

Zur Rolle von Migrationshintergrund und sozialer Herkunft für die Urteilsbildung von Lehrkräften

Ergebnisse aus der experimentellen Begleitforschung zum LeHet-Projekt

Anita Tobisch & Markus Dresel

1. Lehrkrafturteile im Kontext einer heterogenen Schülerschaft

Lehrkräfte sind von essenzieller Bedeutung für den Lernprozess und -erfolg von Schüler*innen. Um einen möglichst adaptiven, auf die Lernenden ausgerichteten Unterricht durchführen zu können, müssen Lehrkräfte zahlreiche lern- und leistungsrelevante Schülermerkmale (z. B. Fähigkeiten oder Motivation) berücksichtigen und sich darüber möglichst akkurate Urteile bilden. Allerdings kann der Prozess der Urteilsbildung von Lehrkräften, wie alle menschlichen Urteilsbildungsprozesse, auch potenziellen Verzerrungen unterliegen.

Das *Modell der Akkuratheit von Lehrkrafturteilen* von Südkamp, Kaiser und Möller (2012) liefert einen umfassenden Rahmen, um Einflussfaktoren auf die Übereinstimmung zwischen Lehrkrafturteil und Schülermerkmalen zu identifizieren. In diesem Modell werden unterschiedliche Merkmale berücksichtigt, die sich auf die Urteilsgenauigkeit von Lehrkräften auswirken können. Dabei berücksichtigen Südkamp et al. (2012) vier zentrale Bereiche, die je nach Ausprägung die Urteilsakkuratheit erhöhen bzw. verringern können. Neben Test- (z. B. Aufgabenschwierigkeit) und Urteilsmerkmalen (z. B. globale vs. spezifische Urteile) wird in diesem Modell auch auf die Bedeutung von Lehrkraft- und Schülermerkmalen verwiesen. Die soziale Herkunft sowie der Migrationshintergrund von Schüler*innen sind, durch die hierbei nach wie vor bestehenden und gut dokumentierten Leistungsdisparitäten und Ungleichheiten im Bildungserfolg (z. B. OECD, 2019), häufig im Fokus der Bildungsforschung. Zahlreiche Studien verweisen darauf, dass der geringere Bildungserfolg dieser Schülergruppen nicht allein durch z. B. geringere Fähigkeiten oder familiäre Unterstützung erklärt werden kann. Auch unter Kontrolle der tatsächlichen Leistungen erhalten diese Schüler*innen z. T. schlechtere Bewertungen oder sind mit niedrigeren Lehrkrafterwartungen konfrontiert (z. B. Bonefeld, Dickhäuser, Janke, Praetorius & Dresel, 2017; Sprietsma, 2013; Tenenbaum & Ruck, 2007). Bereits 1965 konnte Weiss in einer experimentellen Untersuchung mit Hilfe von Fallvignetten zeigen, dass Informationen über den familiären Hintergrund von Schüler*innen (z. B. Berufe der Eltern) einen Einfluss auf die

Bewertung von Mathematik- und Deutscharbeiten haben. Auch in aktuellen experimentellen Untersuchungen konnten diese Urteilsverzerrungen in Abhängigkeit der Herkunft von Schüler*innen nachgewiesen werden (z. B. Bonefeld & Dickhäuser, 2018; Wenz & Hoenig, 2020).

Angenommen wird, dass sich diese herkunftsabhängigen Lehrkrafturteile durch eine mehr oder weniger stereotypengeleitete Informationsverarbeitung ergeben. Um diesen Urteilsbildungsprozess theoretisch zu erklären, bietet das *Kontinuum-Modell der Eindrucksbildung* (Fiske & Neuberg, 1990) einen differenzierten Einblick in Informationsverarbeitungsprozesse, die sich auch auf den schulischen Kontext übertragen lassen. Angenommen wird, dass Urteile mehr oder weniger automatisiert (stereotypengeleitet) oder kontrolliert (individuumsbasiert) gebildet werden. Um zu einer kontrollierteren Informationsverarbeitung zu gelangen, sind u. a. mehr verfügbare kognitive Kapazitäten notwendig. Im Eindrucksbildungsprozess werden demnach Personen zunächst als Vertreter*innen bestimmter sozialer Kategorien (z. B. Geschlecht, Ethnie) wahrgenommen und entsprechende stereotype Eigenschaften automatisch verknüpft. Nach diesem Modell wird die Nutzung individuumsbasierter Informationen wahrscheinlicher, wenn die Person z. B. dem gängigen Stereotyp (mehr oder weniger offensichtlich) nicht entspricht oder man selbst diesem Urteil eine hohe Bedeutung zuschreibt. Eine eher automatisierte Informationsverarbeitung könnte somit auch eine Erklärung für stereotypengeleitete Lehrkrafturteile sein. Die im Folgenden vorgestellten Studien wurden im Kompetenzbereich „Individuelle Beratung und Förderung“ im Projekt *Förderung der Lehrprofessionalität im Umgang mit Heterogenität (LeHet)* durchgeführt und gehen dieser Annahme nach.

2. Zwei Studien aus der Begleitforschung des Kompetenzbereichs „Individuelle Beratung und Förderung“ im LeHet-Projekt

Zentrales Ziel der Untersuchungen war es, verzerrte Lehrkrafturteile näher zu beleuchten und die dadurch gewonnenen Erkenntnisse für die Lehrkraftaus- und -weiterbildung nutzbar zu machen. Zur Analyse herkunftsabhängiger Urteilsverzerrungen wurden zwei experimentelle Untersuchungen durchgeführt, die zum Ziel hatten, die Urteilsbildungsprozesse von (angehenden) Lehrkräften zu untersuchen und dabei den Migrationshintergrund sowie den sozialen Status von Schüler*innen zu berücksichtigen.¹

1 Da sich die größten Disparitäten im Bildungserfolg v. a. für männliche Schüler mit türkischem Migrationshintergrund zeigen, beziehen sich beide hier vorgestellten Studien ausschließlich auf diese Schülergruppe (z. B. Beauftragte der Bundesregierung für Migration,

2.1 Studie 1: Herkunftsassoziierte Lehrkrafturteile

Die erste Studie ging der Frage einer Über- oder Unterschätzung spezifischer Schülergruppen durch Lehrkräfte nach (Tobisch, 2019; Tobisch & Dresel, 2017). Zwar konnte bereits gezeigt werden, dass Urteilsunterschiede im Zusammenhang mit der Herkunft von Schüler*innen auftreten (z. B. Bonefeld et al., 2017), meist blieb in bisherigen Forschungsarbeiten allerdings offen, ob sich herkunftsbedingte Urteilsunterschiede durch eine Über- oder Unterschätzung von Schüler*innen mit spezifischen Herkunftsmerkmalen erklären lassen. Lediglich in Untersuchungen mit Vorschulkindern zeigte sich, dass sowohl die sprachlichen Fähigkeiten von Kindern mit einem hohen sozialen Status überschätzt als auch jene von Kindern mit einem niedrigen sozialen Status unterschätzt werden (Ready & Chu, 2015; Ready & Wright, 2011). Zur Urteilsakkuratheit in Bezug auf Schüler*innen mit und ohne Migrationshintergrund liegen widersprüchliche Befunde vor. So konnte teilweise eine geringere Akkuratheit bei der Leistungsbeurteilung von Kindern mit Migrationshintergrund festgestellt werden (z. B. Pit-ten Cate, Krolak-Schwerdt & Glock, 2016). Kaiser, Südkamp und Möller (2017) berichteten hingegen von einer höheren Urteilsakkuratheit bei Schüler*innen mit Migrationshintergrund. Die hier vorgestellte Studie leistet einen Beitrag zur Aufklärung dieser widersprüchlichen Befundlage.

Methode

In einer experimentellen Online-Studie im *between-subject*-Design wurden $N = 237$ Grundschullehrkräfte (48.9% weiblich) mit einer durchschnittlichen Berufserfahrung von 15.3 Jahren ($SD = 8.4$) um die Beurteilung von Fallvignetten in Form von Grundschulzeugnissen (Halbjahr 4. Jahrgangsstufe) männlicher Schüler gebeten. Dabei enthielten die Zeugnisse sowohl Ziffernnoten für alle Fächer als auch eine Verbalbeurteilung zum Verhalten der Schüler. Die Vignetten zeichneten insgesamt das Bild eines eher guten Schülers. Die Lehrkräfte wurden zufällig drei Versuchsgruppen zugeordnet und sollten nach dem Lesen eines Zeugnisses u. a. ihre Leistungserwartungen für zukünftige Leistungskontrollen in den Hauptfächern in einem Fragebogen angeben. Die Leistungen und Schülerbeschreibungen in den Fallvignetten waren in allen drei Gruppen völlig identisch, lediglich der Vorname (*Julius: kein Migrationshintergrund und hoher sozialer Status*, *Justin: kein Migrationshintergrund und niedriger sozialer Status* und *Murat: Migrationshintergrund und niedriger sozialer Status*) wurde variiert, um bei den teilnehmenden Grundschullehrkräften den Migrationshintergrund und die soziale Herkunft der Schüler*innen zu aktivieren (vgl. hierzu auch

Flüchtlinge und Integration, 2014). Entsprechend wird in der sprachlichen Darstellung lediglich das männliche Geschlecht verwendet.

Bertrand & Mullainathan, 2004; Sprietsma, 2013; Tobisch, 2019; Utech, 2011; Wenz & Hoenig, 2020).²

Zentrale Ergebnisse und Diskussion

Varianzanalysen erbrachten einen signifikanten Haupteffekt der Schülerherkunft auf die Leistungserwartung der Lehrkräfte. Kontrastanalysen deckten dabei auf, dass sowohl die soziale Herkunft als auch der Migrationshintergrund relevante Einflussfaktoren darstellen. So wurden an den Schüler mit einem hohen sozialen Status und ohne Migrationshintergrund (*Julius*) die höchsten Leistungserwartungen gerichtet. Trotz vergleichbarem niedrigen sozialen Status – basierend auf den Einschätzungen der Lehrkräfte im Manipulations-Check – wurde der fiktive Schüler in der Fallvignette mit türkischem Vornamen (*Murat*) schlechter beurteilt als der fiktive Schüler in der Fallvignette mit deutschem Vornamen (*Justin*).

Um zu untersuchen, ob eine herkunftsabhängige Über- oder Unterschätzung der Leistungen bestimmter Schülergruppen vorliegt, wurden die Lehrkräfteeinschätzungen mit dem tatsächlichen Notendurchschnitt der jeweiligen Vignette verglichen. Die hierzu berechneten *t*-Tests zeigten, dass sich die Lehrkrafturteile für den Schüler mit Migrationshintergrund und niedrigem sozialen Status nicht signifikant von den angegebenen tatsächlichen Leistungen unterscheiden. Bei den Schülern ohne Migrationshintergrund, insbesondere bei dem Schüler mit hohem sozialen Status, zeigte sich hingegen eine bedeutsame Abweichung der Leistungseinschätzung vom tatsächlichen Notendurchschnitt: Von diesen Schülern wurden bessere Leistungen erwartet als sie laut Zeugnis bisher gezeigt hatten.

Die Orientierung der Lehrkräfte an den Zeugnisnoten für Vorhersagen zukünftiger Leistungen der Schüler mit Migrationshintergrund und einem niedrigen sozialen Status kann als akkurates Urteil angesehen werden. Die Erwartung einer bedeutsamen Leistungssteigerung bei Schülern ohne Migrationshintergrund und mit hohem sozialen Status deutet hingegen auf eine Überschätzung hin. Neben anderen Aspekten ist insbesondere der geringe Frauenanteil in der Stichprobe als Limitation zu diskutieren. Dennoch sind die Ergebnisse der Studie ein klarer Hinweis darauf, dass v. a. bei Schülern ohne Migrationshintergrund und mit hohem sozialen Status eine positive Urteilsverzerrung zugunsten höherer Leistungserwartungen stattfinden könnte. Daran anknüpfend

2 In beiden hier vorgestellten Studien wurde ein Manipulationscheck durchgeführt, um abzusichern, dass die Versuchspersonen den Migrationshintergrund und den sozialen Status der Schüler entsprechend der geplanten Manipulation wahrnahmen. Die Analysen zeigten dabei, dass die Versuchspersonen, die Herkunft der fiktiven Schüler entsprechend der geplanten Herkunftsmanipulation wahrnahmen.

verfolgte die zweite Studie das Ziel, die zugrundeliegenden kognitiven Prozesse der Urteilsbildung näher zu beleuchten.

2.2 Studie 2: Informationsverarbeitungsprozesse von angehenden Lehrkräften

In Anlehnung an das Kontinuum-Modell der Eindrucksbildung (Fiske & Neuberg, 1990) wurde aufbauend auf den Erkenntnissen der ersten Studie angenommen, dass sich akkuratere Lehrkrafturteile bei Kindern mit Migrationshintergrund und einem niedrigen sozialen Status durch kontrolliertere Informationsverarbeitungsprozesse erklären lassen. Positiv verzerrte Urteile bei Schülern ohne Migrationshintergrund und mit hohem sozialen Status könnten hingegen Ergebnis eher automatisierter Prozesse sein (Tobisch, 2019; Tobisch & Dresel, submitted).

Methode

In der experimentellen Laborstudie im Messwiederholungsdesign wurden $N = 45$ Lehramtsstudierenden (91.1 % weiblich; Alter: $M = 21.5$ Jahre; $SD = 3.1$) drei Fallvignetten von Grundschülern vorgelegt. Der Aufbau der Fallvignetten sowie die Manipulation der Vornamen entsprachen dem Vorgehen der zuvor beschriebenen Studie. Zur Umsetzung des Messwiederholungsdesigns wurden zwei weitere Fallvignetten erstellt, deren Leistungen und Verbalbeurteilungen äquivalent zu denen des Zeugnisses der vorher dargestellten Studie waren. Jede Versuchsperson beurteilte somit drei Zeugnisse und entsprechend auch alle drei Herkunftskombinationen. Um die zugrundeliegenden Prozesse der Urteilsbildung möglichst verhaltensnah untersuchen zu können, wurde mit Hilfe eines Eye-Trackers u. a. der Pupillendurchmesser beim Lesen der Zeugnisse erfasst. Die Pupillenreaktion lässt Rückschlüsse auf den Erregungszustand des zentralen Nervensystems zu (z. B. Bonowski, 2002), welcher eng an psychische Prozesse, wie z. B. die kognitive Aktivität, gekoppelt ist (z. B. Hess & Polt, 1964). Dabei ist anzunehmen, dass eine kontrolliertere Informationsverarbeitung mit einer erhöhten kognitiven Beanspruchung einhergeht. Um Herkunftseffekte auf die kognitiven Prozesse identifizieren zu können, wurden die Pupillendurchmesser der Lehramtsstudierenden in den unterschiedlichen Herkunftskombinationen verglichen. Die Vignetten wurden hierzu in inhaltliche Abschnitte (Areas of Interest) eingeteilt, in denen der individuelle Pupillendurchmesser anschließend gemittelt wurde.

Zentrale Ergebnisse und Diskussion

Varianzanalysen mit Messwiederholung erbrachten insbesondere für die ersten Textabschnitte sowie für einige Noten bedeutsame Unterschiede im Pupillen-

durchmesser in Abhängigkeit von der Herkunft der Schüler. Durch Kontrastanalysen zeigte sich, dass sich diese Unterschiede im Wesentlichen durch den Migrationshintergrund erklären lassen. Beim Lesen der Vignetten des fiktiven Schülers dessen Name mit einem Migrationshintergrund und niedrigem sozialen Status assoziiert wurde, war der Pupillendurchmesser der Teilnehmer*innen durchschnittlich größer als beim Lesen der anderen beiden Vignetten von Schülern ohne Migrationshintergrund. Da eine Erweiterung der Pupille eine erhöhte Aktivität des zentralen Nervensystems und damit einen erhöhten Erregungszustand (Arousal) anzeigt, lässt sich anhand dieser Befunde ableiten, dass die Versuchspersonen beim Lesen der Fallvignette mit einem stereotyp türkischen Vornamen im Vergleich zu den Vignetten, mit typisch deutschen Vornamen, ein höheres Arousal aufweisen. Da ein erhöhtes Arousal als Indikator für kognitive Beanspruchung/Aktivierung angesehen werden kann (z. B. Bonowski, 2002) und anzunehmen ist, dass automatisierte Informationsverarbeitungsprozesse weniger kognitive Kapazitäten fordern als eine kontrolliertere Verarbeitung, würde dieser Befund für einen eher kontrollierten bzw. kognitiv beanspruchenderen Informationsverarbeitungsprozess bei der Urteilsbildung über Schüler mit Migrationshintergrund und mit niedrigem sozialen Status sprechen.

Bei der Interpretation der Befunde ist anzumerken, dass die Erfassung des Pupillendurchmessers lediglich Rückschlüsse auf die Erregung des Nervensystems zulässt. So wären zur Erklärung von Veränderungen im Pupillendurchmesser neben kognitiver Aktivität auch emotionale Reaktionen denkbar, da diese sich ebenfalls in diesen physiologischen Reaktionen niederschlagen. In zukünftigen Studien sollte verstärkt zwischen kognitiven und affektiven Prozessen differenziert werden. Zudem ist einschränkend zu erwähnen, dass Lehramtsstudierende an der Untersuchung teilnahmen. So geht das Modell der adaptiven diagnostischen Kompetenz (Böhmer, Gräsel, Krolak-Schwerdt, Hörstermann & Glock, 2017) davon aus, dass zwar erfahrene Lehrkräfte eher zwischen einer mehr oder weniger automatisierten bzw. kontrollierten Informationsverarbeitung wechseln können, Lehramtsstudierende (die noch Novizen in der Schüler*innenbeurteilung sind) hingegen Informationen eher integrativ verarbeiten.

3. Zusammenfassung und Diskussion der vorgestellten Ergebnisse

Die hier vorgestellten Studien fokussierten herkunftsassoziierte Lehrkrafturteile und untersuchten, inwiefern der Migrationshintergrund und die soziale Herkunft von Schüler*innen die Urteilsbildung von (angehenden) Lehrkräften beeinflussen. In beiden Studien konnten Hinweise für akkuratere Urteile und

eine eher kontrolliertere Informationsverarbeitung bei der Beurteilung von Schülern mit Migrationshintergrund und niedrigem sozialen Status gefunden werden. Gleichzeitig deuten die Ergebnisse auf eine Überschätzung von Schülern ohne Migrationshintergrund und mit einem hohen sozialen Status hin und verweisen hier auf eine eher automatisierte Informationsverarbeitung.

Die Befunde der beiden Studien legen nahe, dass nicht nur – wie häufig angenommen – negative Einstellungen und Stereotype gegenüber Schüler*innen mit Migrationshintergrund und/oder niedrigem sozialen Status zu Urteilsverzerrungen bei Lehrkräften führen können. Auch positive Assoziationen gegenüber Schüler*innen ohne Migrationshintergrund (v. a. mit einem hohen sozialen Status) scheinen eine bedeutsame Rolle für die Entstehung herkunftsabhängiger Disparitäten zu spielen. In einer weiteren Fragebogenstudie (die ebenfalls im LeHet-Kontext entstand) wurden konkrete stereotype Assoziationen untersucht (Tobisch & Dresel, 2019, 2020). Hier bestätigte sich, dass angehende Lehrkräfte v. a. mit Schüler*innen ohne Migrationshintergrund und hohem sozialen Status besonders positive Eigenschaften verknüpften. Die Merkmalszuschreibungen zu Schüler*innen mit Migrationshintergrund und niedrigem sozialen Status fielen zwar im Vergleich negativer aus; insgesamt waren sie jedoch tendenziell neutral bis leicht positiv einzustufen. Auch diese Befunde bestärken die Ergebnisse der beiden in diesem Beitrag vorgestellten Studien.

Einschränkend ist zu erwähnen, dass Vornamen häufig weitere Assoziationen, z. B. zur Attraktivität oder Intelligenz, auslösen können (z. B. Rudolph, Böhm & Lummer, 2007). So sind für zukünftige Studien weitere Namensnormen zu berücksichtigen, die auch Namen unterschiedlicher Kulturkreise umfassen. Einen wichtigen Beitrag zur Ergänzung solcher Namensnormen konnten Kleen und Glock (2020) leisten. Zudem darf nicht vernachlässigt werden, dass nicht nur die soziale Herkunft oder der Migrationshintergrund einen wesentlichen Einfluss auf Lehrkrafturteile haben können. Auch weitere Stereotype (z. B. gegenüber Schüler*innen mit sonderpädagogischem Förderbedarf; Pit-ten Cate & Krischler, 2020) können zu verzerrten Urteilen führen. Aber auch die Bedeutung der Kombination spezifischer Merkmale (z. B. Migrationshintergrund, Religionszugehörigkeit und Leistungsentwicklung; Klapproth, Kärchner & Glock, 2018) für Lehrkrafturteile sollte nicht unterschätzt und in weiteren Untersuchungen berücksichtigt werden.

Durch das experimentelle Design der vorliegenden Studien ist zwar eine hohe interne, allerdings aber auch eine geringe ökologische Validität zu konstatieren. Zwar ist davon auszugehen, dass die analysierten grundlegenden psychischen Mechanismen auch im realen Klassenkontext bedeutsam sind, dennoch lassen sich diese Befunde nicht uneingeschränkt auf den Unterricht übertragen. Nichtsdestotrotz ist es möglich, aus diesen Forschungsarbeiten wichtige Implikationen abzuleiten, die auch in der Lehrkraftaus- und -weiterbildung einen bedeutsamen Stellenwert einnehmen sollten. So wird z. B. deutlich, dass es auch zu positiven Urteilsverzerrungen gegenüber bestimmten Schülergrup-

pen kommen kann, was in der Konsequenz auch zu einer herkunftsabhängigen Diskrepanz in der Beurteilung von Schüler*innen führen kann. Im Folgenden wird exemplarisch an einer Übung dargestellt, wie diese Erkenntnisse für die Aus- und Weiterbildung von Lehrkräften nutzbar gemacht werden.

4. Nutzung der gewonnenen Befunde in der Aus- und Weiterbildung von Lehrkräften

Das zentrale Ziel des LeHet-Projekts stellt die Weiterentwicklung der Aus- und Weiterbildung von Lehrkräften zum Umgang mit Heterogenität in der Schule dar. Entsprechend ist es auch in der Begleitforschung von hoher Relevanz, Methoden und Erkenntnisse zu generieren, die direkt für Maßnahmen in allen Phasen der Lehrkraftausbildung nutzbar gemacht werden können. So wurden auch die Befunde und methodischen Ansätze der zuvor beschriebenen Untersuchungen bereits in Seminare und Fortbildungen integriert.

Zentrale Bestandteile sowohl der Seminare als auch der Fortbildungen stellen zunächst bewährte theoretische Modelle dar, die auch Grundlage der zuvor beschriebenen Studien sind. Hierzu zählen v. a. das *Kontinuum-Modell der Eindrucksbildung* (Fiske & Neuberg, 1990) und das Modell der *Akkuratheit von Lehrkrafturteilen* (Südkamp, Kaiser & Möller, 2012). Diese Grundlagen werden im Verlauf der Lerneinheiten durchgehend durch empirische Befunde (u. a. auch durch Studien aus der LeHet-Begleitforschung) ergänzt und diskutiert. Das Vorgehen ist an die jeweiligen Zielgruppen und ihre Erfahrungen angepasst. Durch Diskussionsanteile werden dabei auch persönliche Erfahrungen aus dem schulischen Kontext aufgegriffen und mit den Theorien und Studienergebnissen in Verbindung gebracht. Ein sensibler Themenkomplex wie dieser (in dem auch die soziale Erwünschtheit eine besondere Rolle spielt) verlangt ein hohes Maß an Fingerspitzengefühl, um (angehende) Lehrkräfte auch mit ihren eigenen Stereotypen und Einstellungen zu konfrontieren. Dies ist essenziell, um bei den Teilnehmer*innen der Veranstaltungen eine persönliche Bedeutsamkeit zu diesem Thema zu generieren, da nur dann auch von einer vertieften und nachhaltigen Informationsverarbeitung ausgegangen werden kann und bedeutsame Lernprozesse angestoßen werden können. Das Material der ersten hier vorgestellten Studie wurde zudem in unterschiedlichen Lehrveranstaltungen mit einem inhaltlichen Schwerpunkt zur Heterogenität der Schülerschaft, aber auch in anderen Veranstaltungen, z. B. zur diagnostischen Kompetenz, genutzt.

Ein Kernelement in diesen Veranstaltungen ist u. a. eine umfangreiche Übung zur diagnostischen Kompetenz und zu herkunftsassoziierten Urteilen. Diese Übung wird regelmäßig im Rahmen universitärer Lehrveranstaltungen und in Lehrkraftfortbildungen durchgeführt. Die Teilnehmer*innen werden um das Lesen einer Fallvignette und eine anschließende Beurteilung des (männli-

chen) Schülers gebeten. Entsprechend des Vorgehens der ersten Studie erhalten zwar alle Studierende identische Vignetten, allerdings mit unterschiedlichen Schülernamen, die den Migrationshintergrund und die soziale Herkunft der Schüler aktivieren sollen. Nachdem die Teilnehmer*innen sich in Einzelarbeit einen Eindruck von dem Schüler bilden konnten und sie ihre Einschätzungen im Fragebogen selbst ausgewertet haben, folgt eine Erklärung der theoretischen Konstrukte, die erfasst wurden. Durch die selbständige Auswertung des eigenen Fragebogens erhalten die Teilnehmer*innen zum einen Einblicke in das Forschungsvorgehen, zum anderen kann hierdurch die soziale Erwünschtheit der Antworten reduziert werden, da nur sie selbst ihre Antworten sehen. Im Plenum werden anschließend die Einschätzungen des Schülerverhaltens, die Eignung des Schülers für eine weiterführende Schulart und Leistungserwartungen diskutiert. Wichtig ist hierbei, zunächst keine Namen aus den Vignetten zu nennen, sondern nur von „dem Schüler“ zu sprechen. In der Diskussion treten häufig implizite Persönlichkeitstheorien und unterschiedliche Erwartungen an den Schüler zu Tage, die häufig auch zu den entsprechenden sozialen Stereotypen passen. Im letzten Teil der Übung wird aufgelöst, dass es sich um identische Vignetten handelt und sich nur die Namen der Schüler unterscheiden. Insbesondere dieses Vorgehen führt meist zu einem großen Überraschungsmoment bei den Teilnehmer*innen. Abschließend werden die Befunde aus der ersten Studie präsentiert, kritisch diskutiert, mit den eigenen Einschätzungen und Stereotypen in Verbindung gebracht sowie Implikationen für die Praxis abgeleitet.

Je nach inhaltlichem Schwerpunkt kann das Material bzw. die Übung im Seminarkontext angepasst werden. Beispielsweise könnten andere Namen eingesetzt werden, um einen anderen Migrationshintergrund zu aktivieren. Auch kann die Bearbeitung unter Zeitdruck erfolgen, um eine automatisierte Informationsverarbeitung zu begünstigen. Diese Übung liefert zudem nicht nur inhaltliche Anknüpfungspunkte zum Thema herkunftsabhängiger Urteilsverzerrungen, sondern u. a. auch zur diagnostischen Kompetenz, der schnellen Eindrucksbildung oder Stereotypen. Informelle Gespräche sowohl mit Studierenden als auch mit erfahrenen Lehrkräften ergaben bislang ein positives Feedback zu dieser Übung, auch wenn selbstverständlich keine direkte Übertragbarkeit auf die schulische Praxis möglich ist. Eine begleitende Evaluation zur Wirksamkeit des eingesetzten Materials (u. a. in Bezug auf die Sensibilisierung und Einstellungsänderung der Studierenden) sollte für zukünftige Veranstaltungen implementiert werden. So können sowohl das Material als auch das didaktische Vorgehen weiter optimiert werden.

5. Zusammenfassung und Perspektiven für die Lehrkraftaus- und -weiterbildung

Der vorliegende Beitrag liefert einen Überblick über zwei experimentelle Studien der LeHet-Begleitforschung, die sich mit Lehrkrafturteilen im Kontext der sozialen Herkunft sowie des Migrationshintergrunds von Schüler*innen auseinandersetzen. Diese hatten zum Ziel, mögliche herkunftsassoziierte Urteilsverzerrungen von (angehenden) Lehrkräften zu untersuchen und hieraus Implikationen für die Lehramtsaus- und -weiterbildung abzuleiten.

Im Kontext herkunftsassoziiierter Disparitäten im Bildungsverlauf und -erfolg von Schüler*innen verweisen die Befunde darauf, dass potenzielle Ursachen nicht nur auf Ebene der Schüler*innen, ihrer Familien oder des Schulsystems zu suchen sind, sondern, dass auch Lehrkrafturteile einen Beitrag zur Erklärung von Disparitäten liefern. So zeigen die Ergebnisse der beiden Studien, dass sich Disparitäten womöglich mehr durch positiv verzerrte Urteile gegenüber Schüler*innen ohne Migrationshintergrund und mit hohem sozialen Status erklären lassen könnten als durch negativ verzerrte Urteile gegenüber Schüler*innen mit Migrationshintergrund und niedrigem sozialen Status. Die negativen Konsequenzen besonders positiver Stereotype und daraus resultierender positiv verzerrter Urteile sind bislang oftmals noch nicht stark genug in der Lehrkräftebildung verankert, sind jedoch von Bedeutung, um Disparitäten abzumildern. Durch die Begleitforschung im LeHet-Projekt sowie die darauf aufbauenden Lehrveranstaltungen und Fortbildungen konnte hierzu ein wichtiger Beitrag geleistet werden.

Literatur

- 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.
- 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
- 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 (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 131–147). Cham, Schweiz: Springer.
- Bonefeld, M. & Dickhäuser, O. (2018). (Biased) grading of students' performance: Students' names, performance level, and implicit attitudes. *Frontiers in Psychology*, 9, 1–13. doi: 10.3389/fpsyg.2018.00481
- 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*. Dissertationschrift, Marburg: Philipps-Universität Marburg.
- 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.
- 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
- Kaiser, J., Südkamp, A. & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, 109, 871–888. doi: 10.1037/edu0000156
- Klapproth, F., Kärchner, H. & Glock, S. (2018). Do students' religion and school absences moderate the effect of ethnic stereotypes on school-placement recommendations? *The Journal of Experimental Education*, 86, 173–194.
- Kleen, H. & Glock, S. (2020). Sag' mir wie du heißt, dann sage ich dir, wie du bist: Eine Untersuchung von Vornamen. In S. Glock & H. Kleen (Hrsg.), *Stereotype in der Schule* (S. 99–131). Wiesbaden: Springer.
- OECD (2019). *PISA 2018 Results: Where all students can succeed* (Vol. II). Paris, France: OECD Publishing.
- Pit-ten Cate, I. & Krischler, M. (2020). Stereotype hinsichtlich Schüler*innen mit sonderpädagogischem Förderbedarf: Lehrkraftüberzeugungen, -erwartungen und -gefühle. In S. Glock & H. Kleen (Hrsg.), *Stereotype in der Schule* (S. 191–224). Wiesbaden: Springer.
- 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
- 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
- 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
- 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
- Spietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, 45, 523–538. doi: 10.1007/s00181-012-0609-x
- 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
- 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.

- Tobisch, A. (2019). *Herkunftsassoziierte Lehrkrafturteile und -erwartungen: Soziale Kognitionen und Urteilsbildungsprozesse im Kontext einer ethnisch und sozial heterogenen Schülerschaft*. (Inaugural Dissertation). Universität Augsburg. Verfügbar unter: urn:nbn:de:bvb:384-opus4-638356
- Tobisch, A. & Dresel, M. (submitted). *Teacher students' automatic and controlled information processing in context of students' ethnic and social background: An eye-tracking study*.
- Tobisch, A. & Dresel, M. (2020). Fleißig oder faul? Welche Einstellungen haben angehende Lehrkräfte gegenüber Schüler*innen aus unterschiedlichen sozialen Schichten? In S. Glock & H. Kleen (Hrsg.), *Stereotype in der Schule* (S. 133–158). Wiesbaden: Springer.
- Tobisch, A. & Dresel, M. (2019). Herkunftsassoziierte Urteilsverzerrungen bei Lehrkräften. Posterpräsentation im Rahmen des Interaktiven Forums „Psychologische Perspektiven in der Qualitätsoffensive Lehrerbildung“ auf der gemeinsamen Tagung der Fachgruppen Entwicklungspsychologie und Pädagogischen Psychologie (PAEPSY) in Leipzig, 2019.
- 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*, 20, 731–752. doi 10.1007/s11218-017-9392-z
- Utech, U. (2011). *Rufname und soziale Herkunft: Studien zur schichtspezifischen Vornamensvergabe in Deutschland*. Hildesheim: Georg Olms.
- Weiss, R. (1965). *Zensur und Zeugnis: Beiträge zu einer Kritik der Zuverlässigkeit und Zweckmäßigkeit der Ziffernbenotung*. Linz: Quirin Haslinger.
- Wenz, S. E. & Hoenig, K. (2020). Ethnic and social class discrimination in education: Experimental evidence from Germany. *Research in Social Stratification and Mobility*, 65. doi: 10.1016/j.rssm.2019.100461