrenzt, dennoch können die Befunde keine Benachteiligung für Grundschüler mit türkischem Migrationshintergrund zeigen, sondern sie verweisen auf eine Bevorzugung von Schülern ohne Migrationshintergrund, v. a. mit hohem sozioökonomischem Status. In zukünftigen Studien sollte jedoch auch die Konzeption einer Fallvignette mit Migrationshintergrund und hohem sozioökonomischem Status berücksichtigt werden. Dies war durch die hohe Konfundierung der ethnischen und sozialen Herkunft sowie die Aktivierung herkunftsassoziiertes Stereotype über Namen im Design der Vignetten dieser Studie nicht möglich. Hierdurch konnte jedoch vertrautes Stimulusmaterial für die Lehrkräfte entwickelt werden und die Schülerherkunft nur durch den Vornamen und ohne zusätzliche Informationen manipuliert werden.

Kritisch zu diskutieren ist jedoch auch die Stichprobe. Bereits der vergleichsweise geringe Anteil weiblicher Grundschullehrkräfte in der Stichprobe verweist darauf, dass es sich hierbei um keine repräsentative Stichprobe handeln kann. Bedingt durch die Rekrutierung in Lehrkraftforen im Internet und die freiwillige Teilnahme der Versuchspersonen, stellt sich die Frage, ob hierdurch nicht bereits deutlich engagiertere Lehrkräfte angesprochen wurden. Dies könnte die akkuratere Einschätzung des Schülers mit Migrationshintergrund und niedrigem sozioökonomischem Status erklären. Angenommen werden könnte in diesem Zusammenhang, dass engagiertere Lehrkräfte sich womöglich bereits bewusster mit herkunftsassoziierten Disparitä-

ten auseinandergesetzt hatten und entsprechend dieser Thematik höhere Aufmerksamkeit schenken, was sich auch auf eine akkurate Urteilsbildung auswirken könnte. Insbesondere Faktoren der Aufmerksamkeit und der Motivation sind dabei maßgebliche Einflussfaktoren auf den Prozess sozialer Informationsverarbeitung (vgl. Kapitel 5). Daher sollten die zugrundeliegenden kognitiven Prozesse bei der Urteilsbildung von Lehrkräften vertieft untersucht werden, um den Einfluss herkunftsassoziierter Stereotype auf diesen Prozess klären zu können. Dabei sollte auch die Problematik einer möglichen verzerrten Stichprobe entgegengewirkt werden und selektive Mechanismen wie eingeschränkte Kommunikationswege durch spezifische Rekrutierungsorte oder eine Vorsortierung nach erhöhter Motivation durch eine freiwillige Teilnahme ausgeschlossen werden. Insbesondere die Hypothese unterschiedlicher kognitiver Prozesse bei der Verarbeitung von Schülerinformationen in Abhängigkeit der sozialen und ethischen Herkunft wird in der zweiten Studie der vorliegenden Arbeit nachgegangen.

9. Studie 2: Herkunftsabhängige Informationsverarbeitungsprozesse von Lehramtsstudierenden

Basierend auf dem Kontinuum-Modell der Eindrucksbildung (Fiske & Neuberg, 1990) und den Befunden aus Studie 1 (vgl. Abschnitt 8.6) stellt sich die Frage, ob unterschiedliche Informationsverarbeitungsprozesse im Zusammenhang mit einer herkunftsassoziierten Urteilsbildung stehen (Fragestellung 2.1). Entsprechend war das Ziel dieser Studie, Hinweise auf unterschiedliche kognitive Prozesse bei der Beurteilung von Schülern unterschiedlicher ethnischer und sozialer Herkunft zu identifizieren. Durch den engen Zusammenhang psychologischer und physiologischer Prozesse wurden im Rahmen einer experimentellen Laborstudie okulomotorische Daten mit einem Eye-Tracker erfasst, um hieraus Rückschlüsse auf die Informationsverarbeitung bei der Schülerbeurteilung ziehen zu können. Zudem wurden implizite Einstellungen zu Personen mit und ohne Migrationshintergrund erfasst, um zu klären, ob implizite Einstellungen die Informationsverarbeitung von Lehramtsstudierenden bei Schülern unterschiedlicher Herkunft beeinflussen (Fragestellung 2.2).

Die Erfassung okulomotorischer Daten als Indikatoren für kognitive Prozesse bei der herkunftsassoziierten Urteils- und Erwartungsbildung von Lehrkräften stellt einen innovativen methodischen Zugang dar. Entsprechend fokussiert der folgende Abschnitt zunächst die physiologischen Grundlagen der Okulomotorik. Dabei erfolgt die Darstellung theoretischer Ansätze und empirischer Befunde zu Zusammenhängen der unterschiedlichen Augenbewegungsarten mit kognitiven Prozessen, was für das Verständnis der methodischen Besonderheiten dieser Studie erforderlich ist.

9.1 Eye-Tracking als Methode zur Erfassung kognitiver Prozesse

„Da nach systemtheoretischen Gesichtspunkten jede auch noch so einfache Verhaltensweise gleichzeitig ... als physiologische Veränderung und als peripherer Ausdruck eines komplexen neurophysiologischen Musters im zentralen Nervensystem angesehen werden kann, sollten auch psychologische Prozesse als emergente Eigenschaften von Mustern neuronaler Aktivität verstanden werden“ (Stemmler, 2001, S. 3). Die Berücksichtigung der engen Verknüpfung und wechselseitigen Beeinflussung physischer und psychischer Prozesse (vgl. Stemmler, 2001) erweist sich als vielversprechender Ansatz der psychologischen Forschung, um einige Einschränkungen von z. B. Selbstberichtsdaten zu überwinden (vgl. z. B. Cacioppo, Lorig, Nusbaum & Berntson, 2004).

Dabei können je nach Untersuchungsziel und Forschungsfrage unterschiedliche physiologische Messmethoden in psychologischen Untersuchungen eingesetzt werden (Überblick z. B. bei Rösler, 1998), so z. B. die Hautleitfähigkeit als Indikator für eine Aktivierung des Nervensystems (z. B. Fowles, Kochanska & Murray, 2000) oder die Elektroenzephalografie (EEG) zur

Erfassung der Aktivität spezifischer Hirnregionen (z. B. Anokhin & Vogel, 1996).^{70, 71} Da in den Untersuchungen der vorliegenden Arbeit mit Fallvignetten in Textform gearbeitet wurde, stellt sich die Frage, welcher methodische Zugang zur Erfassung kognitiver Prozesse bei der Bearbeitung dieses Stimulusmaterials am besten geeignet ist. Krolak-Schwerdt und Kneer (2006) betonen, dass die Erhebung von Blickbewegungsdaten einen lohnenden methodischen Zugang zur Erfassung von Informationsverarbeitungsprozessen bei der Personenbeurteilung darstellt (vgl. auch Krolak-Schwerdt & Wintermantel, 2004). Verglichen mit der Erfassung der Lesezeiten über die Präsentationsdauer eines Stimulus liefern Blickbewegungsdaten exaktere Informationen über die tatsächliche Lesezeit und lassen entsprechend mehr Rückschlüsse auf die Informationsverarbeitung zu. Die Erfassung der Okulomotorik, insbesondere der Analyse der Blickbewegung von Personen als Indikatoren für psychische Prozesse, stellt dabei für die psychologische Forschung nicht erst seit den neuesten technischen Möglichkeiten einen ertragreichen Ansatz dar.

Bereits im späten 19. Jahrhundert entstanden die ersten mechanischen Eye-Tracker (z. B. Delabarre, 1898; Überblick bei Holmqvist et al., 2011). So wurden z. B. Kontaktlinsen mit metallischen Zeigern eingesetzt (z. B. Huey, 1898), die jedoch entsprechend unangenehm für die Versuchspersonen waren (vgl. Bartl-Pokorny et al., 2013). Dodge und Cline (1901) nutzten erstmals die Fotografie der Lichtreflexion der *fovea centralis* (vgl. Abbildung 14) zur Erfassung der Augenbewegungen. Diese nichtinvasive Methode wurde im Verlauf der technischen Weiterentwicklungen zu einer der führenden Technologien zur Aufzeichnung von Augenbewegungen (vgl. Holmqvist et al., 2011). Mittlerweile dominieren nichtinvasive Eye-Tracking-Systeme, die die Reflexion von Infrarotlicht – das von dem Eye-Tracking-Gerät ausgestrahlt wird – der Hornhaut (*cornea*; vgl. Abbildung 14) erfassen (Holmqvist et al., 2011). Infrarotlicht wird dabei aus zwei zentralen Gründen eingesetzt (Holmqvist et al., 2011): Zum einen werden Artefakte natürlicher Reflexionen anderer Lichtquellen verhindert, zum anderen ist Infrarotlicht für das menschliche Auge nicht sichtbar und führt somit zu keiner Ablenkung oder Irritation.

Aktuell lassen sich drei bedeutsame technologische Systeme zur Erfassung der Okulomotorik mit Infrarottechnologie unterscheiden (vgl. u. a. Bartl-Pokorny et al., 2013): (1) Der Kopf der Versuchsperson wird am Eye-Tracker fixiert (*tower-mounted*). (2) Der Eye-Tracker wird am Kopf der Versuchsperson z. B. durch eine Brille angebracht (*head-mounted*). (3) Die Befestigung des Eye-Trackers am Bildschirm, wodurch eine Berührung des Kopfes vermieden werden kann (*remote*). Da Augenbewegungen physiologische Prozesse sind, werden im folgenden Abschnitt zunächst die entsprechenden Grundlagen zusammengefasst, bevor auf psychologische Aspekte der Augenbewegungen eingegangen wird.

⁷⁰ Fowles et al. (2000) konnten in ihrer Untersuchung Zusammenhänge der Hautleitfähigkeit und der Ängstlichkeit von Kindern feststellen.

⁷¹ Anokhin und Vogel (1996) untersuchten mittels EEG Hirnaktivitäten, die im Zusammenhang mit kognitiven Fähigkeiten stehen, und konnten Korrelate zu der Leistung von Versuchspersonen in Intelligenztests feststellen.

9.1.1 Physiologische Grundlagen der Okulomotorik

Die Okulomotorik lässt sich in drei wesentliche Bereiche untergliedern (Galley, 2001): (1) die Blickbewegung, (2) den Lidschlag und (3) die Pupillenmotorik. Diesen unterschiedlichen Augenbewegungen liegen differenzielle physische Prozesse zu Grunde, die eng an den anatomischen Aufbau des Auges geknüpft sind. Wie jede Bewegung des menschlichen Körpers, werden auch Augenbewegungen durch spezifische Muskeln bewirkt, die durch Impulse des Nervensystems aktiviert oder gehemmt werden. Abbildung 14 zeigt die seitliche Ansicht des Auges und die für die unterschiedlichen Augenbewegungen zuständigen Muskeln.

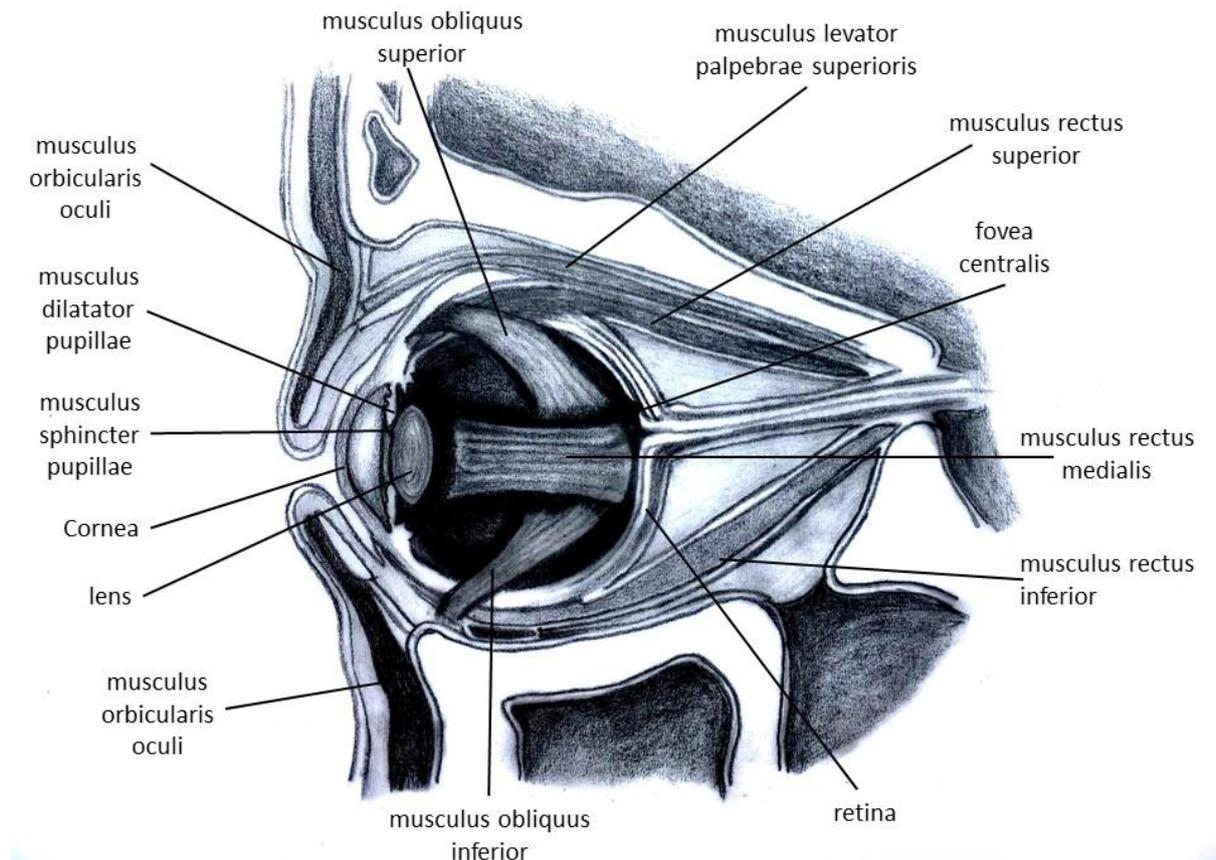


Abbildung 13. Seitliche Ansicht des Auges und der Augenmuskulatur (Matt & Tobisch, 2017 adaptiert nach Drake, Vogl, Mitchell, Tibbitts & Richardson, 2009; Holmqvist et al., 2011; Maidowsky, 2003)

Dabei spielt insbesondere für den Lidschlag und die Pupillenmotorik das autonome Nervensystem, in seiner Untergliederung in sympathische und parasympathische Anteile (Birbaumer & Schmidt, 2010; Gramann & Schandry, 2009), eine besondere Rolle. Sympathische und parasympathische Anteile des autonomen Nervensystems agieren überwiegend antagonistisch (Paulsen & Waschke, 2017). Eine Aktivierung des sympathischen Anteils führt z. B. zu einer erhöhten Herzfrequenz, eine Aktivierung des parasympathischen Anteils dagegen zu einer Abnahme der Herzfrequenz (Birbaumer & Schmidt, 2010). Auch eine vermehrte Schweißbildung (Birbaumer & Schmidt, 2010) – die Grundlage für die Erfassung der Hautleit-

fähigkeit als Indikator für Erregung ist (z. B. Papastefanou, 2013; Schandry, 2008) – erfolgt durch die Aktivierung des sympathischen Anteils des autonomen Nervensystems. Eine Aktivierung des sympathischen Anteils des Nervensystems ist i. d. R. für eine erhöhte Leistungsfähigkeit verantwortlich, eine Aktivierung des parasympathischen Anteils hingegen eher für eine Dämpfung physischer Reaktionen (Gramann & Schandry, 2009).

Blickbewegung

Blickbewegungen sind Augenbewegungen mit einem Objektbezug (Zimmermann, 2013). Für die visuelle Wahrnehmung von Objekten bzw. allgemein visueller Reize ist dabei die *fovea centralis* (der schärfste Punkt des Sehens) von Bedeutung, die sich auf der Netzhaut (*retina*) befindet (vgl. Abadi, Clement & Gowen, 2003). Soll ein Objekt visuell fixiert werden, müssen entsprechend Kopf- und Augenbewegungen so ausgerichtet sein, dass das Objekt genau auf der *fovea centralis* abgebildet wird. Zunächst trifft eine große Anzahl visueller Reize auf der *retina* auf, diese werden jedoch bereits durch die *fovea centralis* eingegrenzt (Harris & Jenkin, 2001). Zur Verarbeitung der Informationen, die auf der *fovea centralis* auftreffen, wird eine gewisse Zeit benötigt, um die weiterhin große Anzahl an Reizen aufzunehmen.

Die Bewegung des Auges wird durch sechs äußere Augenmuskeln (Heben der Sehachse: *musculus rectus superior*, *musculus obliquus inferior*; Senken der Sehachse: *musculus rectus inferior*, *musculus obliquus superior*; Abduktion nach links: *musculus rectus medialis*; Abduktion nach rechts: *musculus rectus lateralis*) gesteuert (Drake et al., 2009; Galley, 2001; vgl. Abbildung 14). Je nach Geschwindigkeit, Amplitude und Dauer lassen sich verschiedene Blickbewegungen unterscheiden (Überblick bei Galley, 2001; Heyden, 1979; Holmqvist et al., 2011). Die längste Dauer weisen dabei Fixationen auf, sie beschreiben den Zustand in dem sich das Auge in einem bestimmten Zeitabschnitt nicht bzw. kaum bewegt (durchschnittliche Dauer: 200 - 300 ms).⁷² Fixationen ermöglichen die visuelle Wahrnehmung, da hierdurch eine Abbildung der Umwelt auf der *fovea centralis* entsteht. Während einer Fixation ist das Auge jedoch nicht vollständig in Ruhe, sondern es bestehen Mikrobewegungen, auch Intrafixationsbewegungen genannt (Holmqvist et al., 2011).⁷³ Die Bewegung des Auges von einer Fixation zur nächsten Fixation wird als Sakkade bezeichnet. Sie ist die schnellstmögliche Bewegung des menschlichen Körpers (Geschwindigkeit: 30 - 500°/s; Amplitude: 4 - 20°; Dauer: 30 - 80 ms) (Gramann & Schandry, 2009; Holmqvist et al., 2011).⁷⁴

⁷² Dabei ist die Fixationsdauer von der jeweiligen Situation abhängig und unterscheidet sich z. B. beim Lesen von Texten oder bei der Suche von Objekten stark (vgl. Zimmermann, 2014).

⁷³ Intrafixationsbewegungen: Der Tremor (auch physiologischer Nystagmus genannt) ist eine oszillierende Bewegung des Auges, die auch „Augenzittern“ genannt wird (Plange, 2017, S. 359). Er hat eine Frequenz von 90 Hz bei einer Amplitude von < 1'. Drifts sind dagegen langsamere Augenbewegungen (Geschwindigkeit: 6-25°/s; Amplitude: 1-60'; Dauer: 200-1000 ms), die den Blick vom Fixationszentrum ablenken und durch Mikrosakkaden (Geschwindigkeit: 15-50°/s; Amplitude: 10-40'; Dauer: 10-30 ms) wieder in die Ursprungsposition zurückgeführt werden (Holmqvist et al., 2011).

⁷⁴ Sakkaden und Mikrobewegungen des Auges lassen sich wegen ihrer hohen Geschwindigkeit nur durch eine entsprechend hohe Abtastrate des Eye-Trackers erfassen.

Lidschlag

Der Lidschlag, als „kurzzeitiges Verschließen des Auges durch die Augenlider“ (Galley, 2001, S. 280), erfüllt die zentrale Funktion der Benetzung der Augen mit Tränenflüssigkeit. Zudem erfüllt der reflektorische Lidschlag eine Schutzfunktion vor, z.B. unerwarteten Berührungen der Wimpern oder Augen durch mechanische oder taktile Reize (vgl. Galley, 2001). Aber auch bei besonders großen Sakkaden und/oder Kopfbewegungen kommt es häufig zu einem Lidschlag, da dieser auch eine ausgleichende Funktion bei der visuellen Reizaufnahme erfüllt (Galley, 2001). Der Lidschlag wird durch die äußeren Augenmuskeln (Lidschließer: *musculus orbicularis oculi*; Lidheber: *musculus levator palpebrae superioris*) gesteuert (Drake et al., 2009; Galley, 2001; vgl. Abbildung 14). Diese Muskeln werden durch die parasymphatisch gespeisten Nerven *nervus oculomotorius* und *nervus facialis* innerviert (Pschyrembel, 1997).⁷⁵ Der *nervus oculomotorius* innerviert den *musculus levator palpebrae* zum Heben des Lids und der *nervus facialis* innerviert den *musculus orbicularis oculi* zum Lidschluss/-schlag (Schünke, Schulte, Schumacher, Voll & Wesker, 2012; Pschyrembel, 1997).⁷⁶

Pupillenmotorik

Zunächst dient die Erweiterung (*Mydriasis*) und Verengung (*Miosis*) der Pupille der Regulation des Lichteinfalls (z. B. Galley, 2001; Gramann & Schandry, 2009). Bedingt durch den *konsensuellen Pupillenreflex* reagieren beide Augen bei Lichtveränderung eines Auges, wobei sich die Pupillenreaktion nach dem Auge mit dem stärkeren Lichteinfall richtet (Gramann & Schandry, 2009; Heyden, 1979).⁷⁷ Aber auch bei der Fixation naher Objekte kommt es zur Verengung der Pupille, um ein scharfes Bild zu ermöglichen (Galley, 2001). Die Öffnung der Pupille ist durch die Kontraktion zweier oppositioneller Muskeln bedingt (Beatty & Lucero-Wagoner, 2000). Gespeist durch parasymphatische Anteile des Nervensystems (*nervus oculomotorius*), sorgt der *musculus sphincter pupillae* für eine Verengung und der durch sympathische Anteile des Nervensystems (*nervus sympathikus*) innervierte *musculus dilatator pupillae* für eine Erweiterung der Pupille (Birbaumer & Schmidt, 2010; Drake et al., 2009; Galley, 2001; Gramann & Schandry, 2009; Heyden, 1979; Plange, 2017; Pschyrembel, 1997; vgl. Abbildung 14). Maßgeblicher Neurotransmitter für die Miosis ist dabei das Acetylcholin, für die Mydriasis das Noradrenalin (Plange, 2017).⁷⁸

⁷⁵ Diese Nerven entspringen direkt dem Gehirn (Galley, 2001; Pschyrembel, 1997).

⁷⁶ Für die Erfassung der Lidschlagrate stehen weitere Methoden zur Verfügung, z. B. Erfassung der elektrischen Spannung zwischen Hornhaut und Netzhaut (Elektrookulogramm) oder des Muskeltonus der Lidheber- und Lidschlussmuskulatur (Elektromyogramm) (Überblick bei Meinold, 2005).

⁷⁷ Der Pupillendurchmesser kann ca. 1.5-9.0 mm betragen (Gramann & Schandry, 2009; Heyden, 1979).

⁷⁸ Acetylcholin führt zu einer Senkung, Noradrenalin zu einer Erhöhung des Blutdrucks (Pschyrembel, 1997).

9.1.2 Psychologische Grundlagen der Okulomotorik

Für die Analyse psychischer Prozesse zeigen sich in der bisherigen Forschung fruchtbare Ansätze bei der Erfassung der Blickbewegung, des Lidschlags sowie der Pupillenmotorik. Im Folgenden werden die Zusammenhänge der jeweiligen Augenbewegungen mit psychischen Prozessen in einem kurzen Überblick dargestellt.

Blickbewegung

Während Sakkaden ruckartige, sprunghafte und sehr schnelle Bewegungen sind, die zwischen Fixationen auftreten, sind Fixationen durch die visuelle Fokussierung auf ein Objekt ein Indikator für psychische Aufmerksamkeit. Dabei wird nach der *Eye-Mind-Hypothese* von Just und Carpenter (1980) davon ausgegangen, dass während einer Sakkade keine Informationen aufgenommen werden können und dagegen eine Verknüpfung zwischen visueller Fixation und kognitiver Informationsverarbeitung besteht. Harris und Jenkin (2001) vermuten, dass die Fixation so lange auf einem Objekt verbleibt, bis alle nötigen Informationen aufgenommen sind. Humphreys und Bruce (1989) gehen ebenfalls davon aus, dass in den meisten Fällen das Objekt der visuellen Fixation Aufmerksamkeit erhält und damit auch zumindest ein gewisser Grad der Informationsverarbeitung stattfindet. Jedoch werden nicht alle Aufmerksamkeitsobjekte fixiert und nicht alle Objekte, die fixiert werden, erhalten Aufmerksamkeit (vgl. Humphreys & Bruce, 1989).

Die Eye-Mind-Hypothese ist u. a. im Bereich der Leseforschung nicht unumstritten (z. B. McConkie & Yang, 2003). Zwar zeigt sich in einigen Untersuchungen, dass eine längere Fixationsdauer z. B. bei Rechtschreibfehlern auftritt (Zola, 1984), was für eine erhöhte Aufmerksamkeit sprechen würde. Jedoch ist ein direkter Zusammenhang von Fixationsdauer und kognitiver Verarbeitung insbesondere bei komplexen Aufgaben wie Lesen und Textverstehen nicht eindeutig nachweisbar (McConkie & Yang, 2003). Auch konnten empirischen Studien die Erkenntnis erbringen, dass Aufmerksamkeit auch auf etwas gerichtet sein kann, ohne dass die Blickbewegung sich an diese Stelle begibt (*covert attention*; Überblick bei Findlay, 2003). Zudem geht Morrison (1984) in seinem *parallel-programming-model* davon aus, dass während einer Fixation auf einem Wort zwar zunächst die Aufmerksamkeit auf diesem Wort liegt, jedoch bereits nach einer kurzen Zeit die nächste Sakkade programmiert wird, um ein folgendes Wort zu fixieren, das durch den *parafovealen Bereich* (Bereich um die fovea centralis) wahrgenommen wurde (vgl. auch Feuß, 2013). Dies würde bedeuten, dass eher automatisierte kognitive Prozesse während einer Fixation ablaufen, die nicht direkt mit der Fixation des Objekts in Zusammenhang stehen, sondern eine präaktionale Programmierung der folgenden Sakkaden und Fixationen beinhalten. Somit ist zwar nicht von einer exakten Übereinstimmung von Fixation und Aufmerksamkeit auszugehen, jedoch sind Fixationen insbesondere beim Lesen als Voraussetzung zum Verstehen eines Textes und der entsprechenden kognitiven Verarbeitung anzusehen, da sie die visuelle Wahrnehmung der Wörter überhaupt erst ermöglichen.

Fixationen werden dabei nicht nur im Kontext der Leseforschung untersucht. Auch im Bereich der Vorurteilsforschung gibt es erste Studien, die die Fixationsdauer auf vorurteilsbehaftetem Stimulusmaterial berücksichtigen (z. B. Kreiner, Sturt & Garrod, 2008). Bean et al. (2012) untersuchten mit Hilfe eines Eye-Trackers, ob sich Personen mit einer hohen ex-

ternalen Motivation zu vorurteilsfreiem Verhalten, im Gegensatz zu Personen mit einer gering ausgeprägten externalen Motivation zu vorurteilsfreiem Verhalten, in ihrer visuellen Aufmerksamkeit bei der Wiedererkennung von Bildern europäischstämmiger und afroamerikanischer Gesichter (die sie aus einer vorherigen Übungsphase kannten) unterscheiden. Theoretische Grundlage für die Annahme bildet das Aufmerksamkeits-Vermeidungs-Muster (*vigilance avoidance pattern*).⁷⁹ Bean et al. (2012) konnten nachweisen, dass bei Personen mit hoher Motivation zu vorurteilsfreiem Verhalten zunächst eine erhöhte Aufmerksamkeit auf den afroamerikanischen Gesichtern lag, wobei anschließend vermieden wurde, dieses Gesicht anzusehen. Das Muster der visuellen Aufmerksamkeit leistet somit einen Beitrag zur Aufklärung herkunftsassoziierter Informationsverarbeitung.

Lidschlagrate (Anzahl der Lidschläge pro Zeiteinheit)

In Experimenten mit Rhesusaffen konnte der Lidschlag durch Mikrostimulationen des *posterioren Parietalkortex*, der u. a. für Aufmerksamkeit verantwortlich ist (Constantinidis, Bucci & Rugg, 2013), ausgelöst werden (Shibutani, Sakata & Hyvärinen, 1984). Dies legt eine physiologische Verknüpfung des Lidschlags mit wahrnehmungsspezifischen Gehirnregionen bei Affen nahe. Stern, Walrath und Goldstein (1984) postulierten bereits in den 1980er Jahren, dass der Lidschlag im Zusammenhang mit kognitiven Prozessen steht. Dabei spielt der Neurotransmitter Dopamin eine wichtige Rolle zur Erklärung der Lidschlagrate (Galley, 2001). Karson (1983) berichtet, dass ein erhöhter Dopaminspiegel bei Affen zu einer erhöhten Lidschlagrate und ein niedriger Dopaminspiegel zu einer Senkung der Lidschlagrate führt. Dies lässt sich durch die aktivierende Wirkung von Dopamin erklären (z. B. Salamone, 2009). Zudem zeigen sich Korrelationen des Lidschlags mit der Hautleitfähigkeit und der Atmung, eine ansteigende Lidschlagrate kann somit im Zusammenhang mit anderen physiologischen Messungen in den Zusammenhang mit einem Erregungslevel des Nervensystems gebracht werden (Überblick bei Meinold, 2005).⁸⁰

Neben der Untersuchung von physiologischen Korrelaten, die Rückschlüsse auf die Aktiviertheit des Nervensystems und den Lidschlag zulassen, wurden bereits einige Untersuchungen durchgeführt, die sich mit der Verknüpfung der Lidschlagrate mit psychologischen Konstrukten wie z. B. kognitiver Aktivierung oder kognitiven Fähigkeiten beschäftigten.⁸¹ Fukuda (2001) konnte in zwei experimentellen Studien zeigen, dass bei der visuellen Präsentation eines Stimulus zunächst eine Hemmung des Lidschlags erfolgt, wobei jedoch die

⁷⁹ Das Aufmerksamkeits-Vermeidungs-Muster beschreibt, dass bei negativen Emotionen wie Angst die Aufmerksamkeit zunächst auf den angstausslösenden Stimulus gerichtet wird (*vigilance*), die visuelle Aufmerksamkeit dann jedoch im Verlauf der Zeit vom Stimulus abgelenkt wird (*avoidance*) (u. a. In-Albon & Schneider, 2012).

⁸⁰ Eine Liderweiterung ist zudem ein Indiz für emotionale Erregung z. B. bei Freude, Angst oder Überraschung (u. a. Galley, 2001; Gramann & Schandry, 2009).

⁸¹ Ladas, Frantzidis, Bamidis und Vivas (2014) konnten für eine klinische Stichprobe sowie der Kontrollgruppe einen negativen Zusammenhang zwischen Lidschlagrate und kognitiven Fähigkeiten feststellen. Eine vergleichsweise erhöhte Lidschlagrate stellt somit einen Indikator für die kognitiven Leistungen dar. Ob die Lidschlagrate jedoch erhöht ist, weil Personen mit geringeren kognitiven Leistungen weniger kognitive Anstrengung bei einem Fähigkeitstest betreiben oder gar im Gegenteil eine hohe kognitive Beanspruchung vorherrscht, die jedoch erfolglos bleibt, ist hierdurch weiterhin ungeklärt.

Lidschlagrate anschließend deutlich ansteigt. Meinold (2005) geht davon aus, dass unterschiedliche Veränderungen der Lidschlagrate in Abhängigkeit des dargebotenen Reizes durch „unterschiedliche bzw. modifizierte kognitive Steuerungsmechanismen im Rahmen situationsspezifischer erforderlicher Aufmerksamkeit“ (S. 78) zu erklären sind. Die Abnahme der Lidschlagrate bei visuellen Reizen wird durch eine potentielle Störung der visuellen Informationsaufnahme durch häufige Lidschlüsse erklärt (Meinold, 2005) und ließ sich in unterschiedlichen empirischen Untersuchungen nachweisen (z. B. Brookings, Wilson & Swain, 1996). So wird die visuelle Reizaufnahme für ca. 150 ms unterbrochen, was bei Aufgabenstellungen mit visuellen Reizen eine geringere Lidschlagrate erklärt (Galley, 2001). Insbesondere beim Lesen von Texten als visueller Reizdarbietung wird darauf hingewiesen, dass Lidschläge an Stellen des potentiell geringsten Informationsverlusts (z. B. beim Sprung in eine neue Zeile oder beim Seitenwechsel) gesetzt werden (Meinold, 2005; Orchard & Stern, 1991). Orchard und Stern (1991) bekräftigen dabei die Annahme, dass Lidschläge an Prozesse der Wahrnehmung und kognitiven Informationsverarbeitung geknüpft sind.

Die Vermutung, dass eine Erregung des Nervensystems direkt durch die Lidschlagrate messbar wäre und damit einen eindeutigen Indikator für z. B. emotionale oder kognitive Aktivierung liefern würde, ließ sich jedoch nicht halten (vgl. Meinold, 2005). Dies liegt u. a. daran, dass das parasympathische Nervensystem nicht nur als passiver Gegenspieler des Sympathikus zu verstehen ist, sondern auch eigenständige Hemmungsprozesse einleiten kann (vgl. z. B. Meinold, 2005; Porges, 2001). Cranach, Schmidt und Vogel (1969) konnten Hinweise für eine erhöhte Lidschlagrate bei erhöhter Aktivierung durch spezifische Aufgabenstellungen in sozialen Situationen finden, jedoch ergaben sich heterogene Befunde in nicht sozialen Versuchsbedingungen. Auch bei emotionalen Belastungen – die mit einer Erregung des sympathischen Nervensystems zusammenhängen – ist ein Anstieg der Lidschlagrate zu verzeichnen. Ponder und Kennedy (1927) berichten von einer erhöhten Lidschlagrate bei Angst, Aufregung, Stress oder erhöhter Aktivierung. Sie vermuten jedoch, dass mentale Aktivierung nicht in jeder Form gleiche Auswirkungen auf die Lidschlagrate haben könnte (Ponder & Kennedy, 1927). So zeigte sich z. B. erwartungskonform eine geringere Lidschlagrate beim Lesen als in einer Phase, in der die Versuchspersonen nichts machten (Karson et al., 1981). Sollten sie jedoch sprechen, stieg die Lidschlagrate deutlich an (Karson et al., 1981). Wurden die Versuchspersonen gebeten zuzuhören, während ihnen ein Text vorgelesen wurde, den sie sich merken sollten, zeigte sich die höchste Lidschlagrate (Karson et al., 1981).

Studien zur kognitiven Belastung (*cognitive load*) erbrachten z. T. ein gegenteiliges Befundmuster (z. B. Holland & Tarlow, 1972). In einer Untersuchung von Ledger (2013) zeigte sich eine geringere Lidschlagrate unter höherer kognitiver Belastung, wobei anzumerken ist, dass in dieser Untersuchung u. a. mit visuellem Stimulusmaterial gearbeitet wurde, was die geringere Lidschlagrate erklären könnte. Tsai, Viirre, Strychacz, Chase und Jung (2007) konnten in einer Untersuchung mit einer sowohl visuellen als auch auditiven Aufgabe hingegen einen Anstieg der Lidschlagrate bei kognitiver Belastung feststellen. Dies bestätigt die zuvor beschriebenen Befunde und deutet auf eine erhöhte Lidschlagrate bei kognitiver Aktivierung hin.

Pupillenmotorik

Die Erfassung der Pupillenmotorik, also der Erweiterung oder Verengung der Pupille, ermöglicht ebenfalls Rückschlüsse auf das Ausmaß der Aktivierung einer Person. So ist sowohl bei kognitiver Beanspruchung als auch bei emotionalen Veränderungen (z. B. Freude oder Angst) eine Pupillendilatation (Erweiterung der Pupille) zu erwarten, wobei auch der Grad der Erregung für das Ausmaß der Pupillenerweiterung verantwortlich ist (Bonowski, 2002). Veränderungen im Pupillendurchmesser durch psychische Prozesse sind dabei i. d. R. relativ gering und betragen, je nach Reizintensität, etwa 0.5 mm (Beatty & Lucero-Wagoner, 2000). Insbesondere die Untersuchungen von Hess und Polt (1964) konnten bereits in den 1960er Jahren Erkenntnisse zum Zusammenhang der Pupillengröße und kognitiver Beanspruchung liefern. So zeigten sie eine Erweiterung der Pupille bei kognitiver Beanspruchung. Zudem konnte das Ausmaß der Erweiterung mit der Aufgabenschwierigkeit in Verbindung gebracht werden. Je höher die kognitive Beanspruchung, desto größer erwies sich die Erweiterung der Pupille (vgl. Hess & Polt, 1964).

Insbesondere für die direkte und objektive Erfassung der kognitiven Beanspruchung hat sich die Messung der prozentualen Veränderung des Pupillendiameters (*percentage change in pupil diameter*, PCPD) als präzise erwiesen (vgl. Kruger, Hefer & Matthew, 2013; Paas, Tuovinen, Tabbers & van Gerven, 2003). Im Vergleich zu anderen physiologischen Messverfahren, wie der Erfassung der Hautleitfähigkeit, besteht ein besonderer Vorteil der Pupillometrie in der geringen Latenzzeit. Dabei beträgt die Latenzzeit der Pupille bei Lichtreaktionen ca. 200 - 250 ms (Gramann & Schandry, 2009).⁸² Bei kognitiver Beanspruchung hingegen sinkt die Latenzzeit der Pupillenreaktion auf ca. 100 - 200 ms (Manzey, 1998). Im Vergleich dazu beträgt die Latenzzeit bei der Erfassung der Hautleitfähigkeit ca. 1000 - 2500 ms, was eine deutlich geringere Genauigkeit bei der Erfassung und Zuordnung von kognitiven Reaktionen auf Reizereignisse zulässt. (Bonowski, 2002; Birbaumer & Schmidt, 2010; Gramann & Schandry, 2009; Schandry, 2003).

Da sowohl der Lidschlag als auch die Pupillendilatation durch das autonome Nervensystem gesteuert werden, ist anzunehmen, dass beide physiologische Reaktionen zu einem gewissen Ausmaß zusammenhängen. Siegle, Ichikawa und Steinhauer (2008) konnten hierfür zeigen, dass zwar ein Zusammenhang zwischen Pupillendilatation und Lidschlagrate besteht, dieser zeitlich aber leicht versetzt ist. Siegle et al. (2008) schließen aus ihren Daten daher, dass kognitive Belastung durch eine Erweiterung der Pupille begleitet wird, eine erhöhte Lidschlagrate jedoch der Pupillenreaktion vorausgeht und anschließend wieder folgt. Dies zeigt auch, dass die Lidschlagrate nur mit bestimmten Prozessabschnitten der Informationsverarbeitung assoziiert ist (vgl. auch Fogarty & Stern, 1989). Nach Ichikawa und Ohira (2004) repräsentiert eine erhöhte Lidschlagrate eher das Ende des Informationsverarbeitungsprozesses, wenn die Belastung im Arbeitsgedächtnis wieder sinkt. Dies würde auch die teilweise inkonsistente Befundlage zum Zusammenhang der Lidschlagrate und kognitiver Beanspruchung erklären, wenn unterschiedliche zeitliche Intervalle zur Erfassung der Lidschlagrate herangezogen werden. Die

⁸² Dabei konnte Keller (2004) einen kleinen aber signifikanten Unterschied ($p < .01$) zwischen den Geschlechtern identifizieren. Die weiblichen Versuchspersonen reagierten dabei mit durchschnittlich 240 ms etwas früher auf einen Lichtstimulus als die männlichen Versuchspersonen (250 ms).

Lidschlagrate kann somit als Indikator für Veränderungen der kognitiven Belastung gesehen werden (Siegle et al., 2008). Diese Zusammenhänge können auch folgendermaßen erklärt werden: Die sympathische Innervation des musculus dilatator pupillae zur Pupillenerweiterung wird direkt mit Beginn (bzw. nach einer kurzen Latenzzeit) der kognitiven Belastung angeregt und führt zur Pupillendilatation. Angenommen werden könnte, dass die parasympathisch innervierten Muskeln zum Lidschluss zu diesem Zeitpunkt (durch die Aktivierung des sympathischen Nervensystems) gehemmt sind und erst nach Absinken der Belastung im Arbeitsgedächtnis parasympathische Impulse den Lidschlag anregen. Durch die vorhergehende Hemmung des Lidschlags, kommt es anschließend zu einer Erhöhung der Lidschlagrate.

Zusammenfassend lässt sich festhalten, dass Fixationen, Lidschlagrate und Pupillendilatation als Indikatoren für kognitive Prozesse angesehen werden können. Dabei müssen allerdings spezifische Annahmen für die einzelnen Augenbewegungen berücksichtigt werden, was dazu führt, dass sie lediglich Hinweise für eine kognitive Aktivierung oder visuelle Aufmerksamkeit liefern. Eine Fixation ist notwendig, um visuelle Informationen aufnehmen zu können, jedoch kann nicht jede Fixation als Aufmerksamkeit gedeutet werden. Fixationen sollten aus diesem Grund zwar nicht als direkte Indikatoren für kognitive Verarbeitung interpretiert werden, sie können aber als eine notwendige Grundlage hierfür angesehen werden und entsprechend Rückschlüsse auf die Aufmerksamkeit zulassen. Empirische Studien zur kognitiven Aktivierung lassen die Rückschlüsse zu, dass eine erhöhte Lidschlagrate sowie eine erweiterte Pupille als Hinweise für eine kognitive Aktivierung interpretiert werden können, sofern durch einen kontrollierten Versuchsaufbau Störvariablen und/oder emotionale Belastung ausgeschlossen werden. Diese Annahme bestätigt sich durch die physiologischen Zusammenhänge beider Augenbewegungen mit dem autonomen Nervensystem.

Da bislang keine Erfahrungen mit der Erfassung okulomotorischer Daten im Kontext einer herkunftsassoziierten Informationsverarbeitung vorlagen und die Fallvignetten so aufbereitet werden mussten, dass sie optimal in das Eye-Tracking-Experiment integriert werden konnten, erfolgte zunächst eine Testung des Versuchsaufbaus.

9.2 Vorstudie zum Einsatz eines Eye-Trackers bei der Schülerbeurteilung

Um den Versuchsaufbau und die Adaption des Stimulusmaterials im Rahmen einer Eye-Tracking-Studie zu testen, wurden in einer experimentellen Laborstudie im November und Dezember 2015 insgesamt 45 Lehramtsstudierende (die im Rahmen ihres Studiums im Basismodul der Psychologie ihre verpflichtenden Versuchspersonenstunden absolvieren mussten) rekrutiert. Bedingt durch Ungenauigkeiten bei der Kalibrierung des Eye-Trackers und eine entsprechend mangelhafte Datenqualität, mussten neun Versuchspersonen aus den folgenden Analysen ausgeschlossen werden. Die Stichprobe betrug somit $N = 36$. Die Studierenden waren im Durchschnitt $M = 21.09$ ($SD = 2.91$) Jahre alt (Fachsemester: $M = 1.19$; $SD = 0.67$). Die Erfassung der okulomotorischen Daten erfolgte mit einem remote Eye-Tracker (SMI red-m) mit einer Abtastrate von 120 Hz. Der Eye-Tracker selbst wurde an einem 22-Zoll-Monitor befestigt. Die Studie wurde von geschulten studentischen Testleiter(inne)n durchgeführt.

Die Studierenden bekamen jeweils drei Grundschulzeugnisse von männlichen Schülern zum Halbjahr der vierten Jahrgangsstufe zu lesen, in denen die Vornamen zur Aktivierung einer spezifischen ethnischen und sozialen Herkunft manipuliert wurden (vgl. Abschnitt 7.1). Nach dem Lesen jedes Zeugnisses wurden die Studierenden in einem Fragebogen um ihre Einschätzung zu den Schülern gebeten (vgl. Abschnitt 7.2). Durch die Anwesenheit der Testleitung bestand die Gefahr, insbesondere bei einem sozial sensiblen Thema wie herkunftsassoziierten Beurteilungen, sozial erwünschtes Antwortverhalten auszulösen. Aus diesem Grund und um visuelle Reize bei der Eye-Tracking-Erhebung zu minimieren, wurde ein weißer Sichtschutz konstruiert, der die Versuchspersonen vom restlichen Raum abschirmte. So konnte die Testleitung die Durchführung der Eye-Tracking-Studie kontrollieren, ohne jedoch im Sichtfeld der Versuchspersonen zu sein und diese abzulenken.

Nach dem Lesen der Fallvignetten wurden die Versuchspersonen anschließend um die Einschätzungen zu Leistungen und leistungsrelevanten Merkmalen gebeten (vgl. Abschnitt 8.4). Hierfür präsentierte die Testleitung nach jedem Zeugnis einen Fragebogen auf dem Experimentalbildschirm, der von den Studierenden ausgefüllt wurde. Dabei hatten die Studierenden eine Laptopaufgabe auf ihrem Schoß, auf dem sich die Tastatur und die Maus befanden. Hierdurch sollte verhindert werden, dass die Studierenden sich zum Ausfüllen des Fragebogens nach vorne bewegen müssen und eine erneute Kalibrierung durchgeführt werden muss. Da das Ziel der Vorstudie der Testung des Settings und Stimulusmaterials galt, wurden lediglich deskriptive Analysen der Eye-Tracking-Daten vorgenommen und v. a. der Prozess der Kalibrierung dokumentiert, um besondere Schwierigkeiten zu erkennen und den Versuchsaufbau und die Durchführung zu optimieren.

Dabei zeigte sich während der eingesetzten neun-Punkt-Kalibrierung, dass diese Schwierigkeiten in der Durchführung bereitete. Die hohe Anzahl an Kalibrierungspunkten sollte zu einer erhöhten Genauigkeit der Messung führen, erwies sich in der Umsetzung jedoch als unpraktikabel. So bereiteten v. a. die Kalibrierungspunkte in den Randbereichen des Bildschirms die größten Probleme, da hier bedingt durch den Blickwinkel und die Monitorgröße die höchste Ungenauigkeit zu verzeichnen war. Nach weiteren Recherchen und dem Austausch mit der Herstellerfirma wurde entsprechend auf eine fünf-Punkt-Kalibrierung umgestellt, da dies für das Stimulusmaterial in der vorliegenden Arbeit zu einer ausreichenden Messgenauigkeit führt.

Da 9 von 45 Versuchspersonen wegen zu geringer Datenqualität ausgeschlossen werden mussten, wurden einige weitere Konsequenzen für die Hauptuntersuchung gezogen. Zum einen wurde der Aufbau der Versuchssituation optimiert. Der Stuhl, auf dem die Versuchspersonen während der Erhebung saßen wurde stabilisiert und die Rückenlehne an einer Rückwand befestigt, um willkürliche und unwillkürliche Bewegungen der Versuchspersonen zu minimieren. Hierzu wurden auch einige zusätzliche Instruktionen zum Verhalten und der Sitzposition der Versuchspersonen in den Erhebungsablauf der Hauptuntersuchung aufgenommen. Zum anderen erfolgte eine Veränderung der Einbettung des Stimulusmaterials in der Experimentalsoftware. So wurde insbesondere darauf geachtet, dass kein Text an den Randbereichen des Bildschirms erscheint, da die Kalibrierungsergebnisse insbesondere hier Probleme bei der Erfassung der Augen zeigten. Entsprechend wurden auch die Kalibrierungspunkte an den neuen Aufbau des Stimulusmaterials angepasst. Des Weiteren wurde für die Hauptuntersuchung ein

19-Zoll-Bildschirm eingesetzt, der ebenfalls durch die kleine Bildschirmgröße eine bessere Erfassung der Lichtreflexion des Auges ermöglicht. Sowohl der Versuchsaufbau im Eye-Tracking-Labor als auch die Präsentation des Stimulusmaterials konnten durch diese Testung der Eye-Tracking-Erhebung optimiert werden. Entsprechend konnten in der darauffolgenden Studie 2 die inhaltlichen Aspekte zu einer herkunftsabhängigen Informationsverarbeitung fokussiert werden.

9.3 Spezifische Fragestellungen und Hypothesen von Studie 2

Basierend auf dem Kontinuummodell der Eindrucksbildung von Fiske und Neuberg (1990) ist davon auszugehen, dass Personen soziale Informationen unter bestimmten Voraussetzungen (z. B. Relevanz des Stimulus für die beurteilende Person) mehr oder weniger automatisiert (d. h. kategoriebasiert) oder kontrolliert (d. h. individuumsbasiert) verarbeiten (vgl. Abschnitt 5.3). Übertragen auf den schulischen Kontext und die Eindrucksbildung von Lehrkräften ist demnach anzunehmen, dass der Grad der Automatisiertheit bzw. Kontrolliertheit der Informationsverarbeitung bei der Schülerbeurteilung ebenfalls variieren kann. Unter Berücksichtigung der ethnischen und sozialen Heterogenität der Schülerschaft im deutschen Bildungssystem (vgl. Abschnitt 2.1) und entsprechenden herkunftsabhängigen Beurteilungen und Erwartungen von Lehrkräften (vgl. Abschnitt 3.3) sowie herkunftsassoziierten Einstellungen und Stereotypen gegenüber bestimmten sozialen Gruppen (vgl. Abschnitt 4.2), besteht eine zentrale Annahme der vorliegenden Arbeit darin, dass die ethnische und soziale Herkunft von Schüler(inne)n und die damit assoziierten Merkmale einen Einfluss auf den Informationsverarbeitungsprozess von Lehrkräften haben können.

Die Befunde der ersten Studie dieser Arbeit zeigen dabei eine akkurate Beurteilung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status sowie eine Überschätzung von Schülern ohne Migrationshintergrund und hohem sozioökonomischem Status (vgl. Abschnitt 8.6). Dies steht weitestgehend im Einklang mit den Befunden von Kaiser et al. (2016) sowie Ready und Chu (2015) und deutet zum einen darauf hin, dass trotz identischer Leistungen und leistungsrelevanter Schülerinformationen die ethnische und soziale Herkunft einen bedeutsamen Einfluss auf die Urteils- und Erwartungsbildung von Lehrkräften hat. Zum anderen lassen die Befunde durch die Akkuratheit der Beurteilung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status auf eine eher kontrollierte Informationsverarbeitung schließen. Die Überschätzung von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status lässt hingegen eine eher automatisierte Informationsverarbeitung vermuten.

Auf dieser Grundlage verfolgte Studie 2 das zentrale Ziel, mit Hilfe okulomotorischer Daten (Fixationen, Lidschlagrate und Pupillendurchmesser) Hinweise auf unterschiedliche Informationsverarbeitungsprozesse bei der Urteils- und Erwartungsbildung in Abhängigkeit der ethnischen und sozialen Schülerherkunft aufzudecken. Theoretische Modelle und empirische Befunde zu herkunftsassoziierten Erwartungen und Stereotypen (vgl. Abschnitt 3.3 und Abschnitt 4.3) von Lehrkräften und Lehramtsstudierenden sowie der theoretischen Annahme, dass Erwartungen die Wahrnehmung beeinflussen (*perceptual confirmation effect*; vgl. Abschnitt 3.2), lassen vermuten, dass sich die visuelle Aufmerksamkeit in Abhängigkeit der her-

kunftsassoziierten stereotypenkonformen und -diskonformen Informationen unterscheidet. Da aus der Leseforschung bekannt ist, dass Personen eine längere Fixationsdauer bei unerwarteten Stimuli (z. B. Rechtschreibfehler; Zola, 1984) aufweisen und Allen et al. (2009) berichten, dass bei geringen kognitiven Kapazitäten eine erhöhte Aufmerksamkeit auf stereotypinkonsistenten Informationen zu finden ist, wurde folgende gerichtete Hypothese formuliert:

H_{2.1}: Bei stereotypinkonsistenten Informationen zeigt sich eine höhere visuelle Aufmerksamkeit (Fixationsdauer) als bei stereotypkonsistenten Informationen.

Zusätzlich könnte angenommen werden, dass Lehramtsstudierende grundlegend motiviert sind alle Schüler(innen) unabhängig von ihrer Herkunft zu beurteilen und entsprechend (unabhängig von stereotypkonformen oder diskonformen Informationen) eine erhöhte visuelle Aufmerksamkeit bei der Beurteilung von Schüler(inne)n mit Migrationshintergrund zeigen um nicht stereotypenbasiert zu urteilen.

Da basierend auf den Ergebnissen der ersten Studie der vorliegenden Arbeit sowie auf den Befunden von Ready und Chu (2015) von einer Überschätzung von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status ausgegangen wird, ist anzunehmen, dass dies an positiven Vorurteilen gegenüber diesen Schülern liegt und sich entsprechend durch eine eher automatisierte bzw. kategoriegeleitete Informationsverarbeitung erklären lässt. Die akkurate Einschätzung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status (vgl. Studie 1; siehe auch Kaiser et al., 2016) deuten hingegen darauf hin, dass die tatsächlichen Leistungen stärker berücksichtigt wurden und somit durch eine eher kontrollierte d. h. individuumsbasierte Informationsverarbeitung zu erklären ist. Entsprechend der Annahme, dass eine eher automatisierte Informationsverarbeitung kognitive Ressourcen einspart, sollte sich in den okulomotorischen Befunden auch eine geringe kognitive Aktivierung bei der Beurteilung von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status auffinden lassen. Eine hohe kognitive Aktivierung sollte durch die aufwändigere individuumsbasierte Informationsverarbeitung bei Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status zu beobachten sein.

H_{2.2}: Beim Lesen der Fallvignette eines Schülers mit Migrationshintergrund und/oder mit niedrigem sozioökonomischem Status zeigt sich eine höhere kognitive Beanspruchung (Pupillendilatation und erhöhte Lidschlagrate) als beim Lesen der Fallvignette eines Schülers ohne Migrationshintergrund und hohem sozioökonomischem Status.

Basierend auf kognitionspsychologischen Annahmen zur sozialen Informationsverarbeitung wird deutlich, dass bereits vorhandene mentale Strukturen den kognitiven Prozess der Informationsverarbeitung beeinflussen können. Zudem ist basierend auf sozialpsychologischen Theorien anzunehmen, dass insbesondere Einstellungen gegenüber sozialen Kategorien die Wahrnehmung und Verarbeitung von sozialen Informationen beeinflussen können. Entsprechend wird in der vorliegenden Arbeit angenommen, dass Einstellungen gegenüber ethnischen Gruppen einen Einfluss auf die Tiefe der Informationsverarbeitung im Sinne des Kontinuum-Modells der Eindrucksbildung (Fiske & Neuber, 1990) haben. Da anzunehmen ist, dass eine eher automatisierte Informationsverarbeitung von Schülern ohne Migrations-

hintergrund auf positiven Einstellungen gegenüber dieser Schülergruppe basiert, sollte zudem folgende Hypothese untersucht werden:

H_{2.3}: Personen mit hohen positiven impliziten Einstellungen gegenüber Personen ohne Migrationshintergrund verarbeiten Informationen von Schülern ohne Migrationshintergrund automatisierter als Personen mit weniger positiven impliziten Einstellungen.

9.4 Untersuchungsdesign und Stichprobe der zweiten Studie

Um den Fragen nach unterschiedlichen kognitiven Prozessen bei der Beurteilung von Schülern unterschiedlicher ethnischer und sozialer Herkunft nachzugehen, wurde im Mai und Juni 2016 eine experimentelle Laborstudie im within-subject-Design durchgeführt. Im Gegensatz zur ersten Studie wurden keine Lehrkräfte für die Untersuchung rekrutiert, sondern Lehramtsstudierende untersucht. So konnten bisherige Untersuchungen auch hier herkunftsassoziierte Beurteilungsunterschiede auffinden (z.B. Glock & Krolak-Schwerdt, 2013). Da Böhmer et al. (2012) darauf verweisen, dass bei unerfahrenen Lehrkräften häufiger eine integrative Informationsverarbeitung abläuft (vgl. Böhmer et al., 2017), ist zudem anzunehmen, dass sie zwar versuchen so viele Informationen wie möglich zu integrieren, jedoch nicht zwischen relevanten und irrelevanten Informationen differenzieren. Die ergänzenden Analysen in Studie 1 konnten zwar Niveauunterschiede zwischen erfahrenen und unerfahrenen Lehrkräften zeigen, jedoch erbrachten die Analysen keine signifikante Interaktion der Berufserfahrung mit der Schülerherkunft (vgl. Abschnitt 8.6). Die herkunftsabhängigen Urteils- und Erwartungsunterschiede in Studie 1 hingen somit nicht mit der Erfahrung von Lehrkräften zusammen. Entsprechend ist davon auszugehen, dass auch Lehramtsstudierende die ethnische und soziale Herkunft von Schüler(inne)n bei der Informationsverarbeitung miteinbeziehen.

Da in Studie 1 der Frauenanteil für eine Lehrkraftstichprobe unter dem erwarteten Anteil lag und die Rekrutierung über Online-Foren (durch die mögliche Teilnahme besonders motivierter Lehrkräfte) zu einer möglichen Stichprobenverzerrung geführt haben könnte, wurde in Studie 2 darauf geachtet, diese Einschränkungen zu vermeiden. Dies gelang durch die Verpflichtung zur Teilnahme durch die jeweiligen Dozierenden der Studierenden in unterschiedlichen Hochschulseminaren. An der Untersuchung nahmen insgesamt 60 Lehramtsstudierende teil. Bedingt durch Kalibrierungsprobleme bei einigen Probanden (u. a. durch dunkle Brillenrahmen) konnten jedoch nicht für alle Versuchspersonen qualitativ hochwertige Eye-Tracking-Daten erhoben werden. Dies führte dazu, dass 14 Versuchspersonen wegen einer unzureichenden Datenqualität ausgeschlossen werden mussten.

Die Studierenden wurden nach ihrem eigenen Migrationshintergrund durch die Erfassung ihres Geburtslands sowie den Geburtsländern ihrer Eltern gefragt (1 [*Deutschland*], 2 [*anderes Land*]). Wurde mindestens ein Elternteil im Ausland geboren, wurde dies als Migrationshintergrund definiert (vgl. Abschnitt 2.1). Entsprechend hatten 21.7% der Studierenden einen Migrationshintergrund. Dabei hatte lediglich eine Person türkische Wurzeln, die übrigen Studierenden mit Migrationshintergrund stammten aus Russland, Kasachstan, Bosnien und Herzegowina, Österreich, Italien, Südkorea und Sri Lanka. Damit zeigten sich äußerst vielfältige Herkunftsregionen der Studierenden, was für weitere Analysen keine differenzierte Berück-

sichtigung der ethnischen Herkunft der Versuchspersonen ermöglichte. Entsprechend der Annahme einer Identifikation mit der ethnischen Ingroup der Versuchspersonen ist davon auszugehen, dass sich, bis auf die Versuchsperson mit türkischen Wurzeln, alle Studierenden nicht mit der Fallvignette, in der ein türkischer Migrationshintergrund aktiviert wurde, identifizierten. Somit wurde die Versuchsperson mit türkischem Migrationshintergrund aus den Analysen ausgeschlossen. Die soziale Herkunft der Studierenden wurde mit einem fünfstufigen Item („Welcher Schicht würden Sie sich bzw. Ihre Familie zuordnen?“) von 1 (*Unterschicht*) bis 5 (*Oberschicht*) erfasst. Die durchschnittliche soziale Schicht der Stichprobe lag bei $M = 3.43$ ($SD = 0.58$). Die Lehramtsstudierenden ordneten sich selbst hauptsächlich der Mittelschicht (54.3 %) oder der oberen Mittelschicht (41.3 %) zu. Jeweils 2.2 % der Stichprobe schätzten die eigene soziale Schicht entweder als Oberschicht oder untere Mittelschicht ein, wobei keine Versuchsperson die soziale Schicht der eigenen Familie als Unterschicht einschätzte.

Durch den Ausschluss von 14 Versuchspersonen wegen geringer Datenqualität und einer Versuchsperson wegen Zugehörigkeit zur türkischen Herkunft, die in diesem Experiment fokussiert wurde, ergab sich eine Analytestichprobe von $N = 45$ Lehramtsstudierenden. Die Studierenden waren im Durchschnitt $M = 21.47$ ($SD = 3.09$) Jahre alt waren (Fachsemester: $M = 3.38$; $SD = 1.97$). Der Frauenanteil lag bei 91.1 % und war damit erwartungsgemäß hoch. Der größte Anteil der Stichprobe bestand mit 46.7 % aus Studierenden des Grundschullehramts (Haupt-/Mittelschullehramt: 15.6 %; Realschullehramt: 13.0 %; Gymnasiallehramt: 24.4 %).

Um intraindividuelle Unterschiede in den Lehrkrafturteilen und -erwartungen sowie kognitiven Prozessen analysieren zu können, wurde das Stimulusmaterial aus Studie 1 (vgl. Abschnitt 7.1.1) auf insgesamt drei Halbjahreszeugnisse der vierten Jahrgangsstufe erweitert. Hierfür wurden durchschnittlich identische Zensuren und Textbausteine mit vergleichbaren Notenäquivalenten zum ersten Grundschulzeugnis für die Verbalbeurteilung ausgewählt (siehe Appendix A). Die ethnische und soziale Herkunft der Schüler wurde über die identischen Vornamen (vgl. Abschnitt 7.1.2) wie in Studie 1 manipuliert.

Manipulations-Check

Im Anschluss an die Beantwortung der Urteile und Erwartungen der Studierenden in Bezug auf die Leistungen und leistungsrelevanten Merkmale wurden die Versuchspersonen zur Prüfung einer erfolgreichen Manipulation der Herkunft um ihre Einschätzung der ethnischen („Was denken Sie, aus welchem Land der Schüler bzw. seine Familie kommt?“; 1 [*Deutschland*], 2 [*anderes Land*]) sowie sozialen („Was denken Sie, welcher sozialen Schicht der Schüler bzw. seine Familie zugehörig ist?“; von 1 (*Unterschicht*) bis 5 (*Oberschicht*)) Herkunft der Schüler gebeten. Dabei zeigte sich eine erfolgreiche Manipulation der ethnischen Herkunft in allen drei Zeugnissen (Zeugnis 1: $\chi^2(2, N = 45) = 36.29$, $p < .001$, Cramer's $V = .898$; Zeugnis 2: $\chi^2(2, N = 45) = 37.24$, $p < .001$, Cramer's $V = .910$; Zeugnis 3: $\chi^2(2, N = 45) = 40.65$, $p < .001$, Cramer's $V = .950$).

Auch die Analyse der sozialen Herkunft erbrachte einen signifikanten Unterschied zwischen den Herkunftskombinationen der Schüler in den Fallvignetten ($F(1.845, 81.199) = 7.792$, $p < .001$, $\eta^2 = .15$). Der Schüler *Julius* wurde entsprechend der geplanten Manipulation einem signifikant höheren sozioökonomischen Status zugeordnet als die beiden anderen Schü-

ler ($F(1,44) = 16.155, p < .001, \eta^2 = .27$). Zwischen dem Schüler mit Migrationshintergrund (*Murat*) und dem Schüler ohne Migrationshintergrund (*Justin*) zeigte sich erwartungsgemäß kein signifikanter Unterschied ($F(1,44) = 2.698, p = .11, \eta^2 = .06$) in der Wahrnehmung der sozialen Schicht.

Fallvignetten und Herkunftskombinationen in Studie 2

Den Studierenden wurde mitgeteilt, dass diese Untersuchung die Blickbewegung beim Lesen von Zeugnissen erfasst. Anschließend wurden die Versuchspersonen um die Bearbeitung von drei Fallvignetten gebeten. Alle Versuchspersonen lasen dabei alle drei Grundschulzeugnisse in identischer Reihenfolge. Jedoch variierte die Reihenfolge der Herkunftskombinationen in den drei Zeugnissen, um einen Einfluss der Abfolge der experimentellen Manipulation zu vermeiden (vgl. Tabelle 11). Durch die Kombination der drei Grundschulzeugnisse mit den drei Herkunftskombinationen ergaben sich somit insgesamt neun Fallvignetten.

Tabelle 11

Übersicht über die neun Fallnetten mit Herkunftsombinationen, entsprechenden Vornamen sowie Anzahl der Versuchspersonen in den sechs möglichen Präsentationsreihenfolgen (Studie 2)

		Mögliche Präsentationsreihenfolgen der Herkunftsombinationen					
Zeugnisse		1	2	3	4	5	6
1. Zeugnis	sozioökonomischer Status hoch; Kein Migrationshintergrund (Julius)	sozioökonomischer Status hoch; Kein Migrationshintergrund (Julius)	sozioökonomischer Status niedrig; Migrationshintergrund (Murat)	sozioökonomischer Status niedrig; Migrationshintergrund (Murat)	sozioökonomischer Status niedrig; Migrationshintergrund (Justin)	sozioökonomischer Status niedrig; Kein Migrationshintergrund (Justin)	sozioökonomischer Status niedrig; Kein Migrationshintergrund (Justin)
2. Zeugnis	sozioökonomischer Status niedrig; Migrationshintergrund (Murat)	sozioökonomischer Status niedrig; Kein Migrationshintergrund (Justin)	sozioökonomischer Status hoch; Kein Migrationshintergrund (Julius)	sozioökonomischer Status hoch; Kein Migrationshintergrund (Julius)	sozioökonomischer Status niedrig; Kein Migrationshintergrund (Justin)	sozioökonomischer Status niedrig; Migrationshintergrund (Murat)	sozioökonomischer Status hoch; Kein Migrationshintergrund (Julius)
3. Zeugnis	sozioökonomischer Status niedrig; Kein Migrationshintergrund (Justin)	sozioökonomischer Status niedrig; Migrationshintergrund (Murat)	sozioökonomischer Status niedrig; Kein Migrationshintergrund (Murat)	sozioökonomischer Status niedrig; Kein Migrationshintergrund (Justin)	sozioökonomischer Status hoch; Kein Migrationshintergrund (Justin)	sozioökonomischer Status hoch; Kein Migrationshintergrund (Julius)	sozioökonomischer Status niedrig; Migrationshintergrund (Murat)
<i>n</i>		7	8	7	7	9	7

Anmerkung: *N* = 45.

Spezifischer Versuchsaufbau zur Erfassung der Okulomotorik

Basierend auf den physiologischen, psychologischen und technischen Voraussetzungen für eine möglichst hohe Datenqualität, wurde ein Eye-Tracking-Labor, soweit möglich, entsprechend der Empfehlungen von Holmqvist et al. (2011) eingerichtet. Da jede Art der sensorischen Reizwahrnehmung eine mögliche Ablenkung der Blickbewegung hervorrufen kann, wurde ein Raum in ruhiger Lage gewählt. Zudem wurden folgende Kriterien berücksichtigt:

- (1) *Lichtverhältnisse:* Um die Lichtverhältnisse im Raum konstant zu halten, wurde ein Labor mit ausschließlich künstlichem Licht gewählt. Das vorhandene Deckenfenster wurde hierfür komplett abgedunkelt. Da Glühbirnen und Halogenlampen mehr störende Vibrationen und Infrarotlicht abstrahlen (Holmqvist et al., 2011), wurden Neonleuchtstoffröhren zur Beleuchtung verwendet.
- (2) *Sichtschutz:* Da ein Versuchsleiter während der Aufzeichnung der Blickbewegungen das Experiment starten sowie überwachen musste und kein angrenzender Raum mit einem verbindenden Einwegspiegel zur Verfügung stand, wurde ein Sichtschutz um den Monitor mit der Eye-Tracking-Kamera aufgestellt. Der Sichtschutz sollte der Versuchsperson einen möglichst geschützten und ruhigen Raum zur individuellen Bearbeitung der Aufgabe ermöglichen. Zudem wurden hierdurch visuelle Ablenkungen durch den Raum oder die Testleitung verhindert.
- (3) *Reizreduktion:* Hierbei wurde darauf geachtet, die Türe sowie das Deckenfenster während der Erhebung geschlossen zu halten, um reflektorische Lidschläge oder eine Veränderung des Pupillendurchmessers, bedingt durch akustische oder optische Reize, zu vermeiden. Somit können Veränderungen der Lidschlagrate oder des Pupillendurchmessers, aber auch spontane Änderungen der Blickbewegung zu einer potentiellen Störquelle weitestgehend ausgeschlossen werden. Auch der Versuchsleiter verhielt sich während der Erhebung ruhig, eventuelle Störungen wurden im Erhebungsprotokoll dokumentiert und bei der Datenaufbereitung berücksichtigt.
- (4) *Stabile Positionierung:* Um eine konstante Datenqualität zu gewährleisten, wurde der Eye-Tracker mit Hilfe von Magneten am unteren Rahmen des Monitors befestigt. Der Monitor selbst wurde auf einer mechanischen Hebebühne fixiert. Diese wurde wiederum auf einem stabilen Tisch befestigt, um gegen mögliche Erschütterungen weitestgehend geschützt zu sein. Die Hebebühne wurde eingesetzt, um die Höhe des Monitors bzw. der Eye-Tracking-Kamera sowie deren Entfernung an die Versuchsperson und jeweilige Größe und Sitzposition anpassen zu können. Hierdurch konnte die Versuchsperson sich in eine möglichst angenehme Sitzposition begeben, was insbesondere für längere Erhebungen hilfreich ist. Außerdem wurde ein möglichst bequemer Stuhl mit Kopfstütze an der Wand des Raumes befestigt, um Bewegungen während der Erhebung zu verhindern sowie die Einstellungen und Kalibrierung zu erleichtern. Da die Versuchspersonen nach jedem Zeugnis einen kurzen Fragebogen zur Einschätzung des Schülers ausfüllen sollten, wurde ihnen (um eine Neukalibrierung zu vermeiden) eine Laptop-Ablage mit der Tastatur und der Maus auf den Schoß gelegt. So konnten die Versuchspersonen ihre Angaben im Fragebogen machen, ohne sich vorbeugen zu müssen.

- (5) *Stimulusmaterial*: Um Lichtschwankungen auszuschließen, wurde das Stimulusmaterial so aufbereitet, dass für jede Seite des Experiments die gleiche Hintergrundfarbe (weiß) und Schriftfarbe (schwarz) genutzt wurde. Zudem wurde auf eine gleichmäßige Verteilung des Textes auf dem Bildschirm geachtet, was für die Verbalbewertung der Fallvignetten sehr gut gelang (vgl. Abbildung 15).

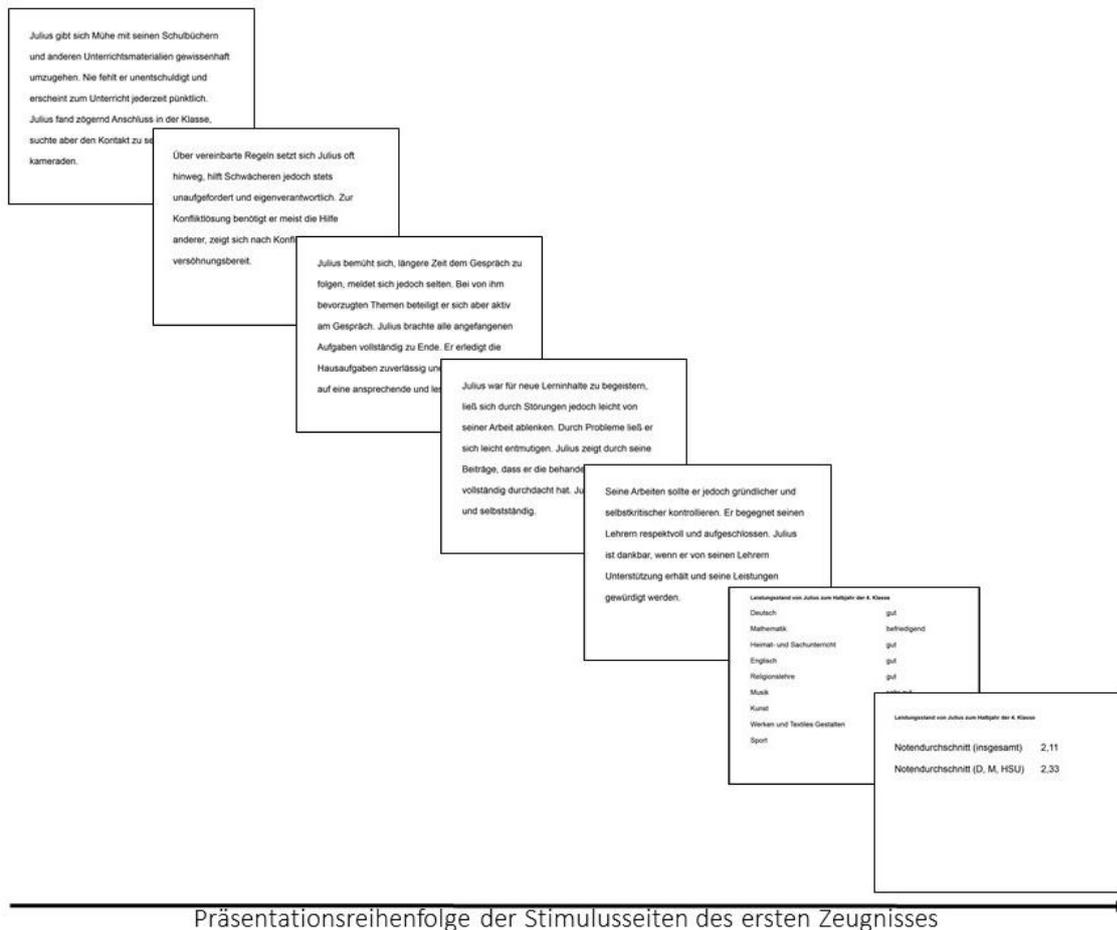


Abbildung 14. Exemplarische Übersicht über den visuellen Aufbau des Stimulusmaterials am Beispiel des ersten Zeugnisses

Die Versuchspersonen wurden über den Einsatz des Eye-Trackers aufgeklärt und um eine ruhige Sitzposition und die Vermeidung von Kopfbewegung gebeten. Anschließend erfolgte vor jeder Erhebung eine fünf-Punkte-Kalibrierung. Die Punkte wurden dabei in der Experimentalsoftware so gesetzt, dass sie den gesamten Bereich des späteren Stimulusmaterials abdeckten. Die Entfernung der Eye-Tracking-Kamera vom Bildschirm und der Winkel der Eye-Tracking-Kamera wurden durch eine Fixierung am unteren Bildschirmrand konstant gehalten. Das Stimulusmaterial wurde auf einem 19-Zoll großen Bildschirm präsentiert. (vgl. Abbildung 16). Über diese Informationen sowie die festgelegten Kalibrierungspunkte und die Entfernung der Versuchsperson von der Eye-Tracking-Kamera berechnet die Experimentalsoftware die Fixationen durch die Verknüpfung der Position der Lichtreflexion des Auges mit den entsprechenden Punkten auf dem Bildschirm.

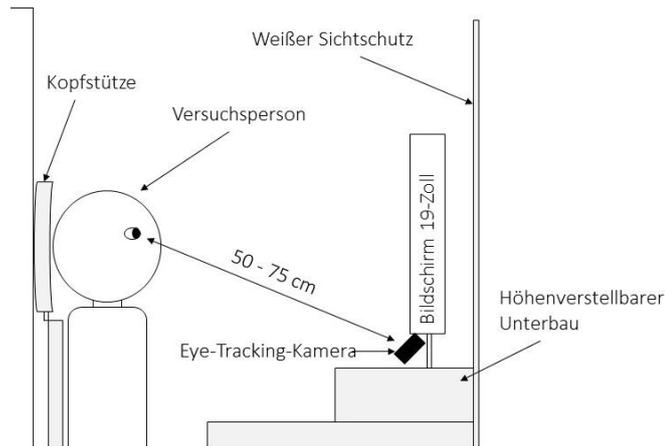


Abbildung 15. Schematische Darstellung des Versuchsaufbaus im Eye-Tracking-Labor

Die Kalibrierung liefert Informationen über die Genauigkeit der Messung als Abweichung der Fixation von der X- und Y-Achse. Die Kalibrierung wurde bei einer Abweichung $\geq 0.5^\circ$ wiederholt, da dies auch ein akzeptabler Abweichungswert aus der Leseforschung ist (z. B. Eichner, 2013). In dieser Studie war zwar keine Messgenauigkeit auf Buchstabenebene, aber auf Wort-/Satzebene erforderlich. Vor einer erneuten Kalibrierung wurden mögliche Störfaktoren (z. B. Veränderung der Entfernung zum Eye-Tracker oder schmutzige Brillengläser) identifiziert und so weit wie möglich behoben.

9.5 Erhebungsinstrumente in Studie 2

Neben der Erfassung der Leistungserwartung ($\alpha = .47$), Leistungsaspiration (adaptiert nach Dresel et al. (2006) und Finsterwald, 2006; $\alpha = .88$), Anstrengungsbereitschaft (adaptiert nach Ramm et al., 2006; $\alpha = .73$), Fähigkeiten (adaptiert nach Dickhäuser et al., 2002; $\alpha = .83$), und Eignung für den Gymnasialbesuch ($M = 3.63$, $SD = 0.64$), deren Messung identisch mit der Erfassung in Studie 1 war (vgl. Abschnitt 7.2 & Abschnitt 8.4), wurden zwei weitere leistungsrelevante Variablen in den Fragebogen integriert. Da im Rahmen des Erwartungs-Wert-Modells insbesondere das Fähigkeitsselbstkonzept der Schüler(innen) einen Einfluss auf die Erfolgserwartung einer Aufgabe hat und diese im Zusammenhang mit der Wertkomponente einen bedeutsamen Einfluss auf die Motivation der Schüler haben (Eccles, 1994; Wigfield & Eccles, 2000), wurde in dieser Studie ebenfalls die Einschätzung der Lehrkraft zum Fähigkeitsselbstkonzept der Schüler erfasst. Hierzu wurde die Subskala des absoluten schulischen Selbstkonzepts aus den Skalen zur Erfassung des schulischen Selbstkonzepts (SESSKO) von Dickhäuser et al. (2002) adaptiert. Hierbei wurden die Versuchspersonen mit fünf Items auf einer fünfstufigen Skala von 1 (*nicht begabt*) bis 5 (*sehr begabt*) um die Angabe gebeten, wie der Schüler sich ihrer Meinung nach selbst einschätzen würde ($\alpha = .91$; Beispielitem: „Ich bin für die Schule...“). Um beide zentralen Komponenten auf die Motivation zu berücksichtigen, wurde ebenfalls erfasst, wie Lehrkräfte den subjektiven schulischen Wert der Schüler einschätzen. Hierfür wurden sechs Items zum subjektiven Wert von Schüler(inne)n nach Ziegler, Dresel, Stöger und Schober (2008) adaptiert und auf einer fünfstufigen Skala von 1 (*stimmt gar nicht*) bis 5 (*stimmt voll und ganz*) erfasst (Beispielitem: „In der Schule gut zu sein ist ihm

wichtig.“). Die Items umfassten dabei sowohl den intrinsischen Wert als auch die Nützlichkeit und Wichtigkeit. Mit $\alpha = .83$ erwies sich die Skala als reliabel. Die Versuchspersonen sollten in dieser Skala sowie in der Skala zum Fähigkeitsselbstkonzept versuchen sich in die Perspektive des Schülers ein zudenken.

Erfassung expliziter und impliziter Einstellungen

Als intraindividuelle Merkmale der Versuchspersonen wurden explizite und implizite Einstellungsmaße erhoben, die im Zusammenhang mit der ethnischen und sozialen Herkunft von Schüler(inne)n stehen. Da ein zentrales Merkmal von Einstellungen die entsprechende Valenz gegenüber Einstellungsobjekten ist, wurden mit jeweils drei bipolaren Items von 1 (*schlecht/negativ/unangenehm*) bis 6 (*gut/positiv/angenehm*) die Einstellung der Lehramtsstudierenden in Bezug auf eine heterogene Zusammensetzung der Schülerschaft sowohl in Bezug auf die kulturelle Herkunft als auch die soziale Herkunft erfasst (Grassinger et al., 03/2016). Mit $\alpha = .84$ für die Valenz gegenüber einer sozial heterogenen Schülerschaft zeigte sich eine gute interne Konsistenz. Für die Skala zur Valenz gegenüber einer kulturell heterogenen Schülerschaft mit $\alpha = .58$ erwies sich die Skala als nicht ausreichend reliabel. Aber nicht nur explizite Einstellungen sind von Bedeutung, sondern auch implizite Einstellungen können die Informationsverarbeitung beeinflussen (vgl. Kapitel 4 & Kapitel 5). Für die Niederlande konnten van den Bergh et al. (2010) zeigen, dass die Lehrkrafteerwartungen zwar nicht mit den expliziten Selbstberichtsdaten der Lehrkräfte zusammenhingen, jedoch mit den impliziten Einstellungen gegenüber ethnischen Minoritäten. Implizite Einstellungen wurden in der vorliegenden Untersuchung durch den *impliziten Assoziationstest* (IAT), der von Greenwald et al. (1998) entwickelte wurde, erfasst. Die Messung erfolgte mit der Software Inquisit 5 (Millisecond Software, 2016). Der IAT ist ein reaktionszeitbasiertes Instrument zur Erfassung impliziter Einstellungen durch die Messung zugrundeliegender automatischer Bewertungen. Die Grundlegende Annahme besteht darin, dass es Personen umso leichter fällt eine Kategorie mit Bewertungen zu kombinieren, je stärker die automatische Assoziation ist. Entsprechend wird mit dem IAT die Stärke der Assoziation zwischen Kategorien mit positiven oder negativen Attributen gemessen, welche zusammen mit den entsprechenden Zielkategorien (in der vorliegenden Studie deutsche und türkische Vornamen) präsentiert werden. Hierfür werden in unterschiedlichen Test- und Übungsblöcken die Attribut- und Zielkategorien in kongruenten – stereotypkonformen – und inkongruenten Blöcken präsentiert (vgl. Abbildung 17).

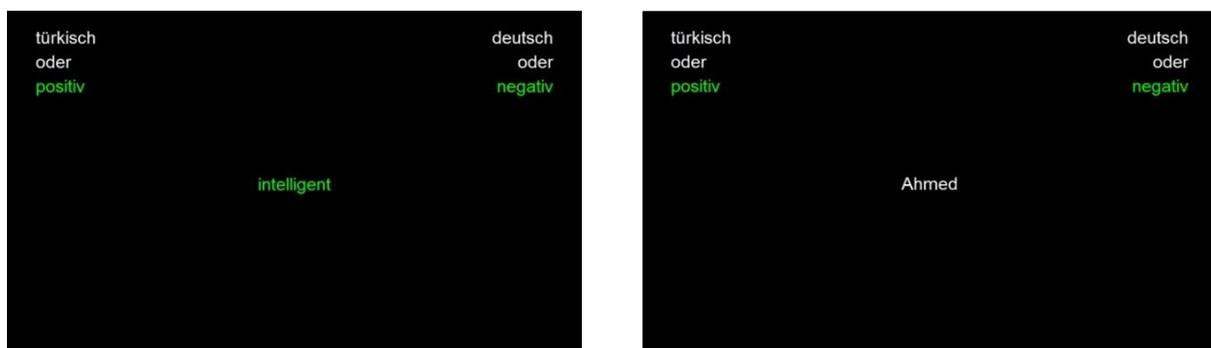


Abbildung 16. Zuordnungsaufgabe einer Attributionskategorie (links) und einer Zielkategorie (rechts) in einem stereotypinkonsistenten Testblock (vgl. Tobisch, 2013)

Die Erfassung der impliziten Stereotype erfolgte dabei über mehrere Übungs- und Testblöcke, in denen die Kombinationsreihenfolge der Kategorien und Attribute variieren (vgl. Tabelle 12). Vor jedem Block erhielten die Versuchspersonen eine genaue Instruktion darüber, welche Taste der Tastatur mit der linken bzw. rechten Bildschirmseite verknüpft ist. Um einen möglichen Einfluss der Kombinationsreihenfolge (u. a. Mierke, 2004) kontrollieren zu können, wurden zwei Reihenfolgevarianten konzipiert, die den Versuchspersonen in randomisierter Reihenfolge zugeordnet wurden (vgl. Tabelle 12).⁸³ Zwischen beiden Reihenfolgekombinationen zeigte sich kein signifikanter Unterschied ($t(42) = 0.357, p = .72$). Es kann entsprechend davon ausgegangen werden, dass die Präsentationsreihenfolge der stereotypkonformen bzw. -diskonformen Attribute keinen Einfluss auf die Assoziationsstärke hat.

Tabelle 9

Reihenfolgevarianten der IAT Übungs- und Testblöcke (vgl. Tobisch, 2013)

Block/ Funktion	Variante 1		Variante 2	
	Linke Bildschirmseite	Rechte Bildschirmseite	Linke Bildschirmseite	Rechte Bildschirmseite
1/Übung	deutsche Namen	türkische Namen	türkische Namen	deutsche Namen
2/Übung	positive Eigenschaften	negative Eigenschaften	positive Eigenschaften	negative Eigenschaften
3/Übung	deutsche Namen oder positive Eigenschaften	türkische Namen oder negative Eigenschaften	türkische Namen oder positive Eigenschaften	deutsche Namen oder negative Eigenschaften
4/Test	deutsche Namen oder positive Eigenschaften	türkische Namen oder negative Eigenschaften	türkische Namen oder positive Eigenschaften	deutsche Namen oder negative Eigenschaften
5/Übung	türkische Namen	deutsche Namen	deutsche Namen	türkische Namen
6/Übung	türkische Namen oder positive Eigenschaften	deutsche Namen oder negative Eigenschaften	deutsche Namen oder positive Eigenschaften	türkische Namen oder negative Eigenschaften
7/Test	türkische Namen oder positive Eigenschaften	deutsche Namen oder negative Eigenschaften	deutsche Namen oder positive Eigenschaften	türkische Namen oder negative Eigenschaften

Der IAT-Effekt (D) berechnet sich aus der Differenz der unterschiedlichen Latenzzeiten für kongruente und inkongruente Testblöcke. Der Wert kann dabei sowohl positive als auch negative Werte annehmen, die je nach Versuchsbedingung die Richtung des Zusammenhangs zwischen bestimmten Attributen und Zielkategorien beschreibt. Der IAT-Wert kann in der vorliegenden Arbeit folgendermaßen interpretiert werden: Bei einem Wert von $D = 0$ ist von

⁸³ Angenommen werden könnte, dass Versuchspersonen die Bearbeitung des zweiten Testblocks (in dem sich die Zuordnung von stereotypkonformen bzw. -diskonformen Attributen ändert) schwerer fällt, da in der ersten Testphase möglicherweise ein Lerneffekt aufgetreten sein könnte.

keiner Präferenz gegenüber einer der beiden Zielkategorien auszugehen. Bei einem hohen positiven Wert besteht eine stärkere Verknüpfung von negativen Attributen mit türkischstämmigen Personen oder eine stärkere Verknüpfung von positiven Attributen mit Personen ohne Migrationshintergrund. In jedem Fall sagt ein hoher positiver Wert eine Präferenz gegenüber Personen ohne Migrationshintergrund im Vergleich zu Personen ohne Migrationshintergrund aus. Ein hoher negativer Wert beschreibt hingegen eine Präferenz gegenüber Personen mit Migrationshintergrund im Vergleich zu Personen ohne Migrationshintergrund. Dies kann an einer starken Assoziation zwischen einem türkischen Migrationshintergrund und positiven Attributen bzw. an einem starken Zusammenhang zwischen Personen ohne Migrationshintergrund und negativen Attributen liegen.

Erfassung okulomotorischer Daten

Das in dieser Arbeit verwendete binokulare remote RED-m Eye-Tracking-System von *Sensomotoric Instruments* (SMI) basiert auf einer Infrarot-Kameratechnologie (SMI, 2012) mit einer Abtastrate (*Samplingrate*) von 60 - 120 Hz und einer Akkuratheit von 0.5° (SMI, 2014a).⁸⁴ Für die Erfassung pupillometrischer Daten wird eine Abtastrate von mind. 120 Hz und die Mittelung der Einzelmessungen empfohlen, um natürliche spontane Schwankungen der Pupillenweite auszugleichen (Gramann & Schandry, 2009). Dementsprechend erfolgte in der Studie der vorliegenden Arbeit eine binokulare Aufzeichnung mit einer Abtastrate von 120 Hz. Nach Herstellerangaben kann eine Bildschirmgröße von bis zu 22-Zoll eingesetzt werden, was unter Umständen jedoch die Datenqualität in den Randbereichen des Bildschirms einschränken kann (vgl. Abschnitt 9.2). Aus diesem Grund wurde der remote Eye-Tracker an einem 19-Zoll Bildschirm angebracht. Bedingt durch unterschiedliche Sitzpositionen und Körpergrößen wurde vor Beginn der Testung zunächst die Höhe und der Abstand des Eye-Tracker zur Versuchsperson so eingestellt, dass der Abstand des Eye-Trackers in der idealen Entfernung von ca. 50 - 75cm (SMI, 2014a) lag. Das Experiment wurde in der Software SMI Experiment Center™ 3.6 (SMI, 2014c) konzipiert und durchgeführt. Dabei wurden die Fallvignetten in Form von Grundschulzeugnissen auf mehrere Seiten verteilt, um die Schriftgröße erhöhen zu können. Anschließend wurden die positiven und negativen Textfragmente nach Inhalten getrennt (z. B. Arbeitsverhalten oder Sozialverhalten) sowie die Ziffernbewertungen der Fallvignetten in *Areas of Interest* (AOIs) eingeteilt. Da Sakkaden eine maximale Dauer von 80 ms aufweisen, wurden in der vorliegenden Untersuchung Fixationen ab einer Dauer von 80 ms und mit einer maximalen Dispersion von 100 px (vergleichbar mit anderen Untersuchungen, z. B. Kruger et al., 2013) als solche definiert.

Die Fixationsdauer wurde für jede AOI in den Fallvignetten erfasst, um die visuelle Aufmerksamkeit der Versuchspersonen auf unterschiedlichen Textabschnitten zu messen. In Anlehnung an die Eye-Mind-Hypothese (Just & Carpenter, 1980) lässt die Fixationsdauer Rückschlüsse auf die Verarbeitungstiefe zu. Längere Fixationen werden entsprechend als Indikator für eine höhere visuelle Aufmerksamkeit interpretiert. Auch eine erweiterte Pupille kann, be-

⁸⁴ Die Verarbeitung visueller Reize erfolgt i. d. R. binokular mit einer synchronen Blickbewegung beider Augen (Ausnahmen hierzu finden sich z. B., wenn ein fixiertes Objekt sich im dreidimensionalen Raum bewegt, aber auch wenn ein Auge dominant und das andere eher passiv ist) (Holmqvist et al., 2011), was bei der Analyse von Blickbewegungsdaten bei binokularen Aufnahmen berücksichtigt werden kann.

dingt durch die Aktivierung des sympathischen Nervensystems (im Gegensatz zu einer Verengung der Pupille als Konsequenz einer Aktivierung des parasympathischen Nervensystems), als Indikator für Stress, Belastung und Erregung (vgl. Bonowski, 2002; Birbaumer & Schmidt, 2010; Gramann & Schandry, 2009) und damit als Hinweis auf eine eher kontrollierte Informationsverarbeitung angesehen werden. Die Versuchspersonen konnten die Zeugnisseiten ohne Zeitdruck lesen und selbst entscheiden, wann sie fortfahren möchten, hierzu konnten sie selbst die Leertaste betätigen und die Aufgaben in ihrem präferierten Tempo bearbeiten. Nach jedem Zeugnis wurden die Versuchspersonen um ihre Einschätzung des Schülers gebeten. Um sich dabei so wenig wie möglich bewegen zu müssen, schaltete die Testleistung nach jedem Zeugnis den Fragebogen zur Beurteilung der Schüler auf den Eye-Tracking-Monitor.

Im Anschluss an die Eye-Tracking-Erhebung und die inhaltlichen Skalen wurden die Versuchspersonen zusätzlich um einige Angaben gebeten, die insbesondere bei der Auswertung der Pupillenreaktion bedeutsam sein können. Da für diese Untersuchung jedoch die Latenzzeit der Pupillenreaktion auf das Stimulusmaterial nicht ausgewertet wurden, konnten die Angaben hierfür zunächst vernachlässigt werden. Bei Problemen während der Kalibrierung oder untypischen Datenmustern konnten diese Informationen aber herangezogen werden, um Schwierigkeiten möglicherweise zu klären. So wurden u. a. Fragen zur Einnahme von Muskelrelaxenzen gestellt oder ob die Versuchspersonen Kontaktlinsen trugen.

Umgang mit fehlenden Werten

Da die Versuchspersonen den Fragebogen zur Einschätzung der Urteile über und Erwartungen an Leistungen und leistungsrelevante Merkmale sowie für die Beantwortung der Fragen zu ihren Einstellungen sowie dem Kognitionsbedürfnis am Computer ausfüllten, konnten hier alle Items als Pflichtitems deklariert werden. Hierdurch lagen für alle Items keine fehlenden Werte vor. Für die Fixationsdaten lag der Anteil fehlender Werte für jede AOI zwischen 0% und 12.2%. (durchschnittlich 5.4%), der mittels EM-Algorithmus geschätzt und imputiert wurde (vgl. Graham, 2009; Lüdtke et al., 2007). Da fehlende Werte der Pupillendaten sich u. a. durch Lidschlüsse ergeben können und durch die Abtastrate von 120 Hz für alle AOIs Pupillendaten vorlagen, die für die weiteren Analysen für die einzelnen AOIs aggregiert wurden, lagen keine fehlenden Werte der Pupillendaten vor. Da Lidschlüsse durch eine fehlende Rücksendung der Lichtreflexion des Auges an das Eye-Tracking-System berechnet werden und diese nicht zu feststehenden Zeitpunkten erfolgen, konnten keine fehlenden Werte der Lidschläge ermittelt werden. Jedoch wurden, wie bereits erwähnt, Versuchspersonen aus den Analysen ausgeschlossen, die durch eine schlechte Datenqualität – u. a. ein längerer Signalverlust der Lichtreflexion des Auges – aufgefallen waren. Hierdurch kann von einer relativ zuverlässigen Erfassung der Lidschläge ausgegangen werden.

9.6 Analysemethoden und Aufbereitung der okulomotorischen Daten der zweiten Studie

Da in dieser Studie ein within-subject-Design vorlag, wurden die Daten mit einer ein-faktoriellen Varianzanalyse mit drei Messwiederholungsstufen (Herkunftskombinationen in den Fallvignetten) analysiert. Dies stellt insbesondere durch die geringe Stichprobengröße eine

notwendige Analyse dar, um die Teststärke im Vergleich zu einem between-subject-Design zu erhöhen. Da hierfür gleiche Varianz in den Versuchsbedingungen angenommen wird sowie eine gleiche Kovarianz zwischen den Bedingungen vorliegen muss (vgl. Field, 2009), sollte der Mauchly-Test auf Sphärizität (Mauchly, 1940) berechnet werden. Da der Mauchly-Test jedoch insbesondere bei kleinen Stichproben – wie im Fall von Studie 2 – nicht exakt ist (hier kann eine Verletzung der Sphärizität vorliegen, obwohl der Mauchly-Test nicht signifikant ist; vgl. Rasch, Friese, Hofmann & Naumann, 2010), wurde für alle Varianzanalysen mit Messwiederholung eine Korrektur der Freiheitsgrade nach Greenhouse und Geisser (1959) vorgenommen. Um eine differenzierte Analyse der Effekte der ethnischen und/oder sozialen Herkunftsmanipulation in den Fallvignetten zu ermöglichen, wurden im Anschluss an die Varianzanalysen mit Messwiederholung a priori Helmert-Kontrastanalysen berechnet (vgl. Abschnitt 8.5; Tabelle 7). Der erste Kontrast verglich dabei den Schüler mit hohem sozioökonomischem Status in den Fallvignetten mit den beiden Schülern mit niedrigem sozioökonomischem Status (*Julius* vs. *Justin* und *Murat*). Der zweite Kontrast verglich hingegen den Einfluss des Migrationshintergrunds bei gleichem sozioökonomischem Status (*Justin* vs. *Murat*). Zu Prüfung der Hypothesen $H_{2.3}$ und $H_{2.4}$ wurden die entsprechenden Variablen der impliziten und expliziten Einstellungen bzw. des Kognitionsbedürfnisses als Kovariaten in die Varianzanalysen mit Messwiederholung miteinbezogen.

Aufbereitung der okulomotorischen Daten

Zunächst wurden die Textfragmente der Zeugnisse in der Analysesoftware SMI BeGaze™ 3.6 (SMI, 2014b) in relevante Regionen eingeteilt, sogenannte *Areas of Interest* (AOIs). Dabei wurden die Textbausteine und deren positive und negative Valenz (vgl. Abschnitt 7.1.2) berücksichtigt. Da einzelne Textbausteine der Zeugnisse aus mehreren Satzteilen bestanden, die unterschiedlich bewertet werden können (z. B. Textfragment zum Arbeitsverhalten aus Zeugnis 1: „...arbeitet sicher und selbständig, seine Arbeiten sollte er jedoch gründlicher und selbstkritischer kontrollieren.“), wurden diese für die Analysen in einzelne Bestandteile zerlegt. Insgesamt ergaben sich für die drei Zeugnisse somit 203 AOIs.

Da die Annahme besteht, dass spezifische Inhalte der AOIs je nach ethnischer und/oder sozialer Herkunft als stereotypkonsistent bzw. inkonsistent angesehen werden, war es notwendig zunächst zu prüfen, ob die AOIs zusätzlich zum entsprechenden Notenäquivalent von Lehramtsstudierenden tatsächlich als positiv oder negativ angesehen werden. Hierzu wurde die Valenz – 1 (*positiv*), 2 (*neutral*), 3 (*negativ*) – der einzelnen AOIs ($N = 203$) durch sechs unabhängige Rater bewertet. Da nominale Daten von mehr als drei Ratern vorlagen, wurde das Fleiss Kappa als Grad der Übereinstimmung berechnet (Fleiss, 1971; Wirtz & Caspar, 2002), wobei sich mit $\kappa = 0.74$ eine substantielle Übereinstimmung zeigte (vgl. Landis & Koch, 1977). Da allerdings nicht alle Textfragmente entsprechend der Notenäquivalente wahrgenommen wurden, wurden einzelne Textabschnitte, bei denen keine eindeutige positive oder negative Zuordnung durch die Rater in Übereinstimmung mit den Notenäquivalenten der Textbausteine vorlag, aus den Analysen entfernt. Anschließend wurden die AOIs nach Inhalt und Valenz zusammengefasst. Daraus ergaben sich vier positive AOIs (regelmäßiger Schulbesuch, Arbeitsverhalten, Sozialverhalten und Verhalten gegenüber der Lehrkraft) und drei negative AOIs (Ordnungsverhalten, Arbeitsverhalten und Sozialverhalten). Zur Analyse der Daten-

qualität wurden in BeGaze™ 3.6 (SMI, 2014b) über die Analysefunktion Scan Path für jede Versuchsperson jede Seite aller Zeugnisse auf mögliche systematische Verschiebungen der Fixationen hin analysiert und soweit eindeutig erkennbare systematische Verzerrungen manuell den richtigen AOIs zugeordnet. War dies nicht zweifelsfrei möglich, wurden die entsprechenden AOIs im Datensatz als fehlende Werte markiert. Anschließend erfolgte der Datenexport der Fixationsdauer sowie des Pupillendurchmessers für jede AOI und der Lidschlagrate für jede Stimulusseite für die Weiterbearbeitung in SPSS 24 (IBM, 2016).

Da die AOIs sich in der Wort- und Satzlänge unterschieden, ist durch die Analyse der Fixationsdauer pro AOI kein Vergleich zwischen stereotypkonsistenten und inkonsistenten Textfragmenten möglich. Aus diesem Grund wurde für die sieben AOIs jeweils die Gesamtgröße in Pixeln (px) ermittelt. Anschließend wurde die individuelle Fixationsdauer (ms) pro AOI durch die Größe der entsprechenden AOI dividiert, daraus ergab sich die relative Fixationsdauer pro Pixel. Da binokulare Daten erfasst wurden, wurden zunächst separat für jedes Auge die Pupillendaten für jede AOI aggregiert. Dies war zunächst erforderlich, da sich die Lesezeiten der Versuchspersonen stark unterschieden und durch die Abtastrate von 120Hz entsprechend viele, jedoch auch in der Anzahl stark interindividuell variierende Angaben zum Pupillendurchmesser für jede AOI vorlagen. Anschließend wurde der Mittelwert des Pupillendiameters aus den Werten beider Augen für jede AOI berechnet. Lag in einzelnen Fällen keine Angabe für beide Augen vor, wurden die Werte des Auges ausgewählt, für das mehr Daten vorhanden waren. Zur Berechnung der prozentualen Veränderung der Pupillendiameter ist zunächst ein Baseline-Wert notwendig, der mit dem Pupillendurchmesser in den entsprechenden Versuchsbedingungen verglichen wird (vgl. Kruger et al., 2013). Wie bereits in anderen Studien wurden als Baseline-Daten der durchschnittliche Pupillendiameter beim Lesen der anfänglich präsentierten Instruktionseiten verwendet und anschließend von den Werten der Pupillengröße in den Versuchsbedingungen abgezogen (vgl. u. a. Kruger et al., 2013).⁸⁵ Daraus ergab sich der durchschnittliche Pupillendiameter pro AOI in Relation zur Baseline. Dabei kann ein Wert von 0 als kein Unterschied im Pupillendiameter der AOI zur Baseline, ein Wert von < 0 als ein kleinerer Pupillendiameter und ein Wert von > 0 als ein erweiterter Pupillendiameter im Vergleich zu Baseline interpretiert werden.

Da bei einem Lidschluss keine Positionsdaten des Auges an den Eye-Tracker übermittelt werden, kann strenggenommen keine direkte Zuordnung des Lidschlusses zu einer AOI vorgenommen werden. Entsprechend wurde die Lidschlagrate, also die Anzahl an Lidschlägen, für das gesamte Stimulusmaterial berechnet. Zur Analyse der Lidschlagrate (Lidschlaganzahl pro Sekunde) wurden zunächst die Anzahl der Lidschläge pro Sekunde der Baseline-Seiten (ohne die erste Seite) addiert und durch die Anzahl der drei letzten Instruktionseiten geteilt. Anschließend wurde für jede Zeugnisseite die individuelle Lidschlagrate in Relation und zu einem Gesamtwert gemittelt und daraufhin in Relation zu Baseline gesetzt.

⁸⁵ Die erste Instruktionseite wurde ausgeschlossen, da davon auszugehen ist, dass die Versuchspersonen sich zunächst an das dargebotene Format, Schrittgröße und Bildschirmhelligkeit gewöhnen müssen.

9.7 Ergebnisse von Studie 2

Deskriptive Befunde

Insgesamt zeigte sich über alle Zeugnisse und Experimentalgruppen hinweg eine eher positive Wahrnehmung der Schüler durch die Lehramtsstudierenden (vgl. Tabelle 13). Sie erwarteten in zukünftigen Proben in den drei Hauptfächern tendenziell eher positive Leistungen von den Schülern und wären mit der Verschlechterung um durchschnittlich eine Notenstufe in den Fächern Deutsch, Mathematik sowie Heimat- und Sachunterricht gerade noch zufrieden. Die Versuchspersonen hielten die Schüler zudem für eher geeignet, das Gymnasium zu besuchen und schätzten auch die Anstrengungsbereitschaft, die schulischen Fähigkeiten sowie das Fähigkeitsselbstkonzept der Schüler tendenziell über dem Skalenmittel ein. Den subjektiven Wert, den die Schüler der Fallvignetten der Schule zuschreiben, schätzten die Studierenden ebenfalls leicht über dem Skalenmittel ein.

Tabelle 10

Mittelwerte (Standardabweichungen) und bivariate Korrelationen der Einschätzungen der Leistungen und leistungsrelevante Merkmale (Studie 2)

	<i>M</i> (<i>SD</i>)	(1)	(2)	(3)	(4)	(5)	(6)
(1) Leistungsaspiration (rekodiert)	3.71 (0.49)	-					
(2) Leistungserwartung (rekodiert)	4.78 (0.20)	.38*	-				
(3) Gymnasialeignung	3.63 (0.64)	-.14	.12	-			
(4) Anstrengungs- bereitschaft	3.71 (0.35)	-.09	.26	.48**	-		
(5) Fähigkeiten	3.58 (0.36)	.13	.18	.68**	.65**	-	
(6) Fähigkeits- selbstkonzept	3.39 (0.48)	.27	.25	.34*	.38**	.62**	-
(7) Subjektiver Wert	3.47 (0.36)	-.00	.28	.16	.55**	.42*	.44*

Anmerkungen. ** $p < .01$, * $p < .05$, $N = 45$.

Die bivariaten Korrelationsanalysen erbrachten einen schwachen positiven Zusammenhang der Leistungsaspiration und der Leistungserwartung der Lehramtsstudierenden (Tabelle 13). Entgegen der Erwartung korrelieren die Einschätzungen der Leistungen nicht mit den leistungsrelevanten Schülermerkmalen. Die Einschätzung der Studierenden zur Gymnasialeignung der Schüler hing zwar nicht mit der Einschätzung des subjektiven Werts zusammen, korrelierte jedoch erwartungsgemäß positiv mit der Einschätzung der Anstrengungsbereitschaft, der Fähigkeiten und des Fähigkeitsselbstkonzepts. Für die weiteren leistungs-

relevanten Merkmale zeigten sich ebenfalls positive Korrelationen. Die Einschätzung der Anstrengungsbereitschaft hing dabei stark mit den eingeschätzten Fähigkeiten sowie dem subjektiven Wert und moderat mit dem eingeschätzten Fähigkeitsselbstkonzept zusammen. Auch für die Einschätzung des Fähigkeitsselbstkonzepts mit den eingeschätzten Fähigkeiten ergab sich eine hohe positive Korrelation.

Einfaktorielle Varianzanalysen mit Messwiederholung erbrachten einen herkunftsabhängigen Unterschied der Leistungserwartung ($F(1,933,85.048) = 3.978, p < .05, \eta^2 = .08$), der sich durch die soziale Herkunft erklären lässt ($F(1,44) = 7.563, p < .01, \eta^2 = .15$). Die Studierenden erwarteten von den beiden Schülern mit niedrigem sozioökonomischem Status – unabhängig von der ethnischen Herkunft ($F(1,44) = 0.481, p = .48, \eta^2 = .01$) – signifikant bessere Leistungen in zukünftigen Proben als von dem Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status. Für die Leistungsaspiration ($F(1,918,84.376) = 0.010, p = .99, \eta^2 = .00$) sowie die Einschätzungen der Gymnasialeignung ($F(1,995,87.779) = 0.935, p = .40, \eta^2 = .02$), der Anstrengungsbereitschaft ($F(1,889,83.121) = 0.937, p = .39, \eta^2 = .02$), der Fähigkeiten ($F(1,955,86.023) = 1.585, p = .21, \eta^2 = .04$), des Fähigkeitsselbstkonzepts ($F(1,977,86.993) = 0.487, p = .61, \eta^2 = .01$) sowie des subjektiven Werts ($F(1,924,84.640) = 2.242, p = .12, \eta^2 = .05$) konnten hingegen keine signifikanten herkunftsabhängigen Unterschiede identifiziert werden. Bei der Einschätzung des subjektiven Werts zeigte sich jedoch ein signifikanter Kontrast der ethnischen Herkunft ($F(1,44) = 3.957, p < .05, \eta^2 = .08$). Die Lehramtsstudierenden schätzten den subjektiven Wert der Schule für den Schüler mit Migrationshintergrund signifikant höher ein als für den Schüler ohne Migrationshintergrund mit vergleichbar niedrigem sozioökonomischem Status. Die Befunde aus Studie 1 konnten somit nicht repliziert werden, die Ergebnisse weisen tendenziell auf eine eher positive Einschätzung des Schülers mit Migrationshintergrund und niedrigem sozioökonomischem Status im Vergleich zu dem Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status hin.

Die *Ergebnisse der Einstellungsmessungen* weisen für implizite und explizite Maße in entgegengesetzte Richtungen. Dabei zeigten die Ergebnisse des IAT-Werts mit $D = 0.58$ ($SD = 0.5$) eine starke Assoziation von negativen Attributen mit türkischstämmigen Personen bzw. einer starken Assoziation von positiven Attributen mit Personen ohne Migrationshintergrund. Die expliziten Einstellungsmaße erbrachten eine positive Einstellung gegenüber einer ethnisch heterogenen Schülerschaft ($M = 4.46, SD = 0.13$) und eine eher positive Einstellung gegenüber einer sozial heterogenen Schülerschaft ($M = 3.85, SD = 0.10$). Dabei hatten die Studierenden der ethnischen Heterogenität in Schulklassen gegenüber eine signifikant positivere Einstellung im Vergleich zu einer sozial heterogenen Schülerschaft ($t(44) = 4.294, p < .001$).

Die deskriptiven *Befunde der okulomotorischen Daten* zeigen zunächst signifikante Unterschiede der relativen Fixationsdauer der einzelnen AOIs ($F(1,817,79.936) = 281.329, p < .001, \eta^2 = .87$). Insbesondere die ersten beiden AOIs zum Ordnungsverhalten ($M = 0.024, SD = 0.007$) und dem regelmäßigen Schulbesuch ($M = 0.014, SD = 0.014$) werden im Vergleich zu den anderen Textabschnitten ($M_{min} = 0.001$ ms/px - $M_{max} = 0.009$ ms/px) deutlich länger fixiert. Insgesamt betrachten die Studierenden die negativen Textteile ($M = 0.013, SD = 0.004$) signifikant länger ($F(1,000,44.000) = 254.673, p < .001, \eta^2 = .85$) als die positiven Abschnitte in den Fallvignetten ($M = 0.007, SD = 0.002$).

Der durchschnittliche Pupillendiameter der Versuchspersonen als Indikator für kognitive Aktivierung lag bei $M = -0.05$ ($SD = 0.08$) und war damit signifikant geringer als bei der Baseline-Messung ($t(44) = -4.276, p < .001$). Jedoch wurde eine große interindividuelle Spannweite der durchschnittlichen Abweichung des Pupillendiameters von der Baseline sichtbar (Min = -0.19, Max = 0.13). Vergleichbar zu den Befunden der Fixationsdauer zeigte sich ebenfalls ein signifikanter Unterschied zwischen den einzelnen Textbausteinen ($F(3.039, 133.699) = 23.205, p < .001, \eta^2 = .35$). Dabei konnte insbesondere bei den ersten beiden AOIs zum Ordnungsverhalten ($M = -0.018, SD = 0.097$) und zum regelmäßigen Schulbesuch ($M = 0.005, SD = 0.100$) der größte Pupillendurchmesser im Vergleich zu den anderen Textabschnitten (Min = -0.044, Max = -0.086) identifiziert werden. Beim Lesen der positiven und negativen AOIs lag kein signifikanter Unterschied des Pupillendurchmessers vor ($F(1.000, 44.000) = 0.004, p = .95, \eta^2 = .00$). Durchschnittlich lag mit $M = 0.08$ ($SD = 0.12$) eine signifikant höhere Lidschlagrate der Lehramtsstudierenden beim Lesen der Fallvignetten im Vergleich zur Baseline vor ($t(44) = 4.105, p < .001$). Auch hier konnte eine große Spannweite zwischen den Versuchspersonen festgestellt werden (Min = -0.22, Max = 0.37).

Die bivariaten Korrelationen der Einstellungsmasse sowie der okulomotorischen Aktivitäten sind Tabelle 14 zu entnehmen. Dabei zeigen keine signifikanten Zusammenhänge der impliziten Einstellungen mit den expliziten Einstellungsmaßen, sowie mit den okulomotorischen Aktivitäten. Beide explizite Einstellungsmaße hingen moderat miteinander zusammen. Die Korrelationen der okulomotorischen Aktivitäten zeigen keinen Zusammenhang der Lidschlagrate mit der Fixationsdauer oder dem Pupillendiameter. Jedoch erbrachten die Analysen einen negativen Zusammenhang zwischen dem Pupillendurchmesser und der durchschnittlichen Fixationsdauer. Eine längere Fixationsdauer geht somit mit einem geringen Pupillendurchmesser einher. Zudem hing der Pupillendurchmesser negativ mit der expliziten Einstellung gegenüber einer sozial heterogenen Schülerschaft zusammen. Je positiver die Einstellungen der Studierenden gegenüber sozialer Heterogenität im Unterricht war, desto geringer war ihr Pupillendiameter beim Lesen aller Fallvignetten. Da der Pupillendiameter als Indikator für kognitive Aktivierung anzusehen ist, bedeutet dies, dass Studierende mit positiven expliziten Einstellungen gegenüber einer sozial heterogenen Schulklasse eine geringere kognitive Beanspruchung bei der Bearbeitung der Fallvignetten hatten.

Tabelle 11

Bivariate Korrelationen der Einstellungen und okulomotorischen Aktivitäten (relative Fixationsdauer Pupillendiameter und Lidschlagrate im Vergleich zur Baseline; über die gesamten Fallvignetten)

	(1)	(2)	(3)	(4)	(5)
Einstellungen					
(1) Implizite Präferenz deutsche/ türkische Herkunft	-				
(2) Explizite Einstellung ethnische Heterogenität	-.14	-			
(3) Explizite Einstellung soziale Heterogenität	-.02	.30*	-		
Okulomotorische Aktivitäten					
(4) Fixationsdauer	.03	-.02	.29	-	
(5) Pupillendiameter	-.24	-.16	-.30*	-.47**	-
(6) Lidschlagrate	.16	-.03	-.08	-.02	.02

Anmerkungen. ** $p < .01$, * $p < .05$, $N = 45$.

Die implizite Präferenz gegenüber der deutschen/türkischen Herkunft wurde durch den IAT erfasst. Hohe positive Werte bedeuten dabei eine Präferenz gegenüber der deutschen Herkunft im Vergleich zur türkischen Herkunft. Hohe negative Werte hingegen eine Präferenz gegenüber der türkischen Herkunft im Vergleich zur deutschen Herkunft.

Hypothesenprüfende Analysen⁸⁶

Hypothese 2.1

Um zu prüfen, ob sich bei stereotypinkonsistenten Informationen eine höhere visuelle Aufmerksamkeit zeigt als bei stereotypkonsistenten Informationen ($H_{2.1}$), wurde die durchschnittliche Fixationsdauer in Relation zur Größe der AOIs in Pixel analysiert. Angenommen wurde, dass positive Textfragmente als stereotypkonsistent für Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status angesehen werden und negative Textfragmente stereotypkonsistent für Schüler mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status sind. Durch einen Vergleich der Mittelwerte zwischen den Herkunftskombinationen in den Fallvignetten, konnten keine Unterschiede in der relativen Fixationsdauer in den einzelnen positiven AOIs festgestellt werden (vgl. Abbildung 18).

⁸⁶ Durch die gerichteten Hypothesen werden einseitige Signifikanzniveaus berichtet.

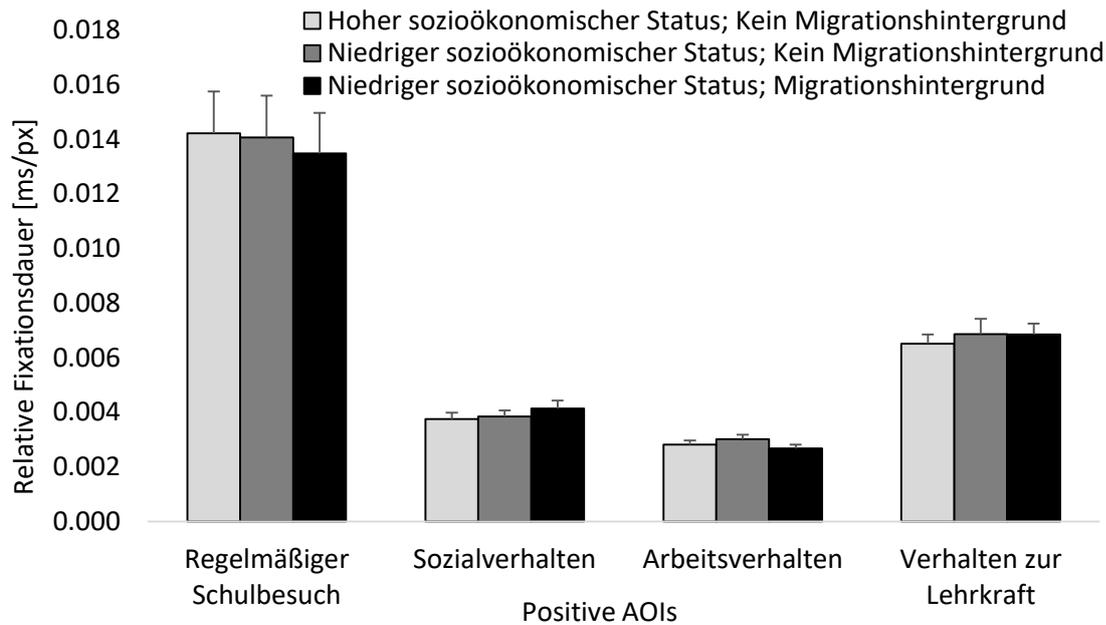


Abbildung 17. Mittelwerte und Standardfehler der relativen Fixationsdauer pro Pixel (ms/px) in positiven AOIs nach Herkunftskombination in den Fallvignetten

Die Varianzanalyse mit Messwiederholung des positiven regelmäßigen Schulbesuchs ($F(2.000,87.983) = 0.052, p = .47, \eta^2 = .00$), des positiven Sozialverhaltens ($F(1.897,83.479) = 0.828, p = .22, \eta^2 = .02$), des positiven Arbeitsverhaltens ($F(1.875,82.518) = 1.923, p = .08, \eta^2 = .04$) sowie des positiven Verhaltens gegenüber der Lehrkraft ($F(1.854,81.597) = 0.253, p = .38, \eta^2 = .01$) bestätigten die deskriptiven Befunde. Lediglich für das positive Arbeitsverhalten zeigte sich eine Tendenz für eine unterschiedliche relative Fixationsdauer nach Herkunftskombination. Die a priori Helmert-Kontrastanalysen erbrachten keinen Effekt des sozioökonomischen Status ($F(1,44) = 0.039, p = .42, \eta^2 = .00$), sondern des Migrationshintergrunds ($F(1,44) = 4.248, p < .05, \eta^2 = .09$). Dabei betrachteten die Lehramtsstudierenden die Informationen zum positiven Arbeitsverhalten signifikant länger bei dem Schüler ohne Migrationshintergrund und niedrigem sozioökonomischem Status als bei dem Schüler mit Migrationshintergrund und niedrigem sozioökonomischem Status. Die Analyse der negativen AOIs erbrachte ebenfalls keine signifikanten Unterschiede in der Fixationsdauer nach Herkunftskombination der Schüler in den Fallvignetten (vgl. Abbildung 19). Weder für Textfragmente zum negativen Ordnungsverhalten ($F(1.867,82.147) = 1.953, p = .08, \eta^2 = .04$), dem negativen Sozialverhalten ($F(1.991,87.591) = 0.089, p = .46, \eta^2 = .00$) noch bei dem negativen Arbeitsverhalten ($F(1.962,86.346) = 0.077, p = .46, \eta^2 = .00$) der Schüler zeichneten sich signifikante Effekte der Schülerherkunft in den Fallvignetten auf die relative Fixationsdauer ab. Für das Ordnungsverhalten ist lediglich eine Tendenz zu erkennen. Anschließende a priori Helmert-Kontrastanalysen erbrachten dabei einen Effekt des sozioökonomischen Status ($F(1,44) = 4.064, p < .05, \eta^2 = .09$), jedoch nicht des Migrationshintergrunds ($F(1,44) = 0.027, p = .44, \eta^2 = .00$). Entgegen der Erwartung zeigte sich eine geringere relative Fixationsdauer beim Lesen der Fallvignette mit hohem Status im Vergleich zu den Vignetten mit niedrigem Status. Zudem wurde für jede Herkunftskombination der Mittelwert für alle positiven sowie alle negativen AOIs berechnet (vgl. Abbildung 20).

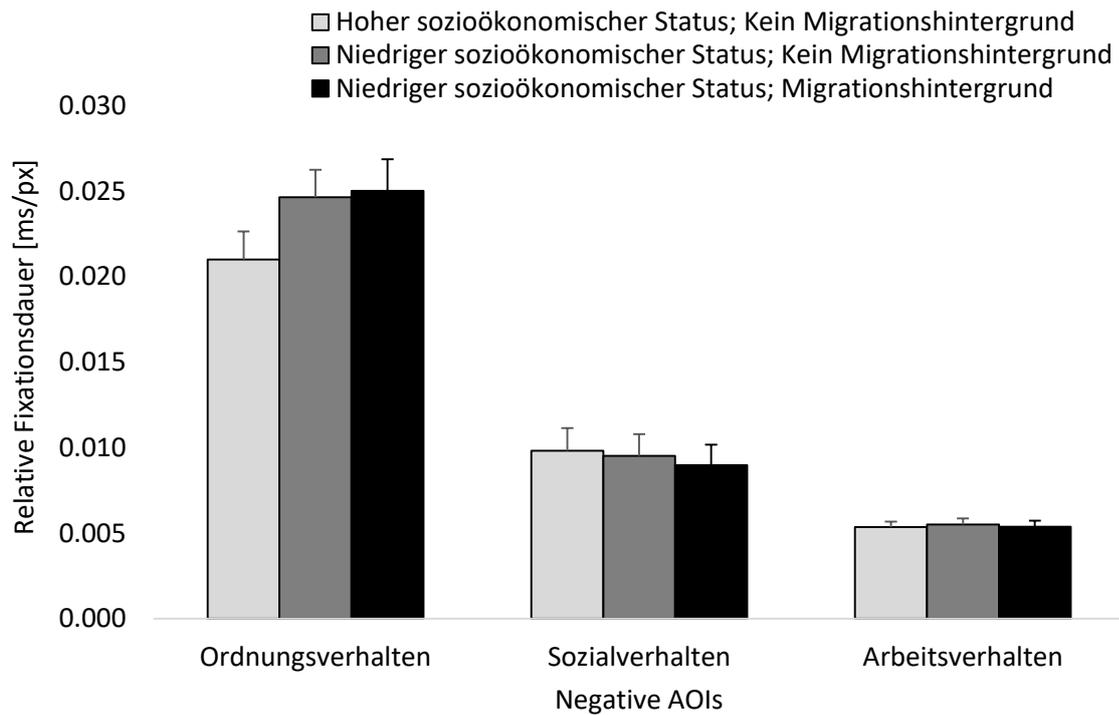


Abbildung 18. Mittelwerte und Standardfehler der Fixationsdauer in negativen AOIs nach Herkunftskombination in den Fallvignetten

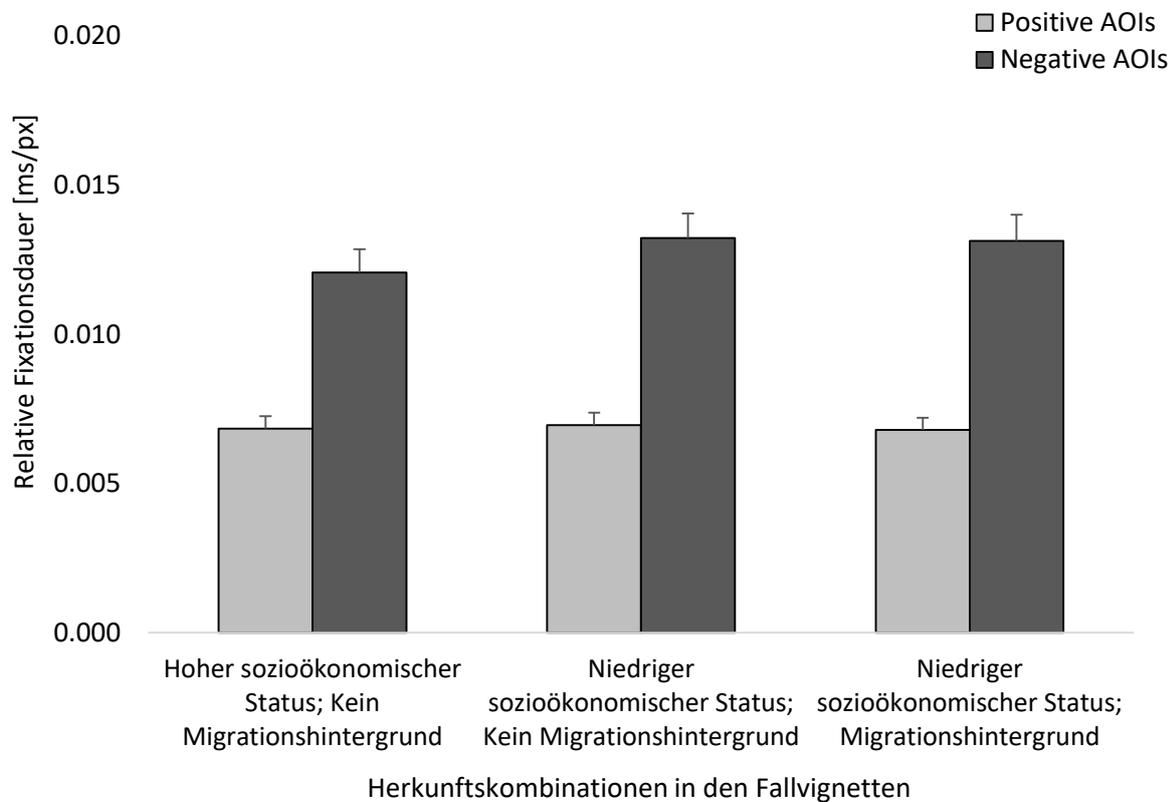


Abbildung 19. Durchschnittliche relative Fixationsdauer und Standardfehler in positiven und negativen AOIs nach Herkunftskombination in den Fallvignetten

Dabei zeigte sich in den Varianzanalysen mit Messwiederholung für alle Herkunftskombinationen eine signifikant längere relative Fixationsdauer bei negativen Textinformationen im Vergleich zu positiven Textinformationen (hoher sozioökonomischer Status, kein Migrationshintergrund: $F(1.000,44.000) = 69.241, p < .001, \eta^2 = .61$; niedriger sozioökonomischer Status, kein Migrationshintergrund: $F(1.000,44.000) = 113.641, p < .001, \eta^2 = .72$, niedriger sozioökonomischer Status, Migrationshintergrund: $F(1.00,44.000) = 112.026, p < .001, \eta^2 = .72$), jedoch kein Unterschied im Kontext der angenommenen herkunftsassoziierten Unterschiede. Die Hypothese zur höheren visuellen Aufmerksamkeit bei stereotypinkonsistenten Informationen kann anhand der relativen Fixationsdauer nicht bestätigt werden.

Hypothese 2.2

Neben der Frage nach der Abhängigkeit visueller Aufmerksamkeit von herkunftsspezifischen stereotypkonformen Textinformationen, verfolgte die Studie das Ziel, die kognitive Aktivierung als Indikator für die Tiefe der Informationsverarbeitung in Abhängigkeit der ethnischen und/oder sozialen Herkunft von Schülern zu analysieren. Die Annahme bestand darin, dass als Zeichen für eine eher kontrollierte Informationsverarbeitung ein größerer Pupillendurchmesser sowie eine erhöhte Lidschlagrate beim Lesen der Fallvignetten der Schüler mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status im Vergleich zu Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status aufzufinden sei ($H_{2.2}$).

Zur Prüfung dieser Hypothese wurden zunächst die Pupillendiameter für die AOIs der Verbal- und Ziffernbewertung in Relation zur Baseline in Abhängigkeit der Herkunftskombinationen der Fallvignetten analysiert. Abbildung 21 sind die durchschnittlichen Abweichungen des Pupillendiameters beim Lesen der verbalen Beurteilungen der Fallvignetten nach AOIs und Herkunftskombination der Fallvignetten zu entnehmen. Insgesamt konnte, nach einem kurzen Anstieg des Pupillendiameters nach den ersten beiden AOIs, eine kontinuierliche Abnahme des Pupillendurchmessers festgestellt werden. Die Analysen erbrachten für die erste AOI der Fallvignetten (in denen das negative Ordnungsverhalten der Schüler beschrieben wurde) einen signifikanten Unterschied nach Schülerherkunft ($F(1.989,87.531) = 3.478, p < .05, \eta^2 = .07$), wobei die Helmert-Kontrastanalysen keinen signifikanten Unterschied ($F(1,44) = 0.826, p = .18, \eta^2 = .02$) nach sozialer Herkunft der Schüler, jedoch nach ethnischer Herkunft der Schüler ($F(1,44) = 6.548, p < .01, \eta^2 = .13$) aufzeigten. Für die zweite AOI der Fallvignetten (positive Informationen über den regelmäßigen Schulbesuch des Schülers) konnte ebenfalls ein signifikanter Effekt der Schülerherkunft auf die durchschnittliche Pupillengröße pro AOI ermittelt werden ($F(1.899,83.546) = 3.064, p < .05, \eta^2 = .07$). Dabei zeigte sich beim Lesen der positiven Informationen zum regelmäßigen Schulbesuch sowohl ein signifikanter Unterschied nach der sozialen Herkunft in den Fallvignetten ($F(1,44) = 3.024, p < .05, \eta^2 = .05$) als auch ein signifikanter Unterschied nach ethnischer Herkunft der Schüler in den Fallvignetten ($F(1,44) = 3.096, p < .05, \eta^2 = .07$). Für die ersten beiden AOIs der Fallvignetten zeigte sich somit, dass der Pupillendurchmesser beim Lesen der Fallvignette der Herkunftskombination türkischer Migrationshintergrund und niedriger sozioökonomischer Status höher war als beim Lesen der vergleichbaren Textinformationen bei Schülern ohne

Migrationshintergrund. Zudem erbrachten die Analysen der zweiten AOI einen Effekt der sozialen Herkunft. Beim Lesen der Fallvignette mit hohem sozioökonomischem Status war der Pupillendiameter geringer als beim Lesen der beiden Fallvignetten mit niedrigem sozioökonomischem Status.

Die Analysen des positiven Sozialverhaltens ($F(1.797,79.083) = 0.186, p = .40, \eta^2 = .00$) sowie des negativen Sozialverhaltens ($F(1.793,78.909) = 1.440, p = .12, \eta^2 = .03$) erbrachten hingegen keine signifikanten Unterschiede nach Schülerherkunft. Auch für das Arbeitsverhalten konnten weder für positive AOIs ($F(1.663,73.177) = 0.189, p = .39, \eta^2 = .00$) noch für negative AOIs ($F(1.700,74.786) = 0.477, p = .24, \eta^2 = .02$) signifikante Unterschiede in dem durchschnittlichen Pupillendiameter ermittelt werden. Die Analysen erbrachten ebenfalls keinen signifikanten Effekt der Herkunft beim Lesen der Informationen zum Verhalten der Schüler gegenüber der Lehrkraft auf den Pupillendurchmesser ($F(1.421,62.512) = 0.161, p = .39, \eta^2 = .00$).

Im Gegensatz zum Verlauf der Abnahme des Pupillendurchmessers beim Lesen der Verbalbeurteilung nahm beim Lesen der Ziffernbewertung und den Informationen zu den Notendurchschnitten der Pupillendiameter zu (vgl. Abbildung 22). Sowohl für das Lesen der Deutschnote ($F(1.957,86.105) = 2.675, p < .05, \eta^2 = .06$), der Englischnote ($F(1.783,78.438) = 2.421, p < .05, \eta^2 = .05$) sowie der Note in Musik ($F(1.958,86.152) = 2.955, p < .05, \eta^2 = .06$) als auch im Fach Werken und textiles Gestalten ($F(1.824,80.245) = 2.680, p < .05, \eta^2 = .06$) war der Effekt der Schülerherkunft auf den Pupillendurchmesser signifikant.

Die Helmert-Kontrastanalysen erbrachten bei den Noten der Fächer Deutsch, Heimat- und Sachunterricht, Englisch sowie Werken und textiles Gestalten einen signifikanten Effekt der ethnischen Herkunft in den Fallvignetten (Deutsch: $F(1,44) = 3.380, p < .05, \eta^2 = .08$; Englisch: $F(1,44) = 3.289, p < .05, \eta^2 = .07$; Werken und textiles Gestalten: $F(1,44) = 3.067, p < .05, \eta^2 = .07$) jedoch nicht der sozialen Herkunft (Deutsch: $F(1,44) = 1.927, p = .09, \eta^2 = .04$; Englisch: $F(1,44) = 0.751, p = .20, \eta^2 = .02$; Werken und textiles Gestalten: $F(1,44) = 2.211, p = .07, \eta^2 = .05$). Lediglich beim Lesen der Note für den Musikunterricht zeigte sich ein Effekt der sozialen ($F(1,44) = 3.783, p < .05, \eta^2 = .08$) Herkunft (Kontrast der ethnischen Herkunft: $F(1,44) = 2.228, p = .07, \eta^2 = .05$).

Hingegen konnten bei der Mathematiknote ($F(1.901,83.636) = 0.054, p = .47, \eta^2 = .00$), der Note im Heimat- und Sachunterricht ($F(1.839,80.936) = 2.227, p < .06, \eta^2 = .05$), der Religionsnote ($F(1.989,87.525) = 1.844, p = .08, \eta^2 = .04$) sowie der Kunst- ($F(1.807,79.511) = 0.293, p = .36, \eta^2 = .01$) und Sportnote ($F(1.958,86.158) = 1.804, p < .09, \eta^2 = .04$) keine signifikanten Herkunftseffekte gefunden werden. Auch die Analyse des Pupillendurchmessers bei der Betrachtung des gesamten Notendurchschnitts ($F(1.751,77.049) = 1.338, p = .13, \eta^2 = .03$) als auch des Notendurchschnitts in den Hauptfächern ($F(1.810,79.623) = 0.397, p = .33, \eta^2 = .01$) erbrachte keine bedeutsamen Differenzen in Bezug auf die Herkunftskombinationen in den drei Fallvignetten.

Für alle Herkunftskombinationen der Fallvignetten konnte zunächst eine Verringerung des Pupillendurchmessers beim Lesen der Verbalbeurteilung festgestellt werden. Jedoch zeigten sich v. a. in den ersten Textfragmenten der Verbalbeurteilung und der Ziffernbewertung

signifikante Unterschiede im Pupillendurchmesser, die sich durch die ethnische Herkunft der Schüler in den Fallvignetten erklären lassen. Für den Schüler mit Migrationshintergrund war der Pupillendurchmesser sowohl beim Lesen der Informationen zum Schülerverhalten als auch beim Lesen der Noten größer als bei den anderen beiden Schülern. Diese Unterschiede waren jedoch nur in den zuvor genannten Abschnitten statistisch bedeutsam.

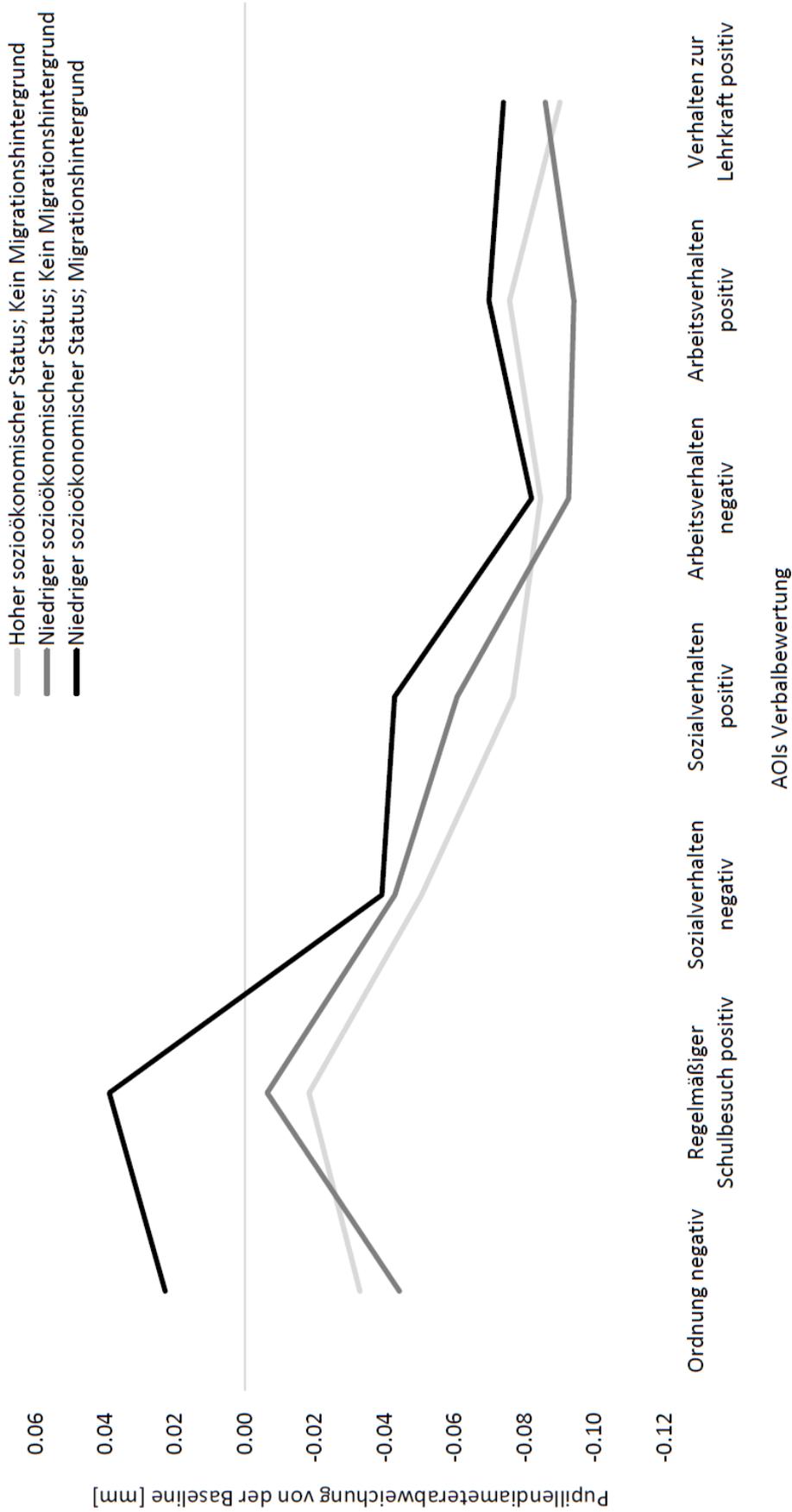


Abbildung 21. Abweichung des durchschnittlichen Pupillendiameters von der Baseline pro AOI in der Verbalbewertung nach Herkunftskombination in den Fallvignetten

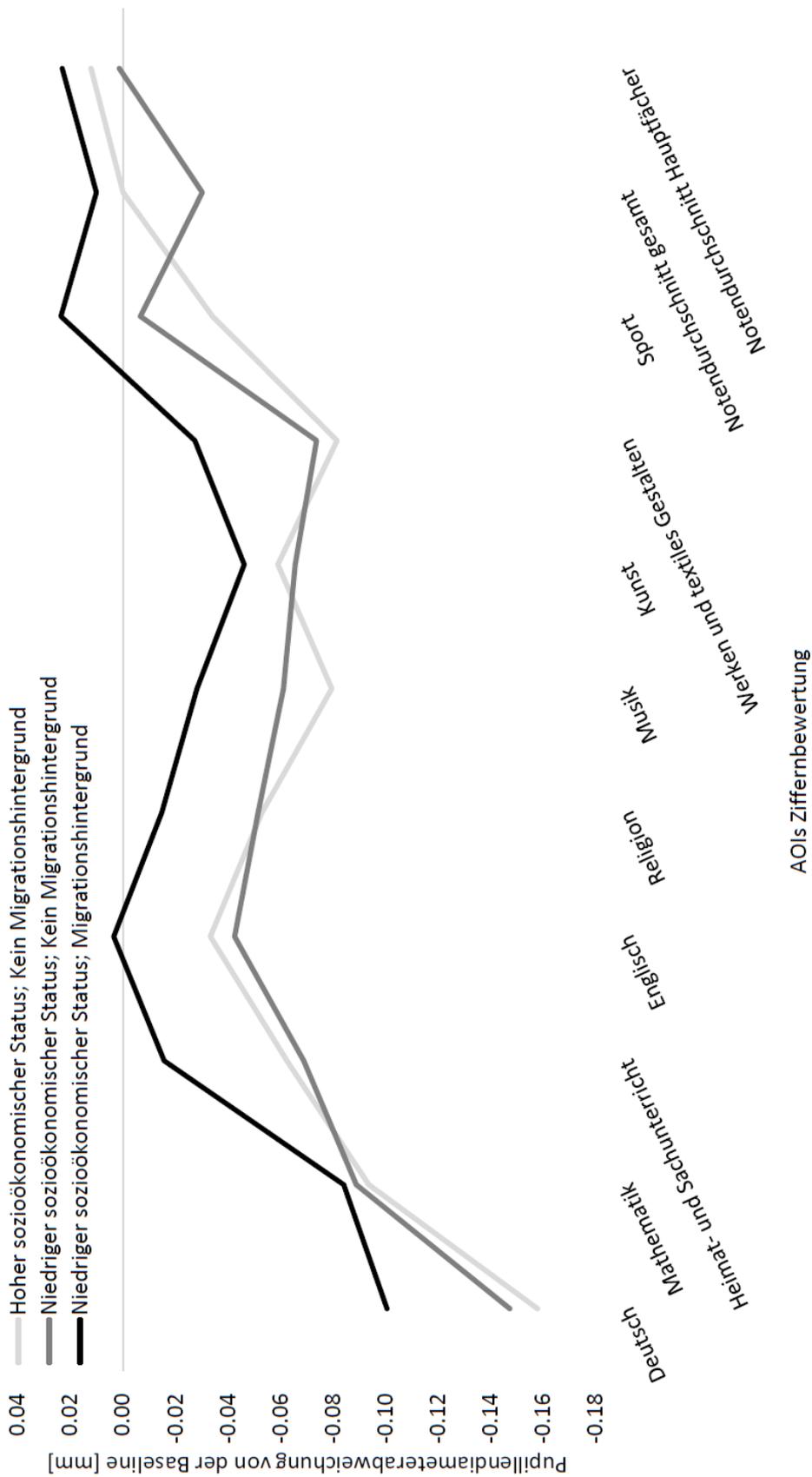


Abbildung 22. Abweichung des durchschnittlichen Pupillendiameters von der Baseline pro AOI in der Ziffernbeurteilung nach Herkunftskombination in den Fallvignetten

Da neben dem Pupillendurchmesser auch die Lidschlagrate als ein Indikator für kognitive Aktivierung und damit auch für die Tiefe der Informationsverarbeitung gilt (vgl. Abschnitt 7.3), wurde diese im Anschluss an die Analysen des Pupillendiameters fokussiert. Zu diesem Zweck wurde für die Analyse herkunftsabhängiger intraindividuelle Unterschiede in der Informationsverarbeitung nach ethnischer und/oder sozialer Herkunft der Schüler Varianzanalysen mit Messwiederholung sowie a priori Helmert-Kontrastanalysen berechnet. Für die Lidschlagrate ergab sich dabei ein bedeutsamer herkunftsabhängiger Unterschied ($F(1,959,86.210) = 3.888, p < .01, \eta^2 = .08$), wobei sich hier ein signifikanter Kontrast der sozialen Herkunft ($F(1,44) = 7.465, p < .01, \eta^2 = .15$) ergab (vgl. Abbildung 23). Dabei zeigte sich für beide Fallvignetten mit niedrigem sozioökonomischem Status eine höhere Lidschlagrate als bei der Fallvignette mit hohem sozialem Status (vgl. Abbildung 23). Dies lässt auf eine höhere kognitive Beanspruchung beim Lesen der Zeugnisse mit niedrigem sozioökonomischem Status, unabhängig von der ethnischen Herkunft, schließen. In Bezug auf die ethnische Herkunft der Schüler in den Fallvignetten, erbrachte die Kontrastanalyse hingegen keine bedeutsamen Befunde ($F(1,44) = 0.121, p = .36, \eta^2 = .00$). Zwar zeigte sich unter allen drei Versuchsbedingungen eine signifikante Erhöhung der Lidschlagrate im Vergleich zur Baseline, jedoch fällt dieser beim Lesen der Vignette des Schülers ohne Migrationshintergrund und mit hohem sozioökonomischem Status ($t(44) = 2.777, p < .01$) im Vergleich zu dem Schüler ohne Migrationshintergrund und mit niedrigem sozioökonomischem Status ($t(44) = 3.754, p < .001$) sowie zu dem mit Migrationshintergrund und mit niedrigem sozioökonomischem Status ($t(44) = 4.461, p < .001$) am geringsten aus.

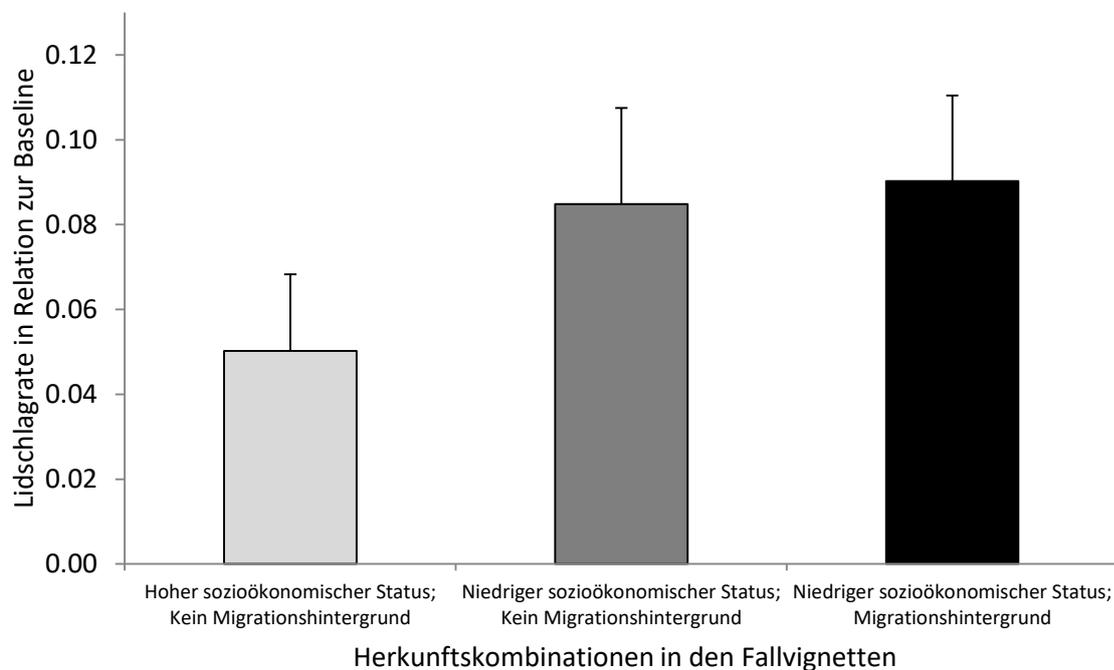


Abbildung 20. Durchschnittliche Lidschlagrate in Relation zur Baseline und Standardfehler nach Herkunftskombination in den Fallvignetten

Die erhöhte Lidschlagrate bei Schülern mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status sowie die teilweise signifikanten größeren Pupillendurchmesser bei Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status bestätigen die Annahme einer höheren kognitiven Beanspruchung bei der Urteils- und Erwartungsbildung teilweise (H_{2.2}). Erwartungsgemäß kann somit von einer höheren kognitiven Beanspruchung beim Lesen der Fallvignetten mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status ausgegangen werden. Die Analyse der Pupillenweite bestätigte eine höhere kognitive Aktivierung beim Lesen der Fallvignetten des Schülers mit Migrationshintergrund im Vergleich zu den Schülern ohne Migrationshintergrund.

Hypothese 2.3

Zusätzlich wurde angenommen, dass sich eine eher automatisierte Informationsverarbeitung beim Lesen einer Fallvignette von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status durch positive implizite Stereotype gegenüber dieser sozialen Gruppe erklären lässt (H_{2.3}). Zur Beantwortung dieser Frage wurden ebenfalls einfaktorielle Varianzanalysen mit drei Faktorstufen berechnet und zusätzlich die impliziten Stereotype, die mit dem IAT erfasst wurden, als Zwischensubjektfaktor eingefügt.⁸⁷ Hierzu wurden das Stereotypenmaß am Median gesplittet und dichotomisiert. Da auch explizite ethnizitäts- und sozialschichtspezifische Einstellungen der Versuchspersonen einen Einfluss haben könnten, wurden diese als Kovariaten berücksichtigt.

Dabei zeigte sich bei der Analyse des durchschnittlichen Pupillendiameters weder in den aggregierten positiven AOIs ($F(1.616,66.246) = 0.998, p = .37, \eta^2 = .05$) noch in den negativen AOIs ($F(1.822,74.708) = 0.689, p = .49, \eta^2 = .03$) eine signifikante Interaktion des Pupillendurchmessers nach Schülerherkunft mit den impliziten Einstellungen. Die Analyse der Lidschlagrate erbrachte hingegen eine signifikante Interaktion zwischen der Lidschlagrate nach Herkunftskombinationen in den Fallvignetten und den impliziten Stereotypen ($F(1.966,80.608) = 2.314, p < .05, \eta^2 = .05$), wobei der Haupteffekt der Herkunft verschwindet. Die Ergebnisse der Kontrastanalyse verdeutlichen, dass es sich um einen Effekt des sozioökonomischen Status ($F(1,41) = 4.756, p < .05, \eta^2 = .10$) und nicht um einen Effekt der ethnischen Herkunft handelt ($F(1,41) = 0.070, p = .40, \eta^2 = .00$). Lehramtsstudierende mit hohen positiven Stereotypen gegenüber Personen ohne Migrationshintergrund bzw. hohen negativen Stereotypen gegenüber Personen mit türkischem Migrationshintergrund hatten beim Lesen der Fallvignetten von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status eine signifikant geringere Lidschlagrate, als Personen mit weniger positiven impliziten Stereotypen (vgl. Abbildung 24). Das Verschwinden des Haupteffekts verdeutlicht, dass die Unterschiede in der Informationsverarbeitung durch die impliziten Stereotype der Studierenden moderiert werden.

⁸⁷ Die Analysen beziehen die expliziten Einstellungsvariablen zwar mit ein, da diese ebenfalls berücksichtigt werden müssen. Da sie jedoch die Einstellung zur Heterogenität und nicht zur spezifischen ethnischen Herkunft erfassen, fokussieren die folgenden Analysen die impliziten Assoziationen, die durch den IAT erfasst wurden.

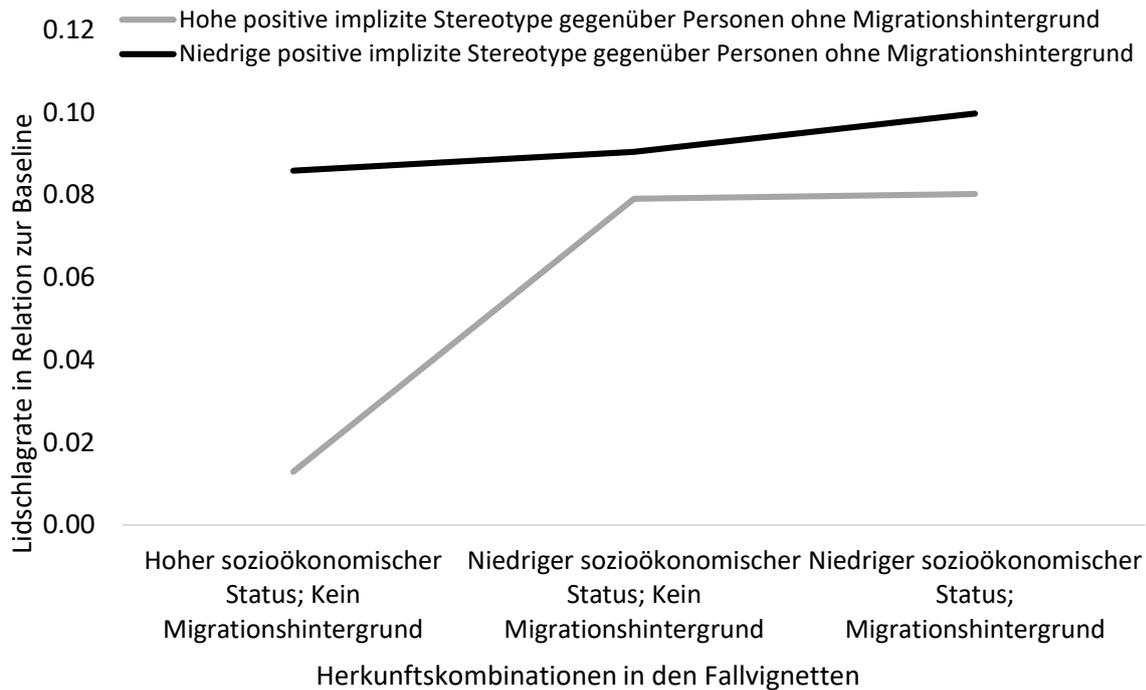


Abbildung 21. Relative Lidschlagrate von Lehramtsstudierenden mit hohen und niedrigen positiven impliziten Stereotypen nach Herkunftscombination in den Fallvignetten

Ergänzende Analysen⁸⁸

Da neben der Prüfung der genannten Hypothesen zusätzlich angenommen wurde, dass implizite Einstellungen auch die visuelle Aufmerksamkeit von Personen beeinflussen, wurden hierzu ergänzende Analysen vorgenommen. Hierfür wurden, vergleichbar mit den hypothesenprüfenden Analysen, einfaktorische Varianzanalysen mit drei Faktorstufen, den dichotomisierten impliziten Einstellungen als Zwischensubjektfaktor und den expliziten Einstellungen zu einer heterogenen Schülerschaft (Migrationshintergrund und soziale Herkunft) berechnet.

Die Analysen für die positiven AOIs zum regelmäßigen Schulbesuch ($F(1.997,81.885) = 0.187, p = .67, \eta^2 = .01$) und Verhalten gegenüber der Lehrkraft ($F(1.899,77.857) = 1.370, p = .25, \eta^2 = .03$) erbrachten keine signifikante Interaktion zwischen der Fixationsdauer und der Schülerherkunft. Für die Fixationsdauer auf dem positiven Sozialverhalten zeigte sich keine signifikante Interaktion mit den impliziten Einstellungen, jedoch ein marginal signifikanter Haupteffekt der Herkunft ($F(1.841,75.475) = 2.823, p = .07, \eta^2 = .06$) sowie ein signifikanter Kontrast der sozialen Herkunft ($F(1,41) = 7.273, p < .01, \eta^2 = .15$). Erwartungsgemäß konnte eine längere Fixationsdauer bei Schüler(inne)n mit niedrigem sozioökonomischem Status (unabhängig von der ethnischen Herkunft) im Vergleich zu Schülern mit hohem sozioökonomischem Status festgestellt werden. Für das positive Arbeitsverhalten zeigte sich unter Berücksichtigung der Einstellungen zwar kein signifikanter Haupteffekt der Schülerherkunft ($F(1.929,79.085) = 1.883, p = .16, \eta^2 = .04$), jedoch ein marginal signifikanter

⁸⁸ Da für die ergänzenden Analysen keine gerichteten Hypothesen formuliert wurden, werden ausschließlich zweiseitige Signifikanzniveaus berichtet.

Interaktionseffekt der Fixationsdauer nach Herkunftskombinationen mit den impliziten Einstellungen ($F(1.929,79.085) = 2.632, p = .08, \eta^2 = .06$). Lehramtsstudierende mit hohen IAT-Werten zeigten dabei eine geringere Fixationsdauer bei Schülern mit hohem sozioökonomischem Status als Lehramtsstudierende mit niedrigeren IAT-Werten. Für die negativen AOIs Ordnungsverhalten (Haupteffekt: $F(1.842,75.522) = 0.637, p = .52, \eta^2 = .02$), Sozialverhalten (Haupteffekt: $F(1.986,81.417) = 0.269, p = .77, \eta^2 = .01$) und Arbeitsverhalten (Haupteffekt: $F(1.950,79.945) = 1.792, p = .17, \eta^2 = .04$) zeigten sich unter Berücksichtigung der Einstellung hingegen keine Tendenzen für die Haupteffekte der Herkunft oder Interaktionseffekte mit den impliziten Einstellungen.

Da angenommen werden könnte, dass implizite Einstellungen auch einen Einfluss auf die Urteile haben können (vgl. z. B. Bargh et al., 1996), wurden zudem eine mögliche Moderation der impliziten Einstellungen auf die Urteile und Erwartungen der Lehrkräfte berücksichtigt. Dabei erbrachten die Analysen jedoch keine signifikanten Interaktionen (vgl. Appendix E; Tabelle E6). Auch die Annahme, Lehramtsstudierende würden ihre visuelle Aufmerksamkeit zwar nicht nach stereotypkonformen Informationen richten, aber bei Schülern mit Migrationshintergrund eine generell höhere visuelle Aufmerksamkeit zeigen um zu vermeiden vorurteilsbehaftet zu urteilen, konnten nicht bestätigt werden. So zeigte sich kein signifikanter Unterschied in der Fixationsdauer nach Schülerherkunft ($F(1.984,87.313) = 1.442, p = .24, \eta^2 = .06$).

9.8 Diskussion von Studie 2

Ziel dieser Studie war die Analyse der Informationsverarbeitungsprozesse von Lehramtsstudierenden bei der Urteilsbildung über Schüler mit und ohne Migrationshintergrund sowie über Schüler mit niedrigem und hohem sozioökonomischem Status. Im Rahmen einer experimentellen Laborstudie im within-subject-Design wurden 45 Lehramtsstudierende um die Beurteilung von jeweils drei Fallvignetten in Form von Grundschulzeugnissen gebeten. Die ethnische und soziale Herkunft der Schüler wurde über die Vornamen in den Fallvignetten manipuliert. Während des Lesens der Fallvignetten wurden über einen Eye-Tracker die Fixationsdauer, die Pupillenmotorik sowie die Lidschlagrate als Indikatoren für kognitive Prozesse erfasst. Zudem bearbeiteten die Studierenden einen IAT, um ihre impliziten Assoziationen zu türkischstämmigen Personen im Vergleich zu Personen ohne Migrationshintergrund zu erfassen. Ergänzend wurden auch die expliziten Einstellungen in Form der Valenz gegenüber einer heterogenen Schülerschaft in Bezug auf die kulturelle und soziale Herkunft von Schüler(inne)n berücksichtigt.

Insgesamt fielen die Urteile über und Erwartungen an Leistungen und leistungsrelevante Merkmale der Schüler in den Fallvignetten positiv aus. Dabei konnte eine akkurate Einschätzung des Schülers mit Migrationshintergrund und niedrigem sozioökonomischem Status sowie eine positive Urteilsverzerrung für Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status wie in Studie 1 (vgl. auch Kaiser et al., 2016; Ready & Chu, 2015) jedoch nicht repliziert werden. Stattdessen erwiesen sich die Einschätzungen der Studierenden weitestgehend unabhängig von der Schülerherkunft, wobei die Studierenden für Schüler mit niedrigem sozioökonomischem Status eine höhere Leistungserwartung zeigten und

Schülern mit Migrationshintergrund ein höherer subjektiver Wert der Schule zugeschrieben. Auch die impliziten Einstellungen hatten keinen moderierenden Einfluss auf die Urteilsbildung der Studierenden. Dies steht allerdings im Einklang mit theoretischen Annahmen, die von einem handlungsmotivierenden, jedoch nicht determinierenden Charakter von Einstellungen ausgehen (z.B. Eagly & Chaiken, 1993). So können u.a. soziale Normen dazu führen, dass kein einstellungskonformes Verhalten gezeigt wird (Ajzen & Fishbein, 1980; Fazio et al., 1989). In Bezug auf die vorliegende Studie, wäre dies ebenfalls denkbar, da soziale Normen eine herkunftsunabhängige Bewertung fordern und dies insbesondere durch die soziale Erhebungssituation (Einzeltermine mit anwesender Testleitung) präsent sein könnte.

Erwartungsgemäß hingen die impliziten Einstellungen gegenüber Personen mit und ohne Migrationshintergrund nicht mit den expliziten Einstellungen zu einer ethnisch und sozial heterogenen Schülerschaft zusammen. Dieser Befund deckt sich mit den Ergebnissen zahlreicher anderer Studien (vgl. Meta-Analyse von Hofmann et al., 2005). In der vorliegenden Arbeit könnte dies in erster Linie daran liegen, dass den Messverfahren unterschiedliche Konstrukte zugrunde liegen (vgl. auch Hofmann et al., 2005). So erfasst der IAT implizite positive oder negative Assoziationen zu Personen ohne Migrationshintergrund oder zu Personen mit türkischem Migrationshintergrund. Die beiden expliziten Maße erfassen hingegen Einstellung zur Heterogenität, also Vielfältigkeit von Schüler(innen) einer Klasse in Bezug auf entweder ihre ethnische oder soziale Herkunft. Zwar besteht eine Verknüpfung zwischen ethnischer Heterogenität und einem türkischen Migrationshintergrund, dennoch beschreibt die Einstellung zur Heterogenität den Aspekt der Ausprägung vielfältiger ethnischer Herkunftsmerkmale im Gegensatz zu positiven bzw. negativen Assoziationen zu einer spezifischen ethnischen Gruppe. Dabei könnten jedoch auch weitere Faktoren einen Einfluss haben, wie z.B. das Personen selbst keinen Zugriff auf implizite Einstellungen haben (Übersicht über diese und weitere Erklärungen zu keinen oder nur geringen Korrelationen zwischen impliziten und expliziten Einstellungsmaßen bei Hofmann et al., 2005).

Die Analysen der relativen Fixationsdauer auf den AOIs zeigten zunächst, dass die Studierenden die ersten Informationen in den Fallvignetten (Ordnungsverhalten und regelmäßiger Schulbesuch) deutlich länger fixierten. Es ist davon auszugehen, dass die ersten Schülerinformationen auch die höchste Bedeutsamkeit für die Eindrucksbildung haben. So könnten die Ergebnisse auch im Einklang mit den Befunden von Anderson (1965) zum primacy-effect interpretiert werden. Zunächst wurde durch die Analyse der durchschnittlichen relativen Fixationsdauer auf positiven und negativen Textfragmenten geprüft, ob Lehramtsstudierende eine längere Fixationsdauer als Indikator für höhere visuelle Aufmerksamkeit auf stereotypinkonsistenten Informationen zeigen als auf stereotypkonsistenten Informationen. Unabhängig von der Schülerherkunft in den Fallvignetten wurden die negativen Textabschnitte signifikant länger betrachtet. Weder die ethnische noch die soziale Herkunft der Schüler hatte dabei jedoch einen Einfluss auf die Fixationsdauer stereotypkonsistenter oder -inkonsistenter Informationen. Die Annahme einer generell höheren visuellen Aufmerksamkeit bei stereotypinkonsistenten Informationen konnte durch diese Analysen somit nicht bestätigt werden. Die ergänzenden Analysen erbrachten unter Berücksichtigung der Einstellungen dabei teilweise Tendenzen für eine längere Fixationsdauer auf stereotypinkonsistenten Textabschnitten für Schüler mit Migrationshintergrund und/oder

niedrigem sozioökonomischem Status bei Personen mit hohen IAT-Werten als bei Personen mit niedrigen IAT-Werten. Für die negativen Textabschnitte konnten keine signifikanten Unterschiede in der Fixationsdauer unter Berücksichtigung der Einstellungen gezeigt werden.

Insgesamt konnte in dieser Studie somit keine Bestätigung für eine herkunftsassoziierte Aufmerksamkeit der Versuchspersonen identifiziert werden. Dies könnte dadurch zu erklären sein, dass Studierende unabhängig von der Schülerherkunft alle Informationen für ihre Urteilsbildung heranziehen. Im Einklang mit dem Modell der adaptiven diagnostischen Kompetenz von Böhmer et al. (2017) könnte davon ausgegangen werden, dass Studierende als Novizen nicht zwischen relevanten und irrelevanten Informationen differenzieren und alle verfügbaren Informationen in den Urteilsbildungsprozess integrieren. Angenommen werden könnte, dass unter eingeschränkten zeitlichen und/oder kognitiven Kapazitäten der Studierenden dennoch eine eher kategoriebasierte Informationsverarbeitung abläuft, die die Aufmerksamkeit auf eher stereotypinkonsistente Informationen lenkt. Entsprechend müsste dies in weiteren Studien durch eine gezielte Manipulation z. B. der Lesezeit geprüft werden. Zudem sollte geklärt werden, ob sich bei erfahrenen Lehrkräften, die unter bestimmten Bedingungen zwischen einer kategorie- und individuumsbasierten Informationsverarbeitung wechseln können (Böhmer et al., 2017), ein herkunftsassoziiertes Aufmerksamkeitsmuster erkennen lässt.

Unabhängig von der Schülerherkunft lag die Lidschlagrate der Studierenden bei der Bearbeitung der Fallvignetten höher als bei der Baseline, was auf ein gewisses Maß an kognitiver Aktivierung hindeutet. Herkunftsabhängige Unterschiede in der kognitiven Aktivierung bei der Beurteilung konnten durch die Lidschlagrate und teilweise durch die Pupillendiameter partiell bestätigt werden. So zeigte sich erwartungskonform eine höhere Lidschlagrate beim Lesen der Fallvignetten von Schülern mit Migrationshintergrund und/oder hohem sozioökonomischem Status im Vergleich zur Lidschlagrate beim Lesen der Fallvignetten von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status. Bei der Urteilsbildung über Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status konnte die geringste kognitive Aktivierung festgestellt werden. Dies spricht für eine eher automatisierte Informationsverarbeitung. Die höhere Lidschlagrate in den beiden anderen Fallvignetten deutet hingegen auf eine eher kontrollierte Informationsverarbeitung hin.

Vor allem die Interaktion der impliziten Stereotype mit der Lidschlagrate nach Herkunftskombination stützen dabei die Annahme, dass die Effekte sich durch eine eher automatisierte Informationsverarbeitung bei Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status erklären lassen. So zeigten Personen mit hohen IAT-Werten die geringste kognitive Aktivierung bei der Bearbeitung der Fallvignette ohne Migrationshintergrund und mit hohem sozioökonomischem Status, hingegen war ihre Lidschlagrate bei der Beurteilung der anderen beiden Schüler signifikant höher. Personen mit niedrigen IAT-Werten wiesen hingegen eine erhöhte Lidschlagrate bei allen drei Herkunftskombinationen auf. Dies lässt sich insbesondere dann plausibel erklären, wenn ein hoher IAT-Wert nicht als starke Assoziation negativer Attribute mit einem türkischen Migrationshintergrund aufgefasst wird, sondern wenn der IAT-Wert als starke Verknüpfung

positiver Merkmale mit Personen ohne Migrationshintergrund angesehen wird.⁸⁹ Dies spricht dafür, dass positive implizite Stereotype zu einer eher automatisierten Informationsverarbeitung führen. Die Befunde des höheren Pupillendiameters bei einigen Textabschnitten bei Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status sprechen für eine höhere kognitive Aktivierung, die sich eher durch einen Effekt der Ethnizität erklären lässt und für eine eher kontrollierte Informationsverarbeitung bei diesen Schülern spricht. Die ausbleibenden Interaktionseffekte mit den impliziten Einstellungen könnten ebenfalls darauf hindeuten, dass der IAT in diesem Fall eher positive Stereotype gegenüber Personen ohne Migrationshintergrund erfasst hat.

Insgesamt konnte Studie 2 einen bedeutsamen Beitrag für die methodische Weiterentwicklung zur Erfassung kognitiver Prozesse bei der Beurteilung von Schüler(inne)n unterschiedlicher Herkunft leisten. Durch den innovativen methodischen Zugang der okulomotorischen Datenerhebung von Lehrenden zeigt sich ein vielversprechender Ansatz, um eher automatisierte und kontrollierte Informationsverarbeitungsprozesse bei der Schülerbeurteilung zu unterscheiden. Zwar ist die technische Ausstattung mit hohen finanziellen Aufwendungen und die Erhebung sowie die Aufbereitung von Eye-Tracking-Daten mit deutlich höherem Aufwand verbunden als z. B. die Erfassung der Präsentationsdauer eines Stimulus als Indikator für die Lesedauer (vgl. auch Krolak-Schwerdt & Kneer, 2006), jedoch zeigen insbesondere die Befunde der Pupillometrie und der Lidschlagrate (die i. d. R. nur durch den Einsatz eines Eye-Trackers erfasst werden können), dass sich dieser Aufwand durch den hohen Informationsgewinn lohnt.

In Bezug auf die Auswertungsmethoden ist anzumerken, dass durch die Aggregation der Pupillendaten nach AOI ein Informationsverlust zu verzeichnen ist. Da keine zeitliche Synchronisierung der Pupillendaten durch individuelles Leseverhalten (unterschiedliche Lesereihenfolge der AOIs und unterschiedliche Lesegeschwindigkeit) möglich war. Strittig ist auch der Umgang mit lidschlussbedingten fehlenden Werten in den Pupillendaten. Einerseits wird die Interpolation der Datenlücken vorgeschlagen (z. B. Hershman, Cohen & Henik, 2016, o. J.), was im Hinblick auf zeitlich kontinuierliche Auswertungen der Pupillendaten entsprechend sinnvoll ist. Für die Analyse der aggregierten Daten auf AOI-Ebene erscheint ein Ausschluss der Lidschlagsequenzen hingegen ebenfalls als adäquate Lösung.

Die Versuchspersonen wurden darum gebeten, sich und insbesondere ihren Kopf während der Erhebung nicht zu bewegen. Hierdurch kann jedoch nicht vollständig ausgeschlossen werden, dass z. T. auch Artefakte der Pupillenreaktion (z. B. durch die Verengung der Pupille bei der Fixation näherer Objekte; Galley, 2001) auftreten. Da der Erhebungsprozess jedoch von geschulten Testleiter(inne)n begleitet wurde und Veränderungen der Sitzposition der Versuchspersonen dokumentiert wurden, ist davon auszugehen, dass dies in der vorliegenden Studie keinen bedeutsamen Einfluss hatte. Ein weiterer vielversprechender

⁸⁹ So kann ein hoher IAT Wert gleichzeitig zwei Dinge bedeuten (vgl. auch Mierke, 2004), in der vorliegenden Studie konnte es einmal eine Assoziation zwischen Personen mit türkischem Migrationshintergrund und negativen Attributen bedeuten, jedoch auch gleichzeitig eine Verknüpfung von positiven Attributen mit Personen ohne Migrationshintergrund. Entsprechend ist denkbar, dass der IAT Wert durch diese unterschiedlichen Bedeutungen zwei unterschiedliche Aspekte gleichzeitig erfasst und je nach Kontext und Versuchsaufbau interpretiert werden muss.

Ansatz ist die detailliertere Analyse der intrapersonellen kontinuierlichen und diskontinuierlichen Schwankungen des Pupillendurchmessers. Der von Marshall (2002) entwickelte Algorithmus zur Berechnung eines Index der kognitiven Aktivierung unter Berücksichtigung natürlicher Pupillendilatationen wurde in dieser Untersuchung jedoch bewusst nicht genutzt, da durch die Patentierung (Marshall, 2000) keine Informationen über das genaue Vorgehen bzw. die Berechnung des Index verfügbar sind. Somit sind nur bedingt Rückschlüsse möglich, da eine differenzierte Betrachtung des Auswertungsalgorithmus nicht erfolgen kann. Aus diesem Grund ist für zukünftige Forschungen eine alternative Analyse-methode zu wählen.

Die durchschnittlich geringeren Pupillenwerte in der Testphase der Untersuchung im Vergleich zur Baseline verweisen darauf, dass die Testsituation als solche bereits zu einer kognitiven Aktivierung führt und sich die Versuchspersonen womöglich erst mit der Situation vertraut machen müssen. Dies würde auch die größeren Pupillendiameter in den ersten präsentierten Textabschnitten erklären und die darauffolgende durchschnittliche Abnahme der Pupillengröße. Entsprechend sollte in zukünftigen Studien ein deutlich längerer Zeitraum für die Baseline-Messung eingeplant werden. Dabei sollte in Erwägung gezogen werden, dass diese nicht an die Instruktion der Aufgaben geknüpft wird, da diese entsprechend auch zu einer kognitiven Aktivierung führt. Somit lassen die Ergebnisse nur Rückschlüsse auf eine relative kognitive Aktivierung der Versuchspersonen im Vergleich zur kognitiven Aktivierung zur Baseline zu. Um zu erfassen, ob eine kognitive Aktivierung im Vergleich zu keiner kognitiven Aktivierung stattfindet, müsste zunächst eine Baseline-Erhebung erfolgen, in der die Versuchspersonen über einen längeren Zeitraum in völliger Ruhe sind und nach einer Gewöhnungsphase an die Versuchssituation in völliger Ruhe die Baseline-Messung erfolgen.

Insgesamt liefert diese Untersuchung vielversprechende Hinweise zur herkunfts-assozierten Informationsverarbeitung. So konnten erste empirische Hinweise für eine eher automatisierte Informationsverarbeitung, die im Zusammenhang mit impliziten positiven Stereotypen gegenüber Personen ohne Migrationshintergrund steht, bei der Beurteilung von Schülern ohne Migrationshintergrund und mit einem hohen sozioökonomischen Status erbracht werden. Diese Befunde stehen auch im Einklang mit der positiven Urteilsverzerrung, die in Studie 1 für Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status identifiziert werden konnten. Entsprechend deuten die Befunde dieser Studie darauf hin, dass negative Stereotype gegenüber Schülern mit Migrationshintergrund und/oder niedrigem sozio-ökonomischem Status nicht die Disparitäten in Lehrkrafturteilen und -erwartungen erklären können, sondern eher positive Stereotype in den Blick genommen werden sollten. Entsprechend wäre eine Replikation der Befunde auch mit erfahrenen Lehrkräften anzustreben.

10. Zusammenfassung und Gesamtdiskussion der empirischen Untersuchungen dieser Arbeit

Zusammenfassung

Zahlreiche empirische Befunde verdeutlichen, dass insbesondere Schüler(innen) mit (v. a. türkischem) Migrationshintergrund und/oder niedrigem sozioökonomischen Status in ihrem Bildungsverlauf und -erfolg deutlich schlechter abschneiden als Schüler(innen) ohne Migrationshintergrund und/oder hohem sozioökonomischen Status (z. B. Mehringer, 2013; Stubbe et al., 2012). Da diese Schülergruppen einen bedeutsamen Anteil der Schülerschaft im deutschen Bildungssystem ausmachen (Maaz et al., 2016; von zur Gathen & Liebert, 2016) und nicht ausschließlich von primären Herkunftseffekten auszugehen ist, hatte die vorliegende Arbeit das Ziel, herkunftsassoziierte Lehrkrafturteile und -erwartungen sowie die zugrundeliegenden kognitiven Prozesse zu analysieren.

Die Fokussierung auf die Bedeutung von Lehrkräften und Lehramtsstudierenden in diesem Zusammenhang erfolgte, da Lehrkräfte einerseits einen bedeutsamen Einfluss auf den Lernerfolg von Schüler(inne)n haben (vgl. z. B. Hattie, 2015; Lipowsky, 2006) und andererseits für einen adaptiven Unterricht sowie eine individuelle Förderung Schülerleistungen und leistungsrelevante Schülermerkmale akkurat einschätzen müssen (vgl. z. B. Helmke, 2009; Schrader & Helmke, 2001). Empirische Studien verweisen jedoch auf eine häufig nur moderat ausgeprägte diagnostische Kompetenz (vgl. z. B. Südkamp et al., 2012) und zeigen, dass Lehrkrafturteile und -erwartungen durch leistungsirrelevante Aspekte verzerrt sein können (z. B. Dunkake et al., 2012).

Bisherige empirische Studien verweisen zudem darauf, dass Lehrkräfte und Lehramtsstudierende für Schüler(innen) mit Migrationshintergrund und/oder niedrigem sozioökonomischen Status negativere Einstellungen und Stereotype aufweisen als für Schüler(innen) ohne Migrationshintergrund und/oder mit hohem sozioökonomischen Status (z. B. Glock & Karbach, 2015; Glock et al., 2016). Des Weiteren deuten Untersuchungen auf unterschiedliche Lehrkrafturteile und -erwartungen an Leistungen in Abhängigkeit der ethnischen und sozialen Herkunft der Schüler(innen) hin (z. B. Glock & Krolak-Schwerdt, 2013; Sprietsma, 2013). Die empirische Befundlage hierzu ist jedoch heterogen und insbesondere im deutschsprachigen Raum noch relativ gering. Aufbauend auf diesen Befunden bestand eine zentrale Annahme der vorliegenden Arbeit darin, dass Lehrkrafturteile und -erwartungen auch durch herkunftsassoziierte Einstellungen und Stereotype verzerrt sein können. Des Weiteren ist durch die hohe Konfundierung der ethnischen und sozialen Herkunft (vgl. z. B. Herwartz-Emden et al., 2010) in vielen Studien zu ähnlichen Fragestellungen nicht immer klar ersichtlich, ob herkunftsassoziierte Lehrkrafturteile und -erwartungen durch die ethnische oder soziale Herkunft erklärt werden können. Aus diesem Grund wurden in der vorliegenden Arbeit beide Herkunftsmerkmale berücksichtigt (vgl. auch Hoenig & Wenz, 2013).

Eine weitere zentrale Annahme dieser Arbeit bestand darin, dass sich herkunftsabhängige Urteile von Lehrkräften durch eine Über- bzw. Unterschätzung bestimmter Schülergruppen

erklären lassen. Für den sozioökonomischen Status konnten bislang lediglich Befunde für den vorschulischen Kontext im US-amerikanischen Raum erbracht werden, die auf eine Unterschätzung von Kindern mit niedrigem sozioökonomischem Status und eine Überschätzung von Kindern mit hohem sozioökonomischem Status hindeuten (Ready & Chu, 2015; Ready & Wright, 2011). Für Deutschland existiert bisher nur eine Studie von Kaiser et al. (2016), die eine akkuratere Beurteilung von Schüler(inne)n mit Migrationshintergrund feststellen konnten. Basierend auf dem Kontinuum-Modell der Eindrucksbildung von Fiske und Neuberg (1990) wurde im Rahmen dieser Arbeit davon ausgegangen, dass eher kategoriegeleitete Informationsverarbeitungsprozesse für verzerrte Urteile verantwortlich sind und eine eher individuumsbasierte Informationsverarbeitung zu akkurateren Urteilen führt. Eine weitere grundlegende Annahme dieser Arbeit bestand darin, dass sich Informationsverarbeitungsprozesse bei der Schülerbeurteilung nach ethnischer und/oder sozialer Herkunft unterscheiden.

Um diese Annahmen zu prüfen, wurden zwei experimentelle Studien durchgeführt, die herkunftsassoziierte Lehrkrafturteile und -erwartungen sowie die kognitiven Prozesse der Urteilsbildung fokussierten. Beide Studien nutzten Fallvignetten in Form von Halbjahreszeugnissen männlicher Grundschüler der vierten Jahrgangsstufe, in denen durch die Variation der Vornamen die ethnische und soziale Schülerherkunft manipuliert wurde. Das Stimulusmaterial wurde in beiden Untersuchungen durch die Versuchspersonen gut angenommen. Auch die Manipulation der Schülerherkunft kann als gelungen betrachtet werden. In beiden Untersuchungen wurde in Bezug auf die ethnische Herkunft der türkische Migrationshintergrund fokussiert, da bei dieser Schülergruppe die größten Differenzen im Bildungserfolg im Vergleich zu Schülern ohne Migrationshintergrund bestehen (vgl. z. B. Mehringer, 2013).

Die erste Studie konnte zeigen, dass sich die Einschätzung der Leistungserwartung und -aspiration, der Gymnasialeignung, der Anstrengungsbereitschaft sowie der Fähigkeiten durch erfahrene Grundschullehrkräfte sowohl in Abhängigkeit der ethnischen als auch sozialen Herkunft von Schülern unterscheiden. Dabei konnten in allen erfassten Merkmalen für Schüler ohne Migrationshintergrund und hohem sozioökonomischem Status im Vergleich zu Schülern mit türkischem Migrationshintergrund und/oder niedrigem sozioökonomischem Status höhere Erwartungen und bessere Urteile festgestellt werden. Die Befunde zu besseren Leistungseinschätzungen für Schüler ohne Migrationshintergrund und hohem sozioökonomischem Status stehen damit im Einklang mit bisherigen Forschungsarbeiten (z. B. Alvidrez & Weinstein, 1999; Darley & Gross, 1983; Rubie-Davis et al., 2006; Tenenbaum & Ruck, 2007; Tom et al., 1984). In Übereinstimmung mit einer Studie von Sprietsma (2013) konnte ebenfalls festgestellt werden, dass Lehrkräfte Schüler mit türkischem Migrationshintergrund für weniger geeignet für einen Gymnasialbesuch halten, als Schüler ohne Migrationshintergrund. Die vorliegende Arbeit konnte hierzu ergänzend zeigen, dass auch Schüler mit niedrigem sozioökonomischem Status als weniger geeignet für das Gymnasium angesehen werden als Schüler mit hohem sozioökonomischem Status (vgl. auch Hoenig & Wenz, 2013). Die vorliegende Arbeit konnte zudem neue Erkenntnisse zur herkunftsabhängigen Einschätzung weiterer leistungsrelevanter Schülermerkmale liefern. So wurden sowohl die Anstrengungsbereitschaft als auch die Fähigkeiten der Schüler in Abhängigkeit der ethnischen und sozialen Herkunft beurteilt. Diese Befunde verdeutlichen dementsprechend, dass herkunftsabhängige Lehrkrafturteile und -

erwartungen sich nicht ausschließlich auf Leistungen beschränken, sondern auch weitere leistungsrelevante Schülermerkmale verzerrt sind. Entsprechend sollten zukünftige Studien weitere Schülermerkmale berücksichtigen, die z.B. in Anlehnung an das Erwartungs-Wert-Modell (Wigfield & Eccles, 2000) bedeutsam für die Motivation von Schüler(inne)n sind. Aber auch herkunftsassoziierte Urteilsunterschiede für relevante Lernemotionen, wie das Hilflosigkeitserleben oder die Leistungsangst, sollten untersucht werden, da empirische Befunde hierbei auf herkunftsassoziierte Ausprägungen verweisen (Tobisch et al., 2016).

Die erste Studie konnte zudem aufdecken, dass es sich bei herkunftsabhängigen Leistungsbeurteilungen nicht um negativ verzerrte Urteile für Schüler mit Migrationshintergrund und niedrigem sozioökonomischem Status handelt, sondern um positiv verzerrte Urteile bei Schülern ohne Migrationshintergrund und dabei v.a. bei Schülern mit hohem sozioökonomischem Status. Die besonders hohen Lehrkrafterwartungen gegenüber Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status stehen dabei im Einklang mit den Ergebnissen von Ready und Chu (2015) für den vorschulischen Bereich. Auch Glock und Karbach (2015) interpretierten den Unterschied zwischen positiveren Einstellungen gegenüber Schülern ohne Migrationshintergrund und negativeren Einstellungen gegenüber Schülern mit Migrationshintergrund ebenfalls eher als eine positive Verzerrung zugunsten der Schüler ohne Migrationshintergrund. Dies deutet darauf hin, dass Lehrkräfte dazu tendieren, positivere Einstellungen gegenüber Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status in ihre Urteilsbildung einfließen zu lassen. In Anlehnung an das Kontinuum-Modell der Eindrucksbildung (Fiske & Neuberg, 1990) lässt dies auf einen eher automatisierten bzw. kategoriegeleiteten Informationsverarbeitungsprozess schließen. Angenommen werden könnte, dass Lehrkräfte durch positive Assoziationen zu Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status diese Schülergruppe eher überschätzen. Dabei könnte diese Überschätzung in Anlehnung an Pettigrew und Meertens (1995) als verdeckte Vorurteile interpretiert werden, da sich keine Abwertung der Fremdgruppe, jedoch eine Aufwertung der Eigengruppe zeigte.

Eine leichte Überschätzung wird in den Diskussionen zur diagnostischen Kompetenz nicht als negativ betrachtet, da sie auch zu besserer Förderung führen kann und somit auch als Anreiz dienen könnte (vgl. hierzu McElvany et al. 2009). Entsprechend wird eine leichte Überschätzung auch von Lehrkräften nicht zwangsläufig als problematisch angesehen, was dazu führen könnte, dass dies nicht kritisch reflektiert wird und entsprechend in eher automatisierten Prozessen resultiert. Allerdings könnte dies zu einer selbsterfüllenden Prophezeiung werden, wenn Lehrkräfte diese höheren Erwartungen in der Interaktion mit den Schülern zeigen und würde entsprechend zu einer tatsächlichen Leistungssteigerung führen. Problematisch ist dies im Vergleich zur akkuraten Einschätzung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status, die diese erwartungskongruente Förderung entsprechend nicht erfahren würden. Dabei besteht möglicherweise die Gefahr durch die Erwartung, zukünftige Leistungen würden sich nicht verbessern, dass dies zu einem sustaining expectation effect führen könnte. Nehmen Lehrkräfte an, dass die bisherigen Leistungen von Schüler(inne)n beibehalten werden, könnten Leistungsveränderungen möglicherweise nicht bemerkt werden.

Aufbauend auf den Befunden der ersten Studie und in Anlehnung an das Kontinuum-Modell der Eindrucksbildung (Fiske & Neuberg, 1990) ging die zweite Studie der Annahme nach, dass sich die akkurate Einschätzung von Schülern mit Migrationshintergrund durch eine eher individuumsbasierte Informationsverarbeitung erklären lässt. Zudem wurde davon ausgegangen, dass die Überschätzung von Schülern ohne Migrationshintergrund durch eine eher kategoriebasierte Informationsverarbeitung zustande kommt. Hierfür wurden okulo-motorische Daten bei der Urteilsbildung von Lehramtsstudierenden mit einem remote Eye-Tracker erfasst. Die Befunde der Lidschlagrate deuten dabei auf eine geringere kognitive Beanspruchung bei der Beurteilung von Schülern ohne Migrationshintergrund hin. Dies kann als Indikator für eine eher automatisierte Verarbeitung angesehen werden, da diese weniger kognitive Anstrengung erfordert. Die Analyse der Lidschlagrate erbrachte zudem Hinweise auf eine höhere kognitive Beanspruchung bei der Beurteilung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status, was für eine eher kontrollierte Verarbeitung spricht. Diese Befunde bestätigen somit die Annahme, dass unterschiedliche kognitive Prozesse bei der Beurteilung von Schülern in Abhängigkeit ihrer ethnischen und sozialen Herkunft ablaufen. Damit verdeutlichen die Befunde, dass es sich nicht um eine stereotypenbasierte Urteilsbildung bei Schülern mit Migrationshintergrund handelt. Bei Schülern ohne Migrationshintergrund und hohem sozioökonomischem Status können die Befunde hingegen als Prozess interpretiert werden, der sich stärker an den sozialen Kategorien orientiert. Studie 2 konnte zudem zeigen, dass dies im Zusammenhang mit positiven impliziten Einstellungen von Lehramtsstudierenden steht. So zeigte sich eine eher automatisierte Informationsverarbeitung bei der Beurteilung von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status v. a. bei Studierenden, die eine hohe implizite Präferenz gegenüber der deutschen Herkunft im Vergleich zur türkischen Herkunft hatten.

Die beiden experimentellen Studien dieser Arbeit konnten somit Hinweise dazu liefern, dass sowohl für die pädagogische Praxis als auch für die empirische Forschung, die Perspektive für positiv verzerrte Urteile und Informationsverarbeitungsprozesse erweitert werden sollte. Im Kontext beider Studien gibt es auch einige Punkte die kritisch diskutiert werden müssen, was im folgenden Abschnitt erfolgt.

Diskussion und Limitationen

Die in der Arbeit vorgestellten Befunde, konnten Hinweise für einige bis dahin noch offene Fragen in diesem Forschungsbereich liefern. So konnte u. a. gezeigt werden, dass herkunftsassoziierte Urteile und Erwartungen auch bei der Einschätzung motivationaler Aspekte und schulischer Fähigkeiten auftreten. Jedoch müssen im Unterrichtsalltag auch weitere Schülermerkmale durch Lehrkräfte eingeschätzt werden, die in der vorliegenden Arbeit nicht berücksichtigt wurden. Dennoch bekräftigen die vorliegenden Ergebnisse die Forderung, über die in Studien meist fokussierten Leistungen hinauszugehen und weitere, z. B. motivationale und emotionale Aspekte zu berücksichtigen.

Da in beiden Studien der vorliegenden Arbeit das zentrale Stimulusmaterial in Form von Grundschulzeugnissen vorlag, muss auch diskutiert werden inwiefern dieses geeignet ist, um stereotypenbasierte Urteilsbildungsprozesse zu untersuchen. Dabei ist zu beachten, dass schriftliches Stimulusmaterial möglicherweise zu anderen Informationsverarbeitungsprozessen

führen könnte als, z.B. die visuelle Erscheinung einer Person. Macrae, Bodenhausen, Milne Thorn und Castelli (1997) zeigten in ihren Untersuchungen, dass keine automatische Stereotypaktivierung bei Fotografien, jedoch bei semantischem Stimulusmaterial erfolgte. Dies würde dafürsprechen, dass der Einsatz einer schriftlichen Personenbeschreibung (in diesem Fall durch ein Zeugnis und entsprechende Vornamen, die eine spezifische Schülerherkunft implizieren) durchaus ein vielversprechender Ansatz für die Untersuchung stereotypenbasierter Informationsverarbeitungsprozesse darstellt. Anzumerken ist allerdings, dass die Zeugnisse als Stimulusmaterial auch Einschränkungen aufweisen. Zum einen limitieren sie in der vorliegenden Arbeit die Generalisierbarkeit auf den Grundschulkontext, zum anderen handelt es sich um die Beschreibung von fiktiven Schülern. So konnte in den durchgeführten Studien dieser Arbeit lediglich ein reduziertes Schülerbild präsentiert werden, das die Komplexität der sozialen Wahrnehmung – durch die Vielzahl an Sinneseinflüssen (Neisser, 1976) – sowie den prozeduralen Charakter sozialer Situationen nicht berücksichtigen kann. In der Realität ist mit einer deutlich umfassenderen Anhäufung von relevanten und irrelevanten Schülermerkmalen zu rechnen, die die Informationsverarbeitung beeinflussen könnten. Prozesse der Wahrnehmung und Urteilsbildung von Lehrkräften sind im schulischen Alltag entsprechend deutlich komplexer und dynamischer, da auch die Komplexität der Schülerinformationen höher ist. Zudem wird die Wahrnehmung sozialer Informationen, deren Interpretation sowie die Antizipation zukünftiger Leistungen oder Verhaltensweisen durch den kontextuellen Rahmen beeinflusst. Krolak-Schwerdt et al. (2017) konnten jedoch zeigen, dass Lehrkrafturteile sich im realen Kontext nicht bedeutsam von Urteilen in Fallvignetten fiktiver Schüler(innen) unterscheiden. Dies spricht dafür, dass Fallvignetten dennoch eine geeignete und ökologisch valide Alternative zum realen Klassenkontext darstellen. So ermöglichten die Zeugnisse als Stimulusmaterial die Kontrolle der Schülerinformationen und die experimentelle Manipulation der Schülerherkunft. Hierdurch konnte eine hohe interne Validität in beiden Untersuchungen ermöglicht sowie kausale Schlüsse über den Einfluss der Schülerherkunft getroffen werden.

Der Einsatz von Vornamen zur Aktivierung von herkunftsassoziierten Stereotypen konnte bereits in zahlreichen Untersuchungen erfolgreich eingesetzt werden (z. B. Glock & Krolak-Schwerdt, 2013; Hoenig & Wenz, 2013; Sprietsma, 2013). Die gelungene Manipulation der Herkunft durch Vornamen bestätigte sich auch in den Studien dieser Arbeit. Anzumerken ist jedoch, dass ausschließlich das männliche Geschlecht sowie die deutsche und türkische Herkunft durch die Namen vermittelt wurden. Bedingt durch die Assoziation der türkischen Herkunft mit einem eher niedrigen sozioökonomischen Status war es nicht möglich, ohne weitere Informationen zu den Schülern den *Schülertyp türkische Herkunft und hoher sozioökonomischer Status* als weitere Herkunftskombination zu integrieren. Dennoch ermöglichte die Fokussierung der Schülerinformationen auf schulische Aspekte eine subtilere Aktivierung der herkunftsassoziierten Stereotype, da diese nicht weiter expliziert werden mussten, sondern ausschließlich durch den Vornamen erfolgte. Dies hatte den Vorteil, dass nicht weitere zusätzliche Kategorien und Assoziationen durch eine ausführlichere Personenbeschreibung aktiviert wurden und möglicherweise die Rückschlüsse auf soziale oder ethnische Stereotype verzerrten.

Kritisch zu diskutieren ist, dass sich die Ergebnisse der ersten Studie zu einer Überschätzung von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem

Status sowie die akkurate Einschätzung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status nicht in der zweiten Studie replizieren lassen konnten. Hierfür könnten unterschiedliche Faktoren verantwortlich sein. Zum einen sind die unterschiedlichen Designs der Untersuchungen in diesem Zusammenhang zu betrachten. So konnten in der ersten Studie die Versuchspersonen ihre Urteile für sich alleine an einem Computer oder mobilen Gerät fällen, während die Proband(inn)en in Studie 2 sich in einer sozialen Situation (durch die Anwesenheit einer Testleitung) befanden. Somit kommt in der zweiten Studie die Problematik der sozialen Erwünschtheit im Antwortverhalten (*social desirability bias*) hinzu, die besonders wenn es der Forschungsfokus auf Stereotypen und Vorurteilen liegt relevant ist (vgl. Stocké, 2004). Dies könnte erklären, dass Schüler mit Migrationshintergrund in der zweiten Studie tendenziell eher besser beurteilt wurden als Schüler ohne Migrationshintergrund. Bei der Beurteilung von Schülern mit Migrationshintergrund konnte eine höhere kognitive Beanspruchung festgestellt werden, die auf eine kontrollierte Informationsverarbeitung verweist. Diese könnte v. a. im Zusammenhang mit der sozialen Erhebungssituation durch die Anwesenheit einer weiteren Person zu sozial erwünschtem Antwortverhalten geführt haben. Entsprechend könnte die eher kontrollierte Informationsverarbeitung in Kombination mit der tendenziell besseren Beurteilung von Schülern mit Migrationshintergrund als eine Art Überkompensation gedeutet werden. Studie 2 konnte (im Einklang mit Glock & Karbach, 2015) durchschnittlich eher negativere implizite Einstellungen gegenüber Personen mit türkischem Migrationshintergrund im Vergleich zu Personen ohne Migrationshintergrund feststellen, was dazu führen könnte, dass die Lehramtsstudierenden versuchten, dies in den expliziten Urteilen durch eine deutliche bessere Beurteilung zu kaschieren.

In diesem Zusammenhang ist ebenfalls zu diskutieren, dass Studie 1 Grundschullehrkräfte und Studie 2 Lehramtsstudierende aller Schularten untersuchte. Auch dies könnte einen Einfluss auf unterschiedliche herkunftsassoziierte Urteile haben. Im Vergleich zu erfahrenen Lehrkräften scheinen Lehramtsstudierende über einen größeren „pädagogische[n] Optimismus“ (van Ophuysen, 2006, S. 159) zu verfügen, der ebenfalls in der besseren Beurteilung von Schülern mit Migrationshintergrund resultieren könnte. Dabei wäre jedoch anzunehmen, dass ein höherer pädagogischer Optimismus für alle drei Herkunftskombinationen auffindbar sein müsste. Da dies nicht der Fall ist, scheint dies keine ausreichende Erklärung für die Unterschiede zwischen den Ergebnissen der beiden Studien zu liefern.

Neben dem Unterschied der Stichproben in Bezug auf die Erfahrung und Spezifizierung der Schulform ist zudem erwähnenswert, dass unterschiedliche Rekrutierungswege gewählt wurden. Die Lehrkräfte in Studie 1 wurden in Online-Foren für Lehrkräfte angeworben und nahmen freiwillig an der Untersuchung teil. Die Studierenden in Studie 2 wurden hingegen durch Dozierende in ihren Seminaren zur Teilnahme verpflichtet. Insbesondere die freiwillige Teilnahme in Studie 1 ist dabei kritisch zu sehen, da hierdurch bereits eine Selektion von z. B. besonders motivierten Lehrkräften stattgefunden haben könnte. Es könnte angenommen werden, dass die akkurate Beurteilung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status daher kommt, dass die Lehrkräfte, die freiwillig an der Studie teilnahmen, generell motivierter sind sich mit Fragen der Beurteilung zu befassen und bereits sensibilisiert für eine vorurteilsfreie Beurteilung sind. Dies könnte dazu geführt haben, dass nach der Stereotypaktivierung durch den türkischen Vornamen eine bewussterer individuums-

basierte Informationsverarbeitung ablief (die wiederum in akkuraten Urteilen resultiert), um eine stereotypenbasierte Urteilsbildung zu vermeiden. Da positive Vorurteile und Stereotype bislang kaum thematisiert werden und im öffentlichen und wissenschaftlichen Diskurs meist nur negative Urteilsverzerrungen angesprochen werden, wäre anzunehmen, dass der Vorname eines Schülers ohne Migrationshintergrund und mit hohem sozioökonomischem Status zu einer Aktivierung positiver Einstellungen führte, die wiederum in einer Überschätzung der Schüler mündete, da kein Anlass für eine Regulation bestand. Um dies zu klären sind weitere empirische Studien notwendig, die v. a. die Informationsverarbeitungsprozesse von erfahrenen Lehrkräften fokussieren sollten, wobei eine repräsentative Stichprobe angestrebt werden sollte. Hierzu hatte sich in Studie 2 die Erfassung okulomotorischer Daten als vielversprechende Möglichkeit zur Messung von Indikatoren zu unterschiedlichen Informationsverarbeitungsprozessen herausgestellt.

Gleichzeitig gehen mit dieser Methode auch Herausforderungen und Einschränkungen einher, die nicht vernachlässigt werden können. So führt die Laboruntersuchung, in der Versuchspersonen wissen, dass ihre Blickbewegung aufgezeichnet wird, wie bereits erwähnt wurde, zu einem möglicherweise sozial erwünschten Antwortverhalten. Jedoch können Blickbewegungen natürlich auch teilweise bewusst gesteuert werden. So könnten die Lehramtsstudierenden durch die Laborsituation eine tendenziell ohnehin eher kontrollierte Informationsverarbeitung ausgelöst haben, die dazu führt, dass die Studierenden sich bewusst bemühen alle Informationen genau zu lesen. Dies könnte erklären, warum in Studie 2 zunächst keine Hinweise auf eine unterschiedliche Fixationsdauer bei positiven und negativen AOIs in Abhängigkeit der Herkunft gefunden werden konnten. Hinzu kommt in diesem Fall, dass nach dem ADCM (Böhmer et al., 2017) Lehramtsstudierende eine integrierende Informationsverarbeitung nutzen und hierdurch die Lesedauer der AOIs bei allen Schülern gleichermaßen bedeutsam waren. Entsprechend müsste diese Studie mit erfahrenen Lehrkräften repliziert werden. Um eine bewusste Kontrolle der Fixationen zu reduzieren, sollte die Darbietung der Schülerinformationen zeitlich stark eingeschränkt werden. Die Studierenden in Studie 2 konnten in ihrem eigenen Tempo lesen und Absätze wiederholen. Würde auf dem Bildschirm sowohl eine stereotypkonsistente als auch eine inkonsistente Information eingeblendet werden, die bereits nach kurzer Zeit verschwindet, könnten womöglich andere Befunde erbracht werden.

In diesem Zusammenhang sollte zudem diskutiert werden, dass die lange Darbietungszeit auch zu einem vergleichbaren Effekt, wie dem Aufmerksamkeits-Vermeidungs-Muster, geführt haben könnte (vgl. Bean et al., 2012; In-Albon & Schneider, 2012). Zwar beschreibt das Aufmerksamkeits-Vermeidungs-Muster Reaktionen auf negative Emotionen und kann deshalb nicht direkt auf die vorliegende Arbeit übertragen werden, dennoch ist denkbar, dass die Lehramtsstudierenden eine AOI zunächst besonders lange fixiert haben und die AOI anschließend mieden. Eine ausschließliche Analyse der durchschnittlichen Fixationsdauer des unterschiedlichen Stimulusmaterials kann durch die Vernachlässigung des zeitlichen Verlaufs somit zu einer Verwässerung der möglichen Effekte führen. Entsprechend sollten zukünftige Studien die Fixationsdauer detaillierter betrachten.

Auch in Bezug auf die Auswertung der Pupillendurchmesser ist anzumerken, dass einige Informationen zu herkunftsassoziierten Informationsverarbeitungsprozessen durch die Berech-

nung des durchschnittlichen Pupillendiameters pro AOI verloren gegangen sein könnten. Da die Latenzzeit der Pupillenreaktion deutlich geringer ist als die durchschnittliche Fixationsdauer auf einer AOI, könnten nach einer Erweiterung durch kognitive Aktivierung auch relativ schnell wieder eine Abnahme des Pupillendurchmessers erfolgen. Diese Schwankungen sind in den aggregierten Pupillendaten nach AOI nicht mehr ersichtlich, könnten jedoch einen Gewinn an detaillierteren Informationen innehaben. In Zukunft müssten diese Schwankungen der Pupillenweite entsprechend deutlich stärker berücksichtigt werden. Zwar liegt ein vermeintlich methodisch sinnvoller Zugang in der Nutzung des *index of cognitive activity* (ICA; Marshall, 2000, 2002), jedoch ist der zugrundeliegende Algorithmus nicht offengelegt und damit nicht nachvollziehbar, wie die entsprechenden Werte (die Informationen über das Ausmaß der kognitiven Beanspruchung liefern sollen) zustande kommen. Entsprechend wurde in der vorliegenden Arbeit bewusst auf die Nutzung des ICA verzichtet. Um die Auswertung der Pupillendaten zu optimieren, müssten eigene Methoden entwickelt werden, die ein zuverlässiges und transparentes Maß für die kognitive Aktivierung darstellen. Dabei müssten v. a. auch natürliche Schwankungen des Pupillendurchmessers berücksichtigt werden. Trotz dieser methodischen Herausforderungen erwies sich der Einsatz eines Eye-Trackers als geeignetes Instrument um verhaltensnähere Daten zur kognitiven Informationsverarbeitung bei der Schülerbeurteilung zu erfassen.

Abschließend kann trotz einiger Limitationen festgehalten werden, dass die vorliegende Arbeit neue Erkenntnisse zu herkunftsassoziierten Lehrkrafturteilen und -erwartungen liefern konnte. So wurde erstmals gezeigt, dass herkunftsassoziierte Unterschiede in Urteilen und Erwartungen sich v. a. durch eine positive Urteilsverzerrung von Lehrkräften bei der Beurteilung von Schülern der Majorität erklären lassen. Somit kann in Anlehnung an das Modell der Akkuratheit von Lehrkrafturteilen (Südkamp et al., 2012) als Schülermerkmal die ethnische und soziale Herkunft als bedeutsamer Faktor für die Urteilsakkuratheit angesehen werden. Zudem konnte dies durch eine eher kategoriebasierte Informationsverarbeitung von Lehramtsstudierenden bei der Urteilsbildung über Schüler ohne Migrationshintergrund und hohem sozioökonomischem Status bestätigt werden. Zum anderen konnte eine akkurate Beurteilung und eine eher individuumsbasierte Informationsverarbeitung der Lehrkräfte und Lehramtsstudierenden bei Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status gezeigt werden. Ferner konnte durch den Einsatz eines Eye-Trackers im Zusammenhang mit herkunftsassoziierten Informationsverarbeitungsprozessen ein innovativer methodischer Zugang erfolgreich erprobt werden. Auf den Befunden dieser Arbeit sowie einigen Einschränkungen lassen sich einige Forschungsanliegen formulieren, die im Folgenden in einem Überblick dargestellt werden.

Ausblick

Ansetzend an den Befunden der vorliegenden Arbeit ergeben sich einige Forschungsdesiderata. Zunächst bedarf es Replikationsstudien zur akkuraten Einschätzung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status sowie zur Überschätzung von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status. Zudem müssen die Ergebnisse einer eher kontrollierten Informationsverarbeitung bei der Beurteilung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status und eher

automatisierten Informationsverarbeitung bei Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status in weiteren Studien repliziert werden.

Des Weiteren sollte in diesen Studien die Kategorie *der Schüler mit türkischem Migrationshintergrund und hohem sozioökonomischem Status* miteinbezogen werden um zu untersuchen, ob der hohe sozioökonomische Status hier ebenfalls einen Einfluss auf eine Überschätzung hat und zu einer eher automatisierten Informationsverarbeitung führt. Hierfür könnten z.B. auch Fotografien von Schüler(inne)n in Fallvignetten eingesetzt werden. So zeigten Dunkake und Schuchart (2015), dass Lehramtsstudierende Rückschlüsse über den sozioökonomischen Status von Schüler(inne)n an Hand deren Kleidung, Gewicht und Frisur auf Fotografien ziehen. Der Einsatz von Fotografien zur Manipulation der Herkunft würde vermeiden, dass in verbalen Beschreibungen die Herkunftsinformationen der Schüler(innen) zu offensichtlich platziert würden und entsprechende Antwortverzerrungen auftreten könnten. Anzunehmen wäre jedoch, dass die stereotypeninkonsistente Zuschreibung eines hohen sozioökonomischen Status eher zu einer erhöhten Aufmerksamkeit und damit zu einer eher kontrollierten Informationsverarbeitung führen könnte. Auf Grund von Geschlechterstereotypen (z. B. Osad'an, 2012) und der unterschiedlichen Wahrnehmung von Mädchen und Jungen mit Migrationshintergrund durch Lehrkräfte (vgl. Weber, 2003) sollte auch das Geschlecht als soziale Kategorie in zukünftigen Studien berücksichtigt werden. Dabei wäre insbesondere eine Verschränkung der ethnischen und sozialen Herkunft unter Berücksichtigung des Geschlechts in spezifischen (geschlechtsstereotypisierten) Domänen zu fokussieren. Hierbei könnte u. a. geklärt werden, ob sich die Befunde der vorliegenden Arbeit auch für Mädchen unterschiedlicher Herkunft replizieren lassen.

Anzunehmen ist zudem, dass sich auch ein Wandel der Wahrnehmung spezifischer ethnischer Gruppen durch eine Veränderung der Migrantengruppen in Deutschland ergibt. In Deutschland sind 44.3 % der Asylbewerber(innen) unter 18 Jahren (BAMF, 2017b). Deren Bleibeaussichten sind zwar in den meisten Fällen ungewiss, dennoch sind diese Kinder und Jugendlichen mittlerweile im deutschen Bildungssystem angekommen. Da diese Schüler(innen) hauptsächlich aus Syrien, dem Irak, Afghanistan oder Eritrea kommen (BAMF, 2017a), ist seitens der Lehrkräfte mit anderen ethnischen Stereotypen und Einstellungen zu rechnen als bei türkischen Migrant(inn)en, was jedoch zunächst empirisch geprüft werden müsste. Jedoch sollten insgesamt weitere ethnische Gruppen in nachfolgenden Untersuchungen berücksichtigt werden, um zu klären, ob es sich um spezifische Ergebnisse für Schüler(innen) mit türkischen Wurzeln handelt, oder ob sich die Befunde auch für weitere ethnische Gruppen zeigen. Entsprechend sollte auch geklärt werden, ob sich ebenfalls positive Urteilsverzerrungen für erfolgreiche Migrantengruppen in Deutschland finden lassen, wie dies für asiatischstämmige Schüler(innen) in den USA der Fall ist (Chen & Weseley, 2011; Tenenbaum & Ruck, 2007).

Sowohl für zukünftige Studien zu herkunftsassoziierten Urteilsverzerrungen als auch zu den zugrundeliegenden kognitiven Prozessen sollte berücksichtigt werden, dass weitere Faktoren, wie z. B. kognitive Kapazitäten oder Motivation (vgl. z. B. Fiske & Neuberg, 1990), einen Einfluss auf die Informationsverarbeitung haben können. So könnten die zeitlichen und kognitiven Kapazitäten experimentell manipuliert und die Motivation zur Verarbeitung ebenfalls berücksichtigt werden. In Anlehnung an das ADCM (Böhmer et al., 2017) sollte auch die Verantwortung Lehrkräfte für eine Entscheidung tragen sowie die Konsistenz der

Informationen berücksichtigt werden. Da sich in der vorliegenden Arbeit, im Gegensatz zu dem ADCM (Böhmer et al., 2017), herkunftsassoziierte Informationsverarbeitungsprozesse bei Lehramtsstudierenden zeigten, sollte ebenfalls untersucht werden unter welchen Bedingungen sich die Informationsverarbeitungsprozesse von Novizen unterscheiden können. Neben der gezielten experimentellen Manipulation der möglichen Einflussvariablen, sollten zudem Feldstudien durchgeführt werden, um zu prüfen ob sich die Befunde einer akkuraten Einschätzung von Schülern mit Migrationshintergrund und niedrigem sozioökonomischem Status sowie einer Überschätzung von Schülern ohne Migrationshintergrund und mit hohem sozioökonomischem Status auch im realen Unterricht zeigen. Dies würde die externe Validität der Befunde deutlich erhöhen.

Insbesondere der Einsatz von Eye-Trackern zur Erfassung okulomotorischer Aktivitäten um hieraus Rückschlüsse auf kognitive Informationsverarbeitungsprozesse zu ziehen, erweist sich als vielversprechender methodischer Zugang, der in zukünftigen Forschungsarbeiten berücksichtigt werden sollte. Dabei sind insbesondere zwei unterschiedliche Aspekte für zukünftige Studien von Interesse: (1) Es sollten weitere experimentelle Laboruntersuchungen durchgeführt werden, die zunächst die Befunde dieser Arbeit mit erfahrenen Grundschullehrkräften replizieren. Zudem sollte das Stimulusmaterial und der Versuchsaufbau optimiert werden, um detailliertere Aussagen über herkunftsassoziierte Informationsverarbeitungsprozesse treffen zu können. Grundschulzeugnisse stellen ein sowohl sprachlich als auch visuell komplexes Stimulusmaterial dar. Dies bietet einerseits den Vorteil der Realitätsnähe durch die Vertrautheit von Lehrkräften und Lehramtsstudierenden mit diesem Material. Andererseits ist derart komplexes Stimulusmaterial als Herausforderung für Eye-Tracking-Studien anzusehen, in denen die Erfassung kognitiver Prozesse im Vordergrund steht. Für zukünftige experimentelle Eye-Tracking-Studien wäre daher eine Reduktion der Komplexität des Stimulusmaterials ratsam. Hierdurch könnten die vielschichtigen Prozessabschnitte der Informationsverarbeitung bei der Beurteilung von Schüler(inne)n besser berücksichtigt werden. Denkbar wäre die Darstellung einzelner stereotypkonsistenter und -inkonsistenter Schülerinformationen (z. B. einzelne Sätze) auf separaten Stimulusseiten in der Experimentalsoftware, die zudem nur für eine zeitlich begrenzte Zeit präsentiert werden. Die zeitliche Begrenzung könnte dabei eine eher automatisierte Informationsverarbeitung begünstigen (vgl. Fiske & Neuberg, 1990). Hierdurch würde für alle Versuchspersonen die Präsentationsdauer konstant gehalten werden und Rücksprünge zu vorherigen Informationen wären nicht möglich. Hierdurch würde für jede Versuchsperson eine exakt gleiche zeitliche Betrachtungsdauer für die einzelnen AOIs vorliegen. Dies würde die Auswertung der Pupillenreaktion im zeitlichen Verlauf ermöglichen und könnte so einen detaillierteren Aufschluss über die kognitive Aktivierung bei der Schülerbeurteilung liefern. (2) Auch der Einsatz von Eye-Trackern in realen oder realitätsnäheren Situationen ist anzustreben, um die Befunde aus Laboruntersuchungen zu herkunftsassoziierten Informationsverarbeitungsprozessen in ökologisch valideren Settings zu untersuchen. Dabei wäre zum einen denkbar, Lehrkräfte mit Eye-Tracking-Brillen auszustatten, um okulomotorische Daten während des regulären Unterrichts zu erfassen. Hierbei müssten selbstverständlich zunächst ethische Aspekte berücksichtigt sowie datenschutzrechtliche Bestimmungen beachtet werden. Jedoch wäre eine akkurate Auswertung des Pupillendiameters erschwert, da die Lichtverhältnisse im Klassen-

raum nur schwer kontrolliert werden können und durch die Vielzahl an akustischen und visuellen Reizen, die zeitgleich auf die Lehrkraft einwirken, Artefakte im Datensatz entstehen könnten, die nicht im Zusammenhang mit den spezifischen Schüler(inne)n stehen, die gerade fixiert werden. Die vergleichbare Problematik würde sich für die Auswertung der Lidschlagrate ergeben. Zwar wären hierbei Schwankungen der Helligkeit nicht von Bedeutung, dennoch kann die Vielzahl an Reizen im Klassenkontext nur bedingt berücksichtigt werden. Ein entsprechendes Setting wäre somit nur durch einen enormen Aufwand und die Dokumentation aller Ereignisse im Klassenraum möglich. Jedoch könnte die Auswertung der Fixationsdauer Hinweise auf die visuelle Aufmerksamkeit der Lehrkraft im Zusammenhang mit bestimmten Schülergruppen liefern. Um den Einschränkungen im realen Unterricht zu entgehen und dennoch die externe Validität in Laboruntersuchungen zu erhöhen, könnte auch realitätsnäheres Stimulusmaterial, wie Videovignetten aus realen Unterrichtssituationen, genutzt werden. Dabei könnte in beiden Varianten untersucht werden, ob sich die Aufmerksamkeit der Lehrkräfte bei Schüler(inne)n unterschiedlicher Herkunft bei stereotypkonsistenten und -inkonsistenten Verhaltensweisen unterscheidet.

Zusammenfassend kann festgehalten werden, dass basierend auf den Befunden der vorliegenden Arbeit zukünftige Studien ihren Fokus zunehmend auch auf positive Einstellungen und Stereotype gegenüber spezifischen Schülergruppen richten sollten und positive Urteilsverzerrungen als Erklärungsansatz für herkunftsassoziierte Disparitäten bei Lehrkrafturteilen und -erwartungen in Erwägung ziehen. Dies schließt nicht aus, dass in spezifischen Fällen auch negative Assoziationen zu bestimmten Schülergruppen zu negativ verzerrten Urteilen führen können, jedoch schränkt die Fokussierung auf diesen Aspekt die Berücksichtigung der unterschiedlichen Konsequenzen beider Urteilsverzerrungen ein. Lassen sich diese Befunde in weiteren Studien replizieren, wäre dies zum einen ein bedeutsamer Erklärungsansatz für herkunftsassoziierte Disparitäten und müsste zum anderen einen höheren Stellenwert in der Ausbildung von Lehramtsstudierenden sowie in der Weiterbildung von erfahrenen Lehrkräften einnehmen.

Praktische Implikationen für die Aus- und Weiterbildung von Lehrkräften

Basierend auf den vorliegenden Befunden und theoretischen Annahmen der sozialen Informationsverarbeitung lassen sich, trotz einiger Limitationen, Implikationen für die pädagogische Praxis ableiten. So deuten die Befunde aus Studie 1 weiterhin auf herkunftsassoziierte Lehrkrafturteile und -erwartungen hin, für die Lehrkräfte entsprechend sensibilisiert werden sollten. Zudem konnten Hinweise gefunden werden, dass Lehrkräfte Schüler mit Migrationshintergrund zwar überwiegend akkurat einschätzen, Schüler ohne Migrationshintergrund und mit hohem sozioökonomischem Status dagegen eher überschätzen und diese dadurch einen Erwartungsvorsprung haben, auf den Schüler mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status verzichten müssen. Dies kann zu einer indirekten Benachteiligung und herkunftsassoziierten Disparitäten führen. Zu niedrige Erwartungen können dabei zu einer Unterschätzung der Lernenden führen (Weinstein, 2002) und entsprechend einer optimalen Förderung im Weg stehen. Aber auch zu hohe Erwartungen können zu einem hohen Druck für die Lernenden führen (Weinstein, 2002) und die Schüler(innen) hierdurch überfordern. Beide Aspekte sollten v.a. unter der Reflexion eigener

Einstellungen und Stereotype berücksichtigt werden und v. a. in der Aus- und Weiterbildung darauf aufmerksam gemacht werden, dass auch positiv verzerrte Urteile zu einer Benachteiligung bestimmter Schülergruppen führen können.

Nach Ferguson (2003) ist es nicht ausreichend, die Forderung nach gleichen Erwartungen an alle Schüler(innen) zu stellen, sondern spezifische Trainingsprogramme anzubieten, die entsprechendes Wissen vermitteln und die Reflexion anregen (z. B. das Programm *Great Expectations*; vgl. Ferguson, 2003; *Great Expectations*, 2017). Diamond et al. (2004) verweisen auf die besondere Bedeutung der Schulleitung, die durch ein hohes Engagement einen Einfluss auf die pädagogischen Überzeugungen der Lehrkräfte haben können. Dabei wurde besonders hervorgehoben, dass sie Lehrkräfte dafür sensibilisieren können, ein Verantwortungsgefühl für das Lernen und die Leistungen der Schüler(innen) zu entwickeln (Diamond et al., 2004). Sind Lehrkräfte davon überzeugt, dass sie einen Einfluss auf das Lernen der Schüler haben und hierfür auch Verantwortung tragen, kann sich dies auf ihre pädagogische Haltung und auch auf ihre Erwartungen gegenüber ethnischen und sozialen Schülergruppen auswirken. Dies steht auch im Einklang mit dem ADCM (Böhmer et al., 2017) sowie mit Untersuchungen von Glock et al. (2012) sowie Krolak-Schwerdt et al. (2013), die zeigen konnten, dass die Schülerherkunft bei einer geringen Verantwortung der Lehrkräfte einen Einfluss auf der Informationsverarbeitung hat. Entsprechend sollte sowohl in der Ausbildung als auch in der Weiterbildung von Lehrkräften verstärkt darauf aufmerksam gemacht werden, dass sie selbst einen entscheidenden Einfluss auf Bildungsverlauf und -erfolg von Schüler(inne)n haben. Hierzu können zum einen theoretische Modelle angeführt werden, die die Bedeutung der Lehrperson aufgreifen (z. B. Helmke, 2009). Zum anderen sollten auch empirische Befunde präsentiert werden, die darauf hinweisen, dass Lehrkräfte einen bedeutsamen Einfluss auf herkunftsassoziierte Leistungsdisparitäten haben (z. B. Dresel et al., 2017).

In diesem Zusammenhang stellt eine stärkere Berücksichtigung der diagnostischen Kompetenz in der Lehramtsausbildung und Weiterbildung von erfahrenen Lehrkräften eine zentrale Voraussetzung für eine akkurate und möglichst objektive Beurteilung von Schüler(inne)n dar (vgl. auch Artelt & Gräsel, 2009).⁹⁰ In Anlehnung an Schrader und Helmke (2001) könnten hierfür die Vorteile des Lehrkrafturteils (Schülerinformationen aus unterschiedlichen Situationen und über einen längeren Zeitraum) mit standardisierten Tests (hohes Maß an Objektivität) kombiniert werden, um ein möglichst objektives und umfassendes Bild der Schüler(innen) und ihrer Leistungen zu erhalten. Auch wenn die Bedeutung einer umfassenden Diagnostik der Schülerleistungen, -fähigkeiten und des Verhaltens für eine adäquate Unterrichtsplanung und -durchführung unabdingbar ist, sollte diese diskret erfolgen, um Schüler(innen) nicht ein Gefühl der ständigen Leistungsprüfung zu vermitteln (Schrader & Helmke, 2001). Dabei sollte für die Schüler(innen) eine klare Trennung von Lern- und Leistungssituationen vorliegen, auch wenn Lehrkräfte in Lernsituationen relevante Informationen für den diagnostischen Prozess erhalten.

⁹⁰ Basierend auf den vorliegenden Befunden können Forderungen nach möglichst genauen, objektiven und validen Instrumenten zur Erfassung von Schülerverhalten und Leistungen, die über die subjektive Einschätzung bzw. alltagsdiagnostischen Methoden der Lehrkraft hinausgehen, sowie bessere Lehrkraftschulungen zur Einschätzung von Schülermerkmalen (z. B. Stang & Urhahne, 2016a, 2016b) nur unterstützt werden.

Zudem gibt es Forderungen nach einer stärkeren Kooperation zwischen Lehrkräften (z. B. Seitz, 2006), wodurch neben didaktischen und pädagogischen Diskussionen ebenfalls ein Austausch über Schüler(innen) und damit auch die Erweiterung der eigenen Perspektive ermöglicht wird. Aber auch die diagnostischen Prozesse der Einschätzung von Leistungen und Fähigkeiten der Schüler(innen) könnten davon profitieren, die große Streuung bei der Einschätzung von Schülerarbeiten (u. a. Birkel, 2005; Birkel & Birkel, 2002; Weiss, 1965) könnte minimiert werden. Jedoch sollte in jedem Fall kriteriengeleitet vorgegangen werden, um zu vermeiden, dass subjektive Fehleinschätzungen und negative Einstellungen sich auf andere Lehrkräfte übertragen.

Entsprechend sollte bereits in der Ausbildung von Lehrkräften auch verstärkt auf eine kriteriale Bewertung von Leistungen mit einem einheitlichen Kriterienkatalog hingewiesen werden (vgl. Birkel & Birkel, 2002). In Bezug auf den Umgang mit spezifischen Lehrkraftherwartungen sollten v. a. individuelle und kriteriale Bezugsnormen bei der Beurteilung angelegt werden und in diesem Zusammenhang eine kontinuierliche Dokumentation individueller Schülerleistungen im Prozess betrachtet werden (Überblick über mögliche Strategien bei Alderman, 2004). Dresel et al. (2017) konnten im Zusammenhang der Bezugsnormorientierung von Lehrkräften zeigen, dass herkunftsassoziierte Leistungsdisparitäten in Klassen mit Lehrkräften, die eine individuelle Bezugsnorm anwenden, deutlich geringer sind. Stang und Urhahne (2016a) konnten zudem einen positiven Zusammenhang zwischen der Rangkomponente und einer ausgeprägten kriterialen Bezugsnormorientierung der Lehrkräfte identifizieren. Je eher Lehrkräfte somit spezifische Kriterien zur Bewertung von Schüler(inne)n heranziehen, desto höher ist auch die Übereinstimmung zwischen Lehrkrafturteil und tatsächlicher Schülerleistung. Entsprechend sinkt die Gefahr einer herkunftsassoziierten positiven oder negativen Urteilsverzerrung bei der Bewertung von Leistungen.

Zumindest sollte in diesem Kontext auf die Vermeidung sozialer Bezugsnormen bei der Beurteilung von Schüler(inne)n und ihren Leistungen hingewiesen werden. Zudem verweist Schrader (2010) darauf, dass „informelle Urteile als diagnostische Hypothesen betrachtet werden [sollten], die der Absicherung und Korrektur durch formelle Diagnose bedürfen.“ (S. 107). Karing und Artelt (2013) verweisen im Kontext ihrer Forschung zur globalen und aufgabenspezifischen Lehrkrafturteilen auf die Vermittlung von hierzu notwendigem diagnostischem Wissen in der Aus- und Weiterbildung von Lehrkräften. Sie betonen jedoch auch, dass fachliches und fachdidaktisches Wissen ebenfalls eine bedeutsame Rolle für die Urteilsakkuratheit von Lehrkräften spielt und entsprechend auch in der Aus- und Weiterbildung von hoher Relevanz ist (Karing & Artelt, 2013).

Ein grundlegendes Ziel der Lehrkraftausbildung und Weiterbildung sollte es zudem sein, negative Einstellungen gegenüber ethnischen und sozialen Minoritäten zu reduzieren (vgl. Ülger, Dette-Hagemeyer & Reichle, 2016), aber auch positive Einstellungen gegenüber bestimmten sozialen Kategorien zu reflektieren. Hierzu ist Wissen über Einflussfaktoren und Verzerrungen der sozialen Informationsverarbeitung und der Beurteilung von Schüler(inne)n und ihren Leistungen (vgl. Glock & Krolak-Schwerdt, 2013; Kaiser et al., 2015; van Ophuysen, 2006) unabdingbar. Da die Unterdrückung von Stereotypen einerseits häufig nicht gelingt und andererseits oftmals zum gegenteiligen Effekt führen kann (vgl. z. B. Bodenhausen & Macrae,

1996), scheint dies keine angemessene Strategie im Umgang mit ethnischen und sozialen Stereotypen im schulischen Kontext zu sein. Durch Übungen und Diskussionen sollte stattdessen ein Bewusstsein für eigene explizite Einstellungen und Stereotype gegenüber bestimmten sozialen Gruppen erfolgen. Des Weiteren könnten durch den wissenschaftlich begleiteten Einsatz des IAT auch implizite Einstellungen bewusstwerden und damit die Gefahr verdeutlicht werden, dass diese sich ebenfalls auf Urteile (insbesondere unter Zeitdruck und geringer kognitiver Ressourcen) auswirken können.

Dabei sind – vergleichbar mit Modellen der Eindrucksbildung (z.B. Brewer, 1988; Devine, 1989) – auch bei Prozessen der Einstellungsänderung v. a. die vorhandenen kognitiven Ressourcen sowie die persönliche Motivation zur Verarbeitung ausschlaggebend für die Verarbeitungstiefe der neuen Informationen (Bless et al., 2004). Zur Einstellungsänderung wurden im Verlauf der psychologischen Auseinandersetzung mit diesem Thema zwei wesentliche Modelle entwickelt (Überblick bei Eagly & Chaiken, 1993): (1) Das *elaboration-likelihood-model of persuasion* (ELM; Petty & Cacioppo, 1986; Petty & Wegener, 1999): Im ELM wird davon ausgegangen, dass die Wahrscheinlichkeit einer Einstellungsänderung von v.a. der Motivation und der Fähigkeit zur Informationsverarbeitung abhängig ist. Unterschieden werden eine zentrale und eine periphere Informationsverarbeitungsroute. Eine zentrale Informationsverarbeitung erfolgt, wenn die Motivation und die Fähigkeiten zur Verarbeitung relativ hoch sind. Zudem sollte u. a. die Information von hoher Qualität sein, um eine Einstellung zu ändern. Durch eine zentrale Informationsverarbeitung steigt die Wahrscheinlichkeit einer stabilen Einstellungsänderung. Eine periphere Verarbeitung von Informationen erfolgt hingegen bei eher niedriger Motivation und geringen Fähigkeiten. Dabei sind qualitative Argumente nicht von besonderer Bedeutung und eine eher oberflächlichere Informationsverarbeitung erfolgt. Die Verarbeitung der Informationen über eine periphere Route führt eher zu instabilen Einstellungsänderungen. (2) Das zweite bedeutsame Modell ist das *heuristic systematic model* (HSM; u. a. Chaiken, Liberman & Eagly, 1989): Im HSM wird ebenfalls von einer vertieften Route der Informationsverarbeitung (*systematic*) ausgegangen, in der alle relevanten Informationen zur Urteilsbildung berücksichtigt werden. Hierfür sind jedoch Anstrengungen und kognitive Kapazitäten erforderlich. Stehen nur geringe kognitive Ressourcen zur Verfügung, steigt die Wahrscheinlichkeit, nur einen gewissen Anteil der zur Verfügung stehenden Informationen zu berücksichtigen und Informationen eher heuristisch zu verarbeiten (*heuristic*). Entsprechend sollte in Maßnahmen zur Einstellungsänderung von Lehrkräften und Lehramtsstudierenden die zentrale Verarbeitung von Informationen angestrebt werden, da es nach dem ELM nur hierdurch zu einer stabilen Einstellungsänderung kommt. Somit sollten entsprechende Seminare und Kurse die Motivation und die Fähigkeiten der Teilnehmer(innen) berücksichtigen, um eine zentrale Informationsverarbeitung zu ermöglichen.

Insbesondere, wenn bestehende Einstellungen nicht mit neuen Erfahrungen übereinstimmen, kann dies ein Anlass zur Einstellungsänderung sein (Allport, 1935). Es wird angenommen, dass Dissonanz zum einen dadurch entsteht, dass Menschen kontinuierlich neuen Informationen ausgesetzt sind, die nicht immer mit dem bisherigen Wissen übereinstimmen müssen (Festinger, 1957; Rothbart & John, 1985). So können angehende Lehrkräfte davon überzeugt sein, dass Schüler(innen) mit Migrationshintergrund schlechtere schulische

Leistungen erbringen oder sprachliche Schwierigkeiten haben, da sie in ihrem Studium häufig mit empirischen Belegen hierzu konfrontiert wurden. In der schulischen Praxis können ihnen jedoch Schüler(innen) mit Migrationshintergrund begegnen, die sie zunächst möglicherweise durch ihre guten Leistungen und sprachliche Fähigkeiten überraschen. Diese erste Überraschung, als Zeichen der Dissonanz der bisherigen und neuen Informationen, kann dann unter Umständen dazu führen, dass Wissen umstrukturiert wird und die neuen Informationen integriert werden (z. B. besonders begabte und engagierte Schüler(innen) werden als Ausnahme von der Regel angesehen). Entsprechend sollten Lehramtsstudierende bereits während ihres Studiums mit positiven Bildungsverläufen von Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status konfrontiert werden und zudem durch die kritische Auseinandersetzung mit Fachartikeln die häufig heterogene Befundlage und die zahlreichen Einflussfaktoren auf den Bildungsverlauf und -erfolg kennenlernen. Hierdurch könnte Dissonanz entstehen, die zu einer Einstellungsänderung führen könnte. Die Reduktion negativer Einstellungen gegenüber sozialen Gruppen kann dabei über unterschiedliche Strategien erfolgen (Überblick bei Ülger et al., 2016).

Bislang existieren zur Reduktion von Stereotypen und Vorurteilen meist Konzepte fokussiert, die sich auf Outgroup-Stereotype fokussieren, wie z. B. die Kontakthypothese oder Dekategorisierung (Allport, 1954; Hewstone, 1996; Pettigrew, 1998; Pettigrew & Tropp, 2008). Diese Maßnahmen sind bei der Reduktion von Stereotypen bzw. bei der Änderung der Kategorisierungsstruktur in Bezug auf die Outgroup von hoher Bedeutung. Allerdings deuten die Befunde der vorliegenden Arbeit darauf hin, dass neben der Stereotypisierung der Outgroup auch eine bedeutsame Stereotypisierung der Ingroup vorliegt, die zu einer indirekten Benachteiligung von Schüler(inne)n mit Migrationshintergrund und/oder niedrigem sozioökonomischem Status führen kann. Hierfür sind die bislang entwickelten Maßnahmen zur Reduktion von Stereotypen und Vorurteilen jedoch eher ungeeignet, da sie die positiven Stereotype zur Eigengruppe vernachlässigen. Aus diesem Grund sind – sofern sich diese Befunde in weiteren empirischen Studien replizieren lassen – weitere Sensibilisierungs- und Reflexionsmaßnahmen notwendig, um bei Lehrkräften das Bewusstsein für das Vorhandensein und die Konsequenzen von positiven Einstellungen und Stereotypen zu schärfen.

Wenn Lehrkräften durch derartige Maßnahmen stärker ins Bewusstsein rückt, dass nicht nur negative Einstellungen und Stereotype zu herkunftsassoziierten Disparitäten beitragen, sondern auch positive Einstellungen und Stereotype zu einer ungleichen Behandlung von Schüler(inne)n führen, wäre ein wichtiger Reflexionsprozess in Gang gesetzt, der in Kombination mit u. a. einer kriterialen Bezugsnorm und einer hohen diagnostischen Kompetenz zur Reduktion von Unterschieden im Bildungsverlauf und -erfolg von Kindern und Jugendlichen unterschiedlicher ethnischer und sozialer Herkunft beitragen kann.

Literaturverzeichnis

- Abadi, R. V., Clement, R. & Gowen, E. (2003). Levels of fixation. In L. Harris & M. Jenkin (Eds.), *Levels of perception* (pp. 213–229). New York, NY: Springer.
- Abdi, H. & Williams, L. J. (2010). Contrast analysis. In N. Salkind (Ed.), *Encyclopedia of research design* (pp. 1–14). Thousand Oaks, CA: Sage.
- Ajzen, I. (2005). *Attitudes, personality and behavior* (2nd ed.). Maidenhead: Open University Press.
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Alba, R. & Johnson, M. (2000). Zur Messung aktueller Einstellungsmuster gegenüber Ausländern in Deutschland. In R. Alba, P. Schmidt & M. Wasmer (Hrsg.), *Deutsche und Ausländer. Freunde, Fremde oder Feinde?* (S. 229–253). Wiesbaden: Westdeutscher Verlag.
- Albarracín, D. & Vargas, P. (2010). Attitudes and persuasion. From biology to social responses to persuasive intent. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of social psychology* (vol. 1, 5th ed., pp. 394–427). Hoboken, NJ: Wiley.
- Alderman, M. K. (2004). *Motivation for achievement. Possibilities for teaching and learning* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Allemann-Ghionda, C. (2006). Klasse, Gender oder Ethnie? Zum Bildungserfolg von Schüler/innen mit Migrationshintergrund. Von der Defizitperspektive zur Ressourcenorientierung. *Zeitschrift für Pädagogik*, 52, 350–362.
- Allen, T. J., Sherman, J. W., Conrey, F. R. & Stroessner, S. J. (2009). Stereotype strength and attentional bias. Preference for confirming versus disconfirming information depends on processing capacity. *Journal of Experimental Social Psychology*, 45, 1081–1087. doi: 10.1016/j.jesp.2009.06.002
- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (vol. 2, reissued 1967, pp. 798–844). New York, NY: Russell & Russell.
- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Alvidrez, J. & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology*, 91, 731–746. doi: 10.1037//0022-0663.91.4.731
- Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences. A meta-analysis. *Psychological Bulletin*, 111, 256–274. doi: 10.1037//0033-2909.111.2.256
- Ambady, N. & Rosenthal, R. (1993). Half a minute. Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64, 431–441. doi: 10.1037//0022-3514.64.3.431
- American Psychological Association. (2012). *Ethnic and racial disparities in education. Psychology's contributions to understanding and reducing disparities*. Verfügbar unter <http://www.apa.org/ed/resources/racial-disparities.aspx>.
- Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, 2, 1–9. doi: 10.1037/h0021966
- Andresen, S., Fegter, S. & Hurrelmann, K. (2013). Wohlbefinden, Armut und Gerechtigkeit aus Sicht der Kinder. In World Vision Deutschland e. V. (Hrsg.), *Wie gerecht ist unsere Welt? Kinder in Deutschland 2013. 3. World Vision Kinderstudie* (S. 26–48). Weinheim: Beltz.
- Anokhin, A. & Vogel, F. (1996). EEG alpha rhythm frequency and intelligence in normal adults. *Intelligence*, 23, 1–14. doi: 10.1016/S0160-2896(96)80002-X
- Antidiskriminierungsstelle des Bundes. (2013). *Diskriminierung im Bildungsbereich und im Arbeitsleben. Zweiter gemeinsamer Bericht der Antidiskriminierungsstelle des Bundes und der in ihrem Zuständigkeitsbereich betroffenen Beauftragten der Bundesregierung und des deutschen Bundestages*. Berlin. Verfügbar unter http://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/BT_Bericht/Gemeinsamer_Bericht_zweiter_2013.pdf?__blob=publicationFile&v=4

- Arnold, K.-H., Bos, W., Richert, P. & Stubbe, T. C. (2007). Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes et al. (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 272–297). Münster: Waxmann.
- Artelt, C. & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23, 157–160. doi: 10.1024/1010-0652.23.34.157
- Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology*, 41 (2), 76–81. doi: 10.1027/1864-9335/a000011
- Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41, 258–290. doi: 10.1037/h0055756
- Asch, S. E. & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology*, 46, 1230–1240. doi: 10.1037/0022-3514.46.6.1230
- Ashmore, R. D. & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–35). Hillsdale, NJ: Lawrence Erlbaum.
- Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K. Fox, M. A., Frohlich, L., Kemp, J., Drake, L. (2010). *The condition of education 2010 (NCES 2010-028)*, National Center for Education Statistics. Verfügbar unter <http://www.nces.ed.gov/pubs2010/2010028.pdf>
- Auernheimer, G. (1996). *Einführung in die interkulturelle Erziehung* (2. Aufl.). Darmstadt: Primus.
- Auernheimer, G. (2001). Anforderungen an das Bildungssystem und die Schule in der Einwanderungsgesellschaft. In G. Auernheimer (Hrsg.), *Migration als Herausforderung* (Interkulturelle Studien, Bd. 7, S. 45–58). Opladen: Leske + Budrich.
- Babad, E. (1990). Measuring and changing teachers' differential behavior as perceived by students and teachers. *Journal of Educational Psychology*, 82, 683–690. doi: 10.1037/0022-0663.82.4.683
- Babad, E. (1992). Teacher expectancies and nonverbal behavior. In R. S. Feldman (Ed.), *Applications of nonverbal behavioral theories and research* (pp. 167–190). Hillsdale, NJ: Lawrence Erlbaum.
- Babad, E. Y., Inbar, J. & Rosenthal, R. (1982). Pygmalion, Galatea, and the Golem. Investigations of biased and unbiased teachers. *Journal of Educational Psychology*, 74, 459–474. doi: 10.1037//0022-0663.74.4.459
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2000). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (9. Aufl.). Berlin: Springer.
- Bade, K. J. (2000). *Europa in Bewegung. Migration vom späten 18. Jahrhundert bis zur Gegenwart*. München: C. H. Beck.
- Bailey, A. L. & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment*, 11, 149–178. doi: 10.1080/10627197.2006.9652988
- Baltes, B. B. & Rudolph, C. W. (2010). Examining the effect of negative Turkish stereotypes on evaluative workplace outcomes in Germany. *Journal of Managerial Psychology*, 25, 148–158. doi: 10.1108/02683941011019357
- Banaji, M. R. & Heiphetz, L. (2010). Attitudes. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of social psychology* (vol. 1, 5th ed., pp. 353–393). Hoboken, NJ: Wiley.
- 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. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122–147. doi: 10.1037//0003-066x.37.2.122
- Bargel, H. & Bargel, T. (2010). *Ungleichheiten und Benachteiligungen im Hochschulstudium aufgrund der sozialen Herkunft der Studierenden* (Arbeitspapier 202). Düsseldorf: Hans-Böckler-Stiftung. Verfügbar unter http://www.boeckler.de/pdf/p_arbp_202.pdf
- Bargh, J. A., Chaiken, S., Govender, R. & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, 62, 893–912. doi: 10.1037//0022-3514.62.6.893
- Bargh, J. A. & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462–479. doi: 10.1037//0003-066x.54.7.462

- Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior. Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230–244. doi: 10.1037//0022-3514.71.2.230
- Baron, J. (1988). *Thinking and deciding*. New York, NY: Cambridge University Press.
- Bar-Tal, D. (1996). Development of social categories and stereotypes in early childhood. The case of “The Arab” concept formation, stereotype and attitudes by Jewish children in Israel. *International Journal of Intercultural Relations*, 20, 341–370. doi: 10.1016/0147-1767(96)00023-5
- Bar-Tal, D. (1997). Formation and change of ethnic and national stereotypes. An integrative model. *International Journal of Intercultural Relations*, 21, 491–523. doi: 10.1016/s0147-1767(97)00022-9
- Bartl-Pokorny, K., Pokorny, F., Bölte, S., Langmann, A., Falck-Ytter, T., Wolin, T. et al. (2013). Eye-Tracking. Anwendung in Grundlagenforschung und klinischer Praxis. *Klinische Neurophysiologie*, 44 (3), 193–198. doi: 10.1055/s-0033-1343458
- Baumert, J., Maaz, K., Gresch, K., McElvany, N., Anders, Y., Jonkmann, K., Neumann, M. et al. (2010). Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten. Zusammenfassung der zentralen Befunde. In K. Maaz, J. Baumert, C. Gresch & N. McElvany (Hrsg.), *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (S. 5–21). Bonn: Bundesministerium für Bildung und Forschung.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520. doi: 10.1007/s11618-006-0165-2
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–410). Opladen: Leske + Budrich.
- Baumert, J., Watermann, R. & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6, 46–71. doi: 10.1007/s11618-003-0004-7
- Bayerische Staatskanzlei. (2008a). § 15 Zwischen- und Jahreszeugnisse. Grundschulordnung (GrSO). Verfügbar unter <http://www.gesetze-bayern.de/Content/Document/BayVSO-15>
- Bayerische Staatskanzlei. (2008b). § 6 Übertritt an ein Gymnasium oder an eine Realschule. Grundschulordnung (GrSO). Verfügbar unter <http://www.gesetze-bayern.de/Content/Document/BayVSO-6>
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst. (o. J.). *Übertritt und Schulwechsel in Bayern. So geht die Schulkarriere weiter*. Verfügbar unter <https://www.km.bayern.de/ministerium/schule-und-ausbildung/schularten/uebertritt-schulartwechsel.html>
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst. (2009). *Beratung und Transparenz in der Übertrittsphase* [Az.: IV.1-5 S 4302-6.64 320]. Verfügbar unter <https://www.verkuendung-bayern.de/kwmb1/jahrgang:2009/heftnummer:13/seite:263>
- Bean, M. G., Slaten, D. G., Horton, W. S., Murphy, M. C., Todd, A. R. & Richeson, J. A. (2012). Prejudice concerns and race-based attentional bias. New evidence from eyetracking. *Social Psychology and Personality Science*, 3, 722–729. doi: 10.1177/1948550612436983
- Beatty, J. & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinay & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge: Cambridge University Press.
- Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration. (2014). *10. Bericht der Beauftragten der Bundesregierung für Migration, Flüchtlinge und Integration über die Lage der Ausländerinnen und Ausländer in Deutschland. (Oktober 2014)*. Berlin: Bonifatius.
- Becker, B. (2010a). *Bildungsaspiration von Migranten. Determinanten und Umsetzung in Bildungsergebnisse*. Arbeitspapiere – Working Papers Nr. 137, 2010. Mannheimer Zentrum für Europäische Sozialforschung. Verfügbar unter <http://www.mzes.uni-mannheim.de/publications/wp/wp-137.pdf>
- Becker, B. (2010b). Ethnische Unterschiede bei der Kindergartenselektion. Die Wahl von unterschiedlich stark segregierten Kindergärten in deutschen und türkischen Familien. In B. Becker & D. Reimer (Hrsg.), *Vom Kindergarten bis zur Hochschule. Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie* (S. 17–48). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Becker, B. & Biedinger, N. (2006). Ethnische Bildungsungleichheit zu Schulbeginn. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 58, 660–684. doi: 10.1007/s11577-006-0261-6
- Becker, R. & Lauterbach, W. (2008). Vom Nutzen vorschulischer Erziehung und Elementarbildung. Bessere Bildungschancen für Arbeiterkinder? In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (3. Aufl., S. 129–159). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A. & Sherman, J. W. (2008). The quadruple process model approach to examining the neural underpinnings of prejudice. *Neuroimage*, 43, 775–783. doi: 10.106/j.neuroimage.2008.08.033
- 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. doi: 10.1037/1045-3830.23.1.43
- Benz, W. (2012). Vorurteile gegen Muslime. Feindbild Islam. In A. Pelinka (Hrsg.), *Vorurteile. Ursprünge, Formen, Bedeutung* (S. 205–220). Berlin: de Gruyter.
- Bartnitzky, H. (1999). Zeugnisse als lernfördernde Rückmeldung. In W. Böttcher, U. Brosch & H. Schneider-Petri (Hrsg.), *Leistungsbewertung in der Grundschule* (Werkstattbuch Grundschule, S. 15–29). Weinheim: Beltz.
- Bertrand, M. & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, 94, 991–1013. doi: 10.3386/w9873
- Bialystok, E., Craik, F. I. M. & Luk, G. (2012). Bilingualism. Consequences for mind and brain. *Trends in Cognitive Sciences*, 16, 240–250. doi: 10.1016/j.tics.2012.03.001
- Biedinger, N. & Becker, B. (2006). *Der Einfluss des Vorschulbesuchs auf die Entwicklung und den langfristigen Bildungserfolg von Kindern. Ein Überblick über internationale Studien im Vorschulbereich*. Arbeitspapiere (Nr. 97). Mannheim: Mannheimer Zentrum für Europäische Sozialforschung. Verfügbar unter <http://www.mzes.uni-mannheim.de/publications/wp/wp-97.pdf>
- Biedinger, N. & Becker, B. (2010). Frühe ethnische Bildungsungleichheit. Der Einfluss des Kindergartenbesuchs auf die deutsche Sprachfähigkeit und die allgemeine Entwicklung. In B. Becker & D. Reimer (Hrsg.), *Vom Kindergarten bis zur Hochschule. Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie* (S. 49–79). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bierhoff, H. W. (1986). *Personenwahrnehmung. Vom ersten Eindruck zur sozialen Interaktion* (Lehr- und Forschungstexte, Bd. 20). Berlin: Springer.
- Billmann-Mahecha, E. & Tiedemann, J. (2006). Migration. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 486–495). Weinheim: Beltz.
- Birbaumer, N. & Schmidt, R. F. (2010). *Biologische Psychologie* (7. Aufl.). Heidelberg: Springer.
- Birkel, P. (2005). Beurteilungsübereinstimmung bei Mathematikarbeiten? *Journal für Mathematik-Didaktik*, 26, 28–51. doi: 10.1007/BF03339005
- Birkel, P. & Birkel, C. (2002). Wie einig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolph Weiss. *Psychologie in Erziehung und Unterricht*, 49, 219–224.
- Birnstein, M. (o. J.). *Schülerbeurteilung. Erstellen von Schülerbeurteilungen durch Auswahl passender Textbausteine*. Verfügbar unter <http://schuelerbeurteilungen.de/beurteilungen.php>
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242–261. doi: 10.1207/s15327957pspr0603_8
- Bless, H., Fiedler, K. & Strack, F. (2004). *Social cognition. How individuals construct social reality*. Hove: Taylor & Francis.
- Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics. Evidence of circadian variations in discrimination. *Psychological Science*, 1, 319–322. doi: 10.1111/j.1467-9280.1990.tb00226.x
- Bodenhausen, G. V. (2005). The role of stereotypes in decision-making processes. *Medical Decision Making*, 25 (1), 112–118. doi: 10.1177/0272989x04273800
- Bodenhausen, G. V. & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies. The impact of task complexity. *Journal of Personality and Social Psychology*, 52, 871–880. doi: 10.1037/0022-3514.52.5.871

- Bodenhausen, G. V. & Macrae, C. N. (1996). The self-regulation of intergroup perception. Mechanisms and consequences of stereotype suppression. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 227–253). New York, NY: Guilford Press.
- Bodenhausen, G. V., Sheppard, L. A. & Kramer, G. P. (1994). Negative affect and social judgment. The differential impact of anger and sadness. *European Journal of Social Psychology*, 24, 45–62. doi: 10.1002/ejsp.2420240104
- Bodenhausen, G. V. & Wyer, R. S. (1985). Effects of stereotypes in decision making and information-processing strategies. *Journal of Personality and Social Psychology*, 48, 267–282. doi: 10.1037//0022-3514.48.2.267
- Bogardus, E. S. (1925a). Analyzing changes in public opinion. *Journal of Applied Sociology*, 9, 372–381. Verfügbar unter https://brocku.ca/MeadProject/Bogardus/Bogardus_1925d.html
- Bogardus, E. S. (1925b). Measuring social distances. *Journal of Applied Sociology*, 9, 299–308. Verfügbar unter https://brocku.ca/MeadProject/Bogardus/Bogardus_925c.html
- Bogardus, E. S. (1925c). Social distance and its origins. *Journal of Applied Sociology*, 9, 216–226. Verfügbar unter https://brocku.ca/MeadProject/Bogardus/Bogardus_1925b.html
- Bogardus, E. S. (1926). Social distance in the city. *Proceedings and Publications of the American Sociological Society*, 20, 40–46. Verfügbar unter https://brocku.ca/MeadProject/Bogardus/Bogardus_1926.html
- Böhmer, I., Gräsel, C., Hörstermann, T. & Krolak-Schwerdt, S. (2012). Die Informationssuche bei der Erstellung der Übergangsempfehlung. Die Rolle von Fallkonsistenz und Expertise. *Unterrichtswissenschaft*, 40, 140–155.
- Böhmer, I., Gräsel, C., Krolak-Schwerdt, S., Hörstermann, T. & Glock, S. (2017). Teachers' school tracking decisions. In D. Leutner, J. Fleischer, J. Grünkorn, E. Klieme & D. Leutner (Eds.), *Competence assessment in education. Research, models and instruments* (pp. 131–147). Cham: Springer International Publishing.
- Böhmer, I., Hörstermann, T., Gräsel, C., Krolak-Schwerdt, S. & Glock, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung. Welcher Urteilsregel folgen Lehrkräfte? *Journal of Educational Research Online*, 7, 59–81. Verfügbar unter www.j-e-r-o.com/index.php/jero/article/viewFile/566/241
- Böhmer, M. (2011). *Expertise und diagnostische Urteilsbildung. Ein sozial-kognitiver Ansatz*. Dissertationsschrift, Universität Luxemburg. Luxemburg.
- Bonefeld, M., Dickhäuser, O., Janke, S., Praetorius, A.-K. & Dresel, M. (2017). Migrationsbedingte Disparitäten in der Notenvergabe nach dem Übergang auf das Gymnasium. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 49, 11–23. doi: 10.1026/0049-8637/a000163
- Bonowski, L. (2002). *Die sympathische Reflexantwort der Pupille*. Dissertationsschrift, Philipps- Universität Marburg. Marburg.
- Bonsen, M., Frey, K. A. & Bos, W. (2008). Soziale Herkunft. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 141–156). Münster: Waxmann.
- Bonsen, M., Kummer, N. & Bos, W. (2008). Schülerinnen und Schüler mit Migrationshintergrund. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 157–175). Münster: Waxmann.
- Borelli, J. L., Crowley, M. J., David, D. H., Sbarra, D. A., Anderson, G. M. & Mayes, L. C. (2010). Attachment and emotion in school-aged children. *Emotion*, 10, 475–485. doi: 10.1037/a0018490
- Bos, W. & Hovenga, N. (2010). Diagnostische Kompetenzen. Besser individuell fördern. *Schule NRW* (8), 383–385. Verfügbar unter <http://www.schulentwicklung.nrw.de/materialdatenbank/nutzersicht/materialeintrag-rrsq.php?matId=5171>
- Bos, W., Schwippert, K. & Stubbe, T. C. (2007). Die Koppelung von sozialer Herkunft und Schülerleistungen im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes et al. (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 225–247). Münster: Waxmann.
- Boudon, R. (1974). *Education, opportunity and social inequality. Changing prospects in western society*. New York, NY: Wiley.
- Bourdieu, P. (1982). *Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft*. Frankfurt a. M.: Suhrkamp.

- Braun, C. & Mehringer, V. (2010). Familialer Hintergrund, Übertrittsempfehlungen und Schulerfolg bei Kindern mit und ohne Migrationshintergrund. In J. Hagedorn, V. Schurt, C. Steber & W. Waburg (Hrsg.), *Ethnizität, Geschlecht, Familie und Schule. Heterogenität als erziehungswissenschaftliche Herausforderung* (S. 55–79). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition. A dual process model of impression formation* (vol. 1, pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum.
- Brewer, M. B. (1996). When stereotypes lead to stereotyping. The use of stereotypes in person perception. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 254–275). New York, NY: Guilford Press.
- Brimi, H. M. (2011). Reliability of grading high school work in english. *Practical Assessment, Research & Evaluation*, 16 (17), 1–12. Verfügbar unter <http://pareonline.net/getvn.asp?v=16&n=17>
- Brookings, J. B., Wilson, G. F. & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42, 361–377. doi: 10.1016/0301-0511(95)05167-8
- Brooks-Gunn, J. & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, 7, 55–71. doi: 10.2307/1602387
- Brooks-Gunn, J., Klebanov, P. K. & Duncan, G. J. (1996). Ethnic differences in children's intelligence test scores. Role of economic deprivation, home environment, and maternal characteristics. *Child Development*, 67, 396–408. doi: 10.2307/1131822
- Brophy, J. E. (1983). Classroom organization and management. *The Elementary School Journal*, 83, 265–285. doi: 10.1086/461318
- Brophy, J. E. & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance. Some behavioral data. *Journal of Educational Psychology*, 61, 365–374. doi: 10.1037/h0029908
- Brophy, J. E. & Good, T. L. (1974). *Teacher-student relationships. Causes and consequences*. New York, NY: Holt, Rinehart and Winston.
- Brückner, G. (2016). Bevölkerung mit Migrationshintergrund. In Statistisches Bundesamt (Hrsg.), *Datenreport 2016. Ein Sozialbericht für die Bundesrepublik Deutschland* (S. 218–235). Bonn: Bundeszentrale für politische Bildung.
- Brühwiler, C., Helmke, A. & Schrader, F.-W. (2017). Determinanten der Schulleistung. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (3. Aufl., S. 291–314). Wiesbaden: Springer.
- Bruner, J. S. (1957a). Going beyond the information given. In J. S. Bruner (Ed.), *Contemporary approaches to cognition. A symposium held at the University of Colorado* (pp. 41–69). Cambridge: Harvard University Press.
- Bruner, J. S. (1957b). On perceptual readiness. *Psychological Review*, 64, 123–152. doi: 10.1037/h0043805
- Bruner, J. S. (1950). Social psychology and group processes. *Annual Review of Psychology*, 1, 119–150. doi: 10.1146/annurev.ps.01.020150.001003
- Bruner, J. S. & Goodman, C. C. (1947). Value and need as organizing factors in perception. *The Journal of Abnormal and Social Psychology*, 42, 33–44. doi: 10.1037/h0058484
- Bruner, J. S. & Postman, L. (1949). On the perception of incongruity. A paradigm. *Journal of Personality*, 18, 206–223. doi: 10.1111/j.1467-6494.1949.tb01241.x
- Bruner, J. S. & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Verfügbar unter <http://www.dtic.mil/get-tr-doc/pdf?AD=AD0024982>
- Büchel, F., Spieß, K. C. & Wagner, G. (1997). Bildungseffekte vorschulischer Kinderbetreuung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 49, 528–539.
- Bundesamt für Migration und Flüchtlinge. (2017a). *Hauptherkunftsländer von Asylbewerbern in Deutschland im Jahr 2017. [Stand: Juli 2017]*. Verfügbar unter <https://de.statista.com/statistik/daten/studie/154287/umfrage/hauptherkunftslaender-von-asylbewerbern/>

- Bundesamt für Migration und Flüchtlinge. (2017b). *Verteilung der Asylbewerber in Deutschland nach Altersgruppen im Jahr 2017. [Stand: Juli 2017]*. Verfügbar unter <https://de.statista.com/statistik/daten/studie/452149/umfrage/asylbewerber-in-deutschland-nach-altersgruppen/>
- Bundesamt für Migration und Flüchtlinge. (2014). *Zuwanderung aus den neuen EU-Mitgliedsstaaten Bulgarien und Rumänien. Forschungsbericht 24*. Verfügbar unter http://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Forschungsberichte/fb24-rumaenien-bulgarien.pdf?__blob=publicationFile
- Bundeskanzleramt. (2012). § 14 Leistungsbeurteilungsverordnung. LBV. Verfügbar unter <https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009375&Artikel=&Paragraf=14&Anlage=&Uebergangsrecht=>
- Bundesministerium für Arbeit und Soziales (2013). *Lebenslagen in Deutschland. Armuts- und Reichtumsberichterstattung der Bundesregierung*. Der vierte Armuts- und Reichtumsbericht der Bundesregierung, Bundesministerium für Arbeit und Soziales. Verfügbar unter <https://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen-DinA4/a334-4-armuts-reichtumsbericht-2013.html>
- Bundesministerium für Bildung und Forschung. (2016). *Neue Wege in der Lehrerbildung. Die Qualitätsoffensive Lehrerbildung*. Verfügbar unter https://www.bmbf.de/pub/Neue_Wege_in_der_Lehrerbildung.pdf
- Cacioppo, J. T., Lorig, T. S., Nusbaum, H. C. & Berntson, G. G. (2004). Social neuroscience. Bridging social and biological systems. In C. Sansone, C. C. Morf & A. T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 383–404). Thousand Oaks, CA: Sage.
- Cantor, N. & Mischel, W. (1979). Prototypes in person perception. *Advances in Experimental Social Psychology*, 12, 3–52. doi: 10.1016/S0065-2601(08)60258-0
- Carlsson, R. & Björklund, F. (2010). Implicit stereotype content. Mixed stereotypes can be measured with the implicit association test. *Social Psychology*, 41, 213–222. doi: 10.1027/1864-9335/a000029
- Casper, C., Rothermund, K. & Wentura, D. (2010). Automatic stereotype activation is context dependent. *Social Psychology*, 41 (3), 131–136. doi: 10.1027/1864-9335/a000019
- Chabris, C. & Simons, J. (2013). *Der unsichtbare Gorilla. Wie unser Gehirn sich täuschen lässt*. München: Piper.
- Chaiken, S., Liberman, A. & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York, NY: Guilford Press.
- Chaiken, S. & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York, NY: Guilford Press.
- Chen, B. & Weseley, A. J. (2011). The effect of student ethnicity on teacher perceptions. *Journal of Research in Education*, 21 (2), 55–65. Verfügbar unter <http://files.eric.ed.gov/fulltext/EJ1098409.pdf>
- Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152. doi: 10.1207/s15516709cog0502_2
- Cikara, M. & Fiske, S. T. (2012). Stereotypes and Schadenfreude. Affective and physiological markers of pleasure at outgroup misfortunes. *Social Psychological and Personality Science*, 3, 63–71. doi: 10.1177/1948550611409245
- Claiborn, W. L. (1969). Expectancy effects in the classroom. A failure to replicate. *Journal of Educational Psychology*, 60, 377–383. doi: 10.1037/h0028320
- Clark, C. M. & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255–296). New York, NY: Macmillan and co.
- Constantinidis, C., Bucci, D. J. & Rugg, M. D. (2013). Cognitive functions of the posterior parietal cortex. *Frontiers in Integrative Neuroscience*, 7, 1–2. doi: 10.3389/fnint.2013.00035
- Cooper, H. M. (1979). Pygmalion grows up. A model for teacher expectation communication and performance influence. *Review of Educational Research*, 49, 389–410. doi: 10.2307/1170137
- Cooper, H. M., Baron, R. M. & Lowe, C. A. (1975). The importance of race and social class information in the formation of expectancies about academic performance. *Journal of Educational Psychology*, 67, 312–319. doi: 10.1037/h0077006

- Cooper, H. M., Findley, M. & Good, T. L. (1982). Relations between student achievement and various indexes of teacher expectations. *Journal of Educational Psychology*, 74, 577–579. doi: 10.1037//0022-0663.74.4.577
- Cortés Núñez, S. A. & Küçük, K. (2016). Armut von Migrantinnen und Migranten in Deutschland. In Der paritätische Gesamtverband (Hrsg.), *Zeit zu handeln. Bericht zur Armutsentwicklung in Deutschland 2016* (S. 57–63). Berlin: Der paritätische Gesamtverband.
- Cranach, M. von, Schmid, R. & Vogel, M. W. (1969). Über einige Bedingungen des Zusammenhanges von Lidschlag und Blickwendung. *Psychologische Forschung*, 33 (1), 68–78. doi: 10.1007/bf00424617
- Crano, W. D. & Mellon, P. M. (1978). Causal influence of teachers' expectations on children's academic performance. A cross-lagged panel analysis. *Journal of Educational Psychology*, 70, 39–49. doi: 10.1037/0022-0663.70.1.39
- Crano, W. D. & Prislin, R. (2006). Attitudes and persuasion. *Annual Review of Psychology*, 57, 345–374. doi: 10.1146/annurev.psych.57.102904.190034
- Crick, N. R. & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, 115, 74–101. doi: 10.1037/0033-2909.115.1.74
- Crisp, R. J., Hutter, R. R. C. & Young, B. (2009). When mere exposure leads to less liking. The incremental threat effect in intergroup contexts. *British Journal of Psychology*, 100 (1), 133–149. doi: 10.1348/000712608x318635
- Cronbach, L. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin*, 52, 177–193. doi: 10.1037/h0044919
- Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.
- Cullen, W. (1797). *Clinical lectures*. London: Messrs. Lee and Hurs.
- Dahl, G. B. & Lochner, L. (2012). The impact of family income on child achievement. Evidence from the earned income tax credit. *American Economic Review*, 102, 1927–1956. doi: 10.1257/aer.102.5.1927
- Darley, J. M. & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, 35, 867–881. doi: 10.1037//0003-066x.35.10.867
- Darley, J. M. & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44, 20–33. doi: 10.1037//0022-3514.44.1.20
- Darley, J. M. & Oleson, K. C. (1993). Introduction to research on interpersonal expectations. In P. D. Blanck (Ed.), *Interpersonal expectations. Theory, research and applications* (pp. 45–63). Cambridge: Cambridge University Press.
- Dasgupta, N. & Greenwald, A. G. (2001). On the malleability of automatic attitudes. Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81, 800–814. doi: 10.1037//0022-3514.81.5.800
- De Boer, H., Bosker, R. J. & van der Werf, M. P. C. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, 102, 168–179. doi: 10.1037/a0017289
- Dee, T. S. (2004). Teachers, race and student achievement in a randomized experiment. *The Review of Economic and Statistics*, 86 (1), 195–210. doi: 10.1162/003465304323023750
- Dee, T. S. (2005). A teacher like me. Does race, ethnicity, or gender matter? *The American Economic Review*, 95, 158–165. doi: 10.1257/000282805774670446
- Delabarre, E. B. (1898). A method of recording eye-movements. *The American Journal of Psychology*, 9, 572–574. doi: 10.2307/1412191
- Demaray, M. K. & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning. A comparison of actual and predicted performances. *School Psychology Quarterly*, 13, 8–24. doi: 10.1037/h0088969
- Devine, P. G. (1989). Stereotypes and prejudice. Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18. doi: 10.1037//0022-3514.56.1.5
- Devine, P. G. & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality & Social Psychology Bulletin*, 21, 1139–1150. doi: 10.1177/01461672952111002
- Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E. & Vance, S. L. (2002). The regulation of explicit and implicit race bias. The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, 82, 835–848. doi: 10.1037//0022-3514.82.5.835

- Diamond, J. B., Randolph, A. & Spillane, J. P. (2004). Teachers' expectations and sense of responsibility for student learning. The importance of race, class, and organizational habitus. *Anthropology & Education Quarterly*, 35, 75–98. doi: 10.1525/aeq.2004.35.1.75
- 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, 844–849. doi: 10.1016/j.lindif.2012.04.004
- Dickhäuser, O. (2006). Fähigkeitsselbstkonzepte. *Zeitschrift für Pädagogische Psychologie*, 20, 5–8. doi: 10.1024/1010-0652.20.12.5
- Dickhäuser, O., Schöne, C., Spinath, B. & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept. Konstruktion und Überprüfung eines neuen Instruments. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23, 393–405. doi: 10.1024//0170-1789.23.4.393
- Diefenbach, H. (2008). Bildungschancen und Bildungs(miss)erfolg von ausländischen Schülern oder Schülern aus Migrantenfamilien im System schulischer Bildung. In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (3. Aufl., S. 221–245). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Diefenbach, H. (2010). *Kinder und Jugendliche aus Migrantenfamilien im deutschen Bildungssystem. Erklärungen und empirische Befunde* (3. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Diehl, C. & Fick, P. (2016). Ethnische Diskriminierung im deutschen Bildungssystem. In C. Diehl & P. Fick (Hrsg.), *Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten* (S. 243–286). Wiesbaden: Springer.
- Diehm, I. & Radtke, F.-O. (1999). *Erziehung und Migration. Eine Einführung* (Grundriss der Pädagogik, Bd. 3). Stuttgart: Kohlhammer.
- Dijker, A. J. & Koomen, W. (1996). Stereotyping and attitudinal effects under time pressure. *European Journal of Social Psychology*, 26, 61–74. doi: 10.1002/(sici)1099-0992(199601)26:1<61::aid-ejsp737>3.0.co;2-#
- Ditton, H. (2016). Der Beitrag von Schule und Lehrern zur Reproduktion von Bildungsungleichheit. In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (5. Aufl., S. 281–312). Wiesbaden: Springer.
- Ditton, H. & Krüsken, J. (2006a). Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 9, 348–372. doi: 10.1007/s11618-006-0055-7
- Ditton, H. & Krüsken, J. (2006b). Sozialer Kontext und schulische Leistungen – zur Bildungsrelevanz segregierter Armut. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 26, 135–157.
- Ditton, H., Krüsken, J. & Schauenberg, M. (2005). Bildungsungleichheit. Der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft*, 8, 285–304. doi: 10.1007/s11618-005-0138-x
- Dodge, R. & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review*, 8, 145–157. doi: 10.1037/h0076100
- Doherty, J. & Conolly, M. (1985). How accurately can primary school teachers predict the scores of their pupils in standardized tests of attainment? A study of some non-cognitive factors that influence specific judgements. *Educational Studies*, 11, 41–60. doi: 10.1080/0305569850110105
- Dollmann, J. (2011). Verbindliche und unverbindliche Grundschulempfehlungen und soziale Ungleichheiten am ersten Bildungsübergang. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 63, 595–621. doi: 10.1007/s11577-011-0148-z
- Dovidio, J. F., Brigham, J. C., Johnsson, B. T. & Gaertner, S. L. (1996). Stereotyping, prejudice, and discrimination. Another look. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 276–319). New York, NY: Guilford Press.
- Downey, D. B. & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, 77, 267–282. doi: 10.1177/003804070407700401
- Drake, R. L., Vogl, A. W., Mitchell, A. W. M., Tibbitts, R. M. & Richardson, P. E. (2009). *Gray's Atlas der Anatomie*. München: Elsevier.
- Dresel, M., Martschinke, S., Kopp, B., Tobisch, A. & Kröner, S. (2017). Zur Bedeutung des Unterrichts für die Koppelung von sozialer Herkunft und Schulleistung. Ergebnisse einer Studie im Grundschulunterricht im Fach Deutsch. *Zeitschrift für Grundschulforschung*, 10, 136–149.

- Dresel, M., Stöger, H. & Ziegler, A. (2006). Klassen- und Schulunterschiede im Ausmaß von Geschlechterdiskrepanzen bei Leistungsbewertungen und Leistungsaspirationen. Ergebnisse einer Mehrebenenanalyse. *Psychologie in Erziehung und Unterricht*, *53*, 44–61.
- Dresel, M. & Ziegler, A. (2006). Langfristige Förderung von Fähigkeitsselbstkonzept und impliziter Fähigkeitstheorie durch computerbasiertes attributionales Feedback. *Zeitschrift für Pädagogische Psychologie*, *20*, 49–63. doi: 10.1024/1010-0652.20.12.49
- Dumont, H., Maaz, K., Neumann, M. & Becker, M. (2014). Soziale Ungleichheiten beim Übergang von der Grundschule in die Sekundarstufe I. Theorie, Forschungsstand, Interventions- und Fördermöglichkeiten. *Zeitschrift für Erziehungswissenschaft*, *17*, 141–165. doi: 10.1007/s11618-013-0466-1
- Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence. Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, *34*, 590–598. doi: 10.1037/0022-3514.34.4.590
- Dunham, Y. & Emory, J. (2014). Of affect and ambiguity. The emergence of preference for arbitrary ingroups. *Journal of Social Issues*, *70* (1), 81–98. doi: 10.1111/josi.12048
- Dunkake, I., Kiechle, T., Klein, M. & Rosar, U. (2012). Schöne Schüler, schöne Noten? Eine empirische Untersuchung zum Einfluss der physischen Attraktivität von Schülern auf die Notenvergabe durch das Lehrpersonal. *Zeitschrift für Soziologie*, *41*, 143–161 doi: 10.1515/zfsoz-2012-0206
- Dunkake, I. & Schuchart, C. (2015). Stereotypes and teacher characteristics as an explanation for the class-specific disciplinary practices of pre-service teachers. *Teaching and Teacher Education*, *50*, 56–69. doi: 10.1016/j.tate.2015.04.005
- Dünnebier, K., Gräsel, C. & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. *Zeitschrift für Pädagogische Psychologie*, *23*, 187–195. doi: 10.1024/1010-0652.23.34.187
- Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J. C., Akande, A. D., Adetoun, B. E. et al. (2012). Nations' income inequality predicts ambivalence in stereotype content. How societies mind the gap. *The British Journal of Social Psychology*, *52*, 726–746. doi: 10.1111/bjso.12005
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*, 1040–1048. doi: 10.1037//0003-066x.41.10.1040
- Eagly, A. H., Ashmore, R. D., Makhijani, M. G. & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109–128. doi: 10.1037//0033-2909.110.1.109
- Eagly, A. H. & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly*, *18*, 585–609. doi: 10.1111/j.1540-4560.2011.01718.x
- Ehmke, T., Hohensee, F., Heidemeier, H. & Prenzel, M. (2004). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland* (S. 225–254). Ergebnisse des zweiten internationalen Vergleichs. Münster: Waxmann.
- Eichner, M. (2013). *Blickbewegung als Marker für kognitive Verarbeitung beim verstehenden Lesen*. Dissertationsschrift, Justus-Liebig-Universität Gießen. Gießen.
- Elashoff, J. D. & Snow, R. E. (1972). *Pygmalion auf dem Prüfstand. Einführung in empirisch-statistische Methoden auf der Grundlage einer kritischen Analyse der Rosenthal-Jacobson-Studie "Pygmalion im Klassenzimmer"*. München: Kösel Verlag.
- Elliott, E. S. & Dweck, C. S. (1988). Goals. An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*, 5–12. doi: 10.1037//0022-3514.54.1.5
- Enck, P., Benedetti, F. & Schedlowski, M. (2008). New insights into the placebo and nocebo responses. *Neuron*, *59*, 195–206. doi: 10.1016/j.neuron.2008.06.030
- Erikson, R. & Jonsson, J. O. (1996). Explaining class inequality in education. The Swedish test case. In R. Erikson & J. O. Jonsson (Eds.), *Can education be equalized? The Swedish case in comparative perspective* (pp. 1–63). Boulder, CO: Westview Press.
- Euwals, R., Dagevos, J., Gijsberts, M. & Roodenburg, H. (2007). *Immigration, integration and the labour market. Turkish immigrants in Germany and the Netherlands*. IZA Discussion Paper No. 2677. Bonn: Forschungsinstitut der Zukunft der Arbeit. Verfügbar unter <http://ftp.iza.org/dp2677.pdf>

- Faber, G. (2007). *Selbstkonzept, Kausalattribution und Leistungsangst im Rechtschreiben. Ansatz und ausgewählte Ergebnisse einer Untersuchungsreihe zu den schulfachspezifisch erfassten Selbsteinschätzungen von Grundschulkindern*. Dissertationsschrift, Universität Hannover. Hannover.
- Farkas, G. (2003). Racial disparities and discrimination in education. What do we know, how do we know it, and what do we need to know? *Teachers College Record*, 105, 1119–1146. doi: 10.1111/1467-9620.00279
- Faulbaum, F. & Stein, P. (2000). Wie homogen sind Einstellungen gegenüber Ausländern? Zur Aufdeckung und Modellierung unbeobachteter Heterogenität in Umfragedaten. In R. Alba, P. Schmidt & M. Wasmer (Hrsg.), *Deutsche und Ausländer. Freunde, Fremde oder Feinde?* (S. 485–518). Wiesbaden: Westdeutscher Verlag.
- Fazio, R. H. (1986). How do attitudes guide behavior? In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition. Foundations of social behavior* (pp. 204–243). New York, NY: Guilford Press.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior. The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (vol. 23, pp. 75–109). New York, NY: Academic Press.
- Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, 25, 603–637. doi: 10.1521/soco.2007.25.5.603
- Fazio, R. H., Effrein, E. A. & Falender, V. J. (1981). Self-perceptions following social interaction. *Journal of Personality and Social Psychology*, 41, 232–242. doi: 10.1037/0022-3514.41.2.232
- Fazio, R. H., Powell, M. C. & Williams, C. J. (1989). The role of attitude accessibility in the attitude-to-behavior process. *Journal of Consumer Research*, 16, 280–288. doi: 10.1086/209214
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C. & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238. doi: 10.1037//0022-3514.50.2.229
- Feinberg, A. B. & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, 18, 52–65. doi: 10.1521/scpq.18.1.52.20876
- Fenning, P. & Rose, J. (2007). Overrepresentation of African American students in exclusionary discipline the role of school policy. *Urban Education*, 42 (6), 536–559. doi: 10.1177/0042085907305039
- Ferguson, R. F. (2003). Teachers' perceptions and expectations and the black-white test score gap. *Urban Education*, 38 (4), 460–507. doi: 10.1177/0042085903038004006
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Feuß, S. (2013). *Auf den ersten Blick. Wie Medieninhalte wahrgenommen und rezipiert werden*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fiedler, K. & Bless, H. (2003). Soziale Kognition. In W. Stroebe, K. Jonas & M. Hewstone (Hrsg.), *Sozialpsychologie. Eine Einführung* (4. Aufl., S. 125–163). Berlin: Springer.
- Field, A. (2009). *Discovering statistics using SPSS. And sex and drugs and rock 'n' roll* (3rd ed.). Los Angeles, CA: Sage.
- Fincke, G. (2009). *Abgehängt, chancenlos, unwillig? Eine empirische Reorientierung von Integrationstheorien zu MigrantInnen der zweiten Generation in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Findlay, J. M. (2003). Eye Movements and visual information processing. Commentary on Section 1. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The mind's eye. Cognitive and applied aspects of eye movement research* (pp. 143–155). Amsterdam: Elsevier.
- Finsterwald, M. (2006). *Motivation und Schulübergang: Bedingungsfaktoren eines erfolgreichen Übergangs auf weiterführende Schulen*. Berlin: Logos.
- Fiske, S. T. (2014). *Social beings. Core motives in social psychology* (3rd ed.). Hoboken, NJ: Wiley.
- Fiske, S. T. & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes. Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (vol. 23, pp. 1–74). New York, NY: Academic Press.
- Fiske, S. T. & Taylor, S. E. (2013). *Social cognition. From brains to culture* (2nd ed.). Los Angeles, CA: Sage.
- Fiske, S. T., Cuddy, A. J. C., Glick, P. & Xu, J. (2002). A model of (often mixed) stereotype content. Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82, 878–902. doi: 10.1037//0022-3514.82.6.878

- Fiske, S. T., Kinder, D. R. & Larter, W. (1983). The novice and the expert. Knowledge-based strategies in political cognition. *Journal of Experimental Social Psychology*, 19, 381–400. doi: 10.1016/0022-1031(83)90029-X
- Fiske, S. T. & Linville, P. W. (1980). What does the schema concept buy us? *Personality & Social Psychology Bulletin*, 6, 543–557. doi: 10.1177/014616728064006
- Fiske, S. T., Lin, M. & Neuberg, S. L. (1999). The continuum model. Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York, NY: Guilford Press.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 378–382. doi: 10.1037/h0031619
- Fogarty, C. & Stern, J. A. (1989). Eye movements and blinks. Their relationship to higher cognitive processes. *International Journal of Psychophysiology*, 8, 35–42. doi: 10.1016/0167-8760(89)90017-2
- Fowles, D. C., Kochanska, G. & Murray, K. (2000). Electrodermal activity and temperament in preschool children. *Psychophysiology*, 37, 777–787. doi: 10.1111/1469-8986.3760777
- Frank, J. D. (1935). Some psychological determinants of the level of aspiration. *The American Journal of Psychology*, 47, 285–293. doi: 10.2307/1415832
- Fuhr, G. (2012). *Armutsgefährdung von Menschen mit Migrationshintergrund. Ergebnisse des Mikrozensus 2010*. Wiesbaden: Statistisches Bundesamt.
- Fukuda, K. (2001). Eye blinks. New indices for the detection of deception. *International Journal of Psychophysiology*, 40, 239–245. doi: 10.1016/s0167-8760(00)00192-6
- Funder, D. C. (1987). Errors and mistakes. Evaluating the accuracy of social judgment. *Psychological Bulletin*, 101, 75–90. doi: 10.1037/0033-2909.101.1.75
- Funder, D. C. (1995). On the accuracy of personality judgment. A realistic approach. *Psychological Review*, 102, 652–670. doi: 10.1037//0033-295x.102.4.652
- Fürstenau, S. & Gomolla, M. (2012). Einführung. Migration und schulischer Wandel: Leistungsbeurteilung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel. Leistungsbeurteilung* (S. 13–24). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Galley, N. (2001). Physiologische Grundlagen, Meßmethoden und Indikatorfunktionen der okulomotorischen Aktivität. In F. Rösler (Hrsg.), *Grundlagen und Methoden der Psychophysiologie* (Enzyklopädie der Psychologie, Bd. 4, S. 237–316). Göttingen: Hogrefe.
- Ganzeboom, H. B. G., de Graaf, P. M. & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1–56. doi: 10.1016/0049-089x(92)90017-b
- Garces, E., Thomas, D. & Currie, J. (2002). Longer-term effects of head start. *American Economic Review*, 92, 999–1012. doi: 10.1257/00028280260344560
- García, E. (2015). *Inequalities at the starting gate. Cognitive and noncognitive skills gaps between 2010-2011 kindergarten classmates* (Economic Policy Institute). Washington, DC. Verfügbar unter <http://files.eric.ed.gov/fulltext/ED560407.pdf>
- Gawronski, B., Alshut, E., Grafe, J., Nespethal, J., Ruhmland, A. & Schulz, L. (2002). Prozesse der Urteilsbildung über bekannte und unbekannte Personen. Wie der erste Eindruck die Verarbeitung neuer Informationen beeinflusst. *Zeitschrift für Sozialpsychologie*, 33, 25–34. doi: 10.1024//0044-3514.33.1.25
- Gawronski, B., Geschke, D. & Banse, R. (2003). Implicit bias in impression formation. Associations influence the construal of individuating information. *European Journal of Social Psychology*, 33, 573–589. doi: 10.1002/ejsp.166
- Gebauer, M. M. & McElvany, N. (2017). Empirische Arbeit. Zur Bedeutsamkeit unterrichtsbezogener heterogenitätsspezifischer Einstellungen angehender Lehrkräfte für intendiertes Unterrichtsverhalten. *Psychologie in Erziehung und Unterricht*, 64, 163. doi: 10.2378/peu2017.art11d
- Gebhardt, M., Rauch, D., Mang, J., Sälzer, C. & Stanat, P. (2013). Mathematische Kompetenzen von Schülerinnen und Schülern mit Zuwanderungshintergrund. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 275–308). Münster: Waxmann.
- Geißler, R. (1994). Soziale Schichtung und Bildungschancen. In R. Geißler (Hrsg.), *Soziale Schichtung und Lebenschancen in Deutschland* (2. Aufl., S. 111–159). Stuttgart: Ferdinand Enke.
- Gershenson, S., Holt, S. & Papageorge, N. (2015). Who believes in me? The effect of student-teacher demographic match on teacher expectations. *SSRN Journal*. Online publication. doi: 10.2139/ssrn.2633993

- Gesellschaft für Deutsche Sprache. (2012). *Vornamensstatistiken Bayern 2010-2012 (unveröffentlichte Statistik)*. Wiesbaden.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Meredith.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 254–267. doi: 10.1037//0033-295x.98.2.254
- Gilbert, D. T. (1999). What the mind's not. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 3–11). New York, NY: Guilford Press.
- Gilbert, D. T. & Hixon, J. G. (1991). The trouble of thinking. Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60, 509–517. doi: 10.1037/0022-3514.60.4.509
- Glock, S. (2016). Does ethnicity matter? The impact of stereotypical expectations on in-service teachers' judgments of students. *Social Psychology of Education*, 19, 493–509. doi: 10.1007/s11218-016-9349-7
- Glock, S. & Karbach, J. (2015). Preservice teachers' implicit attitudes toward racial minority students. Evidence from three implicit measures. *Studies in Educational Evaluation*, 45, 55–61. doi: 10.1016/j.stueduc.2015.03.006
- Glock, S., Kneer, J. & Krolak-Schwerdt, S. (2011). Impression formation or prediction? Category fit and task influence forensic person memory. *Journal of Forensic Psychology Practice*, 11, 391–405. doi: 10.1080/15228932.2011.588529
- 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. doi: 10.1007/s11218-012-9197-z
- Glock, S. & Krolak-Schwerdt, S. (2014). Stereotype activation versus application. How teachers process and judge information about students from ethnic minorities and with low socioeconomic background. *Social Psychology of Education*, 17, 589–607. doi: 10.1007/s11218-014-9266-6
- Glock, S., Krolak-Schwerdt, S. & Hörstermann, T. (2016). The higher the SES the better? Implicit and explicit attitudes influence preservice teachers' judgment of students. In C. Fields (Ed.), *Stereotypes and stereotyping. Misperceptions, perspectives and role of social media* (pp. 1–20). New York, NY: Nova.
- Glock, S., Krolak-Schwerdt, S., Klapproth, F. & Böhmer, M. (2012). Improving teachers' judgments. Accountability affects teachers' tracking decision. *International Journal of Technology and Inclusive Education*, 1, 86–95. doi: 10.20533/ijtie.2047.0533.2012.0012
- Göbel, K. & Hesse, H.-G. (2004). Vermittlung interkultureller Kompetenz im Englischunterricht. Eine curriculare Perspektive. *Zeitschrift für Pädagogik*, 50, 818–834.
- Goldenberg, C. (1992). The limits of expectations. A case for case knowledge about teacher expectancy effects. *American Educational Research Journal*, 29, 517–544. doi: 10.3102/00028312029003517
- Gollwitzer, P. M. & Moskowitz, G. B. (1996). Goal effects on action and cognition. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology. Handbook of basic principles* (pp. 361–399). New York, NY: Guilford Press.
- Gomolla, M. & Radtke, F.-O. (2007). *Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule* (2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Good, T. L. & Brophy J. E. (2003). *Looking in classrooms* (9th ed.). Boston, MA: Pearson Education.
- Graham, J. W. (2009). Missing data analysis. Making it work in the real world. *Annual Review of Psychology*, 60, 549–576. doi: 10.1146/annurev.psych.58.110405.085530
- Gramann, K. & Schandry, R. (2009). *Psychophysiologie. Körperliche Indikatoren psychischen Geschehens*. Weinheim: Beltz.
- Grassinger, R., Dresel, M., Brandl-Bredenbeck, H. P., Ohl, U., Lange-Schubert, K. & Hartinger, A. (2016). *Einstellungen zur Heterogenität der Schülerschaft: Struktur, Korrelate und Fachunterschiede in einer Untersuchung mit Lehramtsstudierenden*. Beitrag auf 4. Tagung der Gesellschaft für empirische Bildungsforschung (GEBF 2016). Freie Universität Berlin.
- Great Expectations. (2017). *Transforming lives through education*. Verfügbar unter <http://www.greatexpectations.org/>
- Greenhouse, S. W. & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112. doi: 10.1007/bf02289823
- Greenwald, A. G. & Banaji, M. R. (1995). Implicit social cognition. Attitudes, self-esteem, and stereotypes. *Psychological Bulletin*, 102, 4–27. doi: 10.1037//0033-295x.102.1.4

- Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition. The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464
- Gregory, A. & Weinstein, R. S. (2008). The discipline gap and African Americans. Defiance or cooperation in the high school classroom. *Journal of School Psychology*, 46, 455–475. doi: 10.1016/j.jsp.2007.09.001
- Grolnick, W. S., Ryan, R. M. & Deci, E. L. (1991). Inner resources for school achievement. Motivational mediators of children's perception of their parents. *Journal of Educational Psychology*, 83, 508–517. doi: 10.1037//0022-0663.83.4.508
- Groos, T. & Jehles, N. (2015). *Der Einfluss von Armut auf die Entwicklung von Kindern. Ergebnisse der Schuleingangsuntersuchung (Arbeitspapiere wissenschaftliche Begleitforschung "Kein Kind zurücklassen!" Werkstattbericht, Bd. 3)*. Gütersloh: Bertelsmann Stiftung.
- Hachfeld, A., Anders, Y., Schroeder, S., Stanat, P. & Kunter, M. (2010). Does immigration background matter? How teachers' predictions of students' performance relate to student background. *International Journal of Educational Research*, 49, 78–91. doi: 10.1016/j.ijer.2010.09.002
- Hachfeld, A., Hahn, A., Schroeder, S., Anders, Y. & Kunter, M. (2015). Should teachers be colorblind? How multicultural and egalitarian beliefs differentially relate to aspects of teachers' professional competence for teaching in diverse classrooms. *Teaching and Teacher Education*, 48, 44–55. doi: 10.1016/j.tate.2015.02.001
- Hagendoorn, L. & Hraba, J. (1989). Foreign, different, deviant, seclusive and working class. Anchors to an ethnic hierarchy in the Netherlands. *Ethnic and Racial Studies*, 12, 441–468. doi: 10.1080/01419870.1989.9993647
- Haller, A. O. (1968). On the concept of aspiration. *Rural Sociology*, 33 (4), 484–487.
- Hamilton, D. L., Sherman, S. J. & Ruvolo, C. M. (1990). Stereotype-based expectancies. Effects on information processing and social behavior. *Journal of Social Issues*, 46 (2), 35–60. doi: 10.1111/j.1540-4560.1990.tb01922.x
- Harackiewicz, J. M., Barron, K. E., Tauer, J. M. & Elliot, A. J. (2002). Predicting success in college. A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology*, 94, 562–575. doi: 10.1037//0022-0663.94.3.562
- Harber, K. D. (1998). Feedback to minorities. Evidence of a positive bias. *Journal of Personality and Social Psychology*, 74, 622–628. doi: 10.1037//0022-3514.74.3.622
- Harris, L. R. & Jenkin, M. (2001). Vision and attention. In M. Jenkin & L. R. Harris (Eds.), *Vision and attention* (pp. 1–17). New York, NY: Springer.
- Harris, M. J. & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects. 31 meta-analyses. *Psychological Bulletin*, 97, 363–386. doi: 10.1037/0033-2909.97.3.363
- Harter, A., Graumann, O. & Grittner, F. (2004). "Grundschul-Numerus Clausus" oder Orientierungsstufe? Auswirkungen verschiedener Übertrittsbedingungen auf Motivationsstile und Leistungsängstlichkeit von Grundschulkindern. *Empirische Pädagogik*, 18, 173–193.
- Harvey, D. G. & Slatin, G. T. (1975). The relationship between child's SES and teacher expectations. A test of the middle-class bias hypothesis. *Social Forces*, 54 (1), 140–159. doi: 10.1093/sf/54.1.140
- Hasselhorn, M., Baethge, M., Füssel, H.-P., Hetmeier, H.-W., Maaz, K., Rauschenbach, T. et al. (2014). *Bildung in Deutschland 2014. Ein indikatorengestützter Bericht mit einer Analyse zur Bildung von Menschen mit Behinderungen*. Bielefeld: Bertelsmann Verlag.
- Hattie, J. (2015). *Lernen sichtbar machen* (3. Aufl.). Baltmannsweiler: Schneider.
- Hechtman, S. B. & Rosenthal, R. (1991). Teacher gender and nonverbal behavior in the teaching of gender-stereotyped materials. *Journal of Applied Social Psychology*, 21, 446–459. doi: 10.1111/j.1559-1816.1991.tb00530.x
- Heller, K. A. & Hany, E. A. (2001). Standardisierte Schulleistungsmessungen. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 87–101). Weinheim: Beltz.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Velber: Klett.

- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung. Voraussetzungen, Bedingungen, Erfahrungen* (S. 119–143). Baltmannsweiler: Schneider Verlag Hohengehren.
- Hershman, R., Cohen, N. & Henik, A. (o.J.). *CHAP. An open source software for processing and analyzing pupillometry data*. Verfügbar unter <http://in.bgu.ac.il/en/Labs/CNL/chap/default.aspx>
- Hershman, R., Cohen, N. & Henik, A. (2016). *CHAP. An open source software for processing and analyzing pupillometry data*. [poster]. Verfügbar unter https://www.researchgate.net/publication/295704976_CHAP_An_open_source_software_for_processing_and_analyzing_pupillometry_data
- Herwartz-Emden, L. (2003). Migrant families in Germany. In L. L. Adler & U. P. Gielen (Eds.), *Migration. Immigration and emigration in international perspective* (pp. 165–182). Westport, CT: Praeger.
- Herwartz-Emden, L., Schurt, V. & Waburg, W. (2010). *Aufwachsen in heterogenen Sozialisationskontexten. Zur Bedeutung einer geschlechtergerechten interkulturellen Pädagogik*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hess, E. H. & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*, 1190–1192. doi: 10.1126/science.143.3611.1190
- Hewstone, M. (1996). Contact and categorization. Social psychological interventions to change intergroup relations. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 323–368). New York, NY: Guilford Press.
- Heyden, T. (1979). Augenbewegungen. In C. Becker-Carus, T. Heyden & G. Ziegler (Hrsg.), *Psychophysiologische Methoden. Eine Einführung* (S. 134–146). Stuttgart: Ferdinand Enke.
- Higgins, E. T., King, G. A. & Mavin, G. H. (1982). Individual construct accessibility and subjective impressions and recall. *Journal of Personality and Social Psychology*, *43*, 35–47. doi: 10.1037/0022-3514.43.1.35
- Hill, N. E. & Wang, M.-T. (2015). From middle school to college. Developing aspirations, promoting engagement, and indirect pathways from parenting to post high school enrollment. *Developmental Psychology*, *51*, 224–235. doi: 10.1037/a0038367
- Hinnant, J. B., O'Brien, M. & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, *101*, 662–670. doi: 10.1037/a0014306
- Hochweber, J., Hosenfeld, I. & Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *Journal of Educational Psychology*, *106*, 289–300. doi: 10.1037/a0033829
- Hödl-Nossberger, F. (1964). *Intelligenzeinschätzung nach dem Gesichtsausdruck* (Schule und Beruf, Bd. 98). Wien: Eugen Kettler.
- Hoening, K. & Wenz, S. E. (2013). *Ethnic and social class discrimination in education. Experimental evidence from Germany*. Presentation at the ASA 2013 annual meeting. New York, NY.
- Hoffmann, L. & Böhme, K. (2013). Wie gut können Grundschullehrkräfte die Schwierigkeit von Deutsch- und Mathematikaufgaben beurteilen? Eine Untersuchung zur Genauigkeit aufgabenbezogener Lehrerurteile auf Klassenebene. *Psychologie in Erziehung und Unterricht*, *61*, 42–55. doi: 10.2378/peu2014.art05d
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H. & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality & Social Psychology Bulletin*, *31*, 1369–1385. doi: 10.1177/0146167205275613
- Hoge, R. D. (1983). Psychometric properties of teacher-judgment measures of pupil aptitudes, classroom behaviors, and achievement levels. *The Journal of Special Education*, *17*, 401–429. doi: 10.1177/002246698301700404
- Hoge, R. D. (1984). The definition and measurement of teacher expectations. Problems and prospects. *Canadian Journal of Education*, *9*, 213–228. doi: 10.2307/1494604
- Hoge, R. D. & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology*, *76*, 777–781. doi: 10.1037//0022-0663.76.5.777
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement. A review of literature. *Review of Educational Research*, *59*, 297–313. doi: 10.2307/1170184

- Hohn, K., Schiepe-Tsika, A., Sälzer, C. & Artelt, C. (2013). Lesekompetenz in PISA 2012. Veränderungen und Perspektiven. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 217–244). Münster: Waxmann.
- Holland, M. K. & Tarlow, G. (1972). Blinking and mental load. *Psychological Reports*, *31*, 119–127. doi: 10.2466/pr0.1972.31.1.119
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. & van de Weijer, J. (2011). *Eye Tracking. A comprehensive guide to methods and measures*. New York, NY: Oxford University Press.
- Hosenfeld, I., Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz. Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. *Zeitschrift für Pädagogik, Beiheft 45*, 65–82.
- Huey, E. B. (1898). Preliminary experiments in the physiology and psychology of reading. *The American Journal of Psychology*, *9*, 575–586. doi: 10.2307/1412192
- Hüfner, G. (2003). *Belastungsprofile einzelner Lehrergruppen. BLLV Befragung Arbeitsbelastung in Schulen*. Verfügbar unter http://www.studienseminar-koblenz.de/medien/pflichtmodule_unterlagen/2004/196/08%20Arbeitsbelastung%20Belastungsprofile.pdf
- Hugenberg, K. & Bodenhausen, G. V. (2004). Ambiguity in social categorization. The role of prejudice and facial affect in race categorization. *Psychological Science*, *15*, 342–345. doi: 10.1111/j.0956-7976.2004.00680.x
- Hülür, G., Wilhelm, O. & Robitzsch, A. (2011). Multivariate Veränderungsmodelle für Schulnoten und Schülerleistungen in Deutsch und Mathematik. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *43*, 173–185. doi: 10.1026/0049-8637/a000051
- Humphreys, G. W. & Bruce, V. (1989). *Visual cognition. Computation, experimental, and neuropsychological perspectives*. Hove: Lawrence Erlbaum.
- IBM. (2016) IBM SPSS Statistics [Computer software]. Armonk, NY.
- Ichikawa, N. & Ohira, H. (2004). Eyeblink activity as an index of cognitive processing. Temporal distribution of eyeblinks as an indicator of expectancy in semantic priming. *Perceptual and Motor Skills*, *98*, 131–140. doi: 10.2466/pms.98.1.131-140
- Imhasly, M.-T. (2004). *Metaevaluation Begabungsförderung*. Universität Zürich. Verfügbar unter <http://www.begabungsforderung.ch/pdf/fundus/netzwerk/imhasly.pdf>
- In-Albon, T. & Schneider, S. (2012). Does the vigilance-avoidance gazing behavior of children with separation anxiety disorder change after cognitive-behavioral therapy? *Journal of Abnormal Child Psychology*, *40*, 1149–1156. doi: 10.1007/s10802-012-9634-7
- Itskowitz, R., Navon, R. & Strauss, H. (1988). Teachers' accuracy in evaluating students' self-image. Effect of perceived closeness. *Journal of Educational Psychology*, *80*, 337–341. doi: 10.1037/0022-0663.80.3.337
- Jackson, L. A., Hunter, J. E. & Hodge, C. N. (1995). Physical attractiveness and intellectual competence. A meta-analytic review. *Social Psychology Quarterly*, *58*, 108. doi: 10.2307/2787149
- Jäkel, J. & Leyendecker, B. (2009). Erziehungsverhalten türkischstämmiger und deutscher Mütter von Vorschulkindern. *Psychologie in Erziehung und Unterricht*, *56*, 1–15.
- Jänsch, A. & Schneekloth, U. (2013). Die Freizeit: vielfältig und bunt aber nicht für alle Kinder. In World Vision Deutschland e. V. (Hrsg.), *Wie gerecht ist unsere Welt? Kinder in Deutschland 2013* (S. 135–167). 3. World Vision Kinderstudie. Weinheim: Beltz.
- Jones, E. E. & Goethals, G. R. (1971). *Order effects in impression formation. Attribution context and the nature of the entity*. New York, NY: General Learning Press.
- Jones, E. E. & Sigall, H. (1971). The bogus pipeline. A new paradigm for measuring affect and attitude. *Psychological Bulletin*, *76*, 349–364. doi: 10.1037/h0031617
- Jowell, R. & Prescott-Clarke, P. (1970). Racial discrimination and white-collar workers in Britain. *Race*, *11*, 397–417. doi: 10.1177/030639687001100401
- Jungermann, H., Pfister, H.-R. & Fischer, K. (1998). *Die Psychologie der Entscheidung. Eine Einführung*. Heidelberg: Spektrum Akademischer Verlag.
- Jürgens, E. & Lissmann, U. (2015). *Pädagogische Diagnostik. Grundlagen und Methoden der Leistungsbeurteilung in der Schule*. Weinheim: Beltz.
- Jussim, L. (1990). Expectancies and social issues. Introduction. *Journal of Social Issues*, *46* (2), 1–8. doi: 10.1111/j.1540-4560.1990.tb01920.x

- Jussim, L., Eccles, J. & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations. Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology*, 28, 281–388. doi: 10.1016/s0065-2601(08)60240-3
- Jussim, L. & Fleming, C. (1996). Self-fulfilling prophecies and the maintenance of social stereotypes. The role of dyadic interactions and social forces. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 161–192). New York, NY: Guilford Press.
- Jussim, L. & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies. Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9, 131–155. doi: 10.1207/s15327957pspr0902_3
- Just, M. A. & Carpenter, P. A. (1980). A theory of reading. From eye fixations to comprehension. *Psychological Review*, 87, 329–354. doi: 10.1037/0033-295x.87.4.329
- Kahneman, D. & Tversky, A. (1972). Subjective probability. A judgment of representativeness. *Cognitive Psychology*, 3, 430–454. doi: 10.1016/0010-0287(72)90016-3
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251. doi: 10.1037/h0034747
- Kahraman, B. & Knoblich, G. (2000). „Stechen statt Sprechen“. Valenz und Aktivierbarkeit von Stereotypen über Türken. *Zeitschrift für Sozialpsychologie*, 31, 31–43. doi: 10.1024//0044-3514.31.1.31
- Kaiser, A. (2009). Vornamen. Nomen est omen? *Oberfränkischer Schulanzeiger* (12), 15–18.
- Kaiser, J., Möller, J., Helm, F. & Kunter, M. (2015). Das Schülerinventar. Welche Schülermerkmale die Leistungsurteile von Lehrkräften beeinflussen. *Zeitschrift für Erziehungswissenschaft*, 18, 279–302. doi: 10.1007/s11618-015-0619-5
- 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
- Kaiser, J., Südkamp, A. & Möller, J. (2016). The effects of student characteristics on teacher' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, 109, 871–888. doi: 10.1037/edu0000156
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23, 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. Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25, 159–172. doi: 10.1024/1010-0652/a000041
- Karing, C. & Artelt, C. (2013). Genauigkeit von Lehrpersonenurteilen und Ansatzpunkte ihrer Förderung in der Aus- und Weiterbildung von Lehrkräften. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31, 166–173. Verfügbar unter http://www.pedocs.de/volltexte/2017/13844/pdf/BZL_2013_2_166_173.pdf
- Karing, C., Pfof, M. & Artelt, C. (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? *Journal for Educational Research Online*, 3 (2), 119–147. Verfügbar unter <http://www.j-e-r-o.com/index.php/jero/article/download/188/132>
- Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain*, 106, 643–653. doi: 10.1093/brain/106.3.643
- Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E. & Wyatt, R. J. (1981). Speaking, thinking, and blinking. *Psychiatry Research*, 5, 243–246. doi: 10.1016/0165-1781(81)90070-6
- Karszen, M., van der Veen, I. & Volman, M. (2017). Diversity among bi-ethnic students and differences in educational outcomes and social functioning. *Social Psychology of Education*, 31. Online publication. doi: 10.1007/s11218-017-9394-x
- Katz, D. & Braly, K. W. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, 28, 280–290. doi: 10.1037/h0074049
- Katz, D. & Braly, K. W. (1935). Racial prejudice and racial stereotypes. *The Journal of Abnormal and Social Psychology*, 30, 175–193. doi: 10.1037/h0059800

- Keller, P. M. (2004). *Untersuchung der Pupillenreaktion und -oszillation an einem Normkollektiv mittels Compact Integrated Pupillograph (CIP) der Firma AMTech*. Inaugural-Dissertation, Ruhr-Universität Bochum. Bochum.
- Kemper, T. (2010). Migrationshintergrund – Eine Frage der Definition. *Die deutsche Schule*, 102, 315–326.
- Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J. et al. (2015). *The condition of education 2015 (NCES 2015-144)*, National Center for Education Statistics. Verfügbar unter <http://www.nces.ed.gov/pubs2015/2015144.pdf>
- Kirchler, E. & Zani, B. (1995). Why don't they stay at home? Prejudices against ethnic minorities in Italy. *Journal of Community & Applied Social Psychology*, 5, 59–65. doi: 10.1002/casp.2450050105
- Klapproth, F., Glock, S., Böhmer, M., Krolak-Schwerdt, S. & Martin, R. (2012). School placement decisions in Luxembourg. Do teachers meet the education ministry's standards? *Literacy Information and Computer Education Journal, Special 1* (1), 856–862. doi: 10.20533/licej.2040.2589.2012.0113
- Klapproth, F., Glock, S., Krolak-Schwerdt, S., Martin, R. & Böhmer, M. (2013). Prädiktoren der Sekundarschulempfehlung in Luxemburg. *Zeitschrift für Erziehungswissenschaft*, 16, 355–379. doi: 10.1007/s11618-013-0340-1
- Klapproth, F. & Schaltz, P. (2015). Who is retained in school, and when? Survival analysis of predictors of grade retention in Luxembourgish secondary school. *European Journal of Psychology of Education*, 30, 119–136. doi: 10.1007/s10212-014-0232-7
- Klauer, K. C. (2008). Soziale Kategorisierung und Stereotypisierung. In L.-E. Peterson & B. Six (Hrsg.), *Stereotype, Vorurteile und Diskriminierung. Theorien, Befunde und Interventionen* (S. 23–32). Weinheim: Beltz.
- Klayman, J. (1995). Varieties of confirmation bias, 32, 385–418. doi: 10.1016/S0079-7421(08)60315-1
- Kleine, L., Birnbaum, N., Zielonka, M., Doll, J. & Blossfeld, H.-P. (2010). Auswirkungen institutioneller Rahmenbedingungen auf das Bildungsstreben der Eltern und die Bedeutung der Lehrerempfehlung. *Journal for Educational Research Online*, 2 (1), 72–93. Verfügbar unter http://www.pedocs.de/volltexte/2011/4568/pdf/JERO_2010_1_Kleine_et_al_Auswirkungen_institutioneller_Rahmenbedingungen_S72_D_A.pdf
- Klink, A. & Wagner, U. (1999). Discrimination against ethnic minorities in Germany. Going back to the field. *Journal of Applied Social Psychology*, 29, 402–423. doi: 10.1111/j.1559-1816.1999.tb01394.x
- Komaraju, M. & Nadler, D. (2013). Self-efficacy and academic achievement. Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences*, 25, 67–72. doi: 10.1016/j.lindif.2013.01.005
- König, J. (2009). Klassenklima und schulbezogene Hilflosigkeit in den Jahrgangsstufen 8 und 9. *Zeitschrift für Pädagogische Psychologie*, 23, 41–52. doi: 10.1024/1010-0652.23.1.41
- Kratzmann, J. (2013). Migrationsgekoppelte Ungleichheit durch niedrigere Erwartungen im Kindergarten? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 65, 73–99. doi: 10.1007/s11577-013-0194-9
- Kreiner, H., Sturt, P. & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution. Evidence from eye-movements. *Journal of Memory and Language*, 58, 239–261. doi: 10.1016/j.jml.2007.09.003
- Kreyenfeld, M. (2004). *Sozialstruktur und Kinderbetreuung*. MPIDR Working Paper WP 2004-009. Rostock: Max-Planck-Institut für demografische Forschung. Verfügbar unter <http://www.demogr.mpg.de/papers/working/wp-2004-009.pdf>
- Kristen, C. (2002). Hauptschule, Realschule oder Gymnasium? Ethnische Unterschiede am ersten Bildungsübergang. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 54, 534–552. doi: 10.1007/s11577-002-0073-2
- Kristen, C. (2003). Ethnische Unterschiede im deutschen Schulsystem. *Aus Politik und Zeitgeschichte* (Bd. 21-22), 26–32.
- Kristen, C. (2005). *School choice and ethnic school segregation. Primary school selection in Germany*. Münster: Waxmann.
- Kristen, C. & Dollmann, J. (2009). Sekundäre Effekte der ethnischen Herkunft. Kinder aus türkischen Familien am ersten Bildungsübergang. *Zeitschrift für Erziehungswissenschaft*, 12, 205–229. doi: 10.1007/978-3-531-92216-4_9

- Kristen, C. & Dollmann, J. (2010). Sekundäre Effekte der ethnischen Herkunft. Kinder aus türkischen Familien am ersten Bildungsübergang. In B. Becker & D. Reimer (Hrsg.), *Vom Kindergarten bis zur Hochschule. Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie* (S. 117–139). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kristen, C. & Granato, N. (2007). The educational attainment of the second generation in Germany. Social origins and ethnic inequality. *Ethnicities*, 7 (3), 343–366. doi: 10.1177/1468796807080233
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. *Zeitschrift für Pädagogische Psychologie*, 23, 175–186. doi: 10.1024/1010-0652.23.34.175
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2012). Leistungsbeurteilungen von Schulkindern. Welche Rolle spielen Ziele und Expertise der Lehrkraft? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44, 111–122. doi: 10.1026/0049-8637/a000062
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance. A social cognitive analysis. *Social Psychology of Education*, 16, 215–239. doi: 10.1007/s11218-013-9215-9
- Krolak-Schwerdt, S., Hörstermann, T., Glock, S. & Böhmer, I. (2017). Teachers' assessments of students' achievements: The ecological validity of studies using case vignettes. *The Journal of Experimental Education*. Online publication. doi: 10.1080/00220973.2017.1370686
- Krolak-Schwerdt, S. & Kneer, J. (2006). Lesezeiten- und Blickbewegungsverfahren als Methoden für die soziale Kognitionsforschung. *Zeitschrift für Sozialpsychologie*, 37, 141–150. doi: 10.1024/0044-3514.37.3.141
- Krolak-Schwerdt, S. & Wintermantel, M. (2004). Verarbeitung von personbeschreibender Information als Textverstehen. *Psychologische Rundschau*, 55, 11–18. doi: 10.1026/0033-3042.55.1.11
- Kröner, S. & Dickhäuser, O. (2009). Was klingt intelligenter - Waltraud oder Matthias? Zeitliche Stabilität von Intelligenz-Wortnormen für Vornamen im Deutschen und ihre Geltung für Lehramtsstudierende. *Psychologie in Erziehung und Unterricht*, 56, 150–157.
- Kruger, J.-L., Hefer, E. & Matthew, G. (2013). Measuring the impact of subtitles on cognitive load. Eye tracking and dynamic audiovisual texts. *Proceedings of Eye Tracking South Africa, Cape Town*, 29-31, 62–66. doi: 10.1145/2509315.2509331
- Kruglanski, A. W. & Freund, T. (1983). The freezing and unfreezing of lay-inferences. Effects on impression primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, 19, 448–468. doi: 10.1016/0022-1031(83)90022-7
- Kuklinski, M. R. & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development*, 72, 1554–1578. doi: 10.1111/1467-8624.00365
- Kultusministerkonferenz. (1971). *Eingliederung von Berechtigten nach dem Bundesvertriebenengesetz (BVFG) in Schule und Berufsbildung*. Beschluss der Kultusministerkonferenz vom 03.12.1971 i.d.F. vom 12.09.1997. Verfügbar unter www.kmk.org/fileadmin/Dateien/pdf/ZAB/Hochschulzugang_Beschluesse_der_KMK/BVFG.pdf
- Kultusministerkonferenz. (1996). *Interkulturelle Bildung und Erziehung in der Schule*. Beschluss der Kultusministerkonferenz vom 25.10.1996 i.d.F. vom 05.12.2013. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1996/1996_10_25-Interkulturelle-Bildung.pdf
- Kultusministerkonferenz. (2002). *Bericht "Zuwanderung"*. Beschluss der Kultusministerkonferenz vom 24.05.2002 i.d.F. vom 16.11.2006. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2002/2002_05_24-Zuwanderung.pdf
- Kultusministerkonferenz. (2004). *Standards für die Lehrerbildung*. *Bildungswissenschaften*. Beschluss der Kultusministerkonferenz vom 16.12.2004. Verfügbar unter http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- Kultusministerkonferenz. (2007). *Integration als Chance – Gemeinsam für mehr Chancengerechtigkeit. Gemeinsame Erklärung der Kultusministerkonferenz und der Organisationen von Menschen mit Migrationshintergrund*. Beschluss der Kultusministerkonferenz vom 13.12.2007. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2007/2007_12_13-Integration.pdf

- Kultusministerkonferenz. (2015). *Darstellung von kultureller Vielfalt, Integration und Migration in Bildungsmedien. Gemeinsame Erklärung der Kultusministerkonferenz, der Organisationen von Menschen mit Migrationshintergrund und der Bildungsmedienverlage*. Beschluss der Kultusministerkonferenz vom 08.10.2015. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_10_08-Darstellung-kultureller-Vielfalt.pdf
- Kunda, Z. & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors. A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284–308. doi: 10.1037/0033-295x.103.2.284
- Ladas, A., Frantzidis, C., Bamidis, P. & Vivas, A. B. (2014). Eye blink rate as a biological marker of mild cognitive impairment. *International Journal of Psychophysiology*, *93*, 12–16. doi: 10.1016/j.ijpsycho.2013.07.010
- Landis, R. J. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi: 10.2307/2529310
- Leacock, E. B. (1969). *Teaching and learning in city schools a comparative study*. New York, NY: Basic Books.
- Ledger, H. (2013). The effect cognitive load has on eye blinking. *The Plymouth Student Scientist*, *6*, 206–223.
- Lee, J. (2002). Racial and ethnic achievement gap trends. Reversing the progress toward equity? *Educational Researcher*, *31*, 3–13. doi: 10.3102/0013189x031001003
- Lepore, L. & Brown, R. (1997). Category and stereotype activation. Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*, 275–287. doi: 10.1037/0022-3514.72.2.275
- Lesaux, N. K. & Harris, J. R. (2013). Children from low-income and non-English speaking families in the US. In N. McElvany, M. M. Gebauer, W. Bos & H. G. Holtappels (Hrsg.), *Sprachliche, kulturelle und soziale Heterogenität in der Schule als Herausforderung und Chance der Schulentwicklung. Daten, Beispiele und Perspektiven* (Jahrbuch der Schulentwicklung, Bd. 17, S. 130–149). Weinheim: Beltz.
- Liebwein, S. (2008). *Erziehung und soziale Milieus. Elterliche Erziehungsstile in milieuspezifischer Differenzierung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. *Zeitschrift für Pädagogik* (Beiheft 51), 47–70.
- Lippmann, W. (1922). *Public opinion* (11. unveränderte Aufl. v. 1949). New York, NY: The Macmillan Company.
- Locksley, A., Hepburn, C. & Ortiz, V. (1982a). On the effect of social stereotypes on judgements of individuals. A comment on Grant and Holmes's "The integration of implicit personality theory schemas and stereotypic images". *Social Psychology Quarterly*, *45*, 270–273. doi: 10.2307/3033923
- Locksley, A., Hepburn, C. & Ortiz, V. (1982b). Social stereotypes and judgments of individuals. An instance of the base-rate fallacy. *Journal of Experimental Social Psychology*, *18*, 23–42. doi: 10.1016/0022-1031(82)90079-8
- Loftus, E. F. & Greene, E. (1980). Warning. Even memory for faces may be contagious. *Law and Human Behavior*, *4*, 323–334. doi: 10.1007/bf01040624
- Lohbeck, A., Nitkowski, D., Petermann, F. & Petermann, U. (2014). Erfassung von Schülerselbsteinschätzungen zum schulbezogenen Sozial- und Lernverhalten – Validierung der Schülereinschätzliste für Sozial- und Lernverhalten (SSL). *Zeitschrift für Erziehungswissenschaft*, *17*, 701–722. doi: 10.1007/s11618-014-0582-6
- Lord, C. G., Umezaki, K. & Darley, J. M. (1990). Developmental differences in decoding the meanings of the appraisal actions of teachers. *Child Development*, *61*, 191–200. doi: 10.1111/j.1467-8624.1990.tb02771.x
- Lorenz, C. & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, *23*, 211–222. doi: 10.1024/1010-0652.23.34.211
- Lorenz, G., Gentrup, S., Kristen, C., Stanat, P. & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *68*, 89–111. doi: 10.1007/s11577-015-0352-3
- Lüdtko, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, *58*, 103–117. doi: 10.1026/0033-3042.58.2.103

- Ludwig, P. H. (1991). *Sich selbst erfüllende Prophezeiungen im Alltagsleben. Theorie und empirische Basis von Erwartungseffekten und Konsequenzen für die Pädagogik, insbesondere für die Gerontagogik*. Stuttgart: Verlag für Angewandte Psychologie.
- Ludwig, P. H. & Ludwig, H. (Hrsg.). (2007). *Erwartungen in himmelblau und rosarot. Effekte, Determinanten und Konsequenzen von Geschlechterdifferenzen in der Schule*. Weinheim: Juventa Verlag.
- Lykken, D. T. (1982). Research with twins. The concept of emergenesis. *Psychophysiology*, *19*, 361–372. doi: 10.1111/j.1469-8986.1982.tb02489.x
- Maaz, K., Baethge, M., Brugger, P., Füssel, H.-P., Hertmeier, H.-W., Rauschenbach, T. et al. (Hrsg.). (2016). *Bildung in Deutschland 2016. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann Verlag.
- Machunsky, M. (2008). Substereotypisierung. In L.-E. Peterson & B. Six (Hrsg.), *Stereotype, Vorurteile und Diskriminierung. Theorien, Befunde und Interventionen* (S. 45–52). Weinheim: Beltz.
- Macrae, C. N. & Bodenhausen, G. V. (2000). Social cognition. Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. doi: 10.1146/annurev.psych.51.1.93
- Macrae, C., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. & Castelli, L. (1997). On the activation of social stereotypes. The moderating role of processing objectives. *Journal of Experimental Social Psychology*, *33*, 471–489. doi: 10.1006/jesp.1997.1328
- Macrae, C., Mitchell, J. P. & Pendry, L. F. (2002). What's in a forename? Cue familiarity and stereotypical thinking. *Journal of Experimental Social Psychology*, *38*, 186–193. doi: 10.1006/jesp.2001.1496
- Madon, S., Jussim, L. & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, *72*, 791–809. doi: 10.1037//0022-3514.72.4.791
- Madon, S., Willard, J., Guyll, M. & Scherr, K. C. (2011). Self-fulfilling prophecies. Mechanisms, power, and links to social problems. *Social and Personality Psychology Compass*, *5*, 578–590. doi: 10.1111/j.1751-9004.2011.00375.x
- Maidowsky, W. (2003). *Anatomie des Auges. Mit einer Einführung in die allgemeine Zellen- und Gewebelehre* (6. Aufl.). Heidelberg: DOZ-Verlag.
- Maio, G. R. & Haddock, G. (2010). *The psychology of attitudes and attitude change*. Los Angeles, CA: Sage.
- Manis, M., Biernat, M. & Nelson, T. F. (1991). Comparison and expectancy processes in human judgment. *Journal of Personality and Social Psychology*, *61*, 203–211. doi: 10.1037//0022-3514.61.2.203
- Manis, M. & Paskewitz, J. R. (1984). Judging psychopathology. Expectation and contrast. *Journal of Experimental Social Psychology*, *20*, 363–381. doi: 10.1016/0022-1031(84)90032-5
- Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In F. Rösler (Hrsg.), *Ergebnisse und Anwendungen der Psychophysiologie* (Enzyklopädie der Psychologie, S. 799–864). Göttingen: Hogrefe.
- Marín, G. (1984). Stereotyping Hispanics. The differential effect of research method, label, and degree of contact. *International Journal of Intercultural Relations*, *8*, 17–27. doi: 10.1016/0147-1767(84)90005-1
- Marks, G. N. (2004). The measurement of socio-economic inequalities in education. A further comment. *Acta Sociologica*, *47*, 91–93. doi: 10.1177/0001699304041553
- Marsella, A. J. & Ring, E. (2003). Human migration and immigration. An overview. In L. L. Adler & U. P. Gielen (Eds.), *Migration. Immigration and emigration in international perspective* (pp. 3–22). Westport, CT: Praeger.
- Marsh, H. W. & Shavelson, R. (1985). Self-concept. Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*, 107–123. doi: 10.1207/s15326985ep2003_1
- Marshall, S. P. (2000). *Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity [US 6090051 A]*. Washington, DC.
- Marshall, S. P. (2002). *The index of cognitive activity. Measuring cognitive workload* (IEEE 7th Human Factors Meeting). Scottsdale, AZ. doi: 10.1109/hfpp.2002.1042860
- Martiny, S. E., Götz, T. & Keller, M. (2013). Emotionsregulation im Kontext von Stereotype Threat. Die Reduzierung der Effekte negativer Stereotype bei ethnischen Minderheiten. In P. Genkova, T. Ringeisen & F. T. L. Leong (Hrsg.), *Handbuch Stress und Kultur. Interkulturelle und kulturvergleichende Perspektiven* (S. 397–415). Wiesbaden: Springer.
- Matt, C. & Tobisch, E. (2017). *Mediale Ansicht des Auges und der Augenmuskulatur*. Augsburg [Unveröffentlichte Zeichnung].

- Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics*, 11, 204–209. doi: 10.1214/aoms/1177731915
- Maykovich, M. K. (1972). Reciprocity in racial stereotypes. White, black, and yellow. *American Journal of Sociology*, 77, 876–897. doi: 10.1086/225229
- McCauley, C., Stitt, C. L. & Segal, M. (1980). Stereotyping. From prejudice to prediction. *Psychological Bulletin*, 87, 195–208. doi: 10.1037/0033-2909.87.1.195
- McCombs, R. C. & Gay, J. (1988). Effects of race, class, and IQ information on judgments of parochial grade school teachers. *The Journal of Social Psychology*, 128, 647–652. doi: 10.1080/00224545.1988.9922918
- McConahay, J. B. (1986). Modern Racism, Ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, Discrimination, and Racism* (pp. 91–125). San Diego, CA: Academic Press.
- McConkie, G. W. & Yang, S.-N. (2003). How cognition affects eye movements during reading. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The mind's eye. Cognitive and applied aspects of eye movement research* (pp. 413–427). Amsterdam: Elsevier.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W. et al. (2009). Diagnostische Fähigkeiten von Lehrkräften. Bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23, 223–235. doi: 10.1024/1010-0652.23.34.223
- McKown, C. & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology*, 32, 159–184. doi: 10.1111/j.1559-1816.2002.tb01425.x
- McKown, C. & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, 74, 498–515. doi: 10.1111/1467-8624.7402012
- Mehringer, V. (2013). *Weichenstellungen in der Grundschule. Sozial-Integration von Kindern mit Migrationshintergrund* (Interkulturelle Bildungsforschung, Bd. 22). Münster: Waxmann.
- Meichenbaum, D. H., Bowers, K. S. & Ross, R. R. (1969). A behavioral analysis of teacher expectancy effect. *Journal of Personality and Social Psychology*, 13, 306–316. doi: 10.1037/h0028470
- Meinold, P. E. (2005). *Psychologie des Lidschlags. Eine literatur- und methodenkritische Studie*. Dissertationsschrift, Universität Köln. Köln.
- Meissel, K., Meyer, F., Yao, E. S. & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments. Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. doi: 10.1016/j.tate.2017.02.021
- Merkens, H. & Wessel, A. (2002). *Zur Genese von Bildungsentscheidungen. Eine empirische Studie in Berlin und Brandenburg* (Jugendforschung aktuell, Bd. 7). Baltmannsweiler: Schneider.
- Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, 8, 193. doi: 10.2307/4609267
- Merton, R. K. (1968). *Social theory and social structure* (3rd ed.). New York, NY: The Free Press.
- Meyer, W.-U. (1984). *Das Konzept von der eigenen Begabung*. Bern: Verlag Hans Huber.
- Mierke, J. (2004). *Kognitive Prozesse bei der indirekten Messung von Einstellungen mit dem impliziten Assoziationstest*. Inaugural Dissertation, Albert Ludwigs Universität. Freiburg im Breisgau. Verfügbar unter <https://freidok.uni-freiburg.de/dnb/download/1591>
- Miller, D. T. & Turnbull, W. (1986). Expectancies and Interpersonal Processes. *Annual Review of Psychology*, 37, 233–256. doi: 10.1146/annurev.ps.37.020186.001313
- Miller, G. A. (1956). The magical number seven, plus or minus two. Some limits on our capacity for processing information. *Psychological Review*, 101, 343–352. doi: 10.1037//0033-295x.101.2.343
- Millisecond Software. (2016) Inquisit 5 Lab (Windows) [Computer software]: Millisecond Software. Verfügbar unter <http://www.millisecond.com/>
- Moors, A. & de Houwer, J. (2006). Automaticity. A theoretical and conceptual analysis. *Psychological Bulletin*, 132, 297–326. doi: 10.1037/0033-2909.132.2.297
- Morrison, R. E. (1984). *Manipulation of stimulus onset delay in reading. Evidence for parallel programming of saccades*. Doctoral dissertation, University of Massachusetts Amherst. Amherst, MA. Verfügbar unter http://scholarworks.umass.edu/dissertations_1/1721

- Müller, K. & Ehmke, T. (2013). Soziale Herkunft als Bedingung der Kompetenzentwicklung. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 245–274). Münster: Waxmann.
- Mummendey, H. D., Bolten, H. G. & Isermann-Gerke, M. (1982). Experimentelle Überprüfung des Bogus-Pipeline-Paradigmas. Einstellungen gegenüber Türken, Deutschen und Holländern. *Zeitschrift für Sozialpsychologie*, *13*, 300–311.
- Neisser, U. (1976). *Cognition and reality. Principles and implications of cognitive psychology*. San Francisco, CA: W. H. Freeman and company.
- Neuberg, S. L. & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology*, *53*, 431–444. doi: 10.1037/0022-3514.53.3.431
- Neumann, U. (2009). Integrationspolitik als Rahmen für den bildungspolitischen Umgang mit Heterogenität – das Beispiel Hamburg. In V. King & H.-C. Koller (Hrsg.), *Adoleszenz – Migration – Bildung. Bildungsprozesse Jugendlicher und junger Erwachsener mit Migrationshintergrund* (2. Aufl., S. 259–278). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Neumann, U. & Reuter, L. R. (2004). Interkulturelle Bildung in den Lehrplänen. Neuere Entwicklungen. *Zeitschrift für Pädagogik*, *50*, 803–817.
- Nicholson-Crotty, J., Grissom, J. A. & Nicholson-Crotty, S. (2011). Bureaucratic representation, distributional equity, and democratic values in the administration of public programs. *The Journal of Politics*, *73*, 582–596. doi: 10.1017/S0022381611000144
- Nisbett, R. E. & Wilson, T. D. (1977). The halo effect. Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250–256. doi: 10.1037/0022-3514.35.4.250
- Oakes, P. J. (1987). The salience of social groups. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher & M. S. Wetherell (Eds.), *Rediscovering the social group. A self-categorization theory* (pp. 117–141). Oxford: Blackwell.
- Oakes, P. J. (1996). The categorization process. Cognition and the group in the social psychology of stereotyping. In W. P. Robinson (Ed.), *Social groups & identities. Developing the legacy of Henri Tajfel* (pp. 95–119). Oxford: Butterworth-Heinemann.
- Organisation for Economic Co-operation and Development. (2016a). *Bildung auf einen Blick 2016. OECD-Indikatoren*. Paris: Bertelsmann Verlag.
- Organisation for Economic Co-operation and Development. (2016b). *PISA 2015 Ergebnisse (Band I). Exzellenz und Chancengerechtigkeit in der Bildung*. Bielefeld: Bertelsmann Verlag.
- Oerke, B., McElvany, N., Ohle, A., Ullrich, M. & Horz, H. (2016). Verbessert sich die diagnostische Urteilsgenauigkeit von Lehrkräften bei längerem Kontakt mit der Klasse? *Psychologie in Erziehung und Unterricht*, *63*, 34. doi: 10.2378/peu2016.art04d
- Olson, M. A. & Fazio, R. H. (2009). Implicit and explicit measures of attitudes. The perspective of the MODE model. In R. E. Petty, R. H. Fazio & P. Briñol (Eds.), *Attitudes. Insights from the new implicit measures* (pp. 19–63). New York, NY: Taylor & Francis.
- Orchard, L. N. & Stern, J. A. (1991). Blinks as an index of cognitive activity during reading. *Integrative Physiological and Behavioral Science*, *26*, 108–116. doi: 10.1007/BF02691032
- Osad'án, R. (2012). Gender stereotypes and elementary school teachers. *Acta Technologica Dubnicae*, *2* (2), 25–29. doi: 10.1515/atd-2015-0057
- Paas, F., Tuovinen, J. E., Tabbers, H. & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63–71. doi: 10.1207/s15326985ep3801_8
- Papastefanou, G. (2013). *Experimentelle Validierung eines Sensor-Armbandes zur mobilen Messung physiologischer Stress-Reaktionen*. Mannheim: Leibniz-Institut für Sozialwissenschaften. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-339493>
- Park, R. E. (1924). The concept of social distance as applied to the study of racial attitudes and racial relations. *Journal of Applied Sociology*, *8*, 339–344. Verfügbar unter https://brocku.ca/MeadProject/Park/Park_1924.html
- Parks, F. R. & Kennedy, J. H. (2007). The impact of race, physical attractiveness, and gender on education majors' and teachers' perceptions of student competence. *Journal of Black Studies*, *37*, 936–943. doi: 10.1177/0021934705285955

- Paulsen, F. & Waschke, J. (2017). *Sobotta Atlas der Anatomie des Menschen. Allgemeine Anatomie und Bewegungsapparat, innere Organe, Kopf, Hals und Neuroanatomie* (23. Aufl.). München: Elsevier.
- Pekrun, R. (1983). *Schulische Persönlichkeitsentwicklung. Theorieentwicklung und empirische Erhebungen zur Persönlichkeitsentwicklung von Schülern der 5. bis 10. Klassenstufe*. Frankfurt a. M.: Peter Lang.
- Pendry, L. F. & Macrae, C. N. (1996). What the disinterested perceiver overlooks. Goal-directed social categorization. *Personality & Social Psychology Bulletin*, 22, 249–256. doi: 10.1177/0146167296223003
- Peterson, L.-E. & Six, B. (Hrsg.). (2008). *Stereotype, Vorurteile und Diskriminierung. Theorien, Befunde und Interventionen*. Weinheim: Beltz.
- Pettigrew, T. F. & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, 25, 57–75. doi: 10.1002/ejsp.2420250106
- Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49, 65–85. doi: 10.1146/annurev.psych.49.1.65
- Pettigrew, T. F. & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, 38, 922–934. doi: 10.1002/ejsp.504
- Petty, R. E. & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (vol. 19, pp. 123–205). New York, NY: Academic Press.
- Petty, R. E. & Wegener, D. T. (1999). The elaboration likelihood model. Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 41–72). New York, NY: Guilford Press.
- Pit-ten Cate, I. M., Krolak-Schwerdt, S. & Glock, S. (2016). Accuracy of teachers' tracking decisions. Short- and long-term effects of accountability. *European Journal of Psychology of Education*, 31, 225–243. doi: 10.1007/s10212-015-0259-4
- Plange, N. (2017). Neuroophthalmologie. In P. Walter & N. Plange (Hrsg.), *Basiswissen Augenheilkunde* (S. 349–372). Berlin: Springer.
- Pollack, I. (1958). The cocktail party revisited. *The Journal of the Acoustical Society of America*, 30, 669–699. doi: 10.1121/1.1929938
- Ponder, E. & Kennedy, W. P. (1927). On the act of blinking. *Journal of Experimental Physiology*, 18, 89–110. doi: 10.1113/expphysiol.1927.sp000433
- Porges, S. W. (2001). The polyvagal theory. Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology*, 42, 123–146. doi: 10.1016/s0167-8760(01)00162-3
- 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, 64–76. doi: 10.1080/00220671.2012.667010
- Praetorius, A.-K., Drexler, K., Rösch, L., Christophel, E., Heyne, N., Scheunpflug, A. et al. (2014). Judging students' self-concepts within 30s? Investigating judgement accuracy in a zero-acquaintance situation. *Learning and Individual Differences*, 37, 231–236. doi: 10.1016/j.lindif.2014.11.015
- Praetorius, A.-K., Greb, K., Lipowsky, F. & Good, M. (2010). Lehrkräfte als Diagnostiker. Welche Rolle spielt die Schülerleistung bei der Einschätzung von mathematischen Selbstkonzepten? *Journal of Educational Research Online*, 2, 121–144. Verfügbar unter http://www.pedocs.de/volltexte/2011/4570/pdf/JERO_2010_1_Praetorius_et_al_Lehrkraefte_als_Diagnostiker_S121_D_A.pdf
- 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. doi: 10.2378/peu2010.art30d
- Prenzel, M., Heidemeier, H., Ramm, G., Hohensee, F. & Ehmke, T. (2004). Soziale Herkunft und mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland* (S. 273–282). Ergebnisse des zweiten internationalen Vergleichs. Münster: Waxmann.
- Pryor, J. B. & Ostrom, T. M. (1981). The cognitive organization of social information. A converging-operations approach. *Journal of Personality and Social Psychology*, 41, 628–641. doi: 10.1037/0022-3514.41.4.628
- Psyhyrembel, W. (Hrsg.). (1997). *Klinisches Wörterbuch* (258. Aufl.). Berlin: de Gruyter.
- Rakhkochkine, A. & Kotthoff, H.-G. (2015). Internationale Lehrermobilität. *Bildung und Erziehung*, 68, 409–416. doi: 10.7788/bue-2015-0402

- Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D. et al. (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Ramm, G., Prenzel, M., Heidemeier, H. & Walter, O. (2004). Soziokulturelle Herkunft. Migration. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland* (S. 254–273). Ergebnisse des zweiten internationalen Vergleichs. Münster: Waxmann.
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2010). *Quantitative Methoden. Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (Bd. 2, 3. Aufl.). Berlin: Springer.
- Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction. A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, 76, 85–97. doi: 10.1037/0022-0663.76.1.85
- 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. doi: 10.3102/0002831210374874
- Ready, D. D. & Chu, E. M. (2015). Sociodemographic inequality in early literacy development. The role of teacher perceptual accuracy. *Early Education and Development*, 26, 970–987. doi: 10.1080/10409289.2015.1004516
- Richert, P. (2012). *Elternentscheidung versus Lehrerdia gnose. Der Übergang von der Grundschule zur Sekundarstufe*. Bad Heilbrunn: Julius Klinkhardt.
- Riley, T. & Ungerleider, C. (2012). Self-fulfilling prophecy. How teachers' attributions, expectations, and stereotypes influence the learning opportunities afforded aboriginal students. *Canadian Journal of Education*, 35, 303–333. Verfügbar unter <http://files.eric.ed.gov/fulltext/EJ975282.pdf>
- Rist, R. (1970). Student social class and teacher expectations. The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, 40, 411–451. doi: 10.17763/haer.40.3.h0m026p670k618q3
- Ritts, V., Patterson, M. L. & Tubbs M. E (1992). Expectations, impressions, and judgments of physically attractive students: A review. *Review of Educational Research*, 62, 413–426. doi: 10.3102/00346543062004413
- Rosenberg, M. J. & Hovland, C. I. (1969). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson & J. W. Brehm (Eds.), *Attitude organization and change. An analysis of consistency among attitude components* (4th ed., pp. 1–14). New Haven: Yale University Press.
- Rosenberg, S. & Jones, R. (1972). A method for investigating and representing a person's implicit theory of personality. Theodore Dreiser's view of people. *Journal of Personality and Social Psychology*, 22, 372–386. doi: 10.1037/h0032891
- Rosenthal, R. (1974). On the social psychology of the self-fulfilling prophecy. Further evidence for pygmalion effects and their mediating mechanisms. In B. E. Collins (Ed.), *Social psychology* (vol. 53, pp. 2–28). New York, NY: MSS Modular Publications.
- Rosenthal, R. (1993). Interpersonal expectations. Some antecedents and some consequences. In P. D. Blanck (Ed.), *Interpersonal expectations. Theory, research and applications* (pp. 3–24). Cambridge: Cambridge University Press.
- Rosenthal, R. (1994). Interpersonal expectancy effects. A 30-year perspective. *Current Directions in Psychological Science*, 3, 176–179. doi: 10.1111/1467-8721.ep10770698
- Rosenthal, R. & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, 8, 183–189. doi: 10.1002/bs.3830080302
- Rosenthal, R. & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3, 16–20. doi: 10.1007/bf02322211
- Rosenthal, R. & Jacobson, L. (1992). *Pygmalion in the classroom. Teacher expectation and pupils' intellectual development* (erweiterte Neuauflage). New York, NY: Irving Publ.
- Rösler F. (Hrsg.). (1998). *Grundlagen und Methoden der Psychophysiologie* (Enzyklopädie der Psychologie (Bd. 4). Göttingen: Hogrefe.
- Rössel, J. & Pape, S. (2010). Was ist ein typischer Arbeiter? Stereotype über soziale Schichten. *Gruppendynamik und Organisationsberatung*, 41, 57–71. doi: 10.1007/s11612-010-0101-z

- Rothbart, M., Fulero, S., Jensen, C., Howard, J. & Birrell, P. (1978). From individual to group impressions. Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology*, 14, 237–255. doi: 10.1016/0022-1031(78)90013-6
- Rothbart, M. & John, O. P. (1985). Social categorization and behavioral episodes. A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues*, 41 (3), 81–104. doi: 10.1111/j.1540-4560.1985.tb01130.x
- Rubie-Davies, C., Hattie, J. & Hamilton, R. (2006). Expecting the best for students. Teacher expectations and academic outcomes. *The British Journal of Educational Psychology*, 76, 429–444. doi: 10.1348/000709905X53589
- Rubie-Davies, C. M. (2007). Classroom interactions. Exploring the practices of high- and low-expectation teachers. *The British Journal of Educational Psychology*, 77, 289–306. doi: 10.1348/000709906X101601
- Rubovits, P. C. & Maehr, M. L. (1971). Pygmalion analyzed. Toward an explanation of the Rosenthal-Jacobson findings. *Journal of Personality and Social Psychology*, 19, 197–203. doi: 10.1037/h0031526
- Rubovits, P. C. & Maehr, M. L. (1973). Pygmalion black and white. *Journal of Personality and Social Psychology*, 25, 210–218. doi: 10.1037/h0034080
- Rudolph, U., Böhm, R. & Lummer, M. (2007). Ein Vorname sagt mehr als 1000 Worte. *Zeitschrift für Sozialpsychologie*, 38, 17–31. doi: 10.1024/0044-3514.38.1.17
- Rudolph, U. & Spörrle, M. (1999). Alter, Attraktivität und Intelligenz von Vornamen. Wortnormen für Vornamen im Deutschen. *Experimental Psychology*, 46, 115–128. doi: 10.1026//0949-3964.46.2.115
- Rüesch, P. (1998). *Spielt die Schule eine Rolle? Schulische Bedingungen ungleicher Bildungschancen von Immigrantenkindern – eine Mehrebenenanalyse*. Berlin: Peter Lang.
- Sagar, H. A. & Schofield, J. W. (1980). Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, 39, 590–598. doi: 10.1037/0022-3514.39.4.590
- Salamone, J. D. (2009). Dopamine, effort, and decision making: theoretical comment on Bardgett et al. (2009). *Behavioral Neuroscience*, 123, 463–467. doi: 10.1037/a0015381
- Sbarra, D. A. & Pianta, R. C. (2001). Teacher ratings of behavior among African American and Caucasian children during the first two years of school. *Psychology in the Schools*, 38, 229–238. doi: 10.1002/pits.1013
- Schaal, B. (1997). *Willentliche Kontrolle stereotypen Denkens. Intentionseffekte auf die Aktivierung von Stereotypen*. Dissertationsschrift, Universität Konstanz. Konstanz. Verfügbar unter https://kops.uni-konstanz.de/bitstream/handle/123456789/11308/262_1.pdf?sequence=1
- Schaarschmidt, U. & Kieschke, U. (2007). *Beanspruchungsmuster im Lehrerberuf. Ergebnisse und Schlussfolgerungen aus der Potsdamer Lehrerstudie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schaller, M., Park, J. H. & Mueller, A. (2003). Fear of the dark. Interactive effects of beliefs about danger and ambient darkness on ethnic stereotypes. *Personality & Social Psychology Bulletin*, 29, 637–649. doi: 10.1177/0146167203029005008
- Schandry, R. (2003). *Biologische Psychologie*. Weinheim: Beltz.
- Schandry, R. (2008). Messmethoden des autonomen Nervensystems. In A. v. Leupoldt & T. Ritz (Hrsg.), *Verhaltensmedizin. Psychobiologie, Psychopathologie und klinische Anwendung* (S. 112–123). Stuttgart: Kohlhammer.
- Schiepe-Tsika, A., Schöps, K., Rönnebeck, S., Köller, O. & Prenzel, M. (2013). Naturwissenschaftliche Kompetenz in Pisa 2012. Ergebnisse und Herausforderungen. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 189–215). Münster: Waxmann.
- Schimpl-Neimanns, B. (2000). Soziale Herkunft und Bildungsbeteiligung. Empirische Analysen zu herkunftsspezifischen Bildungsungleichheiten zwischen 1950 und 1989. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 52, 636–669. doi: 10.1007/s11577-000-0102-y
- Schloz, C. & Dresel, M. (2011). Implizite Fähigkeitstheorien und Fähigkeitsselbstkonzepte im Grundschulalter. Ein Überblick und Ergebnisse einer Studie im Fach Deutsch. In F. Hellmich (Hrsg.), *Selbstkonzepte im Grundschulalter. Modelle, empirische Ergebnisse, pädagogische Konsequenzen* (S. 81–99). Stuttgart: Kohlhammer.

- Schmid Mast, M. & Krings, F. (2008). Stereotype und Informationsverarbeitung. In L.-E. Peterson & B. Six (Hrsg.), *Stereotype, Vorurteile und Diskriminierung. Theorien, Befunde und Interventionen* (S. 33–44). Weinheim: Beltz.
- Schneider, D. J. (1973). Implicit personality theory. A review. *Psychological Bulletin*, 79, 294–309. doi: 10.1037/h0034496
- Schneider, J., Yemane, R. & Weinmann, M. (2014). *Diskriminierung am Ausbildungsmarkt. Ausmaß, Ursachen und Handlungsperspektiven* (Sachverständigenrat deutscher Stiftungen für Integration und Migration, Hrsg.). Verfügbar unter http://www.bosch-stiftung.de/content/language1/downloads/SVR-FB_Diskriminierung-am-Ausbildungsmarkt.pdf.
- Schofield, J. W. & Alexander, K. M. (2012). Stereotype threat, Erwartungseffekte und organisatorische Differenzierung. Schulische Leistungsbarrieren und Ansätze zur Überwindung. In S. Fürstenau & M. Gomolla (Eds.), *Migration und schulischer Wandel. Leistungsbeurteilung* (pp. 65–87). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schöler, H., Guggenmos, J., Hasselbach, P. & Iseke, A. (2005). *Sprachliche Leistungen in der Einschulungsuntersuchung. Ein Vergleich der Jahrgänge 1999 bis 2004 in der Stadt Münster*. Arbeitsbericht aus dem Forschungsprojekt "Differenzdiagnostik" (Nr. 21). Heidelberg: Pädagogische Hochschule Heidelberg. Verfügbar unter http://www01.ph-heidelberg.de/wp/schoeler/datein/nr21_01-01-2006.pdf
- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2012). *SESSKO. Skalen zur Erfassung des schulischen Selbstkonzepts* (2. Aufl.). Göttingen: Hogrefe.
- 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* (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. doi: 10.1024/1010-0652.23.34.237
- Schrader, F.-W. (2010). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 102–108). Weinheim: Beltz.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern. Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27–52.
- Schrader, F.-W. & Helmke, A. (1990). Lassen sich Lehrer bei der Leistungsbeurteilung von sachfremden Gesichtspunkten leiten? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22, 312–324.
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 45–58). Weinheim: Beltz.
- Schuette, R. A. & Fazio, R. H. (2016). Attitude accessibility and motivation as determinants of biased processing. A test of the MODE model. *Personality & Social Psychology Bulletin*, 21, 704–710. doi: 10.1177/0146167295217005
- Schunk, D. H., Pintrich, P. R. & Meece, J. L. (2008). *Motivation in education. Theory, research, and application* (3rd ed.). Upper Saddle River, NJ: Pearson.
- Schünke, M., Schulte, E., Schumacher, U., Voll, M. & Wesker, K. (2012). *Kopf, Hals und Neuroanatomie. Prometheus Lernatlas Anatomie* (3. Aufl.). Stuttgart: Thieme.
- Schwarz, K. A., Pfister, R. & Büchel, C. (2016). Rethinking explicit expectations. Connecting placebos, social cognition, and contextual perception. *Trends in Cognitive Sciences*, 20, 469–480. doi: 10.1016/j.tics.2016.04.001
- Schwarz, N. & Bohner, G. (2001). The construction of attitudes. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology. Intraindividual processes* (pp. 437–457). Malden, MA: Blackwell.
- Schwarz, N. & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being. Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513–523. doi: 10.1037//0022-3514.45.3.513
- Schwarz, N. (2007). Attitude construction. Evaluation in context. *Social Cognition*, 25, 638–656. doi: 10.1521/soco.2007.25.5.638
- Schwippert, K., Bos, W. & Lankes, E.-M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valentin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 265–302). Münster: Waxmann.

- Schwippert, K., Hornberg, S., Freiberg, M. & Stubbe, T. C. (2007). Lesekompetenzen von Kindern mit Migrationshintergrund im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes et al. (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 249–269). 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 im internationalen Vergleich* (S. 191–207). Münster: Waxmann.
- Seaver, W. B. (1973). Effects of naturally induced teacher expectancies. *Journal of Personality and Social Psychology*, 28, 333–342. doi: 10.1037/h0035186
- Seitz, S. (2006). *Migrantenkinder und positive Schulleistungen*. Bad Heilbrunn: Julius Klinkhardt.
- Sensomotoric Instruments. (2012). *RED-m eye tracking system manual. Version 3.2*. Verfügbar unter http://twiki.cis.rit.edu/twiki/pub/MVRL/SmiTracker/RED-m_User_Manual.pdf.
- Sensomotoric Instruments. (2014a). *RED-m. Technical specification*. Verfügbar unter http://www.smivision.com/fileadmin/user_upload/downloads/product_flyer/prod_smi_redmtechspecs.pdf
- Sensomotoric Instruments. (2014b) SMI BeGaze Eye tracking analysis software 3.6 [Computer software]. Berlin.
- Sensomotoric Instruments. (2014c) SMI Experiment Center Software 3.6 [Computer software]. Berlin.
- Shavelson, R. J. & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51, 455. doi: 10.2307/1170362
- Shibutani, H., Sakata, H. & Hyvärinen, J. (1984). Saccade and blinking evoked by microstimulation of the posterior parietal association cortex of the monkey. *Experimental Brain Research*, 55, 1–8. doi: 10.1007/bf00240493
- Siegle, G. J., Ichikawa, N. & Steinhauer, S. (2008). Blink before and after you think. Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45, 679–687. doi: 10.1111/j.1469-8986.2008.00681.x
- Sielschott, S. (2010). „Arbeitslos und keine Ahnung – Hier werden Sie geholfen!“. Zum Framing stereotyper Bewertungen Arbeitsloser in den Regionalzeitungen Nordkurier und Volksstimme. *Journal für Psychologie*, 18 (2), 1–19. Verfügbar unter <https://www.journal-fuer-psychologie.de/index.php/jfp/article/download/72/98>
- Sigall, H. & Page, R. (1971). Current stereotypes. A little fading, a little faking. *Journal of Personality and Social Psychology*, 18, 247–255. doi: 10.1037/h0030839
- Sigel, I. S. (1994). Elterliche Überzeugungen und deren Rolle bei der kognitiven Entwicklung von Kindern. *Unterrichtswissenschaft*, 22, 160–181.
- Simons, D. J. (o. J.). *The monkey business illusion* [Video]. Verfügbar unter <http://www.theinvisiblegorilla.com/videos.html>
- Simons, D. J. (2010). Monkeying around with the gorillas in our midst. Familiarity with an inattentional-blindness task does not improve the detection of unexpected events. *I-Perception*, 1, 3–6. doi: 10.1068/i0386
- Skinner, B. F. (1938). *The behavior of organisms. An experimental analysis*. New York, NY: Appleton-Century-Crofts.
- Solga, H. (2003). Das Paradox der integrierten Ausgrenzung von gering qualifizierten Jugendlichen. *Aus Politik und Zeitgeschichte* (B 21-22), 17–25. Verfügbar unter <http://www.bpb.de/apuz/27622/das-paradox-der-integrierten-ausgrenzung-von-gering-qualifizierten-jugendlichen?p=all>
- Sorhagen, N. S. (2013). Early teacher expectations disproportionately affect poor children's high school performance. *Journal of Educational Psychology*, 105, 465–477. doi: 10.1037/a0031754
- Spearman, C. (1923). *The nature of 'intelligence' and the principles of cognition*. London: Macmillan and co.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85–95. doi: 10.1024/1010-0652.19.12.85
- Spitz, H. H. (1999). Beleaguered pygmalion. A history of the controversy over claims that teacher expectancy raises intelligence. *Intelligence*, 27, 199–234. doi: 10.1016/s0160-2896(99)00026-4
- Sprietsma, M. (2013). Discrimination in grading. Experimental evidence from primary school teachers. *Empirical Economics*, 45, 523–538. doi: 10.1007/s00181-012-0609-x

- Staatsinstitut für Schulqualität und Bildungsforschung. (2008). *Formulierungshilfen zu den Grundschulzeugnissen*. Verfügbar unter <https://www.isb.bayern.de/grundschule/materialien/f/formulierungshilfen-zu-den-grundschul-zeugnissen/>
- Stamm, M. (2007). Begabtenförderung und soziale Herkunft. Befunde zu den verborgenen Mechanismen ihrer Interaktion. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 27, 227–242.
- Stamm, M. (2009). *Begabte Minoritäten*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stanat, P. (2006a). Disparitäten im schulischen Erfolg. Forschungsstand zur Rolle des Migrationshintergrunds. *Unterrichtswissenschaft*, 34, 98–124.
- Stanat, P. (2006b). 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.
- Standesamt München. (o.J.). *Die beliebtesten Vornamen. Statistiken des Standesamts München*. Verfügbar unter <https://www.muenchen.de/rathaus/Stadtverwaltung/Kreisverwaltungsreferat/Standesamt-und-Urkunden/Geburt/Namensgebung.html>
- Standesamt Nürnberg. (o.J.). *Vornamensstatistik*. Verfügbar unter <https://www.nuernberg.de/internet/standesamt/statistik.html>
- Stang, J. & Urhahne, D. (2016a). Stabilität, Bezugsnormorientierung und Auswirkungen der Urteilsgenauigkeit. *Zeitschrift für Pädagogische Psychologie*, 30, 251–262. doi: 10.1024/1010-0652/a000190
- Stang, J. & Urhahne, D. (2016b). Wie gut schätzen Lehrkräfte Leistung, Konzentration, Arbeits- und Sozialverhalten ihrer Schülerinnen und Schüler ein? Ein Beitrag zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 63, 204. doi: 10.2378/peu2016.art18d
- Starbuck, W. H. (1963). Level of aspiration. *Psychological Review*, 70, 51–60.
- Statistisches Bundesamt. (2014). *Bildung und Kultur. Allgemeinbildende Schulen: Schuljahr 2013/14* (Fachserie 11, Bd. 1). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2015). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2014*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2017). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2015*. Wiesbaden: Statistisches Bundesamt.
- Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–329. doi: 10.1037//0003-066x.52.6.613
- Steele, C. M. & Aronson, J. (1995). Stereotype threat and intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811. doi: 10.1037//0022-3514.69.5.797
- Steinmayr, R. & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, 19, 80–90. doi: 10.1016/j.lindif.2008.05.004
- Stemmler, G. (2001). Grundlagen psychophysiologischer Methodik. In F. Rösler (Hrsg.), *Grundlagen und Methoden der Psychophysiologie* (Enzyklopädie der Psychologie, Bd. 4, S. 1–84). Göttingen: Hogrefe.
- Stern, J. A., Walrath, L. C. & Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology*, 21, 22–33. doi: 10.1111/j.1469-8986.1984.tb02312.x
- Steuer, G. (2014). *Fehlerklima in der Klasse. Zum Umgang mit Fehlern im Mathematikunterricht*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stocké, V. (2004). Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie*, 33, 303–320.
- Stojanov, K. (2011). *Bildungsgerechtigkeit. Rekonstruktionen eines umkämpften Begriffs*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stothart, C. R., Wright, T. J., Simons, D. J. & Boot, W. R. (2017). The costs (or benefits) associated with attended objects do little to influence inattention blindness. *Acta Psychologica*, 173, 101–105. doi: 10.1016/j.actpsy.2016.12.012
- Strasser, J. (2016). Pädagogische Professionalität im Zeichen kultureller Vielfalt. In V. Schurt, W. Waburg, V. Mehringer & J. Strasser (Hrsg.), *Heterogenität in Bildung und Sozialisation* (S. 27–52). Opladen: Barbara Budrich.

- Stubbe, T. C. & Bos, W. (2008). Schullaufbahneempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik*, 22, 49–63.
- Stubbe, T. C., Bos, W. & Euen, B. (2012). Der Übergang von der Primar- in die Sekundarstufe. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern im internationalen Vergleich* (S. 209–226). Münster: Waxmann.
- Stubbe, T. C., Buddenberg, I., Hornberg, S. & McElvany, N. (2007). Lesesozialisation im Elternhaus im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes et al. (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 299–327). Münster: Waxmann.
- Stubbe, T. C., Schwippert, K. & Wendt, H. (2016). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 299–316). Münster: Waxmann.
- Stubbe, T. C., Tarelli, I. & Wendt, H. (2012). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 231–246). 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. doi: 10.1037/a0027627
- Südkamp, A. & Möller, J. (2009). Referenzgruppeneffekte im simulierten Klassenraum. *Zeitschrift für Pädagogische Psychologie*, 23, 161–174. doi: 10.1024/1010-0652.23.34.161
- Südkamp, A., Möller, J. & Pohlmann, B. (2008). Der simulierte Klassenraum. *Zeitschrift für Pädagogische Psychologie*, 22, 261–276. doi: 10.1024/1010-0652.22.34.261
- Swann, W. B. & Snyder, M. (1980). On translating beliefs into action. Theories of ability and their application in an instructional setting. *Journal of Personality and Social Psychology*, 38, 879–888. doi: 10.1037/0022-3514.38.6.879
- Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, 25 (4), 79–97. doi: 10.1111/j.1540-4560.1969.tb00620.x
- 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. 247–267). Münster: Waxmann.
- Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83–114). London: Taylor & Francis.
- Taylor, S. E. & Fiske, S. T. (1978). Salience, attention, and attribution. Top of the head phenomena. *Advances in Experimental Social Psychology*, 11, 249–288. doi: 10.1016/S0065-2601(08)60009-X
- Teisl, J. T., Mazzocco, M. M. & Myers, G. F. (2001). The utility of kindergarten teacher ratings for predicting low academic achievement in first grade. *Race*, 34, 286–293. doi: 10.1177/002221940103400308
- Tenenbaum, H. R. & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology*, 99, 253–273. doi: 10.1037/0022-0663.99.2.253
- Tent, L. & Birkel, P. (2010). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 949–958). Weinheim: Beltz.
- Terhart, E. (2015). Heterogenität der Schüler – Professionalität der Lehrer. Ansprüche und Wirklichkeiten. In C. Fischer, C. Fischer-Ontrup, M. Veber & R. Buschmann (Hrsg.), *Umgang mit Vielfalt. Aufgaben und Herausforderungen für die Lehrerinnen- und Lehrerbildung* (S. 13–30). Münster: Waxmann.
- Thomas, W. I. & Znaniecki, F. (1918). *The Polish peasant in Europe and America* (vol. 1, 2nd ed. reprinted 1974). New York, NY: Octagon Books.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25–29. doi: 10.1037/h0071663
- Thorndike, R. L. (1968). Reviews. Rosenthal, Robert, and Jacobson, Lenore. Pygmalion in the classroom. *American Educational Research Journal*, 5, 708–711. doi: 10.3102/00028312005004708

- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529–554. doi: 10.1086/214483
- Thurstone, L. L. (1931). The measurement of social attitudes. *Journal of Abnormal and Social Psychology*, *26*, 249–269. doi: 10.1037/h0070363
- Timmermans, A. C., de Boer, H. & van der Werf, M. P. C. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education*, *19*, 217–240. doi: 10.1007/s11218-015-9326-6
- Timmermans, A. C., Kuyper, H. & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *The British Journal of Educational Psychology*, *85*, 459–478. doi: 10.1111/bjep.12087
- Tobisch, A. (2013). *Ethnische Stereotype und diagnostische Kompetenz von Lehramtsstudierenden*. Unveröffentlichte Masterarbeit, Universität Augsburg. Augsburg.
- Tobisch, A. & Dresel, M. (2016). Herkunftsbezogene Disparitäten bei Lehrkrafturteilen. In V. Schurt, W. Waburg, V. Mehringer & J. Strasser (Hrsg.), *Heterogenität in Bildung und Sozialisation* (S. 73–87). Opladen: Barbara Budrich.
- Tobisch, A. & Dresel, M. (2017). Negatively or positively biased? Dependencies of teachers' judgments and expectations based on students' ethnic and social backgrounds. *Social Psychology of Education*. Online publication. doi: 10.1007/s11218-017-9392-z
- Tobisch, A., Kopp, B., Martschinke, S., Kröner, S. & Dresel, M. (2016). Motivation und emotionales Erleben bei Grundschülerinnen und -schülern unterschiedlicher Herkunft. *Zeitschrift für Grundschulforschung*, *9*, 78–94.
- Tom, D. Y., Cooper, H. & McGraw, M. (1984). Influences of student background and teacher authoritarianism on teacher expectations. *Journal of Educational Psychology*, *76*, 259–265. doi: 10.1037//0022-0663.76.2.259
- Trouilloud, D., Sarrazin, P., Bressoux, P. & Bois, J. (2006). Relation between teachers' early expectations and students' later perceived competence in physical education classes. Autonomy-supportive climate as a moderator. *Journal of Educational Psychology*, *98*, 75–86. doi: 10.1037/0022-0663.98.1.75
- Tsai, Y.-F., Viirre, E., Strychacz, C., Chase, B. & Jung, T.-P. (2007). Task performance and eye activity. Predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine*, *78*, 176–185.
- Tucci, I. (2016). Lebenssituation von Migranten und deren Nachkommen. In Statistisches Bundesamt (Hrsg.), *Datenreport 2016. Ein Sozialbericht für die Bundesrepublik Deutschland* (S. 236–243). Bonn: Bundeszentrale für politische Bildung.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty. Heuristics and biases. *Science*, *185*, 1124–1131. doi: 10.1126/science.185.4157.1124
- Tversky, A. & Kahneman, D. (1973). Availability. A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232. doi: 10.1016/0010-0285(73)90033-9
- Tzelgov, J., Porat, Z. & Henik, A. (1997). Automaticity and consciousness. Is perceiving the word necessary for reading it? *The American Journal of Psychology*, *110*, 429–448. doi: 10.2307/1423567
- Ugwuegbu, D. C. E. (1979). Racial and evidential factors in juror attribution of legal responsibility. *Journal of Experimental Social Psychology*, *15*, 133–146. doi: 10.1016/0022-1031(79)90025-8
- Ülger, Z., Dette-Hagenmeyer, D. E. & Reichle, B. (2016). Reduktion von Vorurteilen und Modifikation von interethnischen Einstellungen. Eine Möglichkeit zur Förderung des Bildungserfolgs von Migrantinnen und Migranten? *Psychologie in Erziehung und Unterricht*, *63*, 169–179. doi: 10.2378/peu2016.art15d
- Unipark. (2016) EFS-Survey [Umfragesoftware] [Computer software]. Köln: Unipark.
- Universität Augsburg. (2016). *LeHet. Lehrerprofessionalität im Umgang mit Heterogenität*. Verfügbar unter <https://www.uni-augsburg.de/projekte/lehet/>
- Urhahne, D., Zhou, J., Stobbe, M., Chao, S.-H., Zhu, M. & Shi, J. (2010). Motivationale und affektive Merkmale unterschätzter Schüler. *Zeitschrift für Pädagogische Psychologie*, *24*, 275–288. doi: 10.1024/1010-0652/a000021
- Utech, U. (2011). *Rufname und soziale Herkunft. Studien zur schichtspezifischen Vornamensvergabe in Deutschland*. Hildesheim: Georg Olms.
- Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M. & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers. Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, *47*, 497–527. doi: 10.3102/0002831209353594

- Van Ophuysen, S. (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnpfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38, 154–161. doi: 10.1026/0049-8637.38.4.154
- Van Twuyver, M. & van Knippenberg, A. (1995). Social categorization as a function of priming. *European Journal of Social Psychology*, 25, 695–701. doi: 10.1002/ejsp.2420250608
- Vervaet, R., D'hondt, F., van Houtte, M. & Stevens, P. A. J. (2016). The ethnic prejudice of Flemish teachers. The role of ethnic school composition and of teachability. *Cultural Diversity & Ethnic Minority Psychology*, 22 (4), 552–562. doi: 10.1037/cdp0000085
- Von zur Gathen, M. & Liebert, J. (2016). Auswirkungen von Armut auf die Lebenswirklichkeit und Entwicklung von Kindern und Jugendlichen. In Der paritätische Gesamtverband (Hrsg.), *Zeit zu handeln. Bericht zur Armutsentwicklung in Deutschland 2016* (S. 35–41). Berlin: Der paritätische Gesamtverband.
- Waller, N. G., Kojetin, B. A., Bouchard, T. J., Lykken, D. T. & Tellegen, A. (1990). Genetic and environmental influences on religious interests, attitudes, and values. A study of twins reared apart and together. *Psychological Science*, 1, 138–142. doi: 10.1111/j.1467-9280.1990.tb00083.x
- Ward, C. & Masgoret, A.-M. (2008). Attitudes toward immigrants, immigration, and multiculturalism in New Zealand. A social psychological analysis. *International Migration Review*, 42, 227–248. doi: 10.1111/j.1747-7379.2007.00119.x
- Wasel, W. & Gollwitzer, P. M. (1997). Willentliche Kontrolle der "automatischen" Stereotypaktivierung. Die Rolle subliminaler vs. supraliminaler Stimulusdarbietung. *Sprache und Kognition*, 16, 198–210. Verfügbar unter http://kops.uni-konstanz.de/bitstream/handle/123456789/11206/97WasGoll_AutStereo.pdf?sequence=1&isAllowed=y
- Watzlawik, P., Bavelas, J. B. & Jackson, D. D. (2011). *Menschliche Kommunikation. Formen, Störungen, Paradoxien* (12. Aufl.). Bern: Huber.
- Weber, M. (2003). *Heterogenität im Schulalltag. Konstruktion ethnischer und geschlechtlicher Unterschiede* (Schule und Gesellschaft, Bd. 30). Opladen: Leske + Budrich.
- Webster, M. & Driskell, J. E. (1983). Beauty as status. *American Journal of Sociology*, 89, 140–165. doi: 10.1086/227836
- Webster, M. & Foschi, M. (1988). *Status generalization. New theory and research*. Stanford, CA: Stanford University Press.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York, NY: Springer.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 17–31). Weinheim: Beltz.
- Weinstein, R. S. (2002). *Reaching higher. The power of expectations in schooling*. Cambridge, MA: Harvard University Press.
- Weir, K. (2016). Inequality at school. *Monitor on Psychology*, 47 (10), 42–47.
- Weiss, R. (1965). *Zensur und Zeugnis. Beiträge zu einer Kritik der Zuverlässigkeit und Zweckmäßigkeit der Ziffernbenotung*. Linz: Quirin Haslinger.
- Wendt, H., Schwippert, K. & Stubbe, T. C. (2016). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 317–331). Münster: Waxmann.
- Wendt, H., Stubbe, T. C. & Schwippert, K. (2012). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern im internationalen Vergleich* (S. 175–190). Münster: Waxmann.
- Wigfield, A. & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68–81. doi: 10.1006/ceps.1999.1015
- Wiley, M. G. & Eskilson, A. (1978). Why did you learn in school today? Teachers' perception of causality. *Sociology of Education*, 51, 261–269. doi: 10.2307/2112364
- Willis, J. & Todorov, A. (2006). First impressions. Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x
- Wilson, J. P., Hugenberg, K. & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability. From size to threat. *Journal of Personality and Social Psychology*, 113, 59–80. doi: 10.1037/pspi0000092
- Wirtz, M. A. (2013). *Dorsch - Lexikon der Psychologie* (16. Aufl.). Bern: Verlag Hans Huber.

- Wirtz, M. & Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen: Hogrefe.
- Wittenbrink, B., Judd, C. M. & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, 72, 262–274. doi: 10.1037//0022-3514.72.2.262
- Wittenbrink, B., Judd, C. M. & Park, B. (2001). Spontaneous prejudice in context. Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81, 815–827. doi: 10.1037//0022-3514.81.5.815
- Wolters, C. A. (1999). The relation between high school students' motivational regulation and their use of learning strategies, effort, and classroom performance. *Learning and Individual Differences*, 11, 281–299. doi: 10.1016/S1041-6080(99)80004-1
- Wolters, C. A. (2003). Regulation of motivation. Evaluating an underemphasized aspect of self-regulated learning. *Educational Psychologist*, 38, 189–205. doi: 10.1207/S15326985EP3804_1
- Word, C. O., Zanna, M. P. & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10, 109–120. doi: 10.1016/0022-1031(74)90059-6
- Wyer, R. S. & Srull, T. K. (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Lawrence Erlbaum.
- Yeo, G. B. & Neal, A. (2004). A multilevel analysis of effort, practice, and performance. Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology*, 89, 231–247. doi: 10.1037/0021-9010.89.2.231
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9, 1–27. doi: 10.1037/h0025848
- Zander, L. & Hannover, B. (2013). Die Bedeutung der Identifikation mit der Herkunftskultur und mit der Aufnahmekultur Deutschland für die soziale Integration Jugendlicher mit Migrationshintergrund in ihrer Schulklasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 45, 142–160. doi: 10.1026/0049-8637/a000092
- Zander, L. & Wolter, I. (2016). Deutsch oder Mathe? Wie die Zusammensetzung der Schulklasse und die Bedeutsamkeit des eigenen Geschlechts geschlechtstypische Aspirationen von Jugendlichen bestimmen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48, 59–68. doi: 10.1026/0049-8637/a000144
- Zanna, M. P., Sheras, P. L., Cooper, J. & Shaw, C. (1975). Pygmalion and galatea. The interactive effect of teacher and student expectancies. *Journal of Experimental Social Psychology*, 11, 279–287. doi: 10.1016/s0022-1031(75)80028-x
- Zebrowitz-McArthur, L. (1981). What grabs you? The role of attention in impression formation and causal attribution. In E. T. Higgins, C. P. Herman & M. P. Zanna (Eds.), *Social cognition. The Ontario symposium* (vol. 1, pp. 201–246). Hillsdale, NJ: Lawrence Erlbaum.
- Zhu, M. & Urhahne, D. (2015). Teachers' judgements of students' foreign-language achievement. *European Journal of Psychology of Education*, 30, 21–39. doi: 10.1007/s10212-014-0225-6
- Zick, A. (1997). *Vorurteile und Rassismus – Eine sozialpsychologische Analyse*. Münster: Waxmann.
- Zick, A., Pettigrew, T. F. & Wagner, U. (2008). Ethnic prejudice and discrimination in Europe. *Journal of Social Issues*, 64 (2), 233–251. doi: 10.1111/j.1540-4560.2008.00559.x
- Ziegler, A., Dresel, M., Stöger, H. & Schober, B. (2008). *Motivationsstestbatterie für Schülerinnen und Schüler der Jahrgangsstufen 5-10 (MTB 5-10)*. Ulm.
- Zimmerman, B. J. (2000). Self-efficacy. An essential motive to learn. *Contemporary Educational Psychology*, 25, 82–91. doi: 10.1006/ceps.1999.1016
- Zimmermann, U. (2013). *Definition: Augenbewegung*, Eye Tracking Kompetenzzentrum. Verfügbar unter <https://eyetracking.ch/glossar-augenbewegung/>
- Zimmermann, U. (2014). *Definition: Fixationsdauer*, Eye Tracking Kompetenzzentrum. Verfügbar unter <https://eyetracking.ch/glossar-fixationsdauer/>
- Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics*, 36, 277–284. doi: 10.3758/bf03206369

Appendix

Appendix A Noten und Textbausteine der Fallvignetten

Im Folgenden werden die eingesetzten Textbausteine der Halbjahreszeugnisse der vierten Jahrgangsstufe inkl. der entsprechenden Noten dargestellt. Das erste Zeugnis wurde in Studie 1 eingesetzt. Für Studie 2 wurden zusätzlich das zweite und dritte Zeugnis konzipiert. Eine Beschreibung der Konstruktion der Fallvignetten findet sich in Abschnitt 7.1.1. Die folgenden Textfragmente wurden mit syntaktischen Verknüpfungen verbunden, um sie in der Form einem realen Zeugnis anzugleichen. Zudem wurde an mehreren Stellen im Zeugnis der experimentell variierte Vorname der Schüler (vgl. Abschnitt 7.1.2) eingesetzt.

Tabelle A1.

Noten der Ziffernbewertungen in den Halbjahreszeugnissen der Fallvignetten

	Ziffernbewertungen in den Fallvignetten		
	Zeugnis 1	Zeugnis 2	Zeugnis 3
Deutsch	2	3	2
Mathematik	3	2	3
Heimat- und Sachunterricht	2	2	2
Englisch	2	3	3
Religion	2	1	2
Musik	1	2	3
Kunst	3	3	1
Werken und textiles Gestalten	3	2	2
Sport	1	2	1

Tabelle A2.

Textbausteine und Notenäquivalent des ersten Zeugnisses (Studie 1 und Studie 2)

Textbaustein	Note
Ordnung & regelmäßiger Schulbesuch	
Gibt sich Mühe, mit seinen Schulbüchern und anderen Unterrichtsmaterialien gewissenhaft umzugehen	3
Nie fehlt er unentschuldigt und erscheint zum Unterricht jederzeit pünktlich	1
Sozialverhalten	
Fand zögernd Anschluss in der Klasse	3
Sucht den Kontakt zu seinen Klassenkameraden	2
Über vereinbarte Regeln setzt er sich oft hinweg	4
Hilft Schwächeren stets unaufgefordert und eigenverantwortlich	1
Zur Konfliktlösung benötigt er meist die Hilfe anderer	3
Zeigt sich nach Konflikten versöhnungsbereit	2
Bemüht sich, längere Zeit dem Gespräch zu folgen, meldet sich jedoch selten	3
Bei von ihm bevorzugten Themen beteiligt er sich aktiv am Gespräch	2
Arbeitsverhalten	
Lässt sich durch Störungen leicht von seiner Arbeit ablenken	3
Bringt alle angefangenen Aufgaben vollständig zu Ende	1
Erledigt die Hausaufgaben zuverlässig	2
Achtet meist auf eine ansprechende und lesbare Schrift	2
Ist für neue Lerninhalte zu begeistern	2
Durch Probleme lässt er sich leicht entmutigen	3
Zeigt durch seine Beiträge, dass er die behandelten Themen vollständig durchdacht hat	1
Arbeitet sicher und selbständig, seine Arbeiten sollte er jedoch gründlicher und selbstkritischer kontrollieren	2
Verhalten gegenüber der Lehrkraft	
Begegnet seinen Lehrern respektvoll und aufgeschlossen	2
Ist dankbar, wenn er von seinen Lehrern Unterstützung erhält und seine Leistungen gewürdigt werden	2

Tabelle A3.

Textbausteine und Notenäquivalent des zweiten Zeugnisses (Studie 2)

Textbaustein	Note
Ordnung & regelmäßiger Schulbesuch	
Sein Bemühen um einen ordentlichen Umgang mit seinen Schulbüchern und Unterrichtsmaterialien ist erkennbar	3
Pünktliches Erscheinen zum Unterricht und ein regelmäßiger Schulbesuch sind für ihn selbstverständlich	1
Sozialverhalten	
Arbeitet nur mit bestimmten Kindern oder bei bestimmten Aufträgen zusammen	3
Schloss Kontakte zu seinen Mitschülern	2
Hält sich auch nach Aufforderung nicht an vereinbarte Regeln	4
Unterstützt andere Mitschüler in ihrem Lernen durch aktive Hilfe	1
Nimmt Lösungsvorschläge in Konflikten nur zögerlich an	3
Verändert gegebenenfalls sein Verhalten	2
Zeigt Hemmungen, vor der Klasse zu sprechen	3
Hört genau zu und geht auf Beiträge von Mitschülern ein	2
Arbeitsverhalten	
Beweist wenig Durchhaltevermögen	3
Zeigt sich äußerst lernwillig und anstrengungsbereit	1
Seine Hausaufgaben erledigt er ordentlich und zuverlässig	2
Arbeitet übersichtlich und sauber	2
Zeigt Interesse an neuen Lerninhalten	2
Benötigt Unterstützung, um sich auch bei Problemen mit dem Thema auseinanderzusetzen	3
Überzeugt durch enormes Vorwissen	1
Mit mehr Mut zu Wortmeldungen und durch eine gründlichere Selbstkontrolle kann er leicht Erfolge erzielen	2
Verhalten gegenüber der Lehrkraft	
In Gesprächen mit seinen Lehrern verhält er sich meistens aufgeschlossen und umgänglich	2
Sucht selbstständig und regelmäßig Kontakt zu seinen Lehrern	2

Tabelle A4.

Textbausteine und Notenäquivalent des dritten Zeugnisses (Studie 2)

Textbaustein	Note
Ordnung & regelmäßiger Schulbesuch	
Geht mit Materialien teilweise gewissenhaft um	3
Den selten versäumten Stoff holt er selbständig nach	1
Sozialverhalten	
Hatte anfangs Schwierigkeiten, sich in die Klasse zu integrieren	3
Knüpft Kontakte zu seinen Mitschülern	2
Den Klassenregeln folgt er nach Ermahnung	4
Verhält sich immer hilfsbereit und teamfähig	1
Konflikte löst er meistens selbständig	3
Ist kritikfähig und nimmt Lösungen gerne an	2
Dem Unterrichtsverlauf schenkt er teilweise Aufmerksamkeit	3
Bei für ihn interessanten Themen zeigt er viel Kommunikationsbereitschaft	2
Arbeitsverhalten	
Überzeugt durch große Selbständigkeit und Ausdauer	1
Schriftliche Arbeiten erledigt er meist zuverlässig	2
Er arbeitet meist sorgfältig	2
Arbeitet aufmerksam und interessiert mit	2
Zeigt ein angemessenes Arbeitstempo	3
Zeigt ein überwiegend gutes Auffassungsvermögen	2
Kann auf Gelerntes zurückgreifen und dies anwenden	1
Ist mit eigenen Beiträgen noch sehr zurückhaltend	3
Verhalten gegenüber der Lehrkraft	
Den Kontakt zur Lehrkraft sucht er meistens von sich aus	2
Zeigt sich in Gesprächen selbstbewusst und aufgeschlossen	2

Appendix B Ziel- und Attributionskategorien im impliziten Assoziationstest

Tabelle B1.

Ziel- und Attributionskategorien im impliziten Assoziationstest

Zielkategorien		Attributionskategorien	
deutsch	türkisch	positiv	negativ
Andreas	Ahmed	ehrgeizig	aggressiv
Christian	Burak	fleißig	ignorant
Felix	Hasan	intelligent	lustlos
Florian	Mehmet	konzentriert	respektlos
Matthias	Muhammed	kreativ	schlampig
Michael	Murat	motiviert	störend
Niklas	Mustafa	neugierig	unzuverlässig
Thomas	Yusuf	teamfähig	vorlaut

Tabelle B1 sind alle Ziel- und Attributionskategorien zu entnehmen, die im impliziten Assoziationstest (IAT) der zweiten Studie der hier vorgestellten Arbeit eingesetzt wurden. Dabei wurden ausschließlich männliche Vornamen ausgewählt, die eindeutig der deutschen oder türkischen Herkunft zugeordnet werden können. Zudem wurden eindeutig positive und negative Adjektive gewählt, die für den schulischen Kontext von Bedeutung sind.

Appendix C Dokumentation der Fragebogenskalen der Studien 1 & 2

Tabelle C1.

Skalenbeschreibung und -kennwerte der Skala Leistungsaspiration

Quelle:	Adaptiert nach Dresel et al., 2006; Finsterwald, 2006	
Beschreibung	Die Skala erfasst die Aspiration der Lehrkraft/des Lehramtsstudierenden an die zukünftigen Leistungen in den Hauptfächern Deutsch, Mathematik und Heimat- und Sachunterricht.	
Itemanzahl:	3	
Skalierung:	1 (<i>sehr gut</i>), 2 (<i>gut</i>), 3 (<i>befriedigend</i>), 4 (<i>ausreichend</i>), 5 (<i>mangelhaft</i>), 6 (<i>ungenügend</i>)	
Skalenkennwerte:	Studie 1: $M = 2.00$ ($SD = 0.66$) Cronbachs $\alpha = .79$	Studie 2: $M = 3.29$ ($SD = 0.49$) Cronbachs $\alpha = .89$
Itemstamm:	„Mit welcher Note wären Sie bei diesem Schüler in der folgenden Probe gerade noch zufrieden?“	
Anmerkungen:	Im Itemstamm der Studie 1 und Studie 2 wurde je nach Herkunftskombination der Fallvignette der jeweilige Vorname des Schülers anstelle „Schüler“ im Itemstamm genannt. Für die Analysen in Studie 1 und in Studie 2 wurden die Items der Skala Leistungsaspiration rekodiert. So entsprechen hohe Werte im empirischen Teil der vorliegenden Arbeit guten Leistungen. In der Skalendokumentation ist die Original-Polung berichtet.	

Tabelle C2

Itemtexte und -kennwerte der Skala Leistungsaspiration

Itemformulierung	Studie 1		Studie 2	
	M (SD)	r_{it}	M (SD)	r_{it}
Deutsch	2.06 (0.74)	.64	3.35 (0.55)	.85
Mathematik	1.92 (0.87)	.55	3.10 (0.52)	.82
Heimat- und Sachunterricht	2.02 (0.76)	.73	3.44 (0.56)	.72

Anmerkungen. $N_{Studie1} = 237$, $N_{Studie2} = 45$.

Tabelle C3

Skalenbeschreibung und -kennwerte der Skala Leistungserwartung

Quelle:	Eigene Entwicklung	
Beschreibung	Die Skala erfasst die Erwartungen der Lehrkraft/des Lehramtsstudierenden an die zukünftigen Leistungen in den Hauptfächern Deutsch, Mathematik und Heimat- und Sachunterricht.	
Itemanzahl:	3	
Skalierung:	1 (<i>sehr gut</i>), 2 (<i>gut</i>), 3 (<i>befriedigend</i>), 4 (<i>ausreichend</i>), 5 (<i>mangelhaft</i>), 6 (<i>ungenügend</i>)	
Skalenkennwerte:	Studie 1: $M = 2.02$ ($SD = 0.56$) Cronbachs $\alpha = .77$	Studie 2: $M = 2.22$ ($SD = 0.20$) Cronbachs $\alpha = .50$
Itemstamm:	„Welche Note würden Sie in der nächsten Probe von diesem Schüler erwarten?“	
Anmerkungen:	Im Itemstamm der Studie 1 und Studie 2 wurde je nach Herkunftskombination der Fallvignette der jeweilige Vorname des Schülers anstelle „Schüler“ im Itemstamm genannt. Für die Analysen in Studie 1 und in Studie 2 wurden die Items der Skala Leistungserwartung rekodiert. So entsprechen hohe Werte im empirischen Teil der vorliegenden Arbeit guten Leistungen. In der Skalendokumentation ist die Original-Polung berichtet.	

Tabelle C4

Itemtexte und -kennwerte der Skala Leistungserwartung

Itemformulierung	Studie 1		Studie 2	
	M (SD)	r_{it}	M (SD)	r_{it}
Deutsch	1.95 (0.63)	.61	2.28 (0.29)	.36
Mathematik	2.05 (0.71)	.53	2.39 (0.27)	.21
Heimat- und Sachunterricht	2.06 (0.66)	.69	2.00 (0.28)	.39

Anmerkungen. $N_{\text{Studie1}} = 237$, $N_{\text{Studie2}} = 45$.

Tabelle C5

Skalenbeschreibung und -kennwerte der Skala Anstrengungsbereitschaft

Quelle:	Adaptiert nach Ramm et al. (2006)	
Beschreibung	Die Skala erfasst die Einschätzung der Lehrkraft/des Lehramtsstudierenden der Bereitschaft des Schülers sich in zukünftigen schulischen Angelegenheiten zu bemühen bzw. anzustrengen.	
Itemanzahl:	5	
Skalierung:	1 (<i>stimmt gar nicht</i>), 2 (<i>stimmt eher nicht</i>), 3 (<i>teils teils</i>), 4 (<i>stimmt eher</i>), 5 (<i>stimmt voll und ganz</i>)	
Skalenkennwerte:	Studie 1: $M = 3.86$ ($SD = 0.66$) Cronbachs $\alpha = .87$	Studie 2: $M = (SD 3.71 = 0.35)$ Cronbachs $\alpha = .78$
Itemstamm:	„Was glauben Sie, wie sich der Schüler in Zukunft verhalten wird? Bitte geben Sie an, wie sehr Sie den folgenden Aussagen zustimmen.“	
Anmerkungen:	Im Itemstamm der Studie 1 und Studie 2 wurde je nach Herkunftskombination der Fallvignette der jeweilige Vorname des Schülers anstelle „Schüler“ im Itemstamm genannt.	

Tabelle C6

Itemtexte und -kennwerte der Skala Anstrengungsbereitschaft

Itemformulierung	Studie 1		Studie 2	
	M (SD)	r_{it}	M (SD)	r_{it}
Er wird sich viel Mühe geben, alles gut zu verstehen.	3.76 (0.61)	.50	4.00 (0.40)	.41
Er wird versuchen, alles so gut wie möglich zu machen.	4.09 (0.87)	.59	3.95 (0.46)	.37
Er wird nicht aufgeben, auch wenn die Aufgaben in der Schule schwierig und umfangreich sind.	3.73 (0.70)	1.00	3.93 (0.45)	.59
Wenn er eine Probearbeit schreibt, wird er sich sehr anstrengen.	4.00 (0.92)	.53	3.49 (0.58)	.60
Wenn er eine Probearbeit schreibt, wird er das letzte aus sich herausholen.	3.72 (0.88)	.61	3.19 (0.58)	.50

Anmerkungen. $N_{Studie1} = 237$, $N_{Studie2} = 45$.

Tabelle C7

Skalenbeschreibung und -kennwerte der Skala Fähigkeiten

Quelle:	Adaptiert nach Dickhäuser et al., 2002	
Beschreibung	Die Skala erfasst die Einschätzung der Lehrkraft/des Lehramtsstudierenden zu den allgemeinen schulischen Fähigkeiten im Sinne einer Begabung des Schülers.	
Itemanzahl:	4	
Skalierung:	Bipolare Skala von 1 (<i>niedrig/schwer</i>) bis 5 (<i>hoch/leicht</i>)	
Skalenkennwerte:	Studie 1: $M = 3.71$ ($SD = 0.47$) Cronbachs $\alpha = .88$	Studie 2: $M = 3.58$ ($SD = 0.36$) Cronbachs $\alpha = .83$
Itemstamm:	„Wie schätzen Sie den Schüler ein?“	
Anmerkungen:	Im Itemstamm der Studie 1 und Studie 2 wurde je nach Herkunftskombination der Fallvignette der jeweilige Vorname des Schülers anstelle „Schüler“ im Itemstamm genannt.	

Tabelle C8

Itemtexte und -kennwerte der Skala Fähigkeiten

Itemformulierung	Studie 1		Studie 2	
	M (SD)	r_{it}	M (SD)	r_{it}
Ich halte seine Begabung für ... (niedrig/hoch)	3.70 (0.54)	.77	3.80 (0.44)	.71
Neues zu lernen fällt ihm ... (schwer/leicht)	3.70 (0.60)	.69	3.41 (0.49)	.57
Seine schulischen Fähigkeiten sind ... (niedrig/hoch)	3.76 (0.47)	.76	3.74 (0.39)	.64
Die Bearbeitung schulischer Aufgaben fällt ihm ... (schwer/leicht)	3.69 (0.56)	.78	3.38 (0.52)	.72

Anmerkungen. $N_{\text{Studie1}} = 237$, $N_{\text{Studie2}} = 45$.

Tabelle C9

Skalenbeschreibung und -kennwerte der Skala Fähigkeitsselbstkonzept

Quelle:	Adaptiert nach Dickhäuser et al., 2002	
Beschreibung	Die Skala erfasst die Einschätzung der Lehrkraft/des Lehramtsstudierenden zum Fähigkeitsselbstkonzept der Schüler.	
Itemanzahl:	5	
Skalierung:	Bipolare Skala von 1 (<i>nicht begabt/schwer/nicht intelligent/wenig</i>) bis 5 (<i>sehr begabt/leicht/intelligent/viel</i>)	
Skalenkennwerte:	Studie 1:	Studie 2:
	-	$M = 3.39$ ($SD = 0.48$)
	-	Cronbachs $\alpha = .91$
Itemstamm:	„Was denken Sie, wie sich der Schüler selbst einschätzen würde?“	
Anmerkungen:	Im Itemstamm der Studie 2 wurde je nach Herkunftskombination der Fallvignette der jeweilige Vorname des Schülers anstelle „Schüler“ im Itemstamm genannt. Die Skala wurde nur in Studie 2 eingesetzt.	

Tabelle C10

Itemtexte und -kennwerte der Skala Fähigkeitsselbstkonzept

Itemformulierung	Studie 1		Studie 2	
	M (SD)	r_{it}	M (SD)	r_{it}
Ich bin für die Schule ... (nicht begabt/begabt)	-	-	3.40 (.53)	.83
Neues zu lernen fällt mir ... (schwer/leicht)	-	-	3.26 (0.53)	.75
Ich bin ... (nicht intelligent/intelligent)	-	-	3.56 (0.62)	.78
Ich kann in der Schule ... (wenig/viel)	-	-	3.56 (0.53)	.84
In der Schule fallen mir viele Aufgaben (schwer/leicht)	-	-	3.16 (0.60)	.69

Anmerkungen. $N_{Studie2} = 45$.

Tabelle C11

Skalenbeschreibung und -kennwerte der Skala subjektiver Wert

Quelle:	Adaptiert nach Ziegler et al., 2008	
Beschreibung	Die Skala erfasst die Einschätzung der Lehrkraft/des Lehramtsstudierenden zum subjektiven Wert (im Sinne der Wichtigkeit, des intrinsischen Werts und der Nützlichkeit) der Schule für den Schüler.	
Itemanzahl:	6	
Skalierung:	1 (<i>stimmt gar nicht</i>), 2 (<i>stimmt eher nicht</i>), 3 (<i>teils teils</i>), 4 (<i>stimmt eher</i>), 5 (<i>stimmt voll und ganz</i>)	
Skalenkennwerte:	Studie 1: - -	Studie 2: $M = (SD = 0.)$ Cronbachs $\alpha = .84$
Itemstamm:	„Wie schätzen Sie den Schüler ein?“	
Anmerkungen:	Im Itemstamm der Studie 1 und Studie 2 wurde je nach Herkunftskombination der Fallvignette der jeweilige Vorname des Schülers anstelle „Schüler“ im Itemstamm genannt. Die Skala wurde nur in Studie 2 eingesetzt.	

Tabelle C12

Itemtexte und -kennwerte der Skala subjektiver Wert

Itemformulierung	Studie 1		Studie 2	
	$M (SD)$	r_{it}	$M (SD)$	r_{it}
Die Schule macht ihm Spaß.	-	-	3.67 (0.40)	.55
In der Schule gut zu sein, ist ihm wichtig.	-	-	3.58 (0.56)	.57
Er glaubt, dass er in der Schule etwas lernt, dass er gut gebrauchen kann.	-	-	3.37 (0.55)	.74
Der Unterricht interessiert ihn.	-	-	3.64 (0.40)	.45
Ihm ist der Unterricht sehr wichtig.	-	-	3.30 (0.43)	.61
Er denkt, dass er den Unterrichtsstoff gut gebrauchen kann.	-	-	3.27 (0.49)	.79

Anmerkungen. $N_{Studie2} = 45$.

Tabelle C13

Itembeschreibung und Kennwerte – Eignung für das Gymnasium

Quelle:	Eigene Entwicklung	
Beschreibung	Das Item erfasst die subjektive Einschätzung der Versuchspersonen zur Eignung des Schülers, das Gymnasium als weiterführende Schulform nach der Grundschule zu besuchen.	
Itemanzahl:	1	
Skalierung:	1 (<i>überhaupt nicht geeignet</i>), 2 (<i>eher nicht geeignet</i>), 3 (<i>teils teils</i>), 4 (<i>eher geeignet</i>), 5 (<i>voll und ganz geeignet</i>)	
Skalenkennwerte:	Studie 1: $M = 3.59$ ($SD = 0.71$) -	Studie 2: $M = 3.63$ ($SD = 0.64$) -
Item:	„Ist der Schüler Ihrer Meinung nach für den Gymnasialbesuch geeignet?“	
Anmerkungen:	Im Itemstamm der Studie 1 und Studie 2 wurde je nach Herkunftskombination der Fallvignette der jeweilige Vorname des Schülers anstelle „Schüler“ im Itemstamm genannt.	

Anmerkungen. $N_{\text{Studie1}} = 237$, $N_{\text{Studie2}} = 45$.

Tabelle C14

Skalenbeschreibung und -kennwerte der Skala Einstellung zur Heterogenität – ethnische Herkunft

Quelle:	Grassinger et al. (2016).	
Beschreibung	Die Skala erfasst die Valenz der Einstellung gegenüber einer heterogenen Schülerschaft in Bezug auf den Migrationshintergrund der Schüler(innen). Die eingesetzte Skala zur Valenz ist eine Subskala aus der Gesamtskala von Grassinger et al. (2016).	
Itemanzahl:	3	
Skalierung:	Bipolare Skala von 1 (<i>schlecht/unangenehm/negativ</i>) bis 6 (<i>gut/angenehm/positiv</i>)	
Skalenkennwerte:	Studie 1: - -	Studie 2: $M = 4.46$ ($SD = 0.88$) Cronbachs $\alpha = .58$
Instruktion:	„Stellen Sie sich vor, Sie sind bereits Lehrkraft und unterrichten...“	
Anmerkungen:	Die Skala wurde nur in Studie 2 eingesetzt.	

Tabelle C15

Itemtexte und -kennwerte der Skala Einstellung zur Heterogenität – ethnische Herkunft

Itemformulierung	Studie 1		Studie 2	
	M (SD)	r_{it}	M (SD)	r_{it}
In meinem Unterricht halte ich die Unterschiedlichkeit der Schüler(innen) in Bezug auf ihren Migrationshintergrund für...“ (schlecht/gut)	-	-	4.56 (1.37)	.39
In meinem Unterricht halte ich die Unterschiedlichkeit der Schüler(innen) in Bezug auf ihren Migrationshintergrund für...“ (unangenehm/angenehm)	-	-	4.22 (1.06)	.35
In meinem Unterricht halte ich die Unterschiedlichkeit der Schüler(innen) in Bezug auf ihren Migrationshintergrund für...“ (negativ/positiv)	-	-	4.60 (1.14)	.43

Anmerkungen. $N_{\text{Studie2}} = 45$.

Tabelle C16

Skalenbeschreibung und -kennwerte der Skala Einstellung zur Heterogenität – soziale Herkunft

Quelle:	Grassinger et al. (2016).	
Beschreibung	Die Skala erfasst die Valenz der Einstellung gegenüber einer heterogenen Schülerschaft in Bezug auf den sozioökonomischen Status der Schüler(innen). Die eingesetzte Skala zur Valenz ist eine Subskala aus der Gesamtskala von Grassinger et al. (2016).	
Itemanzahl:	3	
Skalierung:	Bipolare Skala von 1 (<i>schlecht/unangenehm/negativ</i>) bis 6 (<i>gut/angenehm/positiv</i>)	
Skalenkennwerte:	Studie 1: - -	Studie 2: $M = 3.85$ ($SD = 0.70$) Cronbachs $\alpha = .84$
Instruktion:	„Stellen Sie sich vor, Sie sind bereits Lehrkraft und unterrichten...“	
Anmerkungen:	Die Skala wurde nur in Studie 2 eingesetzt.	

Tabelle C17

Itemtexte und -kennwerte der Skala Einstellung zur Heterogenität – soziale Herkunft

Itemformulierung	Studie 1		Studie 2	
	M (SD)	r_{it}	M (SD)	r_{it}
In meinem Unterricht halte ich die Unterschiedlichkeit der Schüler(innen) in Bezug auf ihren sozioökonomischen Status für...“ (schlecht/gut)	-	-	3.89 (0.80)	.74
In meinem Unterricht halte ich die Unterschiedlichkeit der Schüler(innen) in Bezug auf ihren sozioökonomischen Status für...“ (unangenehm/angenehm)	-	-	3.67 (0.77)	.65
In meinem Unterricht halte ich die Unterschiedlichkeit der Schüler(innen) in Bezug auf ihren sozioökonomischen Status für...“ (negativ/positiv)	-	-	4.00 (0.83)	.74

Anmerkungen. $N_{Studie2} = 45$.

Tabelle C18

Ergänzende Fragen zur Eye-Tracking-Erhebung

Items	Antwortoptionen
Im Folgenden bitten wir Sie noch einige Fragen zu beantworten, die wir zur Analyse der Eye-Tracking-Daten benötigen.	
Wie müde fühlen Sie sich momentan?	1 (<i>sehr müde</i>) - 5 (<i>sehr wach</i>)
Bitte beantworten Sie folgende Fragen zu Ihrer Sehkraft und Ihrem allgemeinen Gesundheitszustand:	
Sind Sie kurzsichtig?	1 (<i>ja</i>), 2 (<i>nein</i>);
Falls ja, wie viel Dioptrien haben Sie?	offenes Antwortfeld
Sind Sie weitsichtig?	1 (<i>ja</i>), 2 (<i>nein</i>);
Falls ja, wie viel Dioptrien haben Sie?	offenes Antwortfeld
Tragen Sie im Moment weiche Kontaktlinsen?	1 (<i>ja</i>), 2 (<i>nein</i>)
Tragen Sie im Moment harte Kontaktlinsen?	1 (<i>ja</i>), 2 (<i>nein</i>)
Tragen Sie im Moment eine Brille?	1 (<i>ja</i>), 2 (<i>nein</i>)
Leiden Sie unter einer der folgenden Krankheiten: Adipositas, Diabetes melitus, Migräne, Anorexia nervosa, Autismus-Spektrum-Störung	1 (<i>ja</i>), 2 (<i>nein</i>)
Nehmen Sie zur Zeit Medikamente mit Wirkung auf das autonome Nervensystem, z. B.: Antidepressiva, Neuroleptika, Muskelrelaxanzen, opiathaltige Medikamente	1 (<i>ja</i>), 2 (<i>nein</i>)
Sind Ihnen andere Gründe bekannt, die Einfluss auf Ihre Pupillenreaktion haben könnten?	1 (<i>ja</i>), 2 (<i>nein</i>);
Falls ja, welche?	offenes Antwortfeld

Appendix D Ergänzende Tabelle: Studie 1

Tabelle D1

Ergebnisse der univariaten Varianzanalysen zur Interaktion der Schülerherkunft in den Fallvignetten und der Berufserfahrung der Lehrkräfte, sowie Haupteffekte der Berufserfahrung

Urteile und Erwartungen	Haupteffekt der Berufserfahrung			Interaktionseffekt: Herkunft; Berufserfahrung		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Leistungen						
Leistungserwartung	1.208	.22	.23	0.902	.65	.21
Leistungsaspiration	1.848	.01	.31	1.002	.48	.23
Leistungsrelevante Schülermerkmale						
Gymnasialeignung	1.615	.02	.29	1.124	.30	.25
Anstrengungsbereitschaft	1.598	.03	.28	0.973	.53	.23
Fähigkeiten	1.943	.01	.32	1.412	.06	.30

Anmerkungen. $df_{\text{Haupteffekt}} = 37$, $df_{\text{Interaktion}} = 45$; $df_{\text{Fehler}} = 150$; $N = 235$.
Zweiseitige Signifikanzwerte.

Appendix E Ergänzende Tabellen: Studie 2

Tabelle E1

Ergebnisse Varianzanalysen mit Messwiederholung für Leistungen und leistungsrelevante Schülermerkmale der Vorstudie

	<i>F</i>	<i>df</i>	<i>df_{Fehler}</i>	<i>p</i>	η^2
Leistungen					
Leistungserwartung	1.243	1.847	64.643	.29	.03
Leistungsaspiration	0.107	1.833	64.147	.88	.00
Leistungsrelevante Merkmale					
Gymnasialeignung	1.927	1.674	58.576	.16	.05
Anstrengungsbereitschaft	1.540	1.959	68.575	.22	.04
Fähigkeiten	0.136	1.879	65.755	.86	.00

Anmerkungen. *df* nach Greenhouse-Geisser Korrektur. *N* = 36.

Zweiseitige Signifikanzwerte.

Tabelle E2

Ergebnisse der a priori Helmert-Kontrastanalysen zu Leistungen und leistungsrelevanten Merkmalen der Vorstudie

	Kontrast 1			Kontrast 2		
	Hoher vs. niedriger sozioökonomischer Status			Kein Migrationshintergrund vs. Migrationshintergrund		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Leistungen						
Leistungserwartung	0.498	.50	.01	2.479	.12	.07
Leistungsaspiration	0.156	.70	.00	0.029	.87	.00
Leistungsrelevante Merkmale						
Gymnasialeignung	0.361	.55	.01	5.385	.03	.13
Anstrengungsbereitschaft	0.085	.77	.00	3.212	.08	.08
Fähigkeiten	0.008	.93	.00	0.34	.57	.01

Anmerkungen. $df = 1$, $df_{\text{Fehler}} = 35$, $N = 36$.

Zweiseitige Signifikanzwerte.

Tabelle E3

Ergebnisse der a priori Helmert-Kontrastanalysen zu Leistungen und leistungsrelevanten Merkmalen

	Kontrast 1			Kontrast 2		
	Hoher vs. niedriger sozioökonomischer Status			Kein Migrationshintergrund vs. Migrationshintergrund		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Leistungen						
Leistungserwartung	7.563	.01	.15	0.505	.48	.01
Leistungsaspiration	0.005	.95	.00	0.017	.90	.00
Leistungsrelevante Merkmale						
Gymnasialeignung	0.612	.44	.01	1.288	.26	.03
Anstrengungsbereitschaft	0.378	.54	.01	1.846	.18	.04
Fähigkeiten	0.426	.52	.01	2.719	.11	.06
Fähigkeitsselbstkonzept	0.860	.36	.02	0.142	.71	.00
Subjektiver Wert	1.049	.31	.02	3.957	.05	.08

Anmerkungen. $df = 1$, $df_{Fehler} = 44$, $N = 45$.

Zweiseitige Signifikanzwerte.

Tabelle E4

Mauchly-Test auf Sphärizität für Fragebogen- und Eye-Tracking-Daten

	Mauchly's W	p	ε
<i>Fragebogendaten</i>			
Leistungserwartung	.965	.458	.966
Leistungsaspiration	.990	.804	.990
Gymnasialeignung	.996	.918	.996
Anstrengungsbereitschaft	.941	.264	.944
Fähigkeiten	.975	.575	.976
Fähigkeitsselbstkonzept	.988	.760	.988
<i>Eye-Tracking-Daten</i>			
Lidschlagrate	.979	.630	.980
Pupillendiameter nach AOI			
Ordnung	.989	.782	.989
Regelmäßiger Schulbesuch	.961	.415	.962
Positives Sozialverhalten	.906	.115	.914
Negatives Sozialverhalten	.915	.143	.922
Positives Arbeitsverhalten	.768	.003	.812
Negatives Arbeitsverhalten	.824	.014	.851
Verhalten zur Lehrkraft	.599	.000	.714
Note Deutsch	.989	.785	.989
Note Mathematik	.944	.279	.947
Note HSU	.907	.116	.915
Note Englisch	.878	.057	.891
Note Religion	.994	.882	.994
Note Musik	.972	.537	.973
Note Kunst	.888	.074	.900
Note WTG	.924	.174	.929
Note Sport	.964	.442	.965
Notendurchschnitt gesamt	.861	.037	.878
Notendurchschnitt Hauptfächer	.892	.082	.903

Anmerkungen. $df = 2$, $N = 45$.

Tabelle E4 (Fortsetzung)

Mauchly-Test auf Sphärizität für Fragebogen- und Eye-Tracking-Daten

	Mauchly's W	p	ϵ
Fixationen nach AOI			
Ordnungsverhalten	.985	.721	.985
Regelmäßiger Schulbesuch	1.000	.994	1.000
Positives Sozialverhalten	.929	.198	.934
Negatives Sozialverhalten	.929	.197	.934
Positives Arbeitsverhalten	.976	.584	.976
Negatives Arbeitsverhalten	.988	.764	.988
Verhalten zur Lehrkraft	.946	.294	.949
Note Deutsch	.855	.032	.873
Note Mathematik	.939	.252	.943
Note Heimat- und Sachunterricht	.908	.121	.916
Note Englisch	.998	.954	.998
Note Religion	.905	.112	.914
Note Musik	.990	.800	.990
Note Kunst	.998	.960	.998
Note Werken und Textiles Gestalten	.953	.346	.955
Note Sport	.740	.001	.794
Notendurchschnitt gesamt	.953	.348	.955
Notendurchschnitt Hauptfächer	.743	.001	.796

Anmerkungen. df = 2, N = 45.

Tabelle E5

Ergebnisse der a priori Helmert-Kontrastanalysen zur relativen Fixationsdauer der positiven und negativen AOIs in den Fallvignetten

Fächer	Kontrast 1			Kontrast 2		
	Hoher vs. niedriger sozioökonomischer Status			Kein Migrationshintergrund vs. Migrationshintergrund		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Positive AOIs						
Regelmäßiger Schulbesuch	0.001	.98	.00	0.061	.81	.00
Sozialverhalten	1.312	.26	.03	0.139	.71	.00
Arbeitsverhalten	0.009	.93	.00	5.002	.03	.10
Verhalten zur Lehrkraft	0.330	.57	.01	0.000	.98	.00
Negative AOIs						
Ordnungsverhalten	4.473	.04	.09	0.043	.84	.00
Sozialverhalten	0.029	.87	.00	0.004	.95	.00
Arbeitsverhalten	0.218	.64	.01	0.555	.46	.01

Anmerkungen. $df = 1$, $df_{Fehler} = 45$, $N = 45$. Zweiseitige Signifikanzwerte.

Tabelle E6

Ergebnisse der a priori Helmert-Kontrastanalysen zur relativen Fixationsdauer der positiven und negativen AOIs in den Fallvignetten

Fächer	Kontrast 1			Kontrast 2		
	Hoher vs. niedriger sozioökonomischer Status			Kein Migrationshintergrund vs. Migrationshintergrund		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Positive AOIs						
Regelmäßiger Schulbesuch	0.001	.98	.00	0.061	.81	.00
Sozialverhalten	1.312	.26	.03	0.139	.71	.00
Arbeitsverhalten	0.009	.93	.00	5.002	.03	.10
Verhalten zur Lehrkraft	0.330	.57	.01	0.000	.98	.00
Negative AOIs						
Ordnungsverhalten	4.473	.04	.09	0.043	.84	.00
Sozialverhalten	0.029	.87	.00	0.004	.95	.00
Arbeitsverhalten	0.218	.64	.01	0.555	.46	.01

Anmerkungen. $df = 1$, $df_{Fehler} = 45$, $N = 45$. Zweiseitige Signifikanzwerte.

Tabelle E7

Ergebnisse der Varianzanalysen mit Messwiederholung mit den impliziten Einstellungen (kontinuierliche Variable) als Kovariate

Urteile und Erwartungen	Haupteffekt der Schülerherkunft			Interaktionseffekt: implizite Einstellungen		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Leistungen						
Leistungserwartung	2.977	.06	.07	0.757	.47	.02
Leistungsaspiration	0.025	.97	.00	0.034	.97	.00
Leistungsrelevante Schülermerkmale						
Gymnasialeignung	0.652	.52	.02	1.024	.36	.02
Anstrengungsbereitschaft	1.795	.17	.04	1.402	.25	.03
Fähigkeiten	0.530	.59	.01	0.522	.59	.01
Fähigkeitsselbstkonzept	0.241	.79	.01	0.849	.43	.02
Subjektiver Wert	1.560	.22	.04	1.016	.37	.02

Anmerkungen. $df = 2$, $df_{\text{Fehler}} = 86$ (nach Greenhouse-Geisser-Korrektur ergaben sich keine Unterschiede zu den hier berichteten Ergebnissen); $N = 45$.
Zweiseitige Signifikanzwerte.