pedocs
Open Access Erziehungswissenschaften

DIPF

McElvany, Nele [Hrsg.]; Grecu, Alyssa Laureen [Hrsg.]; Lorenz, Ramona [Hrsg.]; Becker, Michael [Hrsg.]; Dignath, Charlotte [Hrsg.]; Gaspard, Hanna [Hrsg.]; Lauermann, Fani [Hrsg.]

Jahrbuch der Schulentwicklung. Band 23. 50 Jahre Schulentwicklung - Leitthemen der empirischen Bildungsforschung

Weinheim : Beltz Juventa 2023, 399 S. - (Veröffentlichung des Instituts für Schulentwicklungsforschung)

Quellenangabe/ Reference:
McElvany, Nele [Hrsg.]; Grecu, Alyssa Laureen [Hrsg.]; Lorenz, Ramona [Hrsg.]; Becker, Michael [Hrsg.]; Dignath, Charlotte [Hrsg.]; Gaspard, Hanna [Hrsg.]; Lauermann, Fani [Hrsg.]: Jahrbuch der Schulentwicklung. Band 23. 50 Jahre Schulentwicklung - Leitthemen der empirischen Bildungsforschung. Weinheim : Beltz Juventa 2023, 399 S. - (Veröffentlichung des Instituts für Schulentwicklungsforschung) - URN: urn:nbn:de:0111-pedocs-291903 - DOI: 10.25656/01:29190

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-291903
https://doi.org/10.25656/01:29190

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

http://www.juventa.de

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: http://creativecommons.org/licenses/by-nc-sa/4.0/deed.de - Sie dürfen das Werk bzw. den Inhalt unter folgenden Bedingungen vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen: Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen. Dieses Werk bzw. der Inhalt darf nicht für kommerzielle Zwecke verwendet werden. Die neu entstandenen Werke bzw. Inhalte dürfen nur unter Verwendung von Lizenzbedingungen weitergegeben werden, die mit denen dieses Lizenzvertrages identisch oder vergleichbar sind.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en - You may copy, distribute and transmit, adapt or exhibit the work in the public and alter, transform or change this work as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work. If you alter, transform, or change this work in any way, you may distribute the resulting work only under this or a comparable license.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

pedocs
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der Leibniz-Gemeinschaft

Institut für Schulentwicklungsforschung

Nele McElvany | Alyssa Laureen Grecu |
Ramona Lorenz | Michael Becker |
Charlotte Dignath | Hanna Gaspard |
Fani Lauermann (Hrsg.)

Jahrbuch der Schulentwicklung

Band 23

50 Jahre Schulentwicklung – Leitthemen der empirischen Bildungsforschung

BELTZ JUVENTA

Nele McElvany | Alyssa Grecu | Ramona Lorenz | Michael Becker |
Charlotte Dignath | Hanna Gaspard | Fani Lauermann (Hrsg.)
Jahrbuch der Schulentwicklung. Band 23

Eine Veröffentlichung
des Instituts für Schulentwicklungsforschung
der Technischen Universität Dortmund

Nele McElvany | Alyssa Grecu |
Ramona Lorenz | Michael Becker |
Charlotte Dignath | Hanna Gaspard |
Fani Lauermann (Hrsg.)

Jahrbuch der Schulentwicklung Band 23

50 Jahre Schulentwicklung – Leitthemen der empirischen Bildungsforschung

BELTZ JUVENTA

Mit Online Materialien

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Der Text dieser Publikation wird unter der Lizenz **Creative Commons Namensnennung – Nicht-kommerziell – Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-NC-SA 4.0)** veröffentlicht. Den vollständigen Lizenztext finden Sie unter: https://creativecommons.org/licenses/by-nc-sa/4.0/deed.de.

Verwertung, die den Rahmen der **CC BY-NC-SA 4.0 Lizenz** überschreitet, ist ohne Zustimmung des Verlags unzulässig. Die in diesem Werk enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Quellenangabe/Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Dieses Buch ist erhältlich als:
ISBN 978-3-7799-7718-6 Print
ISBN 978-3-7799-7719-3 E-Book (PDF)

1. Auflage 2023

© 2023 Beltz Juventa
in der Verlagsgruppe Beltz · Weinheim Basel
Werderstraße 10, 69469 Weinheim
Alle Rechte vorbehalten

Herstellung: Myriam Frericks
Satz: Helmut Rohde, Euskirchen
Druck und Bindung: Beltz Grafische Betriebe, Bad Langensalza
Beltz Grafische Betriebe ist ein klimaneutrales Unternehmen (ID 15985-2104-100)
Printed in Germany

Weitere Informationen zu unseren Autor:innen und Titeln finden Sie unter: www.beltz.de

Inhalt

Grußwort der Herausgeber*innen 7

Grußwort des IFS-Beiratsvorsitzenden 11

Soziale Disparitäten im deutschen Bildungssystem
Theoretische Modelle und empirische Befunde
Tobias C. Stubbe & Maria Krieg 14

Differenzielle Überzeugungen von angehenden Lehrkräften und deren Bedeutung für eine diversitätssensible Unterrichtsgestaltung
Miriam M. Gebauer & Ricarda K. Rübben 45

Prozessbezogene Diagnosequalität im Kontext von Schullaufbahnentscheidungen
Stefanie van Ophuysen, Lars Behrmann & Sina Schürer 72

Students' Use of (Online) Self-Assessments in Math: Interindividual Differences in Students' Test Engagement and Test Performance in Math-Intensive Study Programs
Fani Lauermann, Daria K. Benden, Jeffrey M. DeVries & Johanna Heitzer 97

What can Digital Trace Data Tell Us about Post-secondary Students' Academic Success?
An Overview of the Literature and an Illustrative Example
Luise von Keyserlingk, Fani Lauermann, Renzhe Yu, Charlott Rubach, Richard Arum 128

Effekte der (mittleren) Klassenleistung auf die individuelle Leistungsentwicklung: Konzeptionelle Einordnung, aktuelle Debatten und offene Fragen vor dem Hintergrund des deutschen Schulsystems
Michael Becker, Aleksander Kocaj, Malte Jansen, Hanna Dumont, Oliver Lüdtke & Poldi Kuhl 150

Bedingungen der Lesekompetenz von Grundschulkindern – Zentrale Ergebnisse des 15-Jahres-Trends der IGLU-Studie zu Lesebedingungen in Deutschland im internationalen Vergleich
Ramona Lorenz, Ulrich Ludewig, Rahim Schaufelberger, Ruben Kleinkorres & Nele McElvany 180

Quality of ICT Use in German Secondary Schools During the COVID-19 Pandemic: Predictors on the School and Teacher Level
Sittipan Yotyodying & Ramona Lorenz 207

Towards the Measurability of a Competence Model for School-based Civic Education
Pascal Alscher, Alyssa Grecu, Ulrich Ludewig, Nele McElvany 230

Promoting Motivation in Mathematics: Can a Relevance Intervention Contribute to Reducing Educational Disparities?
Hanna Gaspard, Cora Parrisius & Benjamin Nagengast 262

Die Rolle der Lehrkraft im Kontext von Selbstregulation beim Lernen
(Angehende) Lehrkräfte als selbstreguliert Lernende und selbstregulationsfördernde Lehrende
Charlotte Dignath & Antonia Fischer 291

50 Jahre Institut für Schulentwicklungsforschung (IFS) – Beiträge des Instituts zur Empirischen Bildungsforschung
Ramona Lorenz & Alyssa Grecu 315

Autorinnen und Autoren 331

Grußwort der Herausgeber*innen

Mit der 23. Ausgabe des Jahrbuchs für Schulentwicklung rahmt das Institut für Schulentwicklungsforschung (IFS) sein 50-jähriges Bestehen und damit ein besonderes Jubiläum. Fünf Jahrzehnte hat die Forschung am IFS die Schulentwicklung und Empirische Bildungsforschung in zentralen Themenbereichen geprägt sowie Entwicklungen im Bildungsbereich initiiert, begleitet und evaluiert. Als ein führendes Forschungsinstitut in Deutschland hat das IFS herausragende Forschungsarbeiten und Projekte hervorgebracht, die in der schulischen Praxis, der Bildungspolitik und Bildungsadministration sowie in der Wissenschaft für zahlreiche Veränderungen und Weiterentwicklungen sorgten und diese Prozesse begleiteten.

Der vorliegende Band umfasst in der Tradition des Jahrbuchs für Schulentwicklung aktuelle Forschungsarbeiten, die mit dem IFS in Verbindung stehen. Aufgrund des Jubiläums sollte jedoch kein spezifischer inhaltlicher Schwerpunkt leitend für die Zusammenstellung der Beiträge sein, sondern die Forschungsstärke des Instituts in der Vielzahl seiner Forschungsfelder repräsentiert werden. Dazu enthält das Jahrbuch für Schulentwicklungsforschung 2023 ausgewählte Beiträge von Professor*innen, die ehemals am IFS tätig waren. Mit ihren Wurzeln am IFS erweitern sie die Reichweite der Grundlagen und Impulse des Instituts und präsentieren Einblicke in ihre aktuellen Forschungsarbeiten. Zudem bietet das vorliegende Jahrbuch einen Einblick in die breit aufgestellten Forschungsbereiche des IFS, indem es aktuelle Forschungsarbeiten der mit dem Institut verbundenen Professor*innen und Mitarbeiter*innen präsentiert. Die Forschungsstärke des Instituts wird hier durch die Qualität der wissenschaftlichen Forschung und die multiperspektivischen Blickwinkel besonders deutlich. Von Interventionsstudien bis hin zu internationalen Schulleistungsstudien, von Themen auf Individualebene bis hin zur Systemebene sind zentrale Forschungsrichtungen repräsentiert, was die Breite der Forschung am IFS illustrieren kann:

Stubbe und Krieg nehmen Operationalisierungen des sozialen Status von Familien zur Erklärung sozialer Disparitäten im Bildungserfolg in den Blick und benennen damit verbundene Herausforderungen. Unter Bezugnahme auf Bourdieus Kapitaltheorie referieren sie klassische Indikatoren zur Messung sozialen Kapitals und bieten einen Überblick über den Forschungsstand zu sozialen Disparitäten mit einem Fokus auf prominente Panelstu-

dien. Neben deskriptiven Analysen der Operationalisierungsmöglichkeiten sozialen Kapitals stützen sie sich auf vertiefende Trendanalysen von TIMSS 2007 bis 2019.

Gebauer und Rübben fokussieren diversitätsbezogene Einstellungen angehender Lehrkräfte und deren Bedeutung für eine diversitätssensible Unterrichtsgestaltung. Ausgehend von der Annahme, dass die Überzeugungen von Lehrkräften eine diversitätssensible Unterrichtsgestaltung beeinflussen, wurde eine quantitative Fragebogenstudie mit Lehramtsstudierenden einer nordrhein-westfälischen Universität durchgeführt.

van Ophuysen, Behrmann und Schürer untersuchen die Bedeutung von Aspekten der prozessbezogenen Diagnosequalität aus Perspektive von Lehrkräften der Primar- und Sekundarstufe. Der Fokus des Beitrags liegt auf der pädagogischen Diagnostik im Kontext von Schullaufbahnempfehlungen am Ende der Grundschulzeit und am Ende der Erprobungsstufe an Gymnasien. Basierend auf dem Vier-Komponenten-Modell der Diagnosequalität identifizieren die Autor*innen Prozessmerkmale einer qualitativ hochwertigen Diagnostik. Die empirische Basis stellt eine standardisierte Online-Befragung von Lehrkräften an Grundschulen und Gymnasien dar.

Lauermann, Benden, DeVries und Heitzer nehmen interindividuelle Unterschiede bei der Qualität der Testbearbeitung eines Selbsteinschätzungstests für mathematische Grundkenntnisse (WINT-Check) für Studierende mathematikintensiver Studiengänge in den Blick und leisten einen Beitrag zur verbesserten Einschätzung solcher Tests als Informations- und Screeninginstrument für Hochschulen. Mittels Pfadanalysen werden demografische Daten, Leistungsunterschiede und mathematikbezogene Motivation mit verschiedenen Indikatoren der Qualität der Testbearbeitung und Leistung verknüpft.

von Keyserlingk, Lauermann, Yu, Rubach und Arum nutzen Verhaltensspurdaten auf Lernmanagementsystemen wie Moodle, um selbstreguliertes Lernverhalten von Studierenden zu untersuchen. Daten wie das Klickverhalten, die Bearbeitung freiwilliger Testfragen oder die Zeit, die Studierende auf der Lernplattform verbringen, werden vor dem Hintergrund des Kursablaufes und der Charakteristika der Studierenden in Verbindung mit deren Lernerfolg und Noten gebracht.

Becker, Kocaj, Jansen, Dumont, Lüdtke und Kuhl diskutieren die Effekte, die die Komposition von Lerngruppen auf die Leistungsentwicklung der individuellen Gruppenmitglieder haben. Die inhaltlichen und methodischen Probleme in der Konzeption, Analyse und Interpretation von Klassenkompositionseffekten werden vor dem Hintergrund „klassischer“ und aktueller Forschungsbefunde und Kontroversen dargestellt und diskutiert.

Lorenz, Ludewig, Schaufelberger, Kleinkorres und McElvany stellen Trends von 15 Jahren der Internationalen Grundschul-Lese-Untersuchung (IGLU) vor. Besonders im Fokus steht dabei, wie sich zentrale Lernbedingungen der Schülerschaften in Deutschland im internationalen Vergleich zwischen 2001 und 2016 gewandelt haben. Neben Schülerzahlen werden Bedingungsfaktoren des häuslichen und digitalen Umfelds, der Klassenkomposition und der Motivation dargestellt.

Yotyodying und Lorenz erforschen die Prädiktoren für einen qualitätsvollen Einsatz von digitalen Medien im Distanzunterricht vor dem Hintergrund des hybriden Unterrichts während der COVID-19-Pandemie. Berücksichtigt wurden sowohl die Nutzungsmerkmale der Lehrkräfte, als auch die der Schüler*innen. Auf der Basis einer repräsentativen Lehrkräftebefragung wurden Strukturgleichungsmodelle zur Beantwortung der Fragestellungen berechnet.

Alscher, Grecu, Ludewig und McElvany stellen ein Kompetenzmodell für die politische Bildung in Deutschland vor. Das Modell umfasst die Dimensionen des politischen Wissens, der politischen Motivation, der politischen Einstellungen und des politischen Willens. Dieses Modell wird mit Mittelwertsvergleichen und konfirmatorischer Faktorenanalyse empirisch überprüft.

Gaspard, Parrisius und Nagengast untersuchen eine Intervention, die Schüler*innen der neunten Klasse in ihrer wahrgenommenen Relevanz im Mathematikunterricht unterstützen soll. Die Intervention soll insbesondere „Risikogruppen" von Schüler*innen langfristig in ihrer Motivation und ihren Leistungen im Mathematikunterricht fördern. In einer Moderationsanalyse wird untersucht, inwiefern die Intervention erfolgreich ist, diese Risikogruppen anzusprechen.

Dignath und Fischer befassen sich in ihrem Beitrag mit den Rollen von Lehrkräften im Bereich des selbstregulierten Lernens. Lehrkräfte fördern zum einen das selbstregulierende Lernen von Schüler*innen, zum anderen sind sie selbst Lernende und Unterrichtende, die ihr Lernen und das Unterrichten selbst regulieren müssen. Es werden mehrere Forschungsbefunde, die sich mit den verschiedenen Rollen von Lehrkräften befassen, vorgestellt.

Abgerundet wird die Jubiläumsausgabe des Jahrbuchs für Schulentwicklung durch einen Beitrag, in dem die „50 Jahre IFS" von seinen Professor*innen rückblickend und auch mit Blick auf die Zukunft in Interviews besprochen werden. Sie heben zentrale Handlungsfelder, den Einfluss des IFS auf unterschiedliche Akteure und besondere Leistungen hervor und wagen abschließend einen Blick auf zukünftige Entwicklungen.

Ein besonderer Dank gilt den Gutachter*innen, die im Zuge des internen und externen Reviews der Beiträge mit ihrer Expertise maßgeblich zur Sicherung der Qualität der Kapitel beigetragen haben. Dazu bedanken wir uns herzliche neben den Kolleg*innen am IFS bei Prof. Dr. Hermann Josef Abs, Prof. Dr. Michael Besser, Dr. Malte Jansen, Prof. Dr. Michael Kerres, Dr. Christoph König, Prof. Dr. Johannes König, Prof. Dr. Falk Radisch, Prof. Dr. Florian Schmiedek und Prof. Dr. Birgit Spinath für die Begutachtung von Beiträgen.

Die Beiträge des vorliegenden 23. Jahrbuchs der Schulentwicklung sollen einen wertvollen Beitrag für die Bildungsforschung darstellen und nach 50 Jahren Forschung am IFS Impulse für die aktuelle, evidenzbasierte Diskussion zentraler Fragen der Schulentwicklung bieten.

Dortmund, im Juni 2023
Nele McElvany
Alyssa Grecu
Ramona Lorenz
Michael Becker
Charlotte Dignath
Hanna Gaspard
Fani Lauermann

Herausgeber*innen

Grußwort des IFS-Beiratsvorsitzenden

Das Institut für Schulentwicklungsforschung (IFS) an der Technischen Universität Dortmund entwickelte sich nach seiner Gründung 1973 zu einer bedeutenden Größe unter den bundesdeutschen Schul- und Bildungsforschungsinstituten mit wissenschaftlicher Strahlkraft weit über die Grenzen Deutschlands hinaus.

Von Anfang an bildet die empirische Bildungs- und Schulentwicklungsforschung das Kernziel des IFS. Somit bietet das IFS Entscheidungsträgern nicht nur des bundesrepublikanischen Schulsystems sondern auch auf internationaler Ebene substantielle, empirisch evidente Kenntnisse zu aktuellen Fragestellungen im Bereich von Bildung und Schule an. Die bedeutsamen Beiträge des IFS für Wissenschaft und insbesondere die Kultuspolitik werden ermöglicht durch einen interdisziplinären Ansatz, in dem alle heute unter dem Begriff „Bildungswissenschaften" firmierenden Teilgebiete verortet werden können. Seit Jahrzehnten gelingt es dadurch dem IFS, das selbstgesteckte Ziel Lern- und Entwicklungsprozesse, Schulentwicklung und Bildungsergebnisse unter den relevanten Kontextbedingungen aller Ebenen empirisch gesichert zu beschreiben und daraus notwendiges Steuerungswissen für Bildungssysteme zur Verfügung zu stellen, um Bildungsprozesse zu optimieren.

Man kann die TU Dortmund nur dazu beglückwünschen, ein solch forschungsstarkes und national sowie international höchst visibles Institut wie das IFS als ein Aushängeschild des eigenen Profilbereichs „Bildung, Schule und Inklusion" dauerhaft in seinen Strukturen verankert zu haben. Hierin zeigt sich, wie gesellschaftlich bedeutsam und gleichzeitig wissenschaftlich erfolgreich eine dauerhafte Institutionalisierung sein kann, da sowohl in der empirischen Bildungsforschung als auch in der Kultuspolitik das IFS kontinuierlich wichtigste Beiträge lieferte. Hierbei wurde aber auch immer darauf geachtet, nicht erst reaktiv, sondern proaktiv die Erkenntnisfortschritte im bildungswissenschaftlichen Bereich mitzugestalten, was sich unter anderem sehr gut an der methodisch progressiven Arbeitsart und -weise des IFS über die Jahrzehnte erkennen lässt. Dies spiegelt sich auch in der Personalpolitik des IFS wieder, die darauf bedacht ist, aufstrebende, innovative und international visible Forschungspersönlichkeiten an das IFS zu bringen, beziehungsweise dort in ihrer weiteren Entwicklung zur internationalen

Forschungsspitze in dem Feld der empirischen Bildungsforschung zu unterstützen. Durch den Mut des IFS zur Innovation werden immer wieder neue, aktuelle und gesellschaftsrelevante Themenfelder besetzt und bisherige Themen gegebenenfalls dann auch konsequent abgeschlossen, ohne dabei einen eklektizistischen, vom Zeitgeist getriebenen Wissenschaftsansatz zu verfolgen. Vielmehr arbeitet das IFS auf einer klaren wissenschaftlich-empirischen Grundhaltung, basierend auf aktuellsten, vielfältigen Methodiken. Durch dieses wissenschaftliche Streben nach empirisch prüfbarer Wahrheit werden Kontrastpunkte zu vielen weltanschaulich und allein politisch-normativ motivierten Standpunkten im Bildungsbereichen entgegengesetzt. Dies ist keine Selbstverständlichkeit, da hier keine Auftragsforschung im Sinne potenzieller Geldgeber betrieben wurde und wird, sondern wissenschaftliche Methoden als Grundlage zur gesicherten Gestaltung von Bildungszielen eingesetzt werden. Das ist im Zeitalter sogenannter „gefühlter bzw. alternativer Fakten" ein essentielles Korrektiv, um Bildungsprozesse verbunden mit der Ausgestaltung der Bildungsinstitutionen als basale Säule moderner, globalisierter und vor allem demokratischer Wissensgesellschaften weiterzuentwickeln. Betrachtet man die Herausforderungen des 21. Jahrhunderts, so ist der „Kampf um die besten Köpfe" sicherlich einer der wichtigsten Erfolgsfaktoren, um Herausforderungen im globalen Wettbewerb, aber insbesondere in der globalen Kooperation, zu bestehen. Hierüber entscheidet auch die Qualität der Bildungssysteme. Dazu liefert das IFS insbesondere im Bereich Schule substanzielle Erkenntnisse.

Hervorzuheben ist, dass diese Erkenntnisse dabei nicht nur in wissenschaftsüblichen, schriftlichen Formaten kommuniziert werden, sondern es ist zu betonen, dass das IFS gerade in den letzten Jahren in beispielhaftem Maße die rasche Wissenschaftskommunikation relevanter Forschungsergebnisse in die Bildungspraxis befördert. So veranstaltet das IFS jährlich zwei in ihren Schwerpunkt unterschiedlich fokussierte Konferenzen: „Das Dortmunder Symposium der Empirischen Bildungsforschung" und den „IFS-Bildungsdialog". Diese jährlich stattfindenden Veranstaltungen wechseln in ihren Schwerpunktsetzungen und werden ergänzt von der „IFS Virtual Keynote Series", die internationalen Spitzenforscher:innen aus der empirischen Bildungsforschung ein Podium bietet. Dass all diese Wissenschaftskommunikationsaktivitäten gerahmt werden von einem entsprechend informativen Webauftritt, ist nur ein weiterer Beleg für die hohen Standards der Wissenschaftskommunikation, die am IFS gepflegt werden.

Dementsprechend gratuliere ich dem IFS zu fünf Jahrzehnter höchst erfolgreicher interdisziplinärer, multimethodaler und praxisrelevanter Forschung im Bereich der schulischen empirischen Bildungsforschung und

wünsche allen Beteiligten die Kraft und Inspiration, auch in den nächsten Jahrzehnten Bildungsforschung auf diesem Niveau zu betreiben und die Praxis einzubringen.

Frankfurt a. M., im Sommer 2023
Prof. Dr. Holger Horz
Vorsitzender des IFS-Beirats

Soziale Disparitäten im deutschen Bildungssystem

Theoretische Modelle und empirische Befunde

Tobias C. Stubbe & Maria Krieg

Zusammenfassung

Ein unverändert wichtiges Thema der empirischen Bildungsforschung mit einer besonderen gesamtgesellschaftlichen Bedeutung stellt die Analyse von sozialen Disparitäten im deutschen Bildungssystem dar. In diesem Beitrag werden Theorien und empirische Befunde zu den folgenden Fragestellungen aus diesem Themenkomplex präsentiert: Wie lässt sich der soziale Status von Familien operationalisieren? Wie ausgeprägt sind soziale Disparitäten (im Zeitverlauf und im internationalen Vergleich)? Wie lassen sich soziale Disparitäten erklären?

Ein besonderer Fokus liegt dabei auf den Forschungsaktivitäten am Arbeitsbereich des Erstautors: insbesondere unterschiedliche Ansätze der Operationalisierung von sozialem Status sowie sekundäre Herkunftseffekte im Primar- und im Sekundarbereich.

Schlagworte: soziale Disparitäten im Bildungssystem; primäre Herkunftseffekte; sekundäre Herkunftseffekte; Kapitaltheorie nach Bourdieu; Armutsgefährdung; soziales Kapital; Ressourcengenerator

Abstract

The analysis of social disparities in the German education system continues to be an important topic of empirical educational research with particular importance for society as a whole. This paper presents theories and empirical findings on the following questions from this complex of topics: How can the social status of families be operationalized? How pronounced are social disparities (over time and in international comparison)? How can social disparities be explained? A special focus is on the research activi-

ties in the first author's field of work: in particular, different approaches to operationalizing social status as well as secondary effects of origin in primary and secondary education.

Keywords: social disparities in the education system; primary effects of social origin; secondary effects of social origin; Bourdieu's theory of capital; poverty risk; social capital; resource generator

1. Einleitung

Seit Jahrzehnten ist national ebenso wie international empirisch umfassend belegt, dass der Bildungserfolg von Kindern in einem engen Zusammenhang mit dem sozioökonomischen Status (*socio-economic status*; SES) ihrer Familien steht (Eriksson et al., 2021; OECD, 2019; Stubbe, Krieg, et al. 2020; Stubbe, Kleinkorres, et al. 2023; Weis et al., 2019). Der wissenschaftliche und politisch-gesellschaftliche Diskurs zu den Auswirkungen der Coronapandemie im Bildungssystem richtet den Fokus nun erneut auf das altbekannte und überaus wichtige Thema der sozialen Disparitäten. Aktuelle Forschungsergebnisse bestätigen die befürchtete Verstärkung divergierender Teilhabechancen zwischen sozial privilegierten und benachteiligten Kindern als Folge der coronabedingten Schulschließungen (Engzell et al., 2021; Weber et al., 2021b). Diese Entwicklung ist umso dramatischer, da ein gesamtgesellschaftlicher Konsens darüber bestehen dürfte, dass alle Kinder – unabhängig von ihrem Geschlecht, ihrem Migrationshintergrund oder dem sozialen Status ihrer Eltern – die gleichen Bildungschancen haben sollten (Grundgesetz für die Bundesrepublik Deutschland, Artikel 3.3 und The Universal Declaration of Human Rights, Artikel 26.1). Weder die mit der Bildungsexpansion einhergehende gesamtgesellschaftliche Höherqualifizierung noch großangelegte Förderprogramme, wie das Aktionsprogramm „Aufholen nach Corona" (Bundesregierung, 2021), konnten jedoch bislang einen substantiellen Beitrag zur Verringerung der existierenden Ungleichheiten leisten (Blossfeld et al., 2019; Helbig et al., 2022). Umso zentraler ist die kontinuierliche wissenschaftlich fundierte Begleitung sozialer Disparitäten im Bildungssystem, um ungleichheitsverstärkende Mechanismen identifizieren zu können und schulpraktische Implikationen für mehr Chancengleichheit abzuleiten.

In diesem Beitrag werden Theorien und empirische Befunde zu den folgenden Fragestellungen aus diesem Themenkomplex präsentiert: Kapitel 2: Wie lässt sich der soziale Status von Familien operationalisieren? Kapitel 3: Wie ausgeprägt sind soziale Disparitäten (im Zeitverlauf und im internatio-

nalen Vergleich)? Kapitel 4: Wie lassen sich soziale Disparitäten erklären? In Kapitel 5 wird ein kurzes Fazit gezogen.

2. Wie lässt sich der soziale Status von Familien operationalisieren?

Notwendige Voraussetzung für alle Analysen zum Thema SES ist selbstverständlich die reliable und valide Operationalisierung des zunächst theoretisch definierten Konstrukts. Dass dies alles andere als trivial ist, wird in diesem Kapitel verdeutlicht, indem zuerst ein Überblick über die zugrundeliegenden theoretischen Annahmen (Kap. 2.1) und die derzeit in der empirischen Bildungsforschung genutzten Instrumente (Kap. 2.2) gegeben wird. Anschließend wird in Kapitel 2.3 näher auf das theoretische Konstrukt der Einkommensarmut eingegangen, das in der soziologischen Forschung und im politisch-gesellschaftlichen Diskurs eine wichtige Rolle einnimmt, in den großen Schulleistungsstudien bislang jedoch nur vereinzelt Berücksichtigung fand. Bereits im Rahmen des *Programme for International Student Assessment* (PISA) 2000 (Baumert et al., 2000) wurde darauf hingewiesen, dass die Operationalisierung von sozialem Kapital im Rahmen von *Large Scale Assessments* deutlich herausfordernder ist als die von ökonomischem beziehungsweise kulturellem Kapital. In Kapitel 2.4 wird daher näher auf diesen Aspekt eingegangen und es werden Instrumente benannt, die es möglicherweise erlauben, soziales Kapital in der empirischen Bildungsforschung valider zu erfassen, als dies in der Vergangenheit geschehen ist.

2.1. Theoretische Grundlagen

Um den sozialen Status von Familien zu analysieren, hat sich in der empirischen Bildungsforschung die Kapitaltheorie des Soziologen Pierre Bourdieu (1983) etabliert (Becker, 2017; Blossfeld et al., 2019). Die Kapitaltheorie unterscheidet drei Kapitalarten (ökonomisches, kulturelles und soziales Kapital), anhand derer die familiale Ausstattung mit bildungsrelevanten Ressourcen gemessen werden kann. Unter dem ökonomischen Kapital versteht Bourdieu die finanziellen und materiellen Mittel einer Familie. Das kulturelle Kapital mit seinen drei Subfacetten (a) inkorporiertes Kulturkapital, (b) objektiviertes Kulturkapital und (c) institutionalisiertes Kulturkapital beschreibt, inwieweit Personen über (a) Wissen und Fähigkeiten, (b) kulturelle Besitztümer (z. B. Bücher) oder (c) schulische und akademische Titel verfügen. Die dritte Kapi-

talform – das soziale Kapital – gibt Auskunft über das soziale Beziehungsnetzwerk einer Familie. Neben der Anzahl der Beziehungen bestimmt auch die Kapitalausstattung der Personen innerhalb des Netzwerks über die Höhe des sozialen Kapitals der Familien. Ergänzend wird im Bildungskontext häufig auf die Arbeit von Coleman (1988) verwiesen, der zur Beschreibung des sozialen Kapitals explizit einen Fokus auf Bildungsprozesse legt und dessen Bedeutung für den schulischen Kompetenzerwerb hervorhebt. Für alle Kapitalarten gilt, dass diese ineinander transformierbar sind – gegebenenfalls unter Einsatz von Transformationskosten (Becker, 2017; Mikus et al., 2020). Ermöglichen Familien ihren Kindern beispielsweise einen höheren Schulabschluss durch die Finanzierung von Nachhilfe oder den Kauf ergänzender Lernmaterialien, so wird ökonomisches Kapital in institutionalisiertes Kulturkapital umgewandelt. Andererseits besteht ein enger Zusammenhang zwischen dem kulturellen Kapital und dem ökonomischen Kapital: Personen mit mehr kulturellem Kapital sind meist höherqualifiziert und können dadurch in besser bezahlten Berufen tätig sein. Entsprechend verfügen sie häufig auch über mehr ökonomisches Kapital.

2.2. Gängige Indikatoren

In diesem Abschnitt werden die in der empirischen Bildungsforschung zurzeit üblicherweise genutzten Indikatoren kurz dargestellt.

Der Klassiker: Bücher im Haushalt

Es ist sicherlich nicht übertrieben, die Frage nach der Anzahl der im Haushalt vorhandenen Bücher als den Klassiker zur Erfassung von SES in der Schulleistungsforschung zu bezeichnen (Engzell, 2021; Heppt et al., 2022; Schwippert, 2019; Sieben & Lechner, 2019). Erhoben wird diese Information üblicherweise mit Hilfe einer fünfstufigen (z. B. PIRLS (*Progress in International Reading Literacy Study*) und TIMSS (*Trends in International Mathematics and Science Study*)) oder sechsstufigen (z. B. PISA, NEPS (*National Educational Panel Study*)) Likert-Skala.

Aus theoretischer Perspektive erfasst die Variable primär das objektivierte Kulturkapital (Sieben & Lechner, 2019); liefert aber auch einen Hinweis auf das ökonomische Kapital. Indirekt misst die Variable zusätzlich das inkorporierte (und institutionalisierte) Kulturkapital, da ein Zusammenhang zwischen dem Bildungsniveau einer Familie und der Anzahl der im Haushalt verfügbaren Bücher besteht (Heppt et al., 2022).

Empirisch zeigt sich eine enge Koppelung zwischen der Anzahl der Bücher im Haushalt und den erreichten Kompetenzen von Schülerinnen und Schülern in den verschiedenen Domänen (Eriksson et al., 2021; Heppt et al., 2022). Die Korrelation zwischen der im Rahmen von IGLU (*Internationale Grundschul-Lese-Untersuchung*) 2006 erhobenen Lesekompetenz und der Anzahl der Bücher beträgt beispielsweise $r = .32$ (Schülerangabe) beziehungsweise $r = .33$ (Elternangabe). Zudem zeigt sich auch eine relativ hohe Korrelation zwischen der Büchervariable und anderen SES-Indikatoren. So liegt die Korrelation mit dem HISEI (s. u.) für IGLU 2006 bei $r = .33$ (Schülerangabe) beziehungsweise $r = .47$ (Elternangabe) (Stubbe & Goy, 2013). Zu ähnlichen Ergebnissen kommen auch Eriksson et al. (2021), die auf Basis von PISA 2018-Daten die Anzahl der im Haushalt vorhandenen Bücher gegenüber anderen SES-Indikatoren als stärksten Prädiktor der Leistung in Mathematik, Naturwissenschaften und Lesen herausstellen.

Aus erhebungsökonomischer Sicht ist die Frage nach der Anzahl der Bücher somit ausgesprochen effektiv: Sie kann bereits Grundschulkindern gestellt werden, sie kann schnell beantwortet und einfach ausgewertet werden, sie ist theoretisch fundiert und sie weist relativ hohe Korrelationen mit anderen SES-Indikatoren auf. Dennoch sollte keinesfalls außer Acht gelassen werden, dass mit diesem einfachen Merkmal versucht wird, ein komplexes theoretisches Konstrukt zu erfassen, was zwangsläufig einen substanziellen Informationsverlust bedeutet, da die unterschiedlichen Kapitalarten nicht differenziert gemessen werden.

Bildungs- und Berufsabschlüsse

Das institutionalisierte und damit indirekt auch das inkorporierte Kulturkapital kann in Form der erworbenen Abschlüsse gemessen werden. Aufgrund der großen Vielfalt werden Berufsabschlüsse üblicherweise mit offenen Fragen erfasst, während Bildungsabschlüsse auch mit Hilfe von geschlossenen Fragen erhoben werden können. Da beide Konstrukte zunächst ein nominales Skalenniveau aufweisen, müssen diese rekodiert werden, um Aussagen über die Höhe des mit den jeweiligen Abschlüssen verbundenen kulturellen Kapitals treffen zu können. Dazu stehen unterschiedliche Indizes zur Verfügung, die größtenteils für internationale Vergleichsstudien entwickelt wurden.

Für Analysen des Berufsstatus müssen in einem ersten Schritt die offenen Angaben mit Hilfe der Nominalskala *International Standard Classification of Occupations* (ISCO) kodiert werden. Dazu wird jedem Beruf ein vierstelliger Code zugewiesen, der allerdings noch keine Aussagen über den Status

eines Berufes erlaubt (Hoffmann, 2003). Diese Codes lassen sich dann allerdings problemlos in die verschiedenen Indizes des Berufsstatus rekodieren.

Der *International Socio-Economic Index of Occupational Status* (ISEI) gibt an, wie gut ein Beruf geeignet ist, um die Ausbildung einer Person in Einkommen umzuwandeln. Die Skala ist ebenfalls eindimensional und nimmt ganzzahlige Werte zwischen 10 und 90 an, wobei hohe Werte einen hohen sozioökonomischen Status der Berufe kennzeichnen (Ganzeboom et al., 1992; Ganzeboom & Treiman, 1996, 2010). In der empirischen Bildungsforschung wird häufig der höchste ISEI im Haushalt (HISEI) genutzt.

Erikson et al. (1979) teilen Berufe in distinkte Gruppen ein, die sich in Hinblick auf unterschiedliche Dimensionen (z. B. Art der Tätigkeit, angestellte oder selbstständige Tätigkeit, Weisungsbefugnis) durch eine hohe interne Homogenität und eine hohe externe Heterogenität auszeichnen. Bei diesen sogenannten EGP-Klassen handelt es sich um eine Nominalskala.

Welcher dieser Indizes für eine bestimmte Fragestellung genutzt werden sollte, hängt zum einen davon ab, welche theoretischen Annahmen zugrunde gelegt werden und auch davon, welche statistischen Eigenschaften jeweils gewünscht werden (Ganzeboom & Treiman, 2003).

Als überraschend herausfordernd erweist sich die Operationalisierung von Bildungsabschlüssen (Hoffmeyer-Zlotnik, 2003). In der empirischen Bildungsforschung hat sich in den vergangenen Jahren dafür die von der UNECSO entwickelte *International Standard Classification of Education* (ISCED) (UNESCO, 2003) als Standard etabliert, obwohl sie in der Praxis insbesondere für internationale Vergleiche nur eingeschränkt geeignet ist. Schröder und Ganzeboom (2014) kommen in ihrer Zusammenfassung des Forschungsstandes zu dem Schluss, die ISCED-97-Kategorisierung sei nicht differenziert genug, um beispielsweise detailliert zwischen beruflichen und akademischen Bildungsabschlüssen der sekundären und tertiären Bildung zu unterscheiden.

Weitere Variablen

Es gibt eine ganze Reihe weiterer Variablen, die in aktuellen Bildungsstudien zum Einsatz kommen, um den sozialen Status von Schülerinnen und Schülern zu operationalisieren: Besitztümer des Haushalts werden als Indikatoren des ökonomischen Kapitals beziehungsweise des objektivierten Kulturkapitals genutzt; Angaben zu Freizeitaktivitäten können Aufschluss über das inkorporierte Kulturkapital geben; Fragen zu Art und Intensität der Eltern-Kind-Beziehung erlauben Rückschlüsse auf das soziale Kapital inner-

halb der Familie und vereinzelt wird das ökonomische Kapital zusätzlich mit Hilfe des Brutto-Haushaltseinkommens gemessen.

Indizes der sozialen Herkunft: Sozialindex, ESCS

Es sollte deutlich geworden sein, dass es sich beim SES um ein komplexes mehrdimensionales Konstrukt handelt. Ist das Ziel aber nicht eine vergleichende Analyse der unterschiedlichen Teilaspekte des SES, stellt sich die Frage, ob die unterschiedlichen Indikatoren zu einem Index zusammengefasst werden können.

Im Rahmen der Hamburger Längsschnittstudie KESS 4 (*Kompetenzen und Einstellungen von Schülerinnen und Schülern*) wurde dieser Frage empirisch mit Hilfe einer Rasch-Skalierung nachgegangen. Pietsch et al. (2006) konnten zeigen, dass sich die unterschiedlichen Indikatoren auf einer eindimensionalen Skala abbilden lassen. Dieser sogenannte Sozialindex wurde seitdem auch in anderen Studien eingesetzt (z. B. Buddeberg et al., 2009; Stubbe & Bos, 2008; Stubbe, Krieg, et al. 2023).

Auch im Rahmen von PISA kommt mit dem ESCS (*Index of Economic, Social and Cultural Status*) ein Index zum Einsatz, der unterschiedliche Aspekte des SES von Schülerfamilien zusammenfasst (Mang et al., 2021). Der offensichtliche Vorteil dieser Indizes ist es, dass für Analysen eine einzige Variable mit metrischem Skalenniveau zur Verfügung steht, in der unterschiedlichste Aspekte des SES zusammengefasst sind. Die Bildung dieser Indizes ist möglich, weil die verschiedenen Indikatoren relativ hoch miteinander korrelieren (Eriksson et al., 2021; Heppt et al., 2022; Sieben & Lechner, 2019).

Empirische Trennung von ökonomischem und kulturellem Kapital

Nichtsdestotrotz sind auch komplexere Fragestellungen forschungspraktisch relevant, die eine empirische Differenzierung der einzelnen Facetten des SES von Familien notwendig machen. So könnte beispielsweise untersucht werden, ob Kinder aus Akademikerhaushalten, die nur über ein geringes Einkommen verfügen (z. B. wegen Arbeitslosigkeit), höhere oder niedrigere schulische Kompetenzen aufweisen als Schülerinnen und Schüler aus bildungsfernen Elternhäusern, die finanziell gut gestellt sind. Anders ausgedrückt: Ist für den Bildungserfolg von Kindern das ökonomische oder das kulturelle Kapital wichtiger?

Eriksson et al. (2021) konnten diesbezüglich auf Basis von PISA 2018-Daten zeigen, dass in Abhängigkeit vom Entwicklungsstandes eines Staates

der Zusammenhang zwischen dem kulturellen beziehungsweise ökonomischen Kapital und der Leistung von Schülerinnen und Schülern divergiert. In weniger entwickelten Staaten war die Leistung der Jugendlichen demnach weniger stark mit dem kulturellen Kapital verbunden und dafür stärker mit dem ökonomischen Kapital. In hochentwickelten Staaten waren hingegen gegenteilige Effekte zu beobachten.

Bereits Stubbe und Goy (2013) stellten eine Skala vor, die die getrennte Analyse dieser beiden Kapitalarten ermöglicht, indem für eine Liste von 22 Aktivitäten beziehungsweise Besitztümer (z. B. „eine mindestens einwöchige Urlaubsreise im Jahr", „täglich frisches Obst und Gemüse essen", „Klavier", „Brettspiele") gefragt wird, ob die Familien (1) dies tun beziehungsweise haben, (2) sich dies zurzeit nicht leisten können oder (3) dies aus anderen Gründen nicht tun beziehungsweise haben. Diese Skala wurde in den vergangenen Jahren im Rahmen der nationalen Erweiterungen von TIMSS und IGLU eingesetzt (zu empirischen Befunden siehe Kap. 3.2).

2.3. Armutsgefährdung

Spätestens mit den Armuts- und Reichtumsberichten der Bundesregierung (Bundesministerium für Arbeit und Soziales, 2001, 2021) ist das Thema der Einkommensarmut auch in der Bundesrepublik in den Fokus der Öffentlichkeit gerückt. Insbesondere die Frage der Kinderarmut ist dabei von Bedeutung, zumal aktuelle Daten aus dem Jahr 2021 zeigen, dass 16.2 Prozent der Kinder in Deutschland von Armut betroffen sind (Statistisches Bundesamt, 2022a). Bei Kindern, die in Haushalten aufwachsen, in denen kein Erziehungsberechtigter erwerbstätig ist, liegt die Armutsgefährdungsquote sogar bei 68.5 Prozent. In einem Zweiverdienerhaushalt verringert sich die Armutsgefährdungsquote von Kindern auf 1.4 Prozent, wenn beide Erziehungsberechtigte in Vollzeit arbeiten (Bundesministerium für Arbeit und Soziales, 2021, S. 56). Außerdem überproportional häufig von Armut betroffen sind Alleinerziehende sowie Familien mit drei oder mehr Kindern (Bundesministerium für Arbeit und Soziales, 2021, S. 477 ff.).

Trotz der offensichtlichen (bildungs-)politischen Relevanz wurde die Einkommensarmut von Schülerfamilien in den großen Schulleistungsstudien zu Beginn des Jahrtausends nicht berücksichtigt. Erstmals im Jahr 2007 wurde im Rahmen von IGLU Belgien und TIMSS Deutschland das Armutsrisiko im Rahmen von *Large Scale Assessments* erhoben (Bos et al., 2010). Als armutsgefährdet gelten entsprechend der Definition der Europäischen Union (Bardone & Guio, 2005) Haushalte, die über weniger als 60 Prozent

des medianen Nettoäquivalenzeinkommens in dem jeweiligen Staat verfügen (Berechnungsverfahren: Bos et al., 2010).

Auch wenn theoriekonform ein enger Zusammenhang zwischen Armutsgefährdung und anderen SES-Indikatoren vorliegt (Stubbe, Krieg, et al. 2020), besteht ein wissenschaftlicher Mehrwert in der Berücksichtigung dieser Variable, da in Deutschland im Jahr 2021 immerhin 8.6 Prozent der Familien mit hohem kulturellen Kapital von Armut bedroht waren (Statistisches Bundesamt, 2022b).

2.4. Soziales Kapital

Die Operationalisierung von sozialem Kapital ist in der Umfrageforschung deutlich herausfordernder, als die Messung der übrigen Kapitalformen. Soziales Kapital manifestiert sich in den sozialen Beziehungen, über die ein Mensch verfügt. Aus Sicht der empirischen Bildungsforschung sind dabei vor allem jene Beziehungen von Interesse, die zur Steigerung des Bildungserfolges genutzt werden können.

Netzwerkanalyse

Es wäre folglich naheliegend, die sozialen Beziehungsstrukturen von Schülerfamilien zu erfassen und netzwerkanalytisch auszuwerten. Ausgehend von Bourdieus Definition von sozialem Kapital könnten ego-zentrierte Netzwerke erhoben werden, um anschließend das ökonomische und kulturelle Kapital der Alteri zu messen (Jansen, 2006 zum Verfahren der Netzwerkanalyse). Aus diesen Daten ließe sich dann ein Index konstruieren, der angibt, wie viel Kapital insgesamt in Egos Netzwerk vorhanden ist und somit als Indikator für das individuelle soziale Kapital genutzt werden kann. Der Erhebungsaufwand wäre in der Forschungspraxis allerdings unverhältnismäßig groß und im Rahmen von *Large Scale Assessments* überhaupt nicht zu leisten.[1]

Werden hingegen nur sehr kleine Stichproben erhoben, ist die Analyse sozialer Netze durchaus ein geeignetes Verfahren, um soziales Kapital zu untersuchen. Beispielsweise können soziale Netzwerke innerhalb von Schulklassen erfasst werden, also ein kleiner aber aus Sicht der Bildungsforschung wichtiger Ausschnitt aus dem Gesamtnetzwerk der jeweiligen Schülerinnen und Schüler (Lorenz & Stubbe, 2021; Stubbe, Pietsch, et al. 2007). Daten dieser Art erlauben unter anderem Analysen zum Peer Learning, also dem

1 Selbst wenn jede Familie nur 30 Bekannte nennt, würde dies ausgehend von 5000 Befragten bedeuten, dass Daten über 150.000 Alteri erhoben werden müssten.

Unterstützungsverhalten beim Lernen beziehungsweise bei den Hausaufgaben innerhalb einer Klasse.

Positions- und Ressourcengenerator

In der empirischen Sozialforschung gibt es aber auch Verfahren, die das soziale Kapital ohne Einsatz netzwerkanalytischer Verfahren messen.

Der Positionsgenerator (Lin & Dumin, 1986) gibt 20 Berufe vor, für die die Befragten angeben sollen, welche davon im eigenen Bekanntenkreis zu finden sind. Das Ausmaß des verfügbaren sozialen Kapitals einer Person ergibt sich dann aus dem Prestige der genannten Berufe. Zwar erfasst dieses Instrument einen größeren Ausschnitt des sozialen Netzwerkes (als bspw. den engsten Familienkreis), aber dennoch nur Teilaspekte, die sich aus der beruflichen Stellung von Personen ergeben.

Mit dem Ressourcengenerator (Snijders, 1999; van der Gaat & Snijders, 2005) wird hingegen zunächst für 20 Items ermittelt, ob die/der Befragte selbst oder jemand, den sie/er persönlich kennt, bestimmte Eigenschaften besitzt (z. B. „kann geschickt Haushaltsgeräte reparieren", „hat Abitur"). Mit weiteren 17 Items wird anschließend erfasst, ob die/der Befragte jemanden kennt, der ihr/ihm bei bestimmten Dingen helfen kann (z. B. „kann bei einem Umzug helfen", „kann bei einem Konflikt mit Familienmitgliedern beraten"). Zusammengefasst ermöglichen diese Items eine Einschätzung des Kapitals, das einer Person innerhalb ihres sozialen Netzwerkes zur Verfügung steht. Van der Gaat und Snijders (2005) betonen, dass der Ressourcengenerator relativ schwach mit anderen SES-Indikatoren korreliert, was die Vermutung nahelegt, dass dieses Konstrukt in der empirischen Bildungsforschung zusätzliche Erklärungskraft für schulische Kompetenzen besitzen könnte.

Eine von Stubbe und Lorenz (2017) angepasste Version dieser Skala – bestehend aus 36 Items – wurde im Rahmen der nationalen Erweiterungen von TIMSS 2015 und IGLU 2016 erprobt. Auf Basis der empirischen Befunde (siehe Kap. 3.2) wurde für TIMSS 2019 und IGLU 2021 eine auf 18 Items gekürzte Version des Ressourcengenerators entwickelt, so dass inzwischen im Rahmen der empirischen Bildungsforschung auch das soziale Kapital von Familien mit einem vertretbaren Aufwand im Sinne der Theorie von Bourdieu operationalisiert werden kann.

3. Wie ausgeprägt sind soziale Disparitäten (im Zeitverlauf und im internationalen Vergleich)?

In diesem Kapitel wird zunächst ein allgemeiner Überblick über den aktuellen Forschungsstand zum Thema soziale Disparitäten im Bildungssystem gegeben (Kap. 3.1). Im Anschluss werden empirische Befunde zu denjenigen Variablen aus Kapitel 2 präsentiert, die in der Bildungsforschung bislang unterrepräsentiert sind (Kap. 3.2).

3.1. Forschungsstand

Seit Jahrzenten verdeutlichen nationale ebenso wie internationale Ergebnisse der empirischen Bildungsforschung, dass der Bildungserfolg von Kindern in einem engen Zusammenhang mit dem sozialen Status ihrer Herkunftsfamilien steht (Berkemeyer et al., 2017; Eriksson et al., 2021; Hußmann et al., 2017; OECD, 2019; Sirin, 2005; Stubbe, Krieg, et al. 2020; Weis et al., 2019). Die daraus resultierenden Unterschiede hinsichtlich des domänenübergreifenden Kompetenzerwerbs, der Teilhabechancen und damit auch der Gestaltung von Bildungsbiographien wurden ebenfalls breit dokumentiert (Anger & Plünnecke, 2021; Mahler & Kölm, 2019; Müller & Ehmke, 2016; Stubbe, Krieg, et al. 2020; Tophoven et al., 2017; Weis et al., 2019). Maßgeblich für den wissenschaftlichen und öffentlichen Diskurs zu sozialen Ungleichheiten im Bildungssystem sind die Ergebnisse der regelmäßig durchgeführten Schulleistungsstudien, an denen Deutschland seit den späten 1990er-Jahren beteiligt ist: PIRLS/IGLU, TIMSS, PISA, NEPS sowie der Bildungstrend des *Instituts zur Qualitätsentwicklung im Bildungswesen* (IQB).

Für den Primarbereich belegen die Ergebnisse der aktuellen TIMS-Studie, dass die sozialen Disparitäten im Bereich der naturwissenschaftlich-mathematischen Bildung zwischen 2007 und 2019 in Deutschland nicht signifikant verringert werden konnten (Stubbe, Krieg, et al. 2020). Nach wie vor erreichen deutsche Schülerinnen und Schüler aus sozioökonomisch gut gestellten Familien folglich signifikant höhere mathematische und naturwissenschaftliche Kompetenzen als Kinder aus sozioökonomisch benachteiligten Familien. Auch im Fach Deutsch besteht weiterhin eine enge Kopplung zwischen der sozialen Herkunft und der Leseleistung in Deutschland. Im Vergleich zu den Erhebungen von 2001, 2006 und 2011 fielen in IGLU 2016 die Unterschiede in der Lesekompetenz sogar signifikant höher aus. Nur in drei Teilnehmerstaaten (Slowakei, Ungarn und Slowenien) verstärk-

ten sich zwischen 2001 und 2016 die sozialen Disparitäten noch mehr als in Deutschland (Hußmann et al., 2017).

Im Sekundarschulbereich deutet sich im Trend hingegen eine leichte Verringerung der sozialen Unterschiede in der Lesekompetenz an, wie die Ergebnisse der letzten PISA-Erhebung aus dem Jahr 2018 nahelegen. Insbesondere diejenigen Jugendlichen, deren Eltern in Berufen der EGP-Klassen V, VI und VII tätig sind, konnten 2018 ihre Lesekompetenz im Vergleich zur ersten PISA-Erhebung im Jahr 2000 signifikant verbessern. Dennoch waren bei PISA 2018 immer noch 17.2 Prozent der Leistungsunterschiede durch den sozioökonomischen und -kulturellen Status (Prädiktorvariable ESCS[2]) erklärbar (OECD-Durchschnitt: 12.0 %). Unter anderem in Kanada, Norwegen, Finnland und Dänemark konnte hingegen ein geringerer der Anteil der Leistungsdifferenz im Lesen auf den sozioökonomischen und -kulturellen Status zurückgeführt werden (Weis et al., 2019).

Zusätzlich zum internationalen Vergleich im Rahmen von TIMSS, IGLU und PISA ermöglicht der IQB-Bildungstrend Analysen zu sozialen Disparitäten auf Ebene der Länder der Bundesrepublik Deutschland. Als Maß für den Zusammenhang zwischen der domänenspezifischen Kompetenz und dem sozioökonomischen Status der Herkunftsfamilie (HISEI) wird der soziale Gradient (Steigungskoeffizient in der entsprechenden linearen Regression) angeben; wobei höhere Werte eine stärkere Kopplung widerspiegeln. In den Kompetenzbereichen Mathematik und Deutsch (Lesen, Zuhören und Orthografie) am Ende der vierten Jahrgangsstufe haben sich demnach bundesweit zwischen 2011 (bzw. 2016) und 2021 die sozialen Disparitäten signifikant verstärkt (Sachse et al., 2022). Für die Fächer Mathematik, Biologie, Chemie und Physik in der Sekundarstufe I (9. Jahrgangsstufe) liegen aktuell differenzierte Ergebnisse des IQB-Bildungstrends für das Jahr 2018 vor, die Aussagen zum sozialen Gradienten im Ländervergleich zulassen. Demnach variieren 2018 die sozialen Gradienten in den genannten Fächern bundesweit zwischen 34 und 39 Punkten. Das Schlusslicht bildet nominell in allen Domänen Berlin, während sich Brandenburg, Niedersachsen und Thüringen vergleichsweise geringe soziale Disparitäten feststellen lassen (Mahler & Kölm, 2019).

Mit dem Nationalen Bildungspanel (NEPS) liegt eine umfassende längsschnittliche Datenbasis vor, die die Analyse von Bildungsverläufen aus interdisziplinärer Perspektive ermöglicht. Auf Grundlage von NEPS-Daten sind seit Beginn der Studie im Jahr 2009 ebenfalls zahlreiche Ergebnisse zu sozialen Ungleichheiten in den Bildungsverläufen von Kindern, Jugendlichen und Erwachsenen veröffentlicht worden. Bereits im Kleinkind- und Vorschul-

2 Zur Erklärung s. Kapitel 2.2.

alter können soziale Ungleichheitsprozesse in den Fähigkeiten der Kinder beobachtet werden, die sich während ihrer Schulzeit in fächerübergreifenden Kompetenzunterschieden weiter manifestieren und im Schulverlauf weitestgehend stabil bleiben (Attig & Weinert, 2020; Kähler et al., 2021; Linberg et al., 2020; Passaretta et al., 2022; Schneider & Linberg, 2022; Skopek & Passaretta, 2021). Zudem wird der Zusammenhang zwischen Übergangsentscheidungen im Bildungsverlauf und der sozialen Herkunft diskutiert (Esser & Hoenig, 2018; Esser & Seuring, 2020; Lindemann & Gangl, 2019; Lorenz, 2017; Ömerogullari & Gläser-Zikuda, 2021).

Die seit dem Schuljahr 2019/2020 durch die Coronapandemie bedingten Änderungen im Schulalltag (u. a. Schulschließungen, Distanzunterricht) stellen das Bildungssystem vor neue schwerwiegende Herausforderungen; insbesondere im Hinblick auf die potentielle Verstärkung sozialer Disparitäten (Hurrelmann & Dohmen, 2020). Für dieses junge und dynamische Forschungsfeld legen Helm et al. (2022) eine Übersicht zur aktuellen Befundlage zu möglichen Lerneinbußen und steigender Bildungsbenachteiligung während der Pandemie vor (siehe auch Helm et al., 2021 sowie Zierer, 2021). Demnach weisen tatsächlich zahlreiche Studien auf eine Verstärkung divergierender Teilhabechancen zwischen sozial privilegierten und benachteiligten Kindern hin und bestätigen damit die befürchtete Steigerung sozialer Disparitäten (Dietrich et al., 2020; Engzell et al., 2021; Maldonado & De Witte, 2022; Reimer et al., 2021; Rose et al., 2021; Weber et al., 2021a, 2021b; Wößmann et al., 2021). Dietrich et al. (2020) konnten beispielsweise zeigen, dass soziökonomisch benachteiligte Schülerinnen und Schüler während der Schulschließungen im Frühjahr 2020 weniger Zeit fürs Lernen aufbrachten als ihre sozial besser gestellten Klassenkameradinnen und -kameraden. Wößmann et al. (2021) ergänzen, dass sich Kinder aus nicht akademischen Haushalten im Lockdown-bedingten Homeschooling schlechter konzentrieren konnten, weniger effektiv lernten und seltener an Fördermaßnahmen, wie Förder- oder (kostenlosem) Nachhilfeunterricht, teilnahmen. Hinsichtlich möglicher coronabedingter Unterschiede in der Leistungsentwicklung weisen unter anderem Engzell et al. (2021), Rose et al. (2021) und Weber et al. (2021a, 2021b) darauf hin, dass der Lernrückstand sozial benachteiligter Kinder und Jugendlicher substanziell größer ausfiel als der Lernverlust der privilegierten Mitschülerinnen und Mitschüler. Auf Schulebene können Maldonado und De Witte (2020) zeigen, dass an Schulen mit einer sozioökonomisch benachteiligten Schülerklientel größerer Lernverluste zu beobachten waren. In weitaus weniger Arbeiten konnte hingegen keine zusätzliche Bildungsbenachteiligung (z. B. in Form von gestiegenen Leistungsunter-

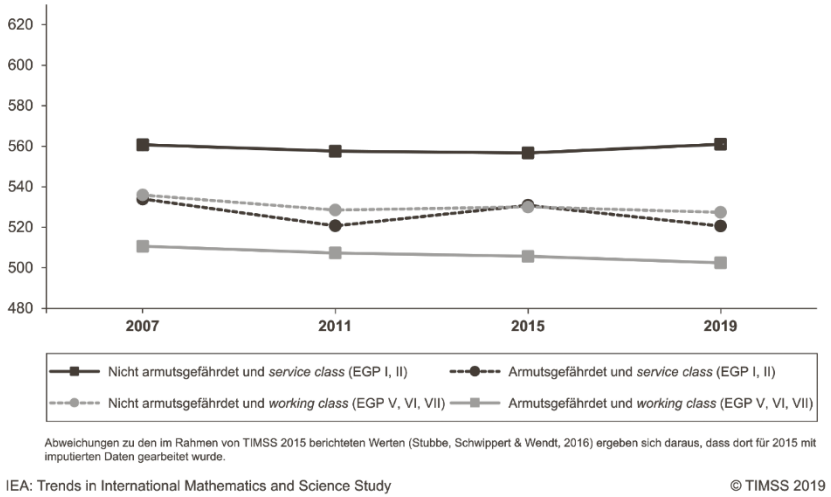
schieden in Abhängigkeit des sozioökomischen Hintergrunds) festgestellt werden (Depping et al., 2021; Ludewig et al., 2022).

Daher verständigte sich die deutsche Kulturministerkonferenz im Februar 2021 darüber, „mit gezielten Fördermaßnahmen in Form von Differenzierung, Intensivierung und … zusätzlichen Angeboten … vorhandene Lernrückstände insbesondere bei benachteiligten Schülerinnen und Schülern auszugleichen“ (KMK, 2021, S. 1). Das nachfolgend vom Bundeskabinett beschlossene Aktionsprogram „Aufholen nach Corona“ (Bundesregierung, 2021) wurde von Helbig et al. (2022) evaluiert. Die Autoren und die Autorin kommen in Hinblick auf die Fokussierung des Aktionsprograms auf die Förderung der besonders benachteiligten Gruppen zu dem Schluss, „dass [bestimmte Programmelemente] – auch innerhalb der einzelnen Schulen – eher nicht den Schüler*innengruppen mit den größten Bedarfen zugutegekommen sind“ (Helbig et al., 2022, S. 273). Als Begründung wird einerseits angeführt, dass die zur Verfügung stehenden Mittel in den meisten Ländern per „Gießkannenprinzip“ verteilt wurden, ohne Leistungs-, Sozial- oder Schulformkriterien (und somit die Bedarfe einzelner Schulen) zu berücksichtigten (ebd.). Zudem setzten verschiedene Fördermaßnahmen (z. B. Lernferien) ein (finanzielles) Engagement der Eltern voraus – wodurch Kinder aus sozial benachteiligten Familien seltener teilnahmen.

3.2. Armutsgefährdung, ökonomisches vs. kulturelles Kapital und soziales Kapital

Stubbe, Krieg, et al. (2020) betrachten die mathematischen Kompetenzen von deutschen Viertklässlerinnen und Viertklässlern in Abhängigkeit von der Armutsgefährdung (siehe Kap. 2.3) der Familie sowie des Berufsstatus der Erziehungsberechtigten (EGP-Klassen) im Trend von TIMSS 2007 bis TIMSS 2019 (siehe Abb. 1). Theoriekonform zeigt sich, dass Kinder aus nicht armutsgefährdeten Familien der sogenannten *service class* (EGP-Klassen I und II) die höchsten und Kinder aus armutsgefährdeten Familien der sogenannten *working class* (EGP-Klassen V, VI, VII) die geringsten Kompetenzen aufweisen.

Abbildung 1: Mittlere mathematische Kompetenzen nach Armutsgefährdung und EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019



Quelle: Stubbe, Krieg, Beese & Jusufi, 2020, Abb. 9.12

Sehr ähnliche Werte im mittleren Kompetenzbereich erreichen diejenigen Schülerinnen und Schüler, die gemessen an einem der beiden Indikatoren einen höheren sozialen Status und nach dem anderen Indikator einen niedrigeren sozialen Status besitzen. Die Analysen zeigen somit, dass die gleichzeitige Betrachtung verschiedener SES-Indikatoren einen zusätzlichen Erkenntnisgewinn bringen kann.

Dies ist beispielsweise auch mit dem von Stubbe und Goy (2013) vorgestellten Instrument (siehe Kap. 2.2) zur empirischen Trennung von ökonomischem und kulturellem Kapital möglich. Erstmals für eine repräsentative Stichprobe wurde die Skala im Rahmen der nationalen Erweiterung von TIMSS 2015 eingesetzt (Stubbe & Lorenz, 2017). Wie auch in der ursprünglichen Studie konnten mit Hilfe von *Latent Class*-Analysen vier ähnlich große Gruppen (23.1 % bis 26.4 %) identifiziert werden, die sich in der Ausprägung des ökonomischen beziehungsweise kulturellen Kapitals unterscheiden (gering/hoch, hoch/gering, gering/gering, hoch/hoch).

Tabelle 1 zeigt die Ergebnisse der durchgeführten Regressionsanalysen zur Erklärung der mathematischen und naturwissenschaftlichen Kompetenzen durch diese Skala (Modelle 1 und 3) sowie durch diese Skala und weitere Indikatoren des sozialen Status (Modelle 2 und 4).

Tabelle 1: Regressionsanalysen zur Erklärung der mathematischen und naturwissenschaftlichen Kompetenzen bei TIMSS 2015 durch unterschiedliche Indikatoren des ökonomischen und kulturellen Kapitals

	Mathematics achievement		Science achievement	
	Model 1	Model 2	Model 3	Model 4
Constant	492.7	490.2	496.5	495.5
Class[1]				
class 1 (only economic capital low)	24.6	10.3	26.3	9.7
class 2 (only cultural capital low)	31.2	17.6	34.5	18.4
class 4 (economic and cultural capital high)	62.5	31.0	69.1	33.0
Parent's highest educational degree[2]				
lower secondary degree		– 12.9		– 18.1
university degree		*n.s.*		*n.s.*
EGP class[3]				
service class		*n.s.*		*n.s.*
working class		*n.s.*		*n.s.*
Highest ISEI		0.5		0.5
Poverty risk		– 14.8		– 16.0
Number of books at home[4]				
at most 10 books		– 29.0		– 36.4
more than 200 books		17.2		16.8
R^2	0.11	0.22	0.12	0.25

All coefficients are unstandardized and statistically significant ($p < .01$).
[1] reference group: class 3 (economic and cultural capital low)
[2] reference group: no university degree but higher than lower secondary degree
[3] reference group: intermediate class
[4] reference group: 11–200 books

Quelle: Stubbe & Lorenz, 2017

Als Referenzgruppe wurden diejenigen Schülerinnen und Schüler ausgewählt, bei denen beide Kapitalformen gering ausgeprägt sind. Werden keine weiteren Variablen berücksichtigt, erreichen Viertklässlerinnen und Viertklässler, in deren Familien nur das ökonomische Kapital gering ausgeprägt ist, eine um 24.6 Punkte höhere Mathematikkompetenz.[3] Für die Gruppe, in der nur das kulturelle Kapital gering ausgeprägt ist, beträgt dieser Wert 31.2. Einen Leistungsvorsprung von 62.5 Punkten – und damit von deutlich mehr als einem Lernjahr (Wendt et al., 2017) – weisen Kinder auf, bei denen zu Hause sowohl ökonomisches als auch kulturelles Kapital in hohem Umfang vorhanden sind. Das Modell 1 erklärt immerhin 11 Prozent der Varianz der abhängigen Variable.

3 Die Ergebnisse für die naturwissenschaftliche Kompetenz fallen ähnlich aus.

Werden weitere Indikatoren des sozialen Status in den Regressionsanalysen berücksichtigt, zeigt sich, dass die Leistungsunterschiede zwischen den vier beschriebenen Gruppen zwar erwartungskonform geringer ausfallen aber weiterhin signifikant sind.

Wie ausführlich dargestellt, stehen in der empirischen Forschung zahlreiche Indikatoren für das ökonomische und kulturelle Kapital zur Verfügung, von denen im Rahmen von *Large Scale Assessment* (insbesondere in den nationalen Ergänzungen) auch regelmäßig mehrere zum Einsatz kommen, während hinsichtlich des sozialen Kapitals bestenfalls nach den innerfamilialen Beziehungen gefragt wird. Im Rahmen von TIMSS 2015 wurde erstmals die von Stubbe und Lorenz (2017) überarbeitete Fassung des Ressourcengenerators für eine repräsentative Stichprobe eingesetzt. Die Analysen haben gezeigt, dass sechs Subskalen dieses Konstruktes unterschieden werden können, je nachdem welche Eigenschaften die Personen im sozialen Netzwerk der Befragten aufweisen (siehe Tab. 2). Im Folgenden werden Analysen mit allen Subskalen präsentiert, wobei insbesondere bei der dritten Dimension (Connections/Beziehungen) einschränkend festgehalten werden muss, dass die Reliabilität unbefriedigend ist.

Tabelle 2: Subskalen des Ressourcengenerators (TIMSS 2015)

Dimension of the construct 'social capital'	Description	No. of items	M	SD	Example item	Cronbach's Alpha
Economic capital	Persons with high economic capital are part of the personal network	6	0.53	0.31	someone who owns shares for at least 10,000 Euros	0.72
Cultural capital	Persons with high cultural capital are part of the personal network	7	0.65	0.27	someone who reads scientific papers or books on a regular basis	0.71
Connections	Persons that have connections to the public sector are part of the personal network	3	0.32	0.33	someone who is active in a political party	0.52
Advice	Persons that can offer advice on different topics are part of the personal network	7	0.65	0.31	someone who can give advice on matters of law	0.77
Everyday support	Persons that can offer support for everday tasks are part of the personal network	6	0.85	0.21	someone who can babysit for your children	0.66
Manual or technical support	Persons that can offer support for manual or technical tasks are part of the personal network	5	0.85	0.22	someone who can repair your car, bike, etc.	0.61

Quelle: Stubbe & Lorenz, 2017

Tabelle 3 zeigt die Korrelationen der sechs Subskalen untereinander, die überwiegend in einem mittleren Bereich liegen. Zudem zeigt sich, dass fünf Subskalen signifikant mit den Kompetenzen der Viertklässlerinnen und

Viertklässler korrelieren. Am größten ist dieser Zusammenhang für die Dimension, die angibt, ob es im sozialen Netzwerk Personen mit hohem kulturellen Kapital gibt. Kein signifikanter Zusammenhang besteht für die Dimension, die angibt, ob es im sozialen Netzwerk Personen mit Fähigkeiten im handwerklichen oder technischen Bereich gibt.

Tabelle 3: Korrelationen zwischen den sechs Subskalen des Ressourcengenerators sowie deren Korrelationen mit den mathematischen und naturwissenschaftlichen Kompetenzen im Rahmen von TIMSS 2015

	Dimension					
	1	2	3	4	5	6
Dimension 1: Economic capital						
Dimension 2: Cultural capital	0.52					
Dimension 3: Connections	0.37	0.34				
Dimension 4: Advice	0.45	0.46	0.35			
Dimension 5: Everyday support	0.31	0.27	0.17	0.37		
Dimension 6: Manual or technical support	0.20	0.17	0.17	0.40	0.35	
Mathematics achievement	0.18	0.24	0.09	0.07	0.14	*n.s.*
Science achievement	0.19	0.26	0.11	0.09	0.16	*n.s.*

All correlations are statistically significant ($p < .05$).

Quelle: Stubbe & Lorenz, 2017

Schließlich zeigt Tabelle 4 die Ergebnisse der durchgeführten Regressionsanalysen zur Erklärung der mathematischen und naturwissenschaftlichen Kompetenzen durch die sechs Dimensionen des Ressourcengenerators (Modelle 1 und 3) sowie durch dieses Konstrukt und weitere Indikatoren des sozialen Status (Modelle 2 und 4).

Wie sich bereits bei den Korrelationsanalysen angedeutet hat, weist die zweite Dimension (kulturelles Kapital) den engsten Zusammenhang mit den erreichten Kompetenzen auf. Aber auch die Dimensionen 5 (Unterstützung im Alltag) und 1 (ökonomisches Kapital) zeigen einen signifikanten positiven Zusammenhang mit den abhängigen Variablen. Hingegen lässt sich kein signifikanter Effekt der Dimension 3 (Beziehungen) und 6 (handwerkliche oder technische Unterstützung) feststellen. Abweichend von den Befunden der bivariaten Analysen weist die vierte Dimension (Ratschläge) in diesem multivariaten Modell einen negativen Zusammenhang mit den Kompetenzen der Schülerinnen und Schüler auf. Unter Kontrolle der übrigen Facetten des sozialen Kapitals geht der Bedarf an Ratschlägen also mit geringeren schulischen Leistungen einher und wirkt möglicherweise kompensatorisch.

Tabelle 4: Regressionsanalysen zur Erklärung der mathematischen und naturwissenschaftlichen Kompetenzen bei TIMSS 2015 durch die sechs Dimensionen des Ressourcengenerators und weitere Indikatoren des sozialen Status

	Mathematics achievement		Science achievement	
	Model 1	Model 2	Model 3	Model 4
Constant	473.6	486.2	467.8	483.6
Ressource generator				
Dimension 1: Economic capital	17.8	*n.s.*	15.9	*n.s.*
Dimension 2: Cultural capital	52.7	19.7	59.0	21.9
Dimension 3: Connections	*n.s.*	*n.s.*	*n.s.*	*n.s.*
Dimension 4: Advice	−19.4	− 13.0	−18.8	− 11.4
Dimension 5: Everyday support	29.4	13.4	37.7	19.5
Dimension 6: Manual or technical support	*n.s.*	*n.s.*	*n.s.*	*n.s.*
Parent's highest educational degree[1]				
lower secondary degree		− 15.1		− 20.2
university degree		*n.s.*		*n.s.*
EGP class[2]				
service class		9.7		8.8
working class		*n.s.*		*n.s.*
Highest ISEI		0.4		0.5
Poverty risk		− 17.2		− 18.2
Number of books at home[3]				
at most 10 books		− 30.7		− 38.1
more than 200 books		19.6		19.3
R^2	0.08	0.20	0.09	0.23

All coefficients unstandardized and statistically significant ($p < .01$).
[1] reference group: no university degree but higher than lower secondary degree
[2] reference group: intermediate class [3] reference group: 11–200 books

Quelle: Stubbe & Lorenz, 2017

Auch in den Modellen, die Indikatoren für das ökonomische und kulturelle Kapital der Familien berücksichtigen, weisen drei der sechs Dimensionen des Ressourcengenerators signifikante Effekte auf. Somit bleibt festzuhalten, dass die adäquate Operationalisierung des sozialen Kapitals für die empirische Bildungsforschung einen Mehrwert darstellen kann.

4. Wie lassen sich soziale Disparitäten erklären?

So wie die Entwicklung von Instrumenten zur Messung von SES notwendige Bedingung für das Beschreiben von sozialen Disparitäten ist, setzt das Erklären von sozialen Disparitäten deren Beschreibung voraus. Zunächst muss das Vorhandensein von sozialen Disparitäten empirisch festgestellt und hinsichtlich Differenzen im Zeitverlauf oder kontextualer Merkmale in

verschiedenen Staaten, Regionen, Personengruppen etc. analysiert werden. Aus wissenschaftlicher Perspektive stellt sich im Anschluss daran konsequenterweise die Frage nach den Ursachen dieser Disparitäten.

Ausgehend von Boudon (1974) wird in der Bildungsforschung bei der Erklärung von sozialen Disparitäten üblicherweise zwischen primären und sekundären Herkunftseffekten unterschieden. Als *primäre Herkunftseffekte* werden demnach die direkten Effekte des SES auf die schulischen Kompetenzen von Schülerinnen und Schülern bezeichnet. Im Gegensatz dazu wird unter *sekundären Herkunftseffekten* der Zusammenhang zwischen SES und Bildungsentscheidungen (unter Kontrolle der primären Herkunftseffekte) verstanden. In Kapitel 4.1 werden mit der *Home Literacy Environment* (HLE) und dem bezahlten Nachhilfeunterricht exemplarisch zwei Erklärungsansätze von primären Herkunftseffekten vorgestellt. Kapitel 4.2 beleuchtet anschließend die Frage von Bildungsentscheidungen und damit die sekundären Herkunftseffekte im deutschen Bildungssystem.

4.1. Primäre Herkunftseffekte: Home Literacy Environment und Nachhilfe

Nach Bourdieu (1983) wird das (ökonomische, kulturelle und soziale) Kapital, das in einer Familie vorhanden ist, unter anderem dazu genutzt, um den Bildungserfolg der Kinder – und damit deren zukünftige soziale Position – zu sichern. Kinder aus bildungsnahen Elternhäusern wachsen folglich in einem anregungsreicheren Umfeld auf, in dem beispielsweise gezielt Aktivitäten stattfinden, die das Lernen fördern (z. B. Vorlesen). Es liegt somit nahe, familiale Maßnahmen der häuslichen Sozialisation als Mediator zwischen SES und Bildungserfolg anzunehmen (Attig & Weinert, 2020).

Die internationalen PIRLS-Datensätze bieten für entsprechende Analysen eine hervorragende Grundlage, da neben der Lesekompetenz der Schülerinnen und Schüler und dem SES der Familien über den Elternfragebogen auch Angaben zur häuslichen Lesesozialisation erhoben wurden. Die Lesesozialisation im Elternhaus stellt einen Teilbereich der sogenannten *Home Literacy Environment* (HLE) dar, also dem Unterstützungsverhalten der Eltern in bildungsbezogenen häuslichen Aktivitäten.

Entsprechend der theoretischen Struktur (McElvany, 2008) lassen sich bei PIRLS 2006 die Teilbereiche leseförderliche Aktivitäten vor der Grundschulzeit, leseförderliche Aktivitäten während der Grundschulzeit, Vorbildverhalten der Eltern und leseförderliche Ressourcen im Elternhaus unterscheiden. Insgesamt stehen 25 Einzelitems zur Verfügung, die eine interne

Konsistenz von *Cronbachs Alpha* = .86 aufweisen. Stubbe, Buddeberg, et al. (2007) konstruieren aus diesen Items einen eindimensionalen Raschscore, der das Ausmaß der häuslichen Leseförderung quantifiziert.

Es kann gezeigt werden, dass sowohl zwischen SES und HLE als auch zwischen HLE und Lesekompetenz ein Zusammenhang besteht (Buddeberg et al., 2008; Stubbe & Buddeberg, 2008; Stubbe, Buddeberg, et al. 2007).

Daran anknüpfend formulieren Tarelli und Stubbe (2010) ein Pfadmodell und schätzen dessen Parameter getrennt für die 29 PIRLS-Staaten, die der EU beziehungsweise der OECD angehören. Zwar zeigt sich theoriekonform in allen Staaten ein indirekter Zusammenhang zwischen SES und Lesekompetenz über HLE; der direkte Pfad von SES zur Lesekompetenz bliebt allerdings dominant (Tarelli & Stubbe, 2010, Table 2). HLE erklärt also einen gewissen Anteil der sozialen Disparitäten, ein substanzieller Anteil bleibt jedoch unerklärt.

Für die dargestellte Form der häuslichen Unterstützung wird vor allem kulturelles Kapital benötigt. Eine Möglichkeit, ökonomisches Kapital zur Steigerung des Bildungserfolges der eigenen Kinder zu nutzen, besteht in bezahlter Nachhilfe. Die empirischen Befunde zum Zusammenhang mit dem sozialen Status sind allerdings uneinheitlich. Während unter anderem Schneider (2005) sowie Hille et al. (2016) auf Basis von Daten des Sozio-oekonomischen Panels (SOEP) herausstellen, dass die Inanspruchnahme von Nachhilfeunterricht mit dem Einkommen der Eltern assoziiert ist, zeigen Lorenz und Stubbe (2020) anhand von NEPS-Daten, dass unter Kontrolle des häuslichen Unterstützungsverhaltens der Eltern kein signifikanter Zusammenhang zwischen dem sozialen Status und der Entscheidung für bezahlte Nachhilfe besteht.

4.2. Sekundäre Herkunftseffekte: Bildungsentscheidungen

Die Analyse von sekundären Herkunftseffekten ist ganz besonders für gegliederte Schulsysteme von Bedeutung, da dort zu festgelegten Zeitpunkten Entscheidungen für bestimmte Bildungsgänge getroffen werden müssen (Skopek & Passaretta, 2021). Entsprechend existieren für die Bundesrepublik zahlreiche Publikationen zu diesem Thema – insbesondere zum Übergang von der Grundschule auf eine weiterführende Schule (Ditton et al., 2017; Dumont et al., 2014; Esser & Seuring, 2020; Stubbe, Kasper, et al. 2020; Stubbe, Schaufelberger, et al. 2023; Zunker & Neumann, 2020).

In praktisch allen Studien zu diesem Thema zeigt sich, dass Schullaufbahnentscheidungen auch unter Kontrolle der schulischen Kompetenzen

einen Zusammenhang mit dem sozialen Status von Schülerfamilien aufweisen. Das heißt, Schülerinnen und Schüler aus den unteren sozialen Lagen haben auch bei gleichen Testleistungen eine geringere Chance, ein Gymnasium zu besuchen, als ihre Mitschülerinnen und Mitschüler aus den oberen sozialen Lagen (Stubbe, 2009b).

Unter Berücksichtigung von Beckers (2000) Forderung, den Einfluss von Lehrkräfteempfehlungen auf elterliche Entscheidungen angemessen zu berücksichtigen, analysieren Pietsch und Stubbe (2007) die Daten zu den Schullaufbahnentscheidungen aus IGLU 2001 unter besonderer Berücksichtigung der Schullaufbahnempfehlungen der Grundschulen und betonen die dreifache soziale Benachteiligung im deutschen Bildungssystem: Kinder aus sozial benachteiligten Familien (1) verfügen über geringere schulische Kompetenzen; (2) haben – im Vergleich zu Mitschülerinnen und Mitschülern mit ähnlichen Leistungen – geringere Chancen von ihrer Grundschule eine Gymnasialempfehlung zu erhalten und (3) haben selbst mit einer Gymnasialempfehlung eine geringere Chance von ihren Eltern tatsächlich auf einem Gymnasium angemeldet zu werden.

Stubbe und Bos (2008) veranschaulichen – ebenfalls mit den Daten aus IGLU 2001 – den Zusammenhang unterschiedlicher Hintergrundvariablen mit der elterlichen Schullaufbahnentscheidung mit Hilfe von Pfadmodellen. Dabei werden neben der Schullaufbahnempfehlung der Lehrkräfte auch die Schulnoten als Mediatoren zwischen den Hintergrundvariablen und der Entscheidung berücksichtigt. Das finale Modell, das von Arnold et al. (2007) für IGLU 2006 und von Stubbe et al. (2012) für IGLU 2011 reproduziert wurde, zeigt, dass auch unter Berücksichtigung weiterer leistungsrelevanter Merkmale (z. B. Anstrengungsbereitschaft) sowohl ein direkter als auch ein indirekter Pfad vom SES zur Schullaufbahnentscheidung beobachtet werden kann.

Stubbe (2009a, 2009b) überträgt die für den Grundschulübergang bewährten Modelle auf Schulformwechselentscheidungen in den ersten Jahren der Sekundarstufe I und stellt auch dort unter Kontrolle der schulischen Leistungen eine Koppelung zwischen SES und Bildungsentscheidungen fest. Daran anknüpfend untersucht Lorenz (2017) den Übergang am Ende der Sekundarstufe I (Entscheidung für oder gegen den Besuch der gymnasialen Oberstufe) und stellt insgesamt ebenfalls ausgeprägte sekundäre Herkunftseffekte fest. Unter Kontrolle der Schulform zeigt sich allerdings, dass dieser Effekt an integrierten Gesamtschulen nicht feststellbar ist, woraus die Autorin ableitet, dass ein eingliedriges Schulsystem einen Beitrag zur Verringerung der sozialen Disparitäten im deutschen Bildungssystem leisten kann.

5. Zusammenfassung und Ausblick

In diesem Beitrag wurde deutlich, dass die Analyse sozialer Disparitäten im Bildungswesen nach wie vor ein hochrelevantes Thema der empirischen Bildungsforschung ist. Aktuelle Herausforderungen, wie die coronabedingten Schulschließungen, aber auch Schulentwicklungsprozesse, wie die Implementation von Profilklassen, können bereits bestehende Mechanismen sozialer Bildungsbenachteiligung weiter verstärken (Engzell et al., 2021; Weber et al., 2022b; Zunker & Neumann, 2020).

Ein besonderer Fokus dieses Beitrags liegt auf der Frage nach einer angemessenen Operationalisierung des sozialen Status. Demnach sollte bei der Gestaltung von Fragebögen in der empirischen Bildungsforschung bezüglich der Auswahl von Variablen zur Erfassung des SES berücksichtigt werden, wer befragt wird und wie viel Bearbeitungszeit für diese Merkmale zur Verfügung steht. Dabei muss aber auch bedacht werden, welche Analysen mit den erhobenen Daten durchgeführt werden sollen. Werden nur Kinder befragt und genügt ein grober Indikator, ist die Frage nach der Anzahl der Bücher im Haushalt weiterhin die erste Wahl. Werden Eltern befragt, wobei der Fragebogen sehr kurz sein soll, bietet es sich an, nach Bildungsabschlüssen und gegebenenfalls nach Berufsabschlüssen zu fragen. Die Ausführungen in diesem Beitrag haben aber auch deutlich gemacht, dass es für differenzierte Analysen der sozialen Disparitäten im Bildungssystem notwendig ist, verschiedene Facetten dieses komplexen theoretischen Konstrukts zu erfassen: Bei der Operationalisierung des ökonomischen Kapitals sollte also explizit die Armutsgefährdung von Haushalten berücksichtigt werden. Bei der Operationalisierung des sozialen Kapitals ist darauf zu achten, nicht nur die Beziehungen innerhalb der Familien zu erfassen (bspw. mit Hilfe des Ressourcengenerators). Und wenn der Vergleich der unterschiedlichen Kapitalarten vorgesehen ist, muss bedacht werden, dass sich kulturelles und ökonomisches Kapital unterscheiden lassen, da viele Indikatoren (beispielsweise Bildungs- und Berufsabschlüsse) beides erfassen.

Neben deskriptiven Befunden zum Ausmaß der sozialen Disparitäten wurden in diesem Beitrag auch verschiedene Ansätze zur Erklärung dieser Bildungsungleichheiten diskutiert. Erst durch die Identifizierung der Mechanismen, die für ungleiche Bildungschancen sorgen, lassen sich bildungspolitische und schulpraktische Handlungsoptionen für mehr Chancengleichheit ableiten.

Literatur

Anger, C. & Plünnecke, A. (2021). *Bildungsgerechtigkeit. Herausforderungen für das deutsche Bildungssystem* (IW-Analysen Nr. 140). Köln: Institut der deutschen Wirtschaft (IW). https://www.econstor.eu/handle/10419/228912

Arnold, K.-H., Bos, W., Richert, P. & Stubbe, T. C. Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 271–297). https://publications.goettingen-research-online.de/handle/2/98351

Attig, M. & Weinert, S. (2020). What Impacts Early Language Skills? Effects of Social Disparities and Different Process Characteristics of the Home Learning Environment in the First 2 Years. *Frontiers in psychology*, *11*, Artikel 557751. https://doi.org/10.3389/fpsyg.2020.557751

Bardone, L. & Guio, A.-C. (2005). In-work poverty. New commonly agreed indicators at the EU level. *Statistics in focus, 5/2005*.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (2000). *Soziale Bedingungen von Schulleistungen: Zur Erfassung von Kontextmerkmalen durch Schüler-, Schul- und Elternfragebögen*. Max-Planck-Institut für Bildungsforschung. https://hdl.handle.net/11858/00-001M-0000-0025-A756-7

Becker, R. (2000). Klassenlage und Bildungsentscheidungen. Eine empirische Anwendung der Wert-Erwartungstheorie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *52*(3), 450–474. https://doi.org/10.1007/s11577-000-0068-9

Becker, R. (2017). Entstehung und Reproduktion dauerhafter Bildungsungleichheiten. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (S. 89–150). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15272-7_4

Berkemeyer, N., Bos, W., Hermstein, B., Abendroth, S. & Semper, I. (2017). *Chancenspiegel – eine Zwischenbilanz. Zur Chancengerechtigkeit und Leistungsfähigkeit der deutschen Schulsysteme seit 2002*. Verlag Bertelsmann Stiftung.

Blossfeld, H.-P., Blossfeld, G. J. & Blossfeld, P. N. (2019). Soziale Ungleichheiten und Bildungsentscheidungen im Lebensverlauf. Die Perspektive der Bildungssoziologie. *Journal for Educational Research Online*, *11*(1), 16–30. https://doi.org/10.25656/01:16785

Bos, W., Stubbe, T. C. & Buddeberg, M. (2010). Gibt es eine armutsbedingte Bildungsbenachteiligung? Die Operationalisierung verschiedener Indikatoren der sozialen Herkunft in der empirischen Bildungsforschung. In D. H. Rost (Hrsg.), *Intelligenz, Hochbegabung, Vorschulerziehung, Bildungsbenachteiligung* (S. 165–208). Waxmann.

Boudon, R. (1974). *Education, opportunity, and social inequality. Changing Prospects in Western Society*. Wiley.

Bourdieu, P. (2012). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In U. Bauer, U. H. Bittlingmayer & A. Scherr (Hrsg.), *Handbuch Bildungs- und Erziehungssoziologie* (S. 229–242). https://doi.org/10.1007/978-3-531-18944-4_15

Buddeberg, I., Stubbe, T. C., & Potthoff, B. (2008). Lesesozialisation im Elternhaus in den Ländern der Bundesrepublik Deutschland. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU-E 2006. Die Länder des Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 127–141). Waxmann.

Buddeberg, M., Dohe, C., Freiberg, M. & Reding, P. (2009). Sozialer Hintergrund und Leseleistungen – Die unterschiedlichen Schulergebnisse im Vergleich. In C. Berg, W. Bos, S. Hornberg, P. Kühn, R. Martin, P. Reding, T. C. Stubbe & R. Valtin (Hrsg.), *LESE-LUX – Lesekompetenzen Luxemburger Schüler und Schülerinnen auf dem Prüfstand. PIRLS-Zusatzstudie Deutsch/Französisch 2008* (S. 151–166). Waxmann.

Bundesministerium für Arbeit und Soziales. (2001). *Lebenslagen in Deutschland. Der erste Armuts- und Reichtumsbericht der Bundesregierung.* Bundesregierung. https://www.armuts-und-reichtumsbericht.de/SharedDocs/Downloads/Berichte/lebenslagen-erster-armuts-reichtumsbericht.pdf?__blob=publicationFile&v=2

Bundesministerium für Arbeit und Soziales. (2021). *Lebenslagen in Deutschland. Der Sechste Armuts- und Reichtumsbericht der Bundesregierung.* Bundesregierung. https://www.armuts-und-reichtumsbericht.de/SharedDocs/Downloads/Berichte/sechster-armuts-reichtumsbericht.pdf?__blob=publicationFile&v=6

Bundesregierung. (2021). *Eckpunkte zu Aktionsprogramm „Aufholen nach Corona*". Bundesregierung. https://www.bundesregierung.de/breg-de/aktuelles/programm-aufholen-nach-corona-1897750

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, 95–120.

Depping, D., Lücken, M., Musekamp, F. & Thonke, F. (2021). Kompetenzstände Hamburger Schüler*innen vor und während der Corona-Pandemie. In D. Fickermann & B. Edelstein (Hrsg.), *Schule während der Corona-Pandemie. Neue Ergebnisse und Überblick über ein dynamisches Forschungsfeld* (S. 51–79). Waxmann. https://doi.org/10.31244/9783830993315.03

Dietrich, H., Patzina, A. & Lerche, A. (2020). Social inequality in the homeschooling efforts of german high school students during a school closing period. *European Societies*, *23*(1), 348–369. https://doi.org/10.1080/14616696.2020.1826556

Ditton, H., Elsäßer, S., Gölz, N., Stahn, V. & Wohlkinger, F. (2017). Soziale Disparitäten im Bildungsverlauf. Schulische Laufbahnen von der 2. Bis zur 7. Jahrgangsstufe. In H. G. Holtappels (Hrsg.), *Entwicklung und Qualität des Schulsystems. Neue empirische Befunde und Entwicklungstendenzen* (S. 143–171). Waxmann.

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. https://doi.org/10.1007/s11618-013-0466-1

Engzell, P. (2021). What Do Books in the Home Proxy For? A Cautionary Tale. *Sociological Methods & Research*, *50*(4), 1487–1514. https://doi.org/10.1177/0049124119826143

Engzell, P., Frey, A. & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(17), Artikel e2022376118. https://doi.org/10.1073/pnas.2022376118

Eriksson, K., Lindvall, J., Helenius, O. & Ryve, A. (2021). Socioeconomic status as a multidimensional predictor of student achievement in 77 societies. *Frontiers in Education*, *6*, Artikel 731634. https://doi.org/10.3389/feduc.2021.731634

Erikson, R., Goldthorpe, J. H. & Portocarero, L. (1979). Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden. *The British Journal of Sociology*, *30*(4), 415–441. https://doi.org/10.2307/589632

Esser, H. & Hoenig, K. (2018). Leistungsgerechtigkeit und Bildungsungleichheit Effekte der Verbindlichkeit der Grundschulempfehlungen beim Übergang auf das Gymnasium. Ein Vergleich der deutschen Bundesländer mit den Daten der „National Educational Panel Study" (NEPS). *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *70*(3), 419–447. https://doi.org/10.1007/s11577-018-0558-2

Esser, H. & Seuring, J. (2020). Kognitive Homogenisierung, schulische Leistungen und soziale Bildungsungleichheit. Theoretische Modellierung und empirische Analyse der Effekte einer strikten Differenzierung nach den kognitiven Fähigkeiten auf die Leistungen in der Sekundarstufe und den Einfluss der sozialen Herkunft in den deutschen Bundesländern mit den Daten der „National Educational Panel Study" (NEPS). *Zeitschrift für Soziologie*, *49*(5-6), 277–301. https://doi.org/10.1515/zfsoz-2020-0025

Ganzeboom, H. B., Graaf, P. M. de & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*(1), 1–56. https://doi.org/10.1016/0049-089x(92)90017-b

Ganzeboom, H. B. G. & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research*, *25*(3), 201–239. https://doi.org/10.1006/ssre.1996.0010

Ganzeboom, H. B. G. & Treiman, D. J. (2003). Three internationally standardized measures for comparative research on occupational status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in cross-national comparison. A European working book for demographic and socio-economic variables* (S. 159–193). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9186-7_9

Ganzeboom, H. B. G. & Treiman, D. (2010). *Occupational status measures for the new International Standard Classification of Occupations ISCO-08; With a discussion of the new classification*. Verfügbar unter: http://www.harryganzeboom.nl/isol/isol2010c2-ganzeboom.pdf

Grundgesetz für die Bundesrepublik Deutschland. (1949). https://www.gesetze-im-internet.de/gg/BJNR000010949.html

Helbig, M., Edelstein, B., Fickermann, D. & Zink, C. (2022). *Aufholen nach Corona? Maßnahmen der Länder im Kontext des Aktionsprogramms von Bund und Ländern*. Waxmann. https://doi.org/10.31244/9783830996033

Helm, C., Huber, S. G. & Loisinger, T. (2021). Was wissen wir über schulische Lehr-Lern-Prozesse im Distanzunterricht während der Corona-Pandemie? – Evidenz aus Deutschland, Österreich und der Schweiz. *Zeitschrift für Erziehungswissenschaft: ZfE*, *24*(2), 237–311. https://doi.org/10.1007/s11618-021-01000-z

Helm, C., Huber, S. G. & Postlbauer, A. (2022). Lerneinbußen und Bildungsbenachteiligung durch Schulschließungen während der Covid-19-Pandemie im Frühjahr 2020. Eine Übersicht zur aktuellen Befundlage. In D. Fickermann, B. Edelstein, J. Gerick & K. Rachenbäumer (Hrsg.), *Schule und Schulpolitik während der Corona-Pandemie: Nichts gelernt?* (S. 59–81). Waxmann. https://doi.org/10.31244/9783830994589.03

Heppt, B., Olczyk, M., & Volodina, A. (2022). Number of books at home as an indicator of socioeconomic status: Examining its extensions and their incremental validity for student achievement. *Social Psychology of Education*, *25*(4), 903–928. https://doi.org/10.1007/s11218-022-09704-8

Hille, A., Spieß, C. K., & Staneva, M. (2016). Immer mehr Schülerinnen und Schüler nehmen Nachhilfe, besonders in Haushalten mit mittleren Einkommen. *DIW Wochenbericht*, *83*(6), 111–121.

Hoffmann, E. (2003). International statistical comparisons of occupational and social structures. Problems, possibilities and the role of ISCO-88. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in cross-national comparison. A European working book for demographic and socio-economic variables* (S. 137–158). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9186-7_8

Hoffmeyer-Zlotnik, J. H. P. (2003). The classification of education as a sociological background characteristic. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in cross-national comparison. A European working book for demographic and socio-economic variables* (S. 245–256). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9186-7_12

Hurrelmann, K., & Dohmen, D. (2020). Die Krise schwächt die Schwachen. Warum die Corona-Pandemie die Bildungsungleichheit verstärkt. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, *40*(3), 313–320.

Hußmann, A., Stubbe, T. C., & Kasper, D. (2017). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195–2017). Waxmann.

Jansen, D. (2006). *Einführung in die Netzwerkanalyse: Grundlagen, Methoden, Forschungsbeispiele*. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-09875-1

Kähler, J., Hahn, I. & Köller, O. (2021). Naturwissenschaftliche Kompetenz in der Grundschule: Effekte von Familienmerkmalen und Klassenkomposition. *Zeitschrift für Pädagogische Psychologie*, 1–15. https://doi.org/10.1024/1010-0652/a000308

Kultusministerkonferenz (KMK). (2021). *Schrittweise Wiederaufnahme des Schulbetriebs ab 15.02.2021.* Beschluss der KMK vom 08.02.2021. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_02_08-Wiederaufnahme-des-Schulbetriebs.pdf

Lin, N., & Dumin, M. (1986). Access to occupations through social ties. *Social Networks*, *8*(4), 365–385. https://doi.org/10.1016/0378-8733(86)90003-1

Linberg, A., Attig, M. & Weinert, S. (2020). Social disparities in the vocabulary of 2-year-old children and the mediating effect of language-stimulating interaction behavior. *Journal for Educational Research Online*, *12*(2), 12–35. https://doi.org/10.25656/01:20971

Lindemann, K. & Gangl, M. (2019). Parental Unemployment and the Transition to Vocational Training in Germany: Interaction of Household and Regional Sources of Disadvantage. *European Sociological Review*, *35*(5), 684–700. https://doi.org/10.1093/esr/jcz027

Lorenz, J. (2017). *Soziale Chancengerechtigkeit durch Gesamtschulen. Können Gesamtschule dazu beitragen sekundäre Herkunftseffekte am Übergang nach der Sekundarstufe I zu reduzieren?* http://dx.doi.org/10.53846/goediss-6187

Lorenz, J., & Stubbe, T. C. (2020). Private tutoring as a means for maintaining social status. *Journal for Educational Research Online*, *12*(2), 89–113.

Lorenz, J., & Stubbe, T. C. (2021). Soziale Integration in der Schule. Anwendungsmöglichkeiten der sozialen Netzwerkanalyse in der empirischen Bildungsforschung. In M. Haas & M. Krieg (Hrsg.), *Forschungsbefunde zu Innovationen in Schule und Universität* (S. 75–94). Universitätsverlag Göttingen.

Ludewig, U., Kleinkorres, R., Schaufelberger, R., Schlitter, T., Lorenz, R., König, C., Frey, A. & McElvany, N. (2022). COVID-19 Pandemic and Student Reading Achievement: Findings from a School Panel Study. *Frontiers in Psychology*, *13*, Artikel 876485. https://doi.org/10.3389/fpsyg.2022.876485

Mahler, N. & Kölm, J. (2019). Soziale Disparitäten. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 265–293). Waxmann.

Maldonado, J. E. & De Witte, K. (2020). The effect of school closures on standardised student test outcomes. *British Educational Research Journal*, *48*(1), 49–94. https://doi.org/10.1002/berj.3754

Mang, J., Seidl, L., Schiepe-Tiska, A., Tupac-Yupanqui, A., Ziernwald, L., Doroganova, A., Weis, M., Diedrich, J., Heine, J.-H., Gonzaléz-Rodríguez, E. & Reiss, K. (2021). *PISA 2018 Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Waxmann.

McElvany, N. (2008). *Förderung von Lesekompetenz im Kontext der Familie*. Waxmann.

Mikus, K., Tieben, N. & Schober, P. S. (2020). Children's conversion of cultural capital into educational success: The symbolic and skill-generating functions of cultural capital. *British Journal of Sociology of Education*, *41*(2), 197–217. https://doi.org/10.1080/01425692.2019.1677454

Müller, K. & Ehmke, T. (2016). Soziale Herkunft und Kompetenzerwerb. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 285–316). Waxmann.

OECD (2019). *PISA 2018 Results (Volume II): Where all students can succeed*. PISA, OECD Publishing. https://doi.org/10.1787/b5fd1b8f-en

Ömeroğullari, M. & Gläser-Zikuda, M. (2021). Entwicklung affektiv-motivationaler Merkmale am Übergang in die Sekundarstufe – Wie ergeht es bildungsbenachteiligten Kindern? *Zeitschrift für Bildungsforschung*, *11*(2), 363–383. https://doi.org/10.1007/s35834-021-00314-6

Passaretta, G., Skopek, J. & van Huizen, T. (2022). Is Social Inequality in School-Age Achievement Generated before or during Schooling? A European Perspective. *European Sociological Review*, *38*(6), 849–865. https://doi.org/10.1093/esr/jcac005

Pietsch, M., Bonsen, M., & Bos, W. (2006). Ein Index sozialer Belastung als Grundlage für die Rückmeldung ‚fairer Vergleiche' von Grundschulen in Hamburg. In W. Bos & M. Pietsch (Hrsg.), *KESS 4 – Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 4 in Hamburger Grundschulen* (S. 225–245). Waxmann.

Pietsch, M. & Stubbe, T. C. (2007). Inequality in the transition from primary to secondary school: School choices and educational disparities in Germany. *European Educational Research Journal*, *6*(4), 424–445. https://doi.org/10.2304/eerj.2007.6.4.424

Reimer, D., Smith, E., Andersen, I. G. & Sortkær, B. (2021). What happens when schools shut down? Investigating inequality in students' reading behavior during Covid-19 in Denmark. *Research in Social Stratification and Mobility*, *71*, Artikel 100568. https://doi.org/10.1016/j.rssm.2020.100568

Rose, S., Twist, L., Lord, P., Rutt, S., Badr, K., Hope, C., & Styles, B. (2021). *Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in key stage 1*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Publications/Covid-19_Resources/Impact_of_school_closures_KS1_interim_findings_paper_-_Jan_2021.pdf

Sachse, K. A., Jindra, C., Schumann, K. & Schipolowski, S. (2022). Soziale Disparitäten. In P. Stanat, S. Schipolowki, R. Schneider, K. A. Sachse, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich* (S. 151–180). Waxmann.

Schneider, T. (2005). Nachhilfe als Strategie zur Verwirklichung von Bildungszielen. Eine empirische Untersuchung mit Daten des Sozioökonomischen Panels (SOEP). *Zeitschrift für Pädagogik*, *51*(3), 363–379. https://doi.org/10.25656/01:4760

Schneider, T. & Linberg, T. (2022). Development of socio-economic gaps in children's language skills in Germany. *Longitudinal and Life Course Studies*, *13*(1), 87–120. https://doi.org/10.1332/175795921X16233448663756

Schröder, H. & Ganzeboom, H. B. G. (2014). Measuring and modelling level of education in European societies. *European Sociological Review*, *30*(1), 119–136. https://doi.org/10.1093/esr/jct026

Schwippert, K. (2019). Was wird aus den Büchern? Sozialer Hintergrund von Lernenden und Bildungsungleichheit aus Sicht der international vergleichenden Erziehungswissenschaft. *Journal for Educational Research Online*, *11*(1), 92–117. https://doi.org/10.25656/01:16789

Sieben, S. & Lechner, C. M. (2019). Measuring cultural capital through the number of books in the household. *Measurement Instruments for the Social Sciences*, *2*(1). https://doi.org/10.1186/s42409-018-0006-0

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Skopek, J. & Passaretta, G. (2021). Socioeconomic Inequality in Children's Achievement from Infancy to Adolescence: The case of Germany. *Social Forces*, *100*(1), 86–112. https://doi.org/10.1093/sf/soaa093

Snijders, T. A. B. (1999). Prologue to the measurement of social capital. *The Tocqueville Review*, *20*(1), 27–44. DOI: 10.3138/ttr.20.1.27

Statistisches Bundesamt (Destatis). (2022a). *Armutsgefährdungsschwelle und Armutsgefährdung (monetäre Armut)*. Statistisches Bundesamt (Destatis). https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Lebensbedingungen-Armutsgefaehrdung/Tabellen/armutsschwelle-gefaehrdung-mz-silc.html

Statistisches Bundesamt (Destatis). (2022b). *Armutsgefährdungsquote (monetäre Armut) nach Sozialleistungen in Deutschland nach dem überwiegenden Erwerbsstatus im Vorjahr und dem Bildungsstand*. Statistisches Bundesamt (Destatis). https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Lebensbedingungen-Armutsgefaehrdung/Tabellen/armutsgef-quote-bildungsstand-mz-silc.html

Stubbe, T. C. (2009a). Bildungsentscheidungen in der Sekundarstufe I. Sekundäre Herkunftseffekte an Hamburger Schulen. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, *29*(4), 419–436.

Stubbe, T. C. (2009b). *Bildungsentscheidungen und sekundäre Herkunftseffekte. Soziale Disparitäten bei Hamburger Schülerinnen und Schülern der Sekundarstufe I*. Waxmann.

Stubbe, T. C., & Bos, W. (2008). Die Koppelung von sozialer Herkunft und Lesekompetenz im internationalen Vergleich und in der Deutschsprachigen Gemeinschaft. In W. Bos, S. Sereni & T. C. Stubbe (Hrsg.), *IGLU Belgien. Lese- und Orthografiekompetenzen von Grundschulkindern in der Deutschsprachigen Gemeinschaft* (S. 111–125). Waxmann.

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 in Deutschland im internationalen Vergleich* (S. 209–226). Waxmann Verlag GmbH.

Stubbe, T. C. & Buddeberg, I. (2008). Lesesozialisation im Elternhaus in der Deutschsprachigen Gemeinschaft im internationalen Vergleich. In W. Bos, S. Sereni & T. C. Stubbe (Hrsg.), *IGLU Belgien. Lese- und Orthografiekompetenzen von Grundschulkindern in der Deutschsprachigen Gemeinschaft* (S. 127–136). Waxmann.

Stubbe, T. C., Buddeberg, 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, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 299–327). Waxmann.

Stubbe, T. C. & Goy, M. (2013). Besitzen wollen und sich leisten können. Analysen zur dimensional getrennten Erfassbarkeit von ökonomischem und kulturellem Kapital in Familien. In K. Schwippert, M. Bonsen & N. Berkemeyer (Hrsg.), *Schul- und Bildungsforschung. Diskussionen, Befunde und Perspektiven* (S. 203–222). Waxmann.

Stubbe, T. C., Kasper, D. & Jentsch, A. (2020). Schullaufbahnpräferenzen am Übergang in die Sekundarstufe und der Zusammenhang mit leistungsrelevanten und sozialen Merkmalen. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 315–330). Waxmann.

Stubbe, T. C., Krieg, M., Beese, C. & Jusufi, D. (2020). Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen von Viertklässlerinnen und Viertklässlern. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 263–289). Waxmann.

Stubbe, T. C., Krieg, M., Haas, M., & Nonte, S. (2023). Anlage und Durchführung von ProBiNi. In T. C. Stubbe, S. Nonte, M. Haas & M. Krieg (Hrsg.), *Musik- und MINT-Profile an Niedersächsischen Gymnasien und integrierten Gesamtschulen. Ergebnisse der Studie ProBiNi.* Waxmann Verlag GmbH. Manuskript in Vorbereitung,

Stubbe, T. C., Kleinkorres, R., Krieg, M., Schaufelberger, R., Schlitter, T. (2023). Soziale und migrationsbedingte Disparitäten in der Lesekompetenz von Viertklässlerinnen Viertklässlern. In N. McElvany, R. Lorenz, A. Frey, F. Goldhammer, A. Schilcher & T. C. Stubbe (Hrsg.), *IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre* (S. 151–177). Waxmann.

Stubbe, T. C., & Lorenz, J. (2017, August, 22–25). *Socially desirable resources and activities. Families' social capital and students' achievement in mathematics and science* [Konferenzpräsentation]. European Conference on Educational Research, Kopenhagen, Dänemark.

Stubbe, T. C., Pietsch, M. & Wendt, H. (2007). Soziale Netze an Hamburger Grundschulen. In W. Bos, C. Gröhlich & M. Pietsch (Hrsg.), *KESS 4 – Lehr- und Lernbedingungen in Hamburger Grundschulen* (S. 71–102). Waxmann.

Stubbe, T. C., Schaufelberger, R., Krieg, M., Kleinkorres, R., & Schlitter, T. (2023). Schullaufbahnpräferenzen am Übergang in die Sekundarstufe und der Zusammenhang mit leistungsrelevanten und sozialen Merkmalen. In N. McElvany, R. Lorenz, A. Frey, F. Goldhammer, A. Schilcher & T. Stubbe (Hrsg.), *IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre* (S. 231–248). Waxmann.

Tarelli, I., & Stubbe, T. C. (2010, Juli, 1–3). *Home literacy environment and reading achievement. A model for determining the relationship between socioeconomic status, home literacy environment and reading achievement* [Konferenzpräsentation]. 4th IEA International Research Conference, Göteborg, Schweden. https://www.iea.nl/sites/default/files/2019-04/IRC2010_Tarelli_Stubbe.pdf

Tophoven, S., Lietzmann, T., Reiter, S., & Wenzig, C. (2017). *Armutsmuster in Kindheit und Jugend. Längsschnittbetrachtungen von Kinderarmut.* Bertelsmann Stiftung.

UNESCO United Nations Educational, Scientific and Cultural Organization. (2003). International Standard Classification of Education, ISCED 1997. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in crossnational comparison. A European working book for demographic and socio-economic variables* (S. 195–220). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9186-7_10

Van der Gaag, M., & Snijders, T. A. B. (2005). The resource generator: Social capital quantification with concrete items. *Social Networks*, *27*, 1–29. https://doi.org/10.1016/j.socnet.2004.10.001

Weber, C., Helm, C., & Kemethofer, D. (2021a). Bildungsungleichheiten durch Schulschließungen? Soziale und ethnische Disparitäten im Lesen innerhalb und zwischen Schulklassen. In D. Fickermann, B. Edelstein, J. Gerick & K. Rachenbäumer (Hrsg.), *Schule und Schulpolitik während der Corona-Pandemie: Nichts gelernt?* (S. 83–99). Waxmann. https://doi.org/10.31244/9783830994589.04

Weber, C., Helm, C. & Kemethofer, D. (2021b). Are Social and Ethnic Reading Inequalities Increasing During School Closure? – The Mediating Role of Parental Involvement in Distance Learning. *Frontiers in Education*, *6*, Artikel 737064. https://doi.org/10.3389/feduc.2021.737064

Weis, M., Müller, K., Mang, J., Heine, J.-H., Mahler, N. & Reiss, K. (2019). Soziale Herkunft, Zuwanderungshintergrund und Lesekompetenz. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018. Grundbildung im internationalen Vergleich* (S. 129–162). Waxmann.

Wendt, H., Kasper, D., Bos, W., Vennemann, M. & Goy, M. (2017). Wie viele Punkte auf der TIMSS-Metrik entsprechen einem Lernjahr? In T. Eckert & B. Gniewosz (Hrsg.), *Bildungsgerechtigkeit* (S. 121–153). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15003-7_8

Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K. & Zierow, L. (2021). Bildung erneut im Lockdown: Wie verbrachten Schulkinder die Schulschließungen Anfang 2021? *Ifo-Schnelldienst*, *74*(5), 36–52.

Zierer, K. (2021). Effects of Pandemic-Related School Closures on Pupils' Performance and Learning in Selected Countries: A Rapid Review. *Education Sciences*, *11*(6), Artikel 252. https://doi.org/10.3390/educsci11060252

Zunker, N. & Neumann, M. (2020). Stratifizierung durch Profilierung? Das Zusammenspiel von inhaltlichen Profilen und Merkmalen der Schülerkomposition an nichtgymnasialen weiterführenden Schulen in Berlin. In C. I. Keßler & N. Nonte (Hrsg.), *(Neue) Formen der Differenzierung. Schul- und Klassenprofilierung im Spannungsfeld gesellschaftlicher Teilhabe und sozialer Ungleichheit* (S. 35–53). Waxmann.

Differenzielle Überzeugungen von angehenden Lehrkräften und deren Bedeutung für eine diversitätssensible Unterrichtsgestaltung

Miriam M. Gebauer & Ricarda K. Rübben

Zusammenfassung

Angesichts sozialer Wandlungsprozesse und stetig steigender Zahlen von Schülerinnen und Schülern mit Migrationshintergrund im deutschen Schulsystem sind Lehrkräfte – zunehmend stärker – dazu angehalten, ihren Unterricht diversitätssensibel zu gestalten. Der Beitrag widmet sich deshalb der Frage, ob und inwiefern Überzeugungen von Lehramtsstudierenden zu kulturell-ethnischer Vielfalt im Besonderen und zu Lehr-und Lerntheorien im Allgemeinen in einem Zusammenhang mit der Intention stehen, prospektiv einen Unterricht zu planen, der kulturell-ethnische Diversität (nicht) berücksichtigt. Datengrundlage bildet eine Fragebogenerhebung mit $N = 345$ Lehramtsstudierenden. Die Analysen zeigen statistisch signifikante Zusammenhänge zwischen den Überzeugungen zu kulturell-ethnischer Vielfalt und der Intention einer diversitätssensiblen Unterrichtsgestaltung, nicht aber zwischen den lehr- und lerntheoretischen Überzeugungen und dieser Intention. Die Ergebnisse werden vor dem Hintergrund ihrer Bedeutung und Implikationen für die universitäre Lehrerbildung und Professionsforschung abschließend diskutiert.

Schlagworte: Diversitätssensibler Unterricht; Lehr-lerntheoretische Überzeugungen von angehenden Lehrkräften; Selbstwirksamkeitsüberzeugung; Vorstellungen in Bezug auf Akkulturationsprozesse

Abstract

In view of social change processes and steadily increasing numbers of students with a migration background in the German school system, teachers

are – increasingly – required to design their lessons in a diversity-sensitive way. This article therefore addresses the question of whether and to what extent student teachers' beliefs about cultural-ethnic diversity in particular and about teaching and learning theories in general are related to their intention to prospectively plan lessons that (do not) take cultural-ethnic diversity into account. Data are based on a questionnaire survey with $N = 345$ student teachers. The analyses show statistically significant correlations between beliefs about cultural-ethnic diversity and the intention of diversity-sensitive instructional design, but not between teaching and learning theory beliefs and this intention. Finally, the results are discussed in light of their significance and implications for university-based teacher education and professional research.

Keywords: diversity-sensitive teaching; beliefs towards acculturation; epistemological beliefs of future teachers; self-efficacy beliefs

1. Einleitung

Der adäquate Umgang mit der Diversität der Schülerinnen und Schüler markiert eine konstitutive Herausforderung für die Schul- und Unterrichtspraxis sowie die Lehrerbildung (Dumont 2019; Trautmann & Wischer, 2011): Es besteht der normative Anspruch, allen Schülerinnen und Schülern, die sich hinsichtlich verschiedener lern- und leistungsrelevanter Merkmale (z. B. Geschlechts-, Kultur- und Religionszugehörigkeit, Motivation, kognitive Fähigkeiten) sowie Bedürfnisse (Emmerich & Moser, 2020) unterscheiden, in ihrer Vielfalt individuell gerecht zu werden (Prengel, 2007). Im Fokus der Diskussion steht immer wieder Migration respektive kulturell-ethnische Zugehörigkeit als ein Merkmal, das in Besonderem mit schulischen Marginalisierungs- und Exklusionsmechanismen verbunden ist und Bildungsbenachteiligung begünstigen kann (Emmerich & Moser, 2020; Stanat & Edele, 2011). Besondere Relevanz erfährt dieses Diversitätsmerkmal vor dem Hintergrund sozialer Wandlungsprozesse und weltweiter Migrations- und Fluchtbewegungen und damit verbunden einem kontinuierlichen Anstieg des Anteils von Schülerinnen und Schülern mit Migrationshintergrund respektive kulturell-ethnischer Diversität in Schulen (Bundesamt für Migration und Flüchtlinge, 2019; Dumont, 2019; Hachfeld & Syring, 2020). Aktuelle Befunde zeigen, dass inzwischen 35,8 % der Schülerinnen und Schüler an allgemein- und berufsbildenden Schulen in Deutschland einen Migrationshintergrund aufweisen (Statistisches Bundesamt, 2021).

Der Anspruch eines diversitätssensiblen Unterrichts ist deshalb neben der Berücksichtigung der oben aufgeführten individuellen lern-und leistungsrelevanten Merkmale der Schülerinnen und Schüler insbesondere auch mit einer Berücksichtigung ihres kulturell-ethnischen Hintergrunds verbunden. Ausgehend vom kompetenztheoretischen Ansatz der Lehrerprofessionalität sind für eine qualitätsvolle Unterrichtsgestaltung neben dem Professionswissen selbstregulative Fähigkeiten und motivationale Orientierungen sowie Einstellungen und Überzeugungen relevant (Baumert & Kunter, 2006; Blömeke, Gustafsson & Shavelson 2015). Einstellungen und Überzeugungen können dabei in Bezug auf das *Selbst*, das Unterrichten, schulische Lehr- und Lernprozess sowie gesellschaftliche Prozesse herausgebildet werden (Fischer & Ehmke, 2019; Fives & Buehl, 2016). Für eine diversitätssensible Unterrichtsgestaltung wird dabei theoretisch angenommen, dass spezifische Überzeugungen zu kulturell-ethnischer Diversität der Schülerinnen und Schüler von Relevanz sind (Tichnor-Wagner et al., 2019). Rezente Forschungsarbeiten weisen darauf hin, dass bereits angehende Lehrkräfte differenzielle Einstellungen und Überzeugungen haben und eine darauf bezogene positive oder ablehnende Haltung bedeutsam für die Unterrichtsgestaltung zu sein scheint (Gebauer & McElvany, 2017; Bello et al., 2017; Fischer & Ehmke, 2019).

Der vorliegende Beitrag knüpft an bestehende Forschungsergebnisse zu Überzeugungen an und beschäftigt sich mit der Bedeutung von verschiedenen – spezifisch kulturell-ethnischen und allgemeinen lehr-lerntheoretischen – Überzeugungen für eine zukünftige diversitätssensible Unterrichtsgestaltung. Dabei sollen folgende Überzeugungen in Bezug auf kulturell-ethnische Diversität in den Fokus gerückt werden: die Intensität der Konsequenzen für die Unterrichtsgestaltung, Akkulturationsprozesse sowie selbstbezogene Kompetenzwahrnehmungen. Zusätzlich sollen lehr-lerntheoretische Überzeugungen ohne einen kulturell-ethnischen Objektbezug sowie die Selbsteinschätzung des im Studiums erworbenen Sachwissens, der eigenen Handlungskompetenz und der Erfahrungen zum Umgang mit kulturell-ethnischer Diversität im Unterricht berücksichtigt werden. Die Betrachtung mehrerer bedeutsamer Überzeugungen stellt damit ein Novum im Bereich der Überzeugungsforschung im Kontext des diversitätssensiblen Unterrichts dar.

2. Forschungsgegenstand und Forschungsdesiderat

2.1. Umgang mit kulturell-ethnischer Diversität im Unterricht

Der Grund für die Auseinandersetzung mit quantifizierten Merkmalsunterschieden zwischen Schülerinnen und Schülern liegt nicht darin, auf Gegensätze oder Verschiedenartigkeit hinzuweisen (Fylkesnes, 2018), sondern die Notwendigkeit adäquater Unterrichtsstrategien zu unterstreichen, um die kulturell-ethnische Vielfalt von Schülerinnen und Schüler wertschätzend zu integrieren (Au, 2009; Darling-Hammond, 2010). In einem diversitätssensiblen Unterricht sollen individuelle Entwicklungsverläufe und Lernprozesse der Schülerinnen und Schüler durch partizipative und kooperative Lernformate angeleitet und gleichzeitig der interkulturelle Austausch gefördert werden, indem die kulturell-ethnische Vielfalt in Unterrichtsinhalte eingebunden wird (Banks, 2016). Dabei werden in der Literatur unterschiedliche Ansätze beschrieben, die hinsichtlich differenzieller Strategien oder Perspektiven kontrastiert werden können:

- Die *multikulturellen* (für weitere Unterscheidungen multikultureller Ansätze siehe z. B. Steinberg & Kincheloe 2001) oder *polykulturellen* Ansätze beschreiben, wie kulturell-ethnische Differenzen wertschätzend gewürdigt werden können, interkultureller Austausch curricular verankert und kulturell-ethnische Wertorientierungen und Traditionen bewahrt werden sollen, wobei in polykulturellen Ansätzen die wechselwirkende Einflussnahme zwischen Kulturen über die Zeit Berücksichtigung finden (Kelley, 1999; Parekh, 1999).
- Demgegenüber steht der *color-blindness* Ansatz mit der Vorstellung, dass kulturell-ethnische Diversität keine Bedeutung für interindividuelle Interaktionen und Kommunikation hat und in Bildungs- oder Arbeitskontexten unberücksichtigt bleiben kann, da der Fokus auf Übereinstimmungen und Gleichheit zwischen den Gruppen und nicht den Differenzen liegt (Markus et al., 2000; Rosenthal & Levy, 2010). Dabei steht dieser Ansatz in der Kritik, Diskriminierung und Machtstrukturen unberücksichtigt zu lassen und somit Benachteiligung zu fördern (Apfelbaum et al., 2012; Gillborn 2019).
- Integrierende theoretische Postulate hingegen sehen mögliche Vorteile in einer Kombination dieser programmatischen Ansätze für Bildungskontexte, in denen der Aspekt der Gleichheit nicht im

Sinne einer Assimilationserwartung interpretiert wird (Rosenthal & Levy, 2010).

Es gibt zudem Untersuchungen zu den verschiedenen Ansätzen, auf die an dieser Stelle lediglich verwiesen werden kann (color-blindness Ansatz z. B.: Gaertner & Dovidio, 2000; Gaertner et al., 1993; Hachfeld et al., 2015; Schütze & Törngren, 2022; multikultureller Ansatz z. B.: Berry & Kalin, 1995; Ryan et al., 2010; Stephan & Stephan, 2001; vergleichende Untersuchungen z. B. Plaut, Thomas & Goren, 2009; Richeson & Nussbaum, 2004; Wolsko et al., 2000).

2.2. (Diversitätssensible) Überzeugungen als Teil professioneller Handlungskompetenz

Einstellungen und Überzeugungen als Teil der professionellen Handlungskompetenz von Lehramtsstudierenden und Lehrkräften spielen für eine diversitätssensible Unterrichtsgestaltung eine bedeutsame Rolle (Tichnor-Wagner et al., 2019). Dabei gelten sowohl implizite als auch explizite Einstellungen und Überzeugungen von (angehenden) Lehrkräften als zentrale Bedingungen für eine erfolgreiche Unterrichtsgestaltung und die Steuerung von Wissenserwerbsprozessen. Die Begriffe Einstellungen und Überzeugungen werden dabei nicht trennscharf verwendet (z. B. Döhrmann et al., 2014; Buchholzer & Pelgrim, 2013): Einstellungen werden typischerweise als kognitive, affektive und behaviorale Gesamtbewertungen der Inhalte definiert (Eagly & Chaiken, 1993; Maio & Haddock, 2009). Einige theoretische Ausführungen gehen davon aus, dass Überzeugungen als kognitive Teilkomponente von Einstellungen verstanden werden können (Baumert & Kunter, 2006; Eagly & Chaiken, 1993). Gleichwohl besteht weitgehend Konsens dahingehend, dass Überzeugungen als selbst-normative (Oser & Blömeke, 2012) konzeptionelle Repräsentationen in Bezug auf ein Objekt beschrieben werden können, die basierend auf Wissen und Erfahrungen neue Erlebnisse filtern, einordnen und infolgedessen handlungsleitend sind (Fives & Buehl, 2012; Hermans et al., 2008; Reusser et al., 2011). Überzeugungen von Lehrkräften können darüber hinaus in Bezug auf das *Selbst*, das Unterrichten und die Unterrichtsplanung, schulische Lehr- und Lernprozesse (Fives & Buehl, 2012; Pajares, 1992) sowie hinsichtlich interkultureller Prozesse herausgebildet werden (Fischer & Ehmke, 2019) und angehende Lehrkräfte haben bereits während ihres Studiums differenzielle Überzeugungen in Bezug auf den Umgang mit Diversität (z. B. Bello et al., 2017).

Überzeugungen, die bedeutsam für einen diversitätssensiblen Unterricht sein können, sind solche, die auf 1) unterrichtsbezogene Konsequenzen von kulturell-ethnischer Diversität, 2) auf Akkulturation, 3) auf Lehr- und Lerntheorien und 4) berufsbezogene Selbstwirksamkeitsüberzeugung Bezug nehmen (z. B. Fischer & Ehmke, 2019):

- zu 1) *Einschätzungen von unterrichtsbezogenen Konsequenzen von Diversität* sind die Überzeugungen von angehenden Lehrkräften, dass die kulturell-ethnische Vielfalt von Schülerinnen und Schülern Aspekte der Unterrichtsplanung beeinflusst (Decristan et al. 2017). Vorausgesetzt wird eine Wechselwirkung zwischen den Merkmalen von Schülerinnen und Schülern und der Unterrichtsgestaltung (Cronbach & Snow, 1977; Snow, 1989). Gleichzeitig wird die Diversität von Schülerinnen und Schülern als relevanter Faktor für die Unterrichtsvorbereitung, -durchführung und -reflexion beschrieben (Hurtado, 1996). Neue Untersuchungen fanden heraus, dass die Wahrnehmung von Diversität bedeutsam für die Einschätzung von zusätzlichem Aufwand für eine diversitätssensible Unterrichtsgestaltung bei angehenden Lehrkräften zu sein scheint (Glock et al., 2019). Hingegen sind negativ bewertete Einschätzungen der Kosten im Sinne eines belastenden zusätzlichen Arbeitsaufwandes nicht von Relevanz für die Unterrichtsvorbereitung, -durchführung und -reflexion (Gebauer & McElvany, 2020).
- zu 2) *Akkulturationsüberzeugungen* sind Vorstellungen, die angehende Lehrkräfte in Bezug auf den Prozess der Akkulturation im schulischen Kontext haben und darüber, inwieweit Veränderungen von beispielsweise Verhaltensweisen durch die Übernahme von anderen kulturell-ethnischen Werten oder Sichtweisen durch interkulturelles Lernen verstärkt oder vermieden werden sollten und gleichzeitig die Herkunftskultur bewahrt werden kann (vgl. Göbel & Buchwald, 2017; Wagner et al., 2001; Weidemann, 2007). In der Akkulturationsforschung werden vier Akkulturationsstrategien beschrieben, die in Integration, Assimilation, Separation und Marginalisierung differenziert werden (Berry, 2005; Weidemann, 2007). Mit Integration und Assimilation sowie Separation und Marginalisierung werden zwei gegenüberstehende Pole beschrieben, die die Bewahrung beziehungsweise das Ablegen der Herkunftskultur und die Anpassung an die Aufnahmekultur beschreiben. Zugleich werden mit der Gegenüberstellung von

Integration und Separation sowie von Assimilation und Marginalisierung zwei gegenüberstehende Punkte definiert, die die Absicht und im anderen Fall die Ablehnung eines interkulturellen Gruppenaustausches beschreiben (Berry, 2005; Sam & Berry, 2006). Wagner et al. (2001) konnten Zusammenhänge zwischen Autoritätshaltung und Akkulturationsüberzeugungen von Lehrkräften sowie zu subgruppenspezifischen Herkunftsmerkmalen von Schülerinnen und Schülern herstellen, untersuchten aber keine Zusammenhänge zu Unterrichtsmerkmalen. Hachfeld et al. (2011) fanden positive Zusammenhänge zwischen Akkulturationsüberzeugungen von Studierenden und multikulturellen Überzeugungen, wobei in dieser Studie die Akkulturationsüberzeugungen Integration und Assimilation nicht als Subskalen, sondern kombiniert in den Analysen berücksichtigt wurden und die multikulturellen Überzeugungen keinen expliziten Bezug auf Vorgehensweisen zur Berücksichtigung von kulturell-ethnischer Diversität im Unterricht nehmen. Darüber hinaus liegen Evidenzen darüber vor, wie Akkulturationsprozesse im Rahmen des Unterrichts unterstützt werden können; inwieweit Überzeugungen von Lehrkräften in Bezug auf schülerseitige Akkulturationsprozesse eine diversitätssensible Unterrichtsgestaltung begünstigen, wird in dieser Untersuchung jedoch nicht betrachtet (Tomcho & Foels, 2002). Hinweise dafür, dass positive Überzeugungen in Bezug auf kulturell-ethnische Vielfalt bedeutsam für diversitätssensible Unterrichtsgestaltung ist, berichten Agostinetto und Bugno (2020). Dabei fanden sie Befunde dafür, dass Lehrkräfte zwar positive multikulturelle Überzeugungen berichten, den effektiven interkulturellen Austausch zwischen Schülerinnen und Schülern jedoch nur in geringem Maße unterstützen.

- zu 3) Unterrichtsbezogene Überzeugungen, die keinen Objektbezug zum Umgang mit kulturell-ethnischer Diversität herstellen, aber von Relevanz für einen qualitätsvollen Unterricht sind, sind *lehr-lerntheoretische Vorstellungen*, die Repräsentationen über Wissensvermittlungs- und Wissensaneignungsprozesse umfassen (Hofer, 2002; Voss et al., 2013). Diese Überzeugungen werden in theoretischen Konzeptionen in transmissive und konstruktivistische Vorstellungen unterschieden, wobei transmissive Vorstellungen als lehrerzentriert beschrieben werden und von einer direkten Wissensübertragung von der Lehrkraft auf die Schülerinnen und Schüler ausgehen (Chan & Elliott, 2004). Konstruktivistische

Annahmen hingegen sehen Lernprozesse als individuelle schülerzentrierte Erkenntnisprozesse, in denen eigenständiges Lernen durch die Lehrenden angeleitet wird (Hofer, 2002). Aktuelle Forschungen weisen auf die Bedeutung von lehr-lerntheoretischen Überzeugungen für Einstellungen in Bezug auf inklusiven Unterricht hin, wobei Zusammenhänge zwischen konstruktivistischen Überzeugungen und positiven Einstellungen in Bezug auf einen inklusiven Unterricht festgestellt wurden (Sheehy et al., 2019; Silverman, 2007). Empirische Evidenzen zum Zusammenhang zwischen lehr-lerntheoretischen Überzeugungen und der Unterrichtsgestaltung zeigen eine variierende Befundlage und dass lehr-lerntheoretische Überzeugungen nicht in jedem Fall zu einer den Überzeugungen entsprechenden Unterrichtsgestaltung führen (Olafson & Schraw, 2006). Inwieweit das auch für diversitätssensiblen Unterricht mit Bezug auf eine kulturell-ethnische Diversität der Schülerinnen und Schüler gilt, ist weitgehend ungeklärt.

- zu 4) Die *berufsbezogene Selbstwirksamkeitsüberzeugung* beschreibt die wahrgenommene Überzeugung in die eigenen Fähigkeiten, in zukünftigen Situationen herausfordernde Aufgaben erfolgreich zu bewältigen (Bandura, 1997; Henson, 2002). Infolgedessen ist die berufsbezogene Selbstwirksamkeitsüberzeugung von Relevanz für die Auswahl von Handlungsstrategien und somit verantwortlich für eine erfolgreiche Aufgabenbewältigung, die für eine qualitätsvolle Unterrichtsgestaltung bedeutsam ist (Klassen & Tze, 2014). Lehrkräfte mit hoch ausgeprägter Selbstwirksamkeitsüberzeugung zeigen mehr Ausdauer in der Unterrichtsvorbereitung, setzen vielfältige und häufiger neue Unterrichtsmethoden ein und bringen mehr Zeit für Interaktionen mit Schülerinnen und Schülern auf (Tschannen-Moran & Woolfolk-Hoy, 2001). Dieses Vertrauen in die eigenen Fähigkeiten kann um den Bezug zum kompetenten Umgang mit kulturell-ethnisch diversen Schülergruppen erweitert werden (Siwatu, 2011a) und Evidenzen weisen auf positive Zusammenhänge mit multikulturellen Überzeugungen hin (Hachfeld et al., 2015; Siwatu, 2011a). Gleichzeitig fanden vorherige Untersuchungen Zusammenhänge zwischen hoher beruflicher Selbstwirksamkeitsüberzeugung und schülerzentrierter Unterrichtsgestaltung (Guo et al., 2010). Inwieweit eine berufsbezogene Selbstwirksamkeitsüberzeugung zur Berücksichtigung kulturell-ethnischer Diversität bedeutsam für eine diversitätssensible Unterrichtsgestaltung ist, ist bisher allerdings ungeklärt.

Forschungsbefunde im Kontext sprachlich-kultureller Heterogenität haben gezeigt, dass für angehende Lehrkräfte Praxiserfahrungen und im Studium erworbenes Wissen bedeutsam für eine positive oder ablehnende Überzeugung sind (Hammer et al., 2016; Fischer et al., 2018). Darüber hinaus scheinen neben praktischer Unterrichtserfahrung insbesondere Erfahrung mit diversen Schülergruppen relevant für die Überzeugungen zu sein (Alisaari et al., 2019; Sharma et al., 2008). Dabei berichten Lehrkräfte jedoch, wenig bis keine Erfahrung im Unterrichten von kulturell-ethnisch diversen Klassen zu haben (Siwatu, 2011b) und haben eher negative Stereotype über Schülerinnen und Schüler (Gao & Mager, 2011; Glock, 2016), die sie mitunter als „anders" im Vergleich zu sich selbst betrachten (Hogg, 2005). Insbesondere das während der Lehrerausbildung erworbene Wissen und die praktischen Erfahrungen tragen dazu bei, dass Lehrkräfte ihre Einstellungen in Bezug auf die Vielfalt der Lernenden und Diversität im Klassenzimmer positiv verändern (Bayraktar, 2011; Mellom et al., 2018).

3. Forschungsfragen

Vor dem Hintergrund der Explikation des Forschungsgegenstandes, des Forschungsstandes und der Forschungsdesiderate, ergeben sich folgende Forschungsfragen:

1) Sind Einstellungen bezogen auf Intensität der Auswirkung von Diversität für die Unterrichtsgestaltung bedeutsam für eine intendierte diversitätssensible Unterrichtsgestaltung oder für die Intention, Diversität nicht zu berücksichtigen?

2) Sind Einstellungen bezogen auf Akkulturationsprozesse von Bedeutung für eine intendierte diversitätssensible Unterrichtsgestaltung oder für eine Intention, Diversität im Unterricht unberücksichtigt zu lassen?

Es wird angenommen, dass eine integrative Akkulturationsüberzeugung, nach der die Herkunftskultur neben der Aufnahmekultur bewahrt und wertgeschätzt wird, mit der Intention, einen diversitätssensiblen Unterricht, gestalten zu wollen, positiv zusammenhängt (Göbel & Buchwald, 2017; Weidemann, 2007). Vice versa wird ein negativer Zusammenhang mit der Intention zu einem Diversität nicht berücksichtigenden Unterricht angenommen. Bei assimilierenden Akkulturationsüberzeugungen seitens der angehenden

Lehrkräfte, die die Aufgabe der Herkunftskultur und Anpassung an die Aufnahmekultur präferieren (Göbel & Buchwald, 2017; Weidemann, 2007) wird erwartet, dass diese mit der Intention, Diversität im Unterricht nicht berücksichtigen zu wollen, korreliert. Vice versa wird angenommen, dass sie negativ mit der Intention einer diversitätssensiblen Unterrichtsgestaltung zusammenhängen.

3*) Sind lehr-lerntheoretische Überzeugungen bedeutsam für eine intendierte diversitätssensible Unterrichtsgestaltung oder für die Intention, Diversität unberücksichtigt zu lassen?*

Es wird davon ausgegangen, dass Überzeugungen, die einen schülerzentrierten Unterricht favorisieren und Lernen als subjektiven Konstruktionsprozess verstehen (Voss et al., 2013), mit der Absicht einer diversitätssensiblen Unterrichtsgestaltung positiv zusammenhängen und negativ mit der einer Diversität nicht berücksichtigenden. Bei transmissiven Vorstellungen wird erwartet, dass diese positiv mit einer Unterrichtsgestaltung zusammenhängen, die die kulturell-ethnische Diversität unberücksichtigt lässt und negativ mit einer diversitätssensiblen Unterrichtsgestaltung.

4) Sind berufsbezogene Selbstwirksamkeitsüberzeugungen bezugnehmend auf den Umgang mit kulturell-ethnischer Vielfalt bedeutsam für eine intendierte diversitätssensible Unterrichtsgestaltung oder für eine Intention, Vielfalt unberücksichtigt zu lassen?

Es wird erwartet, dass eine hoch ausgeprägte selbstbezogene Kompetenzüberzeugung in Bezug auf den Umgang mit kulturell-ethnisch diversen Schülergruppen positiv mit einer diversitätssensiblen Unterrichtsgestaltung zusammenhängt, da sie zur adäquaten Auswahl von Handlungsstrategien und erfolgreichen Handlungsbewältigung führt (Bandura, 1997; Tschannen-Moran & Woolfolk-Hoy, 2001). Für Zusammenhänge mit einem Unterricht, in dem kulturell-ethnische Diversität unberücksichtigt bleibt, wird ein diametrales Ergebnis angenommen.

5) Sind Wissen, praktische Erfahrungen und Erfahrungen mit kulturell-ethnischer Diversität bedeutsam für eine intendierte diversitätssensible Unterrichtsgestaltung oder für eine Intention, die kulturelle-ethnische Diversität nicht zu berücksichtigen?

Es wird erwartet, dass Wissen und Handlungserfahrungen die Intention, einen diversitätssensiblen Unterricht zu gestalten, begünstigen, da neben Überzeugungen (*dispositions*), Wissen (*knowledge*) und Fähigkeiten (*skills*) als Voraussetzung für eine diversitätssensible Unterrichtsgestaltung beschrieben werden (Tichnor-Wagner et al., 2019).

4. Methodische Vorgehensweise

4.1. Stichprobe

Die Stichprobe umfasst $N = 345$ Studierende (davon 72,2 % weiblich) mit einem Altersdurchschnitt von $M = 24{,}51$ Jahren ($SD = 2{,}92$). Die querschnittliche Fragebogenerhebung (Haupterhebung: Sommersemester 2011; Wintersemester 2011/2012 und Nacherhebung: Wintersemester 2013/2014[1]) wurde von geschulten Testleiterinnen und Testleitern an einer nordrhein-westfälischen Universität durchgeführt, deren Teilnahme freiwillig war. Zum Zeitpunkt der Datenerfassung befanden sich 39,4 % der Befragten im Bachelor- und 60,6 % im Masterstudium. 51,6 % der Befragten studierten Grundschullehramt, 16,4 % Haupt- und Realschullehramt und 26,9 % Lehramt für Gesamtschulen und Gymnasien. Bei der Frage zum ersten studierten Fach wurde Deutsch mit 30 % am häufigsten genannt und Mathematik mit rund 24 % am zweithäufigsten. Weitere studierte Fächer waren mit rund 20 % Sport, Kunst oder Musik.[2]

1 Die damals geltende Prüfungsordnung (LABG 2009) wurde im Jahr 2016 novelliert. Eine Prüfung der Dokumente zeigte, dass das Angebot an Lehrveranstaltung zur Thematik „Umgang mit Diversität" für alle Studiengänge und Lehrämter gleichblieb.

2 Weitere Merkmale wie bspw. die Studienabschnittsphase oder Praxiserfahrungen, in der sich die Studierenden zum Befragungszeitpunkt befanden, wurden an anderer Stelle untersucht und sind deshalb für diesen Beitrag nicht relevant (Gebauer & McElvany, 2017; Gebauer, McElvany, & Klukas, 2013).

4.2. Instrumente

Intendiertes Unterrichtsverhalten

Zwei Skalen wurden zur Erfassung des intendierten Unterrichtsverhaltens eingesetzt. Zum einen eine Skala mit fünf Items zur Erfassung der Absicht kulturell-ethnischer Diversität im Unterricht nicht zu berücksichtigen (Beispielitem: „*Ich werde mich in meiner Unterrichtsvorbereitung nicht auf kulturelle Unterschiede bei den Schüler/innen konzentrieren*"). Sowie eine weitere Skala mit sieben Items, die kulturell-ethnisch Diversität berücksichtigt und als Indikator für einen diversitätssensiblen Unterricht steht (Beispielitem: „*Ich werde kulturell-differenzielle lernunterstützende Anreize setzen, damit meine Schüler/innen ihre persönlichen Lernziele finden und strukturiert verfolgen*"). Die theoretisch angenommene Differenz konnte anhand von Faktorenanalysen (Einfaktormodell gegen ein Zweifaktorenmodell) empirisch bestätigt werden ($\Delta\chi^2 = 24$, $\Delta df = 1$, $p < .01$). Die Skalen wurden zum Zweck der Studie entwickelt (Gebauer, McElvany & Klukas, 2013) und die Bedeutung anderer Konstrukte für diese Skalen wurden bereits an anderer Stelle untersucht und publiziert (vgl. Gebauer & McElvany, 2017). Die Items konnten auf einem vierstufigen Antwortformat beantwortet werden von 1 = *trifft überhaupt nicht* zu bis 4 = *trifft voll und ganz zu* (siehe Tabelle 1 für Korrelationen, Mittelwert, Standardabweichung und Reliabilität).

Konsequenzen für Unterrichtsgestaltung

Die Erfassung zur Einschätzung des Ausmaßes von Konsequenzen kulturell-ethnischer Diversität auf Aspekte des Unterrichts (Planung, Durchführung, Nachbereitung) wurde anhand einer für den Zweck dieser Studie entwickelten Skala anhand von fünf Items erfasst (Gebauer, McElvany & Klukas, 2013 in Anlehnung an Maruyama et al., 2000). Das Antwortformat der Items reichte von 1 = *gering bis keine* bis 4 = *sehr groß*. Ein Beispielitem lautet „*Wie groß sind Ihrer Einschätzung nach die Auswirkungen von kultureller Heterogenität in Schulklassen auf die Unterrichtsplanung*?" (siehe Tabelle 1 für Korrelationen, Mittelwert, Standardabweichung und Reliabilität).

Akkulturation

Die Einstellungen zur Akkulturation wurden anhand von zwei Subskalen mit einem bereits in anderen Studien eingesetzten und für diese Studienzwecke adaptierten Instrument erfasst (van Dick, Wagner, Adams & Petzel, 1997). Die Subskalen *Integration* (Beispielitem: „*Lehrkräfte sollten den Kontakt zwischen Schüler/innen verschiedener kultureller Herkunft fördern.*"), gemessen mit drei Items, und *Assimilation* (Beispielitem: „*Lehrkräfte sollten darauf achten, dass Schüler/innen anderer ethnischer Herkunft in den Schulpausen unter sich nur deutsch sprechen.*"), gemessen mit vier Items, ließen sich anhand von Faktorenanalysen empirisch trennen ($\Delta\chi^2 = 44$, $\Delta df = 1$, $p < .01$). Die Items konnten anhand eines sechsstufigen Antwortformats von 1 = *stimme überhaupt nicht zu* bis 6 = *stimme sehr zu* beantwortet werden (siehe Tabelle 1 für Korrelationen, Mittelwert, Standardabweichung und Reliabilität).

Lehr-lerntheoretische Überzeugungen

Zur Erfassung der lehr-lerntheoretischen Überzeugungen wurde ein bereits in anderen Studien eingesetztes Instrument verwendet (Rakoczy, Buff & Lipwosky, 2005). Es wurden zwei Subskalen eingesetzt, die den theoretischen Annahmen und der Unterscheidung in *transmissive* (Beispielitem: „*Schüler/innen können auch ohne Anleitung zu vielen Unterrichtsaufgaben Lösungen finden.*"), gemessen anhand von acht Items und *konstruktivistische* Überzeugungen (Beispielitem: „*Lehrkräfte sollten für das Lösen von Aufgaben detaillierte Vorgehensweisen vermitteln.*"), gemessen mit zehn Items, entspricht und die anhand von Faktorenmodellprüfung bestätig werden konnte (ein Faktormodell gegen ein zwei Faktorenmodell $\Delta\chi^2 = 73$, $\Delta df = 1$, $p < .01$). Die Items konnten auf einem vierstufigen Antwortformat von 1 = *trifft überhaupt nicht* zu bis 4 = *trifft voll und ganz* beantwortet werden (siehe Tabelle 1 für Korrelationen, Mittelwert, Standardabweichung und Reliabilität).

Tabelle 1: Korrelationen, Mittelwerte, Standardsabweichung, Reliabilität und Prozentangaben der fehlenden Werte der in den Analysen genutzte Skalen

	2	3	4	5	6	7	8	9	10	11	M	SD	α	Missing %
01 Konsequenzen Unterricht	.14**	.05	.03	.10	-.10	.02	.02	.05	-.10	.17**	2.62	0.51	.78	4.1
02 Akkulturation (Integration)		-.38**	.25**	-.16**	.26**	-.02	-.09	.20**	-.42**	.41**	4.45	0.76	.70	5.2
03 Akkulturation (Assimilation)			.12*	.33**	-.23**	.03	.13*	-.12*	.32**	-.26**	2.71	0.71	.61	4.3
04 Lehr-Lerntheoretisch (Konstruktivistisch)				-.32**	.12*	.07	.06	.05	-.10	.22**	3.10	0.32	.68	4.6
05 Lehr-Lerntheoretisch (Transmissiv)					-.01	-.05	-.02	.01	.34**	-.13*	2.25	0.57	.86	5.2
06 Berufliche Selbstwirksamkeit						-.31**	-.34**	.10	-.24**	.36**	2.96	0.41	.72	0.3
07 Sachwissen							.65**	-.06	.12*	-.08	3.12	1.17		5.1
08 Handlungswissen								-.16**	.21**	-.12*	3.35	1.30		4.3
09 Erfahrung kulturelle Diversität									-.11	.10	3.18	2.14		23.5
10 Keine Berücksichtigung Diversität											2.04	0.55	.64	0.3
11 Diversitäts-sensibel											2.76	0.53	.79	1.2

Anmerkung: * p <.05; ** p < .01 (2-seitig) signifikant.

Berufliche Selbstwirksamkeitsüberzeugung

Die berufliche Selbstwirksamkeitsüberzeugung wurde anhand eines adaptierten Instruments mit zehn Items erfasst (Schwarzer & Schmitz, 1999). Die Adaption des Instruments bestand darin, dass ein Bezug zur kulturell-ethnischen Diversität der Schülerinnen und Schüler aufgenommen wurde. Das Antwortformat der Items reichte von 1 = *trifft überhaupt nicht* zu bis 4 = *trifft voll und ganz zu*. Ein Beispielitem lautet „*Im Hinblick auf das Unterrichten von kulturell heterogenen Klassen weiß ich, dass ich es schaffen werde, die erwarteten Lerninhalte zu vermitteln, wenn ich mich darum bemühe*." (siehe Tabelle 1 für Korrelationen, Mittelwert, Standardabweichung und Reliabilität).

Die Einschätzung und Bewertung des in der Universität erworbenen *Sachwissens* und *Handlungswissens* wurde mit Notenbewertungen von 1 = *sehr gut* bis 6 = *unbefriedigend* abgefragt.

Die Erfahrungen mit kulturell-ethnisch diversen Klassen konnten die Studienteilnehmerinnen und -teilnehmer mit einer prozentualen Anteilseinschätzung bewerten, die von 0 = 0–9.99 % bis 9 = 90–100 % reichte.

Zur Kontrolle wurde das Geschlecht in die Analysen aufgenommen 1 = *weiblich* und 2 = *männlich*.

4.3. Analyseverfahren

Für die deskriptiven Auswertungen und Vorbereitungen der Daten für weitere Analyseschritte wurde das Programm SPSS 27 (IBM SPSS Statistics, 2020) genutzt. Für die Prüfung der Konstrukte, Überprüfung der Forschungsfragen und angenommenen Zusammenhänge zwischen den Konstrukten wurden latente konfirmatorische Faktorenanalysen und Strukturgleichungsmodelle mithilfe des Programms M*plus* 8 (Muthén & Muthén, 1998–2017) berechnet. Für die Bewertung der Modelle wurden die üblichen Grenzwerte der Fitstatistiken genutzt (Comparative Fit Index [CFI]; Tucker-Lewis-Index [TLI]; Root Mean Square Error of Approximation [RMSEA] und Standardized Root Mean Square Residual [SRMR]). Hier zeigten sich akzeptable Werten für die CFI- und TLI-Werte ab .90 und RMSEA und SRMR unter .08 (Hu & Bentler, 1999) sowie χ^2 und Freiheitsgrade (*df*). Bei der Schätzung dieser Modelle werden Korrelationen der Residualvarianzen zwischen Variablen mit gleichen Itemstämmen zugelassen. Fehlende Werte wurden anhand der Einstellung FIML (Full Information Maximum Likelihood-Estimation, Asparouhov & Muthén, 2010) berücksichtigt. Die dafür vorliegende Eigenschaft der fehlenden Werte *missing completly at random* (MCAR, Rubin, 1976) wurde anhand des Little

Tests (Little, 1988, siehe Tabelle 1 für Prozentangaben der fehlenden Werte) geprüft und bestätigt.

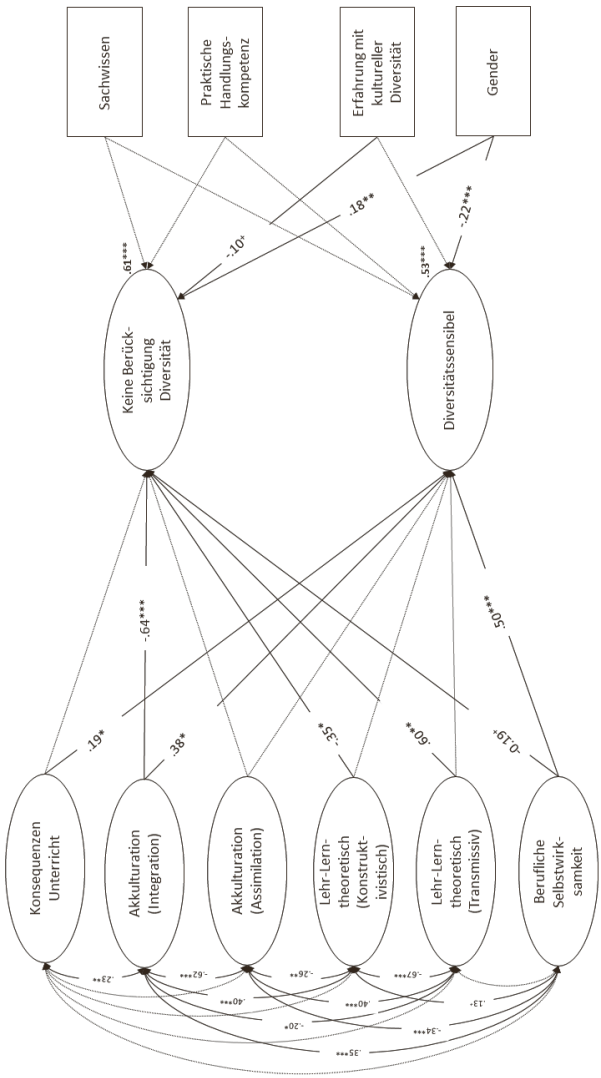
5. Ergebnisse

Anhand der Deskriptiva in Tabelle 1 werden die angenommenen korrelativen Zusammenhänge zwischen den Skalen zur Messung der Überzeugungen und den Skalen zur Messung der intendierten Unterrichtsgestaltung ersichtlich. Es können statistisch signifikante negative Zusammenhänge zwischen der Skala *intendierter Unterricht ohne Berücksichtigung von Diversität* und den Subskalen zur Messung *konstruktivistischer lehr-lerntheoretischer Überzeugungen, Akkulturation* Subskala *Integration* und *berufliche Selbstwirksamkeitsüberzeugung* festgestellt werden und statistisch signifikante positive Zusammenhänge zwischen den Subskalen zur Messung *transmissive lehr-lerntheoretischer Überzeugungen* und *Akkulturation* Subskala *Assimilation*. Statistisch signifikante positive Zusammenhänge können zwischen der Skala zur Messung eines *intendierten diversitätssensiblen Unterrichts*, den Skalen *Konsequenzen für den Unterricht*, *Akkulturation* Subskala *Integration*, *konstruktivistische lehr-lerntheoretische Überzeugungen* und *berufliche Selbstwirksamkeitsüberzeugung* beobachtet werden. Statistisch signifikante negative Zusammenhänge zeigten sich zwischen Subskalen *transmissive lehr-lerntheoretischer Überzeugungen* und *Akkulturation* Subskala *Assimilation*.

Bei simultaner Berücksichtigung der Zusammenhänge in einem Strukturgleichungsmodell mit akzeptablen Fitwerten (χ^2 [970] = 1238.95; CFI = .92; TLI = 91; RMSEA = .03; SRMR = .06) gemessen anhand latenter Variablen können Zusammenhangsgrößen in erwarteter Richtung beobachtet werden, die aufgrund geteilter Varianz geringer ausfallen als die bivariaten korrelativen Zusammenhänge.

Für die Erklärung eines *intendierten Unterrichts ohne Berücksichtigung von Diversität* sind statistisch signifikante negative Pfadkoeffizienten von *Akkulturation* Subskala *Integration, konstruktivistischen lehr-lerntheoretischen Überzeugungen* und *beruflicher Selbstwirksamkeitsüberzeugung* zur abhängigen Variable zu beobachten. Statistisch signifikante positive Pfadkoeffizienten sind zwischen der latenten Variable *intendierter Unterricht ohne Berücksichtigung* und *transmissiven lehr-lerntheoretischer Überzeugungen* festzustellen.

Abbildung 1: Ergebnisse der Analyse Zusammenhänge und Pfadkoeffizienten der untersuchten Konstrukte



Anmerkung: $^{+}p < .10$; $*p < .05$, $**p < .01$, $***p < .001$

Statistisch signifikante positive Pfadkoeffizienten zeigten sich zwischen *Konsequenzen für den Unterricht*, *Akkulturation* Subskala *Integration, berufliche Selbstwirksamkeitsüberzeugung* und der zweiten abhängigen Variable *intendierte diversitätssensible Unterrichtsgestaltung.*

Negative statistisch signifikante Pfadkoeffizienten können zwischen der Erfahrung mit kultureller Diversität und der Variable *intendierter Unterricht ohne Berücksichtigung von Diversität* festgestellt werden sowie zwischen dem *Geschlecht* und der Variable *intendierte diversitätssensible Unterrichtsgestaltung*. Die R^2 Werte weisen darauf hin, dass 61 % und 53 % der Varianz durch die unabhängigen Variable erklärt werden können (Reinecke & Pöge, 2010).

6. Diskussion

Ziel dieser Studie war es, die Überzeugungen von angehenden Lehrkräften in Bezug auf intendiertes Unterrichtsverhalten zu untersuchen. Dabei fanden lehrkraftseitige Überzeugungen Berücksichtigung, die angehende Lehrkräfte bezugnehmend auf die Unterrichtsgestaltung haben, auf schülerseitige Akkulturationsprozesse, auf Wissensaneignungs- und Wissensvermittlungsprozesse sowie in Bezug auf eigene Fähigkeiten, kulturell-ethnische Diversität zu berücksichtigen. Die Intentionen der zukünftigen Unterrichtsgestaltung wurde mit einer Gegenüberstellung von zwei Standpunkten, einen diversitätssensiblen Unterricht zu konzertieren im Gegensatz zu einem Unterricht, in dem Diversität nicht zu berücksichtigt werden soll, ermittelt. Die in dieser Studie gefundenen Ergebnisse entsprachen jedoch nicht allen zuvor formulierten Erwartungen.

Die erste Fragstellung zur Einschätzung des Ausmaßes an Konsequenzen für die Vor-, Nachbereitung und Durchführung von Unterricht für eine kulturell-ethnisch diverse Schülerschaft ist nur für eine diversitätssensible Unterrichtsgestaltung bedeutsam. Angehende Lehrkräfte, die eine realistische Einschätzung dahingehend haben, in welchem Ausmaß die Unterschiede zwischen den Schülerinnen und Schülern die Planung, Nachbereitung und Durchführung von Unterricht beeinflussen, sind diejenigen, die dies im späteren Unterricht umzusetzen beabsichtigen. Dies scheint im Gegensatz zu Studienergebnissen zu stehen, bei denen keine Relevanz der Kosten im Sinne eines zusätzlichen und eher belastend wahrgenommenen Arbeitsaufwandes erfasst wurden (Gebauer & McElvany, 2017; Gebauer & McElvany, 2020). Möglicherweise ist das mit der weniger bewertenden Erfassung der zu erwartenden Konsequenzen für die Unterrichtsgestaltung zu erklären. Gleichzeitig sprechen diese Ergebnisse dafür, dass angehende Lehrkräfte bereits während der universitären Ausbildungsphase eine Einschätzung dahingehend entwickelt haben, welche sozialen Aspekte der Schülerinnen und Schüler für die tägliche schulische Arbeit relevant sind.

Mit Hinblick auf integrative Akkulturationsüberzeugungen der angehenden Lehrkräfte, die mit der zweiten Forschungsfrage in den Blick genommen wurden, zeigen die Daten wie erwartet eine positive Bedeutsamkeit für eine diversitätssensible Unterrichtsgestaltung und negative Zusammenhänge für einen Unterricht, in dem kulturell-ethnische Diversität unberücksichtigt bleibt. Angehende Lehrkräfte, die der Ansicht sind, dass die eigene Herkunftskultur oder die der Eltern und Großeltern neben der Aufnahmekultur bewahrt und wertgeschätzt werden sollte, sind eher dazu bereit, in einem späteren Unterricht die kulturell-ethnischen Unterschiede der Schülerinnen und Schüler wertschätzend im Unterricht wahrzunehmen und einzubinden. Demnach scheint es so zu sein, dass sich dies bei angehenden Lehrkräften, die in Bezug auf Strategien zwischen ethnisch-kulturellen Gruppen eine integrative Vorstellungen haben (Sam & Berry, 2006), in ihrer Unterrichtsgestaltung niederschlägt. Nicht erklärt werden kann, dass trotz statistisch signifikanter bivariater korrelativer Zusammenhänge keine Bedeutsamkeit der assimilierenden Vorstellungen gefunden werden kann. Aus methodischer Sicht könnte ein mediierender Zusammenhang angenommen werden, wenn Pfadkoeffizienten nach Berücksichtigung weiterer Variablen an Bedeutung verlieren (Hayes & Preacher, 2014). Andererseits kann keine theoretische Annahme dahingehend formuliert werden, welche Überzeugung einer anderen Überzeugung übergeordnet ist und diese beeinflusst. Vielmehr besteht weitgehend Konsens dahingehend, dass Überzeugungen eher in wechselseitigen Zusammenhängen stehen und Wissen, Erfahrungen und kontextuelle Faktoren wichtig für die Herausbildung von Überzeugungen sind (Bosse et al., 2017; Levin, 2014). Anschlussfähig sind die Befunde dieser Studie zu anderen Studienergebnissen, die zeigen, dass konstruktivistische lehr-lerntheoretische Vorstellungen zu einer differenzierenden Unterrichtsgestaltung führen.

Sowohl konstruktivistische als auch transmissive lehr-lerntheoretische Überzeugungen der befragten Studierenden, die mit der dritten Forschungsfrage adressiert wurden, sind für eine zukünftige Unterrichtsgestaltung bedeutsam, in der Diversität keine Berücksichtigung findet, nicht aber für eine diversitätssensible Unterrichtsgestaltung. Dabei konnten hier die erwarteten negativen Zusammenhänge zwischen Unterricht, in dem Diversität nicht berücksichtigt werden soll, und konstruktivistischen lehr-lerntheoretischen Überzeugungen sowie positive Zusammenhänge zwischen transmissiven lehr-lerntheoretischen Überzeugungen festgestellt wurden. Die Annahme, dass konstruktivistische Überzeugungen, bei denen Lernprozesse als eigenständige Erkenntnisprozesse angesehen werden, ebenfalls zu einer größeren Berücksichtigung sozialer Merkmale von Schülerinnen

und Schülern im Unterricht führt, kann nicht bestätigt werden. Gleichzeitig beabsichtigen diese Studierenden, Diversität nicht gänzlich unberücksichtigt zu lassen. Dieser Befund ist anschlussfähig zu Studienergebnissen, die lehr-lerntheoretische Überzeugungen auf weitere Aspekte eines erweiterten Inklusionsverständnisses untersucht haben (Börnert-Ringleb et al., 2020). Des Weiteren würde dies für eine differenzierte Vorstellung der angehenden Lehrkräfte in Bezug auf schülerseitige Merkmale sprechen und dafür, dass die Vorstellungen bezugnehmend auf kognitive Lernprozesse der Schülerinnen und Schüler unabhängig von den herkunftsbezogenen Merkmalen gesehen werden. Darüber hinaus sprechen diese Ergebnisse dafür, dass die Ansichten von angehenden Lehrkräften, Diversität zu berücksichtigen oder unberücksichtigt zu lassen, keine austauschbaren Konstrukte sind, bei denen das eine Konstrukt als negativer Pol des anderen betrachtet werden kann (Maggioni & Parkinson, 2008).

Die selbstbezogenen Kompetenzwahrnehmungen in Bezug auf das Unterrichten kulturell-ethnisch diverser Schülergruppen sind für beide Unterrichtsgestaltungen bedeutsam und wurden in der vierten Forschungsfrage adressiert. Dies unterstützt eine Reihe an Forschungsergebnissen, in denen die Bedeutsamkeit der berufsbezogenen Selbstwirksamkeitsüberzeugung für die Unterrichtsgestaltung untersucht wurde (Gebauer & McElvany, 2017; Bosse & Sprörer, 2014). Angehende Lehrkräfte, die eine hohe Kompetenzwahrnehmung haben, sind eher dazu bereit, einen diversitätssensiblen Unterricht anzubieten. Dabei muss relativierend berücksichtigt werden, dass die berufsbezogene Selbstwirksamkeitsüberzeugung bei angehenden Lehrkräften während des Studiums höher ist als bei Lehrkräften in den ersten Jahren im schulischen Dienst (Woolfolk Hoy & Spero, 2005). Dies scheint ebenso für die Selbstwirksamkeitsüberzeugung in Bezug auf inklusives Arbeiten zu gelten (Hecht & Weber, 2020).

Limitationen, die hinsichtlich der Ergebnisse einschränkend berücksichtigt werden müssen, sind unter anderem, dass die Stichprobe dieser Untersuchung keine Aussagen dazu zulässt, ob die angehenden Lehrkräfte ihren Unterricht zukünftig tatsächlich diversitätssensibel gestalten werden. Bekanntlich kann eine Intention lediglich als Indikator für künftiges Verhalten betrachtet werden (Ajzen & Fishbein, 2005). Hier würden Untersuchungen von Lehrkräften im schulischen Dienst in Verbindung mit Unterrichtsbewertungen durch externe Beobachterinnen und Beobachter und schülerseitigen Einschätzungen den Erkenntnisstand in diesem Bereich ergänzen. Darüber hinaus ist bei Befragungen, insbesondere bei normativ aufgeladenen gesellschaftspolitischen Themen wie dem Umgang mit kulturell-ethnischer Diversität, davon auszugehen, dass die Befragten

sozial erwünschte Antworten auf die Fragen geben (Krumpal, 2013). Vergleichende Analysen zwischen impliziten und expliziten Einstellungen in Bezug auf kulturell-ethnische Diversität belegen jedoch keine Unterschiede zwischen explizit und implizit gemessen Einstellungen (Kleen et al., 2019), wobei die zur Diskussion gestellte Validität impliziter Messverfahren relativierend berücksichtigt werden muss (z. B. Kurdi et al., 2021).

Mit den geprüften Konstrukten und Zusammenhängen liefern die Befunde einen Beitrag zur weiteren Untersuchung von Überzeugungen angehender Lehrkräfte. Schließlich kann vor dem Hintergrund reziproker Zusammenhänge untereinander zu weiteren Konstrukten (z. B. Tassell et al., 2020) und über die Zeit (Hecht & Weber, 2020) die Relevanz der universitären Lehrkräfteausbildung, in der Überzeugungen stärker in den Fokus gerückt werden sollten, hervorgehoben werden.

Literaturverzeichnis

Agostinetto, L., & Bugno, L. (2020). Towards congruence between teachers' intentions and practice in intercultural education. *Intercultural Education*, *31*(1), 54–67. https://doi.org/10.1080/14675986.2019.1702261

Alisaari, J., Heikkola, L. M., Commins, N., & Acquah, E. O. (2019). Monolingual ideologies confronting multilingual realities. Finnish teachers' beliefs about linguistic diversity. *Teaching and Teacher Education*, *80*, 48–58. https://doi.org/10.1016/j.tate.2019.01.003

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In: D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *Handbook of attitudes and attitude change* (pp. 173–221). Hillsdale, NJ: Erlbaum.

Apfelbaum, E. P., Norton, M. I., & Sommers, S. R. (2012). Racial color blindness: Emergence, practice, and implications. *Current directions in psychological science*, *21*(3), 205–209. https://doi.org/10.1177/0963721411434980

Asparouhov, T., & Muthén, B. (2010). Multiple imputation with Mplus. *MPlus Web Notes*, 238–246.

Au, K. (2009). Isn't culturally responsive instruction just good teaching?. *Social Education*, *73*(4), 179–183.

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

Banks, J. A. (2016). *Cultural Diversity and Education: Foundations, Curriculum, and Teaching*. New York, NY: Routledge. https://doi.org/10.4324/9781315622255

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

Bayraktar, S. (2011). Turkish preservice primary school teachers' science teaching efficacy beliefs and attitudes toward science: The effect of a primary teacher education program. *School Science and Mathematics, 111*(3), 83–92. https://doi.org/10.1111/j.1949-8594.2010.00065.x

Bello, B., Leiss, D., & Ehmke, T. (2017). Diversitätsbezogene Einstellungen von Lehramtsstudierenden mit und ohne Migrationsgeschichte. *Beiträge zur Lehrerinnen-und Lehrerbildung*, *35*(1), 165–181. https://doi.org/10.25656/01:16799

Berry, J. W. (2005). Acculturation: Living successfully in two cultures. *International Journal of Intercultural Relations*, *29*(6), 697–712. https://psycnet.apa.org/doi/10.1016/j.ijintrel.2005.07.013

Berry, J., & Kalin, R. (1995). Multicultural and ethnic attitudes in Canada: An overview of the 1991 national survey. *Canadian Journal of Behavioural Science*, *27*, 301–320. https://psycnet.apa.org/doi/10.1037/0008-400X.27.3.301

Blömeke, S.; Gustafsson, J.-E.; Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, *223*, 3–13. https://psycnet.apa.org/doi/10.1027/2151-2604/a000194

Bosse, S., Jäntsch, C., Henke, T., Lambrecht, J., Koch, H., & Spörer, N. (2017). Das Zusammenspiel der Offenheit für Innovationen, der Einstellung zum inklusiven Lernen und der Selbstwirksamkeit von Lehrkräften. *Zeitschrift für Bildungsforschung*, *7*(2), 131–146. https://doi.org/10.1007/s35834-017-0185-4

Bosse, S., & Spörer, N. (2014). Erfassung der Einstellung und der Selbstwirksamkeit von Lehramtsstudierenden zum inklusiven Unterricht. *Empirische Sonderpädagogik*, *6*(4), 279–299. https://doi.org/10.25656/01:10019

Börnert-Ringleb, M., Westphal, A., Zaruba, N., Gutmann, F., & Vock, M. (2020). The Relationship Between Attitudes Toward Inclusion, Beliefs About Teaching and Learning, and Subsequent Automatic Evaluations Amongst Student Teachers. *Frontiers in Education 5 (2020)*, *5*, 584464. https://doi.org/10.3389/feduc.2020.584464

Buchholzer, A., & Perlgrims, G. (2013). Editorial. Heterogenität – Zum differenzierten Umgang mit einem komplexen Phänomen. *Schweizerische Zeitschrift für Bildungswissenschaften*, 351(1), 13–19. https://doi.org/10.25656/01:10284

Bundesamt für Migration und Flüchtlinge (2019). Migration, Integration, Asyl Politische Entwicklungen in Deutschland 2018Jährlicher Bericht der deutschen nationalen Kontaktstelle für das Europäische Migrationsnetzwerk (EMN) verfügbar unter: https://www.bamf.de/SharedDocs/Anlagen/DE/EMN/Politikberichte/emn-politikbericht-2018-germany.pdf?__blob=publicationFile&v=10. Gesehen 14.07.2021.

Chan, K. W., & Elliott, R. G. (2004). Relational analysis of personal epistemology and conceptions about teaching and learning. *Teaching and Teacher Education*, *20*(8), 817–831. https://doi.org/10.1016/j.tate.2004.09.002

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*. Oxford, U.K.: Irvington.

Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*, Washington, DC: Center for American Progress. Verfügbar unter: https://files.eric.ed.gov/fulltext/ED535859.pdf. Gesehen 14.07.2021.

Decristan, J., Fauth, B., Kunter, M., Büttner, G., & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research*, *86*, 109–121. https://doi.org/10.1016/j.ijer.2017.09.004

Döhrmann, M., Kaiser, G., & Blömeke, S. (2014). The conceptualization of mathematics competencies in the international teacher education study TEDS-M. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & W. H. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn* (pp. 431–456). https://doi.org/10.1007/978-94-007-6437-8_20

Dumont, H. (2019). Neuer Schlauch für alten Wein? Eine konzeptuelle Betrachtung von in-dividueller Förderung im Unterricht. *Zeitschrift für Erziehungswissenschaft*, *22*(2), 249–277. https://doi.org/10.1007/s11618-018-0840-0

Eagly, A. H., & Chaiken, S. (1993). *The Psychology of Attitudes*. Fort Worth: Harcourt, Brace & Janovich.

Emmerich, M., & Moser, V. (2020). Inklusion, Diversität und Heterogenität in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König, M. Rothland, & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 76–84). Kinkhardt.

Fischer, N., & Ehmke, T. (2019). Empirische Erfassung eines „messy constructs“: Überzeugungen angehender Lehrkräfte zu sprachlich-kultureller Heterogenität in Schule und Unterricht. *Zeitschrift für Erziehungswissenschaft, 22*(2), 411–433. https://doi.org/10.1007/s11618-018-0859-2

Fischer, N., Hammer, S., & Ehmke, T. (2018). Überzeugungen zu Sprache im Fachunterricht: Erhebungsinstrument und Skalendokumentation. In T. Ehmke, S. Hammer, A. Köker, U. Ohm, & B. Koch-Priewe (Hrsg.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 149–184). Münster: Waxmann.

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers’ beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA handbooks in psychology. APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (p. 471–499). American Psychological Association. https://psycnet.apa.org/doi/10.1037/13274-019

Fives, H., & Buehl, M. M. (2016). Teachers’ beliefs, in the context of policy reform. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 114–121. https://doi.org/10.1177/2372732215623554

Fylkesnes, S. (2018). Whiteness in teacher education research discourses: A review of the use and meaning making of the term cultural diversity. *Teaching and Teacher Education*, *71*, 24–33. https://doi.org/10.1016/j.tate.2017.12.005

Gaertner, S., & Dovidio, J. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia, PA: Psychology Press. https://doi.org/10.4324/9781315804576

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorisation and the reduction of intergroup bias. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 4, pp. 1–26). Chichester, UK: Wiley. https://doi.org/10.1080/14792779343000004

Gao, W., & Mager, G. (2011). Enhancing preservice teachers’ sense of efficacy and attitudes toward school diversity through preparation: A case of one US inclusive teacher education program. *International Journal of Special Education*, *26*(2), 92–107.

Gebauer, M. M., McElvany, N., & Klukas, S. (2013). Einstellungen von Lehramtsanwärterinnen und Lehramtsanwärtern zum Umgang mit heterogenen Schülergruppen in Schule und Unterricht. In N. McElvany, M. M. Gebauer, W. Bos, & H. G. Holtappels (Hrsg.), *Sprachliche, kulturelle und soziale Heterogenität in der Schule als Herausforderung und Chance der Schulentwicklung* (IFS-Jahrbuch der Schulentwicklung, Bd. 17, S. 191–219). Juventa.

Gebauer, M. M., & McElvany, N. (2017). Zur Bedeutsamkeit unterrichtsbezogener heterogenitätsspezifischer Einstellungen angehender Lehrkräfte für zukünftiges Unterrichtsverhalten. *Psychologie in Erziehung und Unterricht, 64*(3), 163–180. http://dx.doi.org/10.2378/peu2017.art11d

Gebauer, M. M., & McElvany, N. (2020). Einstellungen und Motivation bezogen auf kulturell-ethnisch heterogene Schülerrinnen- und Schülergruppen und ihre Bedeutung für differenzielle Instruktion im Unterricht. *Zeitschrift für Erziehungswissenschaft*, *23*, 685–708. https://doi.org/10.1007/s11618-020-00956-8

Gillborn, D. (2019). Hiding in Plain Sight. Understanding and Addressing Whiteness and Color-Blind Ideology in Education, *Kappa Delta Pi Record*, *55*(3), 112–117. https://doi.org/10.1080/00228958.2019.1622376

Glock, S. (2016). Does ethnicity matter? The impact of stereotypical expectations on in-service teachers’ judgments of students. *Social Psychology of Education*, *19*(3), 493–509. https://doi.org/10.1007/s11218-016-9349-7

Glock, S., Kovacs, C., & Pit-ten Cate, I. (2019). Teachers’ attitudes towards ethnic minority students: Effects of schools’ cultural diversity. *British Journal of Educational Psychology*, *89*(4), 616–634. https://doi.org/10.1111/bjep.12248

Göbel, K., & Buchwald, P. (2017). *Interkulturalität und Schule: Migration – Heterogenität – Bildung* (Vol. 4642). Paderborn: UTB. https://doi.org/10.36198/9783838546421

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher education*, *26*(4), 1094–1103. https://doi.org/10.1016/j.tate.2009.11.005

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. https://doi.org/10.1016/j.tate.2015.02.001

Hachfeld, A., Hahn, A., Schroeder, S., Anders, Y., Stanat, P., & Kunter, M. (2011). Assessing teachers' multicultural and egalitarian beliefs: The teacher cultural beliefs scale. *Teaching and Teacher education*, *27*(6), 986–996. https://doi.org/10.1016/j.tate.2011.04.006

Hachfeld, A., & Syring, M. (2020). Stichwort: Überzeugungen von Lehrkräften im Kontext migrationsbezogener Heterogenität. *Zeitschrift für Erziehungswissenschaft*, *23*, 659–684. https://doi.org/10.1007/s11618-020-00957-7

Hammer, S., Fischer, N., & Koch-Priewe, B. (2016). Überzeugungen von Lehramtsstudierenden zu Mehrsprachigkeit in der Schule. *Die Deutsche Schule*, Beiheft 13, S. 147–171.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British journal of mathematical and statistical psychology*, *67*(3), 451–470. https://doi.org/10.1111/bmsp.12028

Hecht, P., & Weber, C. (2020). Inklusionsrelevante Selbstwirksamkeitsüberzeugungen und Einstellungen von Studierenden und Lehrkräften im Berufseinstieg – Entwicklung und Zusammenhänge im Längsschnitt. *Zeitschrift für Bildungsforschung*, 1–19.

Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist*, *37*(3), 137–150.

Hermans, R., van Braak, J., & Van Keer, H. (2008). Development of the beliefs about primary education scale: Distinguishing a developmental and transmissive dimension. *Teaching and Teacher Education*, *24*(1), 127–139.

Hofer, B. K. (2002). Epistemological World Views of Teachers: from Beliefs to Practice. *Issues in Education*, *8*(2).

Hogg, M. A. (2005). Uncertainty, social identity and ideology. In S. R. Thye & E. J. Lawler (Eds.), *Advances in Group Processes* (Vol. 22, pp. 203–230). New York: Elsevier.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, *6*(1), 1–55.

Hurtado, S. (1996). How Diversity Affects Teaching and Learning. *Educational Record*, *66*(4), 27–29.

IBM SPSS Statistics (2020). *IBM SPSS Advanced Statistics Version 27*. Verfügbar unter: https://www.ibm.com/docs/en/SSLVMB_27.0.0/pdf/de/IBM_SPSS_Advanced_Statistics.pdf. Gesehen: 14.07.2021.

Kelley, R. D. G. (1999). The people in me. *Utne Reader*, 95, 79–81.

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, *12*, 59–76.

Kleen, H., Bonefeld, M., Glock, S., & Dickhäuser, O. (2019). Implicit and explicit attitudes toward Turkish students in Germany as a function of teachers' ethnicity. *Social Psychology of Education*, *22*(4), 883–899.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47*(4), 2025–2047.

Kurdi, B., Carroll, T. J., & Banaji, M. (2021). Specificity and incremental predictive validity of implicit attitudes: studies of a race-based phenotype. *Cognitive Research: Principles and Implications, 61*(6). 1-21. https://doi.org/10.1186/s41235-021-00324-y

Levin, B. B. (2014). The development of teachers' beliefs. In H. Fives, & M. G. Gill, (Eds.). (2014). *International Handbook of Research on Teachers' Beliefs* (pp. 60–77). New York: Routledge.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Maggioni, L., & Parkinson, M. M. (2008). The role of teacher epistemic cognition, epistemic beliefs, and calibration in instruction. *Educational Psychology Review*, *20*(4), 445–461.

Maio, G. R., & Haddock, G. (2009). The psychology of attitudes and attitude change. Los Angeles: SAGE.

Markus, H. R., Steele, C. M., & Steele, D. M. (2000). Colorblindness as a barrier to inclusion: Assimilation and nonimmigrant minorities. *Daedalus*, *129*, 233–259.

Maruyama, G., Moreno, J. F., Gudeman, R. H., & Marin, P. (2000). *Does diversity make a difference? Three research studies on diversity in college classrooms*. American Council on Education, Washington, DC.; American Association of Univ. Professors, Washington, DC. verfügbar unter: https://files.eric.ed.gov/fulltext/ED444409.pdf. Gesehen: 08.08.2021.

Mellom, P. J., Straubhaar, R., Balderas, C., Ariail, M., & Portes, P. R. (2018). "They come with nothing:" How professional development in a culturally responsive pedagogy shapes teacher attitudes toward Latino/a English language learners. *Teaching and Teacher Education*, *71*, 98–107.

Muthén, L. K., & Muthén, B. (1998-2017). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

Olafson, L., & Schraw, G. (2006). Teachers' beliefs and practices within and across domains. *International Journal of Educational Research*, *45*(1–2), 71–84.

Oser, F., & Blömeke, S. (2012). Überzeugungen von Lehrpersonen. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, *58*(4), 415–421.

Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research, 62*(3), 307–332. https://doi.org/10.3102/00346543062003307

Parekh, B. (1999). A varied moral world. In J. Cohen, M. Howard, & M. Nussbaum (Eds.), *Is multiculturalism bad for women?* (pp. 69–75). Princeton: Princeton University Press.

Plaut, V. C., Thomas, K. M., & Goren, M. J. (2009). Is multiculturalism or color blindness better for minorities? *Psychological Science*, *20*, 444–446.

Prengel, A. (2007). Diversity Education – Grundlagen und Probleme der Pädagogik der Vielfalt. *Diversity studies. Grundlagen und disziplinäre Ansätze*, 49–67.

Rakoczy, K., Buff, A., Lipowsky, F. (2005). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie. „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis". 1. Befragungsinstrumente*. Frankfurt am Main: GFPF u. a.

Reinecke, J., & Pöge, A. (2010). Strukturgleichungsmodelle. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 775–804). Wiesbaden: VS Verlag für Sozialwissenschaften.

Reusser, K., Pauli, C., & Elmer, A. (2011). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz, & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 478–495). Münster: Waxmann.

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color blindness on racial bias. *Journal of Experimental Social Psychology*, *40*, 417–423.

Rosenthal, L., & Levy, S. R. (2010). The colorblind, multicultural, and polycultural ideological approaches to improving intergroup attitudes and relations. *Social Issues and Policy Review*, *4*(1), 215–246.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Ryan, C. S., Casas, J. F., & Thompson, B. K. (2010). Interethnic ideology, intergroup perceptions, and cultural orientation. *Journal of Social Issues*, *66*, 29–44

Sam, D. L., & Berry, J. W. (2006). The Cambridge Handbook of Acculturation Psychology. Cambridge: Cambridge University Press.

Schütze, C., & Törngren, S. O. (2022). Exploring ways of measuring colour-blindness in Sweden. Oerationalisation and theoretical understandings of a US concept in a new context. *Ethnic and Racial Studies*, *45*(16), 637–658.

Schwarzer, R., & Schmitz, G. S. (1999). Skala Lehrer-Selbstwirksamkeitserwartung. In R. Schwarzer & M. Jerusalem (Hrsg.), *Skalen zur Erfassung von Lehrer- und Schülermerkmalen* (S. 60–61). Berlin.

Sharma, U., Forlin, C., & Loreman, T. (2008). Impact of training on pre-service teachers' attitudes and concerns about inclusive education and sentiments about persons with disabilities. *Disability & Society, 23,* 773–785.

Sheehy, K., Budiyanto, Kaye, H., & Rofiah, K. (2019). Indonesian teachers' epistemological beliefs and inclusive education. *Journal of Intellectual Disabilities*, *23*(1), 39–56.

Silverman, J. C. (2007). Epistemological beliefs and attitudes toward inclusion in pre-service teachers. *Teacher Education and Special Education*, *30*(1), 42–51.

Siwatu, K. O. (2011a). Preservice teachers' culturally responsive teaching self-efficacy-forming experiences: A mixed methods study. *The Journal of Educational Research*, *104*(5), 360–369.

Siwatu, K. O. (2011b). Preservice teachers' sense of preparedness and self-efficacy to teach in America's urban and suburban schools: Does context matter? *Teaching and Teacher Education*, *27*(2), 357–365.

Snow, R. (1989). Aptitude-Treatment Interaction as a framework for research on individual differences in learning. In P. Ackerman, R. J. Sternberg, & R. Glaser (ed.), *Learning and Individual Differences.* New York: W.H. Freeman.

Stanat, P., & Edele, A. (2011). Migration und soziale Ungleichheit. In H. Reinders, D. Ditton, G. Gräsel, & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche* (S. 181–192). Wiesbaden: VS Verlag für Sozialwissenschaft.

Statistisches Bundesamt (2021). *3 Bildung. Auszug aus dem Datenreport 2021*. https://www.destatis.de/DE/Service/Statistik-Campus/Datenreport/Downloads/datenreport-2021-kap-3.pdf?__blob=publicationFile

Steinberg, S. R., & Kincheloe, J. L. (2001). Setting the context for critical multi/interculturalism: The power blocs of class elitism, white supremacy, and patriarchy. In S. R. Steinberg (Ed.), *Multi/intercultural conversations* (pp. 2–30). New York: Peter Lang.

Stephan, W. G., & Stephan, C. W. (2001). *Improving intergroup relations*. Thousand Oaks, CA: Sage

Tassell, J., Gerstenschlager, N. E., Syzmanski, T., & Denning, S. (2020). A study of factors impacting elementary mathematics preservice teachers: Improving mindfulness, anxiety, self-efficacy, and mindset. *School Science and Mathematics*, *120*(6), 333–344.

Tichnor-Wagner, A., Parkhouse, H., Glazier, J., & Cain, J. M. (2019). *Becoming a Globally Competent Teacher.* Alexandria, VA: ASCD.

Tomcho, T. J., & Foels, R. (2002). Teaching acculturation: Developing multiple "cultures" in the classroom and role-playing the acculturation process. *Teaching of Psychology*, *29*(3), 226–229.

Trautmann, M., & Wischer, B. (2011). *Heterogenität in der Schule. Eine kritische Einführung.* Wiesbaden: VS.

Tschannen-Moran, M., & Woolfolk Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805.

van Dick, R., Wagner, U., Adams, C., & Petzel, T. (1997). Einstellungen zur Akkulturation: Erste Evaluation eines Fragebogens an sechs deutschen Stichproben. *Gruppendynamik, 28*(1), 83–92.

Voss, T., Kleickmann, T., Kunter, M., & Hachfeld, A. (2013). Mathematics teachers' beliefs. In M. Kunter (Ed.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* (pp. 249–271). Springer: Boston, MA.

Wagner, U., van Dick, R., Petzel, T., & Auernheimer, G. (2001). Der Umgang von Lehrerinnen und Lehrern mit interkulturellen Konflikten. In G. Auernheimer, R. Van Dick, T. Petzel, & U. Wagner (Hrsg.), *Interkulturalität im Arbeitsfeld Schule* (pp. 17–40). Opladen: Leske + Buderich.

Weidemann, D. (2007). Akkulturation und interkulturelles Lernen. In J. Straub, A. Weidemann, & D. Weidemann (Hrsg.), *Handbuch interkulturelle Kommunikation und Kompetenz* (S. 488–498). Stuttgart, Germany: JB Metzler.

Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology*, *78*, 635–654.

Woolfolk Hoy, A., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: acomparison of four measures. *Teaching and Teacher Education*, *21*(4), 343–356.

Prozessbezogene Diagnosequalität im Kontext von Schullaufbahnentscheidungen

Stefanie van Ophuysen, Lars Behrmann & Sina Schürer

Zusammenfassung

Übergänge in der Schullaufbahn sind wichtige Weichenstellungen für die Bildungsbiografie der Schüler*innen. Für lehrkraftseitige Einschätzungen der Schulformeignung ist die Frage nach der Qualität des zugrunde liegenden Diagnoseprozesses von besonderer Bedeutung. Aufbauend auf dem Vierkomponentenmodell der Diagnosequalität (Behrmann & van Ophuysen, 2017) werden Prozessmerkmale einer qualitativ hochwertigen Diagnostik herausgearbeitet. Dabei stehen die Schullaufbahnempfehlungen am Ende von Grundschulzeit sowie Erprobungsstufe als diagnostische Anlässe im Fokus. Der Beitrag geht folgenden Fragen nach: Als wie wichtig erachten Lehrkräfte die Aspekte prozessbezogener Diagnosequalität? Welche Aspekte finden Eingang in ihr diagnostisches Handeln? Wie hoch ist die selbst wahrgenommene Diagnosequalität? Die Daten einer schriftlichen Befragung von Grundschullehrkräften und Lehrkräften aus (Ganztags-)Gymnasien zeigen, dass (Selbst-)Einschätzungen von Wichtigkeit und Qualität in beiden Schulstufen gleichermaßen hoch ausfallen. Schulstufenunterschiede zeigen sich bei der selbstberichteten Umsetzung verschiedener Prozessmerkmale. Die Befunde werden vor dem Hintergrund professionellen Lehrkrafthandelns diskutiert.

Schlagworte: Diagnostik, Diagnosequalität; Übergänge; Grundschule; Erprobungsstufe

Abstract

School transitions are important milestones for the (educational) biography of students. Teachers‘ assessment of students‘ qualifications for different

school tracks and the quality of the underlying diagnostic process are particularly important. Drawing on the Four-Component-Model of diagnostic quality (Behrmann & van Ophuysen, 2017), we identify process features that are characteristic of a high-quality diagnosis. We focus on school track recommendations at the end of primary school and „probationary stage" (after grade 6). This paper asks: How important are features of diagnostic process quality from the teachers' perspective? Which aspects are evident in their diagnostic activities? How well rate teachers their own diagnostics? Data from a survey of primary school and (all-day) high school („Gymnasium") teachers reveal that teachers' ratings of (self-)perceived importance and quality are equally high in both levels of school,. However, school level specific differences emerge in the self-reported implementation of various process features. The findings are discussed in terms of professional teacher action.

Keywords: diagnosis; diagnostic quality; transition; primary school; probational stage

1. Einleitung

Die Bildungsbiografie eines jeden Menschen ist durch zahlreiche Übergänge geprägt. Hierzu zählen bspw. der Übergang von der Kita in die Grundschule, von der Grundschule in die weiterführende Schule, ggf. ein Wechsel in einen anderen Bildungsgang innerhalb der Sekundarstufe I oder der Wechsel von der Schule in die Ausbildung oder das Studium. Diesen Übergängen gehen entsprechende Übergangsentscheidungen voraus, die von Eltern oder Lehrkräften – bestenfalls in Übereinstimmung mit den Kindern – getroffen werden müssen. In unserem Beitrag fokussieren wir die Übergangsentscheidung der Lehrkräfte in Nordrhein-Westfalen (NRW) von der Grund- in die weiterführende Schule sowie die Entscheidung am Ende der Erprobungsstufe in Klasse 6 auf dem Gymnasium.

2. Übergänge in der Bildungslaufbahn

2.1. Rechtliche Rahmenbedingungen

In den meisten Bundesländern, so auch in NRW, wechseln Kinder am Ende der vierten Klasse auf die weiterführende Schule. Je nach Bundesland stellt die Schulformempfehlung der Lehrkraft am Ende von Klasse 4 eine Vorgabe oder

eine Orientierung für die Entscheidung der Eltern dar (Porsch, 2018). In einigen Bundesländern wird die Empfehlung für den Besuch einer bestimmten Schulform basierend auf den Noten in den Hauptfächern getroffen (z. B. Schulordnung für die Grundschulen in Bayern, § 6). In NRW erstellt die Grundschule mit dem Halbjahreszeugnis der Klasse 4 eine begründete Empfehlung für die weiterführende Schulform (SchulG NRW, § 11, Abs. 5; Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (MSB NRW), 2022b, AO-GS, § 8). Charakteristisch für NRW – aber auch für viele andere Bundesländer – ist, dass die Schulformempfehlung der Grundschule nicht bindend ist. Nach der verbindlichen Beratung mit der Klassenlehrkraft zur weiteren schulischen Förderung (MSB NRW, 2022b, AO-GS, § 8) entscheiden die Eltern über den weiteren Bildungsgang ihrer Kinder (SchulG NRW, § 11, Abs. 6.). Wenngleich die Lehrkrafteinschätzung am Ende der Grundschulzeit in den meisten Bundesländern nur empfehlenden Charakter hat, folgen Eltern dieser häufig (MSB NRW, 2022a; Pohlmann, 2009). Insbesondere der Gymnasialempfehlung folgen die meisten Eltern (zusammen mit der eingeschränkten Empfehlung: ~93 %). Am geringsten fällt die Übereinstimmung mit der Empfehlung für die Hauptschule aus (~83 %). Eltern wählen als Alternative zu dieser Schulform zumeist ein integriertes Schulsystem wie die Sekundar- oder Gesamtschule (MSB NRW, 2022a).

In der weiterführenden Schule durchlaufen die Schüler*innen in NRW, aber auch in anderen Bundesländern (z. B. Ministerium für Bildung und Kindertagesstätten des Landes Mecklenburg-Vorpommern, SchulG Mecklenburg-Vorpommern, § 15 Orientierungsstufe), zunächst die sog. Orientierungs- oder Erprobungsstufe in den Klassen 5 und 6. Wie der Name sagt, dienen diese zwei Jahre der „Erprobung, Förderung und Beobachtung der Schülerinnen und Schüler, um in Zusammenarbeit mit den Eltern die Entscheidung über die Eignung der Schülerinnen und Schüler für die gewählte Schulform sicherer zu machen." (SchulG NRW, § 13, Abs. 2). Halbjährlich berät die Erprobungsstufenkonferenz, ob ein Schüler oder eine Schülerin in einer anderen Schulform besser gefördert werden kann und empfiehlt den Eltern ggf. den Wechsel der Schulform (MSB NRW, 2002c, APO-S I NRW, § 11, Abs. 1). Mit dem Ende der Erprobungsstufe steht eine endgültige Überprüfung der Schulformwahl an. Die Erprobungsstufenkonferenz prüft „unter Berücksichtigung des Leistungsstandes, der bisherigen von der Schule durchgeführten Fördermaßnahmen und der zu erwartenden Entwicklung der Schülerin oder des Schülers, ob die gewählte Schulform weiterhin besucht oder die Schulform gewechselt werden soll" (MSB NRW, 2022c, APO-S I NRW, § 11, Abs. 1). Für das Gymnasium bedeutet dies, dass die Klassenkonferenz über die Eignung der Kinder entscheidet und beschließt,

ob die Kinder den gewählten Bildungsgang fortsetzen können (SchulG NRW, § 13, Abs. 2). Bei Nichteignung wechseln die Schüler*innen nach Wahl der Eltern entweder an die Real- oder die Hauptschule. Das Elternwahlrecht ist jedoch dann eingeschränkt, wenn die Versetzungskonferenz feststellt, dass keine Eignung für die Realschule vorliegt. In diesem Fall erfolgt ein Wechsel zur Hauptschule (MSB NRW, 2002c, APO-S I NRW, § 11, Abs. 3).

2.2. Relevanz des Übergangs

Diese beiden Übergänge sind wichtige Weichenstellungen für die Bildungsbiografie der Schüler*innen. Besonders bedeutsam ist der Übergang von der Grundschule zur weiterführenden Schule. Rund 700.000 Schüler*innen in Deutschland wechseln jährlich von der Grund- auf eine weiterführende Schule (Statistisches Bundesamt, 2021), davon allein in NRW über 155.000 Schüler*innen (MSB NRW, 2022a). Die besuchte weiterführende Schulform ist maßgeblich für den Bildungsabschluss und damit für den späteren beruflichen Werdegang der Schüler*innen (z. B. Baumert & Schümer, 2001; Merkens & Wessel, 2002). Je nach weiterführendem Schultyp werden spezifische Lern- und Entwicklungsmöglichkeiten geboten, die sich auf den Kompetenzerwerb auswirken (z. B. Baumert et al., 2009). Maaz et al. (2006, S. 322) bezeichnen den Grundschulübergang als „eine der wichtigsten Statuspassagen im Leben eines jungen Menschen".

Die Formation der Übergangsempfehlung stellt somit eine Aufgabe mit weitreichenden Folgen für die Bildungsbiografie der Schüler*innen dar. Aber auch die Entscheidung über den Verbleib oder das Verlassen der weiterführenden Schule am Ende der Erprobungsstufe stellt eine für den weiteren Bildungsverlauf relevante Eignungseinschätzung durch die Lehrkräfte dar. Neben den langfristigen Folgen sind hier aber insbesondere auch die sozial-emotionalen Konsequenzen einer „Abschulung" in den Blick zu nehmen. Auch wenn mit dem Wechsel vom Gymnasium an eine andere Schulform mittelfristig möglicherweise negative Aspekte des Besuchs des Gymnasiums, wie ein zu hoher Leistungsdruck, wegfallen, so bedeutet die Entscheidung der Nichteignung für das Kind zunächst einmal ein Scheitern und ist häufig mit Selbstzweifeln verbunden. Zudem hängt mit dem Wechsel ein neues soziales Umfeld, konkret eine neue Gruppe von bereits bestehenden Peers zusammen, in welche das Kind als Neuankömmling hineinfinden muss. Studien zu Effekten des Schulformwechselns auf sozial-emotionale Variablen sind den Autor*innen nicht bekannt. Die Situation ist aber vergleichbar mit dem emotionalen Erleben bei Klassenwiederholungen. Die

Befundlage hierzu ist uneindeutig. In einer älteren qualitativen Studie zum Thema Klassenwiederholung zeigt Bynres (1989) negative Emotionen der Klassenwiederholer*innen auf. Diese schämen sich, sind traurig und das Sitzbleiben wird als Versagen gewertet. Auch die Meta-Analyse von Jimerson (2001) zeigt für Klassenwiederholer*innen für sozial-emotionale Variablen negativere Werte im Vergleich zu einer Kontrollgruppe. Sechstklässler*innen zählen Klassenwiederholungen zu den stressreichsten Ereignissen, die sie sich vorstellen können. (Anderson et al., 2005). Brophy (2006, S. 16) fasst die internationale Befundlage mit den Worten zusammen: „School-imposed grade repetition is stressful to students and associated with reduced self-esteem, impaired peer relationships, alienation from school, and sharply increased likelihood of eventual dropout". In einer aktuelleren Studie mit einem großen NEPS-Datensatz wurde der Effekt der Klassenwiederholung auf das schulische Wohlbefinden von Schüler*innen der Klassen 5 bis 12 längsschnittlich untersucht. Entgegen der Erwartungen zeigen sich hier keine negativen Effekte der Klassenwiederholung. Insbesondere für Gymnasiasten zeigt sich ein positiver unmittelbarer, kurz-, mittel- und langfristiger Effekt auf die Schulzufriedenheit (Rathmann et al., 2020). Die Autor*innen erklären dies aber ggf. mit einem hohen Anteil an freiwilligen Rücktritten.

Das Verlassen des Gymnasiums bei einer Nicht-Eignung erfolgt jedoch nicht freiwillig und da nicht nur die Klasse, sondern das komplette schulische Umfeld wechselt, ist eher mit noch stärkeren negativen sozial-emotionalen Konsequenzen zu rechnen, als bei der (unfreiwilligen) Klassenwiederholung.

Aufgrund der hohen Relevanz dieser Übergangsentscheidungen und ihrer Funktion als Gelenkstelle im Bildungssystem wäre zu erwarten, dass diese so akkurat wie möglich nach klaren Regelungen erfolgen. Während es in NRW für den Abschluss der Erprobungsstufe auf den ersten Blick sehr konkrete Vorgaben gibt, sind die Kriterien, auf die Lehrkräfte am Ende von Klasse 4 bei der Schulformempfehlung zurückgreifen können, sehr vage.

Die Versetzungsordnung der Erprobungsstufe regelt in NRW den Übergang in Klasse 7 und einen möglichen Schulformwechsel (Versetzungsordnung in der Erprobungsstufe, APO S I, NRW). Ausschlaggebend sind hierbei die Noten, in welche zumeist schriftliche und „sonstige Leistungen im Unterricht" einfließen. Die weiteren Vorgaben hierzu sind jedoch sehr vage gehalten und es ist unstrittig, dass die Vergabe von Noten ein komplexer Prozess ist, der von Lehrkräften individuell unterschiedlich ausgestaltet wird (Lintorf, 2012; Kap. 4.1). Diskussionen um die Güte von Schulnoten haben eine lange Tradition (vgl. Ingenkamp, 1975, 1995). Auch wenn Schulnoten in Studien zur prognostischen Validität relativ gut abschneiden

(z. B. als Prädiktor des Schul- oder Studienerfolgs), so sind „Korrelationen von r = .50 oder r = .60, wie sie für die prognostische Validität und auch für die Übereinstimmungsvalidität die Regel sind, wenig wünschenswert", wenn man sich die individuellen Konsequenzen der Notengebung vergegenwärtigt (Lintorf, 2012, S. 51).

Für die Empfehlung am Ende der Grundschulzeit führt das Schulgesetz (SchulG NRW, § 11, Abs. 5) als Kriterien den Leistungsstand, die Lernentwicklung und die Fähigkeiten der Schüler*innen an. Das Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK, 2015) spricht von der Empfehlung des Bildungsweges, der der Bildungsfähigkeit der Kinder entspricht und führt weiter aus, dass „auch Eignung, Neigung und Wille des Kindes zu geistiger Arbeit insgesamt zu werten" sind (KMK, 2015). Weiter wird formuliert, dass „nicht nur die Leistungen in Bezug auf die fachlichen Ziele der Lehrpläne, sondern auch die für den Schulerfolg wichtigen allgemeinen Fähigkeiten" (KMK, 2015) zu berücksichtigen sind. Nähere Ausführungen hierzu fehlen. Da es keine konkreteren Hinweise hierzu gibt, ist die Qualität des zugrunde liegenden Diagnoseprozesses von besonderer Bedeutung.

Zusammenfassend ist festzuhalten, dass aufgrund der Bedeutsamkeit der Entscheidungen Lehrkräften eine hohe Verantwortung zukommt. Daher sollten sie eine möglichst qualitätsvolle Diagnostik anstreben.

3. Pädagogische Diagnostik

Die Empfehlungen zum Schulformwechsel am Ende von Grundschulzeit sowie Erprobungsstufe erfordern von den Lehrkräften eine gute Kenntnis des individuellen Kindes. Dieses Wissen sollte jeweils das Ergebnis eines professionellen, pädagogischen diagnostischen Prozesses sein. Diagnostik beschreibt allgemein einen Prozess, bei dem Informationen über Personen erhoben werden, sodass Unterschiede oder Gemeinsamkeiten zwischen verschiedenen Individuen eindeutig festgestellt werden können (Fisseni, 2004, S. 4). Nach Klauer (1982, S. 5) kann diese Informationserfassung dann als „pädagogische" Diagnostik bezeichnet werden, wenn sie „im Dienst aktueller pädagogischer Entscheidungen" steht. Zur pädagogischen Diagnostik zählt ein Akt der Informationsbeschaffung also erst dann, wenn mit der gewonnenen Information unmittelbar pädagogisch gearbeitet wird.

Ingenkamp und Lissmann (2008) spezifizieren die Inhalte und Ziele der Anschlusshandlungen weiter aus und fokussieren dabei insbesondere auf das Lernen der Schüler*innen. Sie verstehen unter pädagogischer Diag-

nostik „alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren" (S. 13). Als weitere Ziele benennen sie „die Zuweisung zu Lerngruppen oder zu individuellen Förderungsprogrammen […] sowie die mehr gesellschaftlich verankerten Aufgaben der Steuerung des Bildungsnachwuchses oder der Erteilung von Qualifikationen" (ebd.). Diese Definition umfasst sowohl Elemente formativer (Optimierung individuellen Lernens) als auch summativer Diagnostik (Qualifikationsentscheidungen).

Lintorf et al. (2016) beschreiben die im Rahmen des diagnostischen Prozesses stattfindenden Tätigkeiten etwas genauer, indem sie pädagogische Diagnostik definieren als *„das systematische, zielbezogene Erfassen („wahrnehmen"), das Dokumentieren und Verfügbarhalten („behalten und erinnern") sowie Analysieren und Verdichten („verarbeiten") von Informationen über individuelle Schüler:innen und Gruppen von Schüler:innen, mit dem Ziel, Entscheidungen im Rahmen der professionellen Aufgaben als Lehrkraft (erziehen, unterrichten, bewerten, beraten) begründet vornehmen zu können"*.

Der Schulalltag birgt eine Vielzahl diagnostischer Situationen, die von Lehrkräften bewältigt werden müssen (vgl. Karst et al., 2017). Die Erteilung der Übergangsempfehlung bzw. die Entscheidung über den Schulformwechsel respektive -verbleib sind konkrete diagnostische Anlässe, bei denen ein komplexes Urteil auf Basis vielfältiger Informationen gefällt wird. So belegt die empirische Forschung zu den Kriterien der Übergangsempfehlung zwar, dass die schulischen Leistungen bzw. Noten als wichtigste Prädiktoren gelten können (Glock et al., 2013). Regressionsanalytische Studien zeigen aber darüber hinaus, dass Arbeitstugend und motivational-affektive Schülermerkmale auch bei Kontrolle der Leistung einen eigenständigen Beitrag zur Vorhersage der Empfehlung leisten (Arnold et al., 2007; Stubbe & Bos, 2008), während das Sozialverhalten keinen inkrementellen Beitrag liefert (Klapproth et al., 2013; Stahl, 2007) und weitere Persönlichkeitsmerkmale kaum untersucht werden (siehe aber Lintorf & van Ophuysen, 2022). Werden die Lehrkräfte hingegen selbst gebeten, diejenigen Kriterien zu nennen, die sie für die Formation der Empfehlung heranziehen, so wird dem Sozialverhalten und Persönlichkeitsmerkmalen des Kindes (z. B. Extraversion, Gelassenheit) durchaus Bedeutsamkeit zugeschrieben (Diebig, 2016; Pohlmann, 2009). Ebenfalls berichten Lehrkräfte, dass sie die wahrgenommenen elterlichen Unterstützungsmöglichkeiten als wichtiges Entscheidungskriterium berücksichtigen (Sneyers et al., 2019), insbesondere bei Kindern mit

uneindeutigem Leistungsprofil (Hollstein, 2008; Wiedenhorn, 2011). Merkmale des familiären Kontextes wie Bildungs- oder Migrationshintergrund werden hingegen nicht als bedeutsame Empfehlungskriterien benannt, obwohl sich zumindest der Bildungshintergrund in regressionsanalytischen Studien wiederholt als wichtiger statistischer Prädiktor der Empfehlung herausgestellt hat (siehe zusammenfassend Glock et al., 2013). Gerade die Interviewstudien mit Lehrkräften verdeutlichen, dass sie die Formation der Übergangsempfehlung als eine Aufgabe ansehen, bei der vielfältige Schüler*innenmerkmale (und zum Teil auch Merkmale der Familie) zu einem Gesamturteil integriert werden müssen. Ein systematischer diagnostischer Prozess ist somit unerlässlich, um die vielfältigen Informationen valide zu erfassen und adäquat zu integrieren. Doch wie gut bewältigen Lehrkräfte diese anspruchsvolle Aufgabe?

4. Diagnostische Kompetenz

Unter dem Stichwort der „diagnostischen Kompetenz" wird in der empirischen Forschung überprüft, wie gut Lehrkräften das Diagnostizieren gelingt. Dabei steht häufig die Urteilsakkuratheit als Indikator der Diagnosekompetenz im Fokus. Dies erscheint auf den ersten Blick naheliegend, da hohe Kompetenz die Wahrscheinlichkeit eines guten – also korrekten, akkuraten – Urteils impliziert, während fehlerhafte Urteile nur sehr selten vorkommen sollten. Als klassisches Beispiel kann die Einschätzung der Fachleistung von Schüler*innen gelten. Diese Lehrkrafturteile werden mit der tatsächlichen Leistung der Kinder verglichen, die diese in der Regel in einem standardisierten Leistungstest erbracht haben. In empirischen Untersuchungen wird die diagnostische Kompetenz der Lehrkräfte typischerweise über drei Komponenten operationalisiert (Schrader, 1989):

- Rangkomponente: Wie gut stimmt die Rangfolge gemäß der Lehrkrafteinschätzung mit der Rangfolge auf Basis der tatsächlich erbrachten Leistung überein? (Korrelation der Messwerte)
- Niveaukomponente: Wie gut stimmen die mittlere Leistungseinschätzung und die mittlere erbrachte Leistung über alle Kinder hinweg überein? (Differenz der Mittelwerte)
- Differenzierungskomponenten: Wie gut stimmt die Streuung der Leistungseinschätzungen mit der Streuung der tatsächlich gezeigten Leistungen überein? (Quotient der Standardabweichungen)

Wenngleich die Komponenten der Urteilsakkuratheit und dabei insbesondere die Rangkomponente sehr häufig in der Forschungspraxis zum Einsatz kamen und kommen (vgl. Südkamp et al., 2012), finden sich einige Einwände gegen die Annahme, die Akkuratheitskomponenten seien adäquate Maße für die diagnostische Kompetenz von Lehrkräften: So zeigen Forschungsbefunde, dass zwischen den Akkuratheitskomponenten keine hohen Zusammenhänge bestehen und zwar sowohl innerhalb eines Merkmals über die Komponenten hinweg als auch innerhalb einer Komponente über verschiedene Merkmale hinweg (Spinath, 2005). Ursächlich mag hier unter anderem sein, dass unterschiedliche Beurteilungsstrategien unterschiedliche Konsequenzen für die Akkuratheitskomponenten haben. So führt beispielsweise die Orientierung an einer sozialen Bezugsnorm zu hohen Werten für die Rangkomponente, selbst wenn das durchschnittliche Niveau gänzlich falsch eingeschätzt wird. Weiter sind die Urteilsakkuratheitskomponenten nach Schrader (1989) hoch situationsspezifisch (vgl. Behrmann & Glogger-Frey, 2017). Es ist also denkbar, dass eine Lehrkraft in der einen Situation ein sehr akkurates Urteil fällen kann, im nächsten Moment aber nicht mehr, weil in diesem anderes und ggfs. nicht hinreichend vorhandenes Wissen erforderlich ist. Es lässt sich somit schlussfolgern, dass die Urteilsakkuratheit an sich nicht optimal ist, um Aussagen über die allgemeinen diagnostischen Kompetenzen von Lehrkräften zu tätigen.

Trotz dieser Schwierigkeiten hat der Einsatz der Urteilsakkuratheit in der Forschung auch zu relevanten Erkenntnissen geführt: So konnte beispielsweise gezeigt werden, dass Lehrkräfte die Fachleistungen ihrer Schüler*innen häufig leicht überschätzen (Niveaukomponenten; vgl. Urhahne & Wijnia, 2021), die Leistungsheterogenität jedoch recht adäquat einschätzen können (Differenzierungskomponente; vgl. van Ophuysen, 2010). Die Fähigkeit von Lehrkräften, die Schulkinder nach ihren Leistungen zu ordnen, liegt typischerweise im mittelhohen Bereich: Bei Hoge und Coladarci (1989) lag der Median der Rangkomponente bei $r = .66$, bei Südkamp et al. (2012) bei $r = .53$. Nach Cohen (1988) wären somit beide Korrelationen als hoch einzuordnen. Da aber Lehrkräfte die Eigenschaften ihrer Schüler*innen eigentlich sehr gut kennen sollten, werden die Werte von uns hier strenger interpretiert. Für motivational-affektive Schülermerkmale fällt die Korrelation zwischen Schüler-Selbsteinschätzung und Lehrkraftbeurteilung deutlich geringer aus. Urhahne und Wijnia (2021) berichten in ihrem systematischen Review beispielsweise von mittleren Korrelationen von $r = .18$ für Lernmotivation sowie $r = .17$ für negative und $r = .31$ für positive Emotionen. Eine mit den Leistungsmerkmalen vergleichbare mittlere Korrelation fanden sie für das Engagement ($r = .51$). Dies könnte darauf zurückzuführen

sein, dass dieses Merkmal enger gefasst und stärker an konkrete Verhaltensweisen gekoppelt ist.

Wenige Studien vergleichen die Urteilsakkuratheit von Lehrkräften verschiedener Schultypen, wobei es Lehrkräften an Grundschulen in der Regel etwas besser abschneiden. So beschreibt Karing (2009) einen direkten Vergleich der Rangkomponenten zwischen Grundschul- und Gymnasiallehrkräften in den Leistungsfeldern Arithmetik, Wortschatz und Textverstehen sowie im Fachinteresse an Mathematik und dem Fach Deutsch. Es zeigte sich hierbei, dass die Grundschullehrkräfte in vier der fünf Vergleiche signifikant akkuratere Werte aufwiesen. Lediglich beim Interesse für das Fach Mathematik zeigte sich der Unterschied nur numerisch, nicht aber statistisch signifikant. Weiter ergab sich in der Untersuchung ein positiver Zusammenhang zwischen der Urteilsgüte der Lehrkräfte und der Leistungsheterogenität in der Klasse. Da die Leistungsheterogenität natürlich in der Grundschule auch größer als am Gymnasium sein sollte, lassen sich hierüber die Ergebnisse von Karings Studie gut erklären. In ähnlicher Weise konnten Harks und Hannover (2017) in Bezug auf ein nicht-leistungsbezogenes Merkmal zeigen, dass Grundschullehrkräfte die Sympathiebeziehungen unter ihren Schüler*innen signifikant besser einschätzen konnten als Lehrkräfte an weiterführenden Schulen. Die Urteilsakkuratheit wurde dabei davon vorhergesagt, inwiefern sich die Lehrkräfte für die Qualität der Peerbeziehungen ihrer Schüler*innen verantwortlich fühlten und diese auch bewusst beobachteten. Dieser Befund konnte von Harks und Hannover darüber begründet werden, dass Grundschullehrkräfte deutlich mehr Zeit in einzelnen Klassen verbringen und somit ein besserer Überblick über die Sozialbeziehungen im Vergleich zu Lehrkräften weiterführender Schulen nicht überraschend sei.

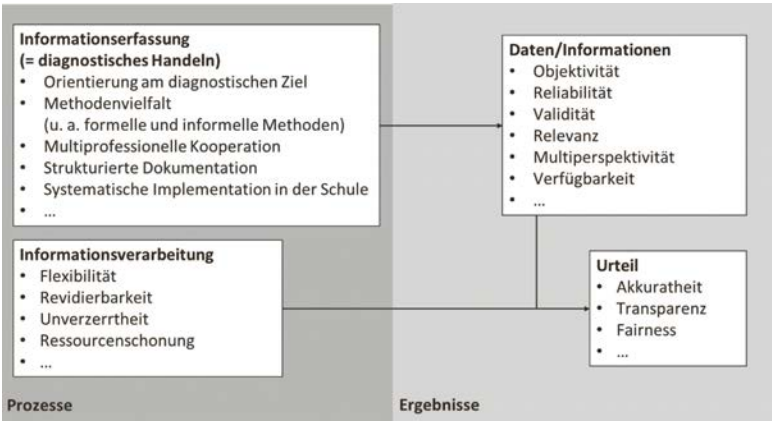
In Studien, in denen die Urteilsakkuratheit erhoben wird, werden typischerweise nur relativ einfach und möglichst objektiv zu erfassende Schüler*innenmerkmale von den Lehrkräften eingeschätzt, da ansonsten die Akkuratheitskomponenten nicht berechenbar wären. Befasst man sich jedoch mit komplexeren pädagogischen Entscheidungen wie der Schulformempfehlung, ist die Akkuratheit der Einschätzung spezifischer Eigenschaften wie Fachleistung oder Sozialkompetenz sicher eine notwendige, nicht aber hinreichende Bedingung, um einer Lehrkraft eine hohe diagnostische Urteilsfähigkeit zu attestieren. Darüber hinaus können die Akkuratheitskomponenten auch nicht unmittelbar auf die lehrkraftseitige Empfehlung angewendet werden, da es für diese – anders als beispielsweise die objektiven, standardisierten Testdaten bei Fachleistungen – keine Referenzgröße gibt, mit der die Lehrkrafteinschätzung abgeglichen und als „korrekt" identifiziert werden kann (vgl. van Ophuysen, 2010).

Gerade aber für solche konsequenzbehafteten Entscheidungen von Lehrkräften wäre es wichtig, zu identifizieren, welche Urteile sich als adäquat erwiesen haben und welche eben nicht. Was aber, wenn das Ergebnis wie im eben skizzierten Fall nicht bewertet werden kann? In so einem Fall bleibt kaum eine andere Möglichkeit, als den der diagnostischen Entscheidung vorgelagerten Prozess ins Visier zu nehmen, damit eine Optimierung des finalen Urteils möglich wird.

5. Das Vier-Komponenten-Modell der Diagnosequalität

Kernannahme des Vier-Komponenten-Modells (Behrmann & van Ophuysen, 2017, S. 39) ist, dass ein „gutes" Urteil mit hoher Wahrscheinlichkeit dann resultiert, wenn es auf verlässlichen Daten basiert, die in angemessener Weise (kognitiv) verarbeitet werden. Eine hohe Datenqualität ihrerseits setzt bestimmte Qualitätsmerkmale des diagnostischen Handelns voraus. Aus dieser Logik heraus kann die Diagnosequalität auch dann ermittelt werden, wenn beispielsweise kein objektives Kriterium vorliegt, um die Urteilsakkuratheit zu bestimmen. Das Vier-Komponenten-Modell beschreibt entsprechend neben der Urteilsqualität zwei Prozess- und zwei Ergebniskomponenten, an denen sich während der Diagnostik orientiert werden sollte.

Abbildung 1: Das Vier-Komponenten-Modell der Diagnosequalität



Quelle: in Anlehnung an Lintorf et al. (2016)

Auch im Vier-Komponenten-Modell ist *Akkuratheit* ein wesentliches Kriterium dafür, dass eine adäquate Diagnostik erfolgt ist. Darüber hinaus wird aber ebenfalls darauf Wert gelegt, dass ein gutes pädagogisches Urteil von Lehrkräften fair sein muss. *Fairness* liegt dann vor, wenn für alle Kinder ein identischer Beurteilungsmaßstab angelegt wird.[1] Dass eine Einschätzung von den beurteilten Kindern als fair erlebt wird, setzt weiter voraus, dass sie den Urteilsprozess nachvollziehen können. Dies wird durch *Transparenz* des Urteils erreicht, die dann vorliegt, wenn eine Lehrkraft grundsätzlich die Kriterien für eine Entscheidung offenlegt und somit für jedes Kind (und seine Eltern) verständlich begründen kann, wie sie zu den einzelnen Urteilen gekommen ist.

Die zweite Ergebniskomponente bezieht sich auf die *Daten und Informationen*. Diese müssen in hoher Qualität vorliegen, damit ein adäquates diagnostisches Urteil überhaupt erst möglich wird. Aus diesem Grund ist es im Vier-Komponenten-Modell wichtig, dass je nach Bedeutsamkeit der aus der Diagnose resultierenden Entscheidung (z. B. Übergangsentscheidung vs. Auswahl Arbeitsblatt) bei der Datenerhebung die drei Hauptgütekriterien der klassischen Testtheorie Reliabilität, Validität und Objektivität in angemessener Weise berücksichtigt werden. Insbesondere in diagnostischen Situationen, in denen das einzuschätzende Schüler*innenmerkmal nicht direkt durch objektive Erhebungsverfahren erfasst werden kann und in Situationen, in denen vielfältige Merkmale oder unterschiedliche situationale Kon-

1 Dies kann auch bedeuten, dass für jedes Kind eine individuelle Bezugsnorm angelegt wird. Zur Differenzierung unterschiedlicher Formen von Gerechtigkeit siehe z. B. van Ophuysen et al. (2015); Falkenberg (2020).

texte zu berücksichtigen sind, ist es sinnvoll, die Ansichten anderer Personen (z. B. Lehrkräfte, pädagogische Fachkräfte, Eltern oder die Schüler*innen selbst) in das Urteil einfließen zu lassen (*Multiperspektivität*). Schließlich ist es wichtig, dass die Daten längerfristig *verfügbar* sind, damit sie auch für zeitlich nachgeordnete Entscheidungen unverzerrt genutzt werden können. Doch selbst multiperspektivische, unverzerrt verfügbare Informationen, die mithilfe reliabler, objektiver und valider Methoden erfasst wurden, sind für die Diagnose nur dann hilfreich, wenn sie für die zu treffende pädagogische Entscheidung *relevant* sind.

Damit Informationen und Daten für die Urteilsbildung schlussendlich eine hohe Qualität aufweisen, muss das diagnostische Handeln bestimmten Qualitätskriterien genügen. Aus diesem Grund beschreibt die erste Prozesskomponente, worauf bei der Informationserfassung geachtet werden sollte: Damit beispielsweise nur für das Urteil relevante Daten generiert werden, muss Diagnostik grundsätzlich *zielorientiert* erfolgen. Darüber hinaus muss pädagogische Diagnostik praktikabel sein: Lehrkräfte werden diese erst dann regelmäßig adäquat umsetzen, wenn es gelingt, sie *systematisch in den Unterricht zu implementieren*. Damit das möglich ist, müssen je nach diagnostischer Situation und unter Berücksichtigung des jeweiligen Anspruchs an die Gütekriterien die passenden Methoden für die Datengenerierung ausgewählt und umgesetzt werden (*Methodenvielfalt*). Um Informationen vollständig und aus unterschiedlichen Sichtweisen erfassen zu können, ist eine (multiprofessionelle) Kooperation mit anderen Pädagog*innen, aber ggf. auch eine gelingende und vertrauensvolle Elternarbeit hilfreich. Zu guter Letzt stellt erst eine sorgfältige *Dokumentation* der gewonnenen Befunde sicher, dass die Informationen auch zukünftig nachvollziehbar bleiben und in zeitlich nachgeordnete Entscheidungen einfließen können.

Ein überlegtes diagnostisches Handeln und die daraus resultierenden hochwertigen Informationen sind notwendige, aber nicht hinreichende Bedingungen für ein akkurates Urteil. Um schlussendlich zu diesem zu gelangen, müssen auch hochwertige Daten sorgfältig verarbeitet und interpretiert werden. Aus diesem Grund bezieht sich die zweite Prozesskomponente auf die *Informationsverarbeitung*. Diese beschreibt verschiedene Eigenschaften kognitiver Aktivität, die das Risiko systematischer Fehlentscheidungen reduzieren. So ist es beispielsweise wichtig, je nach Bedarf verschiedene Informationen *flexibel* aus verschiedenen Quellen zu extrahieren und zu aggregieren. Wenn die Datenerhebung sorgfältig durchgeführt wurde, müssen auch unerwartete empirische Befunde akzeptiert werden, selbst wenn sie vorab bestehenden Überzeugungen der Lehrkraft widersprechen (*Revidierbarkeit*). Weiter sollte den urteilenden Lehrkräften bewusst

sein, dass auch sie leicht Urteilsfehlern wie dem Halo-Effekt oder dem fundamentalen Attributionsfehler unterliegen könnten, sodass hierauf ein besonderes Augenmerk bei der Urteilsbildung gelegt werden sollte (Unverzerrtheit, vgl. van Ophuysen & Lintorf, 2013). Schlussendlich ist es bei allem betriebenen Aufwand für eine möglichst objektive Diagnostik wichtig, auch alle anderen Anforderungen des alltäglichen Unterrichts zu berücksichtigen. Es ist somit elementar, *ressourcenschonend* zu agieren und Komplexität und Aufwand der Diagnostik an die jeweilige Situation und ihre Relevanz für das jeweilige Schulkind anzupassen.

Wird bei der diagnostischen Urteilsfindung auf die Subkomponenten des Vier-Komponenten-Modells geachtet, ist davon auszugehen, dass sich sowohl die Qualität des diagnostischen Prozesses als auch die des Gesamturteils verbessern. Auf diese Weise ist es selbst bei so komplexen Entscheidungen wie der Schulformempfehlung oder der Abwägung auf den Verbleib in der Klasse möglich, zu einem akkuraten und transparenten, fairen Urteil zu gelangen.

6. Fragestellung

Im Rahmen der Forschung zur Formation der Übergangsempfehlung am Ende der Grundschulzeit standen bislang die Ergebnisqualität (erfasst z. B. über den Schulformverbleib, s. van Ophuysen, 2006), die Kriterien der Übergangsempfehlung (Glock et al., 2013; Lintorf & van Ophuysen, 2022) und der Prozess der Informationsverarbeitung (Krolak-Schwerdt et al., 2018) im Fokus. Hingegen sind uns keine Studien bekannt, die untersuchen, wie das diagnostische Handeln im Kontext der Übergangsempfehlung ausgestaltet ist. Untersuchungen zur pädagogischen Diagnostik im Zusammenhang mit der Verbleibsentscheidung am Ende der Erprobungsstufe liegen unseres Wissens gar nicht vor.

Unter Rückgriff auf die Prozessmerkmale des Vier-Komponenten-Modells analysieren wir in unserer empirischen Studie die (F1) Einschätzung des diagnostischen Prozesses aus Sicht der Lehrkräfte. Dazu fragen wir im Einzelnen:

- Als wie wichtig erachten Lehrkräfte die Aspekte prozessbezogener Diagnosequalität?
- Welche Aspekte finden Eingang in ihr diagnostisches Handeln?
- Als wie gut bewerten die LK ihr diagnostisches Handeln?

Darüber hinaus analysieren wir (F2) den Zusammenhang zwischen der eingeschätzten Wichtigkeit des diagnostischen Handelns, dem diagnostischen Handeln selbst und der wahrgenommen Diagnosequalität. Schließlich überprüfen wir, (F3) ob es Unterschiede zwischen den Einschätzungen der Lehrkräfte in Abhängigkeit der Schulstufe (Grundschule vs. Gymnasium) – bzw. des damit einhergehenden diagnostischen Anlasses – gibt.

7. Methode

7.1. Durchführung

Im Rahmen der zweiten Phase des Projektes GanzIn[2] (2015–2019) arbeiteten in drei stufenübergreifenden Schulnetzwerken Lehrkräfte aus je einem Ganztagsgymnasium und mehreren umliegenden Grundschulen zum Thema Übergang Grundschule – Gymnasium zusammen. Die Lehrkräfte dieser sowie drei weiterer Ganztagsgymnasien, die zu einem anderen Themenschwerpunkt arbeiteten, nahmen im Herbst 2016 an einer schriftlichen, standardisierten Lehrkräfte-Befragung teil, die online durchgeführt wurde. Den Link zur Befragung erhielten sie von ihrer Schulleitung. Die Teilnahme war für die Lehrkräfte freiwillig.

7.2. Stichprobe

Der Fragebogen wurde von 89 Lehrkräften aus Grundschulen und 255 Lehrkräften aus Ganztagsgymnasien beantwortet. 16 Grundschul- und 53 Gymnasiallehrkräfte wurden von weiteren Analysen ausgeschlossen, da sie mehr als fünf der im Fokus stehenden 23 Items nicht beantwortet hatten. Die Analysestichprobe umfasst somit 73 Lehrkräfte aus zwölf Grundschulen und 202 Lehrkräfte aus sechs Gymnasien mit (fast) vollständigen Angaben.

Der Frauenanteil unterscheidet sich sehr deutlich zwischen den beiden Schulformen. Während an der Grundschule 93,2 % (68 von 73) der Befragten Frauen sind, liegt der Anteil an den Gymnasien mit 62,9 % (127 von 292) deutlich darunter. Dies deckt sich in etwa mit den Quoten in NRW im Schuljahr 2022/21 (Frauenanteil: Grundschule: 90,1 %, Gymnasium 60,9 %) (Landesbetrieb IT.NRW. Statistik und IT-Dienstleistung, 2021).

35,3 % (97 von 275) der Befragten sind zwischen 31 und 40 Jahren. Diese Altersgruppe bildet die Medianklasse. Die Berufserfahrung der Befragten

2 http://www.ganzin.de/phasezwei/willkommen-2/index.html.

weist eine breite Streuung auf. Während 4,7 % der Befragten noch im Referendariat sind, sind 8,4 % bereits seit mehr als 30 Jahren als Lehrkraft tätig. Die Medianklasse entspricht dabei der Gruppe derjenigen, die zwischen 6 und 10 Jahren im Schuldienst tätig sind. Die teilnehmenden Lehrkräfte von den Grundschulen sind im Schnitt etwas länger im Schuldienst tätig mit einer deutlichen Häufung von 41,1 % in der Gruppe mit 11–20 Jahren Berufserfahrung.

7.3. Instrument

Der Online-Fragebogen erfasst neben wenigen Informationen zur Person und zur Beratungskompetenz insbesondere auch die hier im Fokus stehenden Items zur pädagogischen Diagnostik im Kontext schulischer Übergänge. Bei der Beantwortung der entsprechenden Items wurden die Grundschullehrkräfte gebeten, sich auf den diagnostischen Prozess in Vorbereitung auf die Erteilung der Übergangsempfehlung zu beziehen. Lehrkräfte an Gymnasien sollten die Items mit Blick auf die Entscheidung über den Schulformverbleib am Ende der Erprobungsstufe beantworten.

Wichtigkeit diagnostischer Qualitätsmerkmale. Für sieben verschiedene Qualitätsmerkmale, die aus dem Vierkomponentenmodell abgeleitet wurden (vgl. Tabelle 1), gaben die Lehrkräfte jeweils auf einer vierstufigen Skala an, für wie bedeutsam sie diese Merkmale „für die Erteilung einer guten Empfehlung" erachten (1 = gar nicht wichtig, 4 = sehr wichtig). Dabei wurde nicht weiter ausgeführt, wodurch sich eine „gute Empfehlung" auszeichnet. Scree-Test und Kaiserkriterium sprechen für eine einfaktorielle Lösung, die eine Varianzaufklärung von 44 % ermöglicht. Mit der Eindimensionalität korrespondieren ebenfalls die eher hohen Trennschärfen, die zwischen .428 und .626 liegen, sowie die gute interne Konsistenz der resultierenden Skala (Cronbachs Alpha = .781).

Diagnostisches Handeln. Für die sieben Qualitätsmerkmale, die hinsichtlich ihrer Wichtigkeit beurteilt wurden (vgl. Tabelle 1), gaben die Lehrkräfte ebenfalls auf einer vierstufigen Skala an, inwiefern sie diese in ihrem eigenen diagnostischen Handeln berücksichtigen (1 = stimmt gar nicht, 4 = stimmt genau). Der Scree-Test spricht ebenfalls für eine einfaktorielle Lösung. Wenngleich der Eigenwert der zweiten Komponente minimal über 1 liegt, erscheint die einfaktorielle Lösung, die eine Varianzaufklärung von 43 % erlaubt, aus inhaltlichen Gründen sinnvoll. Trennschärfen zwischen .408 und .598 sowie Cronbachs Alpha = .773 sprechen für die interne Konsistenz der Skala.

Tabelle 1: Ergebnisse von Faktoren- und Reliabilitätsanalyse für die Skalen Wichtigkeit diagnostischer Qualitätsmerkmale und Diagnostisches Handeln.

	Wichtigkeit diagnostischer Qualitätsmerkmale		*Diagnostisches Handeln*	
	Ladung	Trennschärfe	Ladung	Trennschärfe
Bedeutsame diagnostische Kriterien identifizieren	.710	.556	.687	.539
Konkrete Situationen für Diagnose einplanen	.624	.476	.667	.518
Vielfältige Methoden nutzen	.567	.431	.559	.414
Einschätzung anderer Personen erfragen	.683	.535	.569	.408
Informationen langfristig verfügbar halten	.574	.428	.663	.515
Richtigkeit der Einschätzung immer wieder hinterfragen	.775	.626	.754	.598
Über Einfluss persönlicher Voreingenommenheit nachdenken	.699	.524	.673	.494
	Erklärte Varianz: .443	Cronbachs Alpha: .781	Erklärte Varianz: .431	Cronbachs Alpha: .773

Wahrgenommene Diagnosequalität. Die Lehrkräfte schätzten anhand von fünf Merkmalen die Qualität ihrer diagnostischen Einschätzungen ein (1 = gelingt mir gar nicht gut, 4 = gelingt mir sehr gut; vgl. Tabelle 2). Scree-Test und Kaiserkriterium sprechen für eine einfaktorielle Lösung, die eine Varianzaufklärung von 57 % ermöglicht. Die Annahme der Eindimensionalität wird auch durch die Trennschärfen unterstützt, die zwischen .557 und .703 liegen, sowie durch die interne Konsistenz der resultierenden Skala (Cronbachs Alpha = .812).

Tabelle 2: Ergebnisse von Faktoren- und Reliabilitätsanalyse für die Skala Wahrgenommene Diagnosequalität.

	Ladung	Trennschärfe
Objektive und unvoreingenommene Einschätzung	.722	.557
Transparente Begründung der Einschätzung	.733	.574
Präzise und genaue Einschätzung	.728	.569
Faire und gerechte Einschätzung	.837	.703
Ermöglichung bestmöglicher Potentialentfaltung	.757	.602
	Erklärte Varianz: .572	Cronbachs Alpha: .812

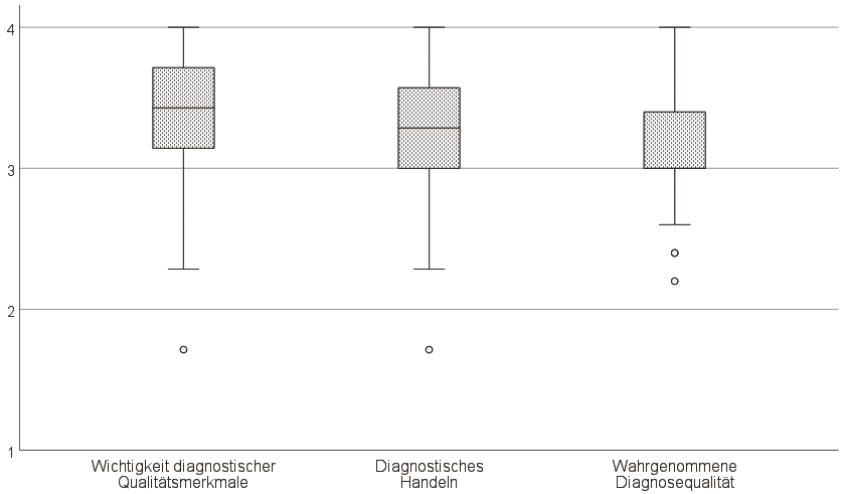
8. Ergebnisse

8.1. Einschätzungen der Lehrkräfte

Insgesamt nehmen die Lehrkräfte im Mittel eine recht hohe Wichtigkeit der diagnostischen Qualitätsmerkmale wahr (*med* = 3.43; *mean* = 3.44; *sd* = 0.401). Nur zwei Lehrkräfte geben einen Wert unterhalb des theoretischen Skalen-

mittels an. Die Werte für diagnostisches Handeln liegen ebenfalls im oberen Skalenbereich (*med* = 3.29; *mean* = 3.26; *sd* = 0.431). Die Werte von sieben der 275 Lehrkräften liegen unterhalb des theoretischen Skalenmittels. Die eigene Diagnosequalität bewerten 110 Lehrkräfte im Mittel mit dem Wert von 3,00. Nur bei drei Personen liegt der Wert unterhalb des theoretischen Skalenmittels. Die Selbsteinschätzungen fallen insgesamt recht homogen aus (*med* = 3.00; *mean* = 3.21; *sd* = 0.369). Die Informationen sind in Abbildung 2 zusammengefasst.

Abbildung 2: Box-Plot für Skalenwerte



8.2. Zusammenhangsanalyse

Alle drei Skalenwerte korrelieren signifikant miteinander. Wichtigkeit und Handeln weisen einen starken positiven linearen Zusammenhang auf (r = .687; p < .001). Je wichtiger die ausgewählten Qualitätsmerkmale des diagnostischen Handelns wahrgenommen werden, desto eher berichten die Lehrkräfte, dass sie die entsprechenden Handlungsweisen praktizieren. Eine mittlere positive Korrelation ergibt sich für Handeln und Diagnosequalität (r = .407; p < .001). Je stärker die Lehrkräfte zustimmen, verschiedene diagnostische Handlungsweisen zu praktizieren, desto positiver bewerten sie in der Tendenz auch ihre Urteilsqualität. Schließlich geht mit hoher eingeschätzten Wichtigkeit der Diagnosemerkmale und auch eine positive Selbsteinschätzung der Diagnosequalität einher (r = .318; p < .001).

8.3. Unterschiede zwischen Schulstufen

Die schulformspezifischen Mittelwerte für die beiden Skalen *Wichtigkeit diagnostischer Qualitätsmerkmale* und *wahrgenommene Diagnosequalität* unterscheiden sich nicht bedeutsam (siehe Tabelle 3). Einen eher schwachen aber signifikanten Schulformeffekt gibt es hingegen für das diagnostische Handeln ($t_{273} = 2.965$; $p = .003$; $d = 0.405$). Lehrkräfte an Grundschulen äußern hier höhere Zustimmung (*mean* = 3.38; *sd* = 0.404) als ihre Kolleg*innen an den Gymnasien (*mean* = 3.21; *sd* = 0.433). Ein Blick in die itemweisen Mittelwerte je Schulform verdeutlicht, dass für vier der sieben Items der Unterschied nach der Konvention von Cohen (1988) als schwacher Schulformeffekt bewertet werden kann. Keine Schulformunterschiede zeigen sich hinsichtlich der Dokumentation diagnostischer Informationen, dem kritischen Hinterfragen der eigenen Einschätzung und der Reflexion hinsichtlich möglicher Voreingenommenheit. Grundschullehrkräfte geben hingegen eher als Gymnasiallehrkräfte an, dass sie im Vorfeld überlegen, welche Informationen für die Empfehlung bedeutsam sind, dass sie planen, in welchen konkreten Situationen sie die relevanten Informationen erfassen wollen, dass sie vielfältige Methoden der Informationserfassung nutzen und dass sie auch andere Personen nach ihrer Einschätzung der Schüler*innen fragen.

Tabelle 3: Mittelwertvergleiche für Skalen und Einzelitems

			mean	sd	t	df	p	d
Skalen	Wichtigkeit diagnostischer Qualitätsmerkmale	GSWS	3.47 3.42	0.400 0.402	0.853	273	.394	0.116
	Wahrgenommene Diagnosequalität	GSWS	3.23 3.20	0.347 0.377	0.620	270	.536	0.085
	Diagnostisches Handeln	GSWS	**3.38** **3.21**	0.404 0.433	2.965	273	.003	0.405
Einzelitems	Bedeutsame diagnostische Kriterien identifizieren	GSWS	**3.46** **3.22**	0.529 0.647	2.760	267	.006	0.380
	Konkrete Situationen für Diagnose einplanen	GSWS	**3.07** **2.84**	0.635 0.739	2.296	269	.022	0.316
	Vielfältige Methoden nutzen	GSWS	**3.42** **3.06**	0.622 0.769	3.596	272	<.001	0.491
	Einschätzung anderer Personen erfragen	GSWS	**3.66** **3.43**	0.506 0.644	2.779	273	.006	0.380
	Informationen langfristig verfügbar halten	GSWS	3.30 3.17	0.681 0.710	1.325	272	.186	0.181
	Richtigkeit der Einschätzung immer wieder hinterfragen	GSWS	3.47 3.40	0.529 0.608	0.868	273	.386	0.119
	Über Einfluss persönlicher Voreingenommenheit nachdenken	GSWS	3.30 3.34	0.617 0.627	-0.413	273	.680	-0.056

Signifikante Mittelwertunterschiede sind fett markiert

9. Diskussion

In beiden Schulstufen werden die abgefragten Aspekte der Ergebnisqualität als sehr wichtig eingeschätzt. Auch die selbsteingeschätzte Qualität des eigenen diagnostischen Handelns ist in beiden Schulstufen in vergleichbarer Höhe ausgeprägt. Deutliche Schulformunterschiede zeigen sich hingegen bei der Umsetzung verschiedener Facetten des diagnostischen Handelns. Auf Ebene der einzelnen Items ist der deutlichste Unterschied bei der Methodenvielfalt zu sehen. Lehrkräfte an Grundschulen stimmten der Aussage, sie nutzten zur Erfassung der empfehlungsrelevanten Informationen vielfältige Methoden, stärker zu als ihre Kolleg*innen an den Gymnasien. Auch die Einbindung verschiedener Informant*innen scheint für Grundschullehrkräfte ein Vorgehen zu sein, das sie häufiger als die Lehrkräfte an den Gymnasien nutzen. Gleiches gilt für die gezielte Planung des diagnostischen Vorgehens und die bewusste Entscheidung über die für die Empfehlung relevanten Informationen. Wenngleich die Qualität der eigenen Diagnose an beiden Schulformen ähnlich hoch eingeschätzt wird, ist davon auszugehen, dass das engagiertere diagnostische Handeln an den Grundschulen zu besserer Datenqualität und reflektierterer Informationsverarbeitung führt, woraus letztlich mit höherer Wahrscheinlichkeit qualitativ hochwertigere Diagnosen (im konkreten Fall passendere Empfehlungen) resultieren sollten.

Der Einsatz vielfältiger Methoden ist zwar kein Selbstzweck, aber unterschiedliche Methoden ermöglichen die flexible Erfassung von Merkmalen. Trifft eine Lehrkraft beispielsweise ihre Empfehlung allein auf Basis von selbst entwickelten, schriftlichen Leistungstests, bleiben anwendungsnahe Kompetenzen (z. B. das Durchführen von Experimenten) unberücksichtigt und das Bild der Schülerin/des Schülers bleibt unvollständig. Ebenfalls kann durch die Einbindung von Informationen anderer Personen (z. B. Kolleg*innen, Mitschüler*innen, Eltern oder weiterem pädagogisches Personal) das Gesamtbild des Kindes vervollständigt werden – gerade bei komplexen diagnostischen Entscheidungen wie einer Übergangsempfehlung. Im Vergleich zu Grundschullehrkräften handeln Lehrkräfte an weiterführenden Schulen – laut Selbstauskunft – weniger mehrperspektivisch. Dies ist plausibel, weil sie primär auf ihr Fach und somit auch eher auf die spezifischen Leistungen in diesem fokussiert sein dürften. Umso wichtiger wäre es, dass zumindest Klassenlehrkräfte an weiterführenden Schulen sich für den Erhalt eines guten Überblicks über die Kinder bemühen, die Informationen von allen in der Klasse unterrichtenden Lehrkräften zusammentragen. Ob eventuell Klassenlehrkräfte (im Vergleich zu reinen Fachlehrkräften) verstärkt diesen mehrperspektivischen Ansatz in der Praxis umsetzen, ist in dieser Untersu-

chung leider nicht nachzuprüfen, da bei der Datenerhebung nicht zwischen Fach- und Klassenlehrkraft unterschieden wurde. Diesbezüglich wäre eine differenziertere Erfassung in zukünftigen Untersuchungen wünschenswert. Es lässt sich aber zumindest konstatieren, dass Lehrkräfte an den weiterführenden Schulen insgesamt die Merkmale des diagnostischen Handelns als ebenso *wichtig* ansehen wie die Grundschullehrkräfte.

Die befragten Lehrkräfte an Grundschulen stimmten stärker der Aussage zu, dass sie sich im Vorfeld genau überlegen, welche Information für die zu treffende Empfehlung überhaupt bedeutsam sind. Eventuell liegt dieser Unterschied darin begründet, dass die Übergangsempfehlung am Ende der Grundschulzeit ein kontrovers diskutiertes Thema in Wissenschaft und Gesellschaft darstellt. Dahingegen wird die Frage nach dem Schulformwechsel vs. -verbleib am Ende der Erprobungsstufe kaum öffentlich diskutiert. Allenfalls wird vereinzelt auf die Problematik verwiesen, dass eine Durchlässigkeit verstärkt „nach unten" besteht, während Schulform-Aufstiege wesentlich seltener zu verzeichnen sind (z. B. Bellenberg, 2012). Gerade die (fehlende) Debatte um mögliche soziale Selektivität an dieser nachgeordneten Gelenkstelle der Bildungsbiografie könnte dazu beitragen, dass auch der Frage, welche Merkmale überhaupt empfehlungsrelevant sein sollten, bei den Gymnasiallehrkräften weniger Aufmerksamkeit zuteilwird.

Es lässt sich zusammenfassen, dass die Ergebnisse der Studie insgesamt einen eher positiven Eindruck der Lehrkräfte und ihrer Einstellungen und Handlungen zur Diagnostik zeichnen. Selbst bei den Lehrkräften der weiterführenden Schulen lagen die Mittelwerte aller Items – zum Teil deutlich – über dem theoretischen Mittel. Genauer gaben Lehrkräfte beider untersuchten Schulformen an, dass für sie die gute Dokumentation von Informationen, kritisches Nachdenken und die Selbstreflexion hinsichtlich der Informationsverarbeitung wichtige und im Alltag umgesetzte diagnostische Handlungsweisen darstellen. Damit ist eine wichtige Voraussetzung erfüllt, um systematisch verzerrte Urteile zu vermeiden.

Allerdings ist zu berücksichtigen, dass es sich bei den Daten in dieser Studie um Selbstauskünfte der Lehrkräfte handelt. Sozial erwünschte Antworttendenzen sind somit zumindest denkbar. Weiter ist es möglich, dass die Ergebnisse durch die Eigenschaften der Stichprobe besonders positiv ausgefallen sind: Viele der befragten Lehrkräfte (insbesondere aus der Grundschule) waren im Rahmen des GanzIn-Projektes aktiv in Schulnetzwerke eingebunden. Dort haben sie sich mit der Übergangsgestaltung sowie mit der Beratung und der Diagnostik im Grundschulübergang beschäftigt. Wenngleich gerade an den Gymnasien auch Lehrkräfte teilgenommen haben, die nicht unmittelbar in die Netzwerkarbeit eingebunden waren, muss bei ihnen

mit möglichen Transfereffekten gerechnet werden, da die Erkenntnisse der Netzwerkarbeit jeweils im gesamten Kollegium vermittelt werden sollten. Eine verstärkte Hinwendung der Lehrkräfte zu diesem Thema ist bei der Beantwortung der Fragen damit aufgrund der selektiven Stichprobe erwartbar.

Für zukünftige Untersuchungen wäre es somit wünschenswert, wenn die Stichprobe heterogener aufgestellt und somit repräsentativer wäre. Weiterhin wurden die Daten in dieser Arbeit nur über einen Selbstauskunfts-Fragebogen erhoben. In Folgeuntersuchungen sollten verschiedene Erhebungsverfahren (Beobachtung, Dokumentenanalyse, Interviews) eingesetzt werden. Auf diese Weise könnte dann überprüft werden, ob die Lehrkräfte im Unterricht auch tatsächlich so handeln, wie sie es in dieser Studie berichtet haben und ob bzw. wie sich dies auf die Qualität der tatsächlichen Urteile auswirkt. Ist die Erfassung der Urteilsqualität wie beim Grundschulübergang oder der Versetzungsentscheidung nicht möglich, wäre zumindest die Überprüfung möglich, ob der Urteilsprozess transparent und fair abgelaufen ist.

Mit unserem Beitrag fokussieren wir das diagnostische Handeln von Lehrkräften im Kontext der Empfehlungen am Ende von Klasse 4 bzw. 6 und setzen damit an einer bedeutsamen Forschungslücke an. Studien zur Prozessqualität sind in der gesamten Forschung zur pädagogischen Diagnostik rar. Mit den oben genannten weiteren Zugängen erscheint uns die Beleuchtung des diagnostischen Handelns auch in Zukunft ein lohnendes Forschungsfeld, da insbesondere das Handeln im Bereich der Informationserfassung die handlungsnahen und somit instruier- und veränderbaren Teile des diagnostischen Prozesses abdeckt, die im Rahmen der Lehrkräfteausbildung und in Fort- und Weiterbildungen adressiert werden können.

Literaturverzeichnis

Anderson, G., Jimerson, S. & Whipple, A. (2005). Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology*, *21*, 1–20.

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, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 271–297). Waxmann.

Baumert, J., Maaz, K., Stanat, P. & Watermann, R. (2009). Schulkomposition oder Institution: Was zählt? Schulstrukturen und die Entstehung schulformspezifischer Entwicklungsverläufe. *Die Deutsche Schule*, *101*(1), 33–46.

Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 232–410). Leske + Budrich.

Behrmann, L. & Glogger-Frey, I. (2017). Produkt- und Prozessindikatoren diagnostischer Kompetenz.: Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 134–141). Waxmann.

Behrmann, L. & van Ophuysen, S. (2017). Das Vier-Komponenten-Modell der Diagnosequalität. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 38–41). Waxmann.

Bellenberg, G. (2012). *Schulformwechsel in Deutschland. Durchlässigkeit und Selektion in den 16 Schulsystemen der Bundesländer innerhalb der Sekundarstufe I*. Bertelsmann Stiftung. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/GP_Schulformwechsel_in_Deutschland.pdf

Brophy, J. (2006). *Grade repetition*. UNESCO.

Bynres, D. (1989). Attitudes of students, parents and educators toward repeating a grade. In L. Shepard & M. Smith (Hrsg.), *Flunking grades: Research and policies on retention* (S. 108–131). The Falmer Press.

Cohen, J. (1988). *The analysis of variance and covariance. Statistical Power Analysis for the Behavioral Sciences, 2nd Edn.* Academic Press.

Diebig, K. (2016). *Die Relevanz der Kriterien für die Übergangsempfehlung: Ein Vergleich unterschiedlicher Erhebungsmethoden* [Dissertation, Westfälische Wilhelms-Universität, Münster]. Deutsche Nationalbibliothek. https://nbn-resolving.org/urn:nbn:-de:hbz:6-88159636088

Falkenberg, K. (2020). *Gerechtigkeitsüberzeugungen bei der Leistungsbeurteilung*. Springer.

Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik* (3. Aufl.). Hogrefe.

Glock, S., Krolak-Schwerdt, S., Klapproth, F. & Böhmer, M. (2013). Prädiktoren der Schullaufbahnempfehlung für die Schulzweige des Sekundarbereichs I. *Pädagogische Rundschau*, *67*(3), 329–347.

Harks, M. & Hannover, B. (2017). Sympathiebeziehungen unter Peers im Klassenzimmer: Wie gut wissen Lehrpersonen Bescheid? *Zeitschrift für Erziehungswissenschaft*, *20*(3), 425–448. https://doi.org/10.1007/s11618-017-0769-8

Hoge, R. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, *59*(3), 297–313.

Hollstein, B. (2008). Der Anteil der Lehrer an der Reproduktion sozialer Ungleichheit. Grundschulempfehlungen und soziale Selektion in verschiedenen Berliner Sozialräumen. In K.-S. Rehberg (Hrsg.), *Verhandlungen des Deutschen Soziologentages: Bd. 33. Die Natur der Gesellschaft: Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel 2006* (S. 2605–2613). Campus.

Ingenkamp, K. (1975). *Pädagogische Diagnostik. Ein Forschungsbericht über Schülerbeurteilung in Europa. Trendbericht im Auftrage des Europarats in Straßburg*. Beltz.

Ingenkamp, K. (Hrsg.). (1995). *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte* (9. Aufl.). Beltz.

Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6. Aufl.). Beltz.

Jimerson, S. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, *30*(3), 420–437.

Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, *23*(34), 197–209.

Karst, K., Klug, J. & Ufer, S. (Hrsg.). (2017). Strukturierung diagnostischer Situationen im inner- und außerunterrichtlichen Handeln von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 102–113). Waxmann.

Klapproth, F., Glock, S., Krolak-Schwerdt, S., Martin, R. & Böhmer, M. (2013). Prädiktoren der Sekundarschulempfehlung in Luxemburg: Ergebnisse einer Large-Scale-Untersuchung. *Zeitschrift für Erziehungswissenschaft*, *16*(2), 355–379.

Klauer, K. J. (Hrsg.). (1982). *Handbuch der pädagogischen Diagnostik*. Schwann.

Krolak-Schwerdt, S., Pit-ten Cate, I. M. & Hörstermann, T. (2018). Teachers' judgments and decision-making: studies concerning the transition from primary to secondary education and their implications for teacher education. In O. Zlatkin-Troitschanskaia, M. Toepper, H. Pant, C. Lautenbach & C. Kuhn (Hrsg.), *Assessment of learning outcomes in higher education. Methodology of educational measurement and assessment.* (S. 73–101). Springer.

Landesbetrieb IT.NRW. Statistik und IT-Dienstleistung. (2021, 17. November). *Weniger als ein Drittel der Lehrkräfte an den NRW-Schulen waren im Schuljahr 2020/21 Männer* [Press release]. Düsseldorf. https://www.it.nrw/weniger-als-ein-drittel-der-lehrkraefte-den-nrw-schulen-waren-im-schuljahr-2020-21-maenner-105528

Lintorf, K. (2012). *Wie vorhersagbar sind Grundschulnoten? Prädiktionskraft individueller und kontextspezifischer Merkmale.* VS.

Lintorf, K., Behrmann, L. & van Ophuysen, S. (2016). Diagnostik im Lehrerberuf. In M. Rothland (Hrsg.), *Beruf Lehrer/Lehrerin* (187–203). Waxmann.

Lintorf, K. & van Ophuysen, S. (2022). Leistung ist nicht alles – Empfehlungskriterien bei sicheren vs. unsicheren Übergangsempfehlungen am Ende der Grundschulzeit. *Empirische Pädagogik*, *36*(3), 5–23.

Maaz, K., Hausen, C., McElvany, N. & Baumert, J. (2006). Stichwort: Übergänge im Bildungssystem. *Zeitschrift für Erziehungswissenschaft*, *9*(3), 299–327.

Merkens, H. & Wessel, A. (2002). *Zur Genese von Bildungsentscheidungen: Eine empirische Studie in Berlin und Brandenburg.* Schneider Verlag Hohengehren.

Ministerium für Bildung und Kindertagesstätten des Landes Mecklenburg-Vorpommern (2021). Schulgesetz für das Land Mecklenburg-Vorpommern (Schulgesetz – SchulG M-V) in der Fassung der Bekanntmachung vom 10. September 2010, zuletzt geändert durch Gesetz vom 2. Dezember 2021.

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2022a). *Das Schulwesen in Nordrhein-Westfalen aus quantitativer Sicht: 2021/2022*. https://www.schulministerium.nrw/system/files/media/document/file/quantita_2021.pdf

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2022b). Verordnung über den Bildungsgang in der Grundschule (Ausbildungsordnung Grundschule – AO-GS) in der Fassung vom 23. März 2005 zuletzt geändert durch Verordnung vom 23. März 2022.

Ministeriums für Schule und Bildung des Landes Nordrhein-Westfalen (2022c). Verordnung über die Ausbildung und die Abschlussprüfungen in der Sekundarstufe I (Ausbildungs- und Prüfungsordnung Sekundarstufe I – APO-S I) in der Fassung vom 2. November 2012 zuletzt geändert durch Verordnung vom 11. November 2022.

Pohlmann, S. (2009). *Der Übergang am Ende der Grundschulzeit. Zur Formation der Übergangsempfehlung aus der Sicht der Lehrkräfte*. Waxmann.

Porsch, R. (2018). Der Übergang von der Grundschule auf weiterführende Schulen als Handlungsfeld für Lehrkräfte. Einleitung. In R. Porsch (Hrsg.), *Der Übergang von der Grundschule auf weiterführende Schulen: Grundlagen für die Lehrerausbildung, Fortbildung und Praxis* (S. 7–12). Waxmann.

Rathmann, K., Loter, K. & Vockert, T. (2020). Critical events throughout the educational career: The effect of grade retention and repetition on school-aged children's well-being. *International Journal of Environmental Research and Public Health, 17(11).*

Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Peter Lang.

Schulgesetz für das Land Nordrhein-Westfalen 15. Februar 2005 (GV. NRW. S. 102) zuletzt geändert durch Gesetz vom 4. Mai 2021 (SGV. NRW. 223). https://bass.schul-welt.de/6043.htm

Schulordnung für die Grundschulen in Bayern 11. September 2008 (GVBl. S. 684) zuletzt geändert durch Gesetz vom 8. Juli 2021 (GVBl. S. 479). https://www.gesetze-bayern.de/Content/Document/BayVSO/true

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2015). *Übergang von der Grundschule in Schulen des Sekundarbereichs I und Förderung, Beobachtung und Orientierung in den Jahrgangsstufen 5 und 6 (sog. Orientierungsstufe)*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_02_19-Uebergang_Grundschule-SI-Orientierungsstufe.pdf

Sneyers, E., Vanhoof, J. & Mahieu, P. (2019). The impact of primary school teachers' expectations of pupils, parents and teachers on teacher track Rrecommendations. *Educational Studies*, *55*(3), 327–345.

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, *19*(1/2), 85–95.

Stahl, N. (2007). Schülerwahrnehmung und -beurteilung durch Lehrkräfte. In H. Ditton (Hrsg.), *Kompetenzaufbau und Laufbahnen im Schulsystem. Ergebnisse einer Längsschnittuntersuchung an Grundschulen* (S. 171–198). Waxmann.

Statistisches Bundesamt (2021). *Statistik der allgemeinbildenden Schulen: 08/2019-07/2020.*

Stubbe, T. C. & Bos, W. (2008). Schullaufbahnempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik*, *22*(1), 49–63.

Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762.

Urhahne, D. & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, *32*, 100374.

van Ophuysen, S. (2006). Zur Problematik der Schulformempfehlung nach der Grundschulzeit und ihrer prognostischen Qualität. In W. Bos, H. G. Holtappels, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung: Daten, Beispiele und Perspektiven* (S. 49–79). Juventa.

van Ophuysen, S. (2010). Professionelle pädagogisch-diagnostische Kompetenz – eine theoretische und empirische Annäherung. In N. Berkemeyer, W. Bos, H. G. Holtappels, N. McElvany & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung* (S. 203–234). Juventa.

van Ophuysen, S. & Lintorf, K. (2013). Pädagogische Diagnostik im Schulalltag. In S.-I. Beutel, W. Bos & R. Porsch (Hrsg.), *Lernen in Vielfalt: Chancen und Herausforderungen für Schul- und Unterrichtsentwicklung* (S. 55–76). Waxmann.

Wiedenhorn, T. (2011). *Die Bildungsentscheidung aus Schüler-, Eltern- und Lehrersicht* (1. Aufl.). VS Verlag für Sozialwissenschaften.

Students' Use of (Online) Self-Assessments in Math: Interindividual Differences in Students' Test Engagement and Test Performance in Math-Intensive Study Programs

Fani Lauermann, Daria K. Benden, Jeffrey M. DeVries & Johanna Heitzer[1]

Abstract

Students' math abilities and skills are critical for their academic success in math-intensive study programs. Thus, intervention efforts to increase students' participation and retention in math-intensive fields often target their math knowledge and skills, for instance, via diagnostic self-tests and tailored instruction. However, students' quality of engagement in these self-tests and interventions is an important prerequisite for their effectiveness. Using data from the Test-M project, which was conducted with beginning postsecondary students enrolled in math-intensive study programs (N = 3,213; 33% female), this study examined students' level of test engagement while using a comprehensive and freely available digital math self-test provided by the NRW state ministry of culture and science. This digital tool assesses students' level of math proficiency, identifies areas in need of improvement, and provides information about relevant skill training opportunities. We find interindividual differences in students' level of test engagement in terms of their subjective evaluations of the tested math content (e.g., motivation while working on the math problems), as well as test persistence (e.g., the proportion of seen but unanswered questions) and performance. These differences favored male and higher-achieving

1 *The project Test-M is funded by the Ministry of Culture and Science of the State of North Rhine Westphalia with two grants awarded to the first author (Test-M-I, 214-5.01.03.02.-149400) and the last author (Test-M-II). The sole responsibility for the content of this publication lies with the authors.*

students, students who were native speakers of German, those who participated in self-testing in-person rather than online, and those with more positive math achievement and math-related motivations before self-testing. Such students were most likely to engage with the self-test and thus potentially benefit from this type of intervention. Higher test engagement was linked to higher test performance, underscoring the importance of test-taking motivation for the utility of self-testing.

Keywords: online self-assessments; self-testing; test engagement; test motivation; STEM; study success; expectancy; values

Zusammenfassung

Die Mathematikkenntnisse der Studierenden sind entscheidend für ihren Erfolg in mathematikintensiven Studiengängen. Daher zielen Interventionsmaßnahmen zur Erhöhung der Teilnahme und des Verbleibs von Studierenden in mathematikintensiven Fächern häufig auf ihre Mathematikkenntnisse ab, z. B. durch solche Unterstützungsangebote wie diagnostische Selbsttests und passgenaue Lerneinheiten. Die Qualität der Nutzung solcher Angebote ist jedoch eine wichtige Voraussetzung für deren Wirksamkeit. Anhand von Daten von Studienanfänger*innen in mathematikintensiven Studiengängen (N = 3,213; 33 % weiblich) aus dem Test-M-Projekt untersucht diese Studie die Qualität der Testbearbeitung bei der Bearbeitung eines umfassenden digitalen und frei verfügbaren Mathematik-Selbsttests, der vom NRW-Landesministerium für Kultur und Wissenschaft zur Verfügung gestellt wurde. Dieses digitale Instrument erfasst die Mathematikkenntnisse der Testteilnehmenden, identifiziert Unterstützungsbedarfe und informiert über passgenaue Lernangebote. Unsere Analysen dokumentieren interindividuelle Unterschiede in der Qualität der Testbearbeitung im Hinblick auf die subjektiven Einschätzungen der getesteten Inhalte (z. B. Motivation während der Bearbeitung der Testaufgaben) sowie die Testausdauer (z. B. Anteil der gesehenen, aber nicht beantworteten Aufgaben) und Leistung. Diese Unterschiede fallen zugunsten von männlichen und leistungsstärkeren Studierenden, Studierenden mit deutscher Muttersprache, Studierenden, die in Präsenz statt online teilgenommen haben, und Studierenden mit vergleichsweise positiveren Mathematikleistungen und motivationalen Einstellungen vor dem Selbsttest aus. Studierende mit solchen Merkmalen lassen sich am ehesten auf diese Art von Selbsttest-Intervention ein und können somit auch am ehesten davon profitieren. Eine höhere Testmotivation war mit einer bes-

seren Testleistung verbunden, was die Bedeutung der Testmotivation für den Nutzen von Selbsttests unterstreicht.

Schlagworte: Online self-assessments; Self-testing; Test engagement; Test motivation; STEM; Study success; Expectancy; Values

1. Introduction

Students' math ability and skills are essential for their academic success in math-intensive study programs in science, technology, engineering, mathematics, and other related fields (STEM; Cromley et al., 2016; Heublein et al., 2022; Pozo & Stull, 2006; Rach & Heinze, 2017). Accordingly, intervention efforts to increase students' participation and retention in math-intensive educational and occupational fields often target their math skills, for instance, via diagnostic self-tests and tailored online or in-person skill training interventions (Ćukušić et al., 2014; Schwerter et al., 2022; Tieben, 2019). However, as we elaborate in greater detail subsequently, students' quantity and quality of engagement in these self-tests and related instructional supports (e.g., e-learning units) is a key prerequisite for their effectiveness in supporting students' academic success (e.g., Schwerter et al., 2022; Thomas et al., 2017). Accordingly, it is important to explore interindividual differences in students' use and quality of engagement in these types of interventions. Of primary interest for the present study are interindividual differences in students' use of (online) self-assessments in the math domain (i.e., self-tests that allow students to assess their math knowledge and review math prerequisites that are relevant for math-intensive study programs in higher education). A key objective of this research is to identify characteristics of students enrolled in STEM fields who are most likely to engage with self-testing of math knowledge and thus potentially benefit from this type of intervention.

1.1. Math Self-Assessments as a Type of Educational Intervention in STEM

Many students enrolled in STEM fields and other math-intensive domains such as economics face severe achievement-related difficulties at the beginning of their studies, experience significant motivational declines during the very first weeks of postsecondary education, report high levels of challenge, especially in gateway math courses, and ultimately face relatively high levels of course dropout (over 50% in some fields; Benden & Lauermann, 2022; Chen,

2013; Cromley et al., 2016; Heublein et al., 2022). Math self-assessments have emerged as a valuable educational resource that can support students' academic success in math-intensive fields, for instance, by providing diagnostic feedback, opportunities to practice math content, and as an information and screening instrument (Chevalier et al., 2018; Pozo & Stull, 2006; Schwerter et al., 2022; Wosnitza & Bürger, 2015; Yang et al., 2021).

Key objectives of online or in-class self-tests in math include (a) activating prior knowledge, (b) offering practice opportunities for previously learned math content, (c) providing diagnostic feedback regarding students' math abilities and potential knowledge gaps to be addressed with further instructional interventions, and (d) informing them about math prerequisites that are typically expected of students in math-intensive study programs (e.g., see Ćukušić et al., 2014; Lauermann, Wosnitza, et al., 2017; Pozo & Stull, 2006; Thomas et al., 2017; Wosnitza & Bürger, 2015; Yang et al., 2021). These objectives are important because prior math knowledge is an essential facilitator for mastering new math content and provides the context for future learning of mathematical concepts (Sidney & Alibali, 2015; Yang et al., 2021). There are also motivational benefits of self-testing because students tend to increase their effort when they are made aware of relevant knowledge gaps, experience how difficult it is to retrieve pertinent information during self-testing (e.g., recall math content learned in secondary school), or feel dissatisfied with their self-tested performance and thus aim to improve in the future by increasing their effort in studying (Ibabe & Jauregizar, 2010; Yang et al., 2021). Self-testing also supports students' self-regulation because it informs their ability self-evaluations (i.e., metacognitive knowledge) and allows them to monitor their learning progress (Ibabe & Jauregizar, 2010).

Finally, math self-assessments can serve as an informational and screening instrument that identifies potential knowledge gaps in math before program enrollment or in the early stages of higher education so that students can address these gaps before they can impede future learning; for instance, students can participate in preparatory math courses or use other educational supports (e.g., see discussion in Tieben, 2019). Thus, the primary use of these math self-assessments and self-testing is to support student learning and knowledge retention and provide formative feedback for future learning. However, prior research points to positive self-selection processes concerning students' participation in preparatory or remedial math courses in STEM study programs, as students with comparatively more positive academic characteristics are typically more likely to participate (e.g., self-selection based on students' levels of aspiration, achievement motivation, and academic self-perceptions; Tieben, 2019). Accordingly, a key objective of

math self-assessments is to provide diagnostic feedback to students who are most in need of learning support and remediation and to encourage them to engage in available skill training opportunities. When self-testing is incentivized or made mandatory, low-achieving students are often most likely to benefit from it (Chevalier et al., 2018; Pozo & Stull, 2006; Schwerter et al., 2022). Therefore, interindividual differences in students' use of interventions such as self-testing are of central interest to the present study.

1.2. Interindividual and Contextual Differences in Test-Taking Motivation and Test Engagement, and Implications for the Usefulness of Math Self-Assessments

Importantly, students' quality and frequency of participation in online or offline self-assessments can vary greatly, especially when these self-testing opportunities are voluntary. Evidence suggests that high-achieving students and students who report high levels of subject-specific motivation are generally more likely to use online self-testing opportunities to prepare for exams or to practice new content (Ćukušić et al., 2014; Ibabe & Jauregizar, 2010; Schwerter et al., 2022; Thomas et al., 2017). Moreover, students' motivation to engage and invest effort in a given self-assessment or test can impact its diagnostic usefulness and thus likely its effectiveness in allowing students to draw valid conclusions about their math ability (Chevalier et al., 2018; Daniel & Broida, 2004; Pozo & Stull, 2006). That is, if students do not perform to the best of their ability due to a lack of motivation, the test can provide only limited diagnostic information about the students' ability and potential knowledge gaps.

The context and modality of implementation can also make a difference. In-class quizzes are more effective in supporting students' learning gains than web-based quizzes unless web-based quizzes are designed in a way to prevent cheating (e.g., students work individually on different problem sets and with imposed time constraints; Daniel & Broida, 2004). Common cheating strategies that are likely to diminish the effectiveness of web-based self-testing include printing and sharing answers with other students, looking up answers during the quiz, and working in groups (Daniel & Broida, 2004; Kibble, 2007). The probability of cheating and thus invalidating the self-testing results increases when students' participation in web-based, unsupervised quizzes is incentivized (Kibble, 2007). This evidence suggests that students' test-taking motivation and test engagement can severely

impact the usefulness of self-testing either as a diagnostic tool or a learning support instrument.

Test-taking motivation has been defined as "giving one's best effort to the test, with the goal being to accurately represent what one knows and can do in the content area covered by the test" (Wise & DeMars, 2005, p. 2). However, as we elaborate in greater detail subsequently, students' motivational beliefs about the test's content (e.g., its perceived usefulness and validity) can also affect their performance (Chan et al., 1997; Ibabe & Jauregizar, 2010). In this study, we use the broader term *test engagement* to describe not only students' behavioral engagement (e.g., test completion and observed effort) but also their emotional (e.g., interest and enjoyment) and cognitive engagement (e.g., the cognitive effort spent by students while working on the test). In the educational literature, students' behavioral, emotional (also called affective), and cognitive engagement typically refer to the quality of a student's involvement with academic activities and goals and have emerged as a powerful predictor of students' academic outcomes such as achievement and persistence on academic tasks (Fredricks et al., 2004; Skinner et al., 2009).

There are different approaches to the operationalization of students' test-taking motivation and engagement that can be used to examine their impact on students' test performance and learning, as well as on the test's diagnostic utility and predictive validity. First, some researchers use within- or between-person experimental designs that incentivize students' participation in self-testing and estimate its effects on students' subsequent learning gains relative to a non-incentivized control condition (Chevalier et al., 2018; Daniel & Broida, 2004; Kibble, 2007; Pozo & Stull, 2006). Typical incentives include making the test count towards a final grade, awarding extra credit for participation, or making students' participation in self-testing compulsory. Overall, this research shows that incentives tend to increase not only students' effort during quizzing (i.e., self-testing) but also their subsequent learning gains and exam performance, for instance, on midterms or final exams. Accordingly, students' test-taking motivation and engagement can significantly impact the usefulness of self-testing in educational settings.

Second, another line of research focuses on the participants' self-reported test-taking motivation (e.g., "I was extremely motivated to do well on these tests") and face validity considerations such as whether the test-takers view the test as a relevant and valid indicator of what the test is supposed to measure (e.g., Chan et al., 1997). This evidence shows that students' test-taking motivation can significantly affect their test performance, especially in low-stakes test environments (Chan et al., 1997; Duckworth et al., 2011; Pohl et al., 2021). That is, low effort is likely to result in reduced performance.

Some studies also show that students' observer-rated test-taking motivation can affect the test's predictive validity; low levels of test-taking motivation and engagement tend to diminish its validity (Duckworth et al., 2011). Furthermore, interindividual differences in perceived face validity and test-taking motivation can contribute to observed group differences in test performance, such as achievement gaps between different ethnic groups, even when differences in prior performance are controlled for (Chan et al., 1997). Accordingly, differences in test-taking motivation and engagement should be taken into account in the design of self-testing opportunities and when interpreting between- or within-person performance differences.

Finally, students' motivational beliefs about the tested content (e.g., its perceived usefulness and validity) and affective-motivational experiences during the test (e.g., interest, cognitive effort) are core indicators of their test engagement that can affect both students' test performance and subsequent effort in learning (Ibabe & Jauregizar, 2010; Penk & Schipolowski, 2015; Sundre & Kitsantas, 2004). Expectancy-value theory (EVT; Eccles & Wigfield, 2020) is a prominent motivational framework that has been used to describe students' test-taking motivation and engagement (Baumert & Demmrich, 2001; Penk & Schipolowski, 2015; Silm et al., 2020; Sundre & Kitsantas, 2004). According to this framework, students' achievement-related choices and behaviors (e.g., effort and persistence while taking a given math test) can be explained by two proximal psychological predictors: students' expected success on the test and their subjective valuing of the tested content or the test outcome. The theory further suggests that students may value achievement-related tasks because they are interesting and enjoyable (*interest value*), because they are useful for other goals (*utility value*), or because they are personally important (*attainment value*).

Most of the reviewed evidence on test-taking motivation, however, has focused on students' self-reported or observed effort while taking a (self-)test rather than their motivational beliefs and experiences such as test-specific self-concept of ability, interest, enjoyment, perceived utility of the tested content, and cognitive engagement (e.g., Silm et al., 2020; Sundre & Kitsantas, 2004). Penk and Schipolowski (2015) further note that relatively few studies to date have examined the interrelations between students' test-related expectancy, valuing of the test content or outcome, test-taking effort, and test performance in the same sample. The evidence that does exist has produced rather mixed results regarding the relative weight of different predictors of test-taking effort and performance (e.g., see Baumert & Demmrich, 2001; Freund & Holling, 2011; Freund et al., 2011; Penk & Schipolowski, 2015).

2. The Present Study

The key objective of the present study was to explore interindividual differences in students' test engagement while taking a freely available self-assessment of math prerequisites for studying in math-intensive study programs in the German state of North Rhine-Westphalia (NRW). In 2015, public universities and universities of applied sciences in NRW agreed on common prerequisites and study requirements for mathematics for students interested in pursuing a degree in math-intensive study programs such as economics, computer science, natural sciences, engineering, and technology (NRW Ministry of Culture and Science, 2015). The test is not specifically targeted to the needs of students pursuing a postsecondary degree in mathematics. These prerequisites refer to ten math content areas, which are tested via the so-called WINT-Check, an online math self-assessment with eleven subtests.[2] Upon completing a given subtest, students receive performance feedback (i.e., the proportion of correctly solved math problems), can review the specific math problems they worked on, and are provided with links to tailored e-learning units for the tested content to address potential knowledge gaps. The test is an open educational resource and can be used for various purposes, including as a screening tool, to practice course content, or as an informational tool about the types of math prerequisites that are typically expected of students in math-intensive study programs before enrollment (for further information, e.g., about the test items, see www.orca.nrw).

In the broader scope of the Test-M project, the WINT-Check was revised by content specialists to ensure curricular validity (for the state's math curriculum), the subtests were shortened so that each subtest could be completed within about 20 minutes, icebreaker questions were added for some of the more challenging subtests (e.g., trigonometry), and the 11 subtests were validated in a sample of students enrolled in math-intensive study programs by linking the students' test results to their course performance in their study program. Of interest for the present study are interindividual differences in the students' test engagement while completing parts of the self-test online or in a pen-and-paper format. In addition, we examined the associations between test engagement and test performance to determine the relevance of our test engagement indicators for students' performance. We use Eccles

2 One of the ten content areas was split into two subtests (elementary functions and higher functions). The 11 subtests are: 1) Basic arithmetic operations, 2) Exponents, roots, logarithms, 3) Terms and equations, 4) Elementary functions, 5) Elementary geometry, 6) Trigonometry, 7) Higher functions, 8) Differential calculus, 9) Integral calculus, 10) Linear equation systems, 11) Vectors and analytic geometry.

et al.'s EVT to operationalize students' test-specific, as well as math-related motivational beliefs. In general, we expected that students with more positive academic characteristics (e.g., prior achievement, math-related motivations) would exhibit higher levels of test engagement, which would then predict greater performance on the self-assessment.

We also explore possible group-level differences by gender, immigration status (whether German is the primary language spoken in the home environment), mode of participation (in-person vs. online), and type of participating institution (regular research-oriented public university vs. university of applied science [UAS]) but we refrain from posing specific predictions. In general, gender differences in mathematics tend to favor male over female students (Cheryan et al., 2017; Lauermann et al., 2020; Lauermann, Tsai, et al., 2017) but students in the present sample have chosen to study in a STEM field, which makes this prediction less certain (see Benden & Lauermann, 2022). In addition, gender differences tend to be domain-specific, as some fields are much more gender-balanced (e.g., mathematics, biology, chemistry) than others (e.g., engineering, physics, computer science; Cheryan et al., 2017). Students who do not speak German as a primary language at home may be at a disadvantage because the self-assessment in the present study is administered in German. At the same time, such students may be particularly highly motivated to do well in higher education, especially students with an international background (Lauermann, 2012). The mode of participation could make a difference as well, as some previous studies have shown that unsupervised web-based self-testing is less effective than in-class testing (Daniel & Broida, 2004; Kibble, 2007). As we elaborate subsequently, however, with very few exceptions, the data collections in the present study were monitored by research assistants, either in person or via video conference, so the mode of participation may be less consequential in this case. Finally, students attending a regular university tend to be academically stronger than those attending a UAS (Tieben, 2019) but the type of institution may be inconsequential once preexisting differences in achievement and math-related motivation are taken into account. Thus, we conduct exploratory analyses to analyze such group-level differences.

In summary, we examine the following research questions (*RQ*): To what extent are differences in students' test engagement and performance linked to (1) students' background characteristics and mode of participation (i.e., gender, family language, type of higher education institution attended, and in-person vs. online participation in the self-assessment; *RQ#1*); (2) preexisting differences in achievement and learning opportunities such as the attendance of advanced placement (AP) math courses in high school (i.e., high-

school GPA, final math grade, and AP math course attendance; *RQ#2*); and (3) preexisting differences in math-related motivations (i.e., math self-concept and interest, *RQ#3*)? We also examine if differences in students' test engagement are linked to corresponding differences in test performance, controlling for preexisting differences in students' background characteristics, achievement, and motivation (*RQ#4*). We expected that higher-quality engagement would result in higher test performance, controlling for differences in background characteristics, prior achievement, and math-related motivations. Finally, drawing on EVT and related research, we examine the predictive effects of students' test-specific motivational beliefs (i.e., test-specific self-concept of math ability and interest/enjoyment) not only on students' test performance but also on their behavioral engagement because students' motivational beliefs are conceptualized as a proximal predictor of effort and persistence (see *RQ#4*).

3. Methods

3.1. Participants and Procedure

Participants in the present study were 3,213 postsecondary students who were taking either a required core course or a preparatory math course for beginning students in a math-intensive field of study (e.g., economics, computer science, natural sciences, and engineering) and who had completed at least one math problem of the revised WINT-Check correctly.[3] Most participating courses focused on math (the remaining 7 out of 36 courses focused on statistics, computer/data science, or quantitative research methods, e.g., for engineering students). A multimatrix design (Frey et al., 2009) with 22 systematically overlapping test booklets of approximately equal length and difficulty was used to reduce the testing time to about 45–60 minutes per participant (completion of all 11 subtests of the revised WINT-Check, in addition to all survey questions included in this study, would have required

3 Of 4,423 initial participations, 4 cases were out of sample (e.g., an instructor), 160 were duplicates (i.e., the same participant was sampled more than once due to attending multiple sampled courses or clicking on the same test link multiple times), 660 cases had insufficient data (no response to any demographics and dropped out prior to reaching the math problems), and 386 participants did not answer any math problems correctly, and thus their performance data could not be analyzed (in most cases, these participants saw at least some of the math problems but did not provide answers). Our final sample consists of 3,213 students enrolled in math-intensive study programs and attending courses for beginning students in their respective study programs.

at least four hours of testing time). Each subtest was split in half (approx. 10 minutes per half), and each participant worked on four subtest halves. Each half was included in four booklets, and each complete subtest was included in three booklets ensuring sufficient overlap in test items to link students' performance across booklets.

Students participated in the self-testing session during class time (in the summer and winter terms of 2021) and were randomly assigned to work on one of the 22 booklets. Students received personalized links to the booklet-specific solutions and a link to freely available e-learning units covering the tested content after all self-tests had been collected. Due to the Covid-19 pandemic, some courses took place in person, others online. Students in in-person courses completed a paper version of the math self-test; students in online courses completed it online, typically during a video conference with the entire course and in the presence of the course instructor. Research assistants were present during the in-person and online data collections to introduce the study, monitor the data collections, and answer questions. Students' participation was voluntary, not incentivized, and they could withdraw from the study at any time. The study was approved by the institutional ethics review board at TU Dortmund University (GEKTUDO_2020_19).

Students completed a consent form and answered questions about their demographic characteristics, prior performance, and math-related expectancy-value beliefs at the beginning of the self-assessment. After each subtest half (i.e., at four time points during the test), students evaluated their test-related motivations and engagement.

Students' Background Characteristics and Mode of Participation

Questions about students' background characteristics were asked at the beginning of the self-assessment and included items about *gender* (N = 3,123 test takers; 67% male, 33% female, <1% diverse), *age* (81% of the sample ≤ 21 years of age), whether German is the *primary language* spoken at home (76% German, 24% another language), and the *academic track* of their secondary school (75% highest academic track, 11% comprehensive school with different tracks, 14% other). The *type of higher education* institution attended by the students was recorded by research assistants (75% attending a regular public university, 25% a university of applied science [UAS]). Data were collected in 8 universities and 9 UAS but universities had larger student populations. Regarding the *mode of participation*, about 71% of the students participated online, and the remaining 29% in person, using a pen-and-paper test format.

Prior Achievement and Advanced Placement Math Course Attendance

The students' high school grade point average (*GPA*) was recoded such that high values indicate high performance on a scale from 1 (*low*) to 4 (*high*), with $M = 2.88$ and $SD = 0.69$ (university: $M = 3.03$, $SD = 0.65$; UAS: $M = 2.43$, $SD = 0.64$). Similarly, students' self-reported *final math grade in high school* was recorded on a scale from 0 (*low*) to 15 (*high*), with $M = 10.68$ and $SD = 3.24$ (university: $M = 11.27$, $SD = 2.96$; UAS: $M = 8.97$, $SD = 3.40$). About 56% of the participants had completed an *advanced placement math course* in high school (university: 62%; UAS: 37%). Participation in such courses suggests both higher math achievement due to self-selection into advanced math courses and exposure to comparatively more advanced math content (i.e., different learning opportunities).

Motivational Beliefs About Math

Students' expectancy beliefs and valuing of math as a subject domain were assessed at the beginning of the study on a scale from 1 (*strongly disagree*) to 6 (*strongly agree*) using previously validated items (Gaspard et al., 2015; Wigfield & Eccles, 2002). The *expectancy component* was captured via three items focusing on students' math ability self-concept (e.g., "I am good at math," $\alpha = .88$), and the *value component* via three items assessing students' math interest (e.g., "I enjoy doing math," $\alpha = .94$). In this study, we did not differentiate between math ability self-evaluations and expectancy of future success because they tend to be highly related in the math domain (Wigfield & Eccles, 2002).

Test-Specific Beliefs and Indicators of Test Engagement and Performance

Before taking the math self-assessment, students were asked to evaluate whether they perceived each of the WINT-Check content areas (i.e., subtests) as *relevant for their study program* so that we could evaluate their preexisting beliefs about its potential utility (on a three point scale: 1 = *not at all relevant*, 2 = *partly relevant*, 3 = *highly relevant*). In addition, four times during the self-testing (i.e., after each of four test parts corresponding to different subtest halves), students were asked to evaluate their performance and test engagement regarding the math problems they had just worked on. The questions referenced the content of the test rather than the test in general because students worked on different booklets. Students' *test-specific ability self-concept* was assessed via two items ("I am very good at the content tested

in Part [1-4]" and "I know the content tested in Part [1-4] at a high level," α = .92 to .94 and r_{tt} = .60 to .69 across the four time points). The perceived relevance of the tested content for one's study program was assessed as an indicator of test *utility* ("Students in my field of study must know the math content tested in Part [1-4]" and "The math knowledge tested in Part [1-4] is expected of students in my study program," α = .90 to .94 and r_{tt} = .61 to .72). To the extent that these questions reflect students' beliefs that the test accurately captures what it is supposed to measure (i.e., math prerequisites for their study program), these items also indicate *face validity*.

Students' *affective engagement* and *intrinsic value* were assessed with two items ("I found the math problems in Part [1-4] interesting" and "I enjoyed working on the math problems in Part [1-4]," α = .90 to .95 and r_{tt} = .59 to .68). Students' *cognitive engagement* was assessed with one item ("The math problems in Part [1-4] were brainteasers that made me think," r_{tt} = .48 to .61). Students' self-reported test motivations were aggregated across all four parts to obtain one booklet-specific overall score for each construct. Finally, students' *behavioral engagement* was inferred from the proportion of math problems they chose to work on (i.e., entered any correct, partly correct, or incorrect answer) relative to the total number of assigned math problems in their booklet. This proportion is reduced by both skipping and test dropout because both factors indicate lower levels of behavioral engagement with the test. Most math problems included multiple parts so that each answer field was counted as a separate response in calculating these proportions. Different booklets contained different subsets of math problems and thus also different numbers of answer fields. We calculated booklet-specific indicators of behavioral engagement by focusing on the booklet-specific proportion of math problems students had worked on.

We calculated the proportion of correct answers relative to all seen math problems (i.e., answer fields) as an indicator of *booklet-specific test performance*. We chose to treat test dropout as missing data in these calculations (rather than wrong responses) because dropout can result from various reasons, including test disengagement (Rose et al., 2017). If test dropout results primarily from a lack of math knowledge, this procedure may lead to somewhat inflated estimates of the proportion of correct responses. Therefore, we conducted a set of supplemental exploratory analyses focusing on *test dropout*, which was treated as a binary variable (0 = *test completed*; 1 = *dropout*). In addition, note that test dropout is included in the operationalization

of behavioral engagement and is thus taken into account as an indicator of disengagement.[4]

4. Results

4.1. Preliminary Analyses

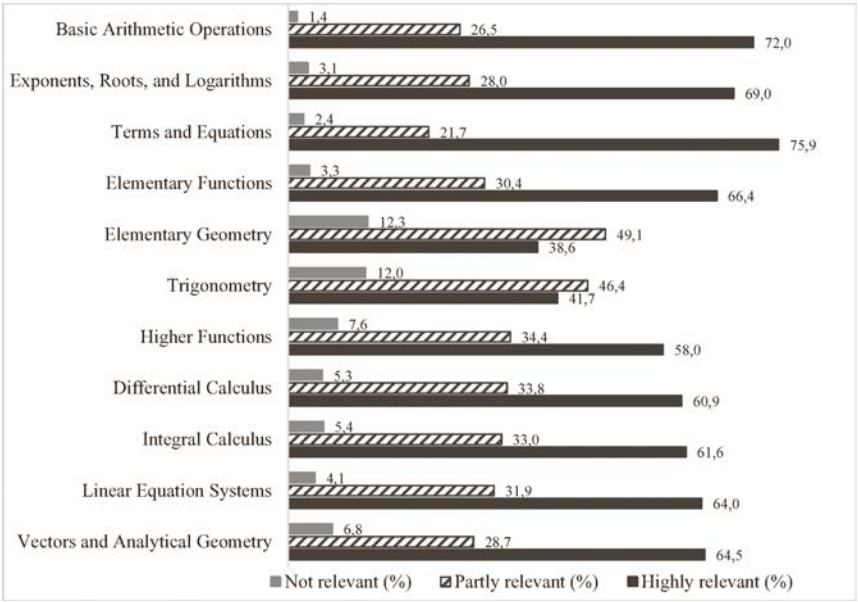
Before attending to our main research questions regarding the predictors of students' test engagement and performance, we conducted a series of exploratory analyses focusing on the test's overall *face validity* (i.e., whether the tested content is viewed as relevant to the students' fields of study), the amount and correlates of *test dropout*, and potential *between-booklet differences* that may need to be taken into account in our main analyses.

Face Validity

Before taking the self-assessment, the vast majority of test-takers (88%-99%) rated the eleven content areas included in the WINT-Check as at least partly to highly relevant for their respective study programs, which supports the test's face validity. More advanced content (e.g., trigonometry) and geometry subtests received somewhat lower ratings than less advanced and non-geometry content areas, which suggests that this content is viewed as more specialized (see **Figure 1**). The students generally viewed the content areas in their assigned booklet and the specific math problems they were assigned to work on as relevant to their field of study (e.g., indicating that students in their field of study are expected to know this type of math content; aggregated utility ratings across the four time points: $M = 4.53$, $SD = 0.97$, $Mdn = 4.5$, $Mode = 6.0$, on a six-point scale; see **Figure 2**).

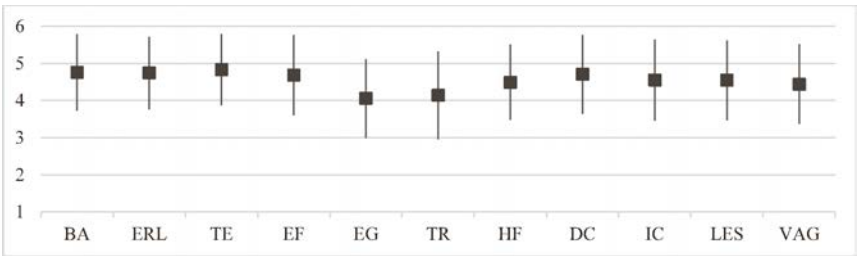
4 In the broader scope of the Test-M project, we tested a second-order generalized partial credit model, which estimated the math ability of each participant across booklets and the 11 tested content areas as a latent variable (i.e., based on an 11-dimensional first-order model and assigning equal weights to each of the 11 subtests to derive a second-order factor of math ability). The correlation between this estimate of math ability and the indicator of booklet-specific test performance used in the present study was—as expected—very high, $r = .89$, $p < .001$. Of primary interest for this study are students' test-taking behaviors, which is why we focused on students' response patterns rather than their latent math ability.

Figure 1: Perceived Relevance of Tested Content Areas Before Self-Testing



Note. The responses are reported in percentage points (0–100%) out of all valid responses.

Figure 2: Average Perceived Relevance of Tested Content Across the Eleven Content Areas (After Self-Testing)



Note. The relevance of each content area was evaluated on a six-point scale (1 = *strongly disagree* to 6 = *strongly agree*) after the students had worked on math problems belonging to that content area in their assigned booklet (average number of participants who worked on a given content area across different booklets: n = 646). BA = Basic Arithmetic Operations, ERL = Exponents, Roots, and Logarithms, TE = Terms and Equations, EF = Elementary Functions, EG = Elementary Geometry, TR = Trigonometry, HF = Higher Functions, DC = Differential Calculus, IC = Integral Calculus, LES = Linear Equation Systems, VAG = Vectors and Analytical Geometry. The black area shows the average rating of relevance (M) and the error bars indicate one standard deviation above and below the mean (M +/- 1SD).

Preliminary Analyses of Test Dropout

Before we attend to our main research questions regarding test engagement and performance, we present a set of exploratory analyses of test dropout (note that test dropout is included in our assessment of behavioral engagement, see Methods). The majority of students (83%) completed the entire self-testing battery assigned to them. Bivariate analyses revealed no significant test dropout differences for gender (male vs. not male: 16% vs. 17%; $\chi^2(1) = 0.65$, $p = .448$). However, we observed significantly higher levels of test dropout among nonnative speakers of German (German vs. not German: 15% vs. 21%, $\chi^2(1) = 11.73$, $p < .001$), UAS students (university vs. UAS: 16% vs. 20%, $\chi^2(1) = 7.36$, $p = .007$), online test-takers (in person vs. online: 7% vs. 21%, $\chi^2(1) = 83.02$, $p < .001$), and students who had not attended an AP math course in high school (AP class vs. regular math class: 13% vs. 21%, $\chi^2(1) = 33.07$, $p < .001$). Students who failed to complete the self-assessment also had somewhat lower high-school performance (GPA: $\Delta M = -0.26$, $SE = 0.03$, $t(3083) = -7.67$, $p < .001$, $d = 0.37$; math grade: $\Delta M = -0.92$, $SE = 0.16$, $t(3025) = -5.84$, $p < .001$, $d = 0.29$) and less positive attitudes toward math (math self-concept: $\Delta M = -0.26$, $SE = 0.05$, $t(3146) = 5.10$, $p < .001$, $d = 0.25$; math interest: $\Delta M = -0.25$, $SE = 0.06$, $t(3143) = 4.48$, $p < .001$, $d = 0.22$). There were no significant differences regarding the perceived utility and relevance of the self-assessment for students' fields of study (aggregated across all four time points: $\Delta M < 0.01$, $SE = 0.05$, $t(3004) = 0.08$, $p = .938$). Low face validity is thus not a likely driver of test dropout.

Bivariate Correlations and Between-Booklet Differences

Descriptive statistics, bivariate associations, and intraclass correlations (ICC) are shown in **Table 1** for all variables of interest. The ICC values reflect the proportion of systematic between-booklet differences in the assessed constructs, relative to their total variance. As is to be expected given the randomized assignment of participants to booklets, all pretest variables had close to zero intraclass correlations (ICC < .01 for demographic characteristics, prior achievement, and math-related motivations). Constructs assessed during the self-assessment (i.e., test-specific ability self-evaluations after students had worked on specific math problems, test engagement indicators, and booklet-specific performance) had slightly higher but still relatively small between-booklet variance (ICC = [.014, .097], see **Table 1**). Despite these rather small differences, we report within-booklet correlations for constructs that referenced booklet-specific content in **Table 1** (below the diagonal) so they can be compared to nonadjusted bivariate correlations (above the diagonal). All

Table 1: Descriptive Statistics, Bivariate Correlations (Above the Diagonal), Within-Booklet Correlations (Below the Diagonal), and Intraclass Correlations (ICC)

	Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
1.	Male		.04*	-.04*	-.03	.14**	-.05**	.02	.11**	.05**	.15**	.05**	.07**	-.01	.06**	.12**
2.	German language	.04*		.00	.01	.10**	.09**	.00	.08**	-.01	-.09**	.01	-.07**	-.12**	.05**	.05**
3.	University (vs. UAS)	-.04**	-.01		.09**	.22**	.38**	.31**	.23**	.20**	.29**	.10**	.18**	-.09**	.15**	.33**
4.	In-person (vs. online)	-.03	.01	.09**		.03	.12**	.12**	.05**	.04*	.01	.02	-.01	-.06**	.15**	.02
5.	AP math course	.14**	.10**	.22**	.03		.25**	.24**	.42**	.34**	.40**	.12**	.29**	-.07**	.28**	.45**
6.	High school GPA	-.05*	.09**	.38**	.12**	.25**		.71**	.43**	.35**	.40**	.15**	.26**	-.14**	.28**	.47**
7.	Final math grade	.02	-.01	.31**	.11**	.24**	.71**		.58**	.46**	.48**	.13**	.30**	-.11**	.28**	.50**
8.	Math SCA	.11**	.07**	.22**	.05**	.42**	.43**	.58**		.71**	.57**	.18**	.43**	-.05**	.28**	.50**
9.	Math INT	.05*	-.02	.20**	.04**	.34**	.34**	.46**	.71**		.48**	.27**	.55**	.07**	.26**	.41**
10.	Test-specific SCA	.16**	-.09**	.29**	.02	.41**	.40**	.49**	.59**	.50**		.36**	.56**	-.08**	.42**	.69**
11.	Test-specific utility	.06**	.02	.10**	.02	.12**	.15**	.14**	.19**	.27**	.34**		.40**	.09**	.16**	.29**
12.	Affective engagement	.07**	-.07**	.17**	.01	.28**	.26**	.31**	.44**	.56**	.56**	.39**		.36**	.34**	.45**
13.	Cognitive engagement	-.01	-.12**	-.09**	-.05**	-.07**	-.14**	-.10**	-.05**	.08**	-.07**	.10**	.36**		-.03	-.13**
14.	Behavioral engagement	.07**	.05*	.15**	.15**	.28**	.29**	.28**	.29**	.27**	.35**	.11**	.33**	.01		.51**
15.	Test performance	.14**	.05*	.35**	.02	.47**	.50**	.53**	.52**	.44**	.68**	.27**	.45**	-.11**	.50**	
	M	0.67	0.76	0.75	0.29	0.56	2.88	10.68	4.17	4.19	3.32	4.53	3.78	3.61	82.33	61.00
	SD	0.47	0.43	0.44	0.45	0.50	0.69	3.24	1.05	1.18	1.08	0.97	1.07	1.04	24.58	21.63
	N	3185	3190	3213	3213	3158	3085	3027	3148	3145	3026	3006	3036	3028	3213	3213
	Possible Range	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[1,4]	[0,15]	[1,6]	[1,6]	[1,6]	[1,6]	[1,6]	[1,6]	[1,100]	[1,100]
	ICC										.044	.043	.014	.017	.026	.097

Note. N = 3,123. Average sample size per booklet k = 146.227. UAS = University of Applied Science, AP = Advanced placement, GPA = Grade Point Average in high school, SCA = self-concept of ability, INT = interest.
*p < .05, **p < .01 (two-sided).

booklet-specific variables (i.e., test-specific self-concept, the perceived utility and relevance of the assigned math problems, all test engagement indicators, and students' test performance on their assigned booklet) were group-mean centered in subsequent analyses to eliminate any differences between booklets.

As shown in **Table 1**, the within-booklet and corresponding nonadjusted bivariate correlations were very similar. All correlations were in the expected direction and revealed positive associations between prior student achievement, motivation, test engagement, and test performance. The only exception was our measure of cognitive engagement, which was positively linked to students' math interest and test-specific affective engagement, but (weakly) negatively to students' high school achievement, math self-concept, test-specific self-concept, and test performance. Both lower-achieving students and students who enjoyed working on the math problems expended higher levels of cognitive effort. Low-achieving students may *need* to expend higher cognitive effort to solve the assigned math problems, whereas intrinsically motivated students may *want* to do so. These two factors are not mutually exclusive.

4.2. Predictors of Test Engagement and Test Performance

We conducted a series of path analyses linking students' demographic characteristics, preexisting achievement differences (high school GPA, final math grade in high school, AP math course attendance), and math-related motivations (math self-concept and interest) to different indicators of test engagement and performance. Multilevel models are not appropriate for our data due to the limited number of clusters (booklets) and the very small ICC values for all outcome measures given our randomized multimatrix booklet design (see **Table 1**). However, we implemented standard error corrections for nested data (students nested within booklets) and used group-mean centering for all constructs referencing booklet-specific content (e.g., questions about the specific math problems a given student worked on). All models were fully saturated and missing data were handled with the full information maximum likelihood algorithm.

We tested four sequential models (M1-M4, **Table 2**) entering demographic data (M1), high school achievement indicators (M2), as well as students' math self-concept and math interest reported prior to taking the self-test (M3) as predictors of students' test engagement and performance. In M4, we also examined the incremental predictive effects of students' test-specific self-concept, affective engagement, and cognitive engagement on their

Table 2: Standardized Path Coefficients for Predictors of Within-Booklet Test Engagement and Performance

	Test-Specific Outcomes																			
	Model 1						*Model 2*						*Model 3*						*Model 4*	
Predictors	*Self-Concept*	*Utility/Relevance*	*Affective ENG*	*Cognitive ENG*	*Behavioral ENG*	*Test Performance*	*Self-Concept*	*Utility/Relevance*	*Affective ENG*	*Cognitive ENG*	*Behavioral ENG*	*Test Performance*	*Self-Concept*	*Utility/Relevance*	*Affective ENG*	*Cognitive ENG*	*Behavioral ENG*	*Test Performance*	*Behavioral ENG*	*Test Performance*
Background																				
Male	.17**	.07**	.08**	-.01	.07**	.15**	.12**	.06**	.05**	-.01	.04**	.10**	.11**	.06**	.04**	.00	.04**	.09**	.02	.04**
German language	-.09**	.02	-.07**	-.12**	.05*	.05*	-.12**	.01	-.09**	-.11**	.02	.01	-.13**	.02	-.08**	-.09**	.02	.00	.06**	.06**
University/UAS	.29**	.10**	.17**	-.09**	.14**	.36**	.10**	.04*	.04**	-.04*	.01	.13**	.09**	.03	.03	-.05*	.01	.13**	-.01	.09**
In-Person/Online	-.01	.01	-.01	-.04*	.14**	-.01	-.04**	.00	-.03*	-.03†	.12**	-.05**	-.04*	.00	-.02	-.03†	.12**	-.05*	.13**	-.03*
Prior Achievement																				
AP math course							.28**	.07**	.22**	-.02	.21**	.32**	.17**	.01	.09**	-.05**	.17**	.26**	.13**	.18**
High school GPA							.08**	.08*	.06	-.09**	.13**	.17**	.08*	.09*	.07*	-.09**	.13**	.17**	.11**	.12**
Final math grade							.33**	.05†	.20**	-.01	.12**	.29**	.13**	-.04	-.01	-.06	.07*	.19**	.05†	.14**
Prior Motivation																				
Math SCA													.31**	-.04	.02	-.11**	.03	.14**	-.02	.01
Math INT													.12**	.28**	.49**	.24**	.10**	.08**	.01	-.01
Test-Specific Beliefs																				
Self-concept																			.15**	.39**
Utility/Relevance																			-.06†	.03**
Affective ENG																			.19**	.11**
Cognitive ENG																			.00	-.07**
R^2	.12**	.02**	.04**	.02**	.05**	.15**	.36**	.03**	.15**	.03**	.15**	.44**	.46**	.08**	.33**	.06*	.17**	.47**	.21**	.58**

Note. UAS = university of applied science, GPA = grade point average, SCA = self-concept of ability, INT = interest, ENG = engagement. All test-specific outcomes are group-mean centered within booklet. † $p < .10$, * $p < .05$, ** $p < .01$ (two-sided).

behavioral engagement and test performance, controlling for preexisting differences in achievement and math-related motivations. These analyses thus examined the extent to which test-related psychological variables such as affective and cognitive engagement (incrementally) predicted students' test-taking behaviors and performance.

For *RQ#1*, path analyses shown in **Table 2** indicated that *group-level differences* in students' test engagement and performance as a function of students' gender, language background, type of institution (university vs. UAS), and mode of participation (in-person vs. online) were generally small but consistent across analyses (M1-M4). *Male students* evaluated their test-specific math abilities more positively than did female students, perceived the test as somewhat more relevant for their studies (utility), reported slightly higher levels of interest in the math problems they worked on (affective engagement), and performed slightly better on the self-assessment, even when we controlled for all other covariates in M4 (**Table 2**). Even though some of these effect sizes are rather negligible (e.g., $\beta = .04$ in M4), the differences consistently favored male participants and thus warrant attention.

Students whose *primary language at home* was not German reported slightly higher test-specific self-concept (i.e., for the math problems they were assigned to work on) and higher affective and cognitive engagement but we observed slightly lower levels of behavioral engagement (due to skipping and test dropout) and lower test performance (see M1 in **Table 2**). Differences in behavioral engagement and test performance became nonsignificant when we controlled for prior achievement (M2) and motivation (M3) but reappeared once differences in test-specific self-concept and engagement were controlled for (M4, **Table 2**). Thus, insufficient German skills may have impeded students' performance, independent of their math-related motivations and affective/cognitive engagement during the test.

University students evaluated their test-specific self-concepts more positively and exhibited higher levels of affective and behavioral engagement than did UAS students, although these differences in test engagement became nonsignificant once preexisting achievement and motivational differences were controlled for (see M3 in **Table 2**). Notably, university students outperformed UAS students, even when all other covariates were included as predictors of test performance (see M4 in **Table 2**). These findings could be a sign of differences in learning opportunities between these two student populations.

Finally, students who participated in the self-assessment in person rather than online had significantly higher levels of behavioral engagement, mainly due to a higher proportion of test dropouts among online participants. Over-

all, very few notable differences emerged between the two *modes of participation*.[5]

For *RQ#2* and *RQ#3*, our analyses showed that students' *high school achievement* (GPA, final math grade, and attendance of an AP math class, see M2-M3 in **Table 2** for *RQ#2*) and *math-related motivations* assessed prior to self-testing (i.e., math self-concept and math interest, see M3 in **Table 2** for *RQ#3*) positively predicted students' test-specific self-concept of math ability, perceived test utility, affective and behavioral engagement, and test performance. Controlling for preexisting achievement differences and AP math course attendance in high school, students' self-concept of math ability no longer significantly predicted their affective and behavioral engagement while taking the test, whereas students' general interest in math, reported before self-testing, did (see M3 in **Table 2**). Thus, students with more positive motivational beliefs about math showed a higher quality of test engagement; they enjoyed working on the math problems more (affective engagement), invested higher levels of cognitive effort (cognitive engagement), and attempted to solve more math problems (behavioral engagement). As noted previously, an interesting and unexpected pattern emerged for students' cognitive engagement (see M3 in **Table 2**), which was negatively predicted by students' prior achievement and math self-concept of ability but positively by students' self-reported math interest reported before self-testing. Students with lower levels of prior achievement and math self-concept may have a higher need to invest cognitive effort to do well on the test; at the same time, students who enjoy doing math are also more willing to invest cognitive effort to solve the assigned math problems.

Finally, for *RQ#4*, our analyses showed that students' *test-specific motivational beliefs* (test-specific self-concept, utility, affective, and cognitive engagement, see M4 in **Table 2**) significantly predicted both students' behavioral engagement and their test performance, controlling for differences in students' background characteristics, high school performance,

5 Small differences in test-taking self-concept and test performance emerged in favor of online participants (β = -.05 to -.03) in our most complex analyses (M2-M4, Table 2), but the corresponding bivariate associations were close to zero and not significant (see Table 1). The likely explanation is that in-person participants had somewhat higher high school achievement and math-related motivation (see Table 1); once high school achievement differences were accounted for in M2-M4 (Table 2), a slight test performance difference emerged favoring online participants. That is, when online and in-person participants had comparable high school achievement, online participants performed slightly better on the test. This could be due to having the ability to look up information during the online self-assessment, but in view of the negligible effect size, we refrain from discussing this finding further.

and math-related motivations assessed before self-testing. Higher levels of perceived test-specific ability and affective engagement during self-testing had unique positive predictive effects on students' behavioral engagement (fewer skipped questions and test dropouts) and test performance. Notably, students' affective engagement during the test was positively related to how many math problems students chose to work on (behavioral engagement) and how many math problems they answered correctly (test performance), even when differences in students' test-specific ability self-evaluations and all remaining achievement- and motivation-related predictor variables were accounted for (see M4 in **Table 2**).

Our final path models (M3 and M4, **Table 2**) explained a substantial amount of variance in students' test-specific self-concept (46%), affective engagement (33%), behavioral engagement (17% in M3, 21% in M4), and test performance (47% in M3, 58% in M4), but much smaller amounts of variance in perceived test utility (8%) and cognitive engagement (6%). Variance in these two constructs thus appeared to be only weakly related to the included predictors. This finding is perhaps not surprising given that students generally viewed the test as relevant for their respective fields of study, and that cognitive engagement could be a mixture of need-based (e.g., due to low achievement) and motivational (e.g., intrinsic interest) factors that likely interact and may make cognitive engagement more difficult to predict.

5. Discussion

Students in math-intensive study programs in higher education often struggle with achievement-related challenges, significant motivational declines, and relatively high levels of course dropout (Benden & Lauermann, 2022; Chen, 2013; Cromley et al., 2016; Heublein et al., 2022; Rach & Heinze, 2017). Accordingly, various interventions have been developed to support students' success and retention in STEM fields, such as diagnostic self-tests and tailored skill-training interventions (Ćukušić et al., 2014; Schwerter et al., 2022; Tieben, 2019; Wosnitza & Bürger, 2015). Math self-assessments, which were of primary interest for the present study, have emerged as a valuable educational resource that can support students' academic success, for instance, by providing diagnostic feedback and opportunities to practice math content (Chevalier et al., 2018; Pozo & Stull, 2006; Schwerter et al., 2022). However, accumulating evidence suggests that students with less advantageous academic backgrounds (e.g., lower prior achievement and academic motivation) are comparatively less likely to take advantage of self-testing and remediation

interventions, despite being most likely to benefit from them (Chevalier et al., 2018; Schwerter et al., 2022; Tieben, 2019). Moreover, students' test-taking motivation and engagement can affect the diagnostic accuracy of self-testing batteries and thus their potential to support students' ability self-evaluations, self-regulation, and monitoring of learning progress (Chevalier et al., 2018; Daniel & Broida, 2004; Pozo & Stull, 2006).

Given this evidence, the present study examined interindividual differences in STEM students' test engagement and performance on a freely available self-assessment designed to assess math prerequisites for math-intensive study programs in NRW (the so so-called WINT-Check).[6] An additional objective was to examine whether differences in students' affective-motivational experiences during self-testing (e.g., test-specific self-concept of ability, affective engagement, and cognitive engagement) may predict corresponding differences in behavioral engagement (i.e., skipping of test items, test dropout) and test performance, controlling for students' demographic characteristics (e.g., gender), prior achievement, and math-related motivations. Thus, we examined if motivational differences may contribute to different test-taking behaviors and performance and therefore potentially affect the test's diagnostic utility. We discuss our main findings in the following sections.

5.1. Students' Background Characteristics and Mode of Participation

Interindividual differences in test engagement and performance emerged as a function of students' gender, type of educational institution (university vs. UAS), language background, and mode of test administration (online vs. in-person). First, although the estimated effect sizes were relatively small, male participants and university students consistently exhibited higher levels of test engagement and had slightly higher test performance, even when all other covariates were included as control variables. Such differences may be due to gendered self-selection processes in different fields of study, as well as disparities in students' access to learning opportunities. Some of the most math-heavy domains included in our study, such as engineering, physics, and computer science, tend to attract substantially more male than female

6 The WINT-Check also offers access to freely available and matching e-learning units, but students' use of these units was beyond the scope of the present research as we were unable to track the students' study behaviors after completing the self-assessment, with the exception of their end-of-term achievement.

students (Cheryan et al., 2017). Thus, on average, male students are likely to be exposed to more advanced math content in their chosen field of study. Previous evidence further suggests a more positive self-selection of boys than girls in math-intensive fields because high-achieving boys are more likely to be interested in these fields than are high-achieving girls (Breda & Napp, 2019; Wang et al., 2013).

Another relevant factor may be the exposure to different learning environments before enrollment. Male-dominated fields of study are likely to attract individuals who have had comparatively higher-quality learning opportunities in math, and the same is true for regular public universities compared to UAS (for comparisons between university vs. UAS students in engineering, see Tieben, 2019). For instance, male students and university students in our study were more likely to have taken an AP math course than were female students and UAS students, respectively. Other unobserved differences in the quality of the experienced learning environments may exist as well and should be considered in future research (e.g., high school selection).

Second, students whose primary language at home was not German were less likely to complete the math self-assessment and underperformed relative to native speakers, despite reporting higher levels of test-specific self-concept, affective engagement, and cognitive engagement. These differences in test performance could be due to differences in the participants' language proficiency. Students' text comprehension can affect their ability to perform well on math assessments, especially for word problems (Fuchs et al., 2018). In addition, anecdotal evidence from the present study (e.g., comments left at the end of the self-assessment) suggests that some online test takers who identified themselves as international students were looking up vocabulary to understand the test instructions and math problems. Accordingly, the extent to which (a) the self-test is well-suited for non-native speakers of German and (b) its validity may be affected by the students' language proficiency should be explored in future research.

Third, the mode of participation—online or in-person—did not seem to have a large effect on students' test engagement and performance (i.e., close-to-zero and nonsignificant correlations), except for a significant effect on students' behavioral engagement. Online participants were significantly less likely than in-person test-takers to complete the self-assessment. As noted in the introduction, previous research has documented significant differences in the frequency and quality of students' self-testing participation between in-class and (unsupervised) web-based self-testing opportunities (Daniel & Broida, 2004; Kibble, 2007). In the present study, both in-class and online

participation was typically supervised which may be why the observed differences between modes of participation were rather small. Cheating is also a cause for concern when tests are implemented online. As mentioned previously, Kibble (2007) observed that cheating was much more likely during unsupervised online quizzes when the test outcome was incentivized. The math self-assessment in the present study was presented as a practice and self-evaluation opportunity and did not have any stakes attached for individual students. Nevertheless, we cannot preclude the possibility that online participants had access to information that was not available to in-person participants, such as sporadic internet searches. Therefore, the mode of participation and possible incentives should be considered carefully in future research or applied settings (for a discussion of different design possibilities, see Chevalier et al., 2018).

5.2. Students' Prior Achievement, Math-Related Motivation, and Test-Specific Beliefs

Consistent with our expectations, students with lower levels of prior achievement and math-related motivations were significantly less likely to complete the assigned math self-assessment, were less engaged while taking the test, and had lower levels of test performance. These differences emerged even though the self-test was administered during class time, its administration was supervised by the course instructors and research assistants, and the tested content was rated, on average, as highly relevant to the students' fields of study by the test-takers.

First, students' prior achievement (i.e., their high school GPA, final math grade, and AP math course attendance) and ability-related beliefs (i.e., domain-specific and test-specific math self-concepts of ability) were among the strongest predictors of their behavioral engagement and test performance. This finding suggests that researchers need to be mindful of the level of challenge students with different ability levels are facing when asked to work on a given self-assessment or test (Anaya et al., 2022; Penk & Schipolowski, 2015). Evidence suggests that confronting students with difficult test questions at the beginning of a given test may result in decreased test-taking motivation for subsequent test items and, consequently, in a less accurate estimate of students' ability. Using two large data sets—a field experiment and data from an international large-scale assessment—Anaya et al. (2022) demonstrated that ordering the questions within a given test from easiest to most difficult reduced the number of test dropouts and increased the proportion of

correct responses. Moreover, students who work on easier math problems before they reach more difficult ones, tend to develop more positive ability self-evaluations even when there are no differences in performance (Bard & Weinstein, 2017). Accordingly, students in the present study who had less positive academic characteristics may have perceived the self-assessment as too challenging, which may have diminished their test-taking motivation and engagement during the test. This evidence calls for further analyses of how the test difficulty and the order in which math problems with different levels of difficulty are presented may affect the performance of test-takers with varying levels of prior ability and motivation.

Notably, second, even when differences in students' prior achievement and self-concepts of ability were controlled for, students' math interest emerged as a potent predictor of students' test engagement, which, in turn, predicted their test performance. Specifically, higher levels of students' math interest prior to self-testing corresponded to a higher quality of engagement during the self-testing session and were the strongest predictor of students' perceived test utility, affective engagement, and cognitive engagement. Moreover, students' affective engagement and test-specific self-concept of ability emerged as the strongest predictors of students' behavioral engagement and both had significant incremental predictive effects on students' test performance. This finding is not trivial, as prior research suggests that students' interest in participating in low-stakes math assessments tends to be rather low (e.g., in representative samples) and that students' test-specific interest can have limited predictive validity for students' test performance (Penk & Schipolowski, 2015). In the present study, however, the participants evaluated the math self-assessment quite positively, which is an important prerequisite for its use in authentic educational contexts. This difference, relative to prior research, may be because our sample included students who had chosen to study a math-intensive subject and the tested math content covered relevant math prerequisites for their chosen field of study (i.e., the test had high face validity).

Finally, students' affective engagement during the test—i.e., their interest and enjoyment of working on their assigned math problems–was a positive predictor of the overall number of math problems students chose to work on and how well they did on these problems even when differences in prior achievement as well as domain- and test-specific achievement-related beliefs were controlled for (e.g., students' test-specific ability self-concepts). This finding underscores the importance of taking into account not only achievement-related but also affective-motivational differences between students to

better understand corresponding differences in students' test-taking behaviors and performance.

5.3. Limitations

Several limitations must be acknowledged in the interpretation of our findings. First, as noted in the introduction, digital and in-class self-assessments can be used in different ways. These include using self-assessments as a self-evaluation instrument designed to identify relevant knowledge gaps and thus support students' self-regulated learning, a practice opportunity that activates prior knowledge and supports students' long-term retention of practiced content, and an information and monitoring tool designed to help students monitor their learning progress (e.g., in preparatory math courses). The WINT-Check, in particular, can be used either as an overall test of math prerequisites for studying in a STEM field, or its subscales and tailored e-learning units can be used as separate stand-alone learning and assessment modules. Furthermore, students can use the WINT-Check and other open educational resources independently or in the context of (math) courses in their respective study programs. Given these varied possibilities, a limitation of our study is that we focused on only one of these possible uses, namely, as a self-evaluation and self-reflection tool in a single supervised session. As noted in the introduction, students' test engagement may vary for different types of use, and students' participation and test engagement may be lower if their self-testing is not supervised.

Second, the main objective of our study was to conduct exploratory analyses of interindividual differences in students' test-taking behaviors and performance. However, although our findings are largely consistent with our expectations and prior evidence, our use of correlational data limits our ability to draw causal inferences about the role of students' test engagement in shaping their test performance. To address this limitation, future research may incentivize students' use of self-testing as a means to increase their test engagement and use random assignment to different types of incentives to allow for causal analyses.

Finally, our analyses focused on a single self-testing session and we used a multimatrix design to cover all relevant math prerequisites across different test booklets. Consequently, we were unable to examine how students' test engagement may change over time, and whether exposure to different types of content over time may affect students' test-taking behaviors and performance. This is an important avenue for future research.

5.4. Conclusions and Outlook

Our analyses demonstrate that the WINT-Check is a promising *open educational resource* that can support students' self-evaluation of math prerequisites for math-intensive study programs. The self-test has high face validity, as rated by students currently enrolled in such programs, and students' overall willingness to complete their assigned math problems was relatively high (i.e., high overall completion rates). However, our analyses also revealed notable heterogeneity in students' level of test engagement in terms of their subjective evaluations of the tested math content (e.g., motivation while working on the math problems) and test persistence (e.g., the proportion of seen but unanswered questions), as well as students' test performance. These differences systematically favored male and higher-achieving students, students who were native speakers of German, those who participated in self-testing in class rather than online, and those with more positive math achievement and math-related motivations before self-testing. Notably, students with lower levels of prior achievement and math-related motivations were significantly less likely to complete the assigned math self-assessment, were less engaged while taking the test, and had lower levels of test performance. Targeted efforts to support such students are needed and may involve, for instance, adjusting the level of challenge to better match students' prior achievement, by using adaptive testing, and by embedding the self-testing sessions in existing courses and academic counseling opportunities.

References

Anaya, L., Iriberri, N., Rey-Biel, P., & Zamarro, G. (2022). Understanding performance in test taking: The role of question difficulty order. *Economics of Education Review, 90*, 102293. https://doi.org/10.1016/j.econedurev.2022.102293

Bard, G., & Weinstein, Y. (2017). The effect of question order on evaluations of test performance: Can the bias dissolve? *Quarterly Journal of Experimental Psychology, 70*(10), 2130–2140. https://doi.org/10.1080/17470218.2016.1225108

Baumert, J., & Demmrich, A. (2001, 2001/09/01). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education, 16*(3), 441. https://doi.org/10.1007/BF03173192

Benden, D. K., & Lauermann, F. (2022). Students' motivational trajectories and academic success in math-intensive study programs: Why short-term motivational assessments matter. *Journal of Educational Psychology, 114*(5), 1062–1085. https://doi.org/10.1037/edu0000708, https://doi.org/10.1037/edu0000708.supp (Supplemental)

Breda, T., & Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences, 116*(31), 15435–15440. https://doi.org/10.1073/pnas.1905779116

Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology, 82*(2), 300–310. https://doi.org/10.1037/0021-9010.82.2.300

Chen, X. (2013). STEM attrition: College students' paths into and out of STEM fields. Statistical Analysis Report (NCES 2014-001). https://nces.ed.gov/pubs2014/2014001rev.pdf

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143*(1), 1. https://doi.org/10.1037/bul0000052

Chevalier, A., Dolton, P., & Lührmann, M. (2018). 'Making it count': Incentives, student effort and performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 181*(2), 323–349. https://doi.org/10.1111/rssa.12278

Cromley, J. G., Perez, T., & Kaplan, A. (2016). Undergraduate STEM achievement and retention: Cognitive, motivational, and institutional factors and solutions. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 4–11. https://doi.org/10.1177/2372732215622648

Ćukušić, M., Garača, Ž., & Jadrić, M. (2014). Online self-assessment and students' success in higher education institutions. *Computers & Education, 72*, 100–109. https://doi.org/10.1016/j.compedu.2013.10.018

Daniel, D. B., & Broida, J. (2004). Using web-based quizzing to improve exam performance: Lessons learned. *Teaching of Psychology, 31*(3), 207–208. https://doi.org/10.1207/s15328023top3103_6

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108*(19), 7716–7720. https://doi.org/10.1073/pnas.1018601108

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Freund, P. A., & Holling, H. (2011). Who wants to take an intelligence test? Personality and achievement motivation in the context of ability testing. *Personality and Individual Differences, 50*(5), 723–728. https://doi.org/10.1016/j.paid.2010.12.025

Freund, P. A., Kuhn, J.-T., & Holling, H. (2011). Measuring current achievement motivation with the QCM: Short form development and investigation of measurement invariance. *Personality and Individual Differences, 51*, 629–634. https://doi.org/10.1016/j.paid.2011.05.033

Frey, A., Hartig, J., & Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measurement: Issues and Practice, 28*(3), 39–53. https://doi.org/10.1111/j.1745-3992.2009.00154.x

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. N. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading, 22*(2), 152–166. https://doi.org/10.1080/10888438.2017.1398259

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology, 107*, 663–677. https://doi.org/10.1037/edu0000003

Heublein, U., Hutzsch, C., & Schmelzer, R. (2022). *Die Entwicklung der Studienabbruchquoten in Deutschland [The development of studen dropout from higher education in Germany].* (Vol. 5). DZHW. https://doi.org/10.34878/2022.05.dzhw_brief

Ibabe, I., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education: The International Journal of Higher Education and Educational Planning, 59*(2), 243–258. https://doi.org/10.1007/s10734-009-9245-6

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: Effects of incentives on student participation and performance. *Advances in Physiology Education, 31*(3), 253–260. https://doi.org/10.1152/advan.00027.2007

Lauermann, F. (2012). To go or not to go: The decision to pursue higher education abroad. In S. A. Karabenick & T. C. Urdan (Eds.), *Transitions Across Schools and Cultures* (Vol. 17, pp. 177–204). Emerald Group Publishing Limited. https://doi.org/10.1108/S0749-7423(2012)0000017010

Lauermann, F., Meißner, A., & Steinmayr, R. (2020). Relative importance of intelligence and ability self-concept in predicting test performance and school grades in the math and language arts domains. *Journal of Educational Psychology, 112*(2), 364–383. https://doi.org/10.1037/edu0000377

Lauermann, F., Tsai, Y.-M., & Eccles, J. S. (2017). Math-related career aspirations and choices within Eccles et al.'s expectancy-value theory of achievement-related behaviors. *Developmental Psychology, 53*(8), 1540–1559. https://doi.org/10.1037/dev0000367

Lauermann, F., Wosnitza, M., Karabenick, S. A., & Katharina, Z. (2017). The potential and limits of free online self-assessments of scholastic abilities to support students' transition to postsecondary education Biannual Conference of the European Association for Learning and Instruction (EARLI), Tampere, Finland.

NRW Ministry of Culture and Science. (2015). WINT-Katalog Mathematik [WINT-Catalog Mathematics]. Open Resources Campus North Rhine-Westphalia [NRW]. https://beta.orca.nrw/media/download/pdfs/handreichungen/WINT-Katalog.pdf

Penk, C., & Schipolowski, S. (2015). Is it all about value? Bringing back the expectancy component to the assessment of test-taking motivation. Learning and Individual Differences, 42, 27–35. https://doi.org/10.1016/j.lindif.2015.08.002

Pohl, S., Ulitzsch, E., & von Davier, M. (2021). Reframing rankings in educational assessments. Science, 372(6540), 338–340. https://doi.org/10.1126/science.abd3300

Pozo, S., & Stull, C. A. (2006). Requiring a math skills unit: Results of a randomized experiment. *American Economic Review, 96*(2), 437–441. https://doi.org/10.1257/000282806777212486

Rach, S., & Heinze, A. (2017). The transition from school to university in mathematics: Which influence do school-related variables have? *International journal of science and mathematics education, 15*(7), 1343–1363. https://doi.org/10.1007/s10763-016-9744-8

Rose, N., von Davier, M., & Nagengast, B. (2017, 2017/09/01). Modeling omitted and not-reached items in IRT models. *Psychometrika, 82*(3), 795–819. https://doi.org/10.1007/s11336-016-9544-7

Schwerter, J., Dimpfl, T., Bleher, J., & Murayama, K. (2022). Benefits of additional online practice opportunities in higher education. *Internet and Higher Education, 53*, Article 100834. https://doi.org/10.1016/j.iheduc.2021.100834

Sidney, P. G., & Alibali, M. W. (2015). Making connections in math: Activating a prior knowledge analogue matters for learning. *Journal of Cognition and Development, 16*(1), 160–185. https://doi.org/10.1080/15248372.2013.792091

Silm, G., Must, O., Täht, K., & Pedaste, M. (2020). Does test-taking motivation predict test results in a high-stakes testing context? *Educational Research and Evaluation, 26*(7–8), 387–413. https://doi.org/10.1080/13803611.2021.1949355

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493–525. https://doi.org/10.1177/0013164408323233

Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29*(1), 6–26. https://doi.org/10.1016/S0361-476X(02)00063-2

Thomas, J. A., Wadsworth, D., Jin, Y., Clarke, J., Page, R., & Thunders, M. (2017). Engagement with online self-tests as a predictor of student success. *Higher Education Research and Development, 36*(5), 1061–1071. https://doi.org/10.1080/07294360.2016.1263827

Tieben, N. (2019). Brückenkursteilnahme und Studienabbruch in ingenieurwissenschaftlichen Studiengängen [Participation in preparatory courses and student dropout in engineering majors]. *Zeitschrift für Erziehungswissenschaft, 22*(5), 1175–1202. https://doi.org/10.1007/s11618-019-00906-z

Wang, M.-T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological science, 24*(5), 770–775. https://doi.org/10.1177/0956797612458937

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of Achievement Motivation* (pp. 91–120). Academic Press. https://doi.org/10.1016/b978-012750053-9/50006-1

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*, 1–17. https://doi.org/10.1207/s15326977ea1001_1

Wosnitza, M., & Bürger, K. (2015). Special issue: Online-self-assessments. *Das Hochschulwesen [Higher Education], 63*(3,4), 73–140. http://www.hochschulwesen.info/inhalte/hsw-3-4-2015.pdf

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 399–435. https://doi.org/10.1037/bul0000309

What can Digital Trace Data Tell Us about Post-secondary Students' Academic Success?

An Overview of the Literature and an Illustrative Example

Luise von Keyserlingk, Fani Lauermann, Renzhe Yu, Charlott Rubach, Richard Arum

Abstract

Self-regulated learning (SRL) is an important facilitator of students' academic success in post-secondary education. We provide an overview of the growing literature that uses digital trace data to investigate students' study behaviors and SRL in post-secondary institutions. Digital trace data such as (changes in) click activity obtained from learning management systems (LMS; e. g., Moodle, Ilias, or Canvas) can be a useful indicator of SRL regarding students' time management and aspects of monitoring behavior. Whereas broad measures of study activities in online environments, such as the number of clicks and time spent on course activities do not consistently predict performance, more fine-grained measures, such as number of clicks before deadlines in courses or using voluntary quizzes to monitor learning, can identify adaptive self-regulated learning strategies. Hereby, information about course design and context is essential for creating such SRL measures. In addition, multi-source data (e. g., digital traces and self-reported learning strategies) are needed to capture not only behavioral but also motivational and meta-cognitive aspects of SRL. In this study, we provide an illustrative example of the type of digital trace data that can be collected via LMS to predict students' academic success using data from the ongoing longitudinal UCI-MUST project.

Keywords: self-regulated learning; digital trace data; post-secondary education; empirical longitudinal data; course performance

Zusammenfassung

Selbstreguliertes Lernen (SRL) ist ein wesentlicher Prädiktor für akademischen Erfolg in postsekundärer Bildung. Das Kapitel gibt einen Überblick über aktuelle Studien, die digitale Verhaltensspurdaten zur Untersuchung von Lernverhalten und SRL von Studierenden nutzen. Digitalen Verhaltensspuren, die zumeist von Lernmanagement Systemen wie Moodle, Ilias oder Canvas gewonnen werden, eignen sich insbesondere für die Untersuchung des Zeitmanagements oder der Selbstevaluation von Studierenden als zentrale Aspekte von SRL. Dabei zeigt sich, dass quantitative Maße von Lernverhalten, wie die gesamte Anzahl an Klickaktivitäten oder die Dauer die Studierende auf einem online Kurs verbringen eher moderate Prädiktoren von Kursleistung darstellen. Spezifischere Maße von SRL, wie z. B. Veränderungen im Klickverhalten von Studierenden vor Deadlines oder die Nutzung regelmäßiger freiwilliger Testfragen zur Überprüfung des Lernzuwachses sind hingegen vielversprechendere Maße um adaptive und erfolgreiche SRL Strategien zu erfassen. Hierbei sind detaillierte Informationen über den Kursablauf und -kontext zu berücksichtigen. Um neben diesen vornehmlich deskriptiven Maßen von SRL auch internale SRL Prozesse identifizieren zu können, sollten zusätzlich zu digitalen Verhaltensspuren auch Selbstberichtdaten zu motivationalen und meta-kognitiven SRL Elementen erhoben werden. Neben einem Überblick über aktuelle Studien nutzen wir Daten der UCI-MUST Studie, um ein Beispiel zur Verwendung von digitalen Verhaltensspuren zur Beschreibung von Lernverhalten und zur Vorhersage von Kursleistungen von Studierenden zu geben.

Schlagworte: Selbstreguliertes Lernen; digitale Verhaltensspurdaten; postsekundäre Bildung; empirische längsschnittliche Daten; Kursleistung

1. Introduction

The ability to plan, monitor, and adjust one's own study activities is important for students' learning success. Particularly in higher education, where learning environments are less structured than in secondary school, and students are afforded more autonomy to choose their courses and structure academic activities, self-regulated learning (SRL) skills, such as planning, monitoring, and adjustment of study behaviors, are particularly important predictors of students' academic success (Broadbent & Poon, 2015; Zimmerman & Schunk, 2011). The challenging years since the outbreak of the COVID-19 pandemic have further shown that SRL-skills are central facilitators of college students'

academic success (Berger et al., 2021; Juriševič et al., 2021). The closure of university campuses and the shift to remote learning during the pandemic drastically changed the learning environments of college students. While structures such as synchronous face-to-face classes, libraries, or study groups on campus were no longer available, students had to develop new study routines with remote learning activities. Learning management systems (LMS), such as Moodle, ILIAS, and Canvas, are designed to facilitate teaching and learning and have become a central element of instruction in post-secondary education settings. During the pandemic, LMS provided a key platform for remote teaching and learning (e. g., for sharing study materials, assignments, and videotaped classes). Moreover, LMS also provide new and promising approaches to investigate students' SLR and learning outcomes in authentic educational contexts. For example, through observations of students' authentic interactions with study materials provided via LMS and their study behaviors such as time management, adherence to deadlines, and click activities.

Self-regulation research has a long tradition of using self-report data to investigate students' SRL (e. g., Pintrich et al., 1991; Weinstein & Palmer, 2002). An advantage of such self-reported data is that they can capture internal cognitive, motivational, and affective processes that are not directly observable in students' behavior (e. g., aspirations and goals, elaboration strategies, outcome expectations, or anxiety and enjoyment during study activities). Furthermore, surveys with established SRL instruments can be administered at relatively low costs and can be applied to any educational context (Wolters & Won, 2018). However, students' self-reports of their study behaviors and learning strategies can be subject to various reporting biases, for instance, due to insufficient memory or the elicitation of socially desirable responses (Baker et al., 2020). Accordingly, self-regulation researchers often rely on multiple data sources to obtain a more comprehensive and precise assessments of SRL. Since the implementation of LMS in educational contexts, SRL researchers increasingly use digital trace data from these LMS to examine students' learning behavior in different courses (Baker et al., 2020; Crompton et al., 2020; Li et al., 2020). Different measures, such as the number of clicks, number of study sessions, or time spent in a course per day can be used to quantitatively describe students' study behaviors, engagement, and learning patterns in a given course. In this chapter, we provide an overview of the growing literature on digital trace data that can be obtained via LMS to investigate students' learning behavior and success. In particular, we focus on studies using digital trace data to measure students' SRL behaviors in post-secondary education, and we outline some of the key advantages and challenges of using such data in SRL research. Second, we use data from

the ongoing longitudinal UCI-MUST project (Arum et al., 2021) to provide an illustrative example of the type of digital trace data that can be collected via LMS to predict students' academic success and potentially aid their SRL.

2. Self-regulated Learning in College

In post-secondary education, students are often required to navigate a challenging curriculum and organize their study activities in several courses that they take simultaneously. Thus, SRL is a central skill that facilitates learning and success in post-secondary education (Broadbent & Poon, 2015). SRL entails multiple cognitive, motivational, behavioral, and affective components, and existing theoretical models on SRL describe and integrate these components in somewhat different ways. Boekaerts (1999), for instance, describes a set of skills and resources that are central to self-regulated learning in a multi-layered model, whereas Zimmerman (1990) describes SRL as a cyclical process where learners use different sets of SRL skills and behaviors in a structured and recurrent manner. Boekaerts' multi-layered model suggests that SRL behaviors unfold on different levels of proximity to the learning content. These include the choice of adaptive cognitive strategies to process the *learning material*, the use of meta-cognitive strategies to regulate the *learning process*, and the choice of suitable goals and recourses to facilitate learning in specific *environments* (Boekaerts, 1999, 2010). Process models, by comparison, describe three central elements of self-regulated learning that occur cyclically: (a) *forethought*, (b) *performance and volitional control*, and (c) *self-reflection* (Zimmerman, 1990). The phase of *forethought* or *planning* entails the setting of specific learning goals and the selection of learning strategies. The second phase of *performance and volitional control* requires students to apply the selected learning strategies and monitor and control their learning process. The phase of *self-reflection* includes self-evaluation and causal attribution of the learning outcomes, for instance, to factors that are internal and controllable or external and not controllable by the student, and informs subsequent planning phases (Zimmerman, 1990; Zimmerman & Schunk, 2011).

Although both theories emphasize different aspects of SRL, they have a large overlap in terms of the described cognitive, meta-cognitive, motivational, and behavioral skills that are essential to SRL. Depending on the chosen theoretical framework, empirical SRL research often focuses either on cyclical aspects of SRL behaviors or on specific cognitive and meta-cognitive skills and strategies of learners. Extensive literature shows that students with

higher SRL skills attain better learning outcomes and are more successful in college (Broadbent & Poon, 2015; Dörrenbächer & Perels, 2016; Kitsantas et al., 2008; Zimmerman, 1990). Most of this literature used well-established survey instruments to measure students' SRL (e. g., Motivated Strategies for Learning Questionnaire – MSLQ: Pintrich et al., 1991; Learning and Study Strategies Inventory – LASSI: Weinstein & Palmer, 2002). Over the past years, however, there has been growing interest in the use of digital trace data from LMS, besides established survey instruments, as a means to investigate SRL (Arizmendi et al., 2022; Bernacki et al., 2020; Cogliano et al., 2022). Such behavioral trace data provide complementary data on students' study behaviors, in addition to their self-reports. However, an open question that warrants careful consideration concerns the interpretability of digital trace data as an indicator of central SRL components, such as cognitive and meta-cognitive skills described by Boekaerts (1999; 2000), or cyclical aspects of SRL behavior described by Zimmerman (1990; 2011).

2.1. Measuring Self-regulated Learning with Digital Trace Data

Digital trace data from learning management systems capture students click activities in log files and, thus, such data allow capturing students' study behavior in authentic contexts (Arizmendi et al., 2022; Bernacki et al., 2020; Crompton et al., 2020). Digital traces of students' study behaviors provide an opportunity to measure certain aspects of self-regulated learning, such as time management, regularity of study efforts, and self-testing (e. g., using voluntary quizzes or assignments to test one's knowledge of course contents). Other central elements of self-regulated learning that do not manifest directly in observable behavior, such as goal-setting, cognitive strategy use (e. g., elaboration of learning material), and emotion regulation are comparatively more difficult to measure with such data (Bernacki, 2018). Current research can be categorized broadly into a) studies that focus on global measures of study behavior, such as the overall number of clicks and time spent on courses, and b) studies that investigate specific aspects of self-regulated behavior, such as active planning behavior, cramming versus spacing, self-testing behavior, help-seeking behavior, and others. The main interest of this chapter is to provide an overview of different approaches that have been used to measure students' learning behaviors and different aspects of SRL with digital trace data. The main objective of this chapter is not to provide an exhaustive review of the existing literature, but rather to review recently developed and applied

approaches to measuring central aspects of SRL behaviors with digital trace data in post-secondary education settings.

3. Links Between Digital Trace Data and Self-regulated Learning Behaviors in Post-secondary Education

3.1. Global Measures of Study Behavior Based on Digital Traces

Several relatively global measures of study behavior can be obtained from digital traces in learning management systems (LMS). Commonly used global assessments include students' overall number of clicks when using the course site via the LMS, time spent navigating through and interacting with course material online, and the number of online study sessions in a course (Baker et al., 2020; Cicchinelli et al., 2018; Greene et al., 2021). Such measures describe the quantity of students' study actions but do not capture the quality or type of study actions. Nevertheless, such measures can predict desirable learning outcomes. For instance, Cicchinelli et al. (2018) found that first-year students in a computer science program who produced more clicks overall had more study sessions, had longer durations of interacting with the course site in the online LMS that was used for their lecture, and attained better grades in quizzes and final exams in the lecture. Similarly, in a distance-learning university in the UK, Nguyen et al. (2020) investigated the associations between the time students spent on their courses online and their course performance. Using data from roughly 150,000 students, they showed that students who spent more time studying for their courses – as inferred by how much time they spent interacting with course contents on LMS – were more likely to pass their exams and obtained better grades. While these associations were observed for all students in the courses, Nguyen et al. (2020) described differences in study activities for students with different demographic backgrounds. On average, female students spent more time on their college courses on LMS compared to their male peers, whereas students from underrepresented ethnicities spent less time on LMS course content and obtained lower course grades. Authors provide different explanations for these results: Students of underrepresented minorities might have competing obligations, such as work in addition to studying, that could constrain their available time and resources for studying. In addition, these students might lack knowledge about effective study and SRL strategies and may therefore show less adaptive study behaviors.

Broad measures, such as the number of clicks, study sessions, and time spent on a course provide additional insights into students' self-regulated

learning when observed on a daily or weekly aggregation level across a course. With such measures, researchers can investigate changes in the number of study activities across specific time periods, the regularity, and the fluctuation in students' study activities. Park et al. (2017), for instance, identified three patterns of changes in students' click activities across the duration of an online and a face-to-face course with LMS at a public university in the US. They categorized students' study behaviors into "increasing click activity," "no change," or "decreasing click activity" across the courses. Students who had an increasing pattern of click activity across the course were more likely to pass the course than students who had a decreasing pattern. Focusing on students' weekly click activity in a course where students regularly received tasks on Mondays with a submission deadline on Fridays, Park et al. (2018) showed that students with a more regular click pattern working on the tasks throughout the week obtained better course grades compared to students who had increased click activity only shortly before the deadline on Fridays. These studies suggest that students who maintain more regular and continuous study activities tend to attain better performance outcomes.

While the above-mentioned studies reported positive associations between broad measures of study activities on LMS and course performance, other studies suggest no or only very small positive associations between the overall number of click activity and time spent on the course sites with course performance outcomes (see, e. g., Greene et al., 2021, You et al., 2016). Mixed findings could be explained by different course designs and types of use of LMS by the instructors. Furthermore, mixed findings could be related to different levels of granularity of behavioral trace measures ranging from variables on the course level (e. g., time spent on the course across the entire semester (Nguyen et al., 2020), daily click activities across the entire semester (Park et al., 2017), or daily click activities per week (Park et al., 2018)). Decisions about the level of aggregation of behavioral trace data should be driven by the research question and available information about the course context. For instance, only if information about the course design and course deadlines is available, researchers can meaningfully interpret increases in study activities before certain dates or can derive measures such as 'time to a deadline' to investigate aspects of students' time management in their courses.

3.2. SRL-specific Measures of Study Behavior Based on Digital Traces

Digital trace data can also be used to infer specific types of self-regulated behaviors, as conceptualized by Boekaerts (1999, 2010) and Zimmerman (1990, 2000). These include, for instance, measures of time management and regularity of study activities that can be linked to cyclical processes of SRL (Zimmerman, 1990; 2000), and to the successful use of meta-cognitive strategies to regulate learning processes (Boekaerts, 1999; 2010). Furthermore, such measures can describe the use of specific meta-cognitive strategies, such as help-seeking behaviors, monitoring of learning outcomes with self-tests and quizzes or monitoring performance outcomes through accessing grade books in a course. We provide an overview of studies that used such measures in the following sections.

Time Management and Procrastination Measured with Digital Trace Data

When relevant information about the course context and course design is available, researchers can generate more SRL-specific variables with students' digital trace data to investigate SRL and performance in college courses. Information about deadlines and due dates, for example, can be used to generate measures for procrastinating behaviors (Li et al., 2020; Rodriguez et al., 2021). Such measures can include the time between the submission of an assignment and the submission deadline and the proportion of assignments and content material accessed before versus on a due date. Using such measures of self-regulated behaviors from an online college course, Li et al. (2020) showed that students who proportionally accessed more study units before the due date than on the due date, and who submitted assignments longer in advance of the deadline, were students who obtained better final course grades. Similarly, Rodriguez et al. (2021) investigated whether regular access versus irregular and delayed access of lecture videos predicted final course grades in an asynchronous online course. The course contained 48 short lecture videos divided into four modules. In each module, students needed to watch the corresponding lecture videos in a pre-specified order until a specific due date. Rodriguez et al. (2021) identified four clusters of study behaviors with students who a) watched nearly all videos before the due date (early planners), b) watched most videos before the due date and only a few on the due date (planners), c) watched most videos on the due date (procrastinators), and d) watched only a few videos and all of them late (low engagers). Students who were 'early planners' and 'planners' obtained better course grades than their

peers who were identified as 'procrastinators' or 'low engagers'. Rodriguez et al. (2021) further investigated if students with certain background characteristics were at particular risk for belonging to a cluster with maladaptive study patterns (i. e., procrastinators or low engagers). Results showed that students of low-income families and first-generation college students more often had low engagement patterns in their courses, compared to their peers, and obtained lower grades in their courses.

Monitoring and Self-evaluation Measured with Digital Trace Data

Digital trace data can also be used to identify study behaviors that are related to planning, monitoring, and self-evaluation, depending on what supplemental information is available about the course (see, e. g., Greene et al., 2021; Huang et al., 2022). For instance, students' use of course calendar functions and visits of course modules that show the course syllabus at the beginning of the course and before course exams can provide information about *planning* behaviors. Course syllabi are required for each college course in the US and typically provide information about course activities, requirements, and grading policies in the course. Students' completion of voluntary quizzes in the course (i. e., opportunities to self-test course content and evaluate knowledge gaps or learning gains) can indicate *monitoring* of the learning progress. Students' regular access of (online) performance feedback from the instructors and gradebooks can indicate *self-evaluation* practices. Latent profile analyses have been used to investigate to what extent students show *planning*, *monitoring*, and *self-evaluating* behaviors when they use course materials provided via LMS (Greene et al., 2021; Hong et al., 2020; Huang et al., 2022; Li & Baker, 2018). Students who regularly participated in ungraded and/or voluntary quizzes attained higher course grades than their peers who showed less quiz-taking behavior (Carvalho et al., 2022; Ifenthaler et al., 2022; Huang et al., 2022; Li & Baker, 2018). A possible explanation provided by the authors is that students who regularly participated in the quizzes were thus able to monitor their learning progress and adjust learning strategies as needed. Greene et al. (2021), for instance, used data from 408 students enrolled in a biology class. They used latent profile analysis to identify SRL behaviors related to planning activities (e. g., accessing the course syllabus and using the course calendar, as well as reading announcements), information acquisition (e. g., attending class meetings, accessing additional course readings), and help-seeking (e. g., reaching out for help, clicking on links to learning support services). Students who showed more planning activities and information acquisition obtained better course grades than their peers. Sim-

ilarly, Hong et al. (2020) used latent profile analysis to investigate if students predominantly used SRL strategies related to planning (by visiting the course syllabus and study guides), monitoring of their learning progress (by taking regular exercises and quizzes), and monitoring and evaluating their performance (by visiting their gradebooks). The sample consisted of digital trace data from 1,326 college students in biology classes at a mid-western university in the US. Most students showed little planning and monitoring behaviors. About 15 % of the students showed more planning behaviors and frequently monitored their performance by visiting the course gradebook. About 10 % of the enrolled students frequently monitored their learning through quizzes. Students who regularly showed behaviors related to monitoring their learning and performance through quizzes and gradebooks outperformed their peers and obtained higher final course grades. These findings indicate that students who showed more study activities that can be linked to critical aspects of SRL were more successful in their courses.

However, the above-described findings derived from digital trace data remain on a rather descriptive level of observable study behavior and associations with desirable performance outcomes. An open question is, whether students enact certain study behaviors, such as regular completion of self-tests and quizzes, because they are encouraged or required to do so by their instructor (i. e., *externally regulated behavior*), or because they voluntarily and purposely used this strategy to monitor their learning progress (i. e., *self-regulated behavior*). Furthermore, these studies did not examine whether students' SRL activities related to monitoring and self-evaluation were predictive of subsequent changes in study behaviors and course performance. A promising approach to further distinguish between self-regulated and externally regulated study activities in (online) course environments is to combine behavioral trace data with (a) self-reported SRL behaviors, and (b) pertinent information about students' learning context (e. g., course requirements). The combination of different sources of information about students' study activities would enable analyses of whether students' study behaviors, as observed via digital trace data, are driven by students' SRL skills and purposefully selected learning strategies, by course requirements, or by a combination of both.

Linking Students' Self-reported SRL with Digital Trace Data

Hence, a relevant question for SRL researchers is to what extent self-reported data and digital trace data on self-regulation overlap, diverge, or complement each other in predicting performance and improving our understanding of SRL in authentic contexts (Baker et al., 2020; Bernacki, 2018). Some of the above-mentioned studies used digital trace data and survey data to investigate students' SRL behaviors in post-secondary education. Results on correlations between self-reported SRL skills and SRL behaviors measured with digital trace data are mixed. Huang et al. (2022) showed that both self-reported self-efficacy and the use of metacognitive strategies (i. e., planning and monitoring) measured with digital trace data predicted course grades. However, the two types of measures were not significantly correlated. Similarly, Cicchinelli et al. (2018) found no significant correlations between students' self-efficacy and overall study activity, time spent on coursework, and monitoring and planning activities. However, they found moderate positive correlations between students' self-reported self-regulation skills and self-regulation measures derived from digital trace data. Li et al. (2020) assessed self-reported self-regulation skills with surveys at the beginning (T1) and end (T2) of a quarter. Time management measured with digital trace data in online lectures correlated positively with self-reported self-regulation skills measured only at T2. Ifenthaler et al. (2022) focused on self-testing strategies and found that students who reported using more self-testing strategies in their courses also engaged in more self-assessment tasks in the LMS of their course. Although not entirely consistent, these findings point to positive associations between self-reported data and digital trace data on SRL skills. A key factor that may contribute to these inconsistencies is the timing of measurement and the level of generality of different types of measures. For instance, students' self-reported broader motivational beliefs such as generalized academic self-efficacy (Cicchinelli et al., 2018; Huang et al., 2022) and self-reported SRL assessed before their course had started (Li et al., 2020) are often not significantly related to students' digital trace data collected during the semester. In contrast, when students report on their SRL after they have already participated in the course for a few weeks, the associations between students' self-reported SRL and their digital trace data tend to be stronger (Cicchinelli et al., 2018; Li et al., 2020). These findings indicate that the time point and specific aspects of SRL (e.g., self-testing to monitor learning progress) measured with self-reported data and digital trace data should be aligned when these measures are being used complementarily.

4. Self-regulated Learning in College: An Illustrative Example of Using Digital Trace Data

In the second part of this chapter, we provide an example of how digital trace data can be used to describe students' study behaviors over time. Based on a collaboration between TU Dortmund and the University of California, Irvine, we were able to use data from the UCI-MUST project (Arum et al., 2021), an ongoing longitudinal study to examine undergraduates' experiences and factors that facilitate college success. We used data from undergraduate students who were enrolled in two large biology lectures in the fall of 2020 and were using course materials that were provided to them via LMS. Importantly, all students were studying remotely due to the COVID-19 pandemic and the social distancing rules that were in place at that time. Consequently, the LMS used at the time includes rich data on students' digital traces and learning behaviors. We focused on a selected subsample of students and examined the pattern of students' study activities across the ten weeks of the academic fall quarter. Our analyses focus on three key research questions:

1. Is variability in study activities on LMS across the quarter associated with critical course events (i. e., midterm exams)?
2. Are global and week-specific measures of study activities on LMS across the quarter associated with students' demographic background variables and final course grades?
3. Is students' self-reported self-efficacy for self-regulated learning related to their study activities on LMS across the quarter?

4.1. Sample and Procedure

We used data from *N1* = 805 undergraduate students who were enrolled in two large biology lectures in their junior year in the fall term of 2020. These biology lectures are usually face-to-face lectures and instructors use the LMS Canvas to provide course materials and assignments. In the fall term of 2020, the lectures were shifted to a fully remote format because of the COVID-19 pandemic. Canvas was used to provide lecture content and administer assignments and midterm exams. The sample consisted of a diverse student population with 44 % first-generation college students, 25 % students who belonged to a historically underrepresented minority (Latino, African American, Pacific Islander), and 67 % female students. We used different data sources to examine the proposed research questions.

Digital Trace Data. We used the overall number of clicks each student used per day in their Canvas course as a quantitative measure of study activity. This measure includes, for example, clicks on course materials, downloads of course material, uploads of assignments, and completion of quizzes and midterm exams. First, we aggregated daily click activities on a weekly level per student. Second, we centered students' weekly click activities at the course mean to remove between-course variance in the two biology lectures (within-course centering), and thus created comparable study activity measures in both lectures. We decided to aggregate daily click activities on a weekly level for two reasons: First, we were interested in the variability in students' click activity across the entire term and during significant week-specific course events (midterm and final exams). Second, the two lectures had the same structure and exams happened in the same weeks, but lectures and exams took place on different weekdays. Aggregation of click activities on a weekly level thus improved comparability of the two courses.

Course Syllabi. Course syllabi are detailed course plans that were available for both lectures. We used the course syllabi to identify the dates of midterm exams. In both lectures, midterm exams took place in weeks 3, 6, and 9 of the fall term of 2020.

Administrative Data. Data on students' demographic backgrounds and final course grades were obtained from students' college records. We used dichotomous variables as indicators of students' first-generation college-going student status (1 = yes; 0 = no), if students belonged to a historically underrepresented minority (URM) on campus (1 = yes; 0 = no), and about students' biological gender (1 = female; 0 = male). Administrative data included an option 'other' to declare students' gender, but all students in the present sample had a record of either female or male gender. Furthermore, we used high school grade point average (GPA) as an indicator of prior achievement. We used final grades as an indicator of course performance. Students received letter grades (A – F) in both lectures. We transformed the letter grade to a numeric variable (A = 12 to F = 0), with higher values indicating better performance.

Survey Data. A small subsample of the 805 students in the selected biology lectures participated in surveys during the UCI-MUST project. In the UCI-MUST project, more than 1.200 undergraduates from all fields of study consented to participate in the survey study of the project. Of those, 25 students were enrolled in the two biology lectures that are presented in the illustrative example of this chapter. These students completed a survey at the beginning of the fall 2020 quarter that included questions about their self-efficacy for self-regulated learning (T1; *N2T1* = 25), and 18 of these stu-

dents completed the same questions again after the fall term of 2020 (T2; *N2T2* = 18). We used five items to measure self-efficacy for self-regulated learning that were based on the self-efficacy scale by Farr et al., (2011). Two example items are: "How good are you at motivating yourself to do schoolwork" and "How good are you at finishing your homework assignments by deadlines". Students responded to the items on a slider scale from 0 – *not at all good* to 100 – *exceptional*. Because of the very small sample sizes, we treat findings of analysis with survey data as preliminary suggestive evidence that needs to be extended and continued with larger samples.

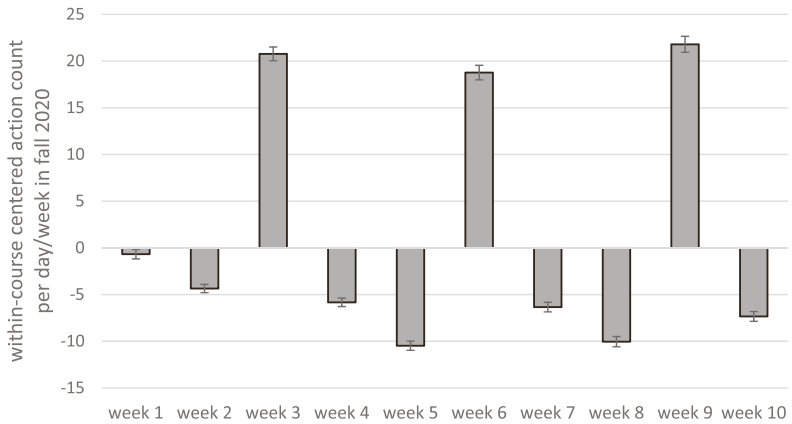
We used descriptive statistics and correlative analyses to describe the pattern of study activities across the quarter and the associations of study activities with demographic variables, course grades, and self-reported self-efficacy for self-regulated learning.

4.2. Results

Associations of Variability in Click Activities Across the Quarter with Critical Course Events (RQ1)

On average, students had 27 action counts on each day in a week in their course. A large standard deviation and a large range from minimum to maximum action counts per day indicate large variability in click activity across days of the quarter (*min* = 1, *max* = 734, *M* = 26.92, *SD* = 28.29). Figure 1 shows the pattern of click activity on a weekly level across the fall 2020 quarter. Overall, the pattern shows a small decline in click activities across the 10 weeks of the quarter with large increases in click activities in weeks 3, 6, and 9 compared to the mean click activity in the course, and decreased click activities in weeks 4, 5, 7, 8, and 10. Information from available course syllabi explained these fluctuations in click activities across the week: Click activity increased in weeks of midterm exams (weeks 3, 6, and 9), and decreased in weeks after the midterm exams. This finding emphasizes the importance of using relevant context information about the courses. Information on relevant deadlines and exams is central for a meaningful interpretation of study patterns.

Figure 1: Behavioral trace data from two biology lectures in fall 2020. N1 = 805 undergraduate students. Within-course centered action counts per day aggregated on a weekly level. Error bars indicate standard errors.



Associations of Click Activities with Students' Demographic Background and Course Performance (RQ2)

In a second step, we explored associations of click activities across the quarter with students' demographic backgrounds and their course performance. Table 1 shows the results of bivariate correlations of these variables. Overall, click activities differed only slightly by students' demographic backgrounds. Female students had slightly more click activities across the quarter (small positive correlations of female gender with fall 2020 overall study activity r = .09; and with study activity in weeks 1, 4, 5, 6, and 10 r = .07 to r = .11). Students of an underrepresented minority had slightly lower study activity in weeks of the midterm exams (r = -.10 to r = -.13). First-generation and continuing-generation college students did not differ systematically in their click activities across the quarter. Furthermore, high school GPA was not significantly associated with click activities during the quarter.

Students' final course grade had a small positive correlation with students' overall click activity across the quarter (r = .15) and small positive correlations in several weeks of the quarter (weeks 2 to 4: r = .11 to r = .17; weeks 6 to 9: r = .12 to r = .14). Hence, click activity in the weeks of the midterm exams, as well as the weeks before and after the midterm exams was significantly correlated with students' final course grades.

Table 1: Bivariate correlations of within-course centered action counts per week with demographic variables and final grade.

	f20 action count	week 1 action counts	week 2 action counts	week 3 action counts	week 4 action counts	week 5 action counts	week 6 action counts	week 7 action counts	week 8 action counts	week 9 action counts	week 10 action counts
Female	**0,09**	**0,08**	0,04	0,04	**0,09**	**0,07**	**0,07**	**0,11**	0,06	0,06	**0,07**
Underrepresented minoriy	-0,04	-0,01	-0,03	**-0,13**	0,00	0,05	**-0,10**	0,01	0,04	**-0,11**	-0,04
First-generation college student	0,01	0,04	0,04	-0,04	-0,03	0,03	-0,01	0,02	0,06	-0,05	0,02
High school GPA	0,03	-0,03	-0,01	0,04	0,01	-0,01	0,02	0,03	0,06	0,04	0,02
Final grade	**0,15**	0,04	**0,11**	**0,17**	**0,12**	0,06	**0,12**	**0,13**	**0,11**	**0,14**	0,05

Note. *N1* = 805 students. Within-course centered weekly action counts.
Greyed cells – weeks with midterm exam (week 3, 6, 9). Bolt font $p < .05$.

Associations Between Self-reported Data and Digital Trace Data (RQ3)

Finally, using data from a small subsample of students who participated in the UCI-MUST project surveys, we examined the associations between students' self-reported self-efficacy for SRL and students' click activities across the quarter (*N2T1* = 25; *N2T2* = 18). Results shown in Table 2 indicated that students with higher self-efficacy for self-regulation at the beginning of the fall 2020 quarter (T1) had higher click activities in several weeks across the quarter. This association was large and statistically significant at the beginning of the quarter and in the weeks around the first two midterm exams (weeks 1 to 3: $r = .41$ to $r = .50$; weeks 5 to 6: $r = .40$ to $r = .65$). The associations between students' self-efficacy for SRL measured at T2 (shortly after the fall 2020 term) and their study activities were positive, but not statistically significant in most weeks of the term. Large standard errors in the small survey sample at T2 (*N2T2* = 18) are likely a contributing factor to the nonsignificant results.

Table 2: Bivariate correlations of within-course centered action counts per week with self-reported self-efficacy for self-regulation.

	final grade	f20 action count	week 1 action counts	week 2 action counts	week 3 action counts	week 4 action counts	week 5 action counts	week 6 action counts	week 7 action counts	week 8 action counts	week 9 action counts	week 10 action counts
T1 SRL self-efficacy	0,13	**0,56**	**0,41**	**0,50**	**0,50**	0,15	**0,65**	**0,41**	0,34	0,16	0,26	*0,39*
T2 SRL self-efficacy	*0,43*	**0,47**	0,22	**0,48**	*0,45*	0,35	0,21	0,32	0,19	-0,17	*0,46*	0,39

Note. *NT1* = 25 students, *NT2* = 18 students. Within-course centered weekly action counts.
Greyed cells – weeks with midterm exam (week 3, 6, 9). Bolt font $p < .05$. Italic font $p < .10$.

4.3. Discussion

Descriptive and correlative findings provided in the example with combined digital trace data, course syllabus data, and survey data from the UCI-MUST project are consistent with previous literature. Variability in click activities across the weeks of the quarter and the increased number of clicks during the midterm exam weeks highlight the importance of considering course design features (i. e., exam weeks) when interpreting students' study patterns in courses. Such information can be obtained, for instance, from the course syllabi. Furthermore, our findings indicated that students' overall click activities were positively associated with final course grades. These findings corroborate previous findings (Cicchinelli et al., 2018; Nguyen et al., 2020; Park et al., 2017). Our findings further suggest that students' weekly click activities might provide valuable information on their SLR behavior. Furthermore, results indicate that click activities in specific weeks – i. e., immediately before, during, and immediately following an exam – might be particularly predictive of students' final grades.

Similar to results from prior studies (Nguyen et al., 2020; Rodriguez et al., 2021), our findings showed that students' click activities varied among students with different demographic characteristics. Female students had slightly more click activities compared to male students, whereas students from historically underrepresented minorities showed fewer click activities in their digital trace data during the weeks of the midterm exams. It is important to note that we cannot infer the causes of differing click activity patterns. Female student' higher click activities might be related to higher levels of conscientiousness. Prior SRL research has shown that more conscientious students are better in managing their time and regulating their effort in education (Douglas et al., 2016; McCrae & Löckenhoff, 2017; Waldeyer et al., 2022), and personality research has shown that female students report higher levels of conscientiousness than their male counterparts (Costa et la., 2001; Schmitt et al., 2008). Female students are also more likely to report higher test anxiety (Cassady & Johnson, 2002; Costa et la., 2001) and hence, higher click activity of female students might also be the result of increased learning activities driven by anxiety before and during exam weeks. Lower levels of click-activities among students of underrepresented minorities could be explained by other obligations, such as jobs besides studying, that might conflict with their time and resources for study activities, or by deficient SRL strategies. Further information on students' characteristics (e. g., personality traits, motivation, and goals) and their study and living situation (e. g., on other responsibilities besides studying) can be obtained through surveys, and are needed to explain variability in click activities among stu-

dents. Thus, these remaining open questions indicate a need to use multiple data sources.

Our findings further showed positive associations between students' self-efficacy for SRL and their amount of click activities in the course overall, as well as in the weeks before and during exams. These findings are consistent with prior evidence of positive associations between specific SRL behaviors assessed through self-reports and digital trace data (e. g., Cicchinelli et al., 2018; Ifenthaler et al., 2022; Li et al., 2020). In our study, click activities across the quarter were positively associated with self-efficacy for SRL measures at the beginning (T1) and end of the quarter (T2), with slightly stronger associations with T1 measures. These findings are in contrast to the findings by Li et al. (2020) who reported stronger associations between self-reported data and behavioral trace data at the end of a course. These authors proposed that students rated their SRL skills based on their real experiences in the course at the end of a quarter, which likely led to more accurate self-reported SRL skills at the second time point. In our study, students were asked about their self-efficacy for SRL in general and not regarding the specific course. This might explain why associations between self-efficacy for SRL and study activities were not stronger towards the end of the quarter.

Overall, positive associations between self-report data and digital trace data on SRL and study activities point to the potential of combining both data sources to investigate study-related behaviors and academic performance: Behavioral trace data can provide measures of students' real-time study behaviors in authentic contexts. Self-report data can provide important information about a) internal cognitive and meta-cognitive aspects of SRL behavior, as well as students' self-efficacy for SRL, and b) self-report data can provide relevant information to validate new SRL measures based on digital trace data.

5. Conclusion

This contribution aimed to provide an overview of current approaches on how to investigate adaptive SRL behaviors of college students with digital trace data. Digital trace data seems particularly useful to measure SRL behaviors related to students' time management and aspects of monitoring behavior (e. g., through self-assessments with quizzes). While broad measures of study activities, such as the overall number of clicks and time spent on courses are moderate predictors of performance, more fine-grained measures, such as changes in click activities towards a deadline or using voluntary quizzes

to monitor learning, are particularly promising to identify students' use of SRL strategies and associations with subsequent performance outcomes. To interpret such measures, it is necessary to take into account information about course design and context. Course syllabi and course plans can be reliable sources to obtain such relevant information (as done in the empirical example described above). Future research on SRL in college should continue to examine how information from survey data and digital trace data can be combined to investigate factors that facilitate or hinder SRL and performance in college. The above-mentioned studies successfully used digital trace data to describe adaptive and maladaptive study patterns. However, a remaining question is whether adaptive study behaviors were driven by external course designs and demands (i. e., *externally regulated* behavior), or through individual and purposefully used SRL strategies (i. e., *self-regulated* behavior). By combining digital trace data, survey data, and course syllabus data, future research could investigate the extent to which intraindividual and course contextual factors contribute to explaining variance in study behaviors and course performance. Multiple source data would allow, for instance, to investigate if regular self-testing behavior and subsequent course performance are the result of course requirements and grading policies, or of students' individual motivation and SRL skills.

References

Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., & Gates, K. M. (2022). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-022-01939-9

Arum, R., Eccles, J. S., Heckhausen, J., Orona, G. A., Keyserlingk, L. von, Wegemer, C. M., Wright, C. E., & Yamaguchi-Pedroza, K. (2021). A Framework for Measuring Undergraduate Learning and Growth. *Change: The Magazine of Higher Learning*, *53*(6), 51–59. https://doi.org/10.1080/00091383.2021.1987810

Baker, R., Di Xu, Park, J., Yu, R., Li, Q., Cung, B., Fischer, C., Rodriguez, F., Warschauer, M., & Smyth, P. (2020). The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: opening the black box of learning processes. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-00187-1

Berger, F., Schreiner, C., Hagleitner, W., Jesacher-Rößler, L., Roßnagl, S., & Kraler, C. (2021). Predicting Coping With Self-Regulated Distance Learning in Times of COVID-19: Evidence From a Longitudinal Study. Frontiers in Psychology, 12, 701255. https://doi.org/10.3389/fpsyg.2021.701255

Bernacki, M. (2018). Examining the cyclical, loosely sequenced, and contingent features of self-regulated learning: Trace data and their analysis. In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 370–387). Routledge Taylor& Francis Group. https://doi.org/10.4324/9781315697048-24

Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. *Computers & Education*, *158*, 103999. https://doi.org/10.1016/j.compedu.2020.103999

Boekaerts, M. (1999). Self-regulated learning: where we are today. *International Journal of Educational Research*, *31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Boekaerts, M. (2010). *Handbook of self-regulation*. Academic Press.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Cassady, J. C., & Johnson, R. E. (2002). Cognitive Test Anxiety and Academic Performance. *Contemporary Educational Psychology*, *27*(2), 270–295. https://doi.org/10.1006/ceps.2001.1094

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. In A. Pardo, K. Bartimote-Aufflick, G. Lynch, S. B. Shum, R. Ferguson, A. Merceron, & X. Ochoa (Eds.), *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 191–200). ACM. https://doi.org/10.1145/3170358.3170381

Cogliano, M., Bernacki, M. L., Hilpert, J. C., & Strong, C. L. (2022). A self-regulated learning analytics prediction-and-intervention design: Detecting and supporting struggling biology students. *Journal of Educational Psychology.* Advance online publication. https://doi.org/10.1037/edu0000745

Costa, P., JR., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322–331. https://doi.org/10.1037//0022-3514.81.2.322

Crompton, H., Bernacki, M., & Greene, J. A. (2020). Psychological foundations of emerging technologies for teaching and learning in higher education. *Current Opinion in Psychology*, *36*, 101–105. https://doi.org/10.1016/j.copsyc.2020.04.011

Dörrenbächer, L., & Perels, F. (2016). Self-regulated learning profiles in college students: Their relationship to achievement, personality, and the effectiveness of an intervention to foster self-regulated learning. *Learning and Individual Differences*, *51*, 229–241. https://doi.org/10.1016/j.lindif.2016.09.015

Douglas, H. E., Bore, M., & Munro, D. (2016). Coping with University Education: The relationships of Time Management Behaviour and Work Engagement with the Five Factor Model Aspects. *Learning and Individual Differences*, *45*, 268–274. https://doi.org/10.1016/j.lindif.2015.12.004

Farr, B., Rotermund, S., Ho, P., Radwin, D., & Robles, J. (2011). Evaluation of the Academy for College Excellence: Year 1 interim report.

Greene, J. A., Plumley, R. D., Urban, C. J., Bernacki, M. L., Gates, K. M., Hogan, K. A., Demetriou, C., & Panter, A. T. (2021). Modeling temporal self-regulatory processing in a higher education biology course. *Learning and Instruction*, *72*, 101201. https://doi.org/10.1016/j.learninstruc.2019.04.002

Hong, W., Bernacki, M. L., & Perera, H. N. (2020). A latent profile analysis of undergraduates' achievement motivations and metacognitive behaviors, and their relations to achievement in science. *Journal of Educational Psychology*, *112*(7), 1409–1430. https://doi.org/10.1037/edu0000445

Huang, X., Bernacki, M. L., Kim, D., & Hong, W. (2022). Examining the role of self-efficacy and online metacognitive monitoring behaviors in undergraduate life science education. *Learning and Instruction*, 101577. https://doi.org/10.1016/j.learninstruc.2021.101577

Ifenthaler, D., & Gibson, D. (Eds.). (2020). *Advances in Analytics for Learning and Teaching. Adoption of Data Analytics in Higher Education Learning and Teaching*. Springer International Publishing. https://doi.org/10.1007/978-3-030-47392-1

Jurišević, M., Lavrih, L., Lišić, A., Podlogar, N., & Žerak, U. (2021). Higher Education Students' Experience of Emergency Remote Teaching during the Covid-19 Pandemic in Relation to Self-Regulation and Positivity. Center for Educational Policy Studies Journal, 11(Sp.Issue). https://doi.org/10.26529/cepsj.1147

Kitsantas, A., Winsler, A., & Huie, F. (2008). Self-Regulation and Ability Predictors of Academic Success During College: A Predictive Validity Study. *Journal of Advanced Academics*, *20*(1), 42–68. https://doi.org/10.4219/jaa-2008-867

Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in Massive Open Online Courses. *Computers & Education*, *127*, 41–65. https://doi.org/10.1016/j.compedu.2018.08.005

Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education*, *45*, 100727. https://doi.org/10.1016/j.iheduc.2020.100727

McCrae, R. R., & Löckenhoff, C. E. (2010). Self-Regulation and the Five-Factor Model of Personality Traits. In R. H. Hoyle (Ed.), *Handbook of personality and self-regulation* (pp. 145–168). Wiley-Blackwell. https://doi.org/10.1002/9781444318111.ch7

Nguyen, Q., Rienties, B., & Richardson, J. T. E. (2020). Learning analytics to uncover inequality in behavioural engagement and academic attainment in a distance learning setting. *Assessment & Evaluation in Higher Education*, *45*(4), 594–606. https://doi.org/10.1080/02602938.2019.1679088

Park, J., Denaro, K., Rodriguez, F., Smyth, P., & Warschauer, M. (2017). Detecting changes in student behavior from clickstream data. In A. Wise, P. H. Winne, G. Lynch, X. Ochoa, I. Molenaar, S. Dawson, & M. Hatala (Eds.), *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 21–30). ACM. https://doi.org/10.1145/3027385.3027430

Park, J., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). *Understanding student procrastination via mixture models.: Paper presentation.* 11th International Conference on Educational Data Mining (EDM), Raleigh, North Carolina.

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for leaning questionnaire (MSLQ)*. University of Michigan, National Center for Reserach to Improve Postsecondary Teaching and Learning.

Rodriguez, F., Lee, H. R., Rutherford, T., Fischer, C., Potma, E., & Warschauer, M. (2021). Using Clickstream Data Mining Techniques to Understand and Support First-Generation College Students in an Online Chemistry Course. In M. Scheffel, N. Dowell, S. Joksimovic, & G. Siemens (Eds.), *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 313–322). ACM. https://doi.org/10.1145/3448139.3448169

Waldeyer, J., Dicke, T., Fleischer, J., Guo, J., Trentepohl, S., Wirth, J., & Leutner, D. (2022). A moderated mediation analysis of conscientiousness, time management strategies, effort regulation strategies, and university students' performance. *Learning and Individual Differences*, *100*, 102228. https://doi.org/10.1016/j.lindif.2022.102228

Weinstein, C. E., & Palmer, D. R. (2002). *Learning and Study Strategies Inventory (LASSI): User's manual* (2nd ed.). H&H Publishing.

Wolters, C. A., & Won, S. (2018). Validity and the use of self-report questionnaires to assess self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 304–322). Routledge Taylor& Francis Group. https://doi.org/10.4324/9781315697048

Zimmerman, B. J. (1990). Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2
Zimmerman, B. J. (2000). Attaining Self-Regulation. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7
Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (1. publ). Routledge. https://doi.org/10.4324/9780203839010

Effekte der (mittleren) Klassenleistung auf die individuelle Leistungsentwicklung: Konzeptionelle Einordnung, aktuelle Debatten und offene Fragen vor dem Hintergrund des deutschen Schulsystems

Michael Becker, Aleksander Kocaj, Malte Jansen, Hanna Dumont, Oliver Lüdtke & Poldi Kuhl[1]

Zusammenfassung

Der vorliegende Beitrag beschäftigt sich mit der seit Jahren diskutierten Frage, welche Effekte die Komposition der Lerngruppe auf die individuelle Leistungsentwicklung Lernender nimmt. Trotz langer Forschungstradition sind die Befunde zu Effekten der Komposition der Lerngruppe auf die individuelle Leistungsentwicklung weiterhin uneindeutig. So wird zum einen darüber debattiert, ob Kompositionseffekte nur statistische Artefakte darstellen und reine „*Phantom-Effekte*" sind. Zum anderen sind die zugrundeliegenden inhaltlichen und methodischen Konzepte wie auch die methodischen Ansätze und Designs in diesem Forschungsfeld sehr heterogen, was einen Teil der Widersprüchlichkeiten in den Befunden begründen mag. Der Beitrag stellt die Debatte um Effekte der Komposition der Lerngruppe auf die individuelle Leistungsentwicklung aus Perspektive des gegliederten deutschen Schulsystems vertieft dar und diskutiert zentrale inhaltliche und methodische Fragen, die Gegenstand künftiger theoretischer Auseinandersetzungen und empirischer Untersuchungen sein sollten.

1 *W*ir danken Marko Neumann und Frauke Steinhäuser für hilfreiche Kommentare und Anmerkungen zu vorangehenden Versionen dieses Manuskriptes.

Schlagworte: Kontexteffekte; Kompositionseffekte; Leistungsgruppierung; Gegliedertes Schulsystem; Leistungsentwicklung; Mehrebenenanalyse; Forschungsdesigns

Abstract

The present study aims to discuss the still "hot topic" of composition effects of schools and classrooms on individual achievement development. Although there is a long research tradition, the effects of school and classroom composition on individual achievement development are still controversially debated. For example, there is a recent debate to which extent composition effects on achievement development are mere statistical artifacts, so-called "phantom effects." Overall, one of the issues in this research area is that the relevant methodological and content-related concepts and definitions as well as research designs are highly heterogeneous, which may explain some of this controversy and even contradictory findings. The article aims to give an overview of the debate of composition effects on individual achievement development, focusing on especially what this means for tracked school systems such as the German one. Furthermore, we discuss both central content-related and methodological questions which we think should be studied further in theoretical and empirical research.

Keywords: context effects; composition effects; ability grouping; tracking; achievement development; multi-level modeling; research designs

1. Einleitung

Dass keine Schulklasse wie die andere ist, ist einerseits eine Binsenweisheit, andererseits eine pädagogische Realität, die unterschiedliche Ursachen haben kann: Beispielweise rekrutieren Schulen mit dem Sprengelprinzip aus sehr unterschiedlichen Regionen und damit sehr unterschiedlichen Schülerpopulationen, Schüler:innen und Eltern wählen mitunter selbst unterschiedliche Schulen aus und nicht zuletzt werden gerade in der Sekundarschulzeit durch fähigkeitsgruppierende Maßnahmen (sogenanntes „*tracking*") Unterschiede zwischen Schulen, aber auch Klassen innerhalb von Schulen erzeugt (vgl. LeTendre, Hofer, & Shimizu, 2003; Steenbergen-Hu, Makel, & Olszewski-Kubilius, 2016). Im deutschsprachigen Schulsystem sind solche *tracking*-Maßnahmen vor allem in Form der Gliederung in unterschiedliche Schulformen zu finden, was – wie grundsätzlich intendiert – deutliche Ausgangsunter-

schiede in der Leistung, aber auch in anderen Schülercharakteristika (z. B. im sozialen Hintergrund) zwischen Schulen und Klassen erzeugt.

Entsprechend ist die Frage, welche Effekte von Unterschieden in der Leistungskomposition von Klassen und Schulen für die individuelle (Leistungs-)Entwicklung von Schüler:innen ausgehen, ein zentrales nationales wie internationales Thema der Bildungsforschung, -politik und -praxis. Im Kern wird also die Frage aufgeworfen, ob sich Schüler:innen mit gleichen individuellen Merkmalen (d. h. auch gleicher Leistung) in Lerngruppen, in denen die Mitschüler:innen unterschiedlich leistungsstark sind (also unterschiedliche Leistungskompositionen vorliegen), unterschiedlich entwickeln. Auf theoretischer Ebene sind vor allem zwei Prozesse anzunehmen, die Unterschiede in der Leistungskomposition für die Individualentwicklung plausibel und relevant erscheinen lassen: *peer-Prozesse* und *Instruktionsprozesse* (für einen Überblick siehe Dumont, 2021; Harris, 2010). Unter *peer*-Prozessen werden im engeren Sinne Effekte verstanden, die durch die Interaktion mit den Mitschüler:innen (*peers*) entstehen. Diese werden auch als *peer contagion* bezeichnet, wenn sich beispielsweise die Motivation der leistungsstarken Mitschüler:innen durch einen anregenden Austausch „überträgt". Dieser Argumentation folgend dienen die Mitschüler:innen im Sinne des sozialen Lernens als Vorbilder und im direkten sozialen Austausch auch als Wissensvermittler:innen (Dishion & Tipsord, 2011; Gottfried, 2015; Müller, Hofmann, Fleischli, & Studer, 2015; Tenenbaum, Winstone, Leman, & Avery, 2020). Zudem wird angenommen, dass die *Qualität der Instruktion* selbst von der Klassenkomposition beeinflusst wird (*Instruktionsprozess*). Dies hat mehrere Facetten: Einerseits können Lehrkräfte direkt auf die unterschiedlichen Kompositionen reagieren, indem leistungsstärkere Klassen schneller und kognitiv anregender unterrichtet werden als weniger leistungsstarke Klassen (Baumert et al., 2010; Dreeben & Barr, 1988; Harker & Tymms, 2004). Auch Aspekte der Klassenführung, wie gut zum Beispiel ein gemeinsames Unterrichtsgeschehen aufrechterhalten werden kann oder wie auf Störungen reagiert wird, spielen hier eine Rolle (Fauth, Atlay, Dumont, & Decristan, 2021).

Obwohl diese Prozesse theoretisch sehr plausibel sind, ist es weiterhin umstritten, ob Unterschiede in der (mittleren Leistungs-)Komposition zu unterschiedlichen Lernergebnissen für einzelne Schüler:innen führen – *unabhängig* von ihren individuellen Lernvoraussetzungen (z. B. Duru-Bellat & Mingat, 1998; Hanushek, Kain, Markman, & Rivkin, 2003; Kang, 2007; Marks, 2010; Nomi & Raudenbush, 2016; Opdenakker & Van Damme, 2001; Opdenakker, Van Damme, De Fraine, Van Landeghem, & Onghena, 2002; Tenenbaum et al., 2020). Jüngst ist diese Debatte wieder aufgeflammt, da

Arbeiten zeigen konnten, dass der differenzielle Umgang mit Messfehlern bei der Schätzung von Kompositionseffekten zu großen Unterschieden führen kann und insbesondere die Gefahr einer Überschätzung der Effekte besteht (Dicke et al., 2018; Pokropek, 2015; Televantou, Marsh, Dicke, & Nicolaides, 2021; Televantou et al., 2015). Entsprechend wird diskutiert, ob überhaupt von (positiven) Effekten der mittleren Leistungskomposition auf die Leistungsentwicklung auszugehen ist – oder ob es sich um reine statistische Artefakte handelt (sogenannte *Phantom-Effekte*; vgl. auch schon Harker & Tymms, 2004). Becker et al. (2022) konnten im Rahmen dieser Debatte zeigen, dass derlei Verzerrungen auch in deutschen Studien für die Ergebnisse zwar relevant sind, gleichzeitig aber nicht grundsätzlich davon auszugehen ist, dass Kompositionseffekte statistische Artefakte darstellen, zumindest nicht im Sinne der aufgeworfenen Debatte.

Der vorliegende Beitrag möchte hieran anschließend die Frage nach *peer*-Effekten der mittleren Klassenleistung noch einmal breiter aufgreifen und im weiteren theoretischen Kontext verorten. Schon in Becker et al. (2022) wurde darauf verwiesen, dass Kompositionseffekte durchaus breiter zu verstehen sind, als dies in den jüngeren Debatten oder auch in der deutschen Bildungsforschung meist diskutiert wird. Der vorliegende Beitrag möchte erläutern, wie diese Diskussion und die jüngsten Ergebnisse zu verorten sind. Entsprechend werden im vorliegenden Beitrag zunächst die grundlegenden Konzeptionen von Kompositionseffekten vorgestellt. Dann wird erläutert, welche Spezifika für das deutsche gegliederte Schulsystem bzw. gegliederte Schulsysteme prinzipiell zu beachten sind und wie sich aus dieser Verortung die gegenwärtige Debatte um Kompositionseffekte und die Ergebnisse aus Becker et al. (2022) kontextualisieren und verstehen lassen. Der Beitrag zieht hierbei unterschiedliche theoretische Ansätze heran, maßgeblich erziehungswissenschaftliche und psychologische Erklärungen, aber auch soziologische und ökonomische. Der Beitrag schließt mit einem Ausblick auf die Diskussion um die Bedeutung von Kompositionseffekten und Implikationen für Forschung und Praxis.

2. Kompositions- und Peer-Effekte: Zentrale Dimensionen

Kompositionseffekte beschreiben ganz allgemein alle Einflüsse, die von der Komposition der Lerngruppe, also den Mitschüler:innen, die ein Individuum umgeben, ausgehen. Da es um die *peers* und ihre Effekte auf die Individualentwicklung geht, werden Kompositionseffekte auch als *peer*-Effekte bezeichnet

(Sacerdote, 2011). Diese beiden Begriffe werden im Folgenden weitgehend synonym gebraucht, wenn es nicht anders vermerkt wird.

Kompositionseffekte werden auch definiert als „all externalities that spill over from peers' or peers' family background or current actions" (Sacerdote, 2011, S. 250). Sie umfassen entsprechend alles, was direkt oder indirekt von den *peers* ausgeht. Mit direkten Effekten wären die oben genannten *peer*-Prozesse gemeint, wenn also Mitschüler:innen sich direkt gegenseitig im Kontakt beeinflussen. Mit indirekten Effekten meint man vor allem die oben beschriebenen *Instruktionsprozesse,* z. B. die Reaktionen der Lehrkräfte auf die Komposition einer Schulklasse, ob schneller oder langsamer unterrichtet wird in Abhängigkeit der Leistungsfähigkeit der Lerngruppe/*peers* etc.[1] Dies kann konzeptuell soweit gehen, dass auch als *peer*-Effekt gelten kann, wenn sich etwa die Eltern der *peers* um einen Wechsel der Lehrkraft und besseren Unterricht bemühen. Ein Effekt der *peers* und der Komposition kann also unterschiedlich proximal oder distal entstehen. Gleichzeitig können die Effekte auf anderen Ebenen auch wieder unterschiedlich interpretiert und verstanden werden – im Falle der vorangehend erwähnten Bemühungen, dass *peer*-Eltern eine andere Lehrkraft und anderen Unterricht herbeiführen, wäre dies auch als Lehrkraft- oder Unterrichtseffekt interpretierbar. Dies ist eine grundlegend definitorische und konzeptuelle Problematik.

Konzeptionell lohnt es sich zu überlegen, welcher Aspekt der Komposition der Mitschüler:innen einen Effekt auf das Individuum haben kann. In der Bildungsforschung und Pädagogischen Psychologie wird besonders prominent der Effekt des Mittelwertes untersucht; ob also vom mittleren Niveau der Gruppe ein Effekt ausgeht (für Übersichten vgl. z. B. Dumont, 2021; Rjosk, 2022; van Ewijk & Sleegers, 2010a, 2010b). Dies wird in der Literatur auch als *linear-in-means*-Modell beschrieben (vgl. Sacerdote, 2011, sowie Tabelle 1). Darüber hinaus lässt sich eine Reihe weiterer Effekte oder Modelle beschreiben, wie die Komposition der Lerngruppe wirken kann, etwa das *bad apple*-Modell, nach dem einzelne *peers* die Entwicklung der Mitschüler:innen stören können, oder der umgekehrte Fall, dass einzelne Personen als leuchtendes Beispiel die Mitschüler:innen positiv beeinflussen (*shining light*-Modell). Diese Effekte können vermutlich auch parallel auftreten. So wäre beispielsweise denkbar, dass sowohl das *linear-in-means*-Mo-

1 Sacerdote (2011) gibt folgende, sehr anschauliche Beispiele hierzu: „[I]f a student's classmates have higher incoming ability and the student learns directly from her classmates, that is a peer effect. If the classmates have higher incoming ability and this enables the teacher to teach at a higher level or a more demanding pace, that is also a peer effect. If the student is disruptive and consumes more of the teacher's attention, thereby reducing her classmates' test scores, that too is a peer effect […] If the student develops an interest in athletics or in shoplifting because of her peers, those are also peer effects." (Sacerdote 2011, S. 250 f.).

dell zutrifft, also die Leistungsentwicklung in Abhängigkeit des mittleren Niveaus der Klasse kovariiert, als auch gleichzeitig *bad-apple*-Effekte auftreten, also einzelne disruptive Schüler:innen die Entwicklung der Mitschüler:innen zusätzlich negativ beeinträchtigen. Mit Blick auf gegliederte Schulsysteme ist vermutlich insbesondere das *boutique*-Modell interessant: Hier wird vermutet, dass Personen vor allem dann besonders gute Lernleistungen zeigen, wenn sie mit leistungsähnlichen *peers* unterrichtet werden, was zumindest theoretisch der Grundintention der schulischen Gliederung entspricht. Nicht zuletzt sind Aspekte der Homo- oder Heterogenität der Lerngruppe von Bedeutung, was in Tabelle 1 unter *focus*- oder *rainbow*-Modellen aufgeführt wird. Empirisch finden sich für diese unterschiedlichen Prozesse in der Literatur empirische Evidenzen, die für die potentielle Relevanz jeder dieser einzelnen Komponenten bzw. Facetten von Komposition sprechen (vgl. u. a. Carrell & Hoekstra, 2010; Feld & Zölitz, 2017; Hoxby, 2000; Sacerdote, 2011). In der deutschen Forschung werden Kompositionseffekte, wie eingangs erwähnt, vor allem aus der Perspektive von *linear-in-means*-Modellen, gelegentlich auch in Hinblick auf Klassenhomo- oder -heterogenität (also *focus*- oder *rainbow*-Modelle) untersucht (für eine knappe Übersicht vgl. jüngst Rjosk, 2022).

Tabelle 1: Unterschiedliche Modelle möglicher *peer*-Effekte

Modell	Effekt-homogenität	Beschreibung
Linear-in-means	Ja	Mittleres Leistungs-/Hintergrundmerkmal der Mitschüler:innen
Bad-Apple	Ja	Disruptives Merkmal stört alle anderen
Shining light	Ja	Herausragende:r Mitschüler:in wirkt als Vorbild für alle anderen
Invidious comparison	Nein	Beeinträchtigungen durch einzelne leistungsstärkere *peers*
Boutique	Nein	Leistungsoptimierung wenn mit gleichen *peers* umgeben
Focus	Ja	Homogenität der Lerngruppe ist prinzipiell gut, unabhängig der individuellen Leistung der Einzelperson
Rainbow	Ja	Heterogenität der Lerngruppe ist prinzipiell gut für alle
Single crossing	Nein	Lernende profitieren (disproportional) von der Leistungsstärke der Mitschüler:innen.

Quelle: Sacerdote (2011, S. 255).

Neben den Arten von Kompositionseffekten werden auch die Konstrukte unterschieden, nach denen sich *peer*-Kompositionen unterscheiden können. Als Merkmale für Klassen- oder Schulkompositionseffekte wird zumindest in der deutschsprachigen Forschung in der Regel die Leistungskomposition herangezogen. Gerade in Deutschland ist die Variation der Leistung vor allem in den Sekundarschulen das zentrale strukturierende Merkmal und entsprechend in der Regel im Fokus des Interesses (Fauth et al., 2021; Traini, Kleinert, & Bittmann, 2021) – gerade (aber nicht nur) weil nach der Grundschule Schüler:innen nach Leistung in unterschiedliche Schulformen aufgeteilt werden.

Entsprechend ist die Leistungsvariabilität zwischen den Sekundarschulen stark ausgeprägt (vgl. z. B. Baumert, Trautwein, & Artelt, 2003). In der internationalen Forschung, insbesondere in den Vereinigten Staaten, sind andere Kompositionsmerkmale prominenter, vor allem die soziökonomische und ethnische Herkunft der Schüler:innen (van Ewijk & Sleegers, 2010a, 2010b), was mindestens durch die starke soziale Segregation von Schulen und Schulbezirken bedingt ist, die in Teilen wiederum auf die vormalige rassistische Trennung in ein „weißes" und ein „schwarzes" Schulsystem zurückgeht (z. B. Reardon & Owens, 2014). Weitere Kompositionsmerkmale, wie etwa die Motivation oder Arbeitsdisziplin der Mitschüler:innen, sind ebenso denkbar und wurden gelegentlich untersucht (z. B. Nikolov & Dumont, 2020; Westphal et al., 2016).

Nicht zuletzt stellt sich die Frage, auf welcher Ebene analysiert wird, ob also Kompositionseffekte auf Klassen- oder Schulebene betrachtet werden. In der Regel wird die Klassenebene als die primäre und vermittelnde Ebene angesehen, in der sich die eigentlichen Lern- und Interaktionsprozesse abspielen; zumindest finden sich tendenziell größere Effekte, wenn auf der Klassenebene im Vergleich zur Schulebene analysiert wird (vgl. u. a. Eccles & Roesser, 2011; van Ewijk & Sleegers, 2010a, 2010b). Dies ist inhaltlich auch dahingehend indiziert und plausibel, da sich Effekte auf Schulebene etwa im Sinne eines Schulklimas denken lassen, das sich ebenso im Klassenzimmer manifestiert. Umgekehrt muss es aber nicht der Fall sein, dass sich Klassenmerkmale auch auf Schulebene auswirken. Gerade in Situationen, in denen innerhalb von Schulen spezifische Leistungsgruppierungen in Form von Basis- und Leistungskursen mit leistungsschwächeren und leistungsstärkeren Schüler:innen gewählt werden (wie z. B. in kooperativen oder integrierten Gesamtschulsystemen), sollte sich dieser Unterschied kaum in der Komposition auf Schulebene abbilden.

3. „Spezialfall" gegliederte Schulsysteme: Deutschland als prototypisches Beispiel

Wie eingangs schon kurz angeführt, kommt es zwischen Klassen und Schulen aus diversen Gründen zu Unterschieden in der Leistungskomposition. Durch unterschiedliche Schuleinzugsgebiete etwa treten einerseits meist implizit Unterschiede zwischen Schulen bzw. deren Klassen auf. In Folge dessen setzt sich beispielsweise die Schülerschaft einer Schule in Duisburg aus einer anderen Klientel zusammen als in Düsseldorf. Regionen und damit Bevölkerungen, aus denen sich Schulklassen rekrutieren, sind ein Grund für Unterschiede

zwischen Schulen, aber auch natürliche Schwankungen von Jahr zu Jahr produzieren Differenzen. Insgesamt wird dies auch als *implizites between-school tracking* bezeichnet (Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006).

Andererseits werden unterschiedliche Lernkontexte auch explizit herbeigeführt, maßgeblich durch sogenannte fähigkeitsgruppierende Maßnahmen (*tracking*), beispielsweise basierend auf Rationalen eines *boutique*-Modells oder *focus-/rainbow*-Modells (vgl. Tab. 1), dass Schüler:innen besser mit ähnlichen Mitschüler:innen lernen bzw. Lehrkräfte besser in leistungshomogenen Gruppen unterrichten können (vgl. auch Hattie, 2002; Kulik & Kulik, 1992; Schofield, 2010). Es lassen sich sehr unterschiedliche Formen von *ability grouping* bzw. *tracking*-Maßnahmen finden (zur Begrifflichkeit siehe u. a. Chmielewski, Dumont, & Trautwein, 2013; Steenbergen-Hu et al., 2016; Trautwein et al., 2006): *course-by-course tracking* sortiert die Schüler:innen innerhalb einer Schule für jeweils einzelne Schulfächer nach ihren Leistungsniveaus (im deutschen Schulsystem etwa in den integrierten Gesamtschulen). Als *within-school tracking* bezeichnet man Leistungsgruppierungen, die konsistent über alle Fächer hinweg innerhalb einer Schule vorgenommen werden (im deutschen Schulsystem in kooperativen Gesamtschulen oder Schnellläuferklassen in Gymnasien). (*Explicit*) *between-school tracking* bezeichnet die Aufteilung der Schüler:innen in getrennte Schulen (in Deutschland die unterschiedlichen Schulformen, wie sie in allen Bundesländern nebeneinander bestehen; Neumann, Maaz, & Becker, 2013).

Für die Analyse von Kompositionseffekten ist diese Ausgangssituation des fähigkeitsbasierten *expliziten between-school trackings* in Form von unterschiedlichen Schulformen im deutschen (Sekundar-)Schulsystem hinsichtlich drei unterschiedlicher Aspekte interessant und gleichzeitig mit großer Umsicht zu interpretieren: Erstens ist die Variabilität der Leistung zwischen den Sekundarschulen relativ hoch; der Großteil der Leistungsvariabilität ist zwischen Schulen (bzw. Schulformen), und nicht innerhalb von Schulen, zu verorten (während z. B. in PISA im OECD-Durchschnitt eher der geringere Teil, rund 40 %, zwischen den Schulen liegt; Brunner, Keller, Wenger, Fischbach, & Lüdtke, 2018). Diese hohe Variabilität zwischen Schulen stellt grundsätzlich eine gute Ausgangsbasis dar, um überhaupt größere Unterschiede in der Leistungskomposition vorzufinden, die potenziell Effekte auf die individuelle Leistungsentwicklung haben können.

Zweitens sind sowohl *peer-* als auch *Instruktionsprozesse* im deutschen Sekundarschulsystem maximiert, da gerade durch das *between-school tracking* in Form der schulischen Gliederung beide Aspekte in ihrer Unterschiedlichkeit forciert werden. Durch das *tracking* werden unterschiedliche disparate Lernmilieus konstruiert (eine Konsequenz des Trackings, die aus

einer soziologischen Perspektive entsprechend kritisch hinterfragt wird, vgl. z. B. Oakes, 1985; van de Werfhorst & Mijs, 2010). Hierdurch wird der Kontakt innerhalb von Lerngruppen intensiviert, da die Schüler:innen in unterschiedliche Schulformen (physisch) getrennt werden, was die Unterschiedlichkeit auf Gruppenebene verstärkt (aus internationaler Perspektive kompakt zusammengefasst in Chmielewski et al., 2013). Ebenso fallen die Instruktionsprozesse in den Schulformen und auch die Lehrkräfteausbildung gerade für die Gymnasien mit einer relativ stärkeren universitären Orientierung deutlich anders aus als in den anderen Schulformen (Baumert et al., 2010; Baumert et al., 2004; Gruehn, 2000; Henschel, Rjosk, Holtmann, & Stanat, 2019). Nicht zuletzt zeichnen sich die Curricula vor allem in den Gymnasien durch eine inhaltlich größere Breite und Komplexität aus (z. B. mit mehr Fremdsprachenunterricht; Becker, 2009).

In diesem Sinne ist als dritter zentraler Punkt für die Interpretation von Leistungskompositionseffekten zu beachten, dass Kompositionseffekte (im Sinne von *peer*-Effekten) mit institutionellen Effekten teilweise konfundiert sind, da ein Teil der Leistungskomposition mit unterschiedlichen institutionellen Arrangements zusammenfällt. Aus theoretischer Perspektive ist dies dahingehend unproblematisch, da der konfundierte Anteil der Leistungskomposition mit der Institution einerseits genau der (intentional herbeigeführten) Variabilität der Leistungskomposition Rechnung trägt, z. B. durch eine kognitiv anregendere Instruktion und ein anspruchsvolleres Curriculum. Dies würde man (situativ) auch als Reaktion auf Unterschiede in der Komposition erwarten; in der Literatur wurde dies auch für „natürliche" Kontextvariabilitäten nachgewiesen, z. B. in der experimentellen Arbeit von Duflo et al. (2011). Andererseits mögen die institutionellen Unterschiede zwischen den Schulformen auch partiell Anderes abbilden als institutionalisierte Reaktionen auf die Leistungskomposition, z. B. durch unterschiedliche Ausbildungen und Klientel der verschiedenen Lehramtsstudiengänge und späteren Lehrkräfte (Klusmann, 2013). Zwar stellt sich hierbei die Frage, inwiefern derlei Merkmalsunterschiede nicht auch indirekt eine Reaktion auf die unterschiedlichen Leistungskompositionen darstellt (z. B. stärker akademisierte Ausbildungen für den späteren Unterricht an Gymnasien), was sich entsprechend im (akademisierten) Unterricht widerspiegelt (Baumert et al., 2010). Traditionell werden derlei Aspekte in der deutschsprachigen Bildungsforschung nicht als Kompositionseffekte im engeren Sinne interpretiert, sondern als Institutionseffekte (Baumert, Stanat, & Watermann, 2006).

4. Spezifische Fragen: Die jüngste Debatte um Effekte der mittleren Leistung auf die Leistungsentwicklung als Artefakt

Es wurde vorangehend schon darauf hingewiesen, dass in der Debatte um Kompositionseffekte in der Bildungsforschung und Pädagogischen Psychologie der Fokus vor allem auf Effekten der mittleren *peer*-Komposition und *linear-in-means* Modellen lag. Regressionsanalytisch wird dies dahingehend konzeptualisiert, dass von einem Kompositionseffekt der Klassen- oder Schulleistung gesprochen wird, wenn nach Kontrolle der individuellen Leistung der Schüler:innen die mittlere aggregierte Leistung weiterhin einen Effekt auf die Leistung erzielt (Harker & Tymms, 2004; van Ewijk & Sleegers, 2010a, 2010b). Normalerweise werden Datensätze mit mindestens zwei Messzeitpunkten analysiert, da so die Ausgangsleistung der Klasse/Schule und der Schüler:in als Prädiktoren für die Individualleistung zum nächsten Messzeitpunkt genutzt werden können. Alternativ wird vereinzelt in querschnittlichen Designs die Leistungskomposition über die durchschnittlichen kognitiven Fähigkeiten operationalisiert, so dass der Effekt auf die individuelle Schulleistung (ggf. nach Kontrolle der individuellen kognitiven Fähigkeiten) geschätzt werden kann (Baumert et al., 2006). In der Regel wird dies als Mehrebenenregression (*multilevel models*, MLM, oder auch synonym als hierarchisch-lineare Modelle, HLM, bezeichnet; Raudenbush & Bryk, 2002) spezifiziert. MLM erlauben, in einem Modell Effekte sowohl auf individueller Ebene (L1) als auch auf Aggregatsebene (L2) zu schätzen; letzteres wäre in der Regel die Klasse oder auch die Schule. Sie sind insbesondere die Methode der Wahl, um Varianzen und Effekte auf den unterschiedlichen Ebenen zu schätzen, d. h. um beispielsweise zu bestimmen, wie die individuellen Leistungen durch die Klassenleistungen vorhergesagt werden können. Genau an dieser Stelle entzündete sich in der Forschung schon vorangehend (Harker & Tymms, 2004) und auch wieder jüngst die Frage, wann man in dieser Art von Modellen „echte" Kompositionseffekte schätzen kann: Einerseits müssen Selektionseffekte adressiert sein, anderseits müssen die Leistungsdaten hinreichend reliabel gemessen sein (Dicke et al., 2018; Harker & Tymms, 2004; Televantou et al., 2015).

Der Frage nach Selektionseffekten kommt grundlegend die größte Bedeutung zu: Um einen regressionsanalytischen mittleren Effekt als „echten" Kompositionseffekt interpretieren zu können, müssen vorangehende individuelle Unterschiede kontrolliert sein. Gerade am gegliederten Schulsystem lässt sich dies gut veranschaulichen: Unterschiede, die aufgrund der leistungsbasierten Zuweisung schon von Beginn an zwischen Schüler:innen in unterschiedlichen Sekundarschulformen bestehen, bevor der Klassen-

kontext überhaupt wirken konnte, müssen entweder per Design oder statistisch kontrolliert werden, damit man das, was man als Effekt der mittleren Leistung schätzt, als Kompositionseffekt interpretieren kann. Andernfalls wird der Kompositionseffekt durch nicht kontrollierte Selektionsprozesse überschätzt (auch als *selection bias* bzw. *pretreatment heterogeneity* bezeichnet; Morgan & Winship, 2015). In der Regel sind relevante konfundierende Merkmale dieser Art neben der Eingangsleistung auch der sozioökonomische und -kulturelle Hintergrund der Familie, das Geburtsland der Eltern oder die Herkunftssprache – Variablen also, die mit der Klassenkomposition bzw. im Falle von Schulformen mit der Eingangsselektion zusammenhängen und gleichzeitig für die Outcomes und ihre Entwicklung von Bedeutung sind (für prototypische Spezifikationen und Diskussion vgl. Dicke et al., 2018; Lauen & Gaddis, 2013; Televantou et al., 2021).

Da in querschnittlichen Studien die Ausgangsmessung des interessierenden Merkmals nicht zur Verfügung steht, sind sie schlechter in der Lage, Selektionseffekte zu kontrollieren. Zwar gibt es Arbeiten, die nahelegen, dass man z. B. mit der Kontrolle des sozialen Hintergrundes und der kognitiven Grundfähigkeiten als Proxy für die Eingangsselektivität zu relativ konsistenten Schätzungen im Vergleich zu längsschnittlichen Studien gelangen kann (vgl. Baumert et al., 2006). Allerdings ist es auch schon in einer längsschnittlichen Studie nicht trivial sicherzustellen, dass eine hinreichende Kontrolle von relevanten Eingangsunterschieden erreicht wurde. So bleiben diese Effekte stets mit einer Unsicherheit behaftet, was aber typisch ist für „kausale" Effekte, die mit nicht-experimentellen Beobachtungsdaten geschätzt wurden (Morgan & Winship, 2015; Reardon & Owens, 2014; Rosenbaum, 2002). Letztendlich muss auf einer inhaltlichen und theoretischen Ebene argumentiert werden, inwiefern die Auswahl der Kontrollvariablen hinreichend ist, um relevante Eingangsunterschiede und damit die Stichprobenselektivität zu kontrollieren. Dies fällt entsprechend in längsschnittlichen Designs mit Kovariaten, die sowohl mit der Selektion als auch mit dem Outcome hoch korrelieren, leichter und lässt sich besser plausibilisieren (Austin, Grootendorst, & Anderson, 2007; VanderWeele, 2019), als dies vor allem für querschnittliche Designs gilt. Prinzipiell konnte in experimentellen Studien gezeigt werden, dass die Ausgangsmessung („pre-treatment") des zu interessierenden (Leistungs-)Merkmals, also bevor das „Treatment", d. h. der Kontext / Komposition gewirkt haben kann, zentral ist bzw. dazu beitragen kann, um mittels Matching-Verfahren oder gar regressionsanalytisch zu einer unverzerrten Schätzung des Treatment-/Kontexteffektes zu gelangen (Cook, Steiner, & Pohl, 2009; Pohl, Steiner, Eisermann, Soellner, & Cook, 2009; P. M. Steiner, Cook, Shadish, & Clark, 2010). Dies ist auf der theo-

retischen Ebene insofern plausibel, da durch die Kontrolle der Ausgangsmessung auch andere Heterogenitätsmerkmale und deren Korrelationen zu dem Merkmal selbst bzw. mit der Selektion implizit kontrolliert werden.[2]

In diesem Punkt wird neben der Kontrolle von Selektionseffekten nun das zweite Thema in der Diskussion um die konsistente Schätzung von Kompositionseffekten relevant, nämlich die Frage nach der Reliabilität der Leistungsmaße auf Ebene der individuellen Messungen (L1) sowie der aggregierten Messungen auf der Kontextebene (L2). Prinzipiell ist in MLM, ähnlich wie in jeder regressionsanalytischen Schätzung, eine reliable Messung notwendig für eine teststarke und unverzerrte Schätzung von Effekten. Mit der Aufteilung der Effekte auf unterschiedliche Analyseebenen (L1 und L2) im MLM ergeben sich unterschiedliche Dimensionen von Messfehlern sowie unterschiedliche Konsequenzen, je nachdem, wo die Fehler zu verorten sind (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Marsh et al., 2009): Neben der klassischen Reliabilität der Messinstrumente (auf L1) ist auch die Reliabilität der aggregierten Messungen auf der Kontextebene (L2) (sog. *sampling*-Reliabilität; Lüdtke et al., 2011) von Bedeutung. *Sampling*-Reliabilität meint hierbei, dass nur eine Teilstichprobe von Individuen aus jedem Kontext (d. h. Schüler:innen in Klassen oder Schulen) berücksichtigt wird. Dies kann stark variieren, da beispielsweise Klassen relativ leicht vollständig gezogen werden können, andererseits ganze Schulen selten vollständig erfasst werden. In PISA beispielsweise werden aus den einzelnen Schulen in der Regel jeweils 25–40 15-jährige Schüler:innen bzw. Neuntklässler:innen gezogen (Baumert, Stanat, & Demmrich, 2001; OECD, 2012), die entsprechend einen mehr oder weniger reliablen Schätzwert für den Schulkontext abgeben, ganz unabhängig davon, wie die eingesetzten Messinstrumente für die jeweiligen Merkmale funktionieren.

In der methodischen Forschung wurde ausführlich dargelegt, dass die Missachtung dieser Aspekte von Reliabilität zu unterschiedlichen Verzerrungen in der Parameterschätzung führen. Bei der Analyse von Kompositionseffekten wurde insbesondere intensiv diskutiert, inwieweit eine mangelnde Reliabilität (bzw. ihre mangelnde Berücksichtigung) auf indi-

2 In diesem Kontext ist auch der Hinweis relevant, dass das einfache Vorliegen längsschnittlicher Daten keine kausal informative Studie bedingt. Eine längsschnittliche Studie bietet lediglich mehr Möglichkeiten, eine kausal informative Analyse durchzuführen. Ob dies gelingt, z. B. ob vorausgehende zentrale *pre-treatment* Kovariaten hinreichend gut erfasst wurden und damit relevante Gruppenunterschiede kontrolliert werden können, ist stets im Einzelfall zu beurteilen und argumentativ zu begründen. Dies wird auch beispielsweise in der Forschung zu *directed acyclic graphs* (Steiner, Kim, Hall, & Su, 2017) betont, einem analytischen Werkzeug, mit dem man anhand von graphischen Veranschaulichungen mögliche kausale Szenarien verdeutlicht und (im Idealfall) nachvollziehbar plausibilisiert.

vidueller Ebene zu einer Überschätzung des Effektes des entsprechenden Merkmals auf Aggregatsebene führt (Harker & Tymms, 2004). Im Sinne des vorangehenden Selektionsproblems würden die individuell bedingten (Eingangs-)Unterschiede aufgrund messfehlerbehafteter Messungen nicht hinreichend korrigiert und (fälschlicherweise) als L2-Effekte im Modell abgebildet (Lüdtke et al., 2011). Im Rahmen latenter MLM können diese Reliabilitätsprobleme auf beiden Ebenen und auch unabhängig voneinander adressiert werden. Wenngleich sich in den klassischen Konstellationen der Schulforschung der Messfehler auf der Individualebene (L1) in der Regel als der problematischere für die Schätzung von Effekten zeigt, hat sich die Berücksichtigung der Messfehler auf beiden Ebenen im Rahmen von *doubly latent models* (DLM) als eine ebenfalls häufig verwendete Methode etabliert (vgl. Abbildung 1).

Abbildung 1: 2x2 Systematik von Multilevel-Regression (nach Lüdtke et al., 2011): Aufteilung nach Berücksichtigung von Messfehlern in den individuellen Messungen (Messreliabiltät; L1) oder in der Ziehung von Individuen aus den Kontexten (Sampling-Reliabilität; L2)

		Berücksichtigung der Sampling-Reliabilität	
		Nein	Ja
Berücksichtigung der Messreliabilität	Nein	Manifest-manifest	Manifest-latent
	Ja	Latent-manifest	Doppelt-latent

Das Problem der Reliabilitätsschätzungen auf den verschiedenen Analyseebenen ist in der Forschung schon länger bekannt (Harker & Tymms, 2004; Lüdtke et al., 2008). Es wurde aber gerade in der jüngeren Forschung mit den neuen Möglichkeiten durch DLM noch einmal zur Diskussion gestellt, inwiefern vorangehende Studien, die die Reliabilität der Messinstrumente und des Samplings nicht explizit modellierten, zu einer Überschätzung von (positiven) Kompositionseffekten auf individuelle akademische Leistung gelangten (Dicke et al., 2018; Pokropek, 2015; Televantou et al., 2021; Televantou et al., 2015). Diese Arbeiten konnten zeigen, dass beide Reliabilitätsprobleme auch in der Forschungspraxis durchaus von Relevanz sein können: Mitunter sind positive Effekte der Komposition auf die Leistungsentwicklung in diesen Analysen verschwunden und sogar negativ geworden (Dicke et al., 2018; Televantou et al., 2021; Televantou et al., 2015). In diesem Sinne wurde gemutmaßt, ob es sich bei Effekten der mittleren Leistungskomposition auf die Leistungsentwicklung um reine „*Phantom-Effekte*" handeln könnte. Gleichzeitig verweisen Dicke et al. (2018) und Televantou et al. (2015, 2021) darauf, dass nicht nur

die Effekte der mittleren Klassenleistung ausbleiben bzw. tendenziell sogar negativ ausfielen, sondern gleichzeitig auch negative Effekte auf psychosoziale Konstrukte, wie das akademische Selbstkonzept, nachzuweisen sind, und sogenannte *Big-Fish-Little-Pond*-Effekte (BFLPE; Marsh, 1989) auftreten: Personen fühlen sich in Klassen mit leistungsstarken *peers* weniger kompetent als mit leistungsschwächeren. Insofern ist hier sehr berechtigt zu überlegen, ob negative psychosoziale Effekte auch die Leistungsentwicklungen negativ dominieren – trotz möglicher positiver *peer contagion* oder anspruchsvolleren *Instruktionsprozessen*.

Effekte der mittleren Leistungskomposition auf die Leistungsentwicklung in Deutschland

Unmittelbar an die Diskussion um die Probleme von Messreliabilität und der Möglichkeit, dass Kompositionseffekte eigentlich „Phantom-Effekte" sein könnten, schließt sich die Frage an, ob derlei Effektmuster auch für das deutsche Schulsystem plausibel sind. So wissen wir z. B. von sehr starken Schulformeffekten, die weitgehend ohne die identifizierten Reliabilitätsprobleme geschätzt wurden (Baumert et al., 2010; Becker, Lüdtke, Trautwein, Köller, & Baumert, 2012; Guill, Lüdtke, & Köller, 2017; Rjosk, Richter, Hochweber, Lüdtke, & Stanat, 2015), gleichzeitig aber Kompositions- und Schulformeffekte im deutschen System zwar stark aufeinander bezogen, aber nicht gleichzusetzen sind. Entsprechend ist konzeptuell und empirisch zu unterscheiden, welche Effekte insgesamt mit der Komposition einhergehen, welche *peer*-Effekte zusammen mit und unabhängig von den Schulformeffekten auftreten.

Die jüngsten zusammenfassenden Analysen hierzu finden sich in der Arbeit von Becker et al. (2022), die auch die vorangehenden Überlegungen zur Messreliabilität aufnehmen und versuchen, diese Art von Effekten in einem einheitlichen Framework zusammenzufassen. In der Arbeit wird einerseits methodisch geprüft, wie die Berücksichtigung der Reliabilitäten (einerseits der Messinstrumente, andererseits des Samplings von L1-Einheiten) die Schätzungen längsschnittlicher Leistungskompositionseffekte (über ein Schuljahr) beeinflussen und welche Rolle andererseits das Spezifikum eines gegliederten Schulsystems spielt. Die Arbeit wurde als eine integrative Datenanalyse (IDA; Curran & Husson, 2009) angelegt und beinhaltet Modellschätzungen auf Basis mehrerer längsschnittlicher *large-scale*-Studien (BIJU, DESI, ELEMENT, PISA-Plus2003-2004, PISA Plus 2012-2013; für Details vgl. Tabelle 2). Für jede dieser Studien wurden Modelle spezifiziert, in denen (1) der Messfehler ignoriert, (2) der Messfehler auf individuel-

ler Ebene bzw. (3) auf individueller und Klassenebene modelliert wurde (vgl. Abbildung 2), sowie zwei weitere Modelle, (4) in denen neben der Ausgangsleistung zusätzlich zentrale individuelle Kovariaten für eine starke Kontrolle der Eingangsselektivität und (5) die Schulform zur Differenzierung der absoluten Kompositionseffekte kontrolliert wurde (vgl. Abbildung 3). Die Analysen der einzelnen Studien wurden dann in der IDA (ähnlich einer Metaanalyse) zusammengefasst.

Tabelle 2: Übersicht über die Studien, die für die Schätzungen von Kompositionseffekten über jeweils ein Schuljahr in Becker et al. (2022) berücksichtigt wurden

Studien-kürzel	Klassen-stufe	Schuljahr	Studienpopulation	Leistungsdomänen
BIJU	7	1991–1992	Schüler:innen der 7. Klasse in allgemeinbildenden Sekundarschulen (Regelschulen) in 4 deutschen Bundesländern (Berlin, Sachsen-Anhalt, Mecklenburg-Vorpommern, Nordrhein-Westphalen)	Biologie, Englisch (nur Nord-Rhein-Westphalen), Mathematik, Physik
DESI	9–10	2003–2004	Schüler:innen der 9. Klasse in allgemeinbildenden Sekundarschulen (Regelschulen) in allen deutschen Bundesländern	Leseverstehen Deutsch, Englisch C-Test, Hörverstehen Englisch
ELEMENT	4–5, 5–6	2003–2005	Schüler:innen der 4., 5. und 6. Klasse in Grundschulen (in Berlin bis 6. Klasse) und grundständigen Gymnasien in Berlin	Leseverstehen Deutsch, Mathematik
PISA-Plus 2003–2004	9–10	2003–2004	Schüler:innen der 10. Klasse in allgemeinbildenden Sekundarschulen (Regelschulen) in allen deutschen Bundesländern	Mathematik, Naturwissenschaften (Gesamtscore)
PISA-Plus 2012–2013	9–10	2012–2013	Schüler:innen der 9. Klasse in allgemeinbildenden Sekundarschulen (Regelschulen) in allen deutschen Bundesländern	Leseverstehen Deutsch, Mathematik, Naturwissenschaften (Gesamtscore)

Anmerkungen: BIJU = Bildungsverläufe und psychosoziale Entwicklung im Jugend- und jungen Erwachsenenalter, DESI = Deutsch Englisch Schülerleistungen International), ELEMENT = Erhebung zum Lese- und Mathematikverständnis: Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin, PISA-Plus = *Programme for International Student Assessment Plus*.

Die Ergebnisse der IDA lassen sich für das deutsche Schulsystem dahingehend zusammenfassen, dass auch hier die unterschiedliche Berücksichtigung der Reliabilität (auf L1 und L2) einen Einfluss hat, diese aber das Auftreten von Kompositionseffekten nicht vollständig erklären kann. Erwartungskonform finden sich größere Kompositionseffekte, wenn Messfehler gänzlich unberücksichtigt bleiben, als wenn Messfehler auf individueller Ebene oder gleichzeitig auf individueller und Kompositionsebene berücksichtigt werden. Abbildung 2 gibt ein Bild davon, wie die Effekte über die Studien und Analysen hinweg variieren. Im Mittel zusammengefasst findet sich ohne Kontrolle von Messfehlern ein Effekt von $b = 0.44$ der mittleren Klassenleistung zu T1 auf die individuelle Leistung zu T2 ein Schuljahr später nach Kontrolle der individuellen Leistung zu T1 (in Abb. 2: manifest-manifest); die Kontrolle des Messfehlers auf L1 macht den größten Unterschied und reduziert den Effekt

um mehr als die Hälfte auf $b = 0.15$ im latent-manifesten Modell. Werden beide Messfehler auf L1 und L2 berücksichtigt, bleibt der Effekt der mittleren Leistungskomposition mit $b = 0.20$ ähnlich (in Abb. 2: latent-latent). Man findet also in der Tat einen deutlichen Effekt der Schätzungen je nach Modellspezifikation und einen deutlichen Hinweis darauf, dass Schätzungen von Kontexteffekten vor allem positiv verzerrt sind, wenn der Messfehler der Individualdaten (L1) nicht berücksichtigt wird.

In weiteren Modellierungen findet sich zudem ein Hinweis darauf, dass auch die zusätzliche Kontrolle weiterer sozialer und kognitiver Unterschiede, und damit eine starke Kontrolle des *selection bias,* noch einmal einen Unterschied bedingt. Dieser (als absoluter Effekt der Leistungskomposition bezeichnete) Effekt fällt dann mit $b = 0.13$ noch einmal kleiner aus als wenn lediglich die Ausgangsleistung regressionsanalytisch kontrolliert wird. Werden schließlich die absoluten Kompositionseffekte weiter ausdifferenziert in einen Anteil der Kompositionsvariabilität, der mit der Schulform (*tracking*) konfundiert ist, und einen *peer*-Effekt im engeren Sinne (ohne Konfundierung mit systemischen Komponenten), so fällt der letztgenannte unter Kontrolle der Schulform mit $b = 0.06$ etwa halb so groß aus wie der absolute Kompositionseffekt (vgl. Abb. 3). Gleichzeitig findet sich nach wie vor ein statistisch bedeutsamer Kompositionseffekt als *tracking*-Effekt mit $b = 0.14$ (nicht in Abbildung 3 dargestellt). Die beiden Effekte bestehen im deutschen Schulsystem parallel nebeneinander. Ein mittlerer *peer*-Effekt ließ sich auch innerhalb der Schulformen nachweisen: sowohl innerhalb der Gymnasien als auch innerhalb der nicht-gymnasialen Schulformen zeigte sich jeweils ein *peer*-Effekt von $b = 0.10$ (nicht in Abbildung 3 dargestellt). Der *peer*-Effekt zeigt sich also nicht nur als gemittelter Effekt über die Schulformen hinweg, sondern darüber hinaus auch innerhalb der Schulformen.

Abbildung 2: Kompositionseffekte der mittleren Leistung auf die individuelle Leistungsentwicklung mit unterschiedlich starker Kontrolle der Messfehler (keine Modellierung der Messfehler = manifest-manifest; Modellierung der Reliabilität der Messinstrumente = latent-manifest, Modellierung der Reliabilität der Messinstrumente und des Samplings = latent-latent; nach Becker et al., 2022, S. 187)

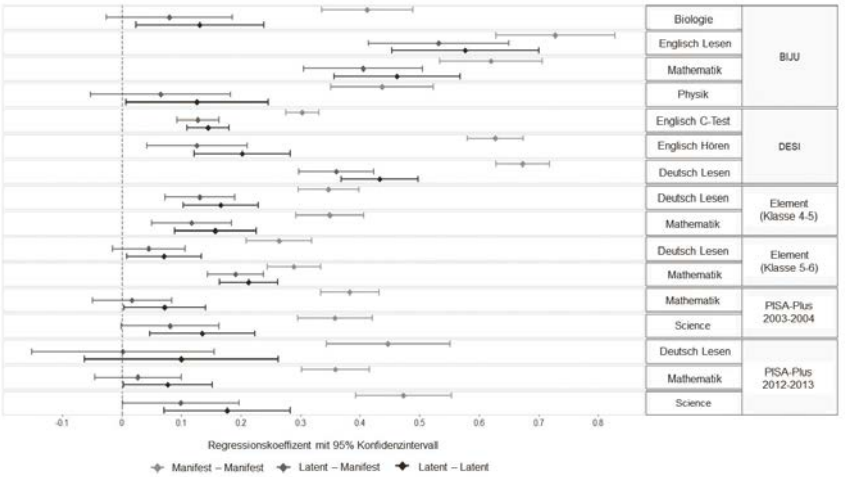
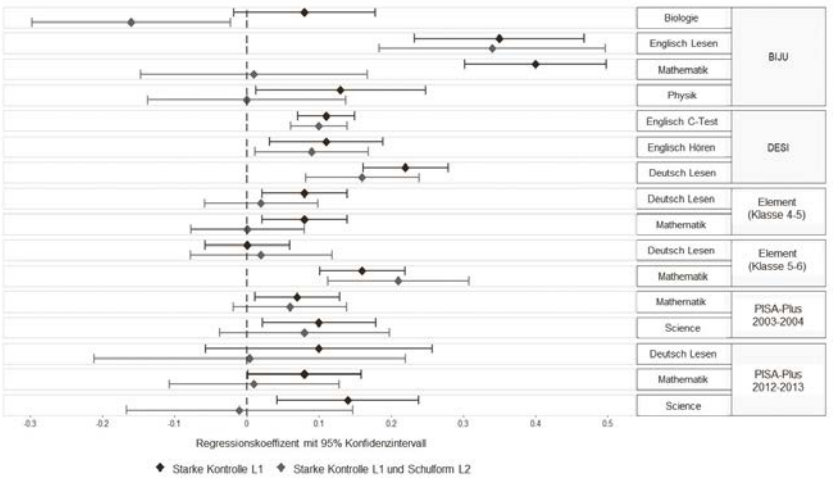


Abbildung 3: Kompositionseffekte der mittleren Leistung auf die individuelle Leistungsentwicklung mit starker Kontrolle der individuellen Eingangsselektion und zusätzlich unter Kontrolle der Schulform (nach Becker et al., 2022, S. 188)



5. Diskussion und Ausblick

5.1. Art und Größe von Kompositionseffekten

Der vorliegende Beitrag hat sich mit der Frage nach den Wirkungen der Klassenkomposition auf die Leistungsentwicklung beschäftigt. Es sollte auf einer theoretischen Ebene verdeutlicht werden, dass die Wirkweise von Kompositionen sehr vielfältig sein kann und multidimensional verstanden werden muss. Vor diesem Hintergrund greift die Frage *des* Effektes der Komposition zu kurz, da z. B. der mittlere *peer*-Effekt (als *linear in means*-Modell) nur *ein* Kompositionseffekt unter mehreren möglichen (und vermutlich parallel existierenden) Effekten ist, deren Existenz nicht zur Diskussion steht, da sie schon an verschiedener Stelle mit viel stärkeren kausaltheoretisch informierten Designs belegt werden konnte (vgl. u. a. Carrell & Hoekstra, 2010; Feld & Zölitz, 2017; Hoxby, 2000; Sacerdote, 2011).

Die Frage nach *dem* Effekt der Komposition als mittlerer *peer*-Effekt greift auch dahingehend zu kurz, als dass mittlere Effekte der Komposition mit der gleichen Methodik in einigen Kontexten nachzuweisen sind (so wie im deutschen Schulsystem), in anderen wie etwa den angelsächsischen Primarschulsystemen aber eher ausbleiben (Dicke et al., 2018; Televantou et al., 2021), und somit nicht unbedingt von einer allgegenwärtigen Existenz auszugehen ist. Hierbei ist eine weitere wichtige Erkenntnis, dass Fragen der Messreliabilität auf verschiedenen Ebenen von großer Bedeutung sind, wenngleich messtheoretische Probleme und der möglicherweise differenzielle Umgang mit Messfehlern hierbei nicht die einzige (und möglicherweise nicht die zentrale) Quelle sind, weshalb Effekte der Komposition unterschiedlich ausfallen können. Die Analysen von Becker et al. (2022) legen zumindest nahe, dass selbst innerhalb des gleichen pädagogischen Kontextes (d. h. im deutschen Schulsystem) und mit der gleichen Methodik deutliche Unterschiede zwischen den Studien auftraten. Diese Variabilität weiter zu erforschen ist, ein zentrales Desiderat, das sich aus diesen Arbeiten weiterführend ableitet.

Sämtliche jüngere Arbeiten zu mittleren Kompositionseffekten auf die Leistungsentwicklungen legen darüber hinaus nahe, dass diese Effekte in jedem Fall relativ klein ausfallen; auch der gemittelte Effekt für das deutsche System stellt hiervon keine Ausnahme dar. Dies betrifft sowohl den absoluten Effekt der Leistungskomposition als auch die einzelnen Effektkomponenten, also die *peer*-Effekte im engeren Sinne und Kompositionseffekte, die mit der Schulform konfundiert sind (Becker et al., 2022). Bemerkenswerterweise entspricht der Schätzwert des *peer*-Effektes im engeren Sinne für das deutsche Schulsystem mit $d = 0.06$ fast exakt dem Schätzwert, den auch

Hattie (2002) in seiner klassischen metaanalytischen Zusammenfassung für Kompositionseffekte auf Klassenebene geschätzt hatte (mit $d = 0.05$).

Dies hat mehrere Implikationen: Nimmt man zunächst den mutmaßlich niedrigen absoluten Betrag der Effekte, so ist es entsprechend wenig überraschend, dass diese Effekte immer wieder auch nicht gefunden werden, denn schon allein aufgrund eines Mangels an statistischer Power ist dies leicht möglich. Gerade MLM benötigen große Stichproben für die Detektion von L2-Effekten (vgl. z. B. Brunner et al., 2018). Das Problem numerisch kleiner Effekte scheint darüber hinaus mit einer großen Effektvariabilität zusammenzukommen, wie zumindest die IDA für das deutsche System nahelegt. Dies könnte durch methodische Faktoren (z. B. Art der Leistungstestung), aber auch inhaltlich bedingt sein (z. B. dass Lesekompetenzen in der Sekundarstufe weniger vom schulischen Kontext abhängen als mathematische Kompetenzen) oder z. B. *peer*- und *tracking*-Effekte je nach Domäne und Jahrgangsstufe in unterschiedlichen Anteilen überlagert sind. Welche dieser Faktoren hier zentral für die Unterschiede verantwortlich sind, kann angesichts der relativ geringen Zahl von Studien nicht abschließend geklärt werden. Es wird die Aufgabe zukünftiger Forschung sein, mit weiteren Arbeiten eine systematischere Analyse dieser Variabilität zu ermöglichen.

Jedoch stellt sich auch die Frage nach der *Größe* des Effekts: Was heißt eigentlich „kleiner" Effekt? Betrachtet man die Effekte zwischen $d = 0.06$ und oder auch $d = 0.13$ verteilungstheoretisch, so überlappen sich die Verteilungen fast vollständig und werden deswegen gemeinhin als kleine Effekte kategorisiert (Cohen, 1988). Andererseits muss bedacht werden, dass dies nur der Effekt eines Schuljahres ist, der sich aber im Verlauf der Schulzeit kumuliert – über die Sekundarstufe I gedacht wäre entsprechend von der 5. bis 10. Klassenstufe ein größerer Effekt zu erwarten. Zudem legt man zumindest in der Schulforschung für die Leistung und Leistungsveränderung noch andere, inhaltlich gut begründbare Maße an, z. B. die Veränderung innerhalb eines Schuljahres. Ein solcher (absoluter) Lernzuwachs wird je nach Altersstufe und Domäne für eher jüngere Schüler:innen in der Grundschule auf Effekte zwischen $d = 0.5$ und 0.8 geschätzt, für eher ältere Schüler:innen in der Sekundarstufe auf $d = 0.2$ und 0.4 (Bloom, Hill, Black, & Lipsey, 2008; Lanahan, McGrath, McLaughlin, Burian-Fitzgerald, & Salganik, 2005).

Relativ hierzu sind Schätzwerte zwischen $d = 0.06$ und $d = 0.13$ für die unterschiedlichen Kompositionseffekte im deutschen Schulsystem wiederum als vergleichsweise groß bzw. bedeutsam zu bewerten. Größere Effekte des Kontextes zu finden, wäre unter dieser Perspektive somit eher nicht zu erwarten. Zwar sind derlei absolute Approximationen für den Lernzuwachs über die Zeit (und konsekutiv ausgedrückt in Schuljahren) auch mit Vorsicht

zu interpretieren, z. B. aufgrund der Problematik, wie Metriken über die Zeit verlinkt werden und dadurch sehr stark variieren können (vgl. z. B. Briggs & Weeks, 2009). Für die Bewertung des Verhältnisses zwischen individuellem Zuwachs und Beitrag des Klassenkontexts sollte diese Kritik aber eher von untergeordneter Bedeutung sein, da nicht die absoluten Effekte interpretiert werden.

5.2. Interpretation der Kompositionseffekte

Becker et al. (2022) liefern einen Hinweis dafür, dass im deutschen Schulsystem einerseits von einem Effekt der Komposition auszugehen ist, andererseits ein relativ großer Anteil mit institutionellen Faktoren wie den unterschiedlichen Curricula an gymnasialen vs. nicht-gymnasialen Schulformen konfundiert ist. Dies geht konform mit der Vermutung, dass Effekte der Schülerkomposition genau dann besonders groß ausfallen, wenn sie mit anderen Faktoren wie z. B. Curriculumsvariationen zusammenkommen (Hattie, 2002; Schofield, 2010). Dies liegt nahe, wenn man bedenkt, dass begleitende Maßnahmen wie Curriculumsdifferenzierung und unterschiedliche Instruktionen gerade deshalb eingeführt werden, damit unterschiedliche Leistungsniveaus und Ausgangsleistungen optimal gefördert werden können. Die Konfundierung dieser Faktoren ist entsprechend kein Nebenprodukt, sondern Hauptziel von *tracking* und somit bewusst hergestellter Variabilität der Komposition – und so wie es mit den Schulformen der gegliederten Schulsysteme im deutschsprachigen Raum explizit intendiert ist und auch international weitgehend, wenngleich in verschiedener Ausgestaltung, der Regelfall zu sein scheint (Chmielewski et al., 2013; Gamoran, 1992; Hallinan & Kubitschek, 1999). Für die deutsche Situation enthält Letzteres wiederum ein breiteres Bündel an Mechanismen und Maßnahmen, die miteinander konfundiert sind. Einige Aspekte der *tracking*-Effekte entsprechen hierbei genau dem, was theoretisch als Wirkung von *peer*-Effekten beschrieben wird – wenn etwa der Unterricht und das Curriculum der Komposition angepasst werden. Duflo et al. (2011) haben gezeigt, dass in „natürlichen" Settings ebenfalls Anpassungen des Curriculums als spontane Reaktionen auftreten, wenn die Lehrkräfte mit unterschiedlichen Fähigkeitsniveaus und Homogenisierungsgraden der Lerngruppen konfrontiert sind. Wenngleich sich diese Aspekte rein konzeptuell unterscheiden lassen, wäre es funktional zu einem gewissen Grad arbiträr, ob eine solche Anpassung institutionell oder spontan-situativ vollzogen wird. Der Unterschied wäre entsprechend eher dort zu suchen, wo die Reaktionen

situativ auftreten oder strukturell präarrangiert sind. Hier zeigen sich diese Effekte kaum voneinander abgrenzbar bzw. ambivalent in ihrer Interpretation.

Andererseits können auch institutionelle Aspekte zu *tracking*-Effekten beitragen, die inhärent nicht direkt oder nur äußerst distal den Mechanismen von *peer*-Effekten entsprächen, z. B. die stärkere Betonung des Fachlehrkraftprinzips an Gymnasien oder Traditionen der Lehrkräftebildung. Mitunter ist auch hier eine gewisse Ambivalenz dahingehend enthalten, ob dies lediglich als institutionelle Differenz auszulegen ist oder nicht eben auch zu Unterschieden im Unterricht in Bezug auf die Komposition führt (vgl. auch Baumert et al., 2010). Empirisch ist es jedoch aufgrund der starken Konfundierung grundsätzlich in diesen Facetten schwierig, *peer*-bezogene und institutionelle Arrangements weiter auszudifferenzieren bzw. analytisch zu separieren. In der weiteren Forschung könnte man sich diesen Fragen beispielsweise nähern, wenn z. B. Schulsysteme durch die Einführung neuer Curricula reformiert werden ohne darüber hinaus weitere Aspekte des Schulsystems zu verändern.

Für die Analyse von *peer*-Effekten verweist das hier Vorgestellte auf ein grundsätzliches Problem dieser Art von *peer*-Forschung, vor allem im MLM-Paradigma: Interessiert man sich für reine *peer*-Effekte und versucht gerade diese kausal eindeutig zu belegen, dann interessiert man sich unter Umständen für einen Teil der Komposition, der eigentlich den geringsten und am wenigsten wirksamen Teil ausmacht: wenn mit Faktoren wie dem Curriculum bzw. den Lernzielen *nicht* auf die Komposition reagiert wird. Gerade wenn das Interesse ausschließlich an diesen engen Effekten besteht, ist möglicherweise das deutsche Schulsystem ein weniger gut geeignetes Feld, um dies zu untersuchen. Gleichzeitig stellt sich die Frage, ob den (kausaltheoretisch) konfundierten Faktoren aus einer pädagogischen Perspektive, die die praktische Arbeit in den Schulen im Blick hat, nicht eigentlich das Hauptaugenmerk gebühren sollte.

Neben den eben genannten Fragen der Interpretation sind auch eine Reihe methodisch-inhaltlicher Fragen offen, insbesondere inwiefern sichergestellt werden kann, dass im MLM-Paradigma Kompositionseffekte valide identifiziert werden. Grundsätzlich gilt zu konstatieren, dass das MLM-Paradigma sowie andere nicht-experimentelle Methoden von einer gewissen Unsicherheit dahingehend begleitet sind, ob die resultierenden Schätzwerte kausal interpretiert werden können. Es kommt stets in Betracht, dass es doch noch weitere unbeobachtete Faktoren gibt, die nicht kontrolliert wurden (Morgan & Winship, 2015; Reardon & Owens, 2014; Rosenbaum, 2002). Becker et al. (2022) diskutieren darüber hinaus, dass der Test von Kompositionseffekten mit MLM auch dahingehend eingeschränkt ist, dass sich mit

linear-in-means-Modellen nur beschränkt unterschiedliche Prozesse voneinander trennen lassen. So bleibt es auch ambivalent, wie genau der Effekt durch die Komposition – auch des *peer*-Effektes im engeren Sinne, wo er nicht mit der schulischen Gliederung konfundiert ist – verursacht wurde, z. B. ob durch *peer contagion,* durch *Instruktionsprozesse* als Reaktion der Lehrkräfte auf die Komposition oder durch eine Wechselwirkung dieser Prozesse (vgl. kritisch Reardon & Owens, 2014). Obwohl dies nicht kausaltheoretisch geklärt werden kann, ist es dennoch möglich, unterschiedliche Varianzquellen zu schätzen, was im Ansatz, den Becker et al. (2022) gewählt haben, durch die Trennung von *tracking*-assoziierten von im engeren Sinne *peer*-bezogenen Effekten realisiert wurde; in der Literatur finden sich hierzu auch andere Beispiele, etwa für die Trennung von Instruktions- und im engeren *peer*-bezogenen Effekten (Lavrijsen, Dockx, Struyf, & Verschueren, 2022). Allerdings bleibt es bei diesen Ansätzen bei einer Zerlegung der Varianz; um die Prozesse selbst näher zu erfassen, müssen andere Methoden hinzugezogen werden, die einzelne Komponenten auf Prozessebene identifizierbar machen. Im Falle von *peer*-Effekten ließen sich etwa Netzwerkanalysen heranziehen, in denen man tatsächlich die Interaktionen und Beziehungen der Schüler:innen nachzeichnet (Rambaran et al., 2017; Ryan, 2001). Potenziell kann hier wiederum die Problematik auftreten, dass die weitere Differenzierung der Effekte dazu führt, dass die Effekte immer kleiner werden – prototypisch belegen dies auch die Analysen aus Becker et al. (2022), wenn der etwas größere absolute Effekt der Leistungskomposition in die zwei genannten Subkomponenten differenziert wird. Möglicherweise führt dies wiederum zu einem *trade-off* zwischen aufwändigeren Designs, die Prozessanalysen erlauben, und der verfügbaren statistischen Power, die durch die Berücksichtigung einer Vielzahl von Faktoren reduziert wird.

Nichtsdestotrotz erscheinen derlei Analysen auch dahingehend lohnenswert, da der Blick auf den Effekt der mittleren Leistung und *linear-in-means* Modellen auch Effektheterogenitäten und Interaktionseffekte verdecken kann (vgl. auch oben, Tab. 1). Die Frage nach Interaktionseffekten, also danach, ob unterschiedliche Gruppen von Schüler:innen differenziell vom Leistungsniveau (oder der Leistungsverteilung generell) profitieren, ist gemäß dieser Ergebnisse eine weitere wichtige Fragestellung nach Kompositionseffekten (vgl. auch Rjosk, 2022). So konnten Lavrijsen et al. (2022) zeigen, dass im flämischen Schulsystem bei sehr geringen mittleren Effekten die Effekte für Subgruppen unterschiedlich ausfielen und leistungsstarke Schüler:innen von leistungsstarken Klassen stärker profitieren konnten, aber eben auf Kosten dessen, dass sich Schüler:innen in leistungsschwachen Klassen weniger günstig entwickelten. Hierdurch neutralisierten sich insge-

samt die differenziellen Effekte zu einem kleineren mittleren Effekt. Im Hinblick auf solche Interaktionen wurde die Literatur tendenziell dahingehend zusammengefasst, dass leistungsschwächere Schüler:innen von gemischten und damit einem insgesamt im Mittel höheren Leistungsniveau profitierten, während leistungsstärkere Schüler:innen von leistungsstärkeren und damit leistungshomogeneren Gruppen profitierten (Kulik & Kulik, 1992; Rjosk, 2022; Sacerdote, 2011; Schofield, 2010). Allerdings ist das Effektmuster sehr heterogen: Es finden sich beispielsweise auch Studien, die zeigen, dass die weniger leistungsstarken Schüler:innen von einem hohen Leistungsniveau profitierten (De Fraine, Van Damme, Van Landeghem, & Opdenakker, 2003; Duflo et al., 2011), ebenso solche, die keinerlei Hinweise auf Interaktionseffekte erbringen (Hanushek et al., 2003) oder auch einige die belegen, dass die Anwesenheit von leistungsstärkeren Schüler:innen zu ungünstigeren Leistungsentwicklungen der leistungsschwächeren Schüler:innen führen (Feld & Zölitz, 2017). Es steht auch zu vermuten, dass diese Interaktionseffekte analog zu den Haupteffekten eine deutliche Variabilität aufweisen (sei es inhaltlich oder methodisch bedingt), was es angesichts der wenigen Studien, die hierzu bislang vorliegen, schwierig macht, ein klares Fazit zu ziehen. Auch hier muss künftige Forschung weitere Beschreibungen und Differenzierungen erbringen, wovon genau diese Effektheterogenität abhängt (z. B. Domäne, Altersgruppen, Art der Kontexte).

5.3. Ausblick

Am Beispiel der jüngeren Analysen zu mittleren Kompositions- und *peer*-Effekten im deutschen System zielte dieser Beitrag darauf ab, die Diskussion um Kompositionseffekte zu weiten und entsprechend breiter zu kontextualisieren. Zu den beschriebenen Perspektiven, vor allem auf die Variabilität der Effekte, schließen sich für die weitere Forschung auch Fragen an, die sich im Wesentlichen auf die Multidimensionalität der Effekte für eine pädagogische Bewertung beziehen. Der vorliegende Beitrag beschränkte sich auf Kompositionseffekte auf die Leistungsentwicklung, um die wesentlichen konzeptionellen und methodischen Punkte zu erläutern und herauszuarbeiten. Geht es in einem nächsten Schritt auch darum zu fragen, welche Bedeutung der Komposition beziehungsweise, breiter gesprochen, dem Kontext für die individuelle Entwicklung zukommt, müssten auch weitere Perspektiven, vor allem hinsichtlich der Bildungserträge, eingenommen werden. Betrachtet man z. B. nicht nur die Leistungen, sondern auch psychosoziale Outcomes, wie etwa das akademische Selbstkonzept, so ist ein komplizierteres Effektmuster zu

erwarten, weil möglicherweise positive Effekte (z. B. auf die Leistungsentwicklung) mit negativen Effekten (z. B. auf das Selbstkonzept) kontrastieren. Hier schließt sich auch die Frage an, wie sich derlei Entwicklungen gemeinsam über die Zeit abbilden – wo möglicherweise positive Effekte längerfristigen Kosten gegenüberstehen, was eine bislang kaum adressierte Frage ist (siehe aber Dumont, Protsch, Jansen, & Becker, 2017; Stäbler, Dumont, Becker, & Baumert, 2017). Ebenso ist weitgehend ungeklärt, wie sich diese Entwicklungen über unterschiedliche Übergänge und Kontexte hinweg abbilden (siehe aber Becker & Neumann, 2018; von Keyserlingk, Becker, & Jansen, 2019) und neben den Prozessen in der Schulzeit selbst auf die nachfolgenden Übergänge und Entwicklungen übertragen (Jansen, Becker, & Neumann, 2021; von Keyserlingk, Becker, Jansen, & Maaz, 2020). Diese eher übergreifenden Fragestellungen wurden bislang insgesamt noch wenig adressiert und müssen in der nachfolgenden Forschung weiter systematisiert werden.

Hier ist auch im Blick zu behalten, welche Effekte tatsächlich von der Komposition (universell) ausgehen und inwiefern Faktoren pädagogisch unterschiedlich adressiert und modifiziert werden können. Zum Beispiel konnten Lavrijsen et al. (2022) zeigen, dass der Umgang der Lehrkräfte mit der Heterogenität der Klasse einen bedeutsameren Effekt hatte als die Heterogenität der Klasse selbst. In weiterem Sinne ist dies ein Echo dessen, worauf auch Hattie (2002, 2009) verwies, dass diese strukturelle Frage möglicherweise weniger bedeutsam ist als der pädagogische Umgang mit ihr im Unterricht. Gerade jüngere Reformen, in denen man sich allein auf die Reform der Strukturen beschränkte, aber wenig für eine pädagogische Umsetzung und Nutzbarmachung vorhielt, zeigten entsprechend dieser Überlegungen eher keine, in Teilen sogar ungünstigere Effekte auf die Entwicklung der Schüler:innen und des Systems (Baumert, Maaz, Neumann, Becker, & Dumont, 2017; Baumert et al., 2019; Becker et al., 2017). Möglicherweise sind diese negativen und ungünstigen Effekte von derlei Strukturreformen transitorisch und allein durch die Kosten der Umstellung begründet. Nichtsdestotrotz verweisen sie einmal mehr auf das Primat der Instruktion und die zentrale Stellung, die der pädagogische Umgang mit Heterogenität im Unterricht einnimmt.

Literatur

Austin, P. C., Grootendorst, P., & Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in Medicine, 26*(4), 734–753. https://doi.org/10.1002/sim.2580

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 1–48. https://doi.org/10.3102/0002831209345157

Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, & U. Schiefele (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (pp. 314–354). Waxmann.

Baumert, J., Maaz, K., Neumann, M., Becker, M., & Dumont, H. (2017). Die Berliner Schulstrukturreform: Hintergründe, Zielstellungen und theoretischer Rahmen. In M. Neumann, M. Becker, J. Baumert, K. Maaz, & O. Köller (Eds.), *Zweigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin* (pp. 9–38). Waxmann.

Baumert, J., Neumann, M., Dumont, H., Becker, M., Bachsleitner, A., Maaz, K., & Köller, O. (2019). Platzierungsentscheidungen beim Übergang in die gymnasiale Oberstufe: Folgen der Umstellung auf Zweigliedrigkeit des Sekundarschulsystems in Berlin. *Zeitschrift für Erziehungswissenschaft, 22*, 721–763. https://doi.org/10.1007/s11618-019-00880-6

Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 15–68). Leske + Budrich.

Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus [School structures and differential learning and developmental environments]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (pp. 95–188). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90082-7_4

Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten: Institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 259–330). Leske + Budrich.

Becker, M. (2009). *Kognitive Leistungsentwicklung in differenziellen Lernumwelten: Effekte des gegliederten Sekundarschulsystems in Deutschland [Cognitive achievement development in differential learning environments: Effects of the tracked secondary school system of Germany].* (PhD), Free University Berlin.

Becker, M., Kocaj, A., Jansen, M., Dumont, H., & Lüdtke, O. (2022). Class-average achievement and individual achievement: Testing achievement composition and peer spillover effects using five German longitudinal studies. *Journal of Educational Psychology, 114*(1), 177–197. https://doi.org/10.1037/edu0000519

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*(3), 682–699. https://doi.org/10.1037/a0027608

Becker, M., & Neumann, M. (2018). Longitudinal big-fish-little-pond effects on academic self-concept development during the transition from elementary to secondary schooling. *Journal of Educational Psychology, 110*(6), 882–897. https://doi.org/10.1037/edu0000233

Becker, M., Neumann, M., Radmann, S., Jansen, M., Nagy, G., Borzikowsky, C., Leucht, M., Maaz, K., Köller, O., & Baumert, J. (2017). Schulleistungen vor und nach der Berliner Schulstrukturreform. In M. Neumann, M. Becker, J. Baumert, K. Maaz, & O. Köller (Eds.), *Zweigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin* (pp. 155–188). Waxmann.

Bloom, H., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*, 289–328. https://doi.org/10.1080/19345740802400072

Briggs, D. C., & Weeks, J. P. (2009). Impact of vertical scaling decisions on growth interpretation. *Educational Measurement: Issues and Practice, 28*, 3–14.

Brunner, M., Keller, U., Wenger, M., Fischbach, A., & Lüdtke, O. (2018). Between-school variation in students' achievement, motivation, affect, and learning strategies: Results from 81 countries for planning group-randomized trials in education. *Journal of Research on Educational Effectiveness, 11*(3), 452–478. https://doi.org/10.1080/19345747.2017.1375584

Carrell, S. E., & Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics, 2*(1), 211–228. https://doi.org/10.1257/app.2.1.211

Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type: An international comparison of students' mathematics self-concept. *American Educational Research Journal, 50*(5), 925–957. https://doi.org/10.3102/0002831213489843

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Lawrence Earlbaum Associates.

Cook, T. D., Steiner, P. M., & Pohl, S. (2009). How bias reduction is affected by covariate choice, unreliability, and mode of data analysis: Results from two types of within-study comparisons. *Multivariate Behavioral Research, 44*(6), 828–847. https://doi.org/10.1080/00273170903333673

Curran, P. J., & Husson, A. M. (2009). Integrative Data Analysis: The Simultaneous Analysis of Multiple Data Sets. *Psychological Methods, 14*(2), 81-100. https://doi.org/10.1037/a0015914

De Fraine, B., Van Damme, J., Van Landeghem, G., & Opdenakker, M. C. (2003). The effects of schools and classes on language achievement. *British Educational Research Journal, 29*, 841–859. https://doi.org/10.1080/0141192032000137330

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology, 110*, 1112–1126. https://doi.org/10.1037/edu0000259

Dishion, T. J., & Tipsord, J. M. (2011). Peer contagion in child and adolescent social and emotional development. *Annual Review of Psychology, 62*(1), 189–214. https://doi.org/10.1146/annurev.psych.093008.100412

Dreeben, R., & Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education, 61*(3), 129–142. https://doi.org/10.2307/2112622

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review, 101*(5), 1739–1774. https://doi.org/10.1257/aer.101.5.1739

Dumont, H. (2021). Auswirkungen der Komposition der Lerngruppe auf die Lern- und Leistungsentwicklung. In M. Kreutzmann, L. Zander, & B. Hannover (Eds.), *Aufwachsen mit Anderen: Peerbeziehungen als Bildungsfaktor* (pp. 93–106). Kohlhammer.

Dumont, H., Protsch, P., Jansen, M., & Becker, M. (2017). Fish swimming into the ocean: How tracking relates to students' self-beliefs and school disengagement at the end of schooling. *Journal of Educational Psychology, 109*(6), 855–870. https://doi.org/10.1037/edu0000175

Duru-Bellat, M., & Mingat, A. (1998). Importance of ability grouping in French "Collèges" and its impact upon pupils' academic achievement. *Educational Research & Evaluation, 4*, 348–368.

Eccles, J. S., & Roesser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence, 21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Fauth, B., Atlay, C., Dumont, H., & Decristan, J. (2021). Does what you get depend on who you are with? Effects of student composition on teaching quality. *Learning and Instruction, 71*, 101355. https://doi.org/10.1016/j.learninstruc.2020.101355

Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics, 35*(2), 387–428. https://doi.org/10.1086/689472

Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review, 57*(6), 812–828. https://doi.org/10.2307/2096125

Gottfried, M. A. (2015). Does classmate ability influence students' social skills? . *School Effectiveness and School Improvement, 26*, 554–585. https://doi.org/10.1080/09243453.2014.988731

Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung*. Waxmann.

Guill, K., Lüdtke, O., & Köller, O. (2017). Academic tracking is related to gains in students' intelligence over four years: Evidence from a propensity score matching study. *Learning and Instruction, 47*, 43–52. https://doi.org/10.1016/j.learninstruc.2016.10.001

Hallinan, M. T., & Kubitschek, W. N. (1999). Curriculum diffrentiation and high school achievement. *Social Psychology of Education, 3*(1-2), 41–62. https://doi.org/10.1023/A:1009603706414

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*, 527–544. https://doi.org/10.1002/jae.741

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15*, 177–199. https://doi.org/10.1080/09243450512331383222

Harris, D. N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers College Record, 112*, 1163–1197. https://doi.org/10.1177/016146811011200404

Hattie, J. (2002). Classroom composition and peer effects. *International Journal of Educational Research, 37*(5), 449–481. https://doi.org/10.1016/S0883-0355(03)00015-6

Henschel, S., Rjosk, C., Holtmann, M., & Stanat, P. (2019). Merkmale der Unterrichtsqualität im Fach Mathematik [Characteristics of instructional quality in mathematics]. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich, & S. Henschel (Eds.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (pp. 335–383). Waxmann. https://doi.org/10.25656/01:18131

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation *NBER Working Papers 7867*: NBER | National Bureau of Economic Research. https://doi.org/10.3386/w7867

Jansen, M., Becker, M., & Neumann, M. (2021). Dimensional comparison effects on (gendered) educational choices. *Journal of Educational Psychology, 113*(2), 330–350. https://doi.org/10.1037/edu0000524

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Econometrics, 61*, 458–495. https://doi.org/10.1016/j.jue.2006.07.006

Klusmann, U. (2013). Individual characteristics of prospective teachers. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers – Results from the COACTIV project* (pp. 311–320). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-5149-5_15

Kulik, J. A., & Kulik, C.-L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36*(2), 73–77. https://doi.org/10.1177/001698629203600204

Lanahan, L., McGrath, D. J., McLaughlin, M., Burian-Fitzgerald, M., & Salganik, L. (2005). Fundamental problems in the measurement of instructional processes: Estimating reasonable effect sizes and conceptualizing what is import to measure. American Institutes for Research.

Lauen, D. L., & Gaddis, S. M. (2013). Exposure to classroom poverty and test score achievement: Contextual effects or selection? *American Journal of Sociology, 118*, 943-979. https://doi.org/10.1086/668408

Lavrijsen, J., Dockx, J., Struyf, E., & Verschueren, K. (2022). Class composition, student achievement, and the role of the learning environment. *Journal of Educational Psychology, 114*(3), 498–512. https://doi.org/10.1037/edu0000709

LeTendre, G., Hofer, B., & Shimizu, H. (2003). What is tracking? Cultural expectations in the U. S., Germany, and Japan. *American Educational Research Journal, 40*, 43–89.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2× 2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods, 16*(4), 444–467. https://doi.org/10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthen, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*(3), 203–229. https://doi.org/10.1037/a0012869

Marks, G. N. (2010). What aspects of schooling are important? School effects on tertiary entrance performance. *School Effectiveness and School Improvement, 21*, 267–287. https://doi.org/10.1080/09243451003694364

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44*(6), 764–802. https://doi.org/10.1080/00273170903333665

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles in social research* (2nd, rev. and enlarged ed.). Cambridge University Press.

Müller, C. M., Hofmann, V., Fleischli, J., & Studer, F. (2015). „Sag' mir, was deine Klassenkameraden tun und ich sage dir, was du tun wirst"?: Zum Einfluss der Klassenzusammensetzung auf die Entwicklung von schulischem Problemverhalten ["Tell me what your peers are doing and I will tell you, what you are doing"? On the influence of class composition on academic prolblem behavior]. *Zeitschrift für Erziehungswissenschaft, 18*, 569–589. https://doi.org/10.1007/s11618-015-0629-3

Neumann, M., Maaz, K., & Becker, M. (2013). Die Abkehr von der traditionellen Dreigliedrigkeit im Sekundarschulsystem: Auf unterschiedlichen Wegen zum gleichen Ziel? *Recht der Jugend und der Bildung, 61*, 274–292. https://doi.org/10.5771/0034-1312-2013-3-274

Nikolov, F., & Dumont, H. (2020). Das Ganze ist mehr als die Summe seiner Teile: Schulkomposition, Schulzufriedenheit und normverletzendes Verhalten [The whole is more than the sum of its parts: School composition, school satisfaction and norm violating behavior]. *Journal of Educational Research Online, 12*(1), 26–47. https://doi.org/10.25656/01:19117

Nomi, T., & Raudenbush, S. W. (2016). Making a success of "Algebra for All": The impact of extended instructional time and classroom peer skill in Chicago. *Educational Evaluation and Policy Analysis, 38*, 431–451. https://doi.org/10.3102/0162373716643756

Oakes, J. (1985). *Keeping track: How schools structure inequality*. Yale University Press.

OECD. (2012). *PISA 2009: Technical Report* Retrieved from https://www.oecd.org/pisa/pisaproducts/50036771.pdf, https://doi.org/10.1787/9789264167872-en

Opdenakker, M.-C., & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal, 27*, 407–432. https://doi.org/10.1080/01411920120071434

Opdenakker, M.-C., Van Damme, J., De Fraine, B., Van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes on mathematics achievement. *School Effectiveness and School Improvment, 13*(4), 399–427. https://doi.org/10.1076/sesi.13.4.399.10283

Pohl, S., Steiner, P. M., Eisermann, J., Soellner, R., & Cook, T. D. (2009). Unbiased causal inference from an observational study: Results of a within-study comparison. *Education & Educational Research, 31*(4), 463–479. https://doi.org/10.3102/0162373709343964

Pokropek, A. (2015). Phantom effects in multilevel compositional analysis: Problems and solutions *Sociological Methods & Research, 44*(4), 677–705. https://doi.org/10.1177/0049124114553801

Rambaran, J. A., Hopmeyer, A., Schwartz, D., Steglich, C., Badaly, D., & Veenstra, R. (2017). Academic functioning and peer influences: A short-term longitudinal study of network-behavior dynamics in middle adolescence. *Child Development, 88*(2), 523–543. https://doi.org/10.1111/cdev.12611

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2 ed.). Sage.

Reardon, S. F., & Owens, A. (2014). 60 years after Brown: Trends and consequences of school segregation. *Annual Review in Sociology, 40*, 199–218. https://doi.org/10.1146/annurev-soc-071913-043152

Rjosk, C. (2022). Dispersion of student achievement and classroom composition. In T. Nilsen, A. Stancel-Piątak, & J. E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education*: Springer. https://doi.org/10.1007/978-3-030-38298-8_47-1

Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., & Stanat, P. (2015). Classroom composition and language minority students' motivation in language lessons. *Journal of Educational Psychology, 107*(1171–1185). https://doi.org/10.1037/edu0000035

Rosenbaum, P. R. (2002). *Observational studies* (2 ed.). New York: Springer. https://doi.org/10.1007/978-1-4757-3692-2

Ryan, A. M. (2001). The peer group as a context for the development of young adolescent motivation and achievement. *Child Development, 72*, 1135–1150. https://doi.org/10.1111/1467-8624.00338

Sacerdote, B. (2011). Peer effects in education: how might they work, how big are they and how much do we know thus far? In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 249–277). Elsevier.

Schofield, J. W. (2010). International evidence on ability grouping with curriculum differentiation and the achievement gap in secondary schools. *Teachers College Record, 112*(5), 1492–1528. https://doi.org/10.1177/016146811011200506

Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish's achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic-self-concept. *Journal of Educational Psychology, 109*(2), 191–207. https://doi.org/10.1037/edu0000135

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: findings of two second-order meta-analyses. *Review of Educational Research, 86*, 849-899. https://doi.org/10.3102/0034654316675417

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychol Methods, 15*(3), 250–267. https://doi.org/10.1037/a0018719

Steiner, P. M., Kim, Y., Hall, C. E., & Su, D. (2017). Graphical models for quasi-experimental designs. *Sociological Methods & Research, 46*(2), 155–188. https://doi.org/10.1177/0049124115582272

Televantou, I., Marsh, H. W., Dicke, T., & Nicolaides, C. (2021). Phantom and big-fish-little-pond-effects on academic self-concept and academic achievement: Evidence from English early primary schools. *Learning and Instruction, 71*, 1–10. https://doi.org/10.1016/j.learninstruc.2020.101399

Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J., & Malmberg, L.-E. (2015). Phantom effects in school composition research: consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement, 26*(1), 75–101. https://doi.org/10.1080/09243453.2013.871302

Tenenbaum, H. R., Winstone, N. E., Leman, P. J., & Avery, R. E. (2020). How effective is peer interaction in facilitating learning? A meta-analysis. *Journal of Educational Psychology, 112*(7), 1303–1319. https://doi.org/10.1037/edu0000436

Traini, C., Kleinert, C., & Bittmann, F. (2021). How does exposure to a different school track influence learning progress? Explaining scissor effects by track in Germany. *Research in Social Stratification and Mobility, 76*, December, 100625. https://doi.org/10.1016/j.rssm.2021.100625

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788

van de Werfhorst, H. G., & Mijs, J. J. B. (2010). Achievement inequality and the institutional structure of educational systems: a comparative perspective. *Annual Review of Sociology, 36*, 407–428. https://doi.org/10.1146/annurev.soc.012809.102538

van Ewijk, R., & Sleegers, P. (2010a). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*, 134–150. https://doi.org/10.1016/j.edurev.2010.02.001

van Ewijk, R., & Sleegers, P. (2010b). Peer ethnicity and achievement: a meta-analysis into the compositional effect. *School Effectiveness and School Improvement, 21*(3), 237–265. https://doi.org/10.1080/09243451003612671

VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology, 34*, 211–219. https://doi.org/10.1007/s10654-019-00494-6

von Keyserlingk, L., Becker, M., & Jansen, M. (2019). Academic self-concept during the transition to upper secondary school. *Contemporary Educational Psychology, 56*, 152–160. https://doi.org/10.1016/j.cedpsych.2019.01.001

von Keyserlingk, L., Becker, M., Jansen, M., & Maaz, K. (2020). Leaving the pond – Choosing the ocean. Effects of student composition on STEM major choices at university. *Journal of Educational Psychology, 112*(4), 751–764. https://doi.org/10.1037/edu0000378

Westphal, A., Becker, M., Vock, M., Maaz, K., Neumann, M., & McElvany, N. (2016). The link between teacher-assigned grades and classroom socioeconomic composition: The role of classroom behavior, motivation, and teacher characteristics. *Contemporary Educational Psychology, 46*, 218–227. https://doi.org/10.1016/j.cedpsych.2016.06.004

Bedingungen der Lesekompetenz von Grundschulkindern – Zentrale Ergebnisse des 15-Jahres-Trends der IGLU-Studie zu Lesebedingungen in Deutschland im internationalen Vergleich

Ramona Lorenz, Ulrich Ludewig, Rahim Schaufelberger, Ruben Kleinkorres & Nele McElvany

Zusammenfassung

Internationale groß angelegte Schulleistungsstudien bieten neben dem Vergleich der Kompetenzen von Schülerinnen und Schülern auch die Möglichkeit, Lernbedingungen zu analysieren und zu vergleichen. Mit dem vorliegenden 15-Jahres-Trend der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) werden die Zusammensetzung der Schülerschaft, häusliche Ressourcen, die Computernutzung in der Schule, Einstellungen zur Schule und die Einstellung zum Lesen als zentrale Lernbedingungen von Viertklässlerinnen und Viertklässlern in Deutschland, Frankreich, den Niederlanden, Schweden und den USA verglichen. Der internationale Vergleich zeigt, dass Deutschland sowohl hinsichtlich digitalisierungsbezogener Aspekte als auch hinsichtlich der Einstellungen zur Schule hinter den anderen Staaten zurückliegt. Die Ergebnisse liefern Hinweise für zukünftige Bemühungen einer Verbesserung der Bedingungen des Lesenlernens in Deutschland.

Schlagworte: Bedingungen des Lesenlernens im Trend und im internationalen Vergleich; Ressourcen häuslicher Lernumgebungen; Unterrichts- und individuelle Merkmale; Einstellungen zur Schule und zum Lesen

Abstract

International large-scale assessments of student achievement not only provide the possibility of comparing student achievement, but also provide the possibility to analyze students learning conditions. Using the available 15-year trend data from the *Progress in International Reading Literacy Study* (PIRLS), important aspects of fourth graders‘ learning conditions regarding student composition, home environmental resources, computer usage at school, attitudes towards school and attitudes towards reading in Germany, France, the Netherlands, Sweden and the USA are compared. The international comparison shows that Germany is lagging behind the other states regarding the aspects of digitalization as well as attitudes towards school. The results provide clues for future efforts on enhancing the learning conditions of reading in Germany.

Keywords: conditions of learning to read in trend and in international comparison; home environmental resources; Instructional and individual characteristics; attitudes towards school and reading

1. Einleitung – 15 Jahre IGLU

Die *Internationale Grundschul-Lese-Untersuchung* (IGLU; international bekannt als *Progress in International Reading Literacy Study* [PIRLS]) wurde im Jahr 2001 zum ersten Mal und seitdem alle fünf Jahre durchgeführt (Bos et al., 2003; Bos et al., 2007; Bos et al., 2012; Hußmann et al., 2017). Im Fokus von IGLU steht die Erfassung der Lesekompetenz von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe im Trend und im internationalen Vergleich. Damit kommt der Studie in besonderem Maße Relevanz zu, da sie die für den weiteren Schulverlauf grundlegende Lesekompetenz an einer wichtigen Gelenkstelle des Schulsystems erfasst und eine solide empirische Grundlage bereitstellt, um die Bedingungen des Lesenlernens in der Grundschule zu beschreiben.

Wie in großangelegten Schulleistungsuntersuchungen mit Fokus auf domänenspezifischen Kompetenzen üblich, liegt IGLU ein umfassend ausgearbeitetes Verständnis von Lesekompetenz zugrunde, das an den aktuellen Stand der Forschung zur Lesekompetenz anknüpft und auf dem anglo-amerikanischen Literacy-Konzept beruht. Der Erwerb von Lesekompetenz wird als Voraussetzung für erfolgreiches Lernen in weiteren Schulfächern sowie für den gesamten weiteren Bildungs- und Lebensweg von Schülerinnen und Schülern betrachtet (Savolainen et al., 2008). In IGLU wird unter

Lesekompetenz die Fähigkeit verstanden, Texte verstehen und nutzen zu können, die gesellschaftlichen und/oder individuell von Bedeutung sind. Bei der Erfassung der Lesekompetenz mittels Lesetests werden zum einen vier Leseverstehensprozesse und zum anderen zwei Leseintentionen berücksichtigt (Bremerich-Vos et al., 2017; Mullis & Martin, 2019). Da der Erwerb und die Entwicklung von Lesekompetenz von Grundschulkindern in vielfältige Bedingungen eingebettet sind, werden in IGLU mittels Fragebögen für die Schülerinnen und Schüler selbst sowie für deren Kontextpersonen (Eltern, Lehrkräfte und Schulleitungen) umfassende Informationen erhoben. Gemäß der theoretischen Rahmenkonzeption von IGLU finden Erwerb und Entwicklung von Lesekompetenz vor dem Hintergrund gesellschaftlicher Ausgangsbedingungen sowohl im schulischen als auch im außerschulischen Kontext unter familiären und individuellen Bedingungen statt. Entsprechend sind die Inhalte der Befragungen breit gefächert und umfassen neben spezifisch lesebezogenen Aspekten wie beispielsweise dem Leseverhalten auch weitere Aspekte der schulischen und außerschulischen Lernumwelt oder demographische Aspekte. Hier knüpft der vorliegende Beitrag an, der das Ziel verfolgt, zentrale Bedingungen des Erwerbs der Lesekompetenz im Trend und im Vergleich zwischen ausgewählten Bildungssystemen zu vergleichen.

Mit den Daten der IGLU-Zyklen 2001 bis 2016 liegt bereits die Möglichkeit vor, Entwicklungen der Lernbedingungen von Schülerinnen und Schülern am Ende der Grundschulzeit über einen Zeitraum von 15 Jahren zu betrachten. Wenngleich IGLU als Trendstudie mit wiederholten Querschnittserhebungen angelegt ist, unterliegen die eingesetzten Erhebungsinstrumente einschließlich der kontinuierlichen Anpassungen, nicht zuletzt auch zur angemessenen Berücksichtigung von Veränderungen in schulischen und gesellschaftlichen Kontexten.

Zusätzlich zu inhaltlichen und methodischen Anpassungen, die seit IGLU 2001 umgesetzt wurden, ist zu beachten, dass die Liste an Teilnehmerstaaten und -regionen der Studie nicht über alle IGLU-Erhebungszyklen hinweg identisch sind. Internationale Vergleiche aus der Perspektive eines Trends über 15 Jahre sind demnach nur mit Fokus auf Teilnehmerstaaten und -regionen sowie zu Untersuchungsgegenständen möglich, die über alle Erhebungszyklen konstant waren. Vor diesem Hintergrund wurden im vorliegenden Beitrag Bedingungsfaktoren des Lesenlernens berücksichtigt, die der theoretischen Rahmenkonzeption von IGLU folgend gesellschaftliche, schulische, familiäre und individuelle Faktoren berücksichtigen, die mit der Leseleistung am Ende der Grundschulzeit verbunden sind. Leitend ist dabei

die Frage nach Veränderungen dieser Bedingungen über die Zeit hinweg und nach Unterschieden im Trend im Vergleich ausgewählter Bildungssysteme.

2. Zentrale Bedingungsfaktoren der Lesekompetenz

Eine Systematisierung von Bedingungsfaktoren, die im Zusammenhang mit der Lesekompetenz stehen, kann in IGLU in unterschiedliche Facetten des gesellschaftlichen, schulischen und familiären sowie individuellen Kontextes unterteilt werden (Ditton, 2016; Hußmann et al., 2017; McElvany, Becker & Lüdtke, 2009). In diesen Bereichen sind relevante Bedingungen zu betrachten, die die Entwicklung der Lesekompetenz von Grundschulkindern beeinflussen können, was auf gesellschaftlicher Ebene beispielsweise Veränderungen in der Schülerschaft oder die fortschreitende Digitalisierung aller Lebensbereiche umfasst. Letztere hat wiederum auch in Schulen und im Unterricht zu Veränderungen beigetragen, die die unterrichtlichen Lernbedingungen wandeln. Zu den familiären sowie individuellen Faktoren zählen neben strukturellen Merkmalen wie z. B. einem möglichen familiären Migrationshintergrund oder dem soziokulturellen Kapital der Familie auch Ressourcen der außerschulischen Lernumwelt oder auch individuelle Einstellungen zur Schule im Allgemeinen oder zum Thema Lesen.

2.1. Zusammensetzung der Schülerschaft

Familiäre Bedingungen gelten als zentrale Determinanten der Schulleistung (Helmke & Weinert, 1997) und es liegen verschiedene Theorien zur Erklärung dieses Verhältnisses vor (z. B. primäre und sekundäre Herkunftseffekte nach Boudon, 1974; Kapitalsorten nach Bourdieu, 1983). Lesebezogene Modelle konnten die Bedeutung der familiären Lesesozialisation für die Kompetenzentwicklung in der Domäne des Lesens herausstellen (Sénéchal & LeFevre, 2002). Es ist anzunehmen, dass diese Unterschiede in der familiären Lesesozialisation einen Beitrag zur Erklärung des international belegten, systematischen Zusammenhangs zwischen der Lesekompetenz und familiären Hintergrundmerkmalen wie dem soziokulturellen Kapital und einem möglichen Migrationshintergrund leisten können (Mullis et al., 2017).

Hinsichtlich des soziokulturellen Kapitals konnten Untersuchungen zeigen, dass Eltern mit höheren Bildungsabschlüssen aufgrund eigener Erfahrungen mit dem Bildungssystem erfolgreicher darin sind, ihre Kinder zu unterstützen, indem sie eine lesesozialisationsförderliche Lernumwelt bieten

(Dong et al, 2020). Die Anzahl der im Haushalt vorhandenen Bücher hat sich in der empirischen Bildungsforschung als ein Indikator für das soziokulturelle Kapital bewährt, der zugleich einfach zu erheben und aussagekräftig ist (Schwippert, 2019). Hinsichtlich eines internationalen Vergleichs im Trend über 15 Jahre bietet sich die Anzahl der im Haushalt vorhandenen Bücher als geeigneter Indikator des soziokulturellen Kapitals an, da diese Information über alle IGLU-Zyklen hinweg in allen Teilnehmerstaaten und -regionen unverändert erfasst wurde.

Neben dem soziokulturellen Kapital der Eltern hängt auch ihr Migrationshintergrund mit der Lesekompetenz von Schülerinnen und Schülern zusammen (Mullis, Martin, Foy & Hooper, 2017). Eltern mit Migrationshintergrund messen der Bildung ihrer Kinder häufig einen hohen Wert bei (Kao & Tienda, 1995). Gleichzeitig verfügen sie aber häufig über weniger soziokulturelles Kapital und Erfahrungen mit dem Bildungssystem des Landes, in dem ihre Kinder die Schule besuchen, und sprechen darüber hinaus häufig mit ihren Kindern eine andere Sprache als die Unterrichtssprache (Kristen & Dollmann, 2012). Die zu Hause gesprochene Sprache ist daher ein weit verbreiteter Indikator für einen Migrationshintergrund, der auch in allen bisherigen IGLU-Zyklen international eingesetzt wurde.

Lesebezogene Leistungsrückstände von Kindern aus Familien mit niedrigerem soziokulturellen Kapital und von Kindern aus Familien mit Migrationshintergrund liegen in den meisten IGLU-Teilnehmerstaaten und -regionen vor (Mullis et al., 2017). In Deutschland ist es seit IGLU 2001 nicht gelungen, die Disparitäten zu reduzieren (Hußmann et al., 2017). Daher ist es von Interesse, die Trends zu diesen relevanten Indikatoren seit IGLU 2001 in Deutschland und im Vergleich in weiteren Staaten zu betrachten.

2.2. Lernbedingungen – Ressourcen häuslicher Lernumwelt

Neben individuellen Merkmalen der Schülerinnen und Schüler sowie deren Familien ist für die Unterstützung von Lernprozessen auch die außerschulische häusliche Ausstattung ein wichtiger Bedingungsfaktor für Lernerträge und die Kompetenzentwicklung (Ditton, 2016; McElvany, Becker & Lüdtke, 2009). Da ein Teil des Lernprozesses beispielsweise im Rahmen der Bearbeitung von Hausaufgaben oder der Vorbereitung auf Tests im häuslichen Umfeld stattfindet, stellt ein Arbeitsplatz, an dem ein Kind ruhig arbeiten kann, eine bedeutsame Ressource dar. So erfasst IGLU im internationalen Vergleich die Angabe der Schülerinnen und Schüler, ob ein Schreibtisch oder anderer Tisch zum Arbeiten für das Kind zu Hause zur Verfügung steht.

Neben dieser Ressource ist in der heutigen schnelllebigen Welt, die auch durch das Vordringen der digitalen Technologien in alle Lebensbereiche gekennzeichnet ist, die Relevanz digitaler Medien in Bildungskontexten gestiegen (Eickelmann et al., 2019; Voogt, Knezek, Christensen & Lai, 2018), was nicht zuletzt auch durch veränderte Lernformen während der Corona-Pandemie beschleunigt wurde (Huber, 2021; Lorenz, Brüggemann, Stang & McElvany, in Druck). Der Einsatz von Computern, Laptops oder Tablets sowie internetbasierten Arbeitsweisen im Unterricht und auch in außerschulischen Settings eröffnet somit neben dem papierbasierten Lesen erweiterte Leseanlässe für Grundschulkinder, die als Bedingungen des Lesenlernens eine wesentliche Rolle einnehmen können und u. a. als motivierend oder unter bestimmten Voraussetzungen der Nutzungsweisen als lernförderlich angesehen werden (Mangen & van der Weel, 2016; Schaumburg, 2018; Zierer, 2020). Die Leseanlässe mit digitalen Medien umfassen Informationen, die schriftsprachlich repräsentiert sind, sodass die Lesekompetenz von Kindern für das Textverstehen, das Lokalisieren relevanter Informationen, das Bewerten von Inhalten oder das Reflektieren der Texte ebenfalls erforderlich ist und dadurch die Weiterentwicklung der Lesekompetenz selbst fördern kann. Für die Beschreibung der Computernutzung sowie den Zugang zu medial vermittelten Informationen als Lernbedingung von Kindern ist eine Internetverfügbarkeit als grundlegender Bedingungsfaktor im Sinne von Ausstattungsmerkmalen zu benennen.

2.3. Lernbedingungen – Computernutzung

Nicht nur die bloße Verfügbarkeit von Computern und Internet als Voraussetzung von Lernprozessen, sondern auch die Nutzung von Computern durch die Kinder selbst ist in den Blick zu nehmen (Lorenz & Bos, 2017). Theoretische Modelle des Leseverstehens betrachten das Zusammenspiel von individuellen Merkmalen, Textinhalt und -gestaltung sowie Leseanweisungen (u. a. McNamara & Magliano, 2009) und empirische Befunde legen Hinweise vor, dass Leseergebnisse durch das Medium beeinflusst werden. Mit ihrer Metaanalyse zeigten Delgado, Vargas, Ackerman und Salmerón (2018), dass mit digitalen Texten im Vergleich zu papierbasierten Lesetexten ein schlechteres Leseverständnis resultierte, was mit der Implikation einer besseren Vorbereitung auf das Lesen in zunehmend digitalen Umgebungen diskutiert wurde. Damit rücken auf der Prozessebene quantitative und qualitative Merkmale der Computernutzung in den Fokus, was sowohl die Nutzungshäufigkeit als quantitatives Merkmal umfasst sowie die Nutzungsweisen und -zwecke

als Indikator der Qualität der Nutzung. Dabei kann unterschieden werden, wo der Einsatz von Computern erfolgt: zu Hause oder in der Schule. Mit Blick auf die Nutzungshäufigkeit von Computern im Unterricht können erste Rückschlüsse auf die Unterrichtsgestaltung bzw. unterrichtlichen Prozesse gezogen werden, sodass die Häufigkeit der Computernutzung in der Schule als eine bedeutsame Lernbedingung angesehen werden kann. Diese kann dazu dienen, die Lernbedingungen der Schülerinnen und Schüler im Kontext der Lesekompetenz zu erfassen.

2.4. Lernbedingungen – Einstellungen zur Schule

Ein weiterer Aspekt, der neben den materiellen Ressourcen, die von der Schule zur Verfügung gestellt werden, eine wichtige Rolle bei Lernprozessen spielt, ist, inwiefern die schulischen Rahmenbedingungen den Bedürfnissen der Schülerinnen und Schüler entsprechen. Ein Indikator für diese Passung, bietet das Wohlbefinden der Schülerinnen und Schüler. Fühlen sie sich in der Schule wohl, kann dies positive Effekte auf deren spätere Leistung haben (Kleinkorres, Stang & McElvany, 2020; Mega, Ronconi & De Beni, 2014). Darüber hinaus wird das schulische Wohlbefinden aber auch als ein inhärentes Ziel von Bildungsarbeit betrachtet (Kanonire, Federiakan & Uglanova, 2020; OECD, 2017). Die Einstellungen der Schülerinnen und Schüler gegenüber dem Kontext Schule stellen einen wichtigen Teilaspekt des schulischen Wohlbefindens dar (Hascher, 2004; Huebner & Gilman, 2006). In IGLU werden die Einstellungen gegenüber der Schule darüber erfasst, wie gerne die Schülerinnen und Schüler zur Schule gehen. Darüber hinaus wird erhoben, wie sicher sie sich in der Schule fühlen.

Die Betrachtung dieser beiden Variablen ermöglicht es, Erkenntnisse über die Passung zwischen den Bedürfnissen der Schülerinnen und Schüler und den schulischen Rahmenbedingungen zu gewinnen bzw. zu untersuchen, ob die Schülerinnen und Schüler sich in einem Umfeld wiederfinden, das ihnen eine angemessene Lernatmosphäre bietet.

2.5. Lernbedingungen – Einstellungen zum Lesen

Wenn es um die Bedingungen geht, die spezifisch für die Lesekompetenz relevant sind, spielen nicht nur globale Bewertungen der Schülerinnen und Schüler in Bezug auf die Institution Schule eine Rolle, sondern auch die motivationalen Überzeugungen und Einstellungen zur Domäne Lesen. Kognitive

und motivationale Faktoren stellen wichtige Voraussetzungen für schulische Leistungen dar (Kriegbaum, Becker & Spinath, 2018). In Bezug auf die Lesekompetenz sind die Lesemotivation und lesebezogene Einstellungen bedeutsame Faktoren, die Auskunft über die individuellen Lernvoraussetzungen der Schülerinnen und Schüler geben. In IGLU wurden diese individuellen Merkmale zwischen 2001 und 2016 anhand zweier Items erhoben. Die Kinder gaben dabei an, wie gerne sie lesen und inwiefern sie Lesen langweilig finden. Die Förderung von Lesemotivation und positiven Einstellungen gegenüber dem Lesen kann darüber hinaus selbst als Kriterium für erfolgreichen Unterricht herangezogen werden.

3. Trends der Bedingungsfaktoren über 15 Jahre im internationalen Vergleich

Die Relevanz der Bedingungsfaktoren des Lesenlernens, die theoretischen Ansätzen folgend in gesellschaftlichen, schulischen, familiären und individuellen Bereichen zu verorten sind, führt zu der Frage, ob im zeitlichen Verlauf Veränderungen in diesen Bereichen festgestellt werden können. Für den vorliegenden Beitrag ist es im Trend über 15 Jahre aufgrund des aufgezeigten engen Zusammenhangs dieser Bedingungsfaktoren mit der Lesekompetenz interessant zu sehen, in welchen Bereichen Veränderungen der Lernbedingungen deutlich werden. Zudem kann es für die Beschreibung der Lernbedingungen in Deutschland aufschlussreich sein, den Trend dieser Bedingungsfaktoren in anderen Bildungssystemen vergleichend heranzuziehen. Die Datengrundlage von IGLU bietet diese Möglichkeit der Analyse von Trends im internationalen Vergleich mit ihren bisher vorliegenden Zyklen.

Datengrundlage

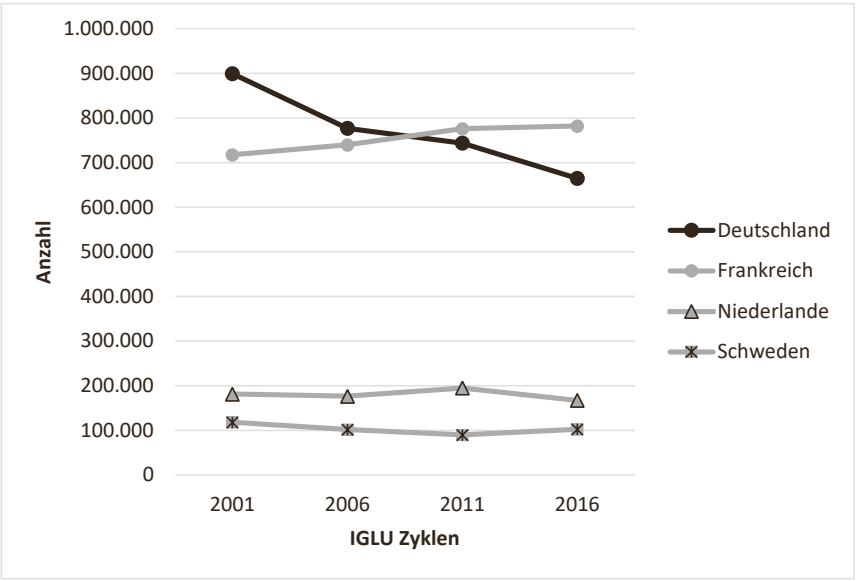
Als Datengrundlange für die Trendanalysen wurden die Daten aus IGLU 2001 bis 2016 verwendet. Die vorliegende Arbeit befasst sich mit Angaben aus den Teilnehmerstaaten Deutschland, Frankreich, den Niederlanden, Schweden und den USA.

Frankreich, die Niederlande und Schweden wurden als Vergleichsstaaten herangezogen, weil es sich um europäische Staaten handelt, die im Hinblick auf Industrialisierungsgrad und Bevölkerungsstruktur viele Ähnlichkeiten zu Deutschland aufweisen. Die mittlere Leseleistung in IGLU 2016 lag in den Niederlanden und Schweden signifikant über dem Mittelwert für Deutschland. Vergleichend wird Frankreich mit einer signifikant geringeren Leseleis-

tung herangezogen. Die USA wurden als zusätzliche Referenz genutzt, da der Kontrast zu den europäischen Bildungssystemen aufschlussreich sein könnte. Die durchschnittliche Lesekompetenz der Viertklässlerinnen und Viertklässler in den USA war in IGLU 2016 signifikant höher im Vergleich zu Deutschland. Die fünf Staaten haben an allen Zyklen teilgenommen. Die Analysen beziehen sich auf Fragen, die zumeist in allen Zyklen abgefragt wurden. Alle Fragen wurden bezüglich ihrer exakten Formulierung geprüft und sind in hinreichender Weise zwischen den Erhebungszyklen vergleichbar.

Abbildung 1 stellt zunächst die Anzahl der Schülerinnen und Schüler in der vierten Klasse für die ausgewählten Teilnehmerstaaten dar und zeigt mit Blick auf Deutschland ein sinkendes Niveau im zeitlichen Verlauf, während für die Niederlande und Schweden ein gleichbleibendes Niveau und für Frankreich leicht zunehmende Schüleranzahlen im Verlauf von 2001 bis 2016 deutlich werden.

Abbildung 1: Schülerzahlen in der vierten Klasse im Trend von 2001 bis 2016



Anmerkungen. Berechnungen der IGLU-Studie basierend auf administrativen Daten (Martin et al., 2017). Der Trend für die USA ist nicht abgebildet, weil die Schülerpopulation um ein vielfaches größer ist. Dort zeigt sich ein steigender Trend.

Das Stichprobenverfahren in allen IGLU-Zyklen folgte einem zweistufigen geschichteten Clusterdesign (d. h. zunächst Schulen und dann Klassen innerhalb der Schulen; Martin et al., 2017). Die Schulen, die an der Studie teilnah-

men, wurden aus einer vollständigen Liste der Grundschulen eines Landes ausgewählt. Dies geschah unter Berücksichtigung von landesspezifischen Strata (z. B. Länder der Bundesrepublik Deutschland). In den Schulen wurde eine Klasse ausgewählt und alle Schülerinnen und Schüler dieser Klasse nahmen an der Erhebung teil. Die Stichproben für die Zyklen und ausgewählten Staaten sind in Tabelle 1 dargestellt. Bei Verwendung der Stichprobengewichte können Ergebnisse basierend auf diesen Stichproben als repräsentativ für die Viertklässlerinnen und Viertklässler der Staaten interpretiert werden. Für die Berechnung aller Werte wurden die für IGLU üblichen Analyseverfahren zur Berechnung von Standardfehlern, Gewichtungsfaktoren sowie Sampling- und Imputationsvarianz berücksichtigt (Martin et al., 2017).

Tabelle 1: Stichprobenbeschreibung

Staat	Zyklus	*N*	Mis (%)	Alter	Weiblich (%)
Deutschland	2001	7633	4.40	10.54	49.30
	2006	7899	19.10	10.46	49.15
	2011	4000	10.10	10.37	50.36
	2016	3959	12.12	10.34	49.89
Frankreich	2001	3538	3.00	10.12	47.67
	2006	4404	8.81	10.01	48.27
	2011	4438	0.70	9.96	49.29
	2016	4767	1.45	9.80	50.47
Niederlande	2001	4112	1.53	10.28	50.06
	2006	4156	1.78	10.24	50.37
	2011	3995	0.80	10.16	50.93
	2016	4206	1.26	10.04	50.53
Schweden	2001	6044	2.38	10.79	49.24
	2006	4394	2.89	10.85	48.39
	2011	4622	2.42	10.74	48.26
	2016	4525	2.45	10.73	49.97
USA	2001	3763	1.25	10.19	50.79
	2006	5190	7.17	10.08	49.78
	2011	12726	1.89	10.22	50.43
	2016	4425	2.69	10.11	50.09

Anmerkungen. *N* = 102796. Mis = Anteil fehlender Angaben im Schülerfragebogen.
Alter = Alter zum Erhebungszeitpunkt.

Fehlende Angaben

In dieser Arbeit wurden Angaben aus den Schülerbefragungen in den Blick genommen. Der Anteil fehlender Angaben variierte zwischen Staaten und Zyklen. Diese Unterschiede zwischen Staaten und Zyklen lassen sich größtenteils auf unterschiedliche Verpflichtungsgrade zur Teilnahme an der Hintergrundbefragung zurückführen. Beispielsweise war die Beantwortung der Fragen in Deutschland 2001 verpflichtend für Schülerinnen und Schüler und erforderte kein Einverständnis der Eltern. In den darauffolgenden Zyklen war

für die Teilnahme an der Hintergrundbefragung eine gültige Einverständniserklärung der Eltern nötig.

Um den Einfluss von fehlenden Angaben auf die Ergebnisse zu reduzieren, wurde eine mehrfache Imputation vorgenommen. Diese erfolgte unter Einbezug aller im Kapitel verwendeter Daten zuzüglich des Zyklus (Faktor), des Staates (Faktor) und der Lesekompetenzdaten, die für alle Schülerinnen und Schüler vorhanden waren. Für die Imputation wurde eine Mehrebenen-Imputation mit prädiktivem Mittelwert-Matching auf Stufe 1 für kontinuierliche Variablen und eine logistische Regression für originär binäre Variablen (z. B. Migrationshintergrund) innerhalb des R-Paketes miceadds (Robitzsch, Grund & Henke, 2022) mit 20 Iterationen und fünf imputierten Datensätzen verwendet. Für die Imputation wurden alle Variablen in ihrer ursprünglichen Metrik verwendet und erst nach der Imputation umgeformt (z. B. dichotomisiert).

Analyse

Die berichteten Mittelwerte und prozentualen Anteile, sowie deren Standardfehler wurden mit der für IGLU üblichen Jackknife-Replications-Methode, Stichprobengewichten und gepoolten Ergebnissen aus den mehrfach imputierten Daten bestimmt. Die geschätzten Standardfehler berücksichtigen die geklumpte Stichproben- und Imputationsvarianz sowie Gewichtung. Zur inferenzstatistischen Absicherung wird zu jedem Schätzwert ein Konfidenzintervall angegeben, dass zwei Standardfehler über und unter dem Schätzwert entspricht. Dies approximiert ein 95 %-iges Konfidenzintervall. Schätzwerte in einem paarweisen Vergleich mit einander nicht überlappenden Konfidenzintervallen können als statistisch signifikant unterschiedlich mit $\alpha < 1$ % gewertet werden. Alle Analyse wurden mit dem Paket BIFIE.survey (BIFIE, Robitzsch & Oberwimmer, 2022) durchgeführt.

3.1. Trends der Zusammensetzung der Schülerschaft

Vor dem Hintergrund des empirisch vielfach aufgezeigten Zusammenhangs des soziokulturellen Kapitals sowie des Migrationshintergrunds mit Leistungsergebnissen über Schulsysteme hinweg, wurden zunächst Trends der Zusammensetzung der Schülerschaft über die Erhebungszeitpunkte hinweg betrachtet.

Die Anzahl der Bücher im Haushalt als Indikator des soziokulturellen Kapitals wurde dichotomisiert mit den Gruppen „maximal 100 Bücher“ und

„mehr als 100 Bücher“ im Haushalt ausgewertet. Abbildung 2 zeigt, dass der Anteil der Schülerinnen und Schüler, die angaben, mehr als 100 Bücher im Haushalt zu besitzen, in Deutschland und den Niederlanden über die Erhebungszyklen hinweg stabil blieb. In Frankreich war nach einem Anstieg zwischen 2001 und 2006 eine Abnahme zu 2011 zu erkennen. Eine nahezu parallel abnehmende Tendenz ist außerdem für Schweden und die USA zu erkennen. In Schweden ist der Anteil der Schülerinnen und Schüler mit mehr als 100 Büchern im Haushalt von einem im Vergleich mit den anderen Teilnehmerstaaten hohen Niveau etwa auf das Niveau der anderen Vergleichsstaaten gesunken.

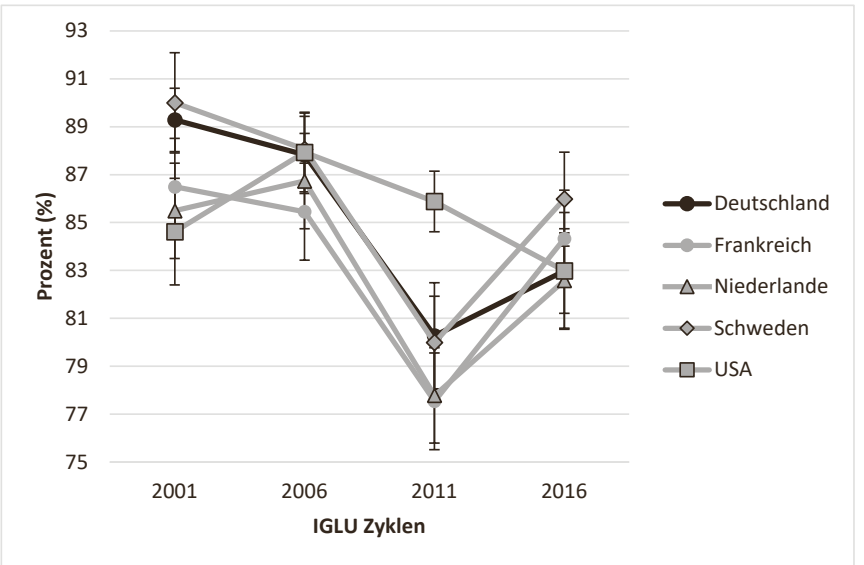
Abbildung 2: Anteile der Schülerinnen und Schüler, die angaben, mehr als 100 Bücher im Haushalt zu haben (Angaben der Schülerinnen und Schüler, Anteile in Prozent)

Anmerkungen. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Die Frage lautete: „Wie viele Bücher gibt es bei dir zu Hause ungefähr?“ Die Punkte in der Abbildung geben an, wie viel Prozent der Kinder darauf antworteten „Genug, um zwei Regale zu füllen (101–200 Bücher)“ oder „Genug, um drei oder mehr Regale zu füllen (über 200 Bücher)“.

Als zweiter Indikator der Schülerzusammensetzung wurden Trends hinsichtlich des Migrationshintergrunds anhand der zu Hause gesprochenen Sprache in den Blick genommen. Für alle betrachteten Staaten sind in Abbildung 3 Veränderungen im Trend erkenntlich. Diese Veränderungen sind jedoch nicht linear. Besonders hervorzuheben sind die in Deutschland, Frankreich, den Niederlanden und Schweden von 2006 zu 2011 gesunkenen Anteile der Schülerinnen und Schüler, die zu Hause immer oder fast immer die Unterrichtssprache sprechen. Für diese Staaten ist zugleich ein Anstieg dieses Anteils

zwischen 2011 und 2016 erkennbar. Für die USA zeigte sich eine Veränderung in einem Anstieg des Anteils der zu Hause immer oder fast immer die Unterrichtssprache sprechenden Schülerinnen und Schüler zwischen 2001 und 2006. Seither zeigt sich über die Erhebungszyklen hinweg ein sinkender Verlauf.

Abbildung 3: Anteile der Schülerinnen und Schüler, die angaben, zu Hause immer oder fast immer die Unterrichtssprache zu sprechen (Angaben der Schülerinnen und Schüler, Angaben in Prozent)



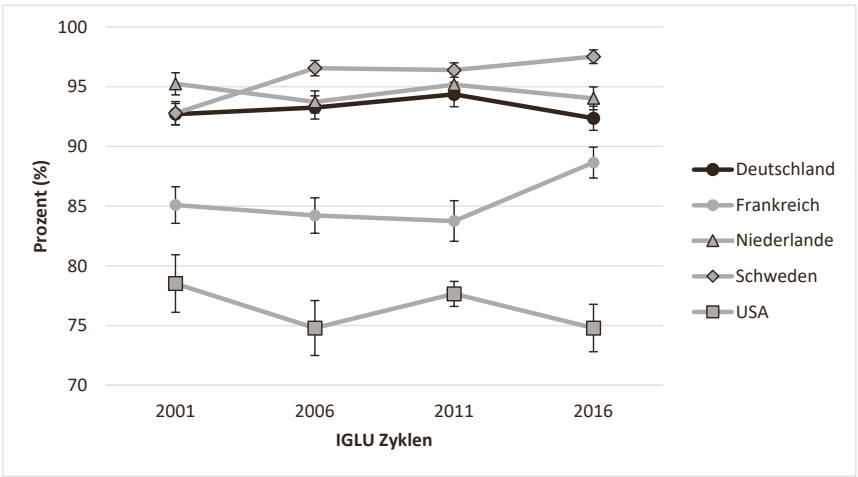
Anmerkungen. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Da sich das Antwortformat zwischen den Zyklen leicht geändert hat, geben in der Abbildung Werte für 2016 an, wie viel Prozent der Kinder auf die Frage „Wie oft sprichst du zu Hause [Unterrichtssprache]?“ mit „Ich spreche zu Hause immer [Unterrichtssprache]“ oder „Ich spreche zu Hause fast immer [Unterrichtssprache]“ antworteten. Für 2001 und 2011 geben die Werte an, wieviel Prozent der Kinder auf diese Frage mit „Ich spreche zu Hause immer oder fast immer [Unterrichtssprache]“ antworteten. Angaben für 2006 sind extrapoliert, da keine vergleichbaren Angaben erhoben wurden.

Mit Blick auf Deutschland lässt sich hinsichtlich der betrachteten Indikatoren der Schülerzusammensetzung herausstellen, dass der soziokulturelle Hintergrund entgegen des sinkenden Trends in den Vergleichsstaaten in Deutschland stabil blieb und der sprachliche Hintergrund der Schülerfamilien in den europäischen Staaten ähnliche Entwicklungen nahm.

3.2. Trends der Bedingungsfaktoren aus Ressourcen häuslicher Lernumwelt

Als ein Indikator für die Bedingungsfaktoren im Bereich der häuslichen Lernumwelt von Viertklässlerinnen und Viertklässlern wurde international vergleichend erfasst, inwiefern den Schülerinnen und Schülern zu Hause ein eigener Schreibtisch bzw. Tisch zur Verfügung steht. In Abbildung 4 sind die Anteile der Schülerinnen und Schüler dargestellt, die angaben, zu Hause einen Schreibtisch oder anderen Tisch nur für sich zu haben. Der Blick auf die Ergebnisse im Trend zeigt für die betrachteten Staaten geringfügige Schwankungen über die Zeit hinweg. Diese Ressource der häuslichen Lernumgebung blieb innerhalb der Bildungssysteme folglich relativ stabil. Der Vergleich zwischen den Staaten zeigt jedoch Niveauunterschiede dahingehend, dass in Deutschland, Schweden und den Niederlanden zu allen Erhebungszeitpunkten höhere Anteile von Schülerinnen und Schülern angaben, zu Hause über einen eigenen Schreibtisch oder Tisch zu verfügen, als in Frankreich und den USA.

Abbildung 4: Anteile der Schülerinnen und Schüler, die angaben, zu Hause einen eigenen Schreibtisch oder anderen Tisch zu haben (Angaben der Schülerinnen und Schüler, Anteile in Prozent)

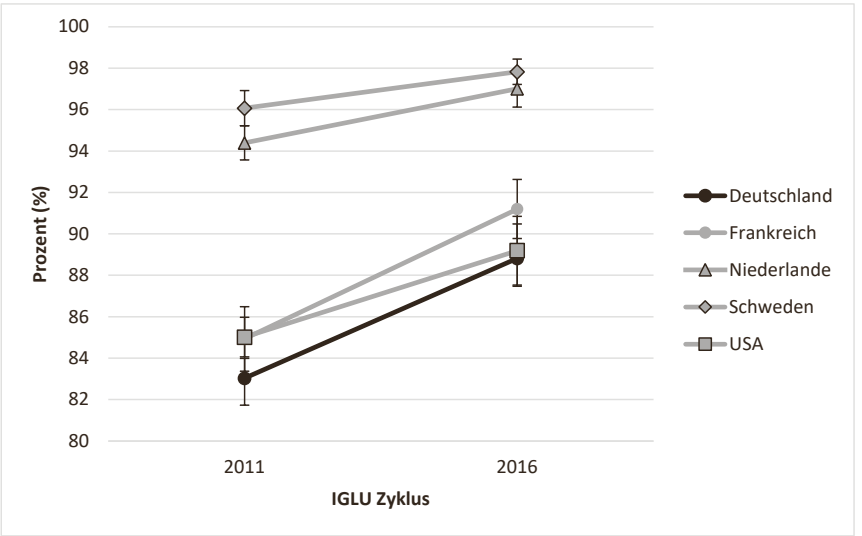


Anmerkungen. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Die Frage lautete: „Gibt es diese Dinge bei dir zu Hause?" Die Punkte geben an, wie viel Prozent der Kinder auf das Item „Einen Schreibtisch oder anderen Tisch nur für dich" mit „ja" antworteten.

Die Internetverfügbarkeit zu Hause als Ressource wurde bei den Viertklässlerinnen und Viertklässlern ab IGLU 2011 erfasst und zeigt insgesamt steigende Tendenzen in allen betrachteten IGLU-Teilnehmerstaaten (Abbildung

5). Insbesondere in den Niederlanden und in Schweden gaben zyklusübergreifend hohe Anteile der Schülerinnen und Schüler an, zu Hause Zugriff auf das Internet zu haben. In Deutschland, Frankreich und den USA sind die Anteile über die Zeit hinweg ebenfalls auf ein hohes Niveau von über 90 Prozent angestiegen, konnten jedoch noch nicht zu den Niederlanden und Schweden aufschließen.

Abbildung 5: Anteile der Schülerinnen und Schüler, die angaben, zu Hause über einen Internet-Zugang zu verfügen (Angaben der Schülerinnen und Schüler, Anteile in Prozent)



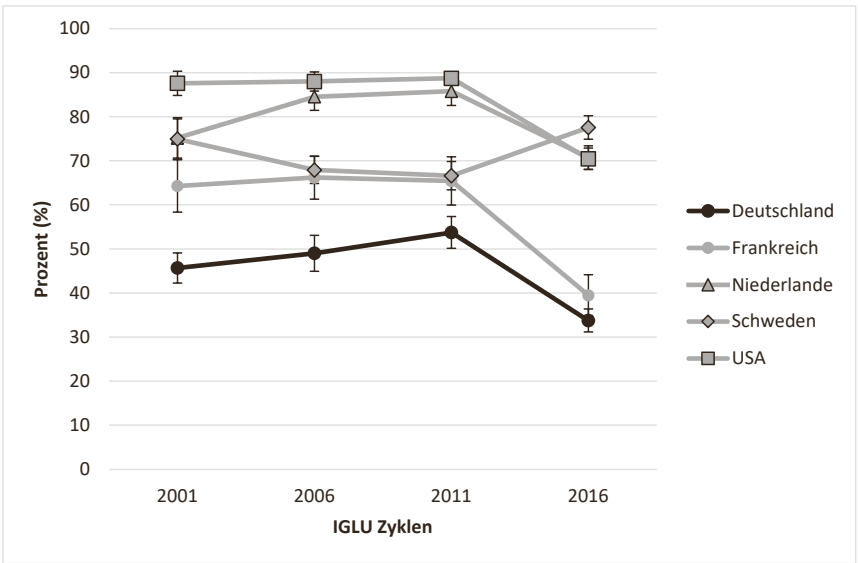
Anmerkungen. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Die Frage lautete: „Gibt es diese Dinge bei dir zu Hause?“ Die Punkte geben an, wie viel Prozent der Kinder auf das Item „Einen Internet-Zugang“ mit „ja“ antworteten.

3.3. Trends des Bedingungsfaktors Computernutzung in der Schule

Hinsichtlich der Computernutzung wurde im Trend die Angabe der Schülerinnen und Schüler erhoben, wie häufig sie in der Schule Computer nutzen. Dazu zeigt Abbildung 6, dass die Nutzungshäufigkeit in Deutschland nahezu durchgängig unter den Werten der anderen betrachteten IGLU-Teilnehmerstaaten lag. Stieg der Anteil der Schülerinnen und Schüler, die angaben, Computer mindestens einmal pro Woche im Unterricht zu nutzen in Deutschland zwischen den Jahren 2001 und 2011 noch an, fiel der Anteil 2016 unter den Wert im ersten Erhebungszyklus zurück. Einen ähnlichen Rückgang verzeichneten auch Frankreich, die Niederlande und die USA. Für Schweden

zeigte sich hingegen von 2011 zu 2016 ein Anstieg und damit ein höherer Anteil von Schülerinnen und Schülern, die angaben, mindestens einmal pro Woche Computer in der Schule zu nutzen. Insgesamt wird im Trend für die Niederlande, Schweden und die USA eine stärkere Implementation von Computern im Unterricht der vierten Klassen im Vergleich zu den weiteren betrachteten Teilnehmerstaaten deutlich.

Abbildung 6: Anteile der Schülerinnen und Schüler, die angaben, in der Schule mindestens einmal pro Woche mit Computern zu arbeiten (Angaben der Schülerinnen und Schüler, Anteile in Prozent)



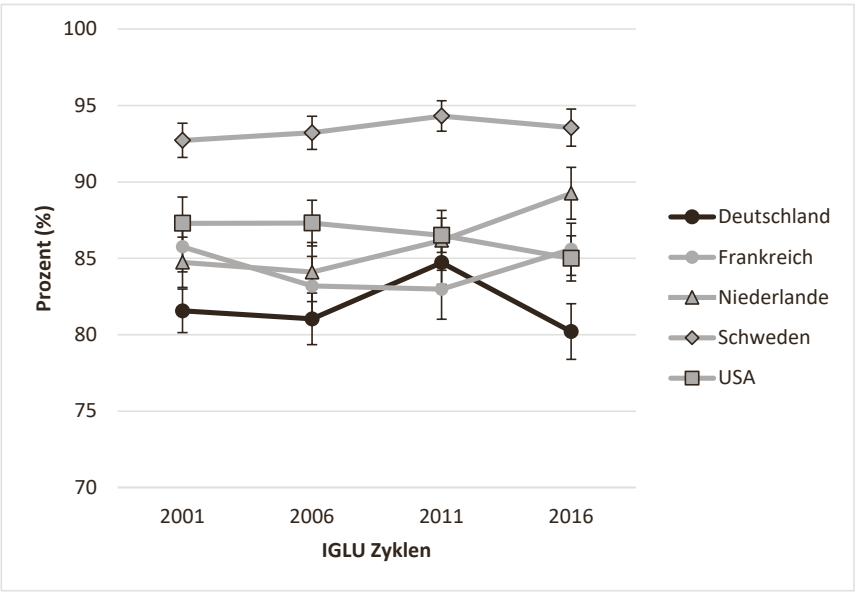
*Anmerkung*en. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Die Frage lautete: „Wie oft benutzt du einen Computer oder ein Tablet an jedem der folgenden Orte für Schularbeiten?". Die Punkte geben an, wie viel Prozent der Kinder auf das Item „In der Schule" mit „Jeden Tag oder fast jeden Tag" oder „Ein bis zwei Mal pro Woche" antworteten.

3.4. Trends der Bedingungsfaktoren der Einstellungen zur Schule

Mit den Einstellungen zur Schule und dem Wohlbefinden der Schülerinnen und Schüler wurde ein weiterer relevanter Bereich von Faktoren im Trend betrachtet, der eng verbunden mit Leistungen von Schülerinnen und Schülern ist. Abbildung 7 zeigt, wie hoch die Anteile der Schülerinnen und Schüler waren, die angaben, sich in der Schule sicher zu fühlen. Auffällig ist dabei, dass die Schülerinnen und Schüler in Schweden im Vergleich zu den weiteren betrachteten Teilnehmerstaaten zu allen Zyklen zu höheren Anteilen ein positives Sicherheitsempfinden in der Schule angaben. Für die Niederlande

lässt sich zudem im zeitlichen Verlauf ein tendenziell steigendes Niveau feststellen. Mit Blick auf Deutschland zeigte sich in den ersten beiden Zyklen ein Anteil von etwas mehr als 80 Prozent der Schülerinnen und Schüler, die angaben, sich in der Schule sicher zu fühlen. Dieser Anteil schwankte in der Folge leicht und lag 2011 etwas höher, fiel 2016 jedoch wieder in etwa auf den Ausgangswert zurück, womit für Deutschland im Vergleich mit den weiteren Staaten geringere Anteile zu verzeichnen sind.

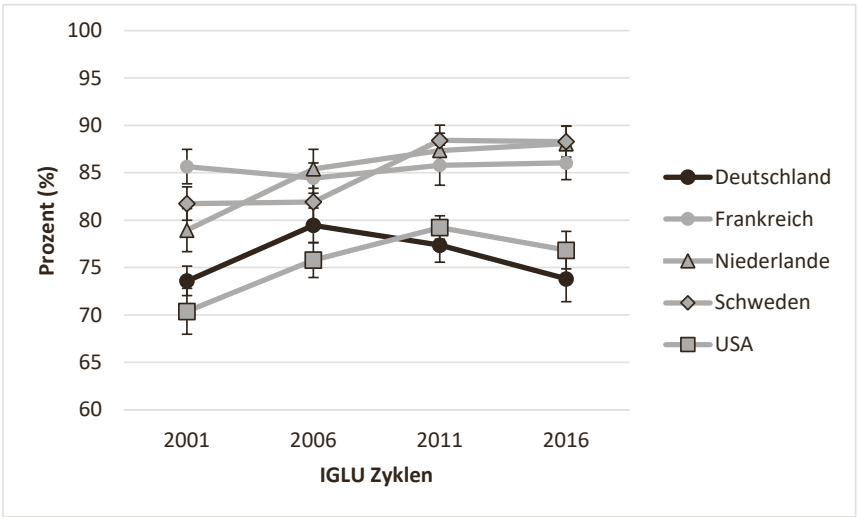
Abbildung 7: Anteile der Schülerinnen und Schüler, die angaben, sich in ihrer Schule sicher zu fühlen (Angaben der Schülerinnen und Schüler, Anteile in Prozent)



Anmerkungen. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Die Frage lautete: „Wie denkst du über deine Schule?". Die Punkte geben an, wie viel Prozent der Kinder auf das Item „Ich fühle mich sicher, wenn ich in der Schule bin" mit „stimme stark zu" oder „stimme einigermaßen zu" antworteten.

Zusätzlich zum schulischen Sicherheitsempfinden konnte im Trend betrachtet werden, wie hoch der Anteil der Viertklässlerinnen und Viertklässler war, die angaben, gerne zur Schule zu gehen (Abbildung 8). Ähnlich wie bei dem vorherigen Bedingungsfaktor zeigte sich für Schweden im Trend ein im Vergleich der Staaten eher hoher Anteil an Schülerinnen und Schülern, die angaben, gerne zur Schule zu gehen, mit tendenziell steigendem Anteil. Frankreich und die Niederlande lagen auf einem vergleichbaren Niveau. Für Deutschland hingegen wurde ein im Vergleich eher geringer Wert mit sinkender Tendenz seit dem Zyklus 2006 erkennbar.

Abbildung 8: Anteile der Schülerinnen und Schüler, die angaben, gerne zur Schule zu gehen (Angaben der Schülerinnen und Schüler, Anteile in Prozent)



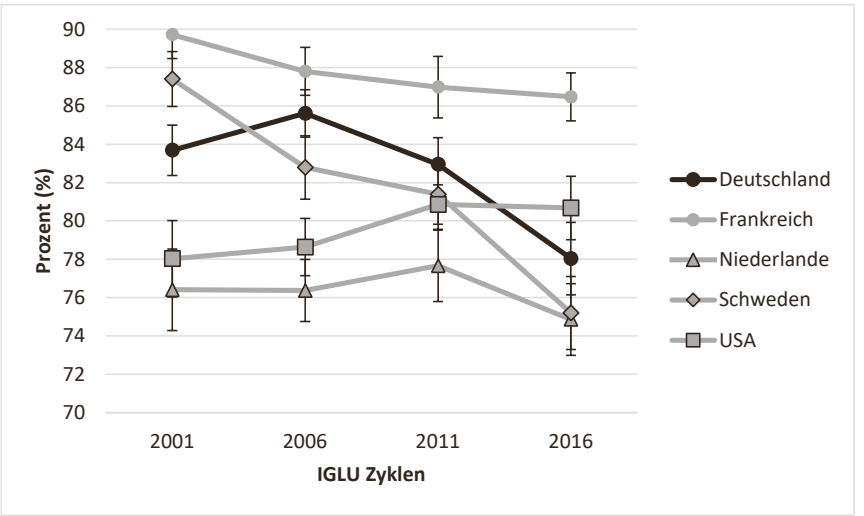
Anmerkungen. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Die Frage lautete: „Wie denkst du über deine Schule?" Die Punkte geben an, wie viel Prozent der Kinder auf das Item „Ich gehe gerne in die Schule" mit „stimme stark zu" oder „stimme einigermaßen zu" antworteten.

Für das Wohlbefinden der Schülerinnen und Schüler in Deutschland scheint sich damit im Vergleich der betrachteten Staaten ein geringeres Sicherheitsgefühl in der Schule und ein geringerer Anteil an Schulfreude abzuzeichnen.

3.5. Trends der Bedingungsfaktoren der Einstellungen zum Lesen

In Abbildung 9 sind die Angaben zu den Einstellungen der Viertklässlerinnen und Viertklässler zum Lesen abgetragen. Die Abbildung stellt dar, inwiefern die Viertklässlerinnen und Viertklässler das Lesen mögen. Der Vergleich der Teilnehmerstaaten zeigt, dass in Frankreich die meisten Schülerinnen und Schüler angaben, das Lesen zu mögen. Diese Tendenz war auch im zeitlichen Verlauf stabil auf hohem Niveau. Für die Niederlande wird hingegen ebenfalls ein stabiler Trend deutlich, jedoch auf geringerem Niveau und, über die Zeit gesehen, mit den geringsten Zustimmungsanteilen im Vergleich der fünf Staaten. Besonders auffällig sind die negativen Trends bezüglich des Anteils an Schülerinnen und Schülern, die das Lesen mögen, für Deutschland und Schweden. Beide Staaten lagen 2016 etwa im Bereich des Wertes für die Niederlande.

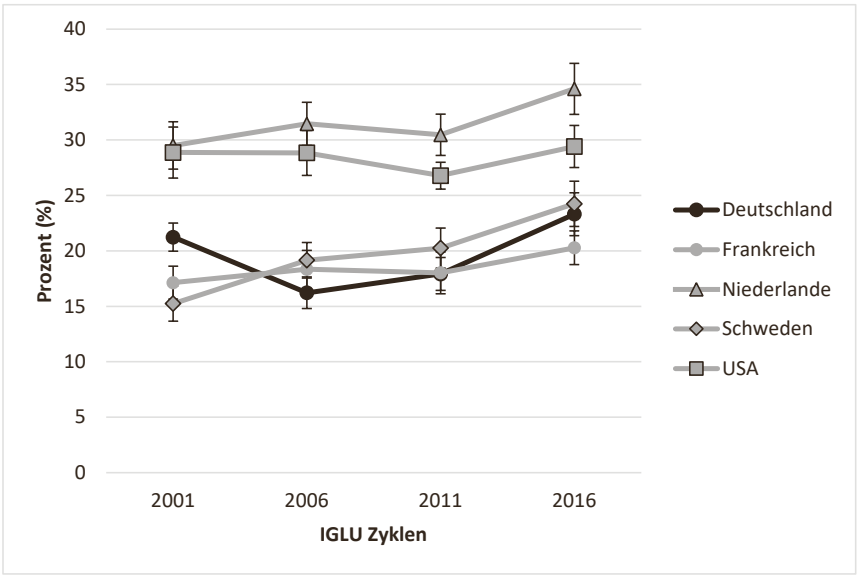
Abbildung 9: Anteile der Schülerinnen und Schüler, die angaben, Lesen zu mögen (Angaben der Schülerinnen und Schüler, Anteile in Prozent)



Anmerkungen. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Die Frage lautete: „Wie denkst du über das Lesen?“ Die Punkte geben an, wie viel Prozent der Kinder auf das Item „Ich lese gerne“ mit „stimme stark zu“ oder „stimme einigermaßen zu“ antworteten.

Zudem wurde die Angabe der Schülerinnen und Schüler zu der Aussage „Ich finde Lesen langweilig“ erfasst. Übereinstimmend mit den Ergebnissen dazu, ob die Kinder das Lesen mögen, verdeutlicht Abbildung 10, dass vergleichsweise hohe Anteile der Schülerinnen und Schüler in den Niederlanden Lesen langweilig fanden, während die entsprechenden Anteile in Frankreich geringer ausfielen. Während Deutschland zwischen 2001 und 2006 einen sinkenden Anteil bzgl. der Zustimmung zu dieser Aussage verzeichnete, stiegen die Werte seitdem tendenziell an.

Abbildung 10: Anteile der Schülerinnen und Schüler, die der Angabe zustimmten, dass Lesen langweilig ist (Angaben der Schülerinnen und Schüler, Anteile in Prozent)



Anmerkungen. Nicht überlappende Konfidenzintervalle (SE*2) zeigen statistisch signifikante Unterschiede an. Die Frage lautete: „Wie denkst du über das Lesen?". Die Punkte geben an, wie viel Prozent der Kinder auf das Item „Ich finde Lesen langweilig" mit „stimme stark zu" oder „stimme einigermaßen zu" antworteten.

4. Zusammenfassung und Diskussion

Internationale groß angelegte Schulleistungsstudien sind ein zentraler Baustein des Bildungsmonitorings und Grundlage für die Qualitätssicherung im Schulsystem (Bos, Gebauer & Postlethwaite, 2018; Grünkorn, Klieme & Stanat, 2019; European Commission, 2018). Da die Bildungssysteme weltweit einem raschen Wandel unterworfen sind, müssen sich auch die internationalen Large-Scale-Assessments weiterentwickeln und diese Bedingungen berücksichtigen (Harju-Luukkainen, McElvany & Stang, 2020). Damit werden neue oder veränderte Konstrukte eingeführt und Konstrukte mit abnehmender Bedeutung entfallen aus den Hintergrundfragebögen. Für die Betrachtung der Ergebnisse im Trend über mehrere Erhebungszyklen hinweg reduziert sich somit das Set an Indikatoren, die für einen internationalen Vergleich von Trends herangezogen werden können. Für den vorliegenden Beitrag konnten für relevante Bereiche zur Beschreibung der Bedingungen des Lesenlernens von Grundschulkindern Indikatoren ausgewählt werden, die eine Betrachtung von Trends ermöglichen und neben den Veränderungen in Deutschland im

Verlauf von 15 Jahren auch einen Vergleich zu den Veränderungen der Bedingungen des Lesenlernens in anderen Bildungssystemen im selben Zeitraum ermöglichen. Damit steht nicht wie zumeist fokussiert die durchschnittliche Lesekompetenz von Schülerinnen und Schülern im Zentrum des Beitrags, sondern die Frage danach wie sich bedeutsame Bedingungsfaktoren der Lesekompetenz am Ende der Grundschulzeit im zeitlichen Verlauf darstellen.

Die Betrachtung der *Schülerkomposition* zeigt für Deutschland eine Besonderheit im Vergleich zu den weiteren herangezogenen Vergleichsstaaten im Hinblick auf die Anzahl der Bücher im Haushalt, die klassischerweise in der empirischen Bildungsforschung primär Hinweise auf das in einer Familie zur Verfügung stehende soziokulturelle Kapital liefert. Während sich für die Vergleichsstaaten eine sinkende Anzahl an Büchern abzeichnete, blieb diese in Deutschland über die IGLU-Zyklen 2001 bis 2016 stabil. Vertiefende Analysen zeigten, dass diese Variable in Deutschland zu den jeweiligen Messzeitpunkten gute Messeigenschaften aufwies (Schwippert, 2019), dies wäre auch für die betrachteten Vergleichsstaaten zu prüfen. Gerade vor dem Hintergrund der gestiegenen Verfügbarkeit digitaler Medien, die Bücher im Haushalt zukünftig weiter ersetzen könnten, sowie des viel diskutierten Digitalisierungsschubs durch die COVID-19-Pandemie ist für internationale Schulleistungsstudien anzuraten, die sich wandelnde Bedeutung gedruckter Bücher im Haushalt als Indikator des soziokulturellen Kapitals gemeinsam mit anderen Indikatoren zu betrachten und insbesondere für den internationalen Vergleich differenzielle Entwicklungsverläufe zu berücksichtigen.

Für die *häuslichen Lernressourcen* zeigte sich eine ähnliche Entwicklung in allen fünf betrachteten Staaten, wobei für die Internetverfügbarkeit jedoch eine bessere Ausstattung in den Niederlanden und Schweden deutlich wurde. Vor dem Hintergrund der Relevanz der häuslichen Ausstattung, die auch insbesondere während der pandemiebedingten Abweichungen vom Präsenzunterricht verstärkt für den Primarbereich und die Lesekompetenz deutlich wurde (Huber, 2021), ist die Sicherstellung der Verfügbarkeit dieser Lernressourcen für Schülerinnen und Schüler ein wichtiges Anliegen, um Lernergebnisse zu sichern. Die Verfügbarkeit digital präsentierter Informationen und auch die Möglichkeiten des Austausches können Lernprozesse fördern.

Für die *Computernutzung in der Schule* ließ sich ein paralleler Verlauf in den betrachteten Staaten erkennen, der insbesondere mit einem Rückgang des Anteils im Jahr 2016 einherging. Lediglich Schweden verzeichnete hierbei einen Anstieg in 2016 und damit eine intensivere Integration digitaler Medien in den Unterricht. Im Zuge der Entwicklungen und bildungspoliti-

schen Anstrengungen der vergangenen Jahre seit 2016 wäre eine Zunahme der Nutzung digitaler Medien im Unterricht zu erwarten. Vielmehr als die quantitative Angabe der Nutzungshäufigkeit wären im internationalen Vergleich jedoch die Nutzungsweisen, und hier insbesondere die gezielte Förderung der Lesekompetenz mit digitalen Medien zu betrachten. Aufschluss darüber zu erhalten, auf welche Weise und zu welchen Zwecken digitale Endgeräte im Unterricht eingesetzt werden, würde die Option eröffnen, Potenziale für die Leseförderung abzuleiten, die schulsystemübergreifend identifiziert werden können.

Die *Einstellungen zur Schule* sind in Deutschland im Vergleich zu den weiteren betrachteten Staaten ungünstiger ausgeprägt mit im Durchschnitt zumeist geringeren Anteilen an Kindern, die angeben, sich in der Schule sicher zu fühlen oder gerne zur Schule zu gehen. Das schulische Wohlbefinden der Schülerinnen und Schüler in Deutschland scheint damit eher geringer zu sein. Dies ist ein Befund, der insbesondere durch die weltweiten Auswirkungen der Pandemie auf den schulischen Betrieb von Interesse ist, und dessen weitere Entwicklung im Trend zu IGLU 2021 besonderen Umständen unterlag. Mit dem IQB-Bildungstrend konnten keine bedeutsamen Veränderungen in der Schulzufriedenheit von Grundschulkindern im Vergleich der Jahre 2016 und 2021 beobachtet werde (Stanat et al., 2021). Für Deutschland lässt sich damit aufgrund bereits ungünstigerer Ausgangslagen im Zyklus 2016 als in den Vergleichsstaaten ableiten, das Wohlbefinden der Schülerinnen und Schüler am Ende der Grundschulzeit und damit vor der wichtigen Gelenkstelle des Übergangs in die weiterführende Schule in den meisten Ländern der Bundesrepublik Deutschland zu beachten.

Die *Einstellungen zum Lesen* der Viertklässlerinnen und Viertklässler in Deutschland verschlechterten sich tendenziell, insbesondere mit Blick auf die IGLU-Zyklen 2006 bis 2016. Hierbei lag Deutschland im Vergleich zu den weiteren betrachteten Staaten 2016 etwa auf demselben Niveau. Dennoch ist es angesichts aktueller Befunde zu geringeren Lesekompetenzen von Kindern (Stanat et al., 2022) ein beunruhigendes Ergebnis. Das Leseinteresse und die Lesemotivation, die eng mit der Lesekompetenz verwoben sind, sollten daher gefördert werden. Eine wichtige Aufgabe der Schule wie auch der Lesesozialisation im Elternhaus scheint es daher zu sein, die Freude der Kinder am Lesen durch gemeinsame Leseaktivitäten und vielfältige, auf ihre Lebenswelt und Interessen abgestimmte Angebote aufrecht zu erhalten. Vertiefend könnte zudem der sich für Deutschland von den Trends der anderen Vergleichsstaaten abhebende gleichbleibende Bücherbesitz im zeitlichen Verlauf bei gleichzeitig sinkendem Leseinteresse der Schülerinnen und

Schüler untersucht werden. Hierbei ist eventuell die stärkere Einbindung digitaler Medien in den Vergleichsstaaten ein zu berücksichtigender Faktor.

4.1. Limitationen

Zur Untersuchung der relevanten Indikatoren von zentralen Bedingungsfaktoren der Lesekompetenz von Viertklässlerinnen und Viertklässlern liegen lediglich wenige Einzelitems unverändert über die gesamte Spanne von 15 Jahren IGLU vor. Wünschenswert wäre eine größere Anzahl von Items oder Skalen hinsichtlich der Bedingungsfaktoren vorzufinden, die diese Bereiche im Zeitverlauf umfassender beschreibbar machen.

4.2. Implikationen

Ein methodisches Anliegen für die Weiterentwicklung im Bereich der Large-Scale-Assessments besteht darin, zentrale Konstrukte über eine längere Zeit unverändert beizubehalten, um das Potenzial von Trendanalysen aufrecht zu erhalten. Dies ist angesichts des raschen Wandels gesellschaftlicher Bedingungen ein anspruchsvolles Feld, das jedoch für die Bedeutsamkeit solch groß angelegter Studien sehr zuträglich ist. Nicht nur querschnittliche Aussagen über Kompetenzen von Schülerinnen und Schülern sowie die Rahmenbedingungen des Erwerbs dieser sind damit möglich, sondern auch das Potenzial, Trends in Bildungssystemen zu verdeutlichen und mit anderen Bildungssystemen zu vergleichen.

Die Ergebnisse zeigten zudem für bildungspolitische Strategien eine evidenzbasierte Grundlage für Entscheidungen, die die Bedingungen des Lesenlernens von Schülerinnen und Schülern am Ende der Grundschulzeit verbessern können. Im Vergleich zu den anderen Staaten wurde für Deutschland insbesondere deutlich, dass sowohl die Internetanbindung zu Hause als auch die mindestens wöchentliche Computernutzung in der Schule in den anderen Teilnehmerstaaten höher ausfallen und Deutschland bezüglich dieser beiden digitalisierungsbezogenen Indikatoren der Bedingungen des Lesenlernens von Grundschulkindern hinter den zum Vergleich herangezogenen Staaten zurücklag. Ähnlich verhält es sich für die Angaben zur Einstellung zur Schule, bei denen Deutschland hinter den betrachteten Vergleichsstaaten zurückblieb. Damit konnte die Betrachtung zentraler Lernbedingungen der Lesekompetenz im internationalen Vergleich und im Trend über die Zeitspanne von 15 Jahren IGLU Bereiche identifizieren, die

für Deutschland Handlungsoptionen zur Optimierung der Lernbedingungen und damit verbunden der Unterstützung der Lesekompetenzentwicklung aufzeigten. Verstärkte Bemühungen hinsichtlich der Optimierung dieser Lernbedingungen sind zur Aufrechterhaltung und Weiterentwicklung der Qualität des Bildungssystems in Deutschland damit zu erwägen, um an internationale Vergleichsstaaten aufzuschließen. Insbesondere bleibt zu betrachten, wie sich die Trends angesichts der Pandemieauswirkungen auf die zentralen Lernbedingungen mit IGLU 2021 fortsetzen. Die Erhebungen etwa ein Jahr nach Beginn der pandemiebedingten Auswirkungen auf den Unterricht und das Lernen können dabei weitere wertvolle Hinweise auf die Entwicklungen in Deutschland und im Vergleich zu anderen Schulsystemen liefern. Die ersten Ergebnisse zu IGLU 2021 werden mit der Berichtslegung in 2023 (McElvany et al., 2023) erfolgen. Mit den Daten aus IGLU 2021 wird erstmalig eine Betrachtung der in diesem Beitrag fokussierten Bedingungen der Lesekompetenz von Schülerinnen und Schülern in Deutschland am Ende der Grundschulzeit über eine Zeitraum von 20 Jahren möglich sein.

Literatur

Bos, W., Gebauer, M. M., & Postlethwaite, T. N. (2018). Möglichkeiten, Grenzen und Perspektiven internationaler Schulleistungsforschung. In R. Tippelt & B. Schmidt-Hertha (Hrsg.), *Handbuch Bildungsforschung* (4. Aufl., S. 325–350). Springer VS. https://doi.org/10.1007/978-3-531-19981-8_13

Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K., & Valtin, R. (Hrsg.). (2007). *IGLU 2006. Lesekompetenzen von Grundschulkindern im Deutschland im internationalen Vergleich*. Waxmann.

Bos, W., Lankes, E.-M., Schwippert, K., Valtin, R., Voss, A., Badel, I., & Plaßmeier, N. (2003). Lesekompetenzen deutscher Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther, & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 69–141). Waxmann.

Bos, W., Tarelli, I., Bremerich-Vos, A., & Schwippert, K. (Hrsg.). (2012). *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann.

BIFIE, Robitzsch, A., & Oberwimmer, K. (2022). BIFIEsurvey: Tools for survey statistics in educational assessment. R package version 3.5-3. https://CRAN.R-project.org/package=-BIFIEsurvey

Boudon, R. (1974). *Education, opportunity, and social inequality: Changing prospects in Western society*. Wiley.

Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 183–198). Schwartz.

Bremerich-Vos, A., Wendt, H., & Bos., W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe, & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 97–142). Waxmann.

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review, 25*, 23–38. https://doi.org/10.1016/j.edurev.2018.09.003

Ditton, H. (2016). Evaluation und Qualitätssicherung im Bildungsbereich. In R. Tippelt & B. Schmidt-Hertha (Hrsg.), *Handbuch Bildungsforschung*. Springer. https://doi.org/10.1007/978-3-531-20002-6_32-1

Dong, Y., Wu, S., Dong, W.-Y., & Tang, Y. (2020). The effects of home literacy environment on children's reading comprehension development: a meta-analysis. *Educational Sciences: Theory and Practice, 20*(2), 63–82. https://doi.org/10.12738/jestp.2020.2.005

Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (Hrsg.). (2019). *ICILS 2018 #Deutschland: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking*. Waxmann. https://doi.org/10.25656/01:18166

European Commission. (2018). *Quality assurance for school development. Guiding principles for policy development on quality assurance in school education*. https://www.schooleducationgateway.eu/downloads/Governance/2018-wgs2-quality-assurance-school_en.pdf

Grünkorn, J., Klieme, E., & Stanat, P. (2019). Bildungsmonitoring und Qualitätssicherung. In O. Köller, M. Hasselhorn, F. W. Hesse, K. Maaz, J. Schrader, & H. Solga (Hrsg.), *Das Bildungswesen in Deutschland – Bestand und Potenziale* (S. 263–298). utb.

Harju-Luukkainen, H. K., McElvany, N., & Stang, J. (Hrsg.). (2020). *Monitoring of student achievement in the 21st century – European policy perspectives and assessment strategies*. Springer.

Hascher, T. (2004). *Wohlbefinden in der Schule*. Waxmann.

Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 71–175). Hogrefe.

Huber, S. G. (2021). Schooling and Education in Times of the COVID-19 Pandemic: Food for Thought and Reflection Derived From Results of the School Barometer in Germany, Austria and Switzerland. *Studies in Educational Administration, 49*(1), 6–17. https://edulead.net/schuba/wp-content/uploads/sites/5/2021/01/Schul-Barometer_Huber_Schooling_and_Education_in_Times_of_the_COVID-19_Pandemic_21-01-10.pdf

Huebner, E. S., & Gilman, R. (2006). Students who like and dislike school. *Applied Research in Quality Of Life, 1*(2), 139–150. https://doi.org/10.1007/s11482-006-9001-3

Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D, Lankes, E.-M., McElvany, N., Stubbe, T. C., & Valtin, R. (Hrsg.). (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern im internationalen Vergleich*. Waxmann. https://doi.org/10.25656/01:15476

Kanonire, T., Federiakin, D. A., & Uglanova, I. L. (2020). Multicomponent framework for students' subjective well-being in elementary school. *School Psychology, 35*(5), 321–331. https://doi.org/10.1037/spq0000397

Kao, G., & Tienda, M. (1995). Optimism and achievement: the educational performance of immigrant youth. *Social Science Quarterly, 76*, 1–19.

Kleinkorres, R., Stang, J., & McElvany, N. (2020). A longitudinal analysis of reciprocal relations between students' well-being and academic achievement. *Journal for Educational Research Online, 12*(2), 114–165. https://doi.org/10.25656/01:20975

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review, 25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Kristen, C., & Dollmann, J. (2012). Migration und Schulerfolg: Zur Erklärung ungleicher Bildungsmuster. In M. Matzner (Hrsg.), *Handbuch Migration und Bildung* (S. 102–117). Beltz.

Lorenz, R., & Bos, W. (2017). Schule digital – der Länderindikator 2017. Theoretisches Rahmenmodell, Überblick über die Befunde des Länderindikators 2017 und Einordnung zentraler Ergebnisse der Erhebungszyklen 2015, 2016 und 2017. In R. Lorenz, W. Bos, M. Endberg, B. Eickelmann, S. Grafe, & J. Vahrenhold (Hrsg.), *Schule digital – der Länderindikator 2017. Schulische Medienbildung in der Sekundarstufe I mit besonderem Fokus auf MINT-Fächer im Bundesländervergleich und Trends von 2015 bis 2017* (S. 11–35). Waxmann. https://doi.org/10.25656/01:15656

Lorenz, R., Brüggemann, T., Stang, J., & McElvany, N. (2023). Unterricht zu Beginn und nach einem Jahr der Corona-Pandemie – Lehrkräftebefragungen zum Lernen mit digitalen Medien im Vergleich. In S. G. Huber & C. Helm (Hrsg.), *Covid-19 und Bildung – Studien und Perspektiven*.

Mangen, A., & van der Weel, A. (2016). The evolution of reading in the age of digitisation: an integrative framework for reading research. *Literacy, 50*(3), 116–124. https://doi.org/10.1111/lit.12086

Martin, M. O., Mullis, I. V. S., & Hooper, M. (Hrsg.). (2017). *Methods and procedures in PIRLS 2016*. TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, MA. https://timssandpirls.bc.edu/publications/pirls/2016-methods.html

McElvany, N., Lorenz, R., Frey, A., Goldhammer, F., Schilcher, A., & Stubbe, T. (Hrsg.). (2023). IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre. Waxmann.

McElvany, N., Becker, M., & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 41*(3), 121–131. https://doi.org/10.1026/0049-8637.41.3.121

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation, 51*, 297–384. https://doi.org/10.1016/S0079-7421(09)51009-2

Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of Educational Psychology, 106*(1), 121–131. https://doi.org/10.1037/a0033546

Mullis, I. V. S., & Martin, M. O. (2019). *PIRLS 2021 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/pirls2021/frameworks/

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. http://timssandpirls.bc.edu/pirls2016/international-results/

OECD – Organisation for Economic Co-operation Development (2017). *PISA 2015. Results (Volume III): Students' Well-Being*. OECD Publishing.

Robitzsch, A., Grund, S., & Henke, T. (2022). *Package 'miceadds'*. https://www.vps.fmvz.usp.br/CRAN/web/packages/miceadds/miceadds.pdf

Savolainen, H., Ahonen, T., Aro, M., Tolvanen, A. & Holopainen, L. (2008). Reading comprehension, word reading and spelling as predictors of school achievement and choice of secondary education. *_Learning and Instruction, 18*(2), 201–210.

Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: a five-year longitudinal study. *Child Development, 73*(2), 445–460. https://doi.org/10.1111/1467-8624.00417

Schaumburg, H. (2018). Empirische Befunde zur Wirksamkeit unterschiedlicher Konzepte des digital unterstützten Lernens. In N. McElvany, F. Schwabe, W. Bos, & H. G. Holtappels, (Hrsg.), *Digitalisierung in der schulischen Bildung. Chancen und Herausforderungen* (S. 27–40). Waxmann.

Schwippert, K. (2019). Was wird aus den Büchern? Sozialer Hintergrund von Lernenden und 915 Bildungsungleichheit aus Sicht der international vergleichenden Erziehungswissenschaft. *Journal for Educational Research Online, 11*(1), 92–117. https://doi.org/10.25656/01:16789

Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., & Henschel, S. (Hrsg.). (2022). *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Waxmann. https://doi.org/10.31244/9783830996064

Voogt, J., Knezek, G., Christensen, R., & Lai, K.-W. (2018). *Second Handbook of Information Technology in Primary and Secondary Education*. Springer. https://doi.org/10.1007/978-3-319-71054-9

Zierer, K. (2020). Die Wirkung digitaler Medien im Schulunterricht – Chancen und Risiken der Digitalisierung aus erziehungswissenschaftlicher Sicht. In R. A. Fürst (Hrsg.), *Digitale Bildung und Künstliche Intelligenz in Deutschland*. Springer. https://doi.org/10.1007/978-3-658-30525-3_15

Quality of ICT Use in German Secondary Schools During the COVID-19 Pandemic: Predictors on the School and Teacher Level

Sittipan Yotyodying & Ramona Lorenz

Abstract

As a consequence of the global spread of COVID-19, the traditional method of on-site teaching and learning has shifted to more hybrid approaches. Thus, ensuring the quality of ICT use in teaching and learning in school has become even more important. The aim of this article was to examine teachers' perceptions regarding predictors of the quality of ICT use in school during the pandemic. We focused on two quality features: organization of teaching and student learning behavior. Furthermore, we distinguished between predictors on the school level (i. e., school's ICT equipment, investment in the school's digital infrastructure, school type) and teacher-side predictors (i. e., ICT self-efficacy, ICT learning opportunities, gender). The data stemmed from a representative survey of secondary school teachers in Germany (N = 1512). Overall, the SEM analyses revealed that ICT equipment, school type, and ICT self-efficacy were significantly associated with both quality features of ICT use. Investment in digital infrastructure and ICT learning opportunities were only significantly associated with organization of teaching, while gender was only significantly related to student learning behaviors. Findings and implications regarding the quality of ICT use in school are discussed.

Keywords: Quality of ICT use in school; school digital equipment, development by investment programs, teacher digital self-efficacy; teaching organization; student learning behaviors; COVID-19 pandemic

Zusammenfassung

Infolge der weltweiten Ausbreitung der COVID-19-Pandemie wandelten sich die traditionellen Präsenz-Lernformate zu hybriden Lehr-Lernansätzen. Daher stieg die Relevanz, die Qualität der Nutzung digitaler Medien an Schulen sicherzustellen. Das Ziel dieses Artikels war es, die von Lehrkräften wahrgenommenen Prädiktoren zu untersuchen, die zur Qualität der Nutzung digitaler Medien in der Pandemiezeit beitragen konnten. Wir fokussierten dabei auf zwei Nutzungsmerkmale: die Unterrichtsorganisation und das Lernverhalten der Schülerinnen und Schüler. Darüber hinaus differenzierten wir zwischen schulseitigen Prädiktoren (ICT-Ausstattung, Entwicklung durch Investitionsprogramme, Schulform) und lehrerseitige Prädiktoren (ICT-Selbstwirksamkeit, ICT-Lernangebote, Geschlecht). Die Datenanalysen basierten auf einer repräsentativen Befragung der Lehrkräfte der Sekundarstufe I in Deutschland (N = 1512). Insgesamt zeigten die SEM-Analysen, dass schulische Ausstattung, Schulform und digitale Selbstwirksamkeit in einem signifikanten Zusammenhang mit beiden Nutzungsmerkmalen standen. Die Entwicklung durch Investitionsprogramme und ICT-Lernangebote standen nur in einem signifikanten positiven Zusammenhang mit der Unterrichtsorganisation, wobei das Geschlecht war nur mit dem Lernverhalten der Schülerinnen und Schüler signifikant verknüpft. Ergebnisse und Implikationen dieser Studie werden im Hinblick auf Nutzungsqualität digitaler Medien diskutiert.

Schlagworte: Qualität der Nutzung digitaler Medien an Schulen; ICT-Ausstattung, Entwicklung durch Investitionsprogramme, ICT-Selbstwirksamkeit; Unterrichtsorganisation; Lernverhalten der Schülerinnen und Schüler; COVID-19-Pandemie

1. Introduction

Information and communication technology (ICT)-related skills *(information, communication and technical)* – along with other skills like collaboration, critical thinking, creativity, and problem-solving – are considered key competences for the *twenty-first century* (Van Laar et al., 2020). Over the past few years, ICT education in Germany has drawn increasing attention from academics and policymakers, resulting in national educational policy frameworks that tasked schools with fostering students' ICT competences in all subjects (KMK, 2017, 2021). A large nationwide investment program beginning in 2019, known as the *Digitalpakt Schule,* made six billion Euro avail-

able for German schools to invest in digital infrastructure, with additional investments also made by the federal states (BMBF, 2021). However, recent representative surveys demonstrate that there is still room for improvement in terms of media education in German schools and students' ICT competences (e. g., Eickelmann et al., 2019; forsa, 2019).

Since early spring 2020, the COVID-19 pandemic has had a major impact on educational systems worldwide, with traditional in-person classroom teaching shifting to remote learning, the hybrid classroom or a mixture of hybrid and on-site classes (Reintjes et al., 2021). Recent studies have also revealed negative impacts of COVID-19 on students' learning and school performance (e. g., Bird et al., 2022; Ludewig et al., 2022) as well as their mental health (e. g., UNICEF, 2020). For this reason, ensuring the quality of teaching and learning processes with digital media in schools has become even more important. Aspects of this include ICT management and support as well as ICT use to support students' learning (Lorenz & Bos, 2017). In the past few years, research has explored possible predictors of the process of teaching and learning with digital media, especially teachers' ICT use for teaching and learning in school, and sought to explain these relationships. Teachers' ICT use has been found to be a crucial predictor of teachers' efforts to foster students' ICT competences (e. g., Lorenz et al., 2019) as well as of students' ICT competences (e. g., Gerick et al., 2017). However, it remains unclear whether these relationships held in the pandemic context.

Based on empirical research, Lorenz and Bos (2017) proposed a complex framework explaining the intercorrelations between various quality dimensions of ICT use in the teaching and learning process. Empirical findings drawing upon this framework (see also Gräsel et al., 2020) have shown that ICT equipment at school as well as teachers' ICT-related attitudes, teachers' participation in ICT trainings, and teachers' ICT-related self-efficacy were positively associated with teachers' ICT use, which in turn fostered students' ICT competences (e. g., Gerick et al., 2017; Hatlevik & Hatlevik, 2018; Lorenz et al., 2019; OECD, 2020). However, it must be noted that past research has typically defined teachers' ICT use in terms of *quantity or frequency of use* (e. g., how often do teachers use ICT in school). However, much less is known about the quality of ICT use in school (*how and in what way*), and most importantly, about factors that may predict different quality levels of ICT use in school. The COVID-19 pandemic has not only changed the requirements for ICT equipment at school, but has also had an impact on teaching and learning processes with digital media (Lorenz & Eickelmann, 2022). To the best of our knowledge, few studies have systematically investigated the question of *how (and in what way)* teachers used ICT in teaching and learn-

ing during the pandemic. To shed light on this issue, this article will provide empirical evidence regarding crucial factors that may contribute to predicting the quality of ICT use in school during the pandemic.

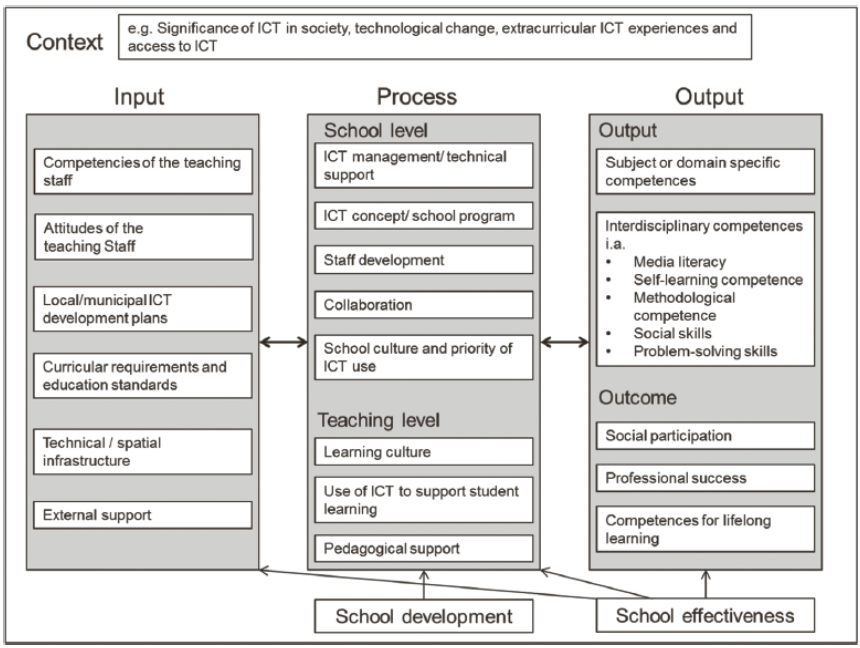
2. Theoretical Background

2.1. Theoretical Framework of Quality Dimensions of Teaching and Learning with Digital Media

Over the past decade, a number of researchers have attempted to explain the relationship between school development and school effectiveness. In doing so, they have proposed different indicators that can be categorized into three levels: input, process, and output (e. g., Creemer & Kyriakides, 2008; Ditton, 2000; Eickelmann & Schulz-Zander, 2008; Scheerens, 1990). Eickelmann and Schulz-Zander (2008) focused on media education in school. In contrast to previous studies, Lorenz and Bos (2017) proposed that indicators on the input, process, and output levels of school development and school effectiveness are *intercorrelated*. In their model (Figure 1), *ICT-related contexts* (i. e., significance of ICT in society, technological change, ICT experiences outside of school, and access to ICT) shape school-related, teaching-related, and individual indicators. At *the input level*, this model focuses on the importance of teachers' ICT-related competences and teachers' attitudes towards media education in school. Apart from teacher factors, the input level also includes school factors, such ICT development plans, curricular requirements and educational standards, ICT infrastructure in school, and support for ICT use.

At *the process level*, a distinction is made between school-level processes and teaching-level processes. The former are concerned with ICT management/support, an ICT vision, staff development as well as the school culture of ICT use. The latter, according to Lorenz and Bos (2017), includes three important factors for teaching with ICT: learning culture, use of ICT to support student learning, and pedagogical support. Likewise, a distinction is made between output and outcome indicators at the output level. Most importantly, it is suggested that school development takes place at the process level, while school effectiveness is related to all three levels.

Figure 1: Theoretical Framework for Quality Dimensions of Teaching and Learning with Digital Media by Lorenz & Bos (2017) (English translation published in Lorenz et al., 2019).



Linking this model to the present study, we aimed to replicate the relationships among input and process factors in the pandemic context. We operationally defined the quality of ICT use to support student learning during the pandemic by focusing on two dimensions: (a) organization of teaching and (b) students' learning behaviors. These quality features were operationally characterized on the basis of the theoretical model by Lorenz and Bos (2017) as well as by recent empirical evidence (e. g., Lorenz & Eickelmann, 2022). *The first quality feature* concerns how teaching with digital media was organized during the pandemic. *The second quality feature* assesses teachers' perceptions of whether students' learning behaviors during the pandemic were consistent with students' *on-site learning behaviors* before the pandemic.

2.2. ICT Equipment at School

ICT equipment is a fundamental resource that plays a central role in processes of teaching and learning with digital media in schools, thereby fostering students' ICT competences as well as other domains of competence (e. g., Lorenz

& Bos, 2017; Lorenz & Endberg, 2017). However, internationally representative research in recent years, such as the *ICILS 2013* (Bos et al., 2014) and *ICILS 2018* (Eickelmann et al., 2019), reported that ICT equipment (e. g., in terms of availability of computers, a Wi-fi connection, and learning management systems) in German schools lagged behind that in other countries. In the literature, there is strong evidence that ICT equipment matters for ICT use in school. As part of *ICILS 2013*, Gerick et al. (2017) examined the connections between school-level predictors and ICT use in four countries: Australia, the Czech Republic, Germany, and Norway. They only found a significantly negative connection between student-computer ratio (whether students need to share a computer at school) and the frequency of teachers' use of computers in the Australian sample. Nevertheless, Germany-wide representative research (forsa, 2019; Yotyodying & Lorenz, 2022) has reported that ICT equipment in schools still needs to be further improved, especially with regard to Internet speed. Analyzing representative teacher data, Lorenz et al. (2019) found that teachers who reported greater availability of ICT equipment at school (e. g., sufficient number of computers, sufficient connectivity) tended to report a greater frequency of ICT use as well.

2.3. Development of Schools' Digital Infrastructure through Investment Programs

In the past few years, the German federal government and federal states (*Bund und Länder*) have provided schools with an enormous funding to improve digital infrastructure in schools (BMBF, 2021). Following the onset of the COVID-19 pandemic, these investment programs were expanded (BMBF, 2021) to ensure high-quality teaching and learning with digital media during the crisis, which involved a shift from on-site classes to hybrid or mixed classes (Reintjes et al., 2021). To the best of our knowledge, the impact of these investment programs on the quality of teaching and learning with digital media has not been systematically investigated, especially with respect to teachers' perceptions of such investment programs to develop digital infrastructure in schools. So far, it has only been revealed that teachers perceive more progress in the areas of learning platforms/learning management system (81 %) and Wi-Fi connectivity (71 %), and less progress in the areas of cloud solutions (61 %), digital textbooks/schoolbooks (58 %), and internet speed (41 %) (Yotyodying & Lorenz, 2022). The present study considered ICT equipment in schools as an assumed predictor of the quality of ICT use in teaching and learning during the pandemic. It is therefore reasonable to examine teachers'

perceptions of the impact of investment programs on digital infrastructure in schools during the pandemic.

2.4. Teachers' ICT Self-efficacy

Apart from sufficient ICT equipment at school, Lorenz and Bos (2017) also proposed that ICT-related competences among the teaching staff are an important antecedent of successful teaching and learning processes with digital media. The *TPACK model* (technological pedagogical and content knowledge: Mishra & Koehler, 2006) has been applied in recent decades to assess teachers' ICT competences or ICT self-efficacy. The TPACK model explains the interactions and intersections among three different types of knowledge related to implementing technology in the classroom: technological knowledge, pedagogical knowledge, and content knowledge. Previous research has provided strong empirical evidence that teachers with greater ICT self-efficacy tend to report more ICT use in the classroom (e. g., Gerick et al., 2017; Hatlevik & Hatlevik, 2018; Katz, 2018). To date, relatively little is known about teachers' ICT self-efficacy and the quality of their ICT use during the pandemic, especially regarding teachers' ability to structure digitally-supported teaching and learning.

2.5. ICT Learning Opportunities

As mentioned earlier, *staff development in ICT*, such as teacher training on ICT-related content, is considered an important quality dimension in the processes of teaching and learning with digital media (Lorenz et al., 2019). Before the COVID-19 crisis, however, empirical evidence from the ICILS 2018 study showed that German teachers' participation in further professional development courses on ICT-related issues was low (Eickelmann et al., 2019). Other studies also found a dissatisfactory level of teacher participation in ICT training (e. g., forsa, 2020; Gerick et al., 2019; GEW, 2020). However, due to the new remote emergency way of digitally-supported teaching and learning during the pandemic, teachers have begun to exhibit greater interest in ICT-related training (e. g., Initiative D21, 2021). In fact, teachers' participation in professional development courses on ICT use can promote teachers' ICT use in their instruction. For instance, Gerick et al. (2017) examined the connection between teachers' participation in professional development courses on ICT use and their actual ICT use in teaching across four countries (Australia, Ger-

many, Norway, and the Czech Republic). The results confirmed a significant positive connection in Australia and Norway: The more ICT-related courses teachers participated in, the more ICT use in teaching they reported.

2.6. Socio-demographic Subgroups and ICT Use in School

The theoretical model by Lorenz and Bos (2017) does not include socio-demographic variables at the school and teacher levels (such as school type and teacher gender) as key factors for school development and school effectiveness in media education. However, earlier studies had revealed socio-demographic subgroup differences with regard to digital equipment in schools and use of digital media in the classroom. For example, previous studies have shown that use of media in the classroom differs between female and male teachers: Higher proportions of male teachers use digital media in the classroom compared to their female colleagues (e. g., Drossel et al., 2019). Sufficient digital infrastructure promotes greater ICT use in school (Lorenz & Bos, 2017). Turning to differences by school type, in Germany, *gymnasium*) as highest track of secondary schools has been found to be better equipped than other types of schools in terms of digital infrastructure (e. g., Eickelmann & Drossel, 2020; forsa, 2019; Lorenz & Bos, 2017). Lorenz and colleagues also found in their nationally representative study that a higher proportion of gymnasium teachers used computers in their lessons compared to teachers in other school types (e. g., Lorenz et al., 2017).

2.7. Research Questions

Before the COVID-19 pandemic, media education in Germany had already begun to receive greater attention from academics and policymakers. Researchers had sought to identify crucial predictors of different quality dimensions of teaching and learning with digital media, which in turn help to foster students' ICT competences. In the present study, we distinguished between predictors *at the school and teacher levels*. Most importantly, earlier studies had already provided empirical evidence that ICT equipment in school, teachers' ICT-related attitudes and ICT-related competences, as well as socio-demographic background various are crucial predictors of ICT use in school, thereby promoting students' ICT competences (e. g., Gerick et al., 2017; Lorenz et al., 2019). However, we knew very little about predictors of ICT use in school *during the pandemic*, and above all, *the quality* of ICT use in

teaching and learning. The aim of the present study was therefore to examine predictors of the quality of ICT use in teaching and learning in school during the pandemic. In doing so, we considered different types of ICT-related factors that encourage use of ICT in school to support students' learning (process: teaching level), input-level factors (e. g., ICT equipment at school, teachers' ICT-related competences) and school-level process (e. g., staff development in terms of ICT-related learning opportunities for teachers). In addition to ICT equipment at school, we also examined investment programs in digital infrastructure, which also became crucially important during the pandemic (Yotyodying et al., 2022). To the best of our knowledge, the present study was the first of its kind to examine such investments in the context of media education in Germany. Most importantly, we aimed to determine which group of variables is a better predictor of the quality of ICT use in school. The following three research questions were addressed:

1. Which predictive factors on *the school level* were associated with the quality of ICT use in teaching and learning during the pandemic?
2. Which predictive factors on *the teacher level* were associated with the quality of ICT use in teaching and learning during the pandemic?
3. Which group of predictive factors (school level versus teacher level) explained more variance in the quality of ICT use in teaching and learning during the pandemic?

We anticipated that our selected predictive factors on the school level (i. e., ICT equipment at school, investment programs to develop digital infrastructure, school type) as well as predictive factors on the teacher level (i. e., ICT self-efficacy, ICT learning opportunities, and teacher gender) would be associated with the quality of ICT use in teaching and learning in school during the pandemic (i. e., organization of teaching, student learning behaviors). The third research question was an exploratory research question. Hence, we did not have any specific predictions. As the background variables school type and teacher gender had previously been found to be related to the study variables (e. g., Drossel et al., 2019; Eickelmann & Drossel, 2020), these two variables were also included as predictors in the model.

3. Method

3.1. Participants and Procedure

The present study was a part of the project "Länderindikator 2021" (Lorenz et al., 2022) funded by Deutsche Telekom Stiftung (project duration: May 2021 to April 2022). The present analyses were based on data from a representative survey of secondary school teachers in the sixteen German federal states (with at least 50 teachers from each federal state). To adjust for the regional distribution of the current survey, a weighting variable was included in the data analyses. Overall, 1,512 secondary school teachers from all school types (excluding special needs schools) participated in the current research (58 % female; 45 % working in Gymnasium schools; average age: 45 years old [*SD* = 11.02]). The teacher survey took place in summer 2021 using fully-structured computer-assisted personal interviews (CAPI).

3.2. Measures

Perceived Sufficient Availability of ICT Equipment at School during the Pandemic

To assess teachers' perceptions regarding the availability of sufficient ICT equipment at school during the pandemic, we used a four-item scale developed by Lorenz et al. (2022) (e. g., "The teachers at my school had sufficient laptops available for distance and alternating teaching"). Teachers were asked to rate their agreement (on a 5-point Likert scale: 1 = *strongly disagree*, 5 = *strongly agree*) on the extent to which the ICT equipment at their school was sufficient in the past eight weeks during the pandemic. The internal consistency of this scale was satisfactory (Cronbach's alpha = .70). Scale items are available in the online supplement.

Development of School Digital Infrastructure via Investment Programs

We used a yes/no scale developed by Lorenz et al. (2022) to assess whether the school's digital infrastructure had been improved as part of an investment program. Teachers were asked whether (or not) the school has improved eight areas of digital infrastructure over the past two years. The eight areas included a) fiberoptic/broadband expansion, b) Wi-Fi, c) learning platforms and learning management systems, d) web conferencing tools, e) cloud solu-

tions, f) licensing of commercial online educational media, g) digital textbooks and schoolbooks, and h) stronger anchoring of media education in the school curriculum. A summative index of "yes" responses per individual (maximum: 8) was used in the analyses.

Teacher ICT Self-efficacy

Teachers' ICT-related self-efficacy was assessed with five items (e. g., "I can design lessons in a way that appropriately combines the subject content, the digital media used and the teaching methods used") based on the TPACK model (e. g., Endberg & Lorenz, 2016; Mishra & Koehler, 2006), with responses on a five-point rating scale (1 = *strongly disagree*, 5 = *strongly agree*). Overall, this scale yielded good internal consistency (Cronbach's alpha = .86). Scale items are available in the online supplement.

ICT Learning Opportunities

We assessed teachers' perceived ICT learning opportunities with four items (e. g., "In professional development courses, I learned how to use digital media as a means of teaching and learning in the classroom"). These items were developed by Lorenz et al. (2017). Responses were made on a five-point rating scale (1 = *strongly disagree*, 5 = *strongly agree*). This scale yielded good internal consistency (Cronbach's alpha = .76). Scale items are available in the online supplement.

Quality of ICT Use in Teaching and Learning during the Pandemic

In this article, we assessed two quality features of ICT use in teaching and learning during the pandemic: teaching organisation and student learning behaviors. The former was assessed with five items (e. g., "In digitally supported lessons, I was able to build on important preliminary work at my school"). The internal consistency of this scale was good (Cronbach's alpha = .75). Responses for the two scales were given on a four-point rating scale, ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). Teachers' perceptions of student learning behaviors during the pandemic were assessed with three items (e. g., "The students reliably completed their tasks in distance learning"). The Cronbach's alpha of this scale was satisfactory (Cronbach's alpha = .66). However, item-total correlations did not suggest deleting any items in order to improve the internal consistency. These three items were newly developed for this study.

Both scales to assess the quality of the teaching and learning process with digital media during the pandemic were developed by Lorenz et al. (2022). To examine whether the two scales captured separate constructs, we performed an exploratory factor analysis (EFA). It was found that eight items loaded onto two factors that accounted for 55.03 % of the variance. EFA loadings on the two factors ranged between .49 and .82. Scale items are available in the online supplement.

3.3. Data Analyses

In the present study, the results of the data analyses will be presented in two parts. To adjust for the regional distribution of the current survey, a weighting variable was included in all data analyses. First of all, we performed an exploratory factor analysis (EFA) using SPSS as well as a confirmatory factor analysis (CFA) using M*plus 8.8* (Muthén, & Muthén, 1998–2017) to test the construct validity of the scales used in this article. Additionally, we analyzed descriptive statistics for the scales and intercorrelations between latent factors. In the second part, we performed an SEM analysis using M*plus 8.8* (Muthén, & Muthén, 1998–2017) to examine our hypotheses. Figure 2 displays an SEM model with significant standardized parameter estimates for the assumed associations among study variables. In this model, two ICT-related predictor constructs and two outcome variables served as latent variables, while teacher gender and school type served as manifest variables. All standardized parameter estimates are also shown in Table 2. Teacher gender was operationalized as a *female teacher* or not, while school type was coded in terms of *Gymnasium teacher* (versus other secondary school types). For each latent construct, the corresponding scale items served as manifest variables.

4. Results

4.1. Testing Construct Validity

It was first necessary to ensure the construct validity of the scales used in the present study. To do so, we first performed an EFA for all items making up the five scales, with the exception of the yes/no scale to assess investments in the school's digital infrastructure. The EFA results showed that scale items loaded onto five factors that accounted for 59.20 % of the variance. EFA factor loadings can be seen in the online supplement. In a second

step, a confirmatory factor analysis (CFA) of the five-factor measurement model was then performed using M*plus 8.8* (Muthén & Muthén, 1998–2017). Overall, the CFA model demonstrated an acceptable fit to the data, χ^2 (136, N = 1512) = 7343.30, *CFI* = .93, *SRMR* = .05, *RMSEA* = .04 (see, Schreiber et al., 2006, for recommendations for the model evaluation). Standardized factor loadings ranged from .40 ($p < .001$) to .80 ($p < .001$), and all of them proved statistically significant.

Table 1: Sample Items, Internal Consistencies, and CFA Factor Loadings

Scale and sample item	*N* item	α	CFA Factor loading
ICT equipment at school during the pandemic			
Sample item: The teachers at my school had sufficient laptops available for distance and alternating teaching.	4	0.70	0.40–0.70
ICT self-efficacy			
Sample item: I can design lessons in a way that appropriately combines the subject content, the digital media used and the teaching methods used.	5	0.86	0.64–0.80
ICT learning opportunities			
Sample item: In professional development courses, I learned how to use digital media for teaching and learning in the classroom.	4	0.76	0.47–0.80
Organization of teaching during the pandemic			
Sample item: In the digitally supported lessons, I was able to build on important preliminary work at my school.	5	0.75	0.49–0.72
Student learning behaviors during the pandemic			
Sample item: The students reliably completed their tasks in distance learning.	3	0.66	0.56–0.71

Note. All standardized CFA factor loadings proved statistically significant ($p < .001$).

4.2. Correlational Analyses among Study Variables

We next performed correlational analyses among study variables. Five study variables served as latent constructs, while three variables served as manifest variables (e. g., ICT investment programs, school type, and gender). The significant standardized estimates for the correlations ranged from .20 ($p < .05$) to .60 ($p < .001$). Two quality features of ICT use in school during the pandemic were significantly positively correlated with one another. Overall, there were significant positive correlations between four out of six predictor constructs and two quality features. We found no significant correlations between gender and either quality feature. There was a significant positive correlation between school type (Gymnasium) and student learning behaviors during the pandemic, but school type was not significantly correlated with organization of teaching during the pandemic.

In addition, we computed scale means of the study variables to consider absolute levels of teachers' perceptions. The results showed that teachers

reported *high levels* of the three following factors: ICT self-efficacy, ICT learning opportunities, and organization of teaching during the pandemic. They reported *low to moderate levels* of ICT equipment at school, ICT investment, and student learning behaviors during the pandemic.

Table 2: Standardized Estimates for Correlations among Study Variables

Study variable	1	2	3	4	5	6	7	8
School-side predictors								
1. ICT equipment at school[a]	–							
2. ICT investment programs[b]	.44***	–						
3. Gymnasium[b]	.03	.02	–					
Teacher-side predictors								
4. ICT self-efficacy[a]	.32***	.23***	.03	–				
5. ICT learning opportunities[a]	.31***	.30***	.04	.60***	–			
6. Female[b]	–.01	–.05	–.02	–.07	–.02	–		
Quality of ICT use								
7. Organization of teaching[a]	.60***	.44***	–.04	.52***	.50***	.02	–	
8. Student learning behaviors[a]	.49***	.29***	.20***	.32***	.25***	.05	.55***	–
Range (min–max)	1–5	0–8	0–1	1–5	1–5	0–1	1–4	1–4
M	3.23	4.62	0.45	3.79	3.72	0.61	2.63	2.39
SD	0.88	2.09	0.50	0.78	0.85	0.49	0.65	0.63

Note. $N = 1512$. a = latent construct. b = manifest variable. ***$p < .001$. **$p < .01$. *$p < .05$.
Model fit: $\chi^2(276, N = 1512) = 8045.75$, *CFI* = .93, *SRMR* = .05, *RMSEA* = .04.

4.3. Predicting Quality of the Teaching and Learning Process with Digital Media During the Pandemic

The relations among study variables were examined by performing a series of SEM analyses with M*plus 8.8* (Muthén, & Muthén, 1998–2017). Overall, the SEM models exhibited acceptable fit to the data. Standardized parameter estimates for the associations among study variables are shown in Table 3. *In the first model*, we focused on *predictors on the school level*, including school ICT-related factors and school type (1 = Gymnasium; 0 = other school types). The results showed that teachers' perception regarding the sufficient availability of ICT equipment at school during the pandemic and teachers' perception of investments in the school's digital infrastructure were significantly positively associated with the two quality features of ICT use in teaching and learning in school during the pandemic. There was also a significant positive correlation between school type and student learning behaviors during the pandemic. *The second model* examined the relationships between *predictors on the teacher level* (i. e., teacher ICT-related factors, gender) and the quality of ICT use in teaching and learning in school. The results demonstrated that teachers' perceived ICT self-efficacy and teachers' perceived ICT learning opportunities were significantly positively associated with organization of

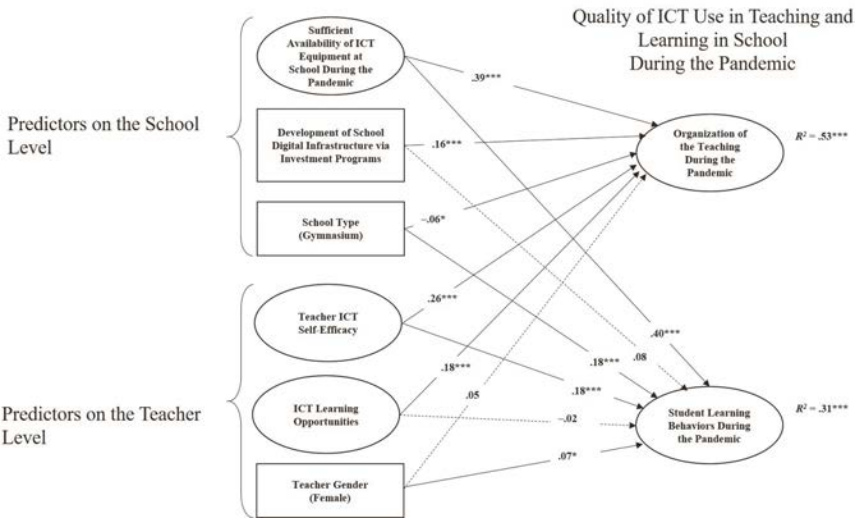
teaching during the pandemic. Perceived ICT self-efficacy and gender were significantly positively associated with student learning behaviours during the pandemic. *In the third model*, we included all six predictor constructs at the school and teacher levels *together in the analyses* (see Figure 2). *The final results* remained mostly unchanged. However, this time, teachers' perception of investment in the school's digital infrastructure was no longer significantly associated with student learning behaviors during the pandemic. However, the association between the Gymnasium school type and organization of teaching became significant. In terms of *R*-squared values (see Table 3), it must be noted that predictors on the school level (Model 1) explained more variance in the two quality features of ICT use in school during the pandemic than predictors on the teacher level (Model 2). Most importantly, the results showed that more variance in organization of teaching during the pandemic was explained in both models.

Table 3: Standardized Parameter Estimates for the Associations among Study Variables

Path coefficient	Organization of teaching during the pandemic		Student learning behaviors during the pandemic	
	Estimate	*SE*	Estimate	*SE*
Model 1: School level				
ICT equipment at school	.50***	.04	.44***	.05
Digital infrastructure investment	.22***	.03	.09*	.04
Gymnasium	−.05	.03	.18***	.03
Model 2: Teacher level				
ICT self-efficacy	.34***	.05	.25***	.06
ICT learning opportunities	.30***	.05	.09	.05
Female	.05	.03	.07*	.03
Model 3: School and teacher levels				
ICT equipment at school	.39***	.04	.40***	.06
Digital infrastructure investment	.16***	.03	.08	.04
Gymnasium	−.06*	.03	.18***	.03
ICT self-efficacy	.26***	.04	.18***	.06
ICT learning opportunities	.18***	.05	−.02	.06
Female	.05	.03	.07*	.03
R^2	Organization of teaching during the pandemic		Student learning behaviors during the pandemic	
	Estimate	*SE*	Estimate	*SE*
Model 1: School level	.40***	.04	.27***	.04
Model 2: Teacher level	.36***	.04	.10***	.03
Model 3: School and teacher levels	.53***	.03	.31***	.05
Model fit	Model 1	Model 2	Model 3	
χ^2	3534.90	6063.311	8027.36	
N	1512	1512	1512	
df	90	153	273	
CFI	.92	.91	.90	
SRMR	.04	.05	.05	
RMSEA	.05	.05	.05	

Note. $N = 1512$. ***$p < .001$. **$p < .01$. *$p < .05$.

Figure 2: SEM Model for Predicting the Quality of Teaching and Learning Process with Digital Media during the Pandemic. For simplicity's sake, factor loadings of the latent constructs are not shown. Estimates for correlations among study variables are presented in Table 2.



Note. *p < .05. **p < .01. ***p < .001.

5. Discussion

In the media education literature, there is strong evidence that ICT-related factors and other relevant factors contribute to predicting the quantity (*how often*) of teachers' ICT use in school. However, predictors of the quality (*how and in what way*) of ICT use have been investigated to a far lesser extent. Most importantly, relatively little is known whether these predictive factors could reliably explain ICT use in school in the context of the COVID-19 pandemic. Inspired by the theoretical framework of quality dimensions of teaching and learning with digital media (Lorenz & Bos, 2017), the present study examined predictive factors that may explain different qualities of ICT use in teaching and learning in school during the COVID-19 pandemic. In addition, we exploratorily examined which group of predictive factors (school level versus teacher level) was able to better explain the quality of ICT use. A total of 1,512 secondary school teachers in Germany participated in the online survey. We assessed two aspects of the quality of ICT use in teaching and learning in school during the pandemic. *The first aspect* assessed how teaching with digital media was organized in school during the pandemic, while *the second aspect* assessed teachers' perceptions of student learning behaviors during the

pandemic. To the best of our knowledge, the present study was the first of its kind to focus on these two quality dimensions of ICT use in school during the pandemic. We performed a series of SEM analyses to examine the associations between the predictive factors and two quality dimensions of ICT use in teaching and learning in school during the pandemic. Overall, the results highlighted how predictors at both the teacher and school level promoted the quality of ICT use under the pandemic context.

5.1. Predictors of the Quality of ICT Use

In line with our predictions, the overall results add to the literature on media education in school for two *major ways.* First, the results confirmed that the six predictor constructs (see Figure 2) derived from the theoretical framework and previous empirical research (e. g., Gerick et al., 2017; Hatlevik & Hatlevik, 2018; Lorenz et al., 2019; OECD, 2020) retained their significance in the pandemic context. Second, these selected predictive factors were able to effectively predict different qualities of ICT use in teaching and learning in school during the pandemic.

On the school level, the results suggest that teachers who reported a more sufficient availability of ICT equipment at school during the pandemic and perceived greater investment in their school's digital infrastructure in the past two years tended to report better organization of teaching in school during the pandemic as well (e. g., digitally supported lessons built on important preliminary work at school, sharing experiences with and knowledge about teaching with digital media with their fellow teachers). This indicates that ICT equipment and digital infrastructure in school increase teachers' ability to organize ICT-related activities with respect to their instruction. In the same vein, teachers who reported a more sufficient availability of ICT equipment at their school during the pandemic tended to more strongly agree that student learning behaviors during the pandemic did not differ between face-to-face and distance learning classes. This indicates that, from teachers' point of view, sufficient ICT equipment in school was essential for ensuring the quality of student learning with digital media during the pandemic. Compared to teachers from other secondary school types, gymnasium teachers tended to report worse organization of teaching during the pandemic, but greater consistency in student learning behaviors between face-to-face and digital learning during the pandemic. These findings expand upon previous studies finding that gymnasium teachers reported greater quantitative use of ICT than teachers in other school types

(e. g., Lorenz et al., 2017). In terms of the quality of ICT use, however, it seems that gymnasium teachers are not convinced of the quality of student learning behaviors with digital media during the pandemic. However, this is not surprising given that gymnasium is *the highest school track in Germany*, involving demanding, intensive instruction to prepare students for higher education (Müller & Schneider, 2013). However, student learning behaviours during the pandemic did not seem to depend on the school's digital infrastructure investments. An explanation for this unexpected finding might be that, in teachers' view, ICT equipment and digital infrastructure in school should affect teachers' use of ICT in their teaching activities rather than students' learning behaviour at home.

On the level of individual teachers, the results suggest that teachers who reported greater ICT self-efficacy and greater ICT learning opportunities tended to report better organization of teaching during the pandemic as well. These results expand previous research in the sense that teachers' competences as well as staff development (e. g., teachers' experiences with ICT-related content from teacher education or professional development courses) not only increase teachers' readiness to use ICT in school more often (e. g., Gerick et al., 2017; Lorenz et al., 2019), but also facilitated greater use of ICT in organizing their teaching during the pandemic. However, organization of teaching did not depend on teacher gender. Teachers who reported greater ICT self-efficacy tended to report greater consistency in student learning behaviors during the pandemic. This indicates that teachers believe their ICT-related skills matter for their students' learning behaviors during the pandemic. Likewise, female teachers tended to report greater consistency in student learning behaviors during the pandemic. Thus, gender should always be taken into consideration when practically implementing the results in schools.

The current research also suggests that predictors *on the school level* explain more variance in both quality features of ICT use in teaching and learning in school during the pandemic. This indicates that school-side factors such as sufficient ICT equipment and digital infrastructure are fundamental for ICT use in school. Taking into account both school- and teacher-level predictors at the same time, the results show that these predictors are more like to explain more variance in organization of teaching during the pandemic than in student learning behaviors during the pandemic. An explanation for these findings might be that we assessed student learning behaviors during the pandemic via teachers' perceptions. In order to more precisely examine the associations between predictors and student learning

behaviors, student-report data on learning behaviors during the pandemic should be included in the model.

5.2. Limitations and Recommendations for Future Research

Although the results of this paper add to the literature, several limitations must also be noted. *First*, the present study employed a cross-sectional research design. Therefore, it is still too early to draw causal conclusions about predictors of the quality of ICT use in teaching and learning in school during the pandemic. Consequently, our working model should be replicated in longitudinal research in order to examine long-term effects of the predictors on the quality of ICT use in teaching and learning in the context of distance or hybrid education. *Second*, we found that the teacher participants reported high levels of ICT self-efficacy (TPACK) on average. When interpreting the results, it was therefore not possible to clearly differentiate between good and very high ICT self-efficacy. Future research may want to integrate students' perceptions of teachers' ICT competences into the working model or even assessment results regarding teachers' competences in using ICT in teaching. The *third limitation* concerned the measure of ICT equipment at school. We did not focus on the amount of ICT equipment, but rather on whether (or not) the ICT equipment at school was considered sufficient by individual teachers. Likewise, we asked teachers to retrospectively estimate whether (or not) they had witnessed investments in the school's digital infrastructure in different areas in the past two years. It might be difficult for teachers to retrospectively rate some areas of investment such as fiber optic/broadband expansion, licensing of commercial online educational media, and stronger anchoring of media education in the school curriculum. In future research, there is a need to optimize assessment of this construct.

Regarding the future recommendations, further studies may want to include further key factors proposed by Lorenz and Bos (2017), such as teachers' attitudes towards media education in school, teachers' collaboration on ICT-related activities, or ICT-related pedagogical support. Including these factors in the working model will help to better identify predictors of the quality of ICT use. Furthermore, it must be noted psychological factors such as teachers' motivation are not included in the theoretical model by Lorenz and Bos (2017) as key factors for processes of teaching and learning with digital media. It might be interesting to take aspects of teachers' motivation into consideration. In this article, we were not able to include students' ICT competences as ultimate outcomes of the quality of ICT use in teaching and

learning in school. This aspect should be definitely taken into consideration in the future.

5.3. Conclusion and Implications

In the wake of the COVID-19 pandemic, schools worldwide should prepare for a new remote emergency way of teaching and learning that can take place either in school or at home. For this reason, *media education in school* has become even more significant than during the pre-pandemic era. The results of the current representative study lead to several recommendations for ensuring the quality of ICT use in teaching and learning in school during pandemic conditions. In the theoretical framework grounding the present study, Lorenz and Bos (2017) postulates that quality dimensions at different levels – ICT equipment, infrastructure, school development, personal development – are actually intercorrelated. To implement this framework – and more importantly, the results of the present study – in practice, all key factors should be implemented *at the same time*. First, to lay a strong foundation for media education in school, it is crucially important to provide sufficient digital infrastructure in schools – of all school types. The enormous funding available as part of the *DigitalPakt schule* (e. g., BMBF, 2021) can support this in every German federal state. Secondly, sufficient digital infrastructure or ICT equipment in schools may not matter if it are not used effectively. Thus, teachers require adequate ICT-related skills in order to use ICT in teaching and learning situations in schools more effectively. Most importantly, apart from ICT use in teaching, the ICT-related content addressed in professional development courses for teachers should focus on how to ensure better-quality ICT use in distance learning (e. g., how to ensure that the learning process is equivalent between on-site and distance classes, how to ensure that students' ICT competences are sufficient to succeed in distance learning).

References

Bird, K. A., Castleman, B. L., & Lohner, G. (2022). Negative impacts from the shift to online learning during the COVID-19 crisis: Evidence from a statewide community college system. *AERA Open, 8*(1), 1–16. https://doi.org/10.1177%2F23328584221081220

BMBF [Bundesministerium für Bildung und Forschung]. (2021). *DigitalPakt Schule*. https://www.digitalpaktschule.de/de/was-ist-der-digitalpakt-schule-1701.html

Brummelhuis, A., & Kuiper, E. (2008) Driving forces for ICT in learning. In: Voogt J. & Knezek G. (eds) *International Handbook of Information Technology in Primary and Secondary Education. Springer International Handbook of Information Technology in Primary and Secondary Education*, vol 20. Springer. https://doi.org/10.1007/978-0-387-73315-9_6

Creemers, B. P. M., & Kyriakides, L. (2008). The dynamics of educational effectiveness. In *A contribution to policy, practice and theory in contemporary schools*. Routledge. https://doi.org/10.4324/9780203939185

Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik, Beiheft 41*, 73–92. https://doi.org/10.25656/01:8486

Drossel, K., Eickelmann, B., Schaumburg, H., & Labusch, A. (2019). Nutzung digitaler Medien und Prädiktoren aus der Perspektive der Lehrerinnen und Lehrer im internationalen Vergleich. In B. Eickelmann, W. Bos, J. Gerick, F. Goldhammer, H. Schaumburg, K. Schwippert, M. Senkbeil, & J. Vahrenhold (Hrsg.), *ICILS 2018 #Deutschland – Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (S. 205–240). Waxmann. https://doi.org/10.25656/01:18325

Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (Hrsg.). (2019). *ICILS 2018 #Deutschland: Computer und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking.* Waxmann. https://doi.org/10.25656/01:18166

Eickelmann, B., & Drossel, K. (2020). *Schule auf Distanz. Perspektiven und Empfehlungen für den Schulalltag. Eine repräsentative Befragung von Lehrkräften in Deutschland*. Vodafone Stiftung.

Endberg, M., & Lorenz, R. (2016). Selbsteinschätzung medienbezogener Kompetenzen von Lehrkräften in Deutschland und im Bundesländervergleich. In W. Bos, R. Lorenz, M. Endberg, B. Eickelmann, R. Kammerl, & S. Welling (Hrsg.), *Schule digital – der Länderindikator 2016. Kompetenzen von Lehrpersonen der Sekundarstufe I im Umgang mit digitalen Medien im Bundesländervergleich* (S. 180–208). Waxmann. https://doi.org/10.25656/01:15134

Eickelmann, B., & Schulz-Zander, R. (2008). Schuleffektivität, Schulentwicklung und digitale Medien. In W. Bos, H. G. Holtappels, H. Pfeiffer, H.-G. Rolff, & R. Schulz-Zander (Eds.), *Jahrbuch der Schulentwicklung* (Vol. 15, pp. 157–193). Juventa.

forsa [forsa Politik- und Sozialforschung GmbH]. (2019). *Die Schule aus Sicht der Schulleiterinnen und Schulleiter – Digitalisierung und digitale Ausstattung. Ergebnisse einer bundesweiten Repräsentativbefragung.* Verband Bildung und Erziehung. https://www.vbe.de/fileadmin/user_upload/VBE/Service/Meinungsumfragen/2019-04-17_forsa-Bericht_SL_Digitalisierung_Bund.pdf

Gerick, J., Eickelmann, B., & Bos, W. (2017). School-level predictors for the use of ICT in schools and students' CIL in international comparision. *Large-scale Assessments in Education, 5*(5), 1–13. https://doi.org/10.1186/s40536-017-0037-7

Gerick, J., Eickelmann, B., & Labusch, A. (2019). Schulische Prozesse als Lern- und Lehrbedingungen in den ICILS-2018-Teilnehmerländern. In B. Eickelmann, W. Bos, J. Gerick, F. Goldhammer, H. Schaumburg, K. Schwippert, M. Senkbeil, & J. Vahrenhold (Hrsg.), *ICILS 2018 #Deutschland: Computer- und informationsbezogene Kompetenzen von Schüle r innen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (S. 173–204). Waxmann. https://doi.org/10.25656/01:18324

GEW [Gewerkschaft Erziehung und Wissenschaft]. (2020). *Digitalpakt Schule und Digitalisierung an Schulen. Ergebnisse einer GEW-Mitgliederbefragung 2020.* https://www.gew.de/fileadmin/media/publikationen/hv/Bildung-digital/202004-Mitgliederbefr-Digitalisierung.pdf

Gräsel, C., Schledjewski, J., & Hartmann, U. (2020). Implementation digitaler Medien als Schulentwicklungsaufgabe. *Zeitschrift für Pädagogik, 66*, 208–224. https://doi.org/10.25656/01:23629

Hatlevik, I. K. R., & Hatlevik, O. E. (2018). Examining the relationship between teachers' ICT self-efficacy for educational purposes, collegial collaboration, lack of facilitation and the use of ICT in teaching practice. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2018.00935

Initiative D21 e. V. (Hrsg.). (2021). *D21 Digital Index 2020/2021: Jährliches Lagebild zur Digitalen Gesellschaft*. https://initiative21.de/d21index/

Katz, Y. J. (2018). The Interaction of Psychological Constructs with Information Technology-Enhanced Teaching and Learning. In J. Voogt, G. Knezek, R. Christensen, & K. Lai (Eds.), *Second Handbook of Information Technology in Primary and Secondary Education* (pp. 293–307). Springer. https://doi.org/10.1007/978-3-319-71054-9_24

KMK. (2017). *Bildung in der digitalen Welt. Strategie der Kultusministerkonferenz: Beschluss der KMK vom 08.12.2016 i. d. F. vom 07.12.2017.* https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2016/2016_12_08-Bildung-in-der-digitalen-Welt.pdf

KMK. (2021). *Lehren und Lernen in der digitalen Welt – Ergänzung zur Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt": Beschluss der Kultusministerkonferenz vom 09.12.2021.* https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_12_09-Lehren-und-Lernen-Digi.pdf

Lorenz, R., Bos, W., Endberg, M., Eickelmann, B., Grafe, S., & Vahrenhold, J. (Hrsg.) (2017). *Schule digital – der Länderindikator 2017. Schulische Medienbildung in der Sekundarstufe I mit besonderem Fokus auf MINT-Fächer im Bundesländervergleich und Trends von 2015 bis 2017.* Waxmann. https://doi.org/10.25656/01:15656

Lorenz, R., & Bos, W. (2017). Schule digital – der Länderindikator 2017. Theoretisches Rahmenmodell, Überblick über die Befunde des Länderindikators 2017 und Einordnung zentraler Ergebnisse der Erhebungszyk-len 2015, 2016 und 2017. In R. Lorenz, W. Bos, M. Endberg, B. Eickelmann, S. Grafe, & J. Vahrenhold (Eds.), *Schule digital – der Länderindikator 2017. Schulische Medienbildung in der Sekundarstufe I im Bundesländervergleich und Trends von 2015 bis 2017* (pp. 11–35). Waxmann. https://doi.org/10.25656/01:15656

Lorenz, R., Bos, W., Endberg, M., Eickelmann, B., Grafe, S., & Vahrenhold, J. (Hrsg.). (2017). *Schule digital – der Länderindikator 2017. Schulische Medienbildung in der Sekundarstufe I mit besonderem Fokus auf MINT-Fächer im Bundesländervergleich und Trends von 2015 bis 2017.* Waxmann. https://doi.org/10.25656/01:15656

Lorenz, R. & Endberg, M. (2017). IT-Ausstattung der Schulen der Sekundarstufe I im Bundesländervergleich und im Trend von 2015 bis 2017. In R. Lorenz, W. Bos, M. Endberg, B. Eickelmann, S. Grafe, & J. Vahrenhold (Hrsg.), *Schulische Medienbildung in der Sekundarstufe I mit besonderem Fokus auf MINT-Fächer im Bundesländervergleich und Trends von 2015 bis 2017* (S. 49–83). Waxmann. https://doi.org/10.25656/01:15656

Lorenz, R., Endberg, M., & Bos, W. (2019). Predictors of fostering students' computer and information literacy – Analysis based on a representative sample of secondary school teachers in Germany. *Education and Information Technologies*. https://doi.org/10.1007/s10639-018-9809-0

Lorenz, R., & Eickelmann, B. (2022). Verlängerung durch die Corona-Pandemie mit Blick auf die Nutzung digitaler Medien in der Sekundarstufe I im Bundesländervergleich – technische Ausstattung, Unterrichtsorganisation und Leistungsbewertung. In R. Lorenz, S. Yotyodying, B. Eickelmann, & M. Endberg (Hrsg.), *Schule digital – der Länderindikator 2021 – Lehren und Lernen mit digitalen Medien in der Sekundarstufe I in Deutschland im Bundesländervergleich und im Trend seit 2017* (S. 161–186). Waxmann. https://doi.org/10.25656/01:24606

Lorenz, R., Yotyodying, S., Eickelmann, B., & Endberg, M., (Hrsg.) (2022). *Schule digital – Länderindikator 2021. Lehren und Lernen mit digitalen Medien in der Sekundarstufe I in Deutschland im Bundesländervergleich und im Trend seit 2017.* Waxmann. https://doi.org/10.31244/9783830995449

Ludewig, U., Kleinkorres, R., Schaufelberger, R., Schlitter, T., Lorenz, R., König, C., Frey, A., & McElvany, N. (2022). COVID-19 pandemic and student reading achievement: Findings from a school panel study. *Frontiers in Psychology.* 13:876485. https://doi.org/10.3389/fpsyg.2022.876485

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A new framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Müller, S., & Schneider, T. (2013). Educational pathways and dropout from higher education in Germany. *Longitudinal and Life Course Studie 2013, 4*(3), 218–241. http://dx.doi.org/10.14301/llcs.v4i3.251

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide*. Muthén & Muthén.

OCED (2020). *Using ICT to support teachers' professional learning.* https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/SR/RD(2020)6&docLanguage=En

Reintjes, C., Porsch, R., & Brahm, G. (Hrsg.). (2021). *Das Bildungssystem in Zeiten der Krise. Empirische Befunde, Konsequenzen und Potenziale für das Lehren und Lernen*. Waxmann. https://doi.org/10.25656/01:22798

Scheerens, J. (1990). School effectiveness and the development of process indicators of school functioning. In B. P. M. Creemers, T. Peters, & D. Reynolds (Hrsg.), *School effectiveness and school improvement, 1,* 61–80. https://doi.org/10.1080/0924345900010106

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research, 99*(6), 323–337. https://doi.org/10.3200/JOER.99.6.323-338

UNICEF. (2020). *The impact of COVID-19 on the mental health of adolescents and youth*. https://www.unicef.org/lac/en/impact-covid-19-mental-health-adolescents-and-youth

Van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2020). Determinants of 21st-Century skills and 21st-century digital skills for workers: A systematic literature review. *SAGE Open*, 1–14. https://doi.org/10.1177/2158244019900176

Yotyodying, S., Eickelmann, B., Endberg, M., & Lorenz, R. (2022). Entwicklungen der schulischen digitalen Infrastruktur für das Lehren und Lernen durch Investitionsprogramme aus Sicht von Lehrkräften der Sekundarstufe I. In R. Lorenz, S. Yotyodying, B. Eickelmann, & M. Endberg (Hrsg.), *Schule digital – der Länderindikator 2021 – Lehren und Lernen mit digitalen Medien in der Sekundarstufe I in Deutschland im Bundesländervergleich und im Trend seit 2017* (S. 145–160). Waxmann. https://doi.org/10.25656/01:24606

Towards the Measurability of a Competence Model for School-based Civic Education

Pascal Alscher, Alyssa Grecu, Ulrich Ludewig, Nele McElvany

Abstract

The adoption of national educational standards in Germany in the early 2000s represented a paradigm shift away from input-driven to output-driven educational practice and research. However, no standards were adopted for civic education. Against this background, we conceptualized and operationalized a civic literacy competence model, which includes a political knowledge test as well as further dispositions in the areas of motivations, attitudes and volitions. With this study, we examine the extent to which the factorial structure of the model is supported by empirical data. In addition, measurement invariance, developmental sensitivity, and convergent and discriminant associations of the instruments were examined. Based on data from 1,047 German students attending 7th (n = 613, M_{age} = 12.47; 47.4 % female; 44.4 % immigrant background) or 10th grade (n = 434, M_{age} = 15.49; 47.7 % female; 35.9 % immigrant background), we calculated confirmatory factor analyses and mean comparisons. Furthermore, we specified and tested a nomological network. The results show that most of the theoretical assumptions are supported by the data. For political efficacy, a distinction should be made between internal and external political efficacy. In addition, the data suggest that a common factor should be assumed for the willingness to politically and the willingness to civically participate scales. The application of the civic literacy framework shall provide important insights for the research and practice of (school-based) civic education.

Keywords: civic education; civic literacy; educational standards; confirmatory factor analysis; measurement invariance

Zusammenfassung

Die Verabschiedung nationaler Bildungsstandards in Deutschland in den frühen 2000er Jahren bedeutete einen Paradigmenwechsel weg von einer inputorientierten hin zu einer outputorientierten Bildungspraxis und -forschung. Für den Politikunterricht wurden jedoch keine Standards verabschiedet. Vor diesem Hintergrund wurde ein Kompetenzmodell zur politischen Bildung konzeptualisiert und operationalisiert, das neben einem Test zum politischen Wissen auch weitere Dispositionen in den Bereichen Motivationen, Einstellungen und Volitionen umfasst. In dieser Studie wird untersucht, inwieweit die faktorielle Struktur des Modells durch empirische Befunde gestützt wird. Darüber hinaus wurden Messinvarianz, Entwicklungssensitivität sowie konvergente und diskriminante Assoziationen der Instrumente untersucht. Basierend auf den Daten von 1.047 deutschen Schülerinnen und Schülern der 7. (n = 613, M_{age} = 12,47; 47,4 % weiblich; 44,4 % Migrationshintergrund) bzw. 10. Klasse (n = 434, M_{age} = 15,49; 47,7 % weiblich; 35,9 % Migrationshintergrund) wurden konfirmatorische Faktorenanalysen und Mittelwertvergleiche gerechnet. Darüber hinaus wurde ein nomologisches Netzwerk spezifiziert und getestet. Die Ergebnisse zeigen, dass die meisten der theoretischen Annahmen durch die Daten gestützt werden. Beim politischen Effektivitätsgefühl sollte zwischen internem und externem politischen Effektivitätsgefühl unterschieden werden. Außerdem legen die Daten nahe, dass ein gemeinsamer Faktor für die Bereitschaft zur politischen und zivilgesellschaftlichen Partizipation angenommen werden sollte. Die Anwendung des Civic Literacy Frameworks soll wichtige Erkenntnisse für die Forschung und Praxis der (schulischen) politischen Bildung liefern.

Schlagworte: Politische Bildung; Civic Literacy; Bildungsstandards; Konfirmatorische Faktorenanalyse; Messinavarianz

1. Introduction

In 2001, the results of the first round of the Programme for International Student Assessment (PISA) were published (OECD, 2001). In Germany, the publication of the results was accompanied by a public outrage about the poor academic performance of German students (Lange, 2002). For weeks, Germany's disappointing ranking dominated the headlines and public discussion. While the poor performance in the Trends in International Mathematics and Science Study (TIMSS) four years earlier already revealed initial deficits, the

debate on the perceived qualitative decline of the German educational system and the demise of German students' international competitiveness reached new heights in the winter of 2001/02 (Waldow, 2009). In the post-'Pisa-shock' period, a variety of reforms were initiated, such as a massive increase in educational spending combined with better access to early childhood education, greater support for children from socioeconomically disadvantaged families, and an increase in the number of schools that offer all-day schooling (Ertl, 2006). Another reform of the post Pisa-shock period was the introduction of national educational standards (*nationale Bildungsstandards*) by the conference of ministers of education (*Kultusministerkonferenz*, KMK) for a number of school subjects (Klieme et al., 2003). Educational standards were introduced for the subjects Mathematics, German, the first foreign language, biology, chemistry and physics. Critics complained that this would ennoble these subjects as core subjects (*Kernfächer*) and declare other subjects unimportant (Köller, 2016). It was thus up to the accompanying didactic research communities of the remaining subjects to formulate educational standards.

In Germany, the Society for Civic Education Didactics and Civic Youth and Adult Education (*Gesellschaft für Politikdidaktik und politische Jugend- und Erwachsenenbildung*, GPJE) presented a proposal for national educational standards in school-based civic education (GPJE, 2004). The proposal subsequently served as the basis for further conceptualizations of educational standards and competency models in civic education in the years after the Pisa-shock, such as May's model for democratic competence (2007). However, the most well known competency model of school-based civic education in Germany is the model of political competence (*Modell der Politikkompetenz*) by Detjen et al. (2012). It is a further development of the GPJE model (2004) and the standards of political knowledge by Weißeno et al. (2010). The model is based on a relatively broad educational-psychological concept of competence (see Weinert, 2001), which includes cognitive as well as affective, motivational, and volitional dispositions. Thus, political knowledge, motivations and attitudes as well as the abilities to make political judgments and to politically participate are dimensions of the model of political competence. The model has made a decisive contribution to the shift from normative to theory-driven and output-oriented didactics in civic education research (e. g., Massing, 2022). Despite a broad agreement within political didactics, the model has so far only been partially empirically tested. One reason for this is the operationalizability and hence the operationalization of the two dimensions 'ability to make political judgments' and 'ability to politically participate'. For the few existing operationalizations, a conflation with political knowledge can be observed (see May et al., 2020). The future task of

civic education research must therefore be to construct competence models that connect the normative ideas of the field with objective and empirical evidence.

While there are standardized tests for the core subjects in Germany for verifying compliance with the educational standards, empirical evidence on the effectiveness of civic education with regard to the acquisition of competence falls short of expectations (e. g., Hedinger, 2018). The civic literacy framework presented in this contribution offers a way to comprehensively and systematically consider competence in the context of civic education (Alscher et al., 2022b). It therefore aims to advance educational monitoring and evaluation in order to benefit civic education research and practice. In the framework, civic literacy is defined as (1) the knowledge and skills to make informed decisions regarding socially and politically relevant issues, and (2) the dispositions for action (i. e., motivations, attitudes and volitions) that are necessary to implement the decisions made. Consequently, the framework includes the four main areas political knowledge, motivations, attitudes, and volitions. In the future, the framework will be the basis for civics assessments in the National Educational Panel Study (*Nationales Bildungspanel*, NEPS), which aims to collect longitudinal data on competence development, educational processes, educational decisions, and educational returns in Germany.

Founded in 2009, the NEPS is a multilocational network of educational research institutions in Germany aiming to provide longitudinal data covering formal, non-formal and informal educational contexts across the whole life course (Blossfeld, Maurice & Schneider, 2019) (NEPS, 2022). Doing so, NEPS is organized in eight pillars covering the different stages of learning in the life course. Major fields of interest are the development of competencies, educational decisions, learning environments, returns to education considering the effects of social inequalities and immigration backgrounds as well as educational transitions. Civic Literacy is one main focus of pillar 4 which centers on lower secondary school as a critical bridge between elementary school and either the general educational or vocational higher secondary domain (Fabian et al., 2019).

The surveys conducted as part of the second version of the Start Cohort Grade 5 (SC 8) of the NEPS will enable scientists, practitioners and policy makers to follow competency trajectories longitudinally throughout lower secondary education and to gain an impression of prognostic conditional relationships.

1.1. Dimensions of Civic Literacy

Important characteristics of competencies are that they are domain-specific and learnable, and that their structure is based on the situation to which the competency is relevant (Hartig & Klieme, 2006). In German national empirical educational research, the model of political competence (Detjen et al., 2012) dominates the discourse. Internationally, the frameworks of the International Civic and Citizenship Education Study (ICCS: e. g., Schulz et al., 2022) and the framework of the Civics Assessment of the National Assessment of Educational Progress (Civic Asssessment of the NAEP: e. g., National Assessment Governing Board, 2018) are of particular importance.

In ICCS 2022, political knowledge is the main aspect of the cognitive domains, which form the Civic Knowledge Framework. In addition, there are the two affective-behavioral domains Attitudes and Engagement that form the Civic Attitudes and Engagement Framework. One of the benefits of the frameworks in the ICCS is that the content is fully operationalized and tested. However, an unavoidable disadvantage of such a large international comparative study is that no reference can be made to the peculiarities of national political systems. Furthermore, assessments in ICCS 2022 are based on two different instead of one consolidated framework.

For the Civics Assessment of the NAEP, the three dimensions (1) knowledge, (2) skills, and (3) dispositions are distinguished (National Assessment Governing Board 2018). The different dimensions are assessed together across different task formats (e. g., multiple choice, short- or long-answer questions). Each of the tasks represents two to three dimensions. The results are combined to an overall score. The determination of an overall score ensures an intuitive and simple interpretation of the results. However, with only an overall score there is no information about the performance in the specific dimensions of the framework.

Using data from ICCS 2009, Hoskins et al. (2015) created a civic competence composite indicator that includes the dimensions knowledge and skills for democracy, social justice values, participatory attitudes and citizenship values. Their competence composite indicator allows calculating and evaluating an overall score as well as the performance in the specific dimensions of the indicator.

While there seems to be broad agreement among scholars that civic literacy is a multidimensional concept, different frameworks and models were developed leading to the investigation of different dimensions (Wray-Lake et al., 2017). One of the differences, for example, is that the abilities to make political judgments and to politically participate from the German political competence model do not seem to play a role in the international frameworks

and models. This is mainly because civic literacy as such is understood as a precondition for engagement (Hoskins et al., 2015) and thus interferes with the concept of ability to politically participate and, to a lesser extent, with the ability to make political judgments. What most conceptualizations have in common, however, is political knowledge, political motivations, political attitudes, and political volitions. These are also the four dimensions of our civic literacy framework (see figure 1). Furthermore, the framework consists of facets. Two facets represent one dimension of the framework. This does not apply to the dimension of political knowledge, which has no separate facets. In contrast to educational achievement tests, the civic literacy framework is not referenced to curricula or instructional objectives (see Messick, 1984, p. 217). The framework shall provide an opportunity to measure and evaluate students' competence in the civic domain.

Political Knowledge

Political knowledge is usually considered a major tool for understanding the political world and everything related to it. It plays a vital role in the perception of the democratic system as well as current political events and is considered a desirable democratic outcome affecting voting behavior (Mondak, 2001). A widely accepted definition of political knowledge is that it encompasses "the range of factual information about politics that is stored in long-term memory" (Delli Carpini & Keeter, 1996, p. 10). As such, political knowledge is an important prerequisite for understanding political discourses and being able to classify their relevance, to process political information, and to gain an understanding about one's own political preferences (Cramer & Toff, 2017). Research finds that knowledgeable citizens engage less often in unlawful political protest activities (Gil de Zúñiga & Goyanes, 2021) and are more likely to cast their vote (Gil de Zúñiga & Diehl, 2018). However, knowledgeable citizens also have a higher tendency for opinion polarization (Herne et al., 2019) and are more likely to vote for populist parties (Stanley & Cześnik, 2022). Based on data from ICCS 2016, it was found that there are significant, positive correlations between students' political knowledge and their willingness to participate in politics in all 24 participating countries (Schulz et al., 2018). The same finding was reported for 7th and 10th grade students from North Rhine-Westphalia in Germany (Alscher et al., 2022a). A recent systematic review of research articles using the ICCS data shows that the cognitive domain has so far received comparably little attention claiming that more in-depth analyses of political knowledge are needed (Myoung & Liou, 2022).

Political Motivations

Empirical evidence informed by the expectancy-value theory (EVT; Steinmayr et al., 2019; Wigfield & Eccles, 2000) suggests that individuals are motivated to engage in activities based on (1) the expectation of success and (2) the value they assign to the activity. In the past, EVT has proven to be a valuable resource in explaining students' effort (Dietrich et al., 2017), career choices (Gaspard et al., 2019), achievement (Lauermann et al., 2017; Meyer et al., 2019) and attitudes – for instance towards future political participation (Levy & Akiva, 2019) or attributes of a 'good' citizen (Liem & Chua, 2013). The expectation of success can be described as one's belief to accomplish something in the future she or he set out to achieve. The difference towards one's self-efficacy, namely one's belief to be *capable* to accomplish something, is very small in civic education research (Levy & Akiva, 2019). The assigned value of an activity is influenced by intrinsic value, attainment value, utility value, and cost (Eccles & Wigfield, 2002). Intrinsic value refers to the experienced enjoyment of an activity and is conceptually closely related to an individual's interest (Wigfield & Eccles, 2000). In civic education research, both political interest and the feeling of political effectiveness are important concepts and powerful predictors of desirable civic outcomes (Rasmussen & Nørgaard, 2018).

Self-efficacy can be understood as a person's conviction that he or she can cope with challenging situations (Bandura, 1977). In civic education research, political efficacy describes the extent to which an individual feels capable of influencing political debates and agenda setting, political decision-making, and political officials (Rasmussen & Nørgaard, 2018). Scholars distinguish between internal and external political efficacy. While internal political efficacy expresses one's own ability to influence the formation of political will, external political efficacy describes the perceived general responsiveness or collective suggestibility of the political system (e. g., Oberle & Wenzel, 2019). Indeed, empirical studies show that political effectiveness is related to expected future political participation (Levy & Akiva, 2019) and actual political participation (McDonnell, 2020).

Political interest is an attitude and thus analytically and empirically to be distinguished from behavior. It can be defined as the extent to which politics arouses a citizen's curiosity (van Deth, 1990). Empirical evidence suggests that political interest appears to be well approximated as a one-dimensional concept with a single underlying dimension capturing most inter-individual differences even when considering different aspects of politics (Prior, 2019). Similar to political efficacy, political interest has proven to be an important prerequisite for expected future political participation (Alscher et al., 2022a)

and actual political participation (Prior & Bougher, 2018; Stattin & Amnå, 2022).

Political Attitudes

Attitudes are mental tendencies expressed by valuing a certain object with a certain degree of affection or rejection (Eagly & Chaiken, 1993). Political attitudes can be understood as acquired, enduring, often more strongly emotional orientations toward classes of persons or objects with a connection to political issues (Detjen et al., 2012). Attitudes toward (state-relevant) objects can be represented by an individual's opinion about democracy while attitudes toward (social groups or classes of) people can be observed through an individual's opinion regarding pluralism (Alscher et al., 2022b). For the opinion about democracy, a distinction must be made between the general agreement with the idea of democracy and its constituent principles on the one hand and the evaluation of the current state of democracy in a particular country and time period on the other hand (Gabriel, 2000).

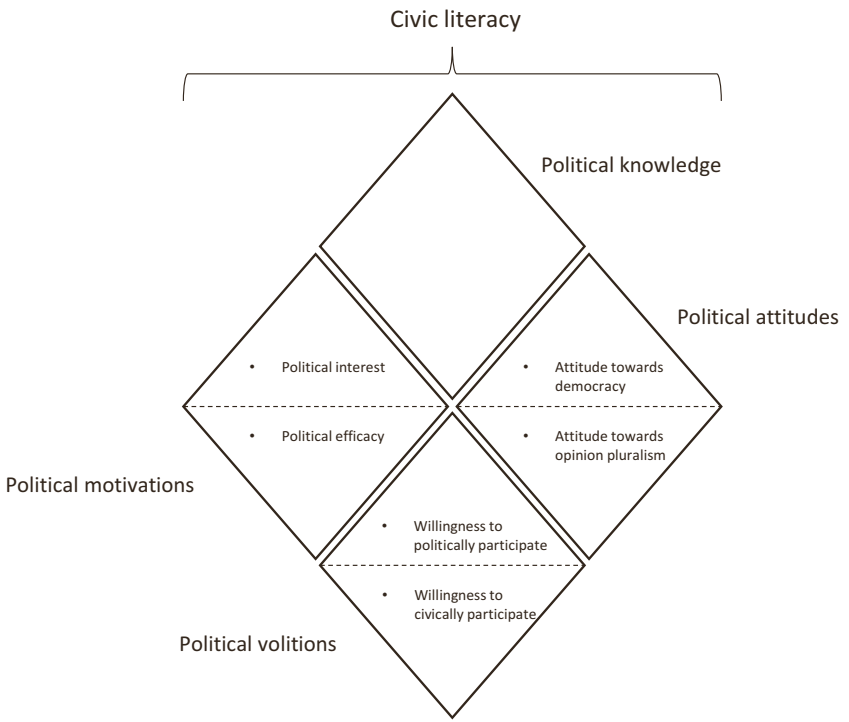
According to ICCS 2009 and 2016 data, adolescents from countries with more inclusive and multiculturalist policies showed more positive attitudes towards ethnic minorities than adolescents from countries with less inclusive and multiculturalist policies (Kim & Byun, 2019). With the exception of a shift towards more inclusive attitudes in terms of racial and gender equality, political attitudes remained rather similar between the 1999 and 2009 CIVED/ICCS cohorts (Barber & Ross, 2019). Furthermore, ICCS 2009 data shows that the introduction of inequality on the basis of achievement through academic segregation impedes positive attitudes towards social cohesion and other democratic values (Kavadias et al., 2017)

Political volitions

Volitions are cognitive processes that are related to the concrete realization of action goals in the context of the preactional and actional phases of action (Achtziger & Gollwitzer, 2010). In general, intentions are considered immediate antecedents of actual behavior (Ajzen, 1991). In civic education research, too, political volitions are viewed as the closest proxy indicator for actual political behavior (Quintelier & Hooghe, 2013). Political volitions represent an individual's willingness to participate in political or civic life including activities such as voting, protesting or joining a political party. Looking at youth's intentions rather than their actual participation is the only way to get a better understanding of minors' future participation. This especially includes

activities that are either not available to them (e. g., voting), for which they need support (e. g., money from parents to be able to donate) or which are very unusual at a young age (e. g., joining a political party) (see McWhirter & McWhirter, 2016). From a theoretical point of view, participation activities can be divided into more civically oriented or latent political participation, and manifest political participation (Ekman & Amnå, 2012). Empirical evidence from the U.S and Italy indeed suggests a strong relationship between voting intentions and actual voting behavior (Ajzen et al., 1982; Roccato & Zogmaister, 2010). Furthermore, CIVED 1999 data suggests that school council involvement and volunteering are associated with greater expectations to become an informed voter and an active citizen (Homana, 2018).

Figure 1: Civic Literacy Framework.



Note. From „Civic literacy — about the theory and measurability of competence in school-based civic education" by Alscher et al., 2022, Zeitschrift für Erziehungswissenschaft (https://doi.org/10.1007/s11618-02201085-0). CC BY 4.0.

2. Current Study

The development of a competency model that enables the extensive assessment and evaluation of students' civic literacy shall make an important contribution to the optimization of civic education and its teaching practices. So far, there are hardly any fully operationalized civic literacy models. In addition, systematic analyses of the dimensions and facets of civic literacy are missing. With this study, we aimed to evaluate the degree to which our data supports the structure and assumptions of the proposed framework (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014) and therefore investigated the following research questions:

1. Is the theoretically assumed structure of the framework better supported by the data than alternative, simpler models?
   1.1 Are the seven facets modeled in the framework empirically separable into distinct factors?
   Given that the various facets of the framework represent complex and theoretically separable concepts, we expected that all facets are separable, but interrelationships (e. g. common factors) exist. In addition, we expected that all items are at least congeneric items, i. e., the factor loadings and error variances of the items may be different, but all items' factor loadings reach substantial magnitude ($\lambda \geq .50$).
   1.2. Is a hierarchical model with facets and dimensions better supported by the data than a non-hierarchical model?
   We assumed that a hierarchical model is better supported by the data than a non-hierarchical model. The reason for this is the distinction between broader dimensions and more concrete facets.
2. What levels of measurement invariance are present for the framework for 7th and 10th grade?
   We expected that configural, metric and scalar invariance holds indicating that the scores from each group have the same unit of measurement (factor loadings) and the same origin (intercepts). Hence, the latent means could be compared across 7th and 10th grade.
3. What differences can be found for the mean scores of the framework's facets between 7th and 10th grade?
   We assumed that all facets are developmentally sensitive and thus

more pronounced (i. e., have higher mean values) at the 10th grade level than at the 7th grade level. In addition, we assumed that political knowledge will be particularly higher, because knowledge transfer is a central goal and recurring component of (civic) education.

4. How closely are the framework's facets related to other facets of the framework (providing convergent evidence) and constructs that are unrelated to civic literacy (providing discriminant evidence)?
   We expected that the different facets of the framework are more closely related to each other than to sense of social affiliation, which is theoretically unrelated to civic literacy. Since the framework distinguishes between facets and dimensions, we also assumed that facets attributed to the same dimension are more strongly related to each other than facets belonging to different dimensions.

3. Methods

3.1. Participants and Procedure

Participants were 1,047 students attending 7th (n = 613, M_{age} = 12.47; 47.4 % female; 44.4 % immigrant background) or 10th grade (n = 434, M_{age} = 15.49; 47.7 % female; 35.9 % immigrant background). Data were collected in fall and winter 2021/2022 in the context of the Study on the Development of Political and Civic Competence in Adolescence (*Studie zur Entwicklung politischer und gesellschaftlicher Kompetenz im Jugendalter*, EPKO). The schools were located in rural, semi-urban and urban areas of West-Germany in the federal state of North Rhine-Westphalia. The response rate was 72 %. Students not reached either did not have informed consent or were absent on the test day due to illness or quarantine measures. The share of students with an immigrant background was with 41.8 % comparable to the official percentage of students with an immigration background in the same federal state in 2020 (40.0 %, MSW, 2022).

The assessments took place in the morning and took about two and a half hours including breaks. At the beginning of the assessments, the students worked on the knowledge test for one hour. After that, students took an 18-minute matrix test and a six-minute reading test to validate the knowledge test. Finally, all participants completed a questionnaire for about 40 minutes. The study was conducted in compliance with the German Research Foundation's (*Deutsche Forschungsgemeinschaft*, DFG) guidelines for good scientific practice. In addition, TU Dortmund university's Joint Ethics Com-

mittee examined and approved the research project. Students could only participate in the study if they presented an informed consent signed by their parents at the beginning of the assessment. Without informed consent, students had to leave the classroom. Each class was awarded 50 EUR for their participation.

3.2. Measures

Political Knowledge

The political knowledge test instrument included 63 items at the 7th grade level and 70 items at the 10th grade level. 34 items were answered by students in both grades, so that a total of 99 items were used. Some of the test items were taken from the POWIS-study (Goll et al., 2011) and adjusted if needed. The instrument targets different facets and content areas of political knowledge (Alscher et al., 2022b). In order to link the test results of the 7th and 10th grade, we calculated a unidimensional multigroup 1-parameter model using R and the "TAM" package (Robitzsch et al., 2021). Before calibration, we detected rapid guessing behavior, which led to the deletion of 6.76 % of the answers. We set the threshold for deletion at seven seconds of processing time. The threshold represents the local minimum between the two modes in the response time distribution (Wise, 2017). For every participant, we estimated a person parameter (θ). Overall and grade-specific reliabilities are shown in table 1.

Political Motivations

Political Interest. We measured political interest with the Short Scale Measuring Political Interest (*Politisches Interesse Kurzskala*, PIKS, Otto & Bacherle, 2011). Due to space constraints and after psychometric consideration of a prior assessment (Alscher et al., 2022a), only the first four items of the scale were used (e. g., "I observe political events with great interest."). The introductory question in the two motivation scales read as follows: "To what extent do the following statements apply to you?" The students were asked to rate each item on a four point scale ranging from 1 = "doesn't apply at all" to 4 = "applies very well".

Political Efficacy. We measured political efficacy with the Political Efficacy Kurzskala (PEKS, Beierlein et al., 2014). The scale consists of two items on

internal (e. g., "I am good at understanding and assessing important political issues.") and two items on external (e. g., "Politicians strive to keep in close touch with the people.") political efficacy.

Political Attitudes

Attitude towards Democracy. We measured attitude towards democracy with a newly constructed four-item-scale. The introductory question read as follows: "When you compare democracy with other forms of government, how much do you agree with the following statements?" The items of the sale are (1) "Democracy is the best form of government there is.", (2) "Democracy best ensures a just and fair society.", (3) "Democracy is the best way to govern a country.", and (4) "Democracy best protects fundamental freedoms."

Attitude towards opinion pluralism. We also measured attitude towards opinion pluralism with a novel four-item-scale. The introductory question read as follows: "To what extent do the following statements apply to you?" The items of the scale are (1) "All people should have the same opportunities to stand up for their opinions.", (2) "The opinion of minorities should also be taken into account in decision-making.", (3) "Dissenting opinions should be accepted and respected.", and (4) "Everyone should have the right to stand up for their opinion, even if the majority disagrees."

Political Volitions

Willingness to politically participate. We measured the willingness to politically participate with a four-item-scale. The introductory question read "If you want to voice your opinion, how likely are you to take the following actions?" For each item, students chose a response on a four-point scale comprising 0 = "not at all likely", 1 = "rather not likely", 2 = "rather likely", and 3 = "very likely". Each item refers to a different political activity, such as voting, protesting, signing a petition.

Willingness to civically participate. We also measured the willingness to civically participate with a four-item-scale. The introductory question again read "If you want to voice your opinion, how likely are you to take the following actions?" Each item refers to a different civic activity, such as donating money, volunteering, boycotting products.

Control Variables

Sense of social affiliation. We measured the sense of social affiliation with a scale from the NEPS (Bömmel et al., 2022) including five items (e. g., I find it easy to make new friends"). The introductory prompt read "For each statement, please tell me how it applies to you using the scale." The students were asked to rate each item on a six point scale ranging from 1 = "doesn't apply at all" to 6 = "applies very well".

School grade. 7th grade was coded as 0 and 10th grade as 1.

3.3. Data Analysis

We conducted all analyses in R, mostly using the "lavaan" and "psych" packages. Confirmatory factor analyses were performed to investigate the internal structure of the model. For this purpose, we compared a series of nested models with each other. The first three models were non-hierarchical models and contained either facets (1), dimensions (2), or a global factor (3). Additionally, three hierarchical models contained either the facets (4) or the dimensions (5) and a global factor, or facets and dimensions but no global factor (6). We set the variance of the factors to one. All models included all items. We disallowed correlations between items in all models. Correlations between factors were allowed only when no other, higher factor was modelled. Prior to the CFAs, we tested the fitness of the dataset for performing factor analyses using the Kaiser-Meyer-Olkin test and the Bartlett's Test of Sphericity. Based on the CFAs' results, the model was adjusted. We used the adjusted model for further analyses.

For the examination of measurement invariance between students in 7th and 10th grade, we performed a series of multiple group confirmatory factor analyses (MGCFA), which test the invariance of estimated parameters of two nested models across different groups (Cheung & Rensvold, 2002). The first MGCFA was specified for each group separately without any equality constraints. For this baseline model, we assessed the overall model fit to examine whether configural invariance holds. In a second model, we constrained factor loadings to be equal across 7th and 10th graders to check for metric invariance. By comparing the fit of the baseline model and the metric model, we can determine whether metric invariance holds. After establishing metric invariance, scalar invariance can be tested by specifying a third model with both constrained factors and intercepts. By comparing the metric model and the scalar model, we can determine whether scalar invariance

holds. A common statistic for assessing differences in fit between two nested models is the Likelihood Ratio Test. However, χ^2 difference tests were criticized for their dependence on sample size (e. g., Lee & Smith, 2020). Hence, we also used differences in the comparative fit index (ΔCFI) to evaluate meaningful change in model fit (Cheung & Rensvold, 2002). When sample size is rather large ($N > 300$) and sample sizes are comparable across groups, changes in the root-mean-square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) can also be evaluated to identify noninvariance (Chen, 2007). For testing metric invariance, a change of ≥ –.010 in CFI, supplemented by a change of ≥ .015 in RMSEA or a change of ≥ .030 in SRMR would indicate noninvariance. For the examination of scalar invariance a change of ≥ –.010 in CFI, supplemented by a change of ≥ .015 in RMSEA or a change of ≥ .010 in SRMR would indicate noninvariance (Chen, 2007).

In contrast to achievement tests, changes in competency measures are not only attributed to instructional sensitivity (Ing, 2018). In competence research, both school and non-school characteristics are considered as possible sources of development (Messick, 1984). Together, these characteristics represent the developmental sensitivity of a test. We tested the developmental sensitivity of the framework by comparing the mean values between 7th and 10th graders using an unpaired two-samples t-test. Before comparison, we weighted students' responses based on type of school.

We obtained convergent and discriminant evidence by specifying and evaluating a nomological network (Cronbach & Meehl, 1955). For this purpose, we identified three nomological orders. The first order represented correlations between facets from the same dimension of the framework. The second order represented correlations between facets from different dimensions of the framework. To this end, the facets of the framework were correlated with political interest, which is a central facet of the framework. Exceptions are political effectiveness and political interest itself. These two facets were correlated with attitude towards democracy. The third order represented correlations between facets of the framework and sense of social affiliation, which is theoretically unrelated to civic literacy. We compared the magnitude of correlations of the three orders with each other. Any differences in correlations were tested for statistical significance.

In this study, we evaluated global model fit with the comparative fit index (CFI), the Tucker-Lewis Index (TLI), the RMSEA, and the SRMR, using the following fit criteria for acceptable fit: CFI ≥ .900, TLI ≥ .900, RMSEA ≤ .100, SRMR < .100 (Weston & Gore, 2006). In addition, χ^2, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were used

for comparative model fit. In all models, we treated political knowledge as a latent variable with a single indicator. We controlled for the clustered structure of the data by using class (n = 69) as the cluster variable, thus enabling the computation of robust standard errors of the parameters estimated in the CFAs. We handled missing data through full information maximum likelihood (FIML) method. Supplemental material (e. g. the full scales, additional analyses) and the script are available at https://osf.io/tr8jp/.

4. Results

4.1. Descriptive Statistics

Table 1 shows the latent correlation coefficients, mean, standard deviations and range as well as the response rate and reliability for each of the constructs. The mean score of the political knowledge is slightly above 0. This is because in the process of scaling the average score of the 7th graders was set to 0. The average score of the 10th graders is 0.53. Average scores for the other variables are a little above the midpoint of the response scales. The only exception is the attitude towards opinion pluralism, whose average value of 3.39 is well above the midpoint and close to the maximum value of 4.00.

All 1,047 participants completed the political knowledge test. The reason for the quantitative variance in the response rates of the different constructs is the positioning of the instruments in the questionnaire. Due to internal school processes and stricter hygiene standards imposed by the pandemic, it was not possible to ensure that all participants were able to complete all questions in every survey for time reasons.

As expected, the bivariate correlations in Table 1 show moderate to strong, statistically significant correlations. The correlations between political interest and political efficacy as well as the correlations between willingness to politically and willingness to civically participate are particularly high. Again, this is not surprising when considering that in both cases both facets belong to the same dimension (i. e., political motivations and political volitions).

The WLE reliability for the knowledge test was good. The McDonald's ω for almost all other scales was acceptable to excellent (≥ .70). The exceptions were attitude towards opinion pluralism across the two grades (ω = .66) and in the 10th grade (ω = .63), and willingness to civically participate in the 10th grade (ω = .68).

Table 1: Means, standard deviations, and correlations.

Variable	1	2	3	4	5	6	7	8
1. Political knowledge	–							
2. Political interest[a]	**.36**	–						
3. Political efficacy[a]	**.42**	**.81**	–					
4. Attitude towards democracy[a]	**.57**	**.38**	**.50**	–				
5. Attitude towards opinion pluralism[a]	**.39**	**.18**	**.25**	**.39**	–			
6. Willingness to politically participate[a]	**.36**	**.42**	**.46**	**.31**	**.30**	–		
7. Willingness to civically participate[a]	**.36**	**.30**	**.31**	**.27**	**.31**	**.98**	–	
8. Sense of social affiliation	.01	**.14**	**.21**	**.14**	**.21**	**.17**	**.18**	–
M	0.23	2.17	2.36	2.95	3.39	2.38	2.49	2.83
SD	0.86	0.76	0.59	0.75	0.68	0.71	0.66	0.55
Actual Range	[b]-1.7–3.1	1–4	1–4	1–4	1–4	1–4	1–4	1–4
n	1047	994	948	947	867	750	753	750
McDonald's ω	[c].89	.88	.78	.89	.66	.78	.72	.72
Grade 7	[c].89	.88	.79	.89	.91	.79	.76	.71
Grade 10	[c].88	.89	.76	.89	.63	.76	.68	.75

Note. M and *SD* represent mean and standard deviation, respectively. Correlations printed in bold reached statistical significance at the 5 % level.

[a]*M* and *SD* are calculated using the row means of the variable indicators. [b]rounded to one decimal place. [c]Instead of McDonald's ω, the WLE reliability is reported.

4.2. Framework Evaluation

We tested the factorability of the data with the Kaiser-Meyer-Olkin test (KMO). The KMO value for the present data is .90 and exceeds the suggested cutoff value of .60, indicating that a factor analysis can probably be conducted (Kaiser & Rice, 1974). The sphericity of the data was tested with Bartlett's Test of Sphericity. The test suggests that there is probably sufficient significant correlation in the data in order to perform a factor analysis ($\chi^2(300) = 7768$, $p < .001$).

For the comparison of the models, model 1 was used as the baseline model. Table 2 shows that model 1 and model 6 outperform all other models with regard to different goodness of fit indicators. The χ^2 test suggests that model 1 fits the data the best among the specified models. The AIC also favors model 1. However, the BIC favors model 6 over model 1 (ΔBIC = -22). In contrast to the AIC, the BIC awards a bigger penalty for additional parameters and thus favors parametrically simpler models. Model 6 has fewer correlations and thus fewer parameters than model 1.

Table 2: Comparison of Fit Indices in different nested CFAs.

Model	χ^2			Robust RMSEA		AIC	BIC	Δ AIC	Δ BIC
	Value	*df*	*p*[a]	Value	90 % CI				
M1: Facets only	989	256	–	.051	[.047, .055]	49121	49587	–	–
M2: Dimensions only	2799	271	< .001	.094	[.091, .097]	50901	51292	1780	1705
M3: G-factor only	5370	275	< .001	.134	[.131, .137]	53565	53936	4444	4349
M4: Facets & g-factor	1706	270	< .001	.070	[.067, .074]	49810	50206	689	619
M5: Dimensions & g-factor	3017	273	< .001	.097	[.094, .101]	51116	51497	1995	1910
M6: Dimensions & facets	1029	265	< .001	.051	[.048, .055]	49144	49565	23	-22

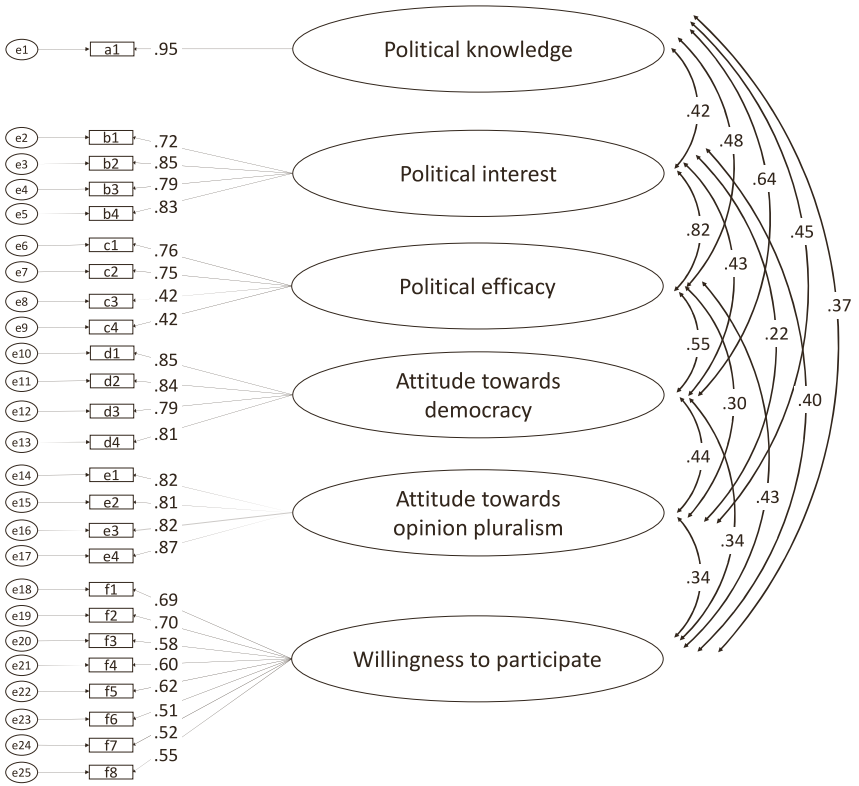
Note. RMSEA = root-mean-square error of approximation; CI = confidence interval; AIC = Akaike information criterion; BIC = Bayesian information criterion.

[a] The χ^2 of Model 1 is used as the benchmark.

Model 1 has an overall acceptable global model fit (CFI: .926; TLI: .913; RMSEA: .052; SRMR: .077; see figure B1). All factor loadings reached significance at the 5 % level. In addition, almost all factor loadings reached substantial magnitude ($\lambda \geq .50$) except for items c3 and c4. While items c1 and c2 loaded highly on political efficacy, items c3 and c4 showed rather low loadings on political efficacy ($\lambda = .41$ and .41 respectively). Eventually, factor loadings for the items of the willingness to civically participate facet were overall only moderate in size ($.50 < \lambda < .61$). Furthermore, the correlation between the two willingness facets is particularly high suggesting that the two factors might measure the same construct. We found similar patterns of results in model 6 (see figure B2). There too, the items of the willingness to civically participate have rather moderate factor loadings. Furthermore, items c1 and c2 again load more strongly on the political efficacy facet than items c3 and c4. Eventually, the modification indices added to the impression, that the political efficacy scale and the two willingness scales cause problems.

Based on these findings, we adjusted model 1. We combined the two willingness facets into one single facet. However, we were unable to separate the political efficacy facet into an internal and an external political efficacy facet, because there are not enough items available in the current framework. The adjusted model (see figure 2) was compared to model 1 and model 6 (see table B1). The χ^2 and AIC still favor model 1, but the BIC now favors the adjusted model. All further analyses will be carried out using the adjusted model.

Figure 2: Confirmatory factor analysis (adjusted model).



Note. CFI: .925; TLI: .914; RMSEA: .052; SRMR: .078.
All factor loadings and correlations reached statistical significance at the 5% level.

For the testing of measurement invariance by class grade, we used the adjusted model as the baseline model. The configural model exhibited an acceptable model fit (see table 2). Regarding metric invariance, the χ^2 test suggests that the configural model does not fit the data statistically significantly better than the metric model. In addition, neither the drop in the CFI (ΔCFI = .000) or the RMSEA (ΔRMSEA = .000), nor the drop in the SRMR (ΔSRMR = -.010) exceed the suggested threshold indicating metric invariance. In terms of scalar invariance, the χ^2 test suggests that the metric model fits the data statistically significantly better than the scalar model. However, the drop in the CFI is below the .007 threshold (ΔCFI = .005) and the CFI itself remains on an acceptable level. Furthermore, the RMSEA (ΔRMSEA = .000) and the SRMR (ΔSRMR = .000) remain stable.

Table 3: Measurement invariance between 7th and 10th grade.

Model	χ^2			CFI	RMSEA	SRMR
	Value	*df*	*p*[a]			
1. Configural model	1338	524		.918	.054	.095
2. Metric model	1357	542	.096	.918	.054	.085
3. Scalar model	1425	561	< .001	.913	.054	.085

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root mean square residual.

[a] The χ^2 of the metric model was compared to the χ^2 of the configural model. The χ^2 of the scalar model was compared to the χ^2 of the metric model.

Next, we tested the developmental sensitivity of the framework by comparing the facets' means between the 7th and 10th grade. All means were higher in the 10th grade than in the 7th grade (see table A1). The differences between the means were statistically significant at the 5 % level for political knowledge, political interest, attitude towards democracy and attitude towards opinion pluralism, and not significant for political efficacy and willingness to participate.

The correlations between the facets of the framework and sense of social affiliation largely correspond to the assumed nomological network (see table A2). The correlational relations of the first order are the largest for all facets while the relations of the third order are the smallest. The only exception is attitude towards democracy, for which the third order relation is larger than the second order relation. All other differences in correlations were significant at the 5 % level with two exceptions. The correlation between attitudes towards democracy and attitude towards opinion pluralism ($r = .39$, $p < .001$) was statistically only insignificantly higher ($\Delta = .01$, $p = .850$) than the correlation between attitude towards democracy and political interest ($r = .38$, $p < .001$). Moreover, the correlation between willingness to participate and political interest ($r = .37$, $p < .001$) was statistically only insignificantly higher ($\Delta = .20$, $p = .920$) than the correlation between willingness to civically participate and sense of social affiliation ($r = .17$, $p = .008$).

5. Discussion

Gradual declines of democratic regime attributes characterize contemporary autocratization (Lührmann & Lindberg, 2019). It is therefore perhaps more important than ever to understand what constitutes a civically literate citizen and to unravel what factors play a role in the development of young people on their way to becoming democratic and active citizens (Oberle, 2022). However, previous research has usually focused on separate areas of civic

literacy, thereby often relegating either cognitive or affective dimensions to a different sphere. Thus, holistic overviews and conceptualizations of youth's civic literacy are uncommon (see Keegan, 2021). In addition, existing conceptualizations are usually rooted solely in theoretical considerations, while empirical examinations of the proposed frameworks are absent. Before data collection, a panel of experts consisting of six professors for political didactics as well as the study director of an international large-scale assessment on civic education assessed the adequacy of the civic literacy framework's content (Alscher et al., 2022b).

The results of this study show that the structure of the suggested civic literacy framework fits the data overall well and most expectations are met. Regarding the internal structure of the model, correlations, and the CFAs give rise to the assumption that internal and external political efficacy should be viewed as two separate constructs. Indeed political efficacy is often understood as a bidimensional belief that includes distinct external and internal components (Ardèvol-Abreu et al., 2020; Oberle, 2018). Previous research has shown that internal and external efficacy have different meanings with regard to the political socialization of young people. For instance, the education effect on internal efficacy is considerably more confounded by personality traits than the educational effect on external efficacy (Rasmussen & Nørgaard, 2018). Future studies should make sure to assess internal and external political efficacy separately and with enough items. Furthermore, the willingness to politically and the willingness to civically participate correlate highly and both load particularly high on the political volitions dimension. In addition, the item loadings are only mediocre. We therefore adjusted the framework accordingly (see figure 3).

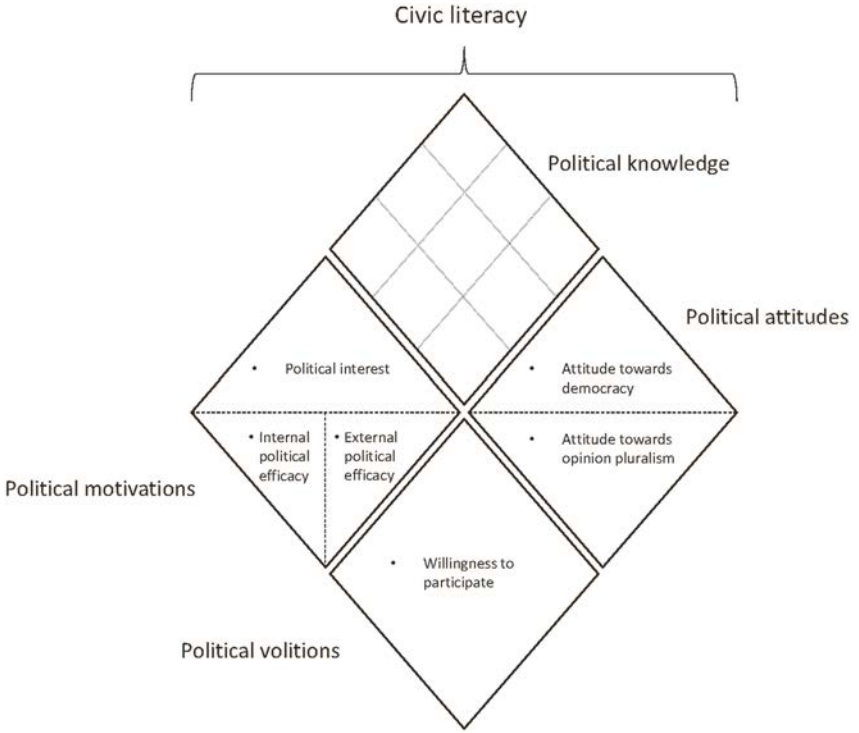
Using the adjusted model, scalar measurement invariance was established according to the ΔCFI, ΔRMSEA and ΔSRMR. Furthermore, the overall model fit remains rather high across the different MGCFAs.

Unlike previous studies, the framework is designed to enable a longitudinal assessment of civic literacy. For this purpose, it is important that the framework is developmentally sensitive and reflects processes of students' political socialization. The comparison of the 7th and 10th grade cohort shows that all framework facets' means are higher in the 10th grade than in 7th grade. However, magnitudes of the differences vary with regard to statistical significance and effect sizes. Especially political knowledge seems to increase from 7th to 10th grade. This is not surprising when one considers that it is primarily cognitive processes and thus knowledge transfer that play a central role in civic education, whereas aspects of political practice are at best of marginal importance (Widmaier & Nonnenmacher, 2011). We

can observe a mediocre difference for political attitudes. However, one of the limitations here is that the identified difference might be confounded by the fact that 7th graders are less familiar with the terms that were used (e. g., democracy, minority/majority) and therefore might have a stronger tendency for the midpoint of the scale. Rather counterintuitively, we found only small differences for political efficacy and willingness to participate. Possible reasons for this could be that younger students might have lower estimates of potential costs for meaningful engagement, assign greater value to engagement or expect higher rewards from actively participating in civic and political life. Moreover, students' decision to become active might not (only) be associated with their absolute but instead their relative level of political knowledge, attitudes or abilities compared to their peers. Hence, participation patterns could be different within a particular class and grade level but rather equal across classes and grade levels. The general mechanism is also known as the 'big fish little pond'-effect, which has already proven helpful in explaining differences in inequality of civic and political engagement (Witschge, 2022). Future research should strive for a better understanding of this relationship.

The investigation of the nomological network suggests that facets of the framework are more related to each other than to a construct that is theoretically unrelated to civic literacy. Furthermore, facets of the same dimension are usually more strongly associated with each other than facets of different dimensions. The framework's constructs thus demonstrate convergence and discrimination. The only exception is attitude towards democracy, which has very similar correlations with attitude towards opinion pluralism and with political interest. Upon closer examination, the bigger problem seems to be the convergence between the two attitude constructs, since the correlation between the two is comparatively weak. This impression is also confirmed by the CFA of model 6 (dimensions and facets), which suggests that the two attitude facets load rather differently on the political motivation dimension. A possible reason for this could be that attitude towards opinion pluralism is less strongly concerned with original political topics. Instead, a social aspect resonates in this facet since the exchange of opinions is something that students experience regularly outside the political arena in their everyday life. This needs to be taken into account in future applications of the framework but also when other political motivations are assessed.

Figure 3: Adjusted Civic Literacy Framework.



Note. From „Civic literacy — about the theory and measurability of competence in school-based civic education" by Alscher et al., 2022, Zeitschrift für Erziehungswissenschaft (https://doi.org/10.1007/s11618-02201085-0). CC BY 4.0.

6. Conclusion

While the empirical assessment of competencies in the school context has gained in importance in the past two decades, there are comparatively few studies with the aim of systematically assessing civic literacy. We therefore aimed to realign the conceptual and empirical foundations of civic literacy research by empirically examining the conceptualized civic literacy framework. The results show that the framework fits the data overall well and most theoretical assumptions are confirmed. However, political efficacy should be distinguished between external and internal political efficacy. In contrast, there is no clear differences between the willingness to politically and the willingness to civically participate. The framework outlined in this study is to be applied in future studies of political socialization as well as empirical educational research and political psychology. This includes, in particular, the

National Educational Panel Study (NEPS; Blossfeld and Roßbach 2019), for which civic literacy was included in the canon of competencies to be assessed in grades 7 and 10. The application of the civic literacy framework, or parts of it, shall provide important insights for future research and practice of (school-based) civic education.

References

Achtziger, A., & Gollwitzer, P. M. (2010). Motivation und Volition im Handlungsverlauf. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation und Handeln* (4th ed., pp. 309–335). Springer VS. https://doi.org/10.1007/978-3-642-12693-2_12

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I., Timko, C., & White, J. (1982). Self-monitoring and the attitude-behavior relation. *Journal of Personality and Social Psychology*, *42*, 426–435. https://psycnet.apa.org/doi/10.1037/0022-3514.42.3.426

Alscher, P., Ludewig, U., & McElvany, N. (2022a). Civic education, teaching quality and students' willingness to participate in political and civic life: Political interest and knowledge as mediators. *Journal of Youth and Adolescence*, *51*(10), 1886–1900. https://doi.org/10.1007/s10964-022-01639-9

Alscher, P., Ludewig, U., & McElvany, N. (2022b). Civic Literacy – zur Theorie und Messbarkeit eines Kompetenzmodells für die schulische politische Bildung. *Zeitschrift Für Erziehungswissenschaft.* Advance online publication. https://doi.org/10.1007/s11618-022-01085-0

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2014). *The standards for educational and psychological testing.* American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Ardèvol-Abreu, A., Gil de Zuniga, H., & Gámez, E. (2020). The influence of conspiracy beliefs on conventional and unconventional forms of political participation: The mediating role of political efficacy. *British Journal of Social Psychology*, *59*(2), 549–569. https://doi.org/10.1111/bjso.12366

Barber, C., & Ross, J. (2018). Cross-Cohort Changes in Adolescents' Civic Attitudes from 1999 to 2009: An Analysis of Sixteen Countries. *Child Indicators Research, 11*, 681–703. https://doi.org/10.1007/s12187-017-9452-0

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1016/0146-6402(78)90002-4

Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2014). *Political Efficacy Kurzskala (PEKS)* (Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)). https://doi.org/10.6102/ZIS34

Blossfeld, H.-P., Maurice, J. von, & Schneider, T. (2019). The national educational panel study: need, main features, and research potential. In H.-P. Blossfeld & H.-G. Roßbach (Eds.), *Edition ZfE: Vol. 3. Education as a Lifelong Process: The German National Educational Panel Study (NEPS)* (pp. 1–16). Springer VS. https://doi.org/10.1007/978-3-658-23162-0_1

Bömmel, N., Kroh, J., Gebel M., & Heineck G. (2022). *Social Inclusion as Return to Education in the National Educational Panel Study (NEPS): Conceptual Framework and Measurement* [Manuscript submitted for publication].

National Educational Panel Study. Doi:XXX

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834
Cheung, G. W., & Rensvold, R. B. (2002) Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance, *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255, https://doi.org/10.1207/S15328007SEM0902_5
Cohen, J. (1992). Quantitative Methods in Psychology: A Power Primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155
Cramer, K. J., & Toff, B. (2017). The fact of experience: Rethinking political knowledge and civic competence. *Perspectives on Politics*, *15*(3), 754–770. https://doi.org/10.1017/S1537592717000949
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957
Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. Yale University Press.
Detjen, J., Massing, P., Richter, D., & Weißeno, G. (2012). *Politikkompetenz – ein Modell*. Springer VS. https://doi.org/10.1007/978-3-658-00785-0
Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction*, *47*, 53–64. https://doi.org/10.1016/j.learninstruc.2016.10.009
Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes* (2.th ed.). Harcourt Brace Jovanovich.
Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153
Ekman, J., & Amnå, E. (2012). Political participation and civic engagement: Towards a new typology. *Human Affairs*, *22*(3), 283–300. https://doi.org/10.2478/s13374-012-0024-1
Ertl, H. (2006). Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education*, *32*(5), 619–634. https://doi.org/10.1080/03054980600976320
Fabian, P., Goy, M., Jarsinski, S., Naujokat, K., Prosch, A., Strietholt, R., Blatt, I., & Bos, W. (2019). Transition and development from lower secondary to upper secondary school. In H.-P. Blossfeld & H.-G. Roßbach (Eds.), *Edition ZfE: Vol. 3. Education as a Lifelong Process: The German National Educational Panel Study (NEPS)* (pp. 231–252). Springer VS. https://doi.org/10.1007/978-3-658-23162-0_13
Gabriel, O. W. (2000). Demokratische Einstellungen in einem Land ohne demokratische Traditionen? Die Unterstützung der Demokratie in den neuen Bundesländern im Ost-West-Vergleich. In J. W. Falter, O. W. Gabriel, & H. Rattinger (Eds.), *Wirklich ein Volk? Die politischen Orientierungen von Ost- und Westdeutschen im Vergleich* (pp. 41–77). Springer VS. https://doi.org/10.1007/978-3-663-05704-8_2
Gaspard, H., Wille, E., Wormington, S. V., & Hulleman, C. S. (2019). How are upper secondary school students' expectancy-value profiles associated with achievement and university STEM major? A cross-domain comparison. *Contemporary Educational Psychology*, *58*, 149–162. https://doi.org/10.1016/j.cedpsych.2019.02.005
Gil de Zúñiga, H., & Diehl, T. (2018). News finds me perception and democracy: Effects on political knowledge, political interest, and voting. *New Media & Society*, *21*(6), 1253–1271. https://doi.org/10.1177/1461444818817548
Gil de Zúñiga, H., & Goyanes, M. (2021). Fueling civil disobedience in democracy: WhatsApp news use, political knowledge, and illegal political protest. *New Media & Society*, 14614448211047850. https://doi.org/10.1177/14614448211047850
GPJE. (2004). *Anforderungen an Nationale Bildungsstandards für den Fachunterricht in der Politischen Bildung an Schulen: Ein Entwurf*. Wochenschau Verlag. http://gpje.de/wp-content/uploads/2017/01/Bildungsstandards-1.pdf

Goll, T., Richter, D., & Weißeno, G. (2011). Politisches Wissen von Schüler/-innen mit und ohne Migrationshintergrund (POWIS): Ergebnisse einer Studie. In D. Lange (Eds.), *Schriftenreihe der DVPB, Wochenschau Wissenschaft, Entgrenzungen: Gesellschaftlicher Wandel und politische Bildung* (pp. 126–131). Frankfurt a. M.: Wochenschau Verlag.

Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik*. Springer Medizin Verlag Heidelberg. https://doi.org/10.1007/3-540-33020-8_9

Hedinger, F. (2018). Erkenntnisse aus der empirischen Politikwissenschaft für das Politikkompetenzmodell. In B. Ziegler & M. Waldis (Eds.), *Politische Bildung in der Demokratie: Interdisziplinäre Perspektiven* (pp. 57–73). Springer VS. https://doi.org/10.1007/978-3-658-18933-4_5

Herne, K., Christensen, H. S., & Grönlund, K. (2019). The influence of political knowledge on opinion polarization in citizen deliberation. *Political Research Exchange*, *1*(1), 1–23. https://doi.org/10.1080/2474736X.2019.1702887

Homana, G. A. (2018). Youth political engagement in Australia and the United States: Student councils and volunteer organizations as communities of practice. *Journal of Social Science Education*, *17*(1), 41–54. https://doi.org/10.4119/jsse-865

Hoskins, B., Saisana, M., & Villalba, C. M. H. (2015). Civic competence of youth in Europe: Measuring cross national variation through the creation of a composite indicator. *Social Indicators Research*, *123*(2), 431–457. https://doi.org/10.1007/s11205-014-0746-z

Ing M. (2018). What About the "Instruction" in Instructional Sensitivity? Raising a Validity Issue in Research on Instructional Sensitivity. *Educational and Psycholical Measurement 78*(4), 635–652. https://doi.org/10.1177/0013164417714846

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177%2F001316447403400115

Kavadias, D., Hemmerechts, K., & Spruyt, B. (2017). Segregation and socialization: Academic segregation and citizenship attitudes of adolescents in comparative perspective? *Journal of Social Science Education*, *16*(2), 30–41. https://doi.org/10.4119/jsse-832

Keegan, P. (2021). Critical affective civic literacy: A framework for attending to political emotion in the social studies classroom. *The Journal of Social Studies Research*, *45*(1), 15–24. https://doi.org/10.1016/j.jssr.2020.06.003

Kenski, K., & Stroud, N. J. (2006). Connections between internet use and political efficacy, knowledge, and participation. *Journal of Broadcasting & Electronic Media*, *50*(2), 173–192. https://doi.org/10.1207/s15506878jobem5002_1

Kim, H., & Byun, S.-Y. (2019) Immigrant integration policy and native adolescents' attitudes towards ethnic minorities: a comparative study of European Countries, *Multicultural Education Review, 11*(3), 172–188. https://doi.org/10.1080/2005615X.2019.1644041

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., & Tenorth, H.-E. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn, Berlin: BMBF. https://doi.org/10.25656/01:20901

Köller, O. (2016). Bildungsstandards. In R. Tippelt & B. Schmidt-Hertha (Eds.), *Handbuch Bildungsforschung* (pp. 625–648). Springer VS. https://doi.org/10.1007/978-3-531-20002-6_26-1

Lange, H. (2002). PISA: Und was nun? *Zeitschrift für Erziehungswissenschaft*, *5*(3), 455–471. https://doi.org/10.1007/s11618-002-0061-3

Lauermann, F., Tsai, Y.-M., & Eccles, J. S. (2017). Math-related career aspirations and choices within Eccles et al.'s expectancy – value theory of achievement-related behaviors. *Developmental Psychology, 53*(8), 1540. https://psycnet.apa.org/doi/10.1037/dev0000367

Lee, H., & Smith, W. Z. (2020). Fit Indices for Measurement Invariance Tests in the Thurstonian IRT Model. *Applied Psychological Measurement 44*(4), 282–295. https://doi.org/10.1177/0146621619893785

Levy, B. L. M., & Akiva, T. (2019). Motivating political participation among youth: An analysis of factors related to adolescents' political engagement. *Political Psychology*, *40*(5), 1039–1055. https://doi.org/10.1111/pops.12578

Liem, G. A. D., & Chua, B. L. (2013). An expectancy-value perspective of civic education motivation, learning and desirable outcomes. *Educational Psychology*, *33*(3), 283–313. https://doi.org/10.1080/01443410.2013.776934

Lührmann, A., & Lindberg, S. I. (2019). A third wave of autocratization is here: what is new about it? *Democratization*, *26*(7), 1095–1113. https://doi.org/10.1080/13510347.2019.1582029

Massing, P. (2022). Kompetenzmodelle in der Politikdidaktik. In G. Weißeno & B. Ziegler (Eds.), *Handbuch Geschichts- und Politikdidaktik* (pp. 17–32). Springer VS. https://doi.org/10.1007/978-3-658-29668-1_2

Maurissen, L. (2020). Political efficacy and interest as mediators of expected political participation among Belgian adolescents. *Applied Developmental Science*, *24*(4), 339–353. https://doi.org/10.1080/10888691.2018.1507744

May, M. (2007). *Demokratiefähigkeit und Bürgerkompetenzen: Kompetenztheoretische und normative Grundlagen der politischen Bildung*. Springer VS. https://doi.org/10.1007/978-3-531-90492-4

May, M., Moritz, B., & Riegel, P. (2020). Politische Urteilskompetenz empirisch – Überlegungen zur Operationalisierung und ein Test zur politischen Werturteilskompetenz. *Zeitschrift Für Didaktik Der Gesellschaftswissenschaften*, *11*(1), 134–146. https://elibrary.utb.de/doi/10.46499/1601.1145

McDonnell, J. (2020). Municipality size, political efficacy and political participation: a systematic review. *Local Government Studies*, *46*(3), 331–350. https://doi.org/10.1080/03003930.2019.1600510

McWhirter, E. H., & McWhirter, B. T. (2016). Critical consciousness and vocational development among Latina/o high school youth: initial development and testing of a measure. *Journal of Career Assessment*, *24*(3), 543–558. https://doi.org/10.1177/1069072715599535

Messick, S. (1984). The Psychology of Educational Measurement. *Journal of Educational Measurement, 21*(3), 215–237. https://doi.org/10.1111/j.1745-3984.1984.tb01030.x

Meyer, J., Fleckenstein, J., & Köller, O. (2019). Expectancy value interactions and academic achievement: Differential relationships with achievement measures. *Contemporary Educational Psychology*, *58*, 58–74. https://doi.org/10.1016/j.cedpsych.2019.01.006

Ministry for Schools and Further Education of the State of North Rhine-Westphalia. (2022). *Das Schulwesen in NRW aus quantitativer Sicht 2021/22: Statistische Übersicht Nr. 417 [The school system in North Rhine-Westphalia from a quantitative perspective. 2016/17. Statistical overview nr. 417].* MKW. https://www.schulministerium.nrw/system/files/media/document/file/quantita_2021.pdf

Mondak, J. J. (2001). Developing valid knowledge scales. *American Journal of Political Science*, 224–238. https://doi.org/10.2307/2669369

Myoung, E., & Liou, P.-Y. (2022): Systematic review of empirical studies on international large-scale assessments of civic and citizenship education. *Scandinavian Journal of Educational Research*, *66*(7), 1269–1291. https://doi.org/10.1080/00313831.2022.2131903

National Assessment Governing Board. (2018). *Civics Framework: for the 2018 National Assessment of Educational Progress*. Washington, DC. National Assessment Governing Board, U.S. Department of Education. https://www.nagb.gov/content/nagb/assets/documents/publications/frameworks/civics/2018-civics-framework.pdf

Oberle, M. (2022). Demokratiebildung in der Schule. In K. Möller, F. Neuscheler, & F. Steinbrenner (Eds.), *Demokratie gestalten! Herausforderungen und Ansätze für Bildungs-und Sozialarbeit*. (pp. 62–73). W. Kohlhammer Verlag.

Oberle, M. (2018). Politisches Effektivitätsgefühl von Schüler/-innen. Struktur, Determinanten und Veränderbarkeit einer motivationalen Facette politischer Kompetenz. In S. Manzel & M. Oberle (Eds.), *Kompetenzorientierung – Potenziale zur Professionalisierung der Politischen Bildung* (pp. 85–97). Springer VS. https://doi.org/10.1007/978-3-658-16889-6_8

Oberle, M., & Wenzel, N. (2019). Politisches Vertrauen und Effektivitätsgefühl von Schüler/innen – Einflussfaktoren und Relevanz für politische Partizipation. In M. Lotz & K. Pohl (Eds.), *Schriftenreihe der GPJE. Gesellschaft im Wandel: Neue Aufgaben für die politische Bildung und ihre Didaktik* (pp. 176–186). Wochenschau Verlag.

OECD [The Organisation for Economic Cooperation and Development] (2001). *Knowledge and skills for life. First results from the OECD Programme for International Student Assessment (PISA) 2000*. OECD Publications.

Otto, L., & Bacherle, P. (2011). Politisches Interesse Kurzskala (PIKS): Entwicklung und Validierung. *Politische Psychologie*, *1*(1), 19–35.

Prior, M. (2019). *Hooked: How politics captures people's interest*. Cambridge University Press. https://doi.org/10.1017/9781108355001

Prior, M., & Bougher, L. D. (2018). "Like they've never, ever seen in this country"? Political interest and voter engagement in 2016. *Public Opinion Quarterly*, *82*(S1), 822–842. https://doi.org/10.1093/poq/nfy002

Quintelier, E., & Hooghe, M. (2013). The relationship between political participation intentions of adolescents and a participatory democratic climate at school in 35 countries. *Oxford Review of Education*, *39*(5), 567–589. https://doi.org/10.1080/03054985.2013.830097

Rasmussen, S. H. R., & Nørgaard, A. S. (2018). When and why does education matter? Motivation and resource effects in political efficacy. *European Journal of Political Research*, *57*(1), 24–46. https://doi.org/10.1111/1475-6765.12213

Robitzsch, A., Kiefer, Thomas, & Wu, Margaret. (2021). *TAM: Test analysis module*. https://cran.r-project.org/web/packages/TAM/index.html

Roccato, M., & Zogmaister, C. (2010). Predicting the vote through implicit and explicit attitudes: A field research. *Political Psychology*, *31*(2), 249–274. http://www.jstor.org/stable/20721288

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *2012*, *48*(2), 36. https://doi.org/10.18637/jss.v048.i02

Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., & Friedman, T. (2018). *Becoming citizens in a changing world: IEA International Civic and Citizenship Education Study 2016: International report*. Springer VS. https://doi.org/10.1007/978-3-319-73963-2

Schulz, W., Fraillon, J., Losito, B., Agrusti, G., Ainley, J., Damiani, V., & Friedman, T. (2022). *International Civic and Citizenship Education Study 2022: Assessment Framework*. Springer VS.

Stanley, B., & Cześnik, M. (2022). Uninformed or informed populists? The relationship between political knowledge, socio-economic status and populist attitudes in Poland. *East European Politics*, *38*(1), 43–60. https://doi.org/10.1080/21599165.2021.1876676

Stattin, H., & Amnå, E. (2022). Basic values transform political interest into diverse political values, attitudes and behaviors. *Journal of Youth and Adolescence*. Advance online publication. https://doi.org/10.1007/s10964-022-01654-w

Steinmayr, R., Weidinger, A. F., Schwinger, M., & Spinath, B. (2019). The importance of students' motivation for their academic achievement – Replicating and extending previous findings. *Frontiers in Psychology*, *10*, 1730. https://doi.org/10.3389/fpsyg.2019.01730

van Deth, J. W. (1990). Interest in politics. In M. K. Jennings & J. W. van Deth (Eds.), *Continuities in political action* (pp. 275–312). De Gruyter. https://doi.org/10.1515/9783110882193.275

Waldow, F. (2009). What PISA did and did not do: Germany after the 'PISA-shock'. *European Educational Research Journal*, *8*(3), 476–483. https://doi.org/10.2304/eerj.2009.8.3.476

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe & Huber Publishers.

Weißeno, G., Detjen, J., Juchler, I., Massing, P., & Richter, D. (2010). *Konzepte der Politik – ein Kompetenzmodell. Schriftenreihe: Vol. 1016*. Bundeszentrale für politische Bildung. https://www.pedocs.de/volltexte/2016/12009/pdf/Weisseno_et_al_2010_Konzepte_der_Politik_.pdf

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, *34*(5), 719–751. https://doi.org/10.1177/0011000006286345.

Widmaier, B., & Nonnenmacher, F. (2011). Von der Politikverdrossenheit zum Wutbürger? Partizipation als Ziel der politischen Bildung – Zur Einführung. In B. Widmaier & F. Nonnenmacher (Eds.): *Partizipation als Bildungsziel. Politische Aktion in der politischen Bildung* (pp. 7–15). Wochenschau Verlag.

Wigfield, A., & Eccles, J. S. (2000). Expectancy – Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wise, S. L. (2017). Rapid-Guessing behavior: Its identification, interpretation, and implications. *Educational Measurement: Issues and Practice*, *36*(4), 52–61. https://doi.org/10.1111/emip.12165

Witschge, J. H. M. (2022). Education systems and inequality of civic and political engagement. [Doctoral thesis, Universiteit van Amsterdam]

Wray-Lake, L., Metzger, A., & Syvertsen, A. K. (2017). Testing multidimensional models of youth civic engagement: Model comparisons, measurement invariance, and age differences. *Applied Developmental Science*, *21*(4), 266–284. https://doi.org/10.1080/10888691.2016.1205495

Appendix A

Convergent and Discriminant Evidence, and Instructional Sensitivity

Table A1: Means of framework constructs in grade 7 and 10.

Framework construct	*M* Grade 7	*M* Grade 10	Δ (*p*)
Political knowledge	0.03	0.75	0.72 (< .001)
Political interest	2.14	2.23	0.10 (.225)
Political efficacy	2.39	2.40	0.01 (.943)
Attitude towards democracy	2.92	3.14	0.22 (.020)
Attitude towards opinion pluralism	3.30	3.51	0.21 (.001)
Willingness to participate	2.42	2.48	0.06 (.467)

Note. *M* represents mean.

Table A2: Correlations within and across framework dimensions and with sense of social affiliation.

Framework construct	Within dimension		PI/ATD[a]		SOSA
Political knowledge	– [b]		**.36**	>	-.01[c]
Political interest	**.81**	>	**.38**[c]	>	**.14**[c]
Political efficacy	**.81**	>	**.50**[c]	>	**.21**[c]
Attitude towards democracy	**.39**	>	**.38**	>	**.14**[c]
Attitude towards opinion pluralism	**.39**	>	**.18**[c]	<	**.21**
Willingness to participate	– [b]	>	**.37**[c]	>	**.17**

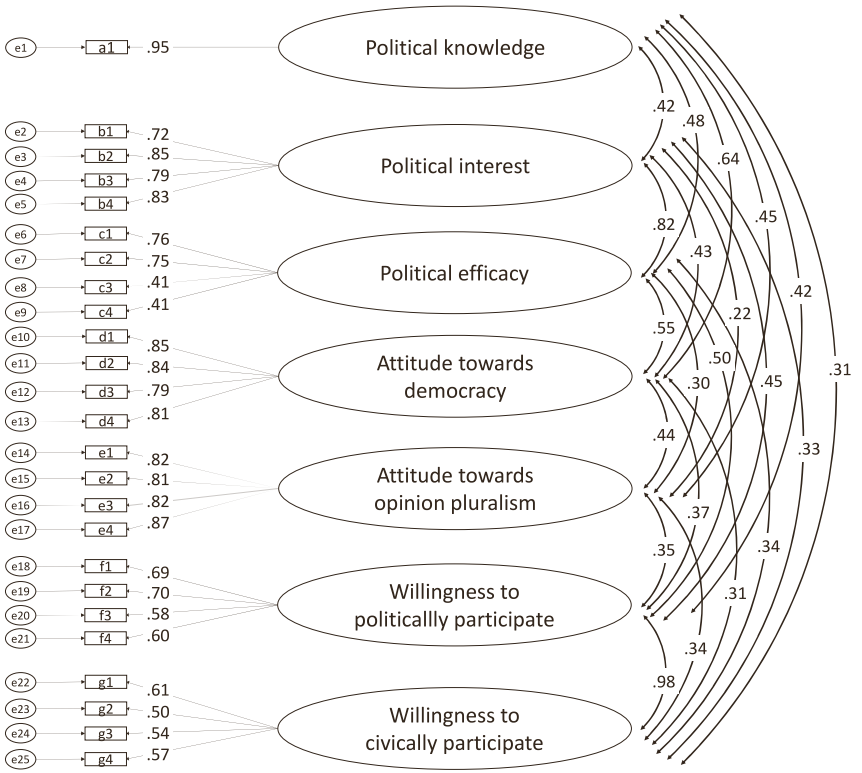
Note. PI = Political interest; ATD = Attitude towards democracy; SOSA = Sense of social affiliation. Correlations printed in bold reached statistical significance at the 5 % level.

[a]For the cross-dimensional correlations, political interest was used. The exceptions were political efficacy and political interest itself, for which attitude towards democracy was used instead. [b]The political knowledge dimension consisted of only one construct. [c]The correlation differs from the correlation to the left of it at the 5 % level of significance.

Appendix B

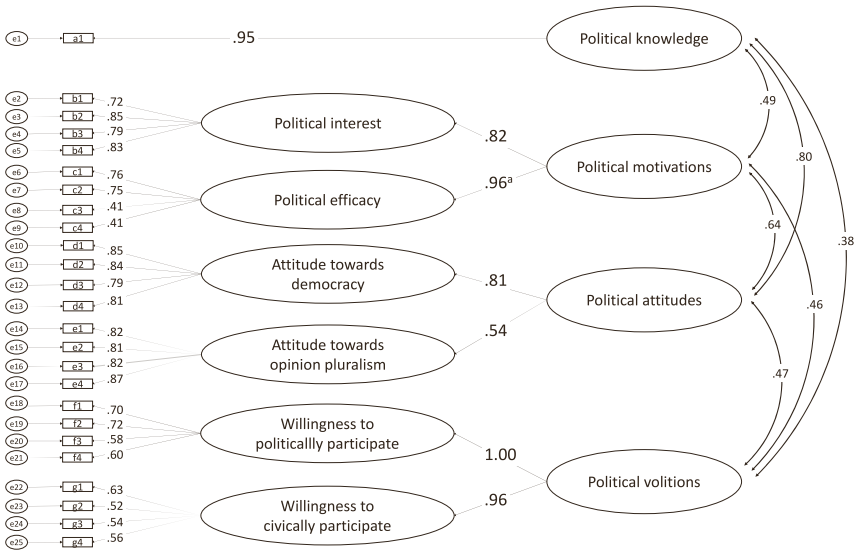
Additional Confirmatory Factor Analyses

Figure B1: Confirmatory factor analysis (model 1).



Note. CFI: .926; TLI: .913; RMSEA: .052; SRMR: .077. All factor loadings and correlations reached statistical significance at the 5% level.

Figure B2: Confirmatory factor analysis (model 6).



Note. CFI: .923; TLI: .913; RMSEA: .052; SRMR: .080. When not reported otherwise, factor loadings/correlations reached statistical significance at the 5 % level.

[a] Factor loading/covariance did not reach significance.

Table B1: Comparison of Fit Indices in different CFAs.

Model	χ^2			Robust RMSEA		AIC	BIC	Δ AIC	Δ BIC
	Value	*df*	*p*[a]	Value	90 % CI				
M1: Facets only	989	256	–	.051	[.047, .055]	49121	49587	–	–
M6: Dimensions & facets	1029	265	< .000	.051	[.048, .055]	49144	49565	23	-22
Adjusted Model	1008	262	.004	.051	[.047, .055]	49129	49564	8	-23

Note. RMSEA = root-mean-square error of approximation; CI = confidence interval; AIC = Akaike information criterion; BIC = Bayesian information criterion.

[a] The χ^2 of Model 1 is used as the benchmark.

Promoting Motivation in Mathematics: Can a Relevance Intervention Contribute to Reducing Educational Disparities?

Hanna Gaspard, Cora Parrisius & Benjamin Nagengast[1]

Abstract

According to expectancy-value theory (Eccles et al., 1983) and corresponding research, when students have high expectancies of success in a domain and attach great value to it, they show more effort and obtain higher achievement. Educational disparities that depend on gender and family background can further be explained through differences in expectations of success and value beliefs. Against this backdrop, interventions to promote the perceived relevance of learning content have been shown to be a potential avenue for promoting motivation and achievement, particularly for "at risk" groups. In this paper, data from a large cluster-randomized trial testing a relevance intervention in mathematics with 78 ninth-grade classrooms (Gaspard et al., 2021) were used to examine whether intervention effects differed by students' demographic characteristics. Small to moderate differences in motivation and achievement that depended on gender, parental education, and migration background were found before the intervention. However, there was little evidence that these student characteristics moderated the effects of the intervention. Therefore, the intervention did not seem to contribute to reducing educational disparities – at least not when applied in the same way for all students as was done here.

Keywords: educational disparities; gender; intervention; mathematics; migration background; motivation; parental education

1 This research was supported by the Eliteprogramme for Postdocs of the Baden-Württemberg Stiftung and the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63). We thank Heide Piesch, Eike Wille, and Ulrich Trautwein for their support in conducting this research.

Zusammenfassung

Wenn Schüler*innen in einem Fach eine hohe Erfolgserwartung haben und ihm einen hohen Wert beimessen, zeigen sie nach der Erwartungs-Wert-Theorie (Eccles et al., 1983) und entsprechender Forschung mehr Anstrengung und bessere Leistungen. Bildungsdisparitäten in Abhängigkeit von Geschlecht und familiärem Hintergrund können zudem durch Unterschiede in Erfolgserwartung und Wertüberzeugungen erklärt werden. Vor diesem Hintergrund haben sich Interventionen zur Förderung der wahrgenommenen Relevanz der Lerninhalte als eine Möglichkeit zur Förderung von Motivation und Leistung erwiesen, insbesondere für „Risikogruppen". Im vorliegenden Beitrag wurde daher anhand von Daten einer großen Cluster-randomisierten Studie zur Testung einer Relevanzintervention in Mathematik mit 78 neunten Klassen (Gaspard et al., 2021) untersucht, ob die Interventionseffekte durch demografische Schüler*innenmerkmale moderiert wurden. Es zeigten sich kleine bis mittlere Unterschiede in Motivation und Leistung vor der Intervention in Abhängigkeit von Geschlecht, elterlicher Bildung und Migrationshintergrund. Jedoch ergab sich wenig Evidenz dafür, dass diese Schüler*innenmerkmale die Effekte der Intervention moderierten. Die Intervention scheint somit nicht dazu beizutragen, Bildungsdisparitäten zu reduzieren – zumindest nicht, wenn sie wie hier für alle Schüler*innen gleichermaßen angewandt wird.

Schlagworte: Bildungsdisparitäten; Geschlecht; elterliche Bildung; Intervention; Mathematik; Migrationshintergrund; Motivation

1. Introduction

When students perceive mathematics as valuable to themselves, they are more engaged in learning and show higher performance in mathematics, and they also choose to pursue math-related courses and careers more often (e. g., Guo et al., 2018; Marsh et al., 2005; Simpkins et al., 2006; Song et al., 2020). However, many adolescents do not see the relevance of mathematics for their lives, and their value beliefs tend to decrease across secondary school (Gaspard et al., 2022; Harackiewicz et al., 2010; Watt, 2004). Given the relevance of mathematical competencies for many careers in- and outside of science, technology, engineering, and mathematics (STEM; Anger et al., 2021; Tenorth, 2001), it is thus important to find ways to help students see the value of mathematics so that they keep motivated and perform up to their potential. Some students are at a particularly high risk for low motivation and

achievement in mathematics. The risk factors that have been identified in prior research include being female (e. g., Gaspard et al., 2022; Jansen & Stanat, 2015), having parents with a relatively low level of education (e. g., Simpkins et al., 2006; Svoboda et al., 2016), and having a migration background (e. g., Jansen & Stanat, 2015; Stanat & Christensen, 2006). Furthermore, differences in students' motivational beliefs, including their task values, according to their gender and family background can help to explain corresponding differences in students' math-related career choices (e. g., Guo et al., 2015, 2018; Kriegbaum & Spinath, 2016). Therefore, promoting students' value beliefs in mathematics can be a possible avenue toward fostering students' educational outcomes in math-related areas in general and toward reducing educational disparities in this domain.

Drawing on (situated) expectancy-value theory (Eccles et al., 1983; Eccles & Wigfield, 2020), researchers have developed and tested interventions that are aimed at promoting students' motivation by targeting the utility value or perceived relevance of the learning content for students' lives. Such relevance interventions (or utility value interventions, see Hulleman & Harackiewicz, 2021) can have positive effects on students' motivation, achievement, and course choices. Moreover, the effects of these interventions have often been found to be larger for students who are at risk for low motivation in the particular context (e. g., Gaspard, Dicke, Flunger, Brisson, et al., 2015; Harackiewicz et al., 2016; Shin et al., 2022; Weidinger et al., 2022). Although these findings suggest that such interventions can be a useful tool for educational practice, many previous studies were conducted with a relatively high level of control exerted by the researchers. Therefore, research is still needed to test whether positive effects can be found under conditions that are closer to educational practice. Furthermore, many previous studies relied on relatively small samples, which make it difficult to investigate the heterogeneity of intervention effects across different student characteristics. Therefore, in this study, we used data from a large effectiveness trial in 78 classrooms to examine whether heterogeneity in the effects of the Motivation in Mathematics (MoMa) intervention, a relevance intervention for ninth-grade math classrooms, depended on students' gender, migration background, or parental education (for the main effects of the intervention, see Gaspard et al., 2021).

1.1. Expectancy-Value Theory as an Approach for Explaining Differences in Achievement-Related Choices

Relevance interventions such as the MoMa intervention can be grounded in Eccles et al.'s (1983) expectancy-value theory, which has recently been renamed situated expectancy-value theory (SEVT; Eccles & Wigfield, 2020). According to this theory, students' academic choices in a specific domain are most directly predicted by their expectancies of success and their subjective task values in this domain. Expectancies of success are conceptually close to other constructs that refer to competence-related beliefs, such as self-concept (a student's subjective evaluation of their ability in a specific domain) and self-efficacy (a student's belief about their ability to perform a specific task at a designated level; Wigfield & Eccles, 2000). However, these different beliefs are typically highly correlated and not always separated in empirical research (Eccles & Wigfield, 2020). In line with our theoretical background, we therefore use the term expectancies to refer to all competence-related beliefs. With respect to subjective task value, Eccles and colleagues (1983) distinguished four major components: intrinsic value (enjoyment of a given domain), attainment value (the personal importance of doing well in a domain), utility value (the perceived usefulness of a domain for achieving one's goals), and cost (perceived negative consequences of engaging in a domain). Whereas the first three components contribute positively to subjective task value, cost reduces it.

Prior research has successfully used SEVT to explain educational disparities in achievement-related behaviors through their motivational beliefs. In particular, much attention has been devoted to gender differences in mathematics. In line with typical gender stereotypes, female adolescents report lower expectancies and also tend to report lower values in this domain than their male peers (e. g., Gaspard et al., 2022; Marsh et al., 2005; Watt, 2004). However, gender differences in math values have been found to depend on the value component that is considered, and the most consistent differences have been found for intrinsic value (e. g., Gaspard, Dicke, Flunger, Schreier, et al., 2015; Watt, 2004). Such gender differences in motivational beliefs can help explain gender differences in math-related course and career choices (e. g., Gaspard et al., 2019; Guo et al., 2015, 2018; Simpkins et al., 2006).

SEVT can also contribute to explaining disparities in students' achievement-related choices on the basis of their family background. Eccles and colleagues postulated that family and parent characteristics (e. g., parental education) impact their children's expectancies, values, and achievement-related choices through parents' beliefs and behaviors (Eccles & Wigfield, 2020). These assumptions have largely been supported. For instance, Svoboda et al. (2016) found that parental education predicted mathematics and

science course taking in high school and that this prediction was partially mediated through parents' and students' motivational beliefs about these domains. Similarly, in a longitudinal addition to the Programme for International Student Assessment (PISA) in Germany, Kriegbaum and Spinath. (2016) found that students' math competence beliefs and interest mediated the associations between parents' socioeconomic status and students' math achievement. However, prior evidence on associations between family characteristics (e. g., parental education and socioeconomic status) and students' motivational beliefs in mathematics remains mixed (studies reporting significant associations: Guo et al., 2015; Häfner et al., 2017; Simpkins et al., 2006; studies reporting nonsignificant assocations: Gladstone et al., 2018; Weidinger et al., 2022).

Another important family background factor concerns the question of whether the students' family has a migration background. In Germany, as in many other countries, students with migration background tend to show pronounced gaps in their math achievement compared with students without migration background (Jansen & Stanat, 2015; Stanat & Christensen, 2006). One might therefore assume that they will also report lower motivational beliefs. However, in an analysis of the PISA data, Stanat and Christensen (2006) found that students with migration background reported higher interest and enjoyment, instrumental motivation, and self-concept in mathematics compared with students without migration background in most countries, especially if they were first-generation immigrants. This pattern of findings can be explained by the immigrant optimism hypothesis, which postulates that individuals with migration background are especially motivated to be successful in school and thus to improve their social position (Kao & Tienda, 1995). However, Stanat and Christensen (2006) also found that students with a migration background reported higher levels of math anxiety and lower self-efficacy in solving specific math tasks. Similarly, in a representative large-scale assessment with ninth-grade students in Germany, Jansen and Stanat (2015) found no differences between students with and without a migration background in their math self-concept, but students with a migration background reported higher math interest. Finally, a study with seventh-grade students in Switzerland (Brandenberger et al., 2017) showed that students with a migration background reported higher intrinsic motivation and enjoyment in mathematics along with higher anxiety and no differences in self-concept.

1.2. Prior Research on Relevance Interventions

In recent years, researchers have developed targeted interventions grounded in SEVT to promote students' motivation and help at-risk students. Such interventions have focused on utility value because it is assumed to be more extrinsic in nature (Eccles & Wigfield, 2020) and should thus be easier to affect through external interventions than the other components (Gaspard, Dicke, Flunger, Brisson, et al., 2015; Hulleman et al., 2010). Even though the main target of these interventions is utility value, broader effects can also occur, including effects on intrinsic and attainment value (Gaspard, Dicke, Flunger, Brisson, et al., 2015; Hulleman et al., 2010). Therefore, we use the term relevance intervention to denote that these interventions rely on mechanisms that go beyond utility value by targeting relevance as "a personally meaningful connection to the individual" (Priniski et al., 2018, p. 12) more broadly.

Heterogeneity in the effects of relevance interventions has been found to depend on students' characteristics. Most studies have found that these interventions work best for students who are traditionally underrepresented in a particular educational context. Concerning gender as a moderator, the effects of a relevance intervention in fifth and sixth grade science classrooms on utility value, appreciation, and science-related career intentions were found to be larger for girls than boys (Shin et al., 2022). Similarly, in the first test of the MoMa intervention in mathematics in secondary school, Gaspard, Dicke, Flunger, Brisson, et al. (2015) found that female adolescents benefitted more from the intervention with respect to their values than male adolescents. In a parent-based intervention aimed at increasing the perceived relevance of mathematics and science for high school students, the intervention was most effective in increasing STEM course-taking for high-achieving girls and low-achieving boys, whereas it did not help low-achieving girls (Rozek et al., 2015).

For family background, findings have also suggested that at-risk students benefit the most from these interventions. In the first MoMa study, Häfner et al. (2017) investigated several family background variables as potential moderators and showed that students from families with lower motivational resources (e. g., lower family interest in mathematics) benefitted more from the intervention than students from other families. Parental education and migration background did not moderate the intervention effects in their study. However, in a similar study that was also conducted in ninth-grade math classrooms, Weidinger et al. (2022) found that the effects of a relevance intervention on attainment value were especially pronounced for academic track students with migration background whose parents did not hold a university entrance certificate. Similarly, in a study of college students,

Harackiewicz et al. (2016) reported that underrepresented minority students whose parents did not attend college showed the largest intervention effects on course grades such that the achievement gap relative to continuing generation majority students was substantially reduced. Finally, in the parent-based relevance intervention targeting mathematics and science at high school described above, the effects did not vary depending on parents' educational background (Harackiewicz et al., 2012).

So far, there is limited research exploring the potential mechanisms for why different student characteristics moderate the effects of relevance interventions. Beyond lower initial motivation and achievement, female students, students with low parental education, and students with a migration background might bring more specific characteristics leading them to reap larger benefits from such interventions. Female students often aspire to careers that are not typically associated with mathematics (e. g., social careers) and they might lack role models in math-related careers (Lauermann et al., 2015; Stout et al., 2011). Moreover, research has shown that female students, students with low parental education, and ethnic minority students are more likely to pursue communal and interdependent goals in educational contexts, which are typically perceived to be less pronounced in STEM fields compared with other domains (Diekman et al., 2010; Harackiewicz et al., 2016; Stephens et al., 2012). Students from families with a low parental education and/or migration background might also talk less often with their parents about the relevance of mathematics because they have less access to this information (e. g., because of their lack of familiarity with the educational system). Finally, female students, students with low parental education, and students with a migration background might respond differently to the intervention materials (e. g., show higher engagement in writing), which could then explain why they benefit more from them (Harackiewicz et al., 2016; Nagengast et al., 2018).

1.3. MoMa Project: Testing Relevance Interventions in Math Classrooms

In this study, we focus on the MoMa intervention, which is a classroom-based relevance intervention designed for ninth-grade academic track students in Germany. This particular age group was chosen because students of this age group typically report low levels of math utility value (Gaspard et al., 2017; Harackiewicz et al., 2010) and – at the same time – should be able to reflect upon the relevance of mathematics for their future lives. The 90-min

intervention consists of a psychoeducational presentation for the whole class and relevance-inducing reflection tasks for the individual students (for more details, see Brisson et al., 2017; Gaspard et al., 2021; Gaspard, Dicke, Flunger, Brisson, et al., 2015). The psychoeducational presentation consists of two main parts. First, information about the importance of effort for achievement in mathematics and research results on students' self-concept of ability are presented. This part targets students' growth mindsets (i. e., their beliefs that one's skills are malleable and can be improved through effort; Dweck & Leggett, 1988) and their expectancies in order to avoid potential threatening effects of relevance information for struggling students. Second, the presentation includes various examples of the utility of mathematics for future education, career opportunities in different fields, and leisure time activities. This part thus directly communicates the usefulness of mathematics in different domains and is meant to prepare students for the relevance-inducing tasks. Finally, students work on the relevance-inducing tasks, in which they are asked to reflect on the personal relevance of mathematics for their lives to personalize and internalize the previously presented content.

In a first randomized controlled trial testing the MoMa intervention (Brisson et al., 2017; Gaspard, Dicke, Flunger, Brisson, et al., 2015), the intervention was implemented by researchers in the classroom. Two reflection tasks were tested in this study: Students were either asked to write an essay about the relevance of mathematics or they were presented with written quotations from interviews with young adults related to the relevance of mathematics and were asked to evaluate those with respect to their personal relevance. The intervention was shown to have positive effects on students' values, expectancies, teacher-rated effort, and achievement in mathematics compared with a waitlist control condition, with the quotations condition showing more extensive effects compared with the essay condition. The intervention was furthermore found to work better for students at risk for low motivation in this context: girls and students from families with low interest in mathematics (Gaspard, Dicke, Flunger, Brisson, et al., 2015; Häfner et al., 2017). The larger intervention effects for girls could potentially be explained by role model effects because the intervention was delivered solely by female researchers in this study.

In a follow-up study (MoMa 2) testing the effectiveness of this intervention under conditions that are closer to educational practice, an optimized version of this intervention including the quotations task was implemented by either trained master's students or the regular math teacher (Gaspard et al., 2021). Although positive effects of both intervention conditions were again found on utility value, the effects were smaller compared with the first

study. In the master's students condition, additional effects were observed for the importance of effort and talent for math achievement and a standardized math test. Unexpectedly, students in both intervention conditions also reported higher perceived cost compared with students in the control condition after the intervention. Overall, the pattern of effects was thus rather mixed. Gaspard et al. (2021) investigated students' expectancies as a moderating variable but did not find any consistent moderating effects. It therefore remains unknown whether any differences in the effects of the intervention depended on other student characteristics (e. g., gender or family background).

1.4. Research Questions

In this study, we used data from the large MoMa 2 study, in which the effectiveness of a relevance intervention was tested in a cluster-randomized trial with a total of 78 classrooms. Because of the large sample size (N = 1,744), the data are well-suited to investigate whether heterogeneity in the intervention effects depended on student characteristics. We examined the following research questions:

1) How do students' motivation and achievement in this context vary by gender, parental education, and migration background? We expected female students and students with a relatively low level of parental education to report lower motivation compared with male students and students with a relatively high level of parental education. On the basis of prior research, we expected students with a migration background to report higher positive value beliefs, but expected them to show no differences or lower levels with respect to their expectancies and to report higher cost compared with students without migration background.
2) Do the effects of the intervention differ by students' gender, parental education, or migration background? If any moderator effects could be found, we expected larger intervention effects for students at risk for low motivation.

2. Method

2.1. Sample and Procedure

Data for the MoMa 2 study were collected in 78 ninth-grade classrooms out of 28 academic track schools in Baden-Württemberg from October 2017 to March 2018 (for more details on the study design, see Gaspard et al., 2021). The sample size was determined by a power analysis for a multisite cluster-randomized trial. Within each school, the participating classrooms were randomly assigned to three conditions: (a) intervention implemented by a master's student, (b) intervention implemented by the regular math teacher, or (c) waitlist control condition. Out of the 70 teachers, eight teachers participated with two classes each. To reduce the risk of diffusion effects, the randomization was based on the math teachers so that classes taught by the same teacher were assigned to the same condition. This randomization procedure resulted in 28 classes in the master's student condition, 25 in the teacher condition, and 25 in the waitlist control condition.

The Ethics Committee for Psychological Research at the University of Tübingen confirmed that the procedures were in line with ethical standards for research with human subjects. Student participation was voluntary and not incentivized, and parents and students had to provide written consent. Overall, 1,744 students participated in the study, corresponding to an 88.7 % participation rate (n = 629 in the master's student condition, n = 569 in the teacher condition, n = 546 in the waitlist control condition). Students' mean age was 14.63 years (SD = 0.48) at the beginning of the study. Overall, 53.8 % of the students were female, with a somewhat higher proportion of female students in the waitlist control condition (59.0 %) compared with the master's student (51.4 %) and teacher conditions (51.7 %). Furthermore, 31.7 % of the students had a migration background (i. e., the student or at least one parent was not born in Germany), and 72.9 % had at least one parent who obtained a general university entrance qualification. These proportions did not vary substantially between the experimental conditions (see Supplemental Materials). For migration background and parents' level of education, our sample was approximately representative of academic track students in Baden-Württemberg (Stanat et al., 2019; Statistisches Bundesamt [Destatis], 2018).

The study consisted of three waves of data collection. Students were administered questionnaires by trained research assistants before the intervention in October 2017 (pretest = T1), an average of 4 weeks (14–40 days) after the intervention in December 2017 (posttest = T2), and an average

of 3 months (11–17 weeks) after the intervention in February 2018 (follow-up = T3). Teachers were asked to rate their students' effort at the same time points.

2.2. Relevance Intervention

As described previously, the intervention was a 90-min lesson, which consisted of an instructor-led psychoeducational presentation for the whole classroom (~45 min) and a relevance-inducing task that the students worked on individually (~40 min; for more information, see Gaspard et al., 2021). In the individual task, the students were asked to read and evaluate six quotations from interviews with young adults describing situations in which mathematics was useful to them.

The master's students and teachers who implemented the intervention in the classroom received all the intervention materials from the research team, and these materials did not differ between conditions (for more information about training in both conditions, see Gaspard et al., 2021). A total of six master's students (five women and one man; age $M = 24.7$, $SD = 1.5$) delivered the intervention in the master's student condition (four to five classes per student). They were trained for this purpose as part of a two-semester class on motivation interventions in the master's program Education Sciences and Psychology at the University of Tübingen. The teachers ($n = 24$; 45.8 % women; age: $M = 40.2$, $SD = 9.8$; years of teaching experience: $M = 11.8$, $SD = 8.9$) assigned to the teacher condition were asked to participate in a 3-hr workshop in small groups to prepare to implement the intervention. One teacher in this condition declined to participate in the workshop and thus did not deliver the intervention in the classroom. We followed the intention-to-treat approach in our analyses and included this class in the teacher condition so that the random assignment remained intact (Sagarin et al., 2014).

2.3. Instruments

We investigated the same set of outcomes as Gaspard et al. (2021), except for a math test that had a large proportion of missing data. Students reported on their motivation in mathematics using the same set of items at pretest, posttest, and (to some extent) the follow-up. All items were rated on a Likert scale ranging from 1 (*completely disagree*) to 4 (*completely agree*). Sample

items and Cronbach's alpha for all rating scales are provided in Table 1 (see Supplemental Materials for the full set of items). Reliabilities were shown to be sufficient across experimental conditions, gender, parental education status, and migration background status (see Supplemental Materials).

Table 1: Sample Items and Reliabilities for Scales in All Waves of Measurement.

Scale	Sample item	# items	α_{T1}	α_{T2}	α_{T3}
Utility value	Math is very useful to me.	12	.88	.89	.89
Attainment value	It is important to me to be good at math.	6	.87	.88	.88
Intrinsic value	Math is fun to me.	3	.93	.93	.93
Cost	Doing math makes me really nervous.	9	.91	.93	.93
Importance of effort	I believe that working diligently is the most important thing in math.	4	.80	.84	–
Importance of talent	To be good at math, you need to have a talent for it.	3	.80	.83	–
Self-concept	I am good at math.	4	.91	.91	.89
Self-efficacy	I am convinced that I can achieve good results on math homework and tests.	4	.85	.90	.88
Effort	I do my best on math tasks.	3	.77	.84	.83
Teacher-rated effort	This student works thoroughly on all of his/her math tasks and homework assignments.	2	.75	.80	.80

Students' *values* were measured with a scale that allows for differentiation across not only the four value components but also several subfacets of utility value, attainment value, and cost (Gaspard, Dicke, Flunger, Schreier, et al., 2015). In this study, we focused on the major value components (for the effects of the intervention on the subscales, see Gaspard et al., 2021). Students' *mindsets* were assessed at pretest and posttest only. Students separately rated the importance of effort and the importance of talent for math achievement. Students' *expectancies* were measured with scales indicating students' self-concept and their self-efficacy in mathematics. Finally, students reported their *effort* in mathematics. Teachers rated individual students' math effort on two items using a Likert scale ranging from 1 (*completely disagree*) to 4 (*completely agree*).

Information on students' *previous math grades* was collected from school records (ranging from 1 = *very good* to 6 = *insufficient*). At the pretest and the follow-up, students worked on a 3 min 30 s normed *speed test* that measured their fluency in solving typical math operations with 50 questions. The sum score was used in the analyses. This speed test is part of the German mathematics test for Grade 9 (Schmidt et al., 2013). Validity studies showed that it is a very good proxy for students' achievement in longer assessments using standardized, curriculum-based math tests. The internal consistency was good at both time points (Kuder-Richardson 20 = .88).

Information about students' *gender* was provided by the schools. Students reported information about their family background. To assess *parents' level of education*, students were asked to report the highest secondary school leaving certificate that their parents had obtained. Parents' secondary school leaving certificates were then coded on the basis of whether they allowed for university entrance. We then combined mothers' and fathers' data into one variable indicating whether students had at least one parent with a general university entrance qualification (7.9 % missing data). Students were also asked about the country in which they and their parents were born. They were assigned a *migration background* if the student or at least one of their parents was not born in Germany (3.8 % missing data). No further information about the parents, such as their beliefs and behaviors, were available in the data.

2.4. Analyses

All analyses were conducted in Mplus 8.7 (Muthén & Muthén, 1998-2017). To deal with missing data, we used the full information maximum likelihood approach, which takes all available information into account. All continuous variables were standardized before running the analyses so that the regression coefficients presented here can be understood as effect sizes with respect to the total variance of the outcome variable. To deal with multiple testing, we applied Benjamini and Hochberg's (1995) procedure to control for a false discovery rate of .05.

To investigate whether mean-level differences in motivation and achievement depended on gender, parental education, and migration background prior to the intervention (Research Question 1), we conducted separate analyses for each indicator and student characteristic in which we regressed the respective indicator (e. g., utility value) on the respective student characteristic (e. g., gender) coded as a dummy variable. In these analyses, we accounted for the nesting of students within classrooms and schools using the design-based correction of standard errors implemented in Mplus (with type = *complex*).

To examine whether intervention effects were moderated by gender, parental education, or migration background (Research Question 2), we conducted two-level regression analyses that were based on the preregistered analyses for testing main effects reported by Gaspard et al. (2021). These analyses considered students at Level 1 and classrooms at Level 2 and additionally accounted for the nesting of classrooms within schools using

the design-based correction of standard errors. Each outcome variable was regressed on two dummy variables indicating the two intervention conditions as compared with the control condition at the classroom level. To increase the precision of the estimated intervention effects, we additionally included the pretest score of the respective outcome variable as a predictor at the student and classroom levels. In addition, we included as covariates any variables for which substantial differences were found between the experimental conditions before the intervention ($\delta > 0.05$; Gaspard et al., 2021; What Works Clearinghouse, 2020). We therefore included prior grades, intrinsic value, cost, self-concept, self-efficacy, effort, achievement score, and teacher-rated effort as covariates at both levels. We added our student characteristics of interest (i. e., gender, parental education, migration background) as predictors on both levels. To test for moderator effects, we additionally included two cross-level interactions between the respective student characteristics and the two intervention conditions. The moderating effects of the different student characteristics were tested in separate models. To keep the models as parallel as possible, all three student characteristics were included as covariates in all models. The effects of the covariates at both levels were freely estimated to account for contextual effects (Marsh et al., 2009). The covariates at the student level were group-mean-centered, and manifest aggregation was used for the class-level predictors (Marsh et al., 2009). To examine whether the study was sufficiently powered to detect such moderator effects, we estimated minimum detectable effect size differences (MDESDs) based on the empirical data in MoMa 2. Our power analyses showed that the MDESDs for all three binary moderator (i. e., gender, parental education, and migration background) and all outcome variables varied between 0.176 and 0.273, with an average of 0.225 (see Supplemental Materials for more details).

3. Results

3.1. Descriptive Statistics

Descriptive statistics (*n, M, SD*) of all study variables depending on the experimental conditions and demographic student characteristics in the different waves of measurement as well as the correlations between all variables can be found in the Supplemental Materials.

3.2. Mean-Level Differences in Motivation and Achievement Depending on Gender, Parental Educational, and Migration Background Prior to the Intervention

To address Research Question 1, we first tested for whether mean-level differences in motivation and achievement depended on students' gender, parental education, and migration background prior to the intervention. With respect to gender (see Table 2), we found that girls reported lower utility and intrinsic value, higher cost, and lower self-concept and self-efficacy than boys. No differences were found for students' mindsets. Girls reported greater effort and were also rated as showing greater effort by their teachers. No significant differences were found for achievement. The effects were small to moderate in size, with the largest differences found for teacher-rated effort and self-efficacy.

Table 2: Mean-Level Differences in Motivation and Achievement Depending on Gender.

	Boys		*Girls*					
Variable	*M*	*SD*	*M*	*SD*	*b*		*SE*	*p*
Utility value	2.88	0.53	2.80	0.47	-0.15	*	0.06	.007
Attainment value	2.89	0.62	2.86	0.59	-0.05		0.06	.376
Intrinsic value	2.54	0.85	2.41	0.87	-0.15	*	0.06	.010
Cost	2.07	0.67	2.17	0.70	0.15	*	0.06	.010
Self-concept	2.93	0.71	2.65	0.79	-0.36	*	0.06	<.001
Self-efficacy	2.91	0.58	2.65	0.63	-0.42	*	0.06	<.001
Importance of effort	2.86	0.69	2.81	0.62	-0.09		0.05	.109
Importance of talent	2.22	0.72	2.26	0.71	0.06		0.06	.334
Effort	3.19	0.60	3.37	0.55	0.30	*	0.06	<.001
Teacher-rated effort	2.82	0.80	3.19	0.73	0.47	*	0.06	<.001
Test score	29.41	8.00	28.71	7.45	-0.09		0.05	.090

Note. Regression coefficients can be interpreted like Cohen's d.

*significant after correcting for a false discovery rate of .05.

Concerning parental education, we found only a few significant differences (see Table 3). Students whose parents held a university entrance qualification reported higher utility value and self-efficacy than students whose parents had no such qualification. There was also a tendency for the same pattern of differences for self-concept and test scores, but these differences were not significant after applying the Benjamini-Hochberg correction.

Table 3: Mean-Level Differences in Motivation and Achievement Depending on Parental Education.

	No university entrance qualification		University entrance qualification					
Variable	*M*	*SD*	*M*	*SD*	*b*		*SE*	*p*
Utility value	2.79	0.50	2.86	0.50	0.13	*	0.05	.009
Attainment value	2.87	0.60	2.88	0.60	0.02		0.05	.782
Intrinsic value	2.42	0.86	2.50	0.85	0.10		0.07	.147
Cost	2.18	0.70	2.09	0.68	-0.12		0.07	.070
Self-concept	2.70	0.74	2.82	0.77	0.16		0.07	.019
Self-efficacy	2.67	0.61	2.81	0.63	0.22	*	0.06	.001
Importance of effort	2.81	0.64	2.84	0.66	0.04		0.06	.439
Importance of talent	2.29	0.73	2.23	0.70	-0.08		0.06	.155
Effort	3.29	0.56	3.30	0.58	0.02		0.05	.768
Teacher-rated effort	2.97	0.79	3.05	0.78	0.10		0.07	.137
Test score	28.35	7.41	29.39	7.75	0.14		0.06	.020

Note. Regression coefficients can be interpreted like Cohen's d.

* significant after correcting for a false discovery rate of .05.

With respect to migration background, we found that students with migration background reported higher attainment value but also higher cost, lower self-concept, and lower self-efficacy compared with students without migration background. Moreover, students with migration background rated the importance of effort as higher and were rated as showing lower effort compared with students without migration background. No significant differences were found for the other variables, including achievement.

Table 4: Mean-Level Differences in Motivation and Achievement Depending on Migration Background.

	No migration background		Migration background					
Variable	*M*	*SD*	*M*	*SD*	*b*		*SE*	*p*
Utility value	2.84	0.50	2.83	0.52	0.00		0.05	.984
Attainment value	2.85	0.60	2.92	0.61	0.13	*	0.05	.017
Intrinsic value	2.49	0.87	2.45	0.85	-0.05		0.05	.384
Cost	2.09	0.68	2.21	0.70	0.18	*	0.06	.002
Self-concept	2.83	0.77	2.67	0.75	-0.21	*	0.05	<.001
Self-efficacy	2.80	0.62	2.69	0.63	-0.17	*	0.05	<.001
Importance of effort	2.80	0.65	2.91	0.67	0.15	*	0.06	.005
Importance of talent	2.27	0.70	2.20	0.74	-0.10		0.06	.085
Effort	3.28	0.57	3.29	0.61	0.01		0.06	.888
Teacher-rated effort	3.07	0.78	2.93	0.78	-0.18	*	0.06	.004
Test score	29.15	7.75	29.02	7.52	-0.02		0.06	.764

Note. Regression coefficients can be interpreted like Cohen's d.

* significant after correcting for a false discovery rate of .05.

3.3. Gender, Parental Educational, and Migration Background as Moderators of the Intervention Effects

To address Research Question 2, we tested whether the effects of the two intervention conditions (as compared with the waitlist control condition) were moderated by students' gender, parental education, and migration background (see Tables 5-7). We report only the regression coefficients of the intervention conditions and the cross-level interaction terms in the tables. When interpreting the effects of the intervention conditions, it needs to be noted that these refer to the group of students coded as zero on the moderator variable (i. e., male students, students with low parental education, or students without migration background). The full models with all predictors at the student and classroom levels can be found in the Online Supplemental Materials. In summary, after we applied the Benjamini-Hochberg correction, no significant interaction terms were found in any of the models. Still, we describe the interaction terms with (uncorrected) $p < .10$ to provide a clearer understanding of the potential pattern of results across models.

Concerning gender, there was a negative interaction with the teacher condition for intrinsic value at follow-up (b = -0.17, p = .024), resulting in a negative effect of the teacher condition for girls (b = -0.15, p = .026) but a nonsignificant effect for boys (b = 0.02, p = .775). Intervention effects of importance of effort in the master's student condition tended to be larger for girls (b = 0.16, p = .052), resulting in significant positive effects for girls (b = 0.23, $p < .001$) but not for boys (b = 0.07, p = .310). There was also a tendency toward a negative interaction with the teacher condition for importance of talent (b = -0.17, p = .96), although the subgroup-specific intervention effects were nonsignificant for girls (b = -0.08, p = .308) and boys (b = 0.10, p = .222). Finally, there was a tendency toward a negative interaction with the master's student condition for teacher-rated effort at posttest (b = -0.17, p = .085) so that the effect of this condition tended to be positive for boys (b = 0.14, p = .064) but not for girls (b = -0.03, p = .636).

Table 5: Interactions Between Gender and the Two Intervention Conditions for the Different Outcomes Assessed at Posttest and Follow-Up.

	Utility value			Attainment value			Intrinsic value			Cost		
Time & Interaction	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest												
Master's student condition	0.08	0.08	.318	0.05	0.08	.543	-0.02	0.08	.771	0.15	0.06	.013
Teacher condition	0.17	0.09	.067	0.06	0.08	.446	0.00	0.08	.969	0.14	0.08	.075
Gender x Master's student	0.11	0.11	.316	-0.03	0.09	.772	-0.10	0.08	.206	-0.08	0.08	.333
Gender x Teacher	0.00	0.13	.997	-0.04	0.09	.658	-0.05	0.08	.491	0.05	0.09	.567
Follow-Up												
Master's student condition	0.10	0.08	.218	-0.01	0.08	.890	-0.08	0.07	.306	0.13	0.06	.032
Teacher condition	0.12	0.08	.110	0.03	0.07	.674	0.02	0.07	.775	0.08	0.06	.181
Gender x Master's student	-0.01	0.11	.926	-0.03	0.09	.779	-0.06	0.08	.465	-0.03	0.08	.690
Gender x Teacher	-0.06	0.10	.560	-0.08	0.08	.310	-0.17	0.08	.024	0.05	0.08	.512

	Self-concept			Self-efficacy			Imp. of effort			Imp. of talent		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest												
Master's student condition	-0.01	0.05	.845	-0.10	0.10	.612	0.07	0.07	.310	-0.04	0.08	.622
Teacher condition	-0.07	0.06	.248	-0.06	0.11	.866	0.16	0.09	.082	0.10	0.08	.222
Gender x Master's student	-0.03	0.06	.555	0.05	0.10	.612	0.16	0.08	.052	-0.13	0.09	.154
Gender x Teacher	0.00	0.09	.958	0.02	0.11	.866	-0.04	0.11	.731	-0.17	0.10	.096
Follow-Up												
Master's student condition	0.05	0.07	.496	-0.01	0.07	.929						
Teacher condition	0.03	0.08	.709	-0.06	0.07	.426						
Gender x Master's student	-0.08	0.08	.316	-0.06	0.11	.615						
Gender x Teacher	-0.11	0.09	.214	-0.03	0.10	.744						

	Effort			Teacher-rated effort			Achievement		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest									
Master's student condition	-0.12	0.10	.226	0.14	0.07	.064			
Teacher condition	-0.11	0.10	.236	0.09	0.07	.169			
Gender x Master's student	0.08	0.10	.421	-0.17	0.10	.085			
Gender x Teacher	0.16	0.10	.111	0.01	0.09	.942			
Follow-Up									
Master's student condition	-0.07	0.08	.381	0.04	0.08	.603	0.08	0.08	.275
Teacher condition	-0.20	0.08	.012	-0.09	0.08	.246	0.10	0.09	.243
Gender x Master's student	0.01	0.08	.937	-0.13	0.10	.217	0.05	0.08	.542
Gender x Teacher	0.12	0.08	.130	0.12	0.10	.214	-0.03	0.08	.704

Note. Imp. = Importance. Gender was coded 0 = male, 1 = female.

There were no interaction terms significant at $p < .10$ for parental education. Concerning migration background, there was a tendency toward a negative interaction with the teacher condition for self-concept at posttest ($b = -0.15$, $p = .058$), resulting in a marginally significant negative intervention effect for students with a migration background ($b = -0.17$, $p = .053$) that could not be observed for students without a migration background ($b = -0.03$, $p = .662$). With respect to achievement as an outcome, interactions with both the master's student and teacher conditions tended to be positive ($b = 0.14$, $p = .065$, and $b = 0.13$, $p = .088$), so that positive intervention effects could be observed for students with migration background ($b = 0.21$, $p = .013$, and $b = 0.18$, $p = .037$) but not for students without migration background ($b = 0.07$, $p = .254$, and $b = 0.05$, $p = .464$).

Table 6: Interactions Between Parental Education and the Two Intervention Conditions for the Different Outcomes Assessed at Posttest and Follow-Up.

	Utility value			Attainment value			Intrinsic value			Cost		
Time & Interaction	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest												
Master's student condition	0.15	0.10	.145	0.06	0.08	.467	-0.09	0.08	.294	0.04	0.04	.542
Teacher condition	0.18	0.10	.061	0.02	0.09	.868	-0.02	0.09	.867	0.12	0.12	.158
Par. Educ. x Master's student	-0.01	0.10	.961	-0.04	0.08	.604	0.01	0.08	.915	0.10	0.08	.215
Par. Educ. x Teacher	-0.01	0.10	.936	0.04	0.10	.715	-0.02	0.08	.754	0.06	0.09	.542
Follow-Up												
Master's student condition	0.15	0.09	.091	-0.01	0.09	.880	-0.19	0.06	.002	0.11	0.07	.105
Teacher condition	0.15	0.09	.078	-0.02	0.09	.830	-0.15	0.07	.043	0.13	0.06	.035
Par. Educ. x Master's student	-0.07	0.10	.439	-0.02	0.09	.812	0.11	0.07	.115	0.01	0.08	.912
Par. Educ. x Teacher	-0.10	0.10	.345	0.01	0.09	.932	0.11	0.09	.232	-0.04	0.07	.604

	Self-concept			Self-efficacy			Imp. of effort			Imp. of talent		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest												
Master's student condition	-0.01	0.07	.866	-0.01	0.08	.922	0.12	0.09	.177	-0.14	0.10	.141
Teacher condition	-0.02	0.08	.791	-0.01	0.09	.884	0.07	0.10	.451	-0.10	0.09	.272
Par. Educ. x Master's student	-0.02	0.07	.721	-0.09	0.08	.263	0.06	0.10	.521	0.04	0.11	.740
Par. Educ. x Teacher	-0.08	0.08	.314	-0.05	0.08	.497	0.11	0.11	.303	0.14	0.11	.173
Follow-Up												
Master's student condition	0.01	0.08	.920	-0.05	0.08	.498						
Teacher condition	-0.07	0.08	.384	-0.12	0.09	.148						
Par. Educ. x Master's student	-0.02	0.08	.851	0.02	0.09	.822						
Par. Educ. x Teacher	0.04	0.08	.576	0.07	0.09	.409						

	Effort			Teacher-rated effort			Achievement		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest									
Master's student condition	-0.03	0.09	.738	0.02	0.08	.811			
Teacher condition	-0.03	0.10	.770	0.04	0.07	.530			
Par. Educ. x Master's student	-0.06	0.10	.590	0.04	0.08	.635			
Par. Educ. x Teacher	0.01	0.12	.927	0.06	0.09	.455			
Follow-Up									
Master's student condition	-0.04	0.10	.716	-0.02	0.09	.793	0.16	0.08	.049
Teacher condition	-0.05	0.09	.572	-0.06	0.08	.471	0.17	0.09	.044
Par. Educ. x Master's student	-0.03	0.09	.718	0.00	0.10	.989	-0.07	0.07	.305
Par. Educ. x Teacher	-0.12	0.11	.283	0.05	0.09	.568	-0.13	0.09	.140

Note. Par. Educ. = Parental Education; Imp. = Importance. Parental education was coded 0 = no university entrance qualification, 1 = university entrance qualification.

Table 7: Interactions Between Migration Background and the Two Intervention Conditions for the Different Outcomes Assessed at Posttest and Follow-Up.

	Utility value			Attainment value			Intrinsic value			Cost		
Time & Interaction	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest												
Master's student condition	0.16	0.07	.028	0.02	0.06	.672	-0.10	0.06	.093	0.13	0.05	.015
Teacher condition	0.21	0.07	.001	0.05	0.05	.363	-0.05	0.07	.486	0.13	0.06	.036
Mig. Back. x Master's student	-0.05	0.11	.648	0.02	0.10	.814	0.07	0.09	.414	-0.05	0.08	.528
Mig. Back. x Teacher	-0.13	0.12	.252	-0.03	0.10	.772	0.05	0.09	.533	0.08	0.09	.378
Follow-Up												
Master's student condition	0.06	0.08	.460	-0.05	0.07	.475	-0.12	0.06	.049	0.10	0.05	.043
Teacher condition	0.08	0.06	.177	-0.01	0.07	.843	-0.08	0.06	.132	0.12	0.05	.008
Mig. Back. x Master's student	0.14	0.11	.214	0.07	0.11	.529	0.05	0.05	.652	0.03	0.09	.728
Mig. Back. x Teacher	0.05	0.13	.702	0.00	0.12	.989	0.04	0.04	.636	-0.04	0.10	.675

	Self-concept			Self-efficacy			Imp. of effort			Imp. of talent		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest												
Master's student condition	-0.01	0.05	.878	-0.09	0.05	.863	0.14	0.06	.019	-0.15	0.05	.004
Teacher condition	-0.03	0.06	.662	-0.04	0.05	.052	0.14	0.08	.065	-0.02	0.05	.648
Mig. Back. x Master's student	-0.07	0.07	.321	0.07	0.09	.424	0.06	0.11	.563	0.14	0.13	.295
Mig. Back. x Teacher	-0.15	0.08	.058	-0.02	0.08	.791	0.04	0.10	.715	0.08	0.13	.567
Follow-Up												
Master's student condition	0.01	0.05	.834	-0.01	0.05	.863						
Teacher condition	-0.05	0.05	.319	-0.11	0.05	.052						
Mig. Back. x Master's student	-0.04	0.08	.645	-0.10	0.09	.312						
Mig. Back. x Teacher	0.04	0.10	.709	0.09	0.09	.294						

	Effort			Teacher-rated effort			Achievement		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Posttest									
Master's student condition	-0.07	0.08	.392	0.04	0.05	.500			
Teacher condition	0.00	0.08	.959	0.06	0.06	.303			
Mig. Back. x Master's student	-0.01	0.10	.906	0.03	0.09	.751			
Mig. Back. x Teacher	-0.09	0.10	.389	0.09	0.10	.384			
Follow-Up									
Master's student condition	-0.08	0.08	.292	-0.06	0.06	.359	0.07	0.06	.254
Teacher condition	-0.09	0.07	.230	-0.06	0.05	.272	0.05	0.06	.464
Mig. Back. x Master's student	0.08	0.12	.504	0.10	0.10	.275	0.14	0.08	.065
Mig. Back. x Teacher	-0.15	0.11	.154	0.10	0.09	.252	0.13	0.08	.088

Note. Mig. Back. = Migration Background; Imp. = Importance. Migration background was coded 0 = *no migration background*, 1 = *migration background*.

4. Discussion

Based on data from a large cluster-randomized trial that tested the effectiveness of a relevance intervention in ninth-grade math classrooms, we investigated whether heterogeneity in the results depended on students' demographic characteristics in their mean levels of motivation and achievement prior to the intervention and in the effects of the intervention on motivation and achievement. We found small to moderate differences in motivation and achievement that depended on students' gender, parental education, and migration background, indicating that female students, students with low levels of parental education, and students with migration background indeed constitute at-risk groups in this context. However, there was little evidence that the intervention effects were moderated through these demographic characteristics. We discuss these findings and their implications in more depth in the following.

4.1. Heterogeneity in Students' Mean Levels of Motivation and Achievement

Altogether, the preintervention motivation and achievement differences that depend on student characteristics were in line with previous research. With respect to gender, prior studies that examined gender differences in students' expectancies and values in German adolescents found similar patterns with more pronounced and consistent differences in students' expectancies (favoring boys) than in students' values (Gaspard et al., 2022; Gaspard, Dicke, Flunger, Schreier, et al., 2015; Marsh et al., 2005). Notably, there were no significant gender differences in test scores in our study so that gender differences in expectancies and values could not be explained by differences in achievement. However, girls showed higher self-reported and teacher-rated effort. The stereotypical picture of the "good girl" might actually undermine girls' perceived competence as they themselves and their socializers attribute their achievement successes to effort rather than to talent (Brandmiller et al., 2020; Tiedemann, 2000).

Concerning parental education, we found only a few, small differences, although the pattern was consistent in favoring students with a high level of parental education. This finding is in line with prior studies showing small differences in a similar context (Häfner et al., 2017). The fact that we did not find more pronounced differences in students' motivation and achievement might also be due to limited variation in our measure of parental education in this context. More than 70 % of the students had at least one parent who

had obtained a university entrance qualification. Information about tertiary education was not available. Future studies might consider using more continuous measures of parental education that also include tertiary education (Svoboda et al., 2016).

Finally, students with migration background reported higher attainment value along with higher cost and lower expectancies compared with students without migration background, whereas we did not observe differences in objective performance. This pattern is somewhat similar to the results of other studies that showed that students with migration background reported higher values and intrinsic motivation but also higher anxiety (Brandenberger et al., 2017; Stanat & Christensen, 2006). In fact, the combination of high values along with rather low expectancies might undermine students' well-being in these courses (Lauermann et al., 2017). Interestingly, students with migration background reported higher importance of effort, a finding that might be explained through the immigrant optimism hypothesis (Kao & Tienda, 1995). A study of college students in Belgium also showed that students with migration background reported a more growth-oriented mindset (Corradi et al., 2019). There were no differences in self-reported effort, but teachers gave lower ratings to the effort of students with migration background. This finding could be interpreted as a bias in teacher ratings (cf. Brandmiller et al., 2020).

4.2. Students' Demographic Characteristics did not Moderate Intervention Effects

Overall, our tests of students' characteristics as moderators of intervention effects yielded very little evidence of such moderator effects. Our findings thus do not support prior findings on more beneficial intervention effects of relevance interventions for girls in mathematics (Gaspard, Dicke, Flunger, Brisson, et al., 2015; Shin et al., 2022) or for students from families with a relatively low level of parental education and from underrepresented ethnicities or with migration background (Harackiewicz et al., 2016; Weidinger et al., 2022). Still, our findings are valuable, as we conducted a rigorous test of such moderator effects by considering multiple outcomes in a large sample. If at all, the moderator effects that tended to show up pointed toward more pronounced intervention effects for students at risk for low motivation and achievement in this context. However, more pronounced also meant stronger adverse and unintended effects of the intervention (e. g., reduced intrinsic values for girls). It thus seems that this "one size fits all" intervention did not

reduce gaps in motivation and achievement related to students' demographic characteristics – at least when implemented as a universal prevention as was done in this study. One could also argue that the intervention worked robustly for students bringing different characteristics. Indeed, the intervention was developed to target all ninth-grade academic track students and included multiple components to be able to target students with different expectancy levels and interests. A possible approach to reduce educational disparities could then be to implement the intervention selectively for students at risk for low motivation and achievement in this context such as female students and students from families with low parental education or migration background. However, it needs to be considered that the main effects of the intervention were rather small in this study, did not extend to all the outcomes considered, and positive effects were accompanied with adverse effects on some outcomes including cost. Furthermore, when selectively implementing interventions in the classroom context, it is important to make sure that the students "at risk" do not suffer from stereotype threat elicited through an expressed need for intervention. An alternative approach to increase the effectiveness of relevance interventions in reducing educational disparities might be to tailor them more specifically towards students at risk for low motivation and achievement. More research is also necessary to understand the mechanisms driving heterogeneity in intervention effects depending on students' characteristics. In the MoMa 1 study, where the intervention was found to be more beneficial for girls than for boys, only female researchers delivered the intervention in the classroom. It is thus possible that role model effects were the reason for this differential efficacy. In this study, the master's students and teachers who delivered the intervention were more diverse in terms of gender composition although most master's students were still female. Taking advantage of such role model effects, one possibility for targeting the intervention more toward individual students might be to present them with quotations from individuals who share the students' demographic characteristics.

4.3. Limitations

Although we conducted our analyses using data from a large cluster-randomized trial, our study also has some limitations. First, whereas the study design and the analyses for the main effects were preregistered, we did not preregister the analyses reported here. They should therefore be considered more exploratory, which is also why we corrected for multiple testing. Second, although the study was adequately powered to detect small main effects (see

Gaspard et al., 2021), the power to detect moderator effects for the considered binary student characteristics at the student level was lower. This means that the difference in the effects between groups would have had to be relatively large for us to be able to detect such differences with sufficient power. It is therefore possible that we did not detect all true moderator effects. Power issues also kept us from exploring intersections between the different student characteristics that we considered. As prior research has pointed towards the importance of considering such intersections (e. g., Harackiewicz et al., 2016; Jansen & Stanat, 2015; Weidinger et al., 2022), this is an important line for future research. Third, only limited information was available on students' family background. In the MoMa 1 study, parents' motivational beliefs related to mathematics were shown to be more closely associated with their students' motivational beliefs compared with broader characteristics such as parental education and these beliefs were also shown to moderate the intervention's effects (Häfner et al., 2017, 2018). However, we were not able to test if similar moderation effects could be found in the MoMa 2 study. Finally, and perhaps most importantly, our sample consisted only of academic track students in the state of Baden-Württemberg so that student diversity was limited. Future studies should therefore adapt and test relevance interventions in other contexts, such as less selective types of schools to test the generalizability of the findings for students who are more diverse in terms of their academic achievement and family background. Although almost one third of the students in our sample had a migration background, there was limited diversity in those students' countries of origin (see Supplemental Materials). This limited diversity kept us from exploring whether students from some countries of origin benefitted more from the intervention than those from others. Such heterogeneity could be explored in future research, for instance building on math-related country-level indicators stemming from international large-scale assessments.

4.4. Conclusion

To conclude, our study shows that there are disparities in academic track students' motivation and achievement in mathematics on the basis of students' gender, parental education, and migration background. However, our rigorous tests of potential moderating effects of a relevance intervention through these demographic characteristics provided little evidence of such moderator effects. It thus seems that this particular intervention was not effective in reducing disparities in motivation and achievement – at least when implemented as a universal prevention for all students as was done in this study.

Future research might therefore have to tailor these interventions more specifically toward different groups of students.

References

Anger, C., Kohlisch, E., Koppel, O., & Plünnecke. (2021). *MINT-Frühjahrsreport 2021 [STEM spring report 2021]*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Brandenberger, C. C., Hagenauer, G., & Hascher, T. (2017). Zum Zusammenspiel von Mathematikleistung, Schüler/innenmotivation und Lernemotionen auf der Sekundarstufe 1 – ein Vergleich zwischen Schüler/innen mit und ohne Migrationshintergrund in der Schweiz [On the interplay between mathematics achievement, student motivation and learning emtions in lower secondary school – a comparison between students with and without migration background in Switzerland]. *Schweizerische Zeitschrift für Bildungswissenschaften*, *39*(2), 375–394. https://doi.org/10.24452/sjer.39.2.5016

Brandmiller, C., Dumont, H., & Becker, M. (2020). Teacher perceptions of learning motivation and classroom behavior: The role of student characteristics. *Contemporary Educational Psychology*, *63*, Article 101893. https://doi.org/10.1016/j.cedpsych.2020.101893

Brisson, B. M., Dicke, A.-L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2017). Short intervention, sustained effects: Promoting students' math competence beliefs, effort, and achievement. *American Educational Research Journal*, *54*(6), 1048–1078. https://doi.org/10.3102/0002831217716084

Corradi, D., Nicolaï, J., & Levrau, F. (2019). Growth mindset and its predictive validity – do migration background and academic validation matter? *Higher Education*, *77*(3), 491–504. https://doi.org/10.1007/s10734-018-0286-6

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, *21*(8), 1051–1057. https://doi.org/10.1177/0956797610377342

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273. https://doi.org/10.1037/0033-295X.95.2.256

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 74–146). W. H. Freeman.

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, Article 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226–1240. https://doi.org/10.1037/dev0000028

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, *107*(3), 663–677. https://doi.org/10.1037/edu0000003

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology*, *48*, 67–84. https://doi.org/10.1016/j.cedpsych.2016.09.003

Gaspard, H., Nagengast, B., Trautwein, U., Jaekel, A.-K., & Göllner, R. (2022). Heterogenität in motivationalen Entwicklungsverläufen in Mathematik und Deutsch in Abhängigkeit von Schulform und Geschlecht [Heterogeneity in motivational developmental trajectories in mathematics and German depending on school type and gender]. *Zeitschrift für Erziehungswissenschaft*, *25*(2), 293–327. https://doi.org/10.1007/s11618-022-01092-1

Gaspard, H., Parrisius, C., Piesch, H., Kleinhansl, M., Wille, E., Nagengast, B., Trautwein, U., & Hulleman, C. S. (2021). The potential of relevance interventions for scaling up: A cluster-randomized trial testing the effectiveness of a relevance intervention in math classrooms. *Journal of Educational Psychology*, *113*(8), 1507–1528. https://doi.org/10.1037/edu0000663

Gaspard, H., Wille, E., Wormington, S. V., & Hulleman, C. S. (2019). How are upper secondary school students' expectancy-value profiles associated with achievement and university STEM major? A cross-domain comparison. *Contemporary Educational Psychology*, *58*, 149–162. https://doi.org/10.1016/j.cedpsych.2019.02.005

Gladstone, J. R., Häfner, I., Turci, L., Kneißler, H., & Muenks, K. (2018). Associations between parents and students' motivational beliefs in mathematics and mathematical performance: The role of gender. *Contemporary Educational Psychology*, *54*, 221–234. https://doi.org/10.1016/j.cedpsych.2018.06.009

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences*, *37*, 161–168. https://doi.org/10.1016/j.lindif.2015.01.008

Guo, J., Wang, M.-T., Ketonen, E. E., Eccles, J. S., & Salmela-Aro, K. (2018). Joint trajectories of task value in multiple subject domains: From both variable- and pattern-centered perspectives. *Contemporary Educational Psychology*, *55*, 139–154. https://doi.org/10.1016/j.cedpsych.2018.10.004

Häfner, I., Flunger, B., Dicke, A.-L., Gaspard, H., Brisson, B. M., Nagengast, B., & Trautwein, U. (2018). The role of family characteristics for students' academic outcomes: A person-centered approach. *Child Development*, *89*, 1405–1422. https://doi.org/10.1111/cdev.12809

Häfner, I., Flunger, B., Dicke, A. L., Gaspard, H., Brisson, B. M., Nagengast, B., & Trautwein, U. (2017). Robin Hood effects on motivation in math: Family interest moderates the effects of relevance interventions. *Developmental Psychology*, *53*(8), 1522–1539. https://doi.org/10.1037/dev0000337

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology*, *111*(5), 745–765. https://doi.org/10.1037/pspp0000075

Harackiewicz, J. M., Hulleman, C. S., Rozek, C. S., Katz-Wise, S., & Hyde, J. S. (2010). Parents' understanding of the utility value of STEM courses for high school students. *Paper Presented at the 2010 Biennial Meeting of the Society for Research on Adolescence*.

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science*, *23*, 899–906. https://doi.org/10.1177/0956797611435530

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*(4), 880–895. https://doi.org/10.1037/a0019506

Hulleman, C. S., & Harackiewicz, J. M. (2021). The utility-value intervention. In G. M. Walton & A. J. Crum (Eds.), *Handbook of wise interventions: How social psychology can help people change* (pp. 100–125). Guilford Press.

Jansen, M., & Stanat, P. (2015). Achievement and motivation in mathematics and science: The role of gender and immigration background. *International Journal of Gender, Science and Technology*, *8*(1), 4–18.

Kao, G., & Tienda, M. (1995). Optimism and achievement: The educational performance of immigrant youth. *Social Science Quarterly*, *76*(1), 1–19.

Kriegbaum, K., & Spinath, B. (2016). Explaining social disparities in mathematical achievement: The role of motivation. *European Journal of Personality*, *30*(1), 45–63. https://doi.org/10.1002/per.2042

Lauermann, F., Chow, A., & Eccles, J. S. (2015). Differential effects of adolescents' expectancy and value beliefs about math and English on math/science-related and human-services-related career plans. *International Journal of Gender, Science and Technology*, *7*(2), 205–228.

Lauermann, F., Eccles, J. S., & Pekrun, R. (2017). Why do children worry about their academic achievement? An expectancy-value perspective on elementary students' worries about their mathematics and reading performance. *ZDM*, *49*(3), 339–354. https://doi.org/10.1007/s11858-017-0832-1

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764–802. https://doi.org/10.1080/00273170903333665

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide. Eighth edition.* Muthén & Muthén.

Nagengast, B., Brisson, B. M., Hulleman, C. S., Gaspard, H., Häfner, I., & Trautwein, U. (2018). Learning more from educational intervention studies: Estimating complier average causal effects in a relevance intervention. *Journal of Experimental Education*, *86*(1), 105–123. https://doi.org/10.1080/00220973.2017.1289359

Priniski, S. J., Hecht, C. A., & Harackiewicz, J. M. (2018). Making learning personally meaningful: A new framework for relevance research. *Journal of Experimental Education*, *86*(1), 11–29. https://doi.org/10.1080/00220973.2017.1380589

Rozek, C. S., Hyde, J. S., Svoboda, R. C., Hulleman, C. S., & Harackiewicz, J. M. (2015). Gender differences in the effects of a utility-value intervention to help parents motivate adolescents in mathematics and science. *Journal of Educational Psychology*, *107*(1), 195–206. https://doi.org/10.1037/a0036981

Sagarin, B. J., West, S. G., Ratnikov, A., Homan, W. K., Ritchie, T. D., & Hansen, E. J. (2014). Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychological Methods*, *19*(3), 317–333. https://doi.org/10.1037/met0000013

Schmidt, S., Ennemoser, M., & Krajewski, K. (2013). *Deutscher Mathematiktest für 9. Klassen [German mathematics test for Grade 9]*. Hogrefe.

Shin, D. D., Lee, M., Jung, S. J., & Bong, M. (2022). Relative effects of classroom utility value intervention on the science motivation of girls and boys. *Research in Science Education*, Advance online publication. https://doi.org/10.1007/s11165-022-10070-w

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, *42*(1), 70–83. https://doi.org/10.1037/0012-1649.42.1.70

Song, J., Gaspard, H., Nagengast, B., & Trautwein, U. (2020). The Conscientiousness × Interest Compensation (CONIC) model: Generalizability across domains, outcomes, and predictors. *Journal of Educational Psychology*, *112*(2), 271–287. https://doi.org/10.1037/edu0000379

Stanat, P., & Christensen, G. (2006). *Where immigrant students succeed: A comparative review of performances and engagement in PISA 2003*. OECD.

Stanat, P., Schipolowski, S., Mahler, N., Weirich, S., & Henschel, S. (Eds.). (2019). *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich [IQB Educational Trend 2018: Mathematical and scientific competencies at the end of lower secondary school in the second state comparison]*. Waxmann. https://doi.org/10.25656/01:18131

Statistisches Bundesamt (Destatis). (2018). *Bildungsstand der Bevölkerung: Ergebnisse des Mikrozensus 2017 [Educational level of the population: Results of the microcensus 2017]*.

Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, *102*(6), 1178–1197. https://doi.org/10.1037/a0027143

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *100*(2), 255–270. https://doi.org/10.1037/a0021385

Svoboda, R. C., Rozek, C. S., Hyde, J. S., Harackiewicz, J. M., & Destin, M. (2016). Understanding the relationship between parental education and STEM course taking through identity-based and expectancy-value theories of motivation. *AERA Open*, *2*(3), 233285841666487. https://doi.org/10.1177/2332858416664875

Tenorth, H.-E. (Ed.). (2001). *Kerncurriculum Oberstufe: Mathematik – Deutsch – Englisch [Core curriculum upper school: Mathematics – German – English]*. Beltz.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, *92*(1), 144–151. https://doi.org/10.1037//0022-0663.92.U44

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, *75*(5), 1556–1574. https://doi.org/10.1111/j.1467-8624.2004.00757.x

Weidinger, A. F., Gaspard, H., Harackiewicz, J. M., Paschke, P., Bergold, S., & Steinmayr, R. (2022). Utility-value intervention in school: Students' migration and parental educational backgrounds as moderators. *The Journal of Experimental Education*, *90*(2), 364–382. https://doi.org/10.1080/00220973.2020.1855407

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook*. U.S. Department of Education, Institute of Education Sciences, National Center for Education and Regional Assistance. https://doi.org/10.1037/e578392011-004

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81. https://doi.org/10.1006/ceps.1999.1015

Die Rolle der Lehrkraft im Kontext von Selbstregulation beim Lernen

(Angehende) Lehrkräfte als selbstreguliert Lernende und selbstregulationsfördernde Lehrende

Charlotte Dignath & Antonia Fischer

Zusammenfassung

Lehrkräfte spielen eine zentrale Rolle für die Förderung von Selbstregulation beim Lernen (SRL). In Anlehnung an Modelle professioneller Lehrkräftekompetenz lassen sich darum Befunde zusammentragen zum Wissen von Lehrkräften über SRL und SRL-Förderung, sowie zu ihren professionellen Überzeugungen und ihren motivationalen Orientierungen mit Blick auf SRL-Förderung. Lehrkräfte haben im Kontext von SRL mehrere Rollen. So sind sie einerseits Fördernde der SRL ihrer Schüler*innen. Andererseits sind sie auch selbst lebenslange selbstregulierte Lernende, und – mit Blick auf ihre Unterrichtsplanung und -durchführung – auch selbstreguliert Unterrichtende. Es werden Forschungsbefunde dargestellt, die diese multiplen Rollen untersuchen. So zeigt eine Interventionsstudie mit Lehramtsstudierenden, wie sie ihre Selbstregulationskompetenzen im Zuge eines Kurztrainings weiterentwickeln. Eine Zusammenhangsstudie erklärt außerdem Unterschiede in der SRL-Förderung von Lehrkräften anhand von Aspekten professioneller SRL-Kompetenz sowie von Kontextmerkmalen der Klasse und der Schule. Das Kapitel endet mit Implikationen für weitere Forschung in diesem Bereich, sowie für die verschiedenen Phasen der Lehrkräftebildung, die Schulentwicklung und die schulische Praxis.

Schlagworte: Selbstregulation; Selbstreguliertes Lernen; Lehrkräfte; Kompetenz; Training

Abstract

Teachers play a central role in promoting self-regulation in learning (SRL). Therefore, following models of professional teacher competence, findings can be compiled on teachers' knowledge of SRL and SRL promotion, as well as on their professional beliefs and motivational orientations with regard to SRL promotion. Teachers have multiple roles in the context of SRL. On the one hand, they are promoters of their students' SRL. On the other hand, they are also lifelong self-regulated learners, and – with regard to their lesson planning and teaching – self-regulated teachers. Research findings that examine these multiple roles are presented. For example, an intervention study with pre-service teachers shows how they develop their SRL competencies in the course of a brief training. A correlational study also explains differences in teachers' SRL promotion based on aspects of professional SRL competence as well as classroom and school contextual characteristics. The chapter concludes with implications for further research in this area, as well as for the various stages of teacher education, school development, and school practice.

Keywords: self-regulation; self-regulated learning; teachers; competence; training

1. Einleitung

Mittlerweile gibt es zahlreiche Befunde, die deutlich machen, dass Selbstregulation beim Lernen (SRL) mit der Leistung zusammenhängt (Dent & Koenka, 2016) und einen positiven Einfluss auf Lernerfolg (Jansen et al., 2019) und Lernmotivation (Efklides, 2011) hat. Umso mehr stellt sich die Frage, warum SRL noch nicht standardmäßig im Unterricht der Grund- und weiterführenden Schule sowie an der Universität oder in der beruflichen Bildung gefördert wird und auch nicht fest in der Ausbildung von Lehrkräften verankert ist (Dignath & Veenman, 2021). Eine zentrale Rolle bei der Förderung von SRL spielen die Lehrkräfte (Hattie, 2008). In diesem Kapitel wird darum die professionelle Kompetenz von Lehrkräften bzgl. SRL genauer beleuchtet und ihre Rolle sowohl als selbstregulierte Lernende als auch als Lehrkräfte, die SRL fördern, untersucht. Hierzu werden sowohl empirische Befunde zur SRL von angehenden Lehrkräften als auch zur Vorhersage der SRL-Förderung durch verschiedene Aspekte der professionellen Kompetenz von Lehrkräften exemplarisch dargestellt. Zuletzt werden Implikationen für

die SRL-Forschung mit Lehrkräften sowie praktische Implikationen für Lehrkräfteaus- und Fortbildung dargestellt.

1.1. Selbstregulation beim Lernen

Bei SRL werden sowohl kognitive als auch affektiv-motivationale und behaviorale Prozesse beim Lernen aktiviert, um die Zielerreichung zu unterstützen (Pintrich, 2001). So orientieren sich selbstreguliert Lernende zu Beginn des Lernprozesses zunächst, um die Anforderungen der Lernaufgabe zu identifizieren und mit ihren Lernvoraussetzungen abzugleichen. Im nächsten Schritt setzen sie sich ein konkretes Ziel, das sie erreichen möchten, und planen die Schritte, die für die Zielerreichung erforderlich sind. Während der Aufgabenbearbeitung überwachen selbstreguliert Lernende ihren Lernfortschritt, monitoren ihre Aufmerksamkeit, und sorgen dafür, dass sie die Lernhandlung nicht unterbrechen, sondern zielführend weiterarbeiten. Nach Ablauf der Aufgabenbearbeitung evaluieren diese Lernenden ihren Lernfortschritt und vergleichen ihr Lernergebnis mit dem Lernziel. Dabei reflektieren sie, ob sie das Ziel erreicht haben oder noch weitere Arbeitsschritte notwendig sind (Zimmerman, 2000). Selbst sehr effiziente und leistungsstarke Lernende sind hin und wieder auf solche Regulationsprozesse angewiesen, um ihr Lernen am Laufen zu halten und zu verbessern (Greene, 2021).

Häufig setzen Lernende Strategien ein, um ihre Regulationsprozesse zu unterstützen (Dinsmore & Fryer, 2019). Dabei wirken sich kognitive Strategien direkt auf den Informationsverarbeitungsprozess aus, während metakognitive Strategien sich nur indirekt darauf auswirken, indem sie den Lernprozess initiieren, aufrechterhalten und kontrollieren. Daneben wirken motivationale und volitionale Strategien, die stark mit Emotionen verbunden sind, auf den Lernprozess ein (Boekaerts, 1999). Auch Strategien zur Emotionsregulation spielen zudem eine Rolle für eine effektive SRL (Dignath, 2022).

Der Einsatz von Selbstregulationsstrategien unterstützt das Lernen dabei über die Lebensspannen hinweg, bereits vom frühen Schulalter an, aber auch noch beim universitären bzw. beruflichen Lernen im Erwachsenenalter (Dent & Koenka, 2016).

1.2. Förderung von Selbstregulation beim Lernen

Was können Lehrende tun, um die SRL der Lernenden zu fördern? SRL kann sowohl indirekt aktiviert (situativer Ansatz) als auch direkt gefördert (kognitiver Ansatz) werden (Dignath, 2021; Dignath & Veenman, 2021). Aus kognitiver Sicht beinhaltet das Lernen die Speicherung und den Abruf von Wissen im und aus dem Langzeitgedächtnis sowie die Verarbeitung im Arbeitsgedächtnis. Die Theorie des *Cognitive Load* geht davon aus, dass komplexe Lernumgebungen, die zu viel Information enthalten, dem Lernen abträglich sind (Sweller & Chandler, 1991). Folglich sollte der Unterricht explizite Anleitung, direkte Instruktion und Scaffolding beinhalten, um eine kognitive Überlastung zu vermeiden (Kirschner et al., 2006). Im Hinblick auf die Förderung von SRL impliziert dies eine direkte Instruktion von SRL-Strategien. Die Forschung hat gezeigt, dass SRL-Strategietraining besonders effektiv ist, wenn die Lehrkräfte explizit den Nutzen von SRL-Strategien betonen und metakognitiv mit den Schüler*innen über deren Strategieeinsatz reflektieren (Dignath & Büttner, 2008). Zahlreiche Interventionsstudien konnten zeigen, dass Schüler*innen solche SRL-Strategien erlernen können und dass sich dies positiv auf ihren Lernerfolg auswirkt. Dabei profitieren auch Schüler*innen mit Lernschwierigkeiten (Hattie et al, 1996; Donker et al., 2014) und sehr junge Lernende (Dignath et al., 2008) bereits von Strategietrainings, aber auch bei Studierenden finden sich hohe Effekte für Selbstregulationstrainings (Theobald, 2021). Insbesondere eine Kombination verschiedener Strategiearten sowie eine explizite Vermittlung des Nutzens dieser Strategien haben sich dabei als besonders wirksam herausgestellt (Dignath & Büttner, 2008).

Daneben können Lehrkräfte die Lernumgebung derart gestalten, dass von den Schüler*innen SRL erfordert wird. Diese situativen Ansätze beinhalten Merkmale konstruktivistischer Lernumgebungen, die schülerzentriert sind und die Ko-Konstruktion von Wissen durch den Lernenden berücksichtigen. Mehrere Merkmale konstruktivistischer Lernumgebungen aktivieren indirekt die SRL. So sollte zum einen das Vorwissen der Schüler*innen aktiviert werden, um ihnen zu helfen, anspruchsvolle Ziele zu setzen und zu verfolgen. Außerdem sollte das Lernen in einem Anwendungskontext stattfinden. Bei abstrakten Lernangeboten ist es herausfordernd, die in der Schule gelernten Inhalte auf alltägliche Anwendungskontexte zu übertragen. Kooperative Lernangebote unterstützen zudem den Austausch unterschiedlicher Perspektiven. Insbesondere bei komplexen Lernaufgaben, die hohe mentale Anforderungen an die Schüler*innen stellen, lassen sich diese Anforderungen in kooperativen Settings auf die kognitiven Ressourcen mehrerer verteilen. Darüber hinaus geben schülerzentrierte Lernumgebungen den Lernenden die Möglichkeit, sich aktiv an Planung und

Durchführung der Lernaktivitäten zu beteiligen. Wenig vorstrukturierte Lernumgebungen gewähren ihnen somit mehr Autonomie durch freie Wahl des Lerninhalts, des Lernorts, der Kooperationspartner, des Schwierigkeitsgrads oder des Lerntempos (Perry & Rahim, 2011). Damit schaffen sie Übungsmöglichkeiten, die Lernende indirekt dazu ermutigen, ihr Lernen selbst zu regulieren.

Damit Schüler*innen und Studierende selbstregulierende Lernende werden, brauchen sie beides – eine direkte Instruktion von Selbstregulationsstrategien sowie Übungsgelegenheiten, um diese Strategien anwenden zu können (Dignath & Veenman, 2021). Zudem sollten kognitive und situative Lernansätze miteinander verbunden werden, um einen Conceptual Change bei den Lernenden zu unterstützen (Vosniadou, 2007). So hat sich beispielsweise problembasiertes Lernen als Unterrichtsstrategie als vorteilhaft für die Förderung von Conceptual Change erwiesen, wobei sich gezeigt hat, dass die Lerneffekte in problemorientierten Lernumgebungen höher sind, wenn diese auch eine direkte Instruktion von Strategien beinhalten (Wijnia et al., 2014). Damit Lernende bereit sind, das modellierte Verhalten nachzuahmen, müssen sie sich zudem ihrer naiven Überzeugungen und Fehlkonzepte bewusstwerden, und die Diskrepanz zu den wissenschaftlichen Konzepten erkennen, mit denen sie konfrontiert werden. Nur dann kann eine bewusste Veränderung von Konzepten stattfinden (Vosniadou, 2007). Nur wenig Forschung hat allerdings bisher die Rolle von Konzepten und Überzeugungen für SRL-Trainings bei Schüler*innen und Studierenden untersucht.

2. Professionelle Kompetenz von Lehrkräften bzgl. Selbstregulation

2.1. Selbstregulationskompetenz von Lehrkräften

Zur Einordnung der professionellen Kompetenz von Lehrkräften hinsichtlich der SRL nutzen wir das COACTIV-Modell (Baumert & Kunter, 2013) als ein allgemeines Modell der Lehrkräftekompetenz, das darauf abzielt, Unterschiede in der Unterrichtspraxis von Lehrkräften zu erklären und potenziellen Fortbildungsbedarf zu identifizieren. Während die *Unterrichtspraxis* als die Initiierung und Unterstützung von Lernprozessen durch die Lehrkräfte konzeptualisiert wird, die den Schüler*innen hilft, bestimmte Lernziele zu erreichen, wird *Unterrichtskompetenz* als ein multidimensionales Konstrukt betrachtet, das sowohl kognitive als auch motivational-affektive Aspekte umfasst, die Unterrichtspraxis vorhersagen (Baumert & Kunter, 2013). Zur

Erklärung der Unterrichtspraxis werden vier Kompetenzaspekte herangezogen: (1) das professionelle *Wissen* der Lehrkräfte, (2) ihre pädagogischen *Überzeugungen* und (3) *motivationale* Orientierungen, sowie (4) die *Selbstregulation* der Lehrkräfte (Kunter et al., 2013).

Lehrkräfte spielen eine entscheidende Rolle bei der Förderung von SRL, aber erst seit Kurzem untersucht die Forschung, ob und wie Lehrkräfte SRL im Unterricht fördern, und welche Merkmale von Lehrkräften ihre SRL-Praktiken beeinflussen. So ist bisher noch wenig darüber bekannt, was die Förderung von SRL vorhersagt (Lawson et al., 2019). Für eine systematische Untersuchung dieser Frage beziehen wir das COACTIV-Modell auf den Kontext der SRL-Förderung (Dignath, 2021), um SRL-Kompetenz von Lehrkräften umfassend zu untersuchen.

2.2. Professionelles Wissen über SRL und SRL-Förderung

Das professionelle Wissen von Lehrkräften zum Thema SRL lässt sich unterteilen in *inhaltliches Wissen* über das Konstrukt SRL sowie *didaktisches Wissen* über die Förderung von SRL (Dignath, 2021; Zohar et al., 2001). Die wenigen Studien, die das inhaltliche Wissen von Lehrkräften über SRL untersucht haben, zeigen, dass viele Lehrkräfte den funktionalen Wert von SRL-Strategien nicht erkennen, dass es ihnen generell an strategischem Wissen mangelt (Askell-Williams et al., 2012), dass Lehrkräfte viele Fehlkonzepte und fragmentiertes Wissen über SRL-Strategien haben (Glogger-Frey et al., 2018), und über nur geringes Wissen über metakognitive Strategien verfügen (Dignath & Büttner, 2018).

Mit Blick auf das didaktische Wissen von Lehrkräften über die Förderung von SRL zeigt sich, dass einige Lehrkräfte SRL hauptsächlich für leistungsstarke Schüler*innen als relevant erachten, obwohl Befunde zeigen, dass leistungsschwache Schüler*innen besonders von einer direkten Instruktion von SRL-Strategien profitieren (Zohar & Peled, 2008). Schließlich verfügen viele Lehrkräfte nur über begrenztes Wissen über pädagogische Konzepte zur Förderung von SRL (Zohar & Ben-Ari, 2022) und kennen zur SRL-Förderung vor allem indirekte Förderansätze, während sie nur selten an eine direkte Strategieinstruktion zur Förderung von SRL denken (Michalsky, 2014).

2.3. Professionelle Überzeugungen über SRL und SRL-Förderung

Die professionellen Überzeugungen zur Selbstregulation lassen sich einteilen in Überzeugungen, die konsistent und inkonsistent mit der wissenschaftlichen Vorstellung von SRL sind (Vosniadou et al., 2020). Solche Überzeugungen beinhalten auch die Vorstellung, ob SRL für Lernende geeignet ist. Neben diesen *Überzeugungen zu SRL als eine Form des Lernens* können Überzeugungen von Lehrkräften sich aber auch im Sinne pädagogischer *Überzeugungen auf die Förderung von SRL* beziehen (Lombaerts et al., 2009). Diese pädagogischen Überzeugungen lassen sich wiederum unterteilen in Überzeugungen zur direkten und zur indirekten SRL-Förderung. Sie beinhalten unter anderem Vorstellungen darüber, ob SRL für die Zielgruppe der eigenen Schüler*innen geeignet ist.

Untersuchungen zu den Überzeugungen von (angehenden) Lehrkräften über SRL zeigen, dass sie gleichzeitig sowohl Überzeugungen, die mit der SRL-Theorie vereinbar sind, als auch damit inkonsistente Überzeugungen (Fehlkonzepte) haben können (Darmawan et al., 2020). Die meisten Lehrkräfte haben eine positive Einstellung zur indirekten Förderung von SRL (Lombaerts et al., 2009), wobei solche konstruktivistischen Überzeugungen oft mit einem geringen Wissen über didaktische Mittel zur Förderung von SRL verbunden sind (Zohar et al., 2001).

2.4. Motivationale Orientierungen bzgl. SRL und SRL-Förderung

Die Überzeugungen von Lehrkräften hängen eng mit ihrer Motivation zusammen, selbstreguliert vorzugehen und Selbstregulation zu vermitteln. Nur sehr wenig Forschung hat sich bisher mit der Motivation von Lehrkräften für SRL und SRL-Förderung beschäftigt. Am häufigsten wurde bisher stellvertretend für Motivation die Selbstwirksamkeitserwartung von Lehrkräften, SRL erfolgreich zu fördern, untersucht (z. B. De Smul et al., 2018; Dignath, 2016, 2021; Perry & Rahim, 2011). In unserer aktuellen Forschung erweitern wir diesen Fokus und betrachten die SRL-Motivation aus einer Erwartungs-Wert-Perspektive (Eccles et al., 1983). Die Erwartungs-Wert-Theorie besagt, dass der Wert, den Lernende einer Lernaktivität zuschreiben, sowie der erwartete Erfolg bei der Ausführung der Aktivität das leistungsbezogene Verhalten vorhersagen. Aus der Erwartungs-Wert-Perspektive heraus sagen also die Selbstwirksamkeitserwartung der Lernenden, SRL-Strategien effektiv nutzen zu können, sowie ihr wahrgenommener Wert der Nutzung solcher SRL-Strategien voraus, ob Lernende SRL-Strategien anwenden (Zimmerman & Risemberg, 1997). Ebenso nehmen wir dies für die SRL-Förderung an: Nur,

wenn eine Lehrkraft eine ausreichend hohe Selbstwirksamkeitserwartung hat, SRL in ihrem Unterricht effektiv fördern zu können, und SRL zudem für ihre Schüler*innen für nützlich hält, sind die motivationalen Voraussetzungen dafür gegeben, dass die Lehrkraft SRL in ihrem Unterricht auch fördert.

Selbstwirksamkeitserwartungen von Lehrkräften umfassen ihren Glauben an die eigene Fähigkeit, komplexe Aufgaben zu bewältigen, und gelten als wichtiger Faktor, der das menschliche Handeln im Allgemeinen (Bandura, 1997) und das Unterrichtsverhalten der Lehrkräfte sowie viele andere damit verbundene Ergebnisse im Besonderen beeinflusst (Tschannen-Moran et al., 1998). Die Selbstwirksamkeitserwartungen einer Lehrkraft bestimmen, wie viel Mühe und Ausdauer sie in das Erreichen eines bestimmten Ziels investiert (Gregoire, 2003), und beziehen sich im Hinblick auf die Förderung von SRL auf die Überzeugung der Lehrkräfte, dass sie in der Lage sind, ihre Schüler*innen für SRL zu begeistern (De Smul et al., 2018). Sie stellen somit einen wichtigen Motivationsfaktor für das Unterrichtsverhalten von Lehrkräften dar (Dignath, 2016). Die Forschung zeigt, dass die Selbstwirksamkeit von Lehrkräften zu den stärksten Prädiktoren für die SRL-Praxis von Lehrkräften gehört (Dignath, 2021). Lehrkräfte, die SRL nicht oder nur unzureichend fördern, fühlen sich oft nicht selbstwirksam genug, SRL anzuregen (Vandevelde et al., 2012). Trotz positiver Überzeugungen zu SRL fühlen sich viele Lehrkräften nicht in der Lage, die SRL ihrer Schüler*innen zu unterstützen (Perry et al., 2008). Das Wissen der Lehrkräfte kann sich auf die Entwicklung ihrer Selbstwirksamkeit auswirken: Lehrkräfte, die der Meinung sind, dass sie über grundlegendes Wissen verfügen, entwickeln mit größerer Wahrscheinlichkeit ein hohes Gefühl der Selbstwirksamkeit (Morris et al., 2017).

Neben der Selbstwirksamkeitserwartung spielen zudem auch die Erwartungen einer Lehrkraft über den Nutzen von SRL eine Rolle. Im Gegensatz zur zunehmenden Forschungsbasis zur Lehrkräfte-Selbstwirksamkeit im Kontext von SRL gibt es bisher kaum Forschung dazu, für wie nützlich Lehrkräfte SRL für sich und für ihre Schüler*innen halten. Es gibt allerdings eine aktuelle Studie, die die Bedeutung des erwarteten Nutzens von SRL-Strategien für den SRL-Strategieeinsatz bei Lernenden zeigen konnte (Karabenick et al., 2021).

2.5. Selbstregulation: Lehrkräfte als selbstreguliert Lernende und Unterrichtende

Die Selbstregulation der Lehrkräfte kann sich sowohl auf die Regulation des eigenen Lernens als auch auf die Regulation des Unterrichtens beziehen. Lehrkräfte müssen selbst gute selbstregulierte Lernende sein – das heißt ähnliche SRL-Strategien einsetzen wie ihre Schüler*innen (z. B. Hilfe suchen von Peers und Mentoren) – um ihren Schüler*innen SRL-Strategien glaubwürdig und effektiv vermitteln zu können (Kramarski & Kohen, 2017; Peeters et al., 2014). Auch wenn die Trennung teilweise nicht deutlich ist, lassen sich dabei SRL-Strategien unterscheiden, die eher berufsorientiert sind (d. h. sich auf die Arbeit des Unterrichtens beziehen) von solchen, die eher lernorientiert sind (d. h. sich auf das Lernen für und vom Unterrichten beziehen; Randi, 2004). So sind auch Lehrkräfte lebenslange Lernende, die sich beispielsweise durch Hinzuziehen professioneller Literatur weiterbilden, wobei sie SRL-Strategien einsetzen können, in dem sie ihre Lernhandlungen planen, überwachen und reflektieren. Zudem können sie bei der Planung und der Durchführung ihres Unterrichts selbstreguliert vorgehen, indem sie sich konkrete Ziele setzen, passende Lernhandlungen planen, Instruktionsstrategien nutzen, um diese Handlungen anzuleiten und die Ziele zu erreichen, den Lernfortschritt der Schüler*innen überwachen und evaluieren, und ihr Vorgehen eventuell anpassen (Butler et al., 2004; Kramarski & Michalsky, 2009).

Insgesamt zeigt die bisher dünne Forschungsbasis zur Selbstregulation von Lehrkräften, dass mehr Forschung erforderlich ist, um den Vorhersagewert von Wissen, Überzeugungen, Selbstwirksamkeit und Selbstregulation von Lehrkräften im Kontext von SRL zu verstehen. Allerdings legen die Befunde nahe, dass die Förderung von SRL bereits im Lehramtsstudium zentral ist, damit Lehramtsstudierende von Beginn ihrer Ausbildung an selbstregulierte Lernende und später auch selbstregulierte Lehrende werden (Kramarski & Heaysman, 2021).

2.6. Befunde zum Zusammenspiel professioneller SRL-Kompetenz von Lehrkräften

Obwohl die meisten Studien zur SRL-Kompetenz von Lehrkräften bisher eher einen variablenzentrierten Ansatz verfolgt haben, indem sie spezifische Aspekte der Lehrerkompetenz separat untersuchten (z. B. Zohar & Ben-Ari, 2022 für Wissen; Lombaerts et al., 2009 für Überzeugungen; De Smul et al., 2018 für Selbstwirksamkeit; Kramarski & Kohen, 2017 für Selbstregulation), gibt es mittlerweile erste Studien, die das Zusammenspiel zwischen verschie-

denen Kompetenzaspekten prüfen. Diese Befunde legen nahe, dass manche Kompetenzaspekte, wie das Wissen der Lehrkräfte und ihre Überzeugungen über die Förderung von SRL, nicht unbedingt zusammenhängen (Dignath, 2016; Spruce & Bol, 2015), während andere, wie das Wissen der Lehrkräfte und ihre Selbstwirksamkeitserwartung, deutlichere Zusammenhänge aufweisen (Dignath, 2016; Karlen et al., 2020).

Neben dem Zusammenwirken einzelner Kompetenzaspekte spielt zudem die Frage eine Rolle, welche Kompetenzaspekte die Förderung von SRL im Unterricht vorhersagen können. Hier ist die Evidenz bisher recht uneinheitlich. Während die Überzeugungen von Lehrkräften in einigen Studien ein guter Prädiktor sind für ihre selbstberichtete (z. B. Thomas et al., 2022) bzw. die beobachtete SRL-Förderung (z. B. Depaepe et al., 2010) sind, zeigt die Überzeugung von Lehrkräften in anderen Studien keine Vorhersagekraft (z. B. Dignath & Büttner, 2018). Übereinstimmend stellen die Forscher fest, dass die Selbstwirksamkeit der Lehrkräfte bei der Umsetzung von SRL ihre selbstberichtete SRL-Praxis vorhersagt (Karlen et al., 2020). Überraschenderweise wurde dies nicht für das Wissen der Lehrkräfte über die Förderung von SRL festgestellt (Dignath & Van der Werf, 2012). Stattdessen scheint die Selbstwirksamkeitserwartung von Lehrkräften ein guter Prädiktor für ihre selbstberichtete (Dignath, 2016) und die von Schüler*innen berichtete SRL-Förderung (Dignath, 2021) zu sein.

3. Forschungsbefunde zur Rolle von Lehrkräften im Kontext von SRL

3.1. Ein Forschungsbeispiel zur Rolle von Lehrkräften als selbstregulierende Lernende

Modelling Examples und Refutation zur Anregung von Conceptual Change

Eine aktuelle Studie unserer Arbeitsgruppe zur Rolle von Lehrkräften als selbstregulierende Lernende zielte darauf ab, die SRL-Kompetenz von Lehramtsstudierenden zu verbessern und zu untersuchen, welche Form von Instruktion für diese Zielgruppe besonders wirksam ist. Als Interventionsinhalte wurden zentrale metakognitive SRL-Strategien (Planung, Überwachung, Evaluation) mithilfe direkter Strategieinstruktion vermittelt. Dabei kamen zielgruppengerechte Instruktionsvideos zum Einsatz, bei denen die Strategien

über ein Modell situationsspezifisch anhand der Situation des Vorbereitens einer Unterrichtsskizze eingeführt wurden.

Die Aktivierung von Modelllernen mit Hilfe von Vorbildern kann besonders zu Beginn des Studiums von Nutzen sein, da sich eine modellbasierte Instruktion als besonders effektiv für Novizen erwiesen hat (Renkl, 2014). Während sich Mastery-Modelle, bei denen ein Peer den Lernenden das korrekte Vorgehen demonstriert, eher für fortgeschrittene Lernende als wirksam erwiesen haben, haben sich Coping-Modelle insbesondere für unerfahrene Lernende bewährt, da sie eine leichtere Identifikation mit dem Rollenmodell ermöglichen, indem hier ein Peer das fehlerhafte Vorgehen darstellt, dies aber letztendlich ablegt und das korrekte Vorgehen vormacht (Van Gog & Rummel, 2010). Coping-Modelle bieten daher eine Möglichkeit, Fehlkonzepte zu aktivieren, um sie dann im Anschluss argumentativ zu widerlegen.

Im Gegensatz zu einseitiger Information zielt widerlegende Information (*Refutation*) darauf ab, einen Conceptual Change, d. h. einen konzeptionellen Wandel, bei den Lernenden anzuregen. Hierzu befasst sich die Instruktion explizit mit allgemein verbreiteten Fehlkonzepten zum Thema, die allerdings sofort widerlegt werden, und es wird eine geeignetere Erklärung angeboten (Hynd, 2001). Dabei werden die Fehlvorstellungen der Lernenden und die korrekte Information gleichzeitig aktiviert und die Lernenden dabei unterstützt, die Inkonsistenz zwischen ihren vorherigen Konzepten und der neuen Information zu erkennen. Dies führt schließlich zu einer Überarbeitung der Fehlvorstellungen und zu einem Conceptual Change (Kendeou et al., 2011). Darüber hinaus bieten Refutationen direkt eine Alternative, um die falsche Vorstellung zu ersetzen, indem sie den Lernenden ermöglichen, ihre Vorstellungen zu aktualisieren (Ecker et al., 2010). Schließlich ist Refutation dann effektiver, wenn erklärt wird, warum die Fehlvorstellung falsch war (Lewandowsky et al., 2005).

Für diese Studie haben wir uns sowohl die Mastery- und Coping-Modellierung als auch das Prinzip von Refutation zu Eigen gemacht, um den Nutzen des Aktivierens und Widerlegens von Fehlkonzepten über das Lernen für einen Conceptual Change bzgl. der SRL zu testen. Bei der Strategieinstruktion wurde das Adressieren von Fehlkonzepten bzgl. des Lernens variiert. Hierzu wurde in der einen Interventionsbedingung ein „*Mastery* Model" gezeigt, d. h. eine Studierende, die SRL-Strategien sehr gut einsetzt. In der anderen Interventionsbedingung wurde ein „*Coping* Model" gezeigt, d. h. eine Studierende, die zunächst über ihre uneffektiven Vorstellungen über das universitäre Lernen reflektiert (Aktivierung von Fehlkonzepten), und erst im zweiten Schritt von einer Freundin SRL-Strategien an die Hand

bekommt, die sie dann einsetzt und damit ihr Lernen erfolgreich zu Ende bringt (Refutation).

Hypothesen

Unsere Hypothesen umfassen, dass (1) die Teilnehmenden in den Interventionsgruppen (*Mastery* und *Coping Modelling*) nach dem Anschauen des Videos über eine ausgeprägtere SRL-Kompetenz verfügen, d. h. (a) mehr über SRL wissen, (b) mehr mit wissenschaftlichen Theorien konsistente Überzeugungen, (c) weniger inkonsistente Überzeugungen aufweisen (d. h. weniger Fehlkonzepte haben), (d) selbstwirksamer hinsichtlich der eigenen Nutzung von SRL-Strategien sind und (e) SRL-Strategien als nützlicher bewerten als die Kontrollgruppe.

Im Vergleich der beiden Interventionsgruppen vermuten wir (2) höhere Effekte für die *Coping*-Gruppe. Basierend auf Befunden zu Fehlkonzepten von (angehenden) Lehrkräften über das Lernen (z. B. Lawson et al., 2019) erwarten wir, dass die Teilnehmenden in der *Coping*-Gruppe (a) einen größeren Wissenszuwachs bzw. (b) eine Zunahme von Überzeugungen zum Lernen, die mit der wissenschaftlichen Theorie konsistent sind, bei (c) einer gleichzeitigen Abnahme inkonsistenter Fehlkonzepte aufweisen. Aufgrund der Aktivierung der Fehlkonzepte sowie der anschließenden Refutation im *Coping*-Video lässt sich davon ausgehen, dass hier ein ausgeprägterer Conceptual Change stattfindet als in der *Mastery*-Gruppe, in der nur die Strategien vermittelt werden, ohne zuvor die Fehlkonzepte zu adressieren (Gregoire-Gill et al., 2022). Darüber hinaus gehen wir (d) von einem höheren Anstieg der Selbstwirksamkeitsüberzeugungen zur Nutzung von SRL in der *Coping*-Gruppe aus. Dies ist darauf zurückzuführen, dass die Teilnehmenden in der *Coping*-Gruppe die Schwierigkeiten des Modells, von dem sie lernen, direkt miterleben, sich dadurch besser mit diesem identifizieren können und den Lernerfolg des Modells dann auch leichter für sich selbst als möglich erachten. Im Gegensatz dazu haben es die Studierenden in der *Mastery*-Gruppe schwerer, sich mit dem Modell zu identifizieren (Van Gog & Rummel, 2010). Zudem nehmen wir an, dass (e) die Studierenden der *Coping*-Gruppe SRL-Strategien für ihr Lernen als nützlicher erachten, da sie den Nutzen der Strategien beim Modell direkt beobachten konnten.

Methodisches Vorgehen

Um diese Hypothesen zu prüfen, wurde eine Interventionsstudie mit drei Gruppen, zwei Experimentalgruppen und einer aktiven Kontrollgruppe, durchgeführt. 157 Studierende (v. a. Lehramt, aber auch Psychologie, Erziehungswissenschaften und Soziologie) nahmen an der Studie teil. Im Schnitt waren diese 22.66 Jahre alt, 81 % waren weiblich.

Die Teilnehmenden beantworteten im Labor zunächst Fragebögen und Fragen mit offenem Antwortformat zu den verschiedenen Aspekten ihrer SRL-Kompetenz. Im Anschluss sahen die Studierenden eins von drei Videos. In den beiden Interventionsvideos wurde die Nutzung von SRL-Strategien, Überzeugungen in Bezug auf SRL-Strategien, die Nützlichkeitserwartung sowie die Selbstwirksamkeitserwartung in Bezug auf die Nutzung solcher Strategien durch beispielbasiertes Modelllernen (Van Gog & Rummel, 2010) thematisiert. Das Video der *Mastery*-Gruppe (n = 56) zeigte eine Studentin, die eine Aufgabe für ihr Lehramtsstudium sehr selbstreguliert angeht – d. h., sich zunächst einen Überblick über die Aufgabenstellung verschafft, ihr Vorgehen plant, ihr Verhalten während der Aufgabenbearbeitung überwacht und ihren Fortschritt im Nachhinein reflektiert. In der *Coping*-Gruppe (n = 53) sahen die Teilnehmenden dieselbe Studentin, die ihre Aufgabe zu Beginn jedoch wenig selbstreguliert angeht und sich viel unnötige Arbeit macht, was zu Frustration führt, woraufhin sie nach Lösungen sucht. Sie lernt etwas über SRL-Strategien, probiert diese aus und kommt dann zu einem zufriedenstellenden Ergebnis. Die Kontrollgruppe (n = 48) sah das Video einer Lehramtsstudentin, die von ihrem Praktikum in der Grundschule erzählt. Dieses Video sollte bewusst keinen Bezug zu SRL haben, aber eine vergleichbare Alternativ-intervention bieten. Im Anschluss erfassten wir erneut die SRL-Kompetenzen der Studierenden. Vier Wochen nach der Datenerhebung im Labor füllten die Teilnehmenden im Rahmen einer Follow-Up-Befragung erneut die Fragebögen zu ihren SRL-Kompetenzen online aus (N = 139).

Ergebnisse

Die Daten wurden mittels ANOVA mit Messwiederholung analysiert, bei der die drei Messzeitpunkte (Prä-, Post- und Follow-Up-Testung) als Innersubjektfaktor und die drei Gruppen als Zwischensubjektfaktor berücksichtigt wurden. Es fanden sich die folgenden signifikanten paarweisen Vergleiche ($p < 0.05$): (a) die eigene Nutzung von SRL-Strategien unterschied sich in der Coping-Gruppe signifikant zwischen Prä- und Posttestung (die Nutzung nahm zu), (b) die Nutzenüberzeugungen der Studierenden in Bezug

auf SRL-Strategien unterschied sich signifikant zwischen der Prä- und Posttestung in der Coping-Gruppe (die Nutzenüberzeugungen nahmen zu), (c) in der Kontrollgruppe stiegen die inkonsistenten Überzeugungen zu SRL der Studierenden sowohl zwischen Prä- und Post-Testung als auch zwischen Post- und Follow-Up-Testung signifikant an, während in den Interventionsgruppen ein signifikanter Rückgang zwischen Prä- und Post-Testung, aber wiederum ein signifikanter Anstieg zwischen Post- und Follow-Up-Testung zu verzeichnen war. Bezüglich (d) des Wissens schnitt die Coping-Gruppe zum Post- und Follow-Up-Test signifikant besser ab als die beiden anderen Gruppen ($p < 0.05$). Auch stieg in den Interventionsgruppen das Wissen über SRL zum Posttest signifikant an ($p < 0.05$), um dann zur Follow-Up-Testung wieder signifikant ($p < 0.05$) zurückzugehen. Die verschiedenen Studiengänge (Lehramt vs. Erziehungswissenschaft, Psychologie und Soziologie) unterschieden sich dabei bloß signifikant in ihren inkonsistenten Überzeugungen: Zum Prä- und Postzeitpunkt waren diese bei Lehramtsstudierenden geringer ausgeprägt als bei Studierenden der anderen Studiengänge.

3.2. Ein Forschungsbeispiel zur Rolle von Lehrkräften als Fördernde der Selbstregulation

Erklärungsfaktoren für die direkte und indirekte SRL-Förderung

Neben Fragen zur Entwicklung von SRL bei angehenden Lehrkräften untersuchen wir in unserer Forschung, wie die SRL-Kompetenz von Lehrkräften im Schuldienst mit ihrer Förderung von SRL im Unterricht zusammenhängt. Nur wenige Studien haben bisher systematisch untersucht, inwieweit das professionelle Wissen über SRL die SRL-Förderung erklärt. Dabei scheint es einen Unterschied zu machen, ob es um das Inhaltswissen über SRL geht oder um das didaktische Wissen, wie SRL gefördert werden kann. So finden sich in Studien, die das Inhaltswissen von Lehrkräften über SRL erfassten, keine oder nur geringe Zusammenhänge mit der SRL-Förderung der Lehrkräfte (Dignath, 2016), während das didaktische Wissen über die Vermittlung von SRL vereinzelt mit der SRL-Förderung der Lehrkräfte zusammenhängt (Karlen et al., 2020). Auch bzgl. des Zusammenhangs von Überzeugungen und SRL-Förderung sind die Befunde inkonsistent, was wiederum mit der Art von Überzeugungen und ihrer Operationalisierung zusammenhängen könnte. So finden sich signifikanten Zusammenhänge, wenn sowohl die Überzeugungen als auch die SRL-Förderung mit Fragebogen erfasst wurde (Heirweg et al., 2020), während sich in Unterrichtsbeobachtungsstudien nur geringe

bis keine Zusammenhänge fanden (Dignath & Büttner, 2018; Spruce & Bol, 2015). Demgegenüber sind die Befunde für die Selbstwirksamkeitserwartung von Lehrkräften, SRL effektiv fördern zu können, konsistent. So zeigen sich durchgängig positive Zusammenhänge zwischen der Selbstwirksamkeit und der selbstberichteten SRL-Förderung der Lehrkräfte (Dignath, 2016; Karlen et al., 2020). Von Lehrkräften wird erwartet, dass sie ein gutes Vorbild für ihre Schüler*innen sind, indem sie selbstregulierende Prozesse im Klassenzimmer vorleben (Çetin, 2015). Dennoch scheint es, dass eine selbstregulierte Lehrkraft nicht zwangsläufig gute SRL-Förderung betreibt: Chumworatayee (2012) stellte beispielsweise fest, dass manche Lehrkräfte zwar viele SRL-Strategien anwenden, aber nur begrenzt Strategien in ihrem Unterricht vermitteln. Um eine selbstregulationsfördernde Lehrkraft zu werden, bedarf es vermutlich mehr als nur eigene SRL.

Hypothesen

Um zu verstehen, wie Unterschiede in der SRL-Förderung von Lehrkräften erklärt werden können, haben wir die folgenden Hypothesen getestet: (1) Die SRL-Förderung von Lehrkräften wird positiv durch (a) Wissen über SRL, (b) konsistente Überzeugungen bzgl. SRL, (c) eigene Selbstregulation, (d) Selbstwirksamkeitsüberzeugung, SRL zu fördern, und negativ (e) durch ihre inkonsistenten Überzeugungen bzgl. SRL vorhergesagt. (2) Die SRL-Förderung der Lehrkräfte wird außerdem durch (a) die derzeitigen SRL-Fähigkeiten ihrer Schüler*innen, den Prozentsatz der Schüler*innen mit (b) Lernschwierigkeiten und (c) Deutsch als Zeitsprache, (d) die Klassengröße, (e) die Klassenstufe und (f) die SRL School Vision der Schule beeinflusst.

Methodisches Vorgehen und Ergebnisse

Hierzu untersuchten wir in einer Online-Befragung anhand der Angaben von $N = 257$ Lehrkräften der 1. bis 8. Klasse, welche Aspekte der SRL-Kompetenz der Lehrkräfte ihre direkte und indirekte Förderung von SRL im Unterricht vorhersagen. Die Lehrkräfte waren im Schnitt 37.55 Jahre alt und hatten 10.59 Jahre Lehrerfahrung. 83 % von ihnen waren weiblich. Wir erfragten das professionelle Wissen der Lehrkräfte zu SRL, ihre Überzeugungen bzgl. SRL und Fehlkonzepte, sowie die Selbstwirksamkeitserwartung und die eigene SRL der Lehrkraft. Zudem wurden die Lehrkräfte gebeten, ihre direkte und indirekte Förderung anzugeben. Um auch zu ermitteln, ob Lehrkräfte ihre SRL-Förderung an die Merkmale ihrer Schüler*innen anpassen, wurden zudem Variablen auf Klassen- und Schulebene erhoben. Da die Studie wäh-

rend der Corona-Pandemie durchgeführt wurde, wurden die Daten online erhoben.

Mittels eines Strukturgleichungsmodells zur Vorhersage der direkten bzw. der indirekten SRL-Förderung wurden Bedingungsfaktoren ermittelt, die Unterschiede in der von den Lehrkräften berichteten SRL-Förderung erklären. Die Ergebnisse zeigen, dass die Fehlkonzepte von Lehrkräften zum Lernen (d. h. ihre Überzeugungen, die inkonsistent zur SRL-Theorie sind) die direkte Förderung von Selbstregulation vorhersagen. Lehrkräfte, die mehr Fehlkonzepte aufweisen, berichten, weniger direkte SRL-Förderung in ihrem Unterricht durchzuführen. Ein weiterer Kompetenzaspekt, der die direkte SRL-Förderung vorhersagte, war die Selbstwirksamkeitserwartung. Je selbstwirksamer eine Lehrkraft ist, desto mehr direkte SRL-Förderung wurde berichtet. Mit Blick auf die Frage, inwieweit die SRL-Förderung von Lehrkräften mit Merkmalen ihrer Klasse oder Schule zusammenhängt, zeigte sich, dass der Anteil der Schüler*innen mit Lernstörungen in der Klasse sowie die *SRL School Vision* Prädiktoren für die direkte SRL-Förderung waren. SRL School Vision meint dabei die Vision, die die Schule der Lehrkraft für die Förderung von SRL hat (Heirweg et al., 2020). So berichteten Lehrkräfte, mehr direkte SRL-Förderung umzusetzen, wenn sie aus Schulen mit einer schwachen SRL School Vision kamen, und wenn sie einen höheren Anteil an Schüler*innen mit Lernstörungen in der Klasse hatten.

Für die indirekte Förderung von SRL erwies sich von allen Kompetenzaspekten nur die Selbstwirksamkeitserwartung der Lehrkräfte als signifikanter Prädiktor. Hinsichtlich der Kontextvariablen zeigte sich, dass die indirekte Aktivierung von SRL durch die Klassenstufe signifikant vorhergesagt wurde. Je höher die Klassenstufe, umso selbstregulationsförderlicher gestalteten die Lehrkräfte die Lernumgebungen der Schüler*innen.

Da sich zeigte, dass die Selbstwirksamkeitserwartung der Lehrkräfte hohe Korrelationen mit den anderen Kompetenzaspekten aufwies, führten wir zusätzlich Moderationsanalysen mit der Selbstwirksamkeitserwartung als Moderator und den Kompetenz- bzw. Kontextaspekten als Prädiktoren durch. Wir fanden eine negative Interaktion zwischen der Selbstwirksamkeitserwartung und der SRL School Vision für die direkte Förderung von SRL. Scheinbar kann die SRL-Vision der Schule die negativen Auswirkungen einer geringen Selbstwirksamkeitsüberzeugung auf die direkte SRL-Förderung der Lehrkräfte abfedern. Des Weiteren fanden wir einen Interaktionseffekt von Selbstwirksamkeitsüberzeugungen und den durch die Lehrkraft eingeschätzten SRL-Fähigkeiten der Schüler*innen auf die indirekte SRL-Förderung der Lehrkräfte. Lehrkräfte, die ihre Schüler*innen bereits für selbstreguliert halten, scheinen SRL eher indirekt zu fördern. Dies gilt

jedoch nur für Lehrkräfte mit einer hohen Selbstwirksamkeit zur Förderung von SRL-Strategien.

Eine Limitation beider Studien ist jedoch, dass sowohl die Selbstregulation (Studie 1) als auch die SRL-Förderung (Studie 2) der Teilnehmenden im Selbstbericht erfasst wurde und die Ergebnisse aufgrund sozialer Erwünschtheit der Teilnehmenden damit verzerrt sein könnten. Dennoch liefern die Ergebnisse erste Hinweise über Zusammenhänge professioneller Kompetenz und der SRL sowie der SRL-Förderung von (angehenden) Lehrkräften und es lassen sich Implikationen für die Aus- und Fortbildung von Lehrkräften sowie die schulische Praxis ableiten. Weitere Forschung sollte die hier gefundenen Zusammenhänge und Effekte mit Fremdberichts- bzw. Beobachtungsverfahren für SRL und SRL-Förderung validieren.

4. Implikationen für Forschung und Praxis

4.1. Implikationen für weitere Forschung

Ein erst in sehr wenigen Studien adressierter Aspekt professioneller Kompetenz von Lehrkräften sind ihre Fehlkonzepte (siehe Vosniadou et al., 2020). In unseren beiden aktuellen Studien zeigte sich, dass Fehlkonzepte über SRL sowohl in der Ausbildung von angehenden Lehrkräften als auch bei der Vorhersage von SRL-Förderung bei Lehrkräften eine wichtige Rolle spielen. So berichten Lehrkräfte, die mehr Fehlkonzepte aufweisen, weniger direkte SRL-Förderung in ihrem Unterricht durchzuführen. Zudem haben sich Fehlkonzepte von Lehramtsstudierenden als veränderungssensitiv im Zuge einer Kurzintervention herausgestellt, auch wenn die Effekte über einen längeren Zeitraum dann nicht stabil waren. Diese Befunde werfen die Frage auf, wie Fehlkonzepte auch langfristig verändert werden können.

In einigen Studien hat sich mittlerweile gezeigt, dass die Selbstwirksamkeitserwartung von Lehrkräften der stärkste Prädiktor unter den Kompetenzaspekten ist, um die SRL-Förderung vorherzusagen (Dignath, 2016; 2021; Karlen et al., 2020). Darüber hinaus zeigen die Ergebnisse unserer aktuellsten Studie, dass die Selbstwirksamkeit zudem als Moderator fungiert. So spielt die SRL School Vision der eigenen Schule für wenig selbstwirksame Lehrkräfte eine deutlich größere Rolle für ihre SRL-Förderung als für hoch selbstwirksame Lehrkräfte. Diese Ergebnisse zeigen, wie wichtig es ist, die verschiedenen Kompetenzaspekte simultan zu erfassen, um mögliche Interaktionen identifizieren zu können.

Solche Interaktionseffekte zwischen Aspekten professioneller SRL-Kompetenz spielen aber nicht nur eine Rolle für das Verstehen und Vorhersagen von SRL-Förderung im Unterricht, sondern auch für die Entwicklung der SRL-Kompetenz bei (angehenden) Lehrkräften. Hierzu braucht es experimentelle oder quasi-experimentelle Interventionsstudien, um nicht nur Zusammenhänge sondern Kausalschlüsse ableiten zu können. Nur wenig Forschung hat sich mit der Entwicklung professioneller SRL-Kompetenz beschäftigt und dabei Effekte auf die SRL-Förderung untersucht. In einer Interventionsstudie zeigte sich auch hier die Bedeutung der Selbstwirksamkeit von Lehrkräften, die die Wirksamkeit einer Lehrkräfte-Intervention zur Förderung des SRL moderierte: Lehrkräfte, die bereits zu Beginn der Intervention über eine hohe Selbstwirksamkeit verfügten, profitierten auch mehr von der Intervention (Dignath, 2021).

Auch wenn wir solch einen Effekt der Selbstwirksamkeit in unserer aktuellsten Studie für die Kurzintervention für Lehramtsstudierende nicht finden konnten, zeigte sich auch hier, dass Aspekte der motivationalen Orientierung mit der Intervention adressiert werden konnten. So verbesserte die Kurzintervention zwar nicht die Selbstwirksamkeit, aber die Wahrnehmung des Nutzens von SRL-Strategien. Weitere Interventionsforschung sollte das Wechselspiel von Selbstwirksamkeitserwartung und Wertzuschreibung bzgl. SRL weiter untersuchen, um detaillierte Erkenntnisse zu liefern, welche Kompetenzaspekte sich in welcher Reihenfolge verändern, wenn (angehende) Lehrkräfte über SRL lernen.

4.2. Implikationen für die Phasen der Lehrkräftebildung, die Schulentwicklung und die schulische Praxis

Neben den Erkenntnissen, die die Theorie zur professionellen SRL-Kompetenz von Lehrkräften weiterbringen, lassen sich aus den hier besprochenen Ergebnissen unmittelbar Implikationen für Aus- und Weiterbildung, Schulentwicklung und die schulische Praxis ableiten. Insbesondere die Befunde von Interventionsforschung können dabei Impulse geben für die Aus- und Weiterbildung von Lehrkräften zum Thema SRL. So weisen die Befunde unserer Interventionsstudie darauf hin, dass es hilfreich ist, (angehenden) Lehrkräften nicht nur Wissen über SRL zu vermitteln, sondern ihre Fehlkonzepte zu aktivieren und zu entkräften. So bieten sich mit Blick auf die Lehrkräfteausbildung viele Möglichkeiten, SRL bereits im Studium zu fördern und dabei vorhandene Fehlkonzepte zu adressieren. Die Ergebnisse von Studie 1 legen dabei nahe, dass sich das Wissen von Lehramtsstudierenden über

SRL-Strategien gut fördern lässt, wobei der Wissensanstieg über einen längeren Zeitraum nicht nachhaltig und für eine SRL-Förderung nicht ausreichend ist. Tiefergehende Maßnahmen sind notwendig, um neben dem Aufbau von Wissen auch Fehlkonzepte zu bearbeiten und Überzeugungen zu verändern.

Zwar haben in unserer Studie in beiden Interventionsgruppen die Fehlkonzepte zum Posttest hin signifikant abgenommen; allerdings zeigte sich im Follow-Up-Test nur in der Coping-Bedingung eine signifikante Zunahme des wahrgenommenen Nutzens von SRL-Strategien sowie des Einsatzes von SRL-Strategien. Dieser Befund deutet darauf hin, dass die Reduzierung von Fehlkonzepten scheinbar nicht ausreicht, um tatsächlich den Einsatz von SRL-Strategien zu verbessern, sondern – im Sinne eines *Hot Conceptual Change* (Sinatra, 2005) – auch die Motivation vorhanden sein muss, SRL-Strategien einzusetzen (Karabenik et al., 2021). Für die Aus- und Weiterbildung von Lehrkräften bedeutet dies, dass hier sowohl Fehlkonzepte adressiert als auch der Nutzen von SRL deutlich gemacht werden sollte, um langfristige Effekte im Sinne einer Verhaltensänderung erwarten zu können.

Hinsichtlich Implikationen für die Lehrkräfteweiterbildung hat Studie 2 gezeigt, dass Fehlkonzepte von Lehrkräften über SRL auch hier eine Rolle spielen sollten. Darüber hinaus legen die Ergebnisse von Studie 2 nahe, dass insbesondere die Selbstwirksamkeit von Lehrkräften gestärkt werden sollte, damit sie sich die Förderung von SRL im Unterricht zutrauen – insbesondere und auch bei Lernenden, die noch wenig selbstreguliert sind. Auch hier könnte das Lernen am Modell mithilfe von Instruktionsvideos wirksam sein, wenn die Videos nicht (nur) das Lernen mit SRL-Strategien, sondern vor allem die Förderung von SRL im Unterricht beispielhaft vorführen. Ebenso ist eine explizite Vermittlung der Förderstrategien für SRL angezeigt, um Fehlkonzepte von Lehrkräften gezielt anzusprechen und zu verändern. Darüber hinaus zeigen die Ergebnisse von Studie 2 Implikationen für die Schulentwicklung auf. So wurde der schulischen Vision über SRL in der SRL-Forschung bisher nur sehr wenig Beachtung geschenkt; auch in der schulischen Praxis wird sie in Deutschland bisher nur sehr vereinzelt berücksichtigt. Mit Blick auf Schulentwicklung könnten Schulen jedoch sehr von der Formulierung einer gemeinsamen Schulvision auf SRL profitieren, da in einer Schulgemeinschaft mit einer starken SRL-Vision auch Lehrkräfte mit einer niedrigen Selbstwirksamkeitserwartung SRL in ihrem Unterricht fördern. Eine schulweite Vision für das Thema scheint daher die Selbstwirksamkeitserwartung einer einzelnen Lehrkraft kompensieren zu können und sollte somit im Schulentwicklungskonzept verankert werden.

Mit Blick auf Implikationen für die schulische Praxis lassen sich die Ergebnisse von Studie 1 auch auf den Schulkontext und das Lernen von

Schüler*innen über SRL übertragen. So bieten Instruktionsvideos auch im Schulkontext eine gute Möglichkeit, SRL am Modell einzuführen und damit Fehlkonzepte von Schüler*innen über SRL zu verändern.

Um Fehlkonzepte von Lernenden und Lehrenden aktivieren zu können, ist es zunächst allerdings notwendig, ein genaues Bild dieser Fehlkonzepte zu haben. Allerdings gibt es bisher kaum Forschung, in der Fehlkonzepte über SRL aufgedeckt wurden. Die Identifizierung solcher im Widerspruch zu wissenschaftlichen Theorien stehenden Überzeugungen kann einen Beitrag dazu leisten, die Richtung des Aus- und Weiterbildungsprogramms von Lehrkräften und der Schulentwicklung zu bestimmen (Vosniadou et al., 2020).

Literatur

Askell-Williams, H., Lawson, M. J., & Skrzypiec, G. (2012). Scaffolding cognitive and metacognitive strategy instruction in regular class lessons. *Instructional Science*, *40*(2), 413–443. https://doi.org/10.1007/s11251-011-9182-5

Bandura, A. (1997). Self-efficacy: The exercise of control. Macmillan.

Baumert, J., & Kunter, M. (2013). Professionelle Kompetenz von Lehrkräften. In Stichwort: Zeitschrift für Erziehungswissenschaft (pp. 277–337). Springer. https://doi.org/10.1007/978-3-658-00908-3_13

Boekaerts, M. (1999). Self-regulated learning: Where we are today. International Journal of Educational Research, 31(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Butler, D. L., Lauscher, H. N., Jarvis-Selinger, S., & Beckingham, B. (2004). Collaboration and self-regulation in teachers' professional development. Teaching and Teacher Education, 20(5), 435–455. https://doi.org/10.1016/j.tate.2004.04.003

Cetin, B. (2015). Academic motivation and self-regulated learning in predicting academic achievement in college. Journal of International Education Research, 11(2), 95–106. https://doi.org/10.19030/jier.v11i2.9190

Chumworatayee, T. (2012). The Relationships among EFL-Teacher Trainee Metacognitive Awareness, Instructional Beliefs, and Instructional Practices Regarding Reading Strategies. rEFLections, 15, 44–69.

Darmawan, I. G. N., Vosniadou, S., Lawson, M. J., Van Deur, P., & Wyra, M. (2020). The development of an instrument to test pre-service teachers' beliefs consistent and inconsistent with self-regulation theory. British Journal of Educational Psychology, 90(4), 1039–1061. https://doi.org/10.1111/bjep.12345

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. Educational Psychology Review, 28(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Depaepe, F., De Corte, E., & Verschaffel, L. (2010). Teachers' approaches towards word problem solving: Elaborating or restricting the problem context. Teaching and Teacher Education, 26(2), 152–160. https://doi.org/10.1016/j.tate.2009.03.016

De Smul, M., Heirweg, S., Van Keer, H., Devos, G., & Vandevelde, S. (2018). How competent do teachers feel instructing self-regulated learning strategies? Development and validation of the teacher self-efficacy scale to implement self-regulated learning. Teaching and Teacher Education, 71, 214–225. https://doi.org/10.1016/j.tate.2018.01.001

Dignath, C., Biegel, M., Veenman, M., & Büttner, G. (2022). Assessing How Teachers Enhance Self-regulated Learning. Frankfurt am Main: Forschungsdatenzentrum Bildung am DIPF. https://doi.org/10.7477/795:378:1

Dignath, C. (2021). For unto every one that hath shall be given: teachers' competence profiles regarding the promotion of self-regulated learning moderate the effectiveness of short-term teacher training. Metacognition and Learning, 16(3), 555–594. https://doi.org/10.1007/s11409-021-09271-x

Dignath, C. (2016). What determines whether teachers enhance self-regulated learning? Predicting teachers' reported promotion of self-regulated learning by teacher beliefs, knowledge, and self-efficacy. Frontline Learning Research, 4(5), 83–105. https://doi.org/10.14786/flr.v4i5.247

Dignath, C., & Veenman, M. V. (2021). The role of direct strategy instruction and indirect activation of self-regulated learning – Evidence from classroom observation studies. Educational Psychology Review, 33(2), 489–533. https://doi.org/10.1007/s10648-020-09534-0

Dignath, C., & Büttner, G. (2018). Teachers' direct and indirect promotion of self-regulated learning in primary and secondary school mathematics classes – insights from video-based classroom observations and teacher interviews. Metacognition and Learning, 13(2), 127–157. https://doi.org/10.1007/s11409-018-9181-x

Dignath-van Ewijk, C., & Van der Werf, G. (2012). What teachers think about self-regulated learning: Investigating teacher beliefs and teacher behavior of enhancing students' self-regulation. Education Research International, 2012. https://doi.org/10.1155/2012/741713

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. Metacognition and Learning, 3(3), 231–264. https://doi.org/10.1007/s11409-008-9029-x

Dignath, C., Buettner, G., & Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively?: A meta-analysis on self-regulation training programmes. Educational Research Review, 3(2), 101–129. https://doi.org/10.1016/j.edurev.2008.02.003

Dinsmore, D. L., & Fryer, L. K. (2019). Developing learners' cognitive strategies and the motivation to use them: Rethinking education policy. Policy Insights from the Behavioral and Brain Sciences, 6(2), 107–114. https://doi.org/10.1177/2372732219860862

Donker, A. S., De Boer, H., Kostons, D., Van Ewijk, C. D., & van der Werf, M. P. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. Educational Research Review, 11, 1–26. https://doi.org/10.1016/j.edurev.2013.11.002

Eccles, W. J., & Eccles, W. J. (1983). The Canadian Frontier, 1534-1760. UNM Press.

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. Memory & Cognition, 38(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. Educational Psychologist, 46(1), 6–25. https://doi.org/10.1080/00461520.2011.538645

Gill, M. G., Trevors, G., Greene, J. A., & Algina, J. (2022). Don't take it personally? The role of personal relevance in conceptual change. The Journal of Experimental Education, 90(1), 1–22. https://doi.org/10.1080/00220973.2020.1754152

Glogger-Frey, I., Ampatziadis, Y., Ohst, A., & Renkl, A. (2018). Future teachers' knowledge about learning strategies: Misconcepts and knowledge-in-pieces. Thinking Skills and Creativity, 28, 41–55. https://doi.org/10.1016/j.tsc.2018.02.001

Greene, J. A. (2021). Teacher support for metacognition and self-regulated learning: a compelling story and a prototypical model. Metacognition and Learning, 16(3), 651–666. https://doi.org/10.1007/s11409-021-09283-7

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. Educational Psychology Review, 15(2), 147–179. https://doi.org/10.1023/A:1023477131081

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. Review of Educational Research, 66(2), 99–136. https://doi.org/10.3102/00346543066002099

De Smul, M., Heirweg, S., Devos, G., & Van Keer, H. (2020). It's not only about the teacher! A qualitative study into the role of school climate in primary schools' implementation of self-regulated learning. School Effectiveness and School Improvement, 31(3), 381–404. https://doi.org/10.1080/09243453.2019.1672758

Hynd, C. R. (2001). Refutational texts and the change process. International Journal of Educational Research, 35(7-8), 699–714. https://doi.org/10.1016/S0883-0355(02)00010-1

Jansen, R. S., Van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. Educational Research Review, 28,. https://doi.org/10.1016/j.edurev.2019.100292

Karabenick, S. A., Berger, J. L., Ruzek, E., & Schenke, K. (2021). Strategy motivation and strategy use: Role of student appraisals of utility and cost. Metacognition and Learning, 16(2), 345–366. https://doi.org/10.1007/s11409-020-09256-2

Karlen, Y., Hertel, S., & Hirt, C. N. (2020). Teachers' professional competences in self-regulated learning: An approach to integrate teachers' competences as self-regulated learners and as agents of self-regulated learning in a holistic manner. Frontiers in Education, 5, 159. https://doi.org/10.3389/feduc.2020.00159

Kendeou, P., Muis, K. R., & Fulton, S. (2011). Reader and text factors in reading comprehension processes. Journal of Research in Reading, 34(4), 365–383. https://doi.org/10.1111/j.1467-9817.2010.01436.x

Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why unguided learning does not work: An analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. Educational Psychologist, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kramarski, B., & Heaysman, O. (2021). A conceptual framework and a professional development model for supporting teachers' "triple SRL–SRT processes" and promoting students' academic outcomes. Educational Psychologist, 56(4), 298–311. https://doi.org/10.1080/00461520.2021.1985502

Kramarski, B., & Kohen, Z. (2017). Promoting preservice teachers' dual self-regulation roles as learners and as teachers: Effects of generic vs. specific prompts. Metacognition and Learning, 12(2), 157–191. https://doi.org/10.1007/s11409-016-9164-8

Kramarski, B., & Michalsky, T. (2009). Investigating preservice teachers' professional growth in self-regulated learning environments. Journal of Educational Psychology, 101(1), 161–175. https://doi.org/10.1037/a0013101

Lawson, M. J., Vosniadou, S., Van Deur, P., Wyra, M., & Jeffries, D. (2019). Teachers' and students' belief systems about the self-regulation of learning. Educational Psychology Review, 31(1), 223–251. https://doi.org/10.1007/s10648-018-9453-7

Lewandowsky, S., Stritzke, W. G., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. Psychological Science, 16(3), 190–195. https://doi.org/10.1111/j.0956-7976.2005.00802.x

Lombaerts, K., De Backer, F., Engels, N., Van Braak, J., & Athanasou, J. (2009). Development of the self-regulated learning teacher belief scale. European Journal of Psychology of Education, 24(1), 79–96. https://doi.org/10.1007/BF03173476

Michalsky, T. (2014). Developing the SRL-PV assessment scheme: Preservice teachers' professional vision for teaching self-regulated learning. Studies in Educational Evaluation, 43, 214–229.

Morris, D. B., Usher, E. L., & Chen, J. A. (2017). Reconceptualizing the sources of teaching self-efficacy: A critical review of emerging literature. Educational Psychology Review, 29(4), 795–833. https://doi.org/10.1007/s10648-016-9378-y

Peeters, J., De Backer, F., Reina, V. R., Kindekens, A., Buffel, T., & Lombaerts, K. (2014). The role of teachers' self-regulatory capacities in the implementation of self-regulated learning practices. Procedia-Social and Behavioral Sciences, 116, 1963–1970. https://doi.org/10.1016/j.sbspro.2014.01.504

Perry, N. E., Hutchinson, L., & Thauberger, C. (2008). Talking about teaching self-regulated learning: Scaffolding student teachers' development and use of practices that promote self-regulated learning. International Journal of Educational Research, 47(2), 97–108. https://doi.org/10.1016/j.ijer.2007.11.010

Perry, N. E., & Rahim, A. (2011). Studying self-regulated learning in classrooms. In B. J. Zimmerman & D. J. Schunk (Eds.), Handbook of self-regulation of learning and performance (pp. 136–150). Routledge. https://doi.org/10.4324/9780203839010

Pintrich, P. R., & De Groot, E. V. (1991). Motivated strategies for learning questionnaire. Journal of Educational Psychology.

Randi, J. (2004). Teachers as self-regulated learners. Teachers College Record, 106(9), 1825–1853. https://doi.org/10.1111/j.1467-9620.2004.00407.x

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. Cognitive Science, 38(1), 1–37. https://doi.org/10.1111/cogs.12086

Sinatra, G. M. (2005). The "warming trend" in conceptual change research: The legacy of Paul R. Pintrich. Educational Psychologist, 40(2), 107–115. https://doi.org/10.1207/s15326985ep4002_5

Spruce, R., & Bol, L. (2015). Teacher beliefs, knowledge, and practice of self-regulated learning. Metacognition and Learning, 10(2), 245–277. https://doi.org/10.1007/s11409-014-9124-0

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. Cognition and Instruction, 8(4), 351–362. https://doi.org/10.1207/s1532690xci0804_5

Thomas, V., Peeters, J., De Backer, F., & Lombaerts, K. (2022). Determinants of self-regulated learning practices in elementary education: a multilevel approach. Educational Studies, 48(1), 126–148. https://doi.org/10.1080/03055698.2020.1745624

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. Review of Educational Research, 68(2), 202–248. https://doi.org/10.3102/00346543068002202

Vandevelde, S., Vandenbussche, L., & Van Keer, H. (2012). Stimulating self-regulated learning in primary education: Encouraging versus hampering factors for teachers. Procedia-Social and Behavioral Sciences, 69, 1562–1571. https://doi.org/10.1016/j.sbspro.2012.12.099

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. Educational Psychology Review, 22(2), 155–174. https://doi.org/10.1007/s10648-010-9134-7

Vosniadou, S. (2007). The cognitive-situative divide and the problem of conceptual change. Educational Psychologist, 42(1), 55–66. https://doi.org/10.1080/00461520709336918

Vosniadou, S., Lawson, M. J., Wyra, M., Van Deur, P., Jeffries, D., & Ngurah, D. I. G. (2020). Pre-service teachers' beliefs about learning and teaching and about the self-regulation of learning: A conceptual change perspective. International Journal of Educational Research, 99, 101495. https://doi.org/10.1016/j.ijer.2019.101495

Wijnia, L., Loyens, S. M., van Gog, T., Derous, E., & Schmidt, H. G. (2014). Is there a role for direct instruction in problem-based learning? Comparing student-constructed versus integrated model answers. Learning and Instruction, 34, 22–31. https://doi.org/10.1016/j.learninstruc.2014.07.006

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In B. J. Zimmerman, M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation (pp. 13–39). Academic press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., & Risemberg, R. (1997). Self-regulatory dimensions of academic learning and motivation. In G. D. Phye (Ed.), Handbook of academic learning (pp. 105–125). Academic Press. https://doi.org/10.1016/B978-012554255-5/50005-3

Zohar, A., & Ben-Ari, G. (2022). Teachers' knowledge and professional development for metacognitive instruction in the context of higher order thinking. Metacognition and Learning, 17(3), 855–895. https://doi.org/10.1007/s11409-022-09310-1

Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low-and high-achieving students. Learning and Instruction, 18(4), 337–353. https://doi.org/10.1016/j.learninstruc.2007.07.001

Zohar, A., Degani, A., & Vaaknin, E. (2001). Teachers' beliefs about low-achieving students and higher order thinking. Teaching and Teacher Education, 17(4), 469–485. https://doi.org/10.1016/S0742-051X(01)00007-5

50 Jahre Institut für Schulentwicklungsforschung (IFS) – Beiträge des Instituts zur Empirischen Bildungsforschung

Ramona Lorenz & Alyssa Grecu

1. Das IFS von der Gründung bis heute

Das heutige Institut für Schulentwicklungsforschung wurde mit Erlass des Ministeriums für Wissenschaft und Forschung des Landes NRW vom 27. Juni 1973 als „Arbeitsstelle für Schulentwicklungsforschung" (AFS) an der damaligen Pädagogischen Hochschule Ruhr offiziell errichtet und war damit das erste Forschungsinstitut in diesem Bereich an einer deutschen Hochschule. Seinen heutigen Namen trägt das IFS seit 1980, als der Begriff „Arbeitsstelle" durch den des „Instituts" ersetzt wurde. Der Gründer, Professor em. Hans-Günter Rolff, leitete das Institut und prägte durch seine langjährige Arbeit die Entwicklung des IFS bis zu seiner Emeritierung im Jahre 2005 maßgeblich. Professor i. R. Heinz Günter Holtappels, ebenfalls ein ausgewiesener Experte im Bereich der Schulentwicklungsforschung, übernahm im Jahr 2002 die Institutsleitung bis zur Berufung und Wahl zum Geschäftsführenden Direktor von Professor i. R. Wilfried Bos im Jahr 2005. Mit Professor i. R. Wilfried Bos gewann das Institut insbesondere im Kontext internationaler Large-Scale-Assessments an Bekanntheit und Ansehen hinzu. Seit dem Jahr 2014 ist Professorin Nele McElvany die Geschäftsführende Direktorin des IFS und erweiterte die Schwerpunkte des Instituts mit Perspektiven der Pädagogischen Psychologie an der Schnittstelle zur Empirischen Bildungsforschung auf individuelle Bildungsprozesse, die Unterrichtsqualität und die Lehr- und Lernforschung.

Auch die Mitarbeiter*innen, deren Anzahl im Laufe der Jahre stetig wuchs – bei der Gründung noch mit neun Stellen bis heute mehr als 50 Mitarbeiter*innen – prägten das Institut. Sie trugen aktiv zu einer national

und international bedeutsamen Forschungsstärke und Visibilität des IFS in zentralen Themenbereichen der Empirischen Bildungsforschung bei. Nicht zuletzt ist dies allen Professor*innen und Mitarbeitenden zu verdanken, die mit ihren Forschungsbereichen am Institut in Wissenschaft, Bildungspolitik, Schule und der breiten Öffentlichkeit dem IFS zu merklichen Erfolgen verhalfen. Wer könnte den Beitrag des IFS zur Empirischen Bildungsforschung daher besser beschreiben, als seine Geschäftsführenden Direktor*innen und Professor*innen? An sie richten wir unsere Fragen zu 50 Jahren IFS.

2. „Mit welchen Forschungsschwerpunkten hat das IFS die Empirische Bildungsforschung national und international im Laufe der Zeit geprägt?“

Beim Blick auf 50 Jahre Forschung am Institut für Schulentwicklungsforschung beschreibt die Geschäftsführende Direktorin Nele McElvany aus heutiger Sicht die wichtigsten Schwerpunkte des IFS wie folgt:

> *Das IFS steht rückblickend auf die vergangenen 50 Jahre für theoriebasierte praxisnahe Schulentwicklungsforschung beispielsweise zu Fragen der Schulstruktur, der Organisationsentwicklung oder des Ganztags, aber auch für Schulleistungsstudien mit dem Ziel der objektiven und vergleichenden Messung von Bildungsergebnissen und dem Stand des Bildungssystems. Gleichzeitig waren und sind Themen wie soziale Ungleichheiten und individuelle und institutionelle Einflussfaktoren auf Bildungsprozesse und Bildungserfolge ein wichtiger Fokus.*
>
> Prof. Dr. Nele McElvany, Direktorin des IFS seit 2014, Professorin am IFS seit 2009

Das IFS wuchs im Laufe der Jahre zu einem Institut mit national und international sichtbaren Forschungsbeiträgen heran. Doch bereits in seinen Anfangsjahren leistete das Institut einen herausragenden Beitrag für Entwicklungen im Schulsystem und prägte den Begriff der Schulentwicklung, wie der Gründungsdirektor rückblickend beschreibt:

> *Damals hat das IFS in der Tat den Begriff und die Sache der Schulentwicklung in Deutschland etabliert. In der allerersten Runde, die ersten drei bis vier Jahre, war Schulentwicklung so etwas wie Schulentwicklungsplanung. […] Und ich habe gemerkt, dass es einen riesigen Forschungsbedarf gab.*
>
> Prof. em. Dr. Hans-Günter Rolff, Geschäftsführender Direktor des IFS 1973 bis 2002, Professor am IFS bis 2005

Das neu definierte Verständnis der Schulentwicklungsforschung in Deutschland stellte mit zahlreichen Forschungsprojekten in den Jahren nach der Gründung einen wichtigen Grundpfeiler für das Institut dar. Die Organisation Schule wurde mit interdisziplinären Ansätzen erforscht, um Organisations- und Entwicklungsmodelle zu entwickeln, die den Beteiligten auf den verschiedenen Handlungsebenen bei der Lösung von Schulentwicklungsproblemen halfen.

Mit dem Ziel der Unterstützung von Schulen, die nun selber professionell Prozesse steuern können sollten, stellte und stellt die Identifikation von Gelingensbedingungen und Wirksamkeitsanalysen einen wichtigen Schwerpunkt des IFS dar.

> *Das Wichtigste ist, glaube ich, dass wir in eine Richtung forschen, wo es um Gelingensbedingungen auch von Schulentwicklungsprozessen und -vorhaben geht. Das heißt: Es sind nicht nur Schulentwicklungsvorhaben, die von der Bildungsadministration oder der Bildungspolitik kommen, sondern Schulen haben mittlerweile ja auch selbst die Aufgabe – weil sie mehr Schulautonomie haben, mehr Gestaltungsautonomie –, sich weiter zu entwickeln. Und es ist ganz wichtig zu untersuchen: Was sind denn die gelingenden und die misslingenden Bedingungen für Schulentwicklung? Was für Voraussetzungen brauchen Schulen? Welche Unterstützung brauchen Schulen? Und wie kann man die Wirksamkeit von Schulen dabei verbessern? Teamhandeln, Schulleitungshandeln, das innovativ ist aber eben auch „Leadership", also Führung bedeutet. Und auch die Steuerung von Schulentwicklungsprozessen. Das wurde dann auch erforscht und ich denke, dass wir da auch die Forschung ein Stück vorangebracht haben.*
>
> Prof. i. R. Dr. Heinz Günter Holtappels, Professor am IFS von 2001 bis 2020

Dabei war es nicht nur wichtig, die Forschung voranzutreiben, sondern auch die Forschungsergebnisse in die Praxis zu tragen und die Schulen bei Entwicklungsprozessen zu begleiten:

> *Ich habe bedeutende Forschungsprojekte zu Schulautonomie, zu Ganztagsschulen, zu Schulen in herausfordernden Lagen durchgeführt. Wenn ich zurückblicke, waren es immer innovative Schulentwicklungsansätze die ich erforscht habe: verschiedene Projekte auch zu Steuergruppen in Schulen, zu Schulprogrammarbeit, zu Schulautonomie, neue Grundschulmodelle und natürlich zur Entwicklung der Ganztagsschule. Und man könnte es auf den Punkt bringen: Ich glaube, das Bedeutendste, was ich eingebracht habe, war die Forschung zu Schulentwicklungsprozessen in Kombination zur Schulwirksamkeit, also „School Effectiveness" und „School Improvement". Diese Verbindung habe ich eingebracht und mich um die gute Organisation der Schule bemüht: eine gute Schul- und Lern-*

organisation, aber eben auch um Organisationslernen. Das heißt also auch, um eine Schule, die sich weiterentwickelt, die lernt.

Prof. i. R. Dr. Heinz Günter Holtappels, Professor am IFS von 2001 bis 2020

Auch Fragen der Schulstruktur finden bis heute ihren Platz in der Arbeit der Professor*innen am IFS. So forscht seit 2020 Professor Michael Becker am IFS zu individuellen Entwicklungsverläufen und schulischen Rahmenbedingungen. Damit bearbeitet er zentrale Themen des IFS im Kontext von Bedingungsfaktoren des Kompetenzerwerbs und setzt dabei in seiner aktuellen Arbeit einen Schwerpunkt auf langfristige Entwicklungsverläufe:

Die Entwicklung von Menschen über die Lebensspanne hängt in hohem Maße von individuellen Merkmalen und von Umweltfaktoren ab. Die Schule zählt zu den wichtigen Umweltfaktoren für die kognitive Entwicklung von Kindern und Jugendlichen. Wir untersuchen die Bedeutung schulischer Rahmenbedingungen, individueller Faktoren und das Wechselspiel dieser beiden für die Entwicklung über die Lebensspanne.

Prof. Dr. Michael Becker, Professor am IFS seit 2020

Einen bedeutsamen Meilenstein hinsichtlich der Forschungsschwerpunkte des IFS stellte die Einrichtung einer Professur dar, die das Forschungsfeld auf den Schwerpunkt informations- und kommunikationstechnologische Grundbildung erweiterte und 1992 mit Professorin Renate Schulz-Zander besetzt wurde.

Mit dem IT-Bereich ist ein neuer Schwerpunkt am IFS entstanden und damit Schulentwicklung durch diese Perspektive erweitert worden. Schulentwicklung war zunächst ein neues Thema für mich. Ich habe es als Herausforderung angenommen, in meine Forschungen einbezogen und dies als einen wichtigen Aspekt bei der Einführung digitaler Medien in den Schulbereich und die Lehrerbildung bewertet. Meine Forschungsschwerpunkte waren Medien und IT in Bildung, Erziehung und Unterricht, Lernen und Lehren mit digitalen Medien in der Schule und Hochschule, Schul- und Unterrichtsforschung, Genderforschung sowie digitale Medien in der Lehrerbildung.

Prof. i. R. Dr. Renate Schulz-Zander, Professorin am IFS von 1992 bis 2010

An der Schnittstelle zur Erziehungswissenschaft war dies in Deutschland eine der ersten Professuren, die sich mit digitalen Medien und Bildungsprozessen befasste, sodass das IFS auch hier eine Vorreiterrolle einnahm. Zudem gestaltete Professorin Renate Schulz-Zander auch international gut vernetzt

die Forschung in diesem für die damalige Zeit ganz neuen und innovativen Forschungsfeld mit.

> *Meine internationalen Kooperationen führten unter anderem dazu, dass ich als Vertreterin für Deutschland Mitglied des internationalen Ausschusses COMMITT (COmmittee on MultiMedia in Teacher Training) des niederländischen Ministeriums für Bildung und Erziehung wurde und an einer gemeinsamen Forschungsreise zum Besuch einschlägiger Einrichtungen zur Nutzung von Informationstechnologien für Bildungszwecke wie dem Peabody College der Vanderbilt University in Nashville, dem Media Lab des MIT in Cambridge, der BBN (Bolt Beranek and Newman Cooperation) in Boston und dem Schulzentrum in Houston, Texas teilnahm. COMMITT hat darauf basierend Empfehlungen zur Lehrerbildung publiziert, die prägend für meine weiteren Forschungen am IFS waren. Mit der Gründung und Schriftleitung der fachdidaktischen Zeitschrift LOG IN und der Zeitschrift Computer und Unterricht habe ich maßgebliche Wirkungsfelder für die Umsetzung digitaler Bildung in der Schule geschaffen.*
>
> Prof. i. R. Dr. Renate Schulz-Zander, Professorin am IFS von 1992 bis 2010

Die Forschungsergebnisse zum schulischen Einsatz digitaler Medien hatten Auswirkungen auf die Lehrkräftebildung, auf landesweite curriculare Entwicklungen und fanden bildungspolitisch Beachtung. Auch heute noch ist das Thema der Bildung im Kontext einer zunehmend digitalisierten Welt ein wichtiges Forschungsthema am IFS, das unter anderem von PD Dr. Ramona Lorenz in Projekten wie *Schule digital – der Länderindikator* oder der digitalen Erfassung der Lesekompetenz bei der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) bearbeitet wird.

Schon früh zeichnete sich das IFS durch eine internationale Orientierung aus, wie beispielsweise Aufenthalte des Gründungsdirektors als Visiting Professor an der Stanford University oder Gastprofessuren in Österreich und der Schweiz dokumentieren.

Die Forschung des IFS sollte zudem – dies Ziel setzten sich die Professor*innen des IFS um die Jahrtausendwende – noch sichtbarer und auch international stärker wahrgenommen werden. Neben der umfassenden Vernetzung des IFS, der Teilnahme an internationalen Konferenzen sowie der Einladung internationaler Forscher*innen an das Institut gewann die internationale Perspektive mit Professor i. R. Wilfried Bos eine neue Prägung – nämlich über die international vergleichende Schulleistungsforschung, mit der die Qualität des deutschen Schulsystems weltweit verglichen werden konnte:

Kern meiner Arbeit waren die Schulleistungsstudien, die habe ich gewissermaßen in Deutschland mit eingeführt oder großgemacht. […] Weil ich mich bei Neville Postlethwaite, der international weltweit die Schulleistungsstudien seit den 60er Jahren implementierte, in Hamburg habilitiert habe, besaß ich ein Know-How, über das außer mir fast niemand verfügte. Die Studien und Daten wurden natürlich alle international erhoben und ich saß jedes Mal dabei, wenn international diskutiert wurde. Ich sagte auch „IGLU muss kommen" und war daran beteiligt, denn internationale Grundschuluntersuchungen gab es noch nicht.

Prof. i. R. Dr. Wilfried Bos, Geschäftsführender Direktor des IFS von 2005 bis 2014, Professor am IFS bis 2019

Mit seinem Schwerpunkt im Bereich der Large-Scale-Assessment-Projekte hat Professor i. R. Wilfried Bos nationale und internationale Schulleistungsstudien wie die *Internationale Grundschul-Lese-Untersuchung* (IGLU), die *Trends in International Mathematics and Science Study* (TIMSS), Etappe 4 des *Nationalen Bildungspanels* (NEPS) und die *International Computer and Information Literacy Study* (ICILS) am das IFS durchgeführt. Damit wurde das IFS federführend in diesem Bereich, fand Antworten auf drängende Fragen der Bildungsforschung und legte eine wichtige Basis für die weitere Forschung des Instituts. Ein spezieller Fokus war Professor Wilfried Bos dabei besonders wichtig, der bildungspolitisch weitreichend wahrgenommen wurde und unter anderem auch zu einer Revision von Ressourcenverteilungen an Schulen führte:

Für mich war die Koppelung des Sozialindex an den Bildungserfolg wichtig, also: Aus welchen Elternhäusern kommen Kinder? Wie ist die Schichtabhängigkeit vom Bildungsergebnis? Da haben wir mit unseren Studien wirklich wegweisende Antworten gegeben und das war für mich persönlich mit das Wichtigste, weil diese Koppelung in Deutschland extrem hoch ist. Objektivität im Bildungswesen, wenn man die überhaupt einführen kann, lässt sich im Prinzip über Schulleistungsstudien erhöhen. Auch Schulentwicklung ist mit empirisch belastbaren Daten auf eine ganz andere Basis gestellt worden.

Prof. i. R. Dr. Wilfried Bos, Geschäftsführender Direktor des IFS von 2005 bis 2014, Professor am IFS bis 2019

Die Schulentwicklungsforschung am IFS konnte damit auf der Grundlage solcher repräsentativen Daten einen entscheidenden Beitrag zur Weiterentwicklung des Schulsystems leisten. Das datengestützte Monitoring hielt auch in der Bildungspolitik nachhaltig Einzug und wurde fester Bestandteil der Strategie der Kultusministerkonferenz zum Bildungsmonitoring in Deutschland.

Die *Internationale Grundschul-Lese-Untersuchung* (IGLU) wird ebenso wie die wissenschaftliche Leitung der Etappe 4 des *Nationalen Bildungspanels* (NEPS) am IFS inzwischen von Professorin Nele McElvany weitergeführt, die im Jahr 2009 berufen wurde und seit 2014 als Geschäftsführende Direktorin die Entwicklung des IFS maßgeblich gestaltet. Mit ihr gewann das Institut eine Forschungsperspektive hinzu, die erweiternd auf individuelle Bildungsprozesse und Unterrichtsqualität ausgerichtet ist:

> *Bildungsprozesse und der Bildungserfolg im Kontext der Lesekompetenz von Schülerinnen und Schülern bilden den Kern meiner Forschungsarbeit. Dabei sind nicht nur die Messung der Lesekompetenz sowie ihre Entwicklung und Förderung leitende Anliegen, sondern auch die Untersuchung der Bedingungsfaktoren. Die Lesemotivation auf Schülerseite oder die Unterrichtsqualität sind Beispiele für Bedingungsfaktoren, die bedeutsam für die Lesekompetenz sind, wobei die Lesemotivation selbst auch als Bildungsziel beschrieben werden kann. Darüber hinaus stehen auch der Wortschatzerwerb – gerade bei Kindern mit Migrationshintergrund – oder auch grundsätzlich die Frage sozialer Ungleichheiten im Bildungskontext im Fokus von Forschungsprojekten.*
>
> Prof. Dr. Nele McElvany, Direktorin des IFS seit 2014, Professorin am IFS seit 2009

Die Interventionsforschung (z. B. Motivationsförderung im schulischen Kontext) ist ein Schwerpunkt, den Professorin Hanna Gaspard seit 2020 am Institut vertritt. Zudem unterstützt sie die fundierte empirische Ausrichtung des IFS im Rahmen einer Professur für Educational Data Science und betont vor diesem Hintergrund die Relevanz von Forschungsarbeit am IFS:

> *In der Empirischen Bildungsforschung wurden am IFS bereits zahlreiche bedeutsame Befunde erarbeitet, doch nun wird diese Forschung methodisch erweitert. Mit dem Profilbereich „From Prediction to Agile Intervention Research in the Social Sciences" (FAIR) ist federführend durch das IFS ein interdisziplinärer Forschungsbereich entstanden, in dem die zunehmend größeren und komplexeren zur Verfügung stehenden Datensätze genutzt werden, um Vorhersagemodelle in den Sozialwissenschaften zu optimieren. Damit kann die Wirkung verfügbarer Interventionen zur Förderung von akademischem Erfolg und Wohlbefinden erhöht werden, indem sie basierend auf einem besseren Verständnis von Ursachen und Wirkungen datengestützte, individualisierte Anpassungen ermöglichen. Die Entwicklung und Anwendung innovativer Forschungsmethoden verbunden mit der Nutzung großer Datenmengen ist sehr wertvoll für aktuelle gesellschaftliche Herausforderungen im Bildungswesen.*
>
> Prof. Dr. Hanna Gaspard, (Vertretungs-)Professorin am IFS von 2020 bis 2023

Auch weitere zwischenzeitlich an das IFS berufene Professor*innen setzten ihre Akzente: Neben Professorin Aileen Edele, die den Schwerpunkt der gesellschaftlichen Heterogenität und sozialen Ungleichheit verstärkt einbrachte, beschäftigte sich Professorin Fani Lauermann mit Fragen zu Bildungsverläufen und Determinanten von Unterrichtsqualität und initiierte den Profilbereich FAIR maßgeblich. Außerdem unterstützte sie die Weiterentwicklung der Internationalisierung des IFS durch internationalen Austausch, Engagement in internationalen Kontexten und herausragende internationale Publikationen umfassend.

So baute das IFS seine Grundpfeiler der Forschung über die Jahre und Jahrzehnte immer weiter intensiv aus und entwickelte sich und seine zentralen Forschungsbereiche gleichzeitig dynamisch weiter. Die Forschungsqualität trägt dabei in entscheidendem Maße zum nationalen und internationalen Renommee des Instituts bei. Schließlich ist diese Qualität nicht nur für die Forschung relevant, sondern ist auch ein wichtiger Baustein für Förderung und Qualifikation des wissenschaftlichen Nachwuchses:

> *Die Forschungsstärke des IFS ist beeindruckend. Mit mehr als 200 Projekten seit der Gründung des Instituts bis heute besteht hier ein reicher Erfahrungsschatz in der Durchführung unterschiedlicher Studien: schulnahe Entwicklungsprojekte, Interventionsstudien bis hin zu großen Datensätzen und internationalen Schulleistungsstudien. Diese Projekte sind vor allem methodisch sehr anspruchsvoll und werden auf hohem Niveau unter Anwendung komplexer statistischer Verfahren durchgeführt. Dies ist auch für die Weiterqualifikation des wissenschaftlichen Nachwuchses äußerst wertvoll.*
>
> Prof. Fani Lauermann, PhD, Professorin am IFS von 2018 bis 2023

Der intensive Austausch zwischen den Arbeitsgruppen bereichert damals wie heute sowohl die Forschung als auch die Qualifikation des wissenschaftlichen Nachwuchses. Durch die gelungene internationale Vernetzung, Sichtbarkeit der Forschungsergebnisse sowie internationale Forschungsprojekte kann das IFS nach 50 Jahren auf wichtige Erfolge zurückblicken und resümierend kann festgehalten werden:

> *Das IFS zeichnet sich durch seine multiperspektivische Forschung aus: Von Grundlagenforschung bis hin zu konkreten Innovationen im Bildungswesen, von der Systemebene bis hin zu individuellen Lehr-Lernprozessen ist es eine ganzheitliche Perspektive. Damit konnte das IFS in den vergangenen 50 Jahren zahl-*

reiche wertvolle Befunde zur Verbesserung der Qualität und Entwicklung des Schulsystems beitragen.

Prof. Dr. Nele McElvany, Direktorin des IFS seit 2014, Professorin am IFS seit 2009

3. „Wie hat das IFS auf die Bildungspolitik, Bildungsadministration und schulische Praxis gewirkt?“

Das IFS verfügt über ein breites Repertoire an Forschungsschwerpunkten, zu denen es empirische Studien mit essenziellen Befunden vorweisen kann. Doch welche Reichweite haben diese Befunde über die Forschung hinaus? Wir fragen die Professor*innen des Instituts, wie das IFS auf zentrale Entscheidungsstrukturen auf Ebene der Bildungspolitik und Bildungsadministration und in die schulische Praxis gewirkt hat.

Noch heute beeinflusst der Ursprungsgedanke der ‚Schulentwicklung‘ die Forschung und die pädagogische Praxis weitreichend. Der theoretische Ansatz der Schulentwicklung fand am IFS so seine Wurzeln:

*Wir haben in den ersten Jahren des Instituts Schulentwicklungspläne erstellt. […] Doch bald wurde mir klar, in der Strukturfrage kommen wir nicht sehr viel weiter und diese löst auch nicht alle Probleme. Wir müssen innere Schulreform konzipieren und erforschen. So kam das neue Konzept der Schulentwicklung als Trias von der Organisationsentwicklung, Personalentwicklung und Unterrichtsentwicklung mit dem Bezugspunkt ‚lernende Schüler*innen‘. Und diesen Begriff, diese ‚Trias von den drei Wegen‘ haben fast alle Landesinstitute und alle Referenzrahmen für Schulqualität übernommen.*

Prof. em. Dr. Hans-Günter Rolff, Geschäftsführender Direktor des IFS 1973 bis 2002, Professor am IFS bis 2005

Eine der deutlichsten Wirkungen geht zudem auf die Befunde der internationalen Schulleistungsstudien zurück, mit denen Professor i. R. Wilfried Bos mit seiner Arbeitsgruppe wichtige Erkenntnisse für die Qualitätssicherung im Bildungswesen erarbeitet hat. Die Teilnahme an internationalen Schulleistungsstudien ist in die Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring eingegangen, sodass das IFS mit der Durchführung solcher Studien auf bildungspolitischer Ebene eine herausragende Rolle einnimmt und die Qualitätssicherung im Bildungswesen sowie die Schulentwicklung maßgeblich unterstützt. Dabei bekräftigt auch Professor Michael Becker, dass insbesondere die Relevanz der Datenbasis und empirisch gestützte Entscheidungen eines der wesentlichen Merkmale der Wirkung des IFS ist. Die

internationale Schulleistungsforschung hat nicht zuletzt dazu beigetragen, dass das IFS ein international wahrgenommener Forschungspartner ist. Insgesamt kann eine ganzheitliche Perspektive auf das Schulsystem eingenommen werden:

> *Das IFS setzt seit Jahrzehnten auf sehr unterschiedlichen Ebenen national und international Akzente. Das reicht von ganz grundlegenden Aspekten, wie etwa evidenzbasierte Forschung und Forschungs-Praxis-Kooperationen voranzutreiben, über die Bearbeitung zentraler Themen der Schulentwicklung, wie beispielsweise dem Ausbau der Ganztagsschule, bis hin zur Gestaltung international vergleichender Large-Scale-Studien wie PIRLS/IGLU als Baustein des Systemmonitorings. Interessant finde ich insbesondere, wie das IFS die Forschung von der Arbeit an der konkreten Einzelschule bis hin zum nationalen und internationalen Systemmonitoring unter einem Dach vereint. Aber ich würde sagen, dass mir die Rolle des IFS, bildungspolitische Diskussionen stärker evidenzbasiert zu gestalten, besonders zentral erscheint.*
>
> Prof. Dr. Michael Becker, Professor am IFS seit 2020

Die Zusammenarbeit mit dem Bundesministerium für Bildung und Forschung (BMBF), der Kultusministerkonferenz (KMK), Ministerien der Länder, Landesinstituten, Bezirksregierungen oder Bildungsbüros war stets relevant und diese Vernetzung hat dazu beigetragen, auf der Systemebene die Qualität der Unterstützung der Schulen zu verbessern. So sind auch Qualitätskriterien und Orientierungsrahmen entstanden, die diese Institutionen in ihre Arbeit haben einfließen lassen. Ergebnisse der Forschung am IFS sind somit auf diesem Weg, aber auch in der direkten Arbeit mit Schulen, über Fortbildungen für Schulleitungen und Lehrkräfte in die Schulen gelangt.

> *Für mich war es immer wichtig, dass unsere Ergebnisse in die schulbezogene Fortbildung und in die Beratung einmünden und die Ergebnisse zum Schulsystem natürlich auch in die regionale, in die kommunale Schulentwicklungsplanung eingehen.*
>
> Prof. i. R. Dr. Heinz Günter Holtappels, Professor am IFS von 2001 bis 2020

Neben den bildungspolitischen Debatten, für die das IFS eine empirische Grundlage geliefert hat, sind die Ergebnisse auch in die Lehrkräftebildung eingeflossen. Die Phasen der Lehrkräftebildung haben Veränderungen erfahren, die unter anderem mithilfe von Befunden des Instituts gestützt wurden, wie etwa zur Unterrichtsqualität oder Kompetenzförderung. Die Gründung der Deutschen Akademie für Pädagogische Führungskräfte (DAPF) der TU Dortmund durch Professor em. Hans-Günter Rolff im Jahr 2005 und die enge

Kooperation mit dem IFS hat maßgeblich dazu beigetragen, Wissenschaft und Praxis im Bereich der Fortbildung für Lehrkräfte und Schulleitungen eng miteinander zu verbinden.

Der Transfer wissenschaftlicher Erkenntnisse in die Praxis stand am IFS immer im Fokus und wurde mit unterschiedlichen Formaten unterstützt. Große Projekte wie *Schulen im Team*, *GanzIn – Mit Ganztag mehr Zukunft. Das neue Ganztagsgymnasium NRW*, *Studie zur Entwicklung von Ganztagsschulen: Wirkungen von außerunterrichtlichen Förderarrangements in Verbindung mit dem Unterricht an Ganztagsgrundschulen* (*StEG*) oder *Qualitätssicherung von zentralen Prüfungen – Zentralabitur NRW* (*AbiTÜV*) – um nur einige Beispiele zu nennen – waren einflussreich und lieferten wegweisende Erkenntnisse. Dabei galt es nicht nur Einzelschulen zu erreichen, sondern Befunde in die Breite zu tragen. Zeitschriften, die speziell an Schulen gerichtet sind, waren eines dieser Mittel, Wissenschaft und Schule stärker zusammenzubringen, wie zum Beispiel *Computer + Unterricht*. Die Vernetzung und Diskussion bildungswissenschaftlicher Fragestellungen wurde mit der Gründung des Open Access erscheinenden *Journal for Educational Research Online* (JERO) als internationalem und interdisziplinärem Journal vorangetrieben. Neben Zeitschriften hat das IFS im Laufe der Jahre verschiedene Herausgeberschaften von Reihen verantwortet – unter anderem die des vorliegenden Jahrbuchs der Schulentwicklung. Auch die über mehrere Jahre durchgeführte, repräsentative IFS-Umfrage stellte Ergebnisse zu Entwicklungen von Urteilen und Einstellungen der Bevölkerung in Deutschland zu Schul- und Bildungsfragen vor. Um den gezielten Austausch des IFS mit verschiedenen Akteur*innen zu intensivieren, sind zudem regelmäßige Tagungsformate etabliert worden wie das *Dortmunder Symposium der Empirischen Bildungsforschung* für den interdisziplinären wissenschaftlichen Austausch und der *IFS-Bildungsdialog*, bei dem Vertreter*innen aus der Bildungspolitik, Bildungsadministration, Schulen oder relevanten Unterstützungssystemen mit der Wissenschaft in Austausch treten. Das IFS lädt renommierte Wissenschaftler*innen, Vertreter*innen der Bildungspolitik und schulische Akteur*innen ein, im Rahmen dieser Formate miteinander zu diskutieren, voneinander zu lernen und gemeinsam über die verschiedenen Perspektiven nachzudenken.

Daneben sind Formate wie das webbasierte Praxisportal oder Transferhandreichungen, wie jüngst in IGLU entwickelt, wichtige Maßnahmen. Zusammenfassend kann also festgehalten werden:

*Viele Ergebnisse der Schulentwicklungsforschung haben eine direkte praktische Relevanz. Studien werden häufig in Kooperation mit der schulischen Praxis durchgeführt, so dass die Ergebnisse hier direkt einfließen können. Auch die Bildungsadministration und Bildungspolitik haben großes Interesse an den Forschungsergebnissen. Formate wie der IFS-Bildungsdialog richten sich direkt an die Kommunikation mit diesen Akteur*innen und erfahren eine sehr positive Resonanz.*

Prof. Dr. Hanna Gaspard, (Vertretungs-)Professorin am IFS von 2020 bis 2023

Des Weiteren wird über die Institutsgrenzen hinaus ein vielfältiges Publikum mit regelmäßig stattfindenden Ringvorlesungen und der „Virtual Keynote Series" erreicht. Über die Ergebnisse der Forschung wird die interessierte Öffentlichkeit zudem durch aktive Pressearbeit sowie verschiedene Social-Media-Aktivitäten informiert.

Das IFS hat eine starke internationale Strahlkraft und wird nicht nur national, sondern auch international als „Powerhouse" für Empirische Bildungsforschung anerkannt. Mit international sichtbaren Veranstaltungsformaten wie den „IFS Virtual Keynotes" trägt das IFS zum akademischen und öffentlichen Diskurs zu wichtigen Bildungsthemen bei, wie z. B. „Interventions that support students' academic success", „Instructional design innovations", „Civic literacy", „Education in the digital world" und „Education and societal diversity".

Prof. Fani Lauermann, PhD, Professorin am IFS von 2018 bis 2023

Nicht zuletzt ist die Strahlkraft des IFS auch weit über das Institut hinaus dadurch wahrnehmbar, dass im Laufe der 50 Jahre zahlreiche Personen am Institut eine Promotion oder Habilitation abgeschlossen und einflussreiche Positionen übernommen haben, unter anderem auf Professuren, in Ministerien, Landesinstituten und Bildungsbüros. Auch die herausragende Leistung dieser ehemaligen IFS-ler*innen baut auf die Grundlagen der Arbeit des Instituts auf.

4. „Welche sind die zukünftigen Perspektiven, mit denen das IFS die empirische Bildungsforschung zentral prägen wird? Wohin geht die Reise des IFS in den nächsten zehn Jahren?"

Nach dem Rückblick auf wesentliche Meilensteine der letzten 50 Jahre des Instituts, richten wir nun den Blick auf die Zukunft und fragen, welche Perspektiven sich für das IFS skizzieren lassen. „Quo vadis IFS?" fragen wir.

Die besondere Stärke des Instituts im Bereich der klassischen Schulentwicklungsforschung wurde in den vergangenen zwei Jahrzehnten erweitert und es rückten zunehmend große Large-Scale-Assessment-Studien sowie pädagogisch-psychologische und interdisziplinäre Fragestellungen der Bildungsforschung in den Fokus. Diese Vielfalt der Forschung und methodischen Zugänge, die interdisziplinäre Zusammenarbeit und den Transfergedanken stellt auch die Geschäftsführende Direktorin als weiterhin wichtige Zukunftsperspektive heraus:

> *Wir werden als großes, interdisziplinäres Team weiterhin die wichtigen Fragen der aktuellen Debatten um schulische Bildungsprozesse und das Schulsystem mit empirischen Befunden aus Studien mit unterschiedlichen Themenschwerpunkten und Forschungsdesigns unterstützen. Themen wie der vielfältige Kompetenzerwerb von Kindern und Jugendlichen angesichts verschiedener individueller, institutioneller und sozialer Bedingungsfaktoren, die Digitalisierung oder guter Unterricht werden sicherlich leitend bleiben und das IFS als Größe in der Empirischen Bildungsforschung sichtbar bleiben lassen. Das IFS wird dabei auch zukünftig dafür stehen, dass das Institut nicht nur wichtige Erkenntnisse für die wissenschaftliche Community liefert, sondern auch ein wichtiger Partner auf allen anderen Ebenen – für die Politik, Bildungsadministration, Praxis und interessierte Öffentlichkeit – ist und so die Empirische Bildungsforschung und Bildungspraxis prägt.*
>
> Prof. Dr. Nele McElvany, Direktorin des IFS seit 2014, Professorin am IFS seit 2009

Diese Einschätzung teilen die Professor*innen des IFS und sehen das Institut ebenfalls an einer wichtigen Schlüsselstelle:

> *Es wird sicherlich eine Reihe von relevanten Entwicklungen geben. Ich denke, das IFS ist eine der zentralen Institutionen der deutschen Bildungslandschaft, die viele Forschungslinien mitgeprägt hat und mit der großen Vielfalt an sehr unterschiedlichen Studien und einem großen interdisziplinären Forscher*innen-Team tief in die Bildungslandschaft hineinwirkt – sowohl in die Forschung im engeren Sinne als auch in den Forschungs-Praxis-Austausch.*
>
> Prof. Dr. Michael Becker, Professor am IFS seit 2020

Konkrete Entwicklungslinien sind zwar immer schwierig vorauszusagen, dennoch gibt es prägende Themen, mit denen das IFS eng verbunden ist und die weiterhin prägsam bleiben werden. Die Antwort auf die Frage, wo die Reise des IFS in den nächsten zehn Jahren wohl hingeht, identifiziert mehrere thematische Entwicklungslinien:

Ich würde eher von Reisen im Plural sprechen, da vielfältige Entwicklungslinien am IFS verfolgt werden, die alle wichtig und zentral erscheinen, unter anderem Themen wie die Veränderung und Förderung von kognitiven und sozioemotionalen Kompetenzen, Lehren und Lernen im digitalen Zeitalter, Heterogenität und Ungleichheit im Bildungswesen, und nicht zuletzt die Transformation und Entwicklung von Schulen und des Bildungssystems.

Prof. Dr. Michael Becker, Professor am IFS seit 2020

Dazu ergänzt Professorin Charlotte Dignath auch die Entwicklungspotenziale des IFS im Bereich von Interventionen und nachhaltiger Bildung mit Blick auf längere Zeitspannen und Bildungsphasen:

Ich kann mir auch gut vorstellen, dass das IFS die Forschung zu Interventionen im Bildungsbereich, zur Wirksamkeit von Instruktion und zur Rolle von Bildung für längerfristige Entwicklungen zentral mitprägen wird. Gerade für solche aktuellen Fragen im Bildungskontext ist die enge Zusammenarbeit der Arbeitsgruppen am IFS sehr gewinnbringend.

Prof. Dr. Charlotte Dignath, Professorin am IFS seit 2022

Die Vielfalt der Forschungsthemen und Herausforderungen im Bildungsbereich bedingt, dass eine weitere Stärke des IFS auch in Zukunft Bestand haben wird: auf methodisch höchstem Niveau unterschiedliche Designs, Erhebungsmethoden und Analysestrategien in den Studien des Instituts zu implementieren:

Wir werten Daten unter anderem aus international vergleichenden Large-Scale-Assessment-Studien, experimentellen Designs in Schulen oder im Labor, Längsschnittstudien, videographierten Unterrichtsbeobachtungen, Interviews, Kompetenztests, Fragebogendaten oder digitalen Verhaltensspuren mit komplexen Methoden und innovativen Ansätzen aus, um unsere Prädiktionsmodelle und Wirksamkeitsanalysen bestmöglich zu spezifizieren und Antworten auf die wichtigen Forschungsfragen zu finden.

Prof. Dr. Hanna Gaspard, (Vertretungs-)Professorin am IFS von 2020 bis 2023

Bei all den zahlreichen Leistungen, die im vorliegenden Kapitel lediglich punktuell benannt werden konnten, muss das Engagement aller Beteiligten im Laufe der vergangenen 50 Jahre hervorgehoben werden, ohne deren Zutun sich die heutige Präsenz und Wirksamkeit des IFS nicht hätte entwickeln können:

*Ein großer Dank sei allen Ehemaligen, Affiliierten und aktuellen Mitarbeiter*innen in Wissenschaft und Verwaltung ausgesprochen, deren Arbeit, Engagement, innovatives Denken und kollegiales Miteinander sich wie ein Puzzlestück in das große Gesamtbild des IFS eingefügt hat! Es ist eine große Freude und mir persönlich eine Ehre mit so vielen tollen Menschen zusammenarbeiten zu dürfen. Für die Zukunft bleibt an dieser Stelle nun dem IFS ein ebenso gutes Gelingen bei allen weiteren Vorhaben zu wünschen.*

Prof. Dr. Nele McElvany, Direktorin des IFS seit 2014, Professorin am IFS seit 2009

Autorinnen und Autoren

Pascal Alscher, M.Sc., wissenschaftlicher Mitarbeiter, Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Politische Bildung in der Adoleszenz, kognitive und motivationale Faktoren politischer Beteiligung, Testkonstruktion und Testentwicklung.
Pascal Alscher, M.Sc.
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
pascal.alscher@tu-dortmund.de

Prof. Dr. Richard Arum, Professor of Sociology and Education at University of California, Irvine. Research Interests: Social Stratification, Sociology of Organizations, Sociology of Education.
Richard Arum, PhD
School of Education
3231 Education
Irvine, CA 92697
Richard.arum@uci.edu

Prof. Dr. Michael Becker, seit 2020 Professor für Empirische Bildungsforschung schulische Bildungsprozesse am Institut für Schulentwicklungsforschung der Technischen Universität Dortmund und Research Fellow am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. Forschungs- und Arbeitsschwerpunkte: Akademische und psychosoziale Entwicklung und Bildungsverläufe, Entstehung von Bildungsungleichheiten unter besonderer Berücksichtigung des Schulkontextes, Entwicklung und Veränderung von Schulsystemen, quantitative Methoden der Sozialforschung
Prof. Dr. Michael Becker
Institut für Schulentwicklungsforschung (IFS)
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
michael.becker@tu-dortmund.de

Dr. Lars Behrmann, Studienrat im Hochschuldienst am Institut für Erziehungswissenschaft der Universität Münster. Forschungs- und Arbeitsschwerpunkte: Professionalität von Lehrkräften mit Fokus auf Pädagogische Diagnostik/diagnostische Kompetenz; Lehrkräftebildung mit Fokus auf Forschendes Lernen.
Dr. Lars Behrmann
Institut für Erziehungswissenschaft
Universität Münster
Georgskommende 33
48143 Münster
lars.behrmann@uni-muenster.de

Dr. Daria K. Benden, bis 31.03.2023 wissenschaftliche Mitarbeiterin am Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Seit 01.04.2023 wissenschaftliche Mitarbeiterin am Bonner Zentrum für Lehrerbildung (BZL), Universität Bonn. Forschungs- und Arbeitsschwerpunkte: Motivationsentwicklung im MINT-Bereich, Bildungsverläufe, Bildungs- und Berufsentscheidungen im MINT-Bereich.
Daria K. Benden
Bonner Zentrum für Lehrerbildung
Universität Bonn
Poppelsdorfer Allee 15
53115 Bonn
daria.benden@uni-bonn.de

Dr. Jeffrey M. DeVries, wissenschaftlicher Mitarbeiter am Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Test Entwicklung, Inklusion, mathematische und geistige Entwicklung.
Institut für Schulentwicklungsforschung
Dr. phil. Jeffrey M. DeVries
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
jeffrey.devries@tu-dortmund.de

Prof. Dr. Charlotte Dignath, Universitätsprofessorin für Empirische Bildungsforschung mit Schwerpunkt Förderung kognitiver und motivational-affektiver Lernprozesse, Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Selbstregulation beim Lernen, individuelle Förderung, Instructional Design, Lehrkräftekompetenz, Meta-Analysen.
Prof. Dr. Charlotte Dignath
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
charlotte.dignath@tu-dortmund.de

Prof. Dr. Hanna Dumont, seit 2021 Professorin für Pädagogische Psychologie mit dem Schwerpunkt schulische Lehr-Lern-Prozesse an der Universität Potsdam. Forschungs- und Arbeitsschwerpunkte: Entstehung von Bildungsungleichheiten, Umgang mit Heterogenität in Schule und Unterricht, Adaptiver Unterricht und individuelle Förderung.
Prof. Dr. Hanna Dumont
Universität Potsdam
Department Erziehungswissenschaft
Karl-Liebknecht-Str. 24–25
14476 Potsdam
hanna.dumont@uni-potsdam.de

Antonia Fischer, M.Sc., wissenschaftliche Mitarbeiterin, Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Selbstregulation beim Lernen, Lehrkräftekompetenz.
Antonia Fischer, M.Sc.
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
antonia2.fischer@tu-dortmund.de

Prof. Dr. Hanna Gaspard, Universitätsprofessorin für Educational Data Science, Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Motivationsentwicklung, Interventionen zur Förderung von Motivation und Bildungserfolgen, Evaluation von Interventionen im Bildungskontext, Unterrichtsqualität und Motivation von Lehrenden und Lernenden.
Prof. Dr. Hanna Gaspard
Institut für Schulentwicklungsforschung
TU Dortmund
Martin-Schmeißer-Weg 13
44227 Dortmund
hanna.gaspard@tu-dortmund.de

Prof. Dr. Miriam Marleen Gebauer, Universitätsprofessorin für Schulpädagogik am Institut für Erziehungswissenschaft, Otto-Friedrich-Universität Bamberg. Forschungs- und Arbeitsschwerpunkte: Professionalisierung von Lehrkräften und Lehrkräftebildung, Diversität in Schule und Unterricht, sozial-kognitive Fähigkeiten als Determinanten von Lehr-Lernprozessen.
Prof. Dr. Miriam Marleen Gebauer
Lehrstuhl für Schulpädagogik
Otto-Friedrich-Universität Bamberg
Markusplatz 3
96047 Bamberg
miriam-marleen.gebauer@uni-bamberg.de

Dr. Alyssa Laureen Grecu, Akademische Rätin (a. Z.), Institut für Schulentwicklungsfroschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Soziale Schichtung und Bildungsungleichheiten, Lernen in formalen und informalen Bildungsumfeldern, Schulentfremdung, Bildungsverläufe.
Dr. Alyssa Laureen Grecu
Institut für Schulentwicklungsforschung (IFS)
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
Alyssa.Grecu@tu-dortmund.de

Prof. Dr. Johanna Heitzer, Universitätsprofessorin der Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen für das Lehr- und Forschungsgebiet Didaktik der Mathematik. Forschungs- und Arbeitsschwerpunkte: Mathematikdidaktik, Aus- und Weiterbildung von Mathematiklehrkräften, Unterrichtsforschung im Bereich der Mathematik.
Prof. Dr. Johanna Heitzer
RWTH Aachen
Kreuzherrenstraße 2–4
52062 Aachen
johanna.heitzer@matha.rwth-aachen.de

PD Dr. Dr. habil. Malte Jansen, seit 2017 wissenschaftlicher Leiter des Forschungsdatenzentrums am Institut zur Qualitätsentwicklung im Bildungswesen (IQB), An-Institut an der Humboldt-Universität zu Berlin. Forschungs- und Arbeitsschwerpunkte: Motivationale Schüler*innenmerkmale, soziale Beziehungen in der Schule, Reproduktion, Replikation und Sekundärdatenanalysen in der Bildungsforschung.
Dr. Malte Jansen
Humboldt-Universität zu Berlin
Institut zur Qualitätsentwicklung im Bildungswesen
Unter den Linden 6
10099 Berlin
Malte.Jansen@IQB.HU-Berlin.de

Ruben Kleinkorres, M.Sc., wissenschaftlicher Mitarbeiter am Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Wohlbefinden im schulischen Kontext, kompetenzbedingende Faktoren.
Ruben Kleinkorres
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
ruben.kleinkorres@tu-dortmund.de

Dr. Aleksander Kocaj, seit 2012 wissenschaftlicher Mitarbeiter am Forschungsdatenzentrum des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB), An-Institut an der Humboldt-Universität zu Berlin. Forschungs- und Arbeitsschwerpunkte: Schüler*innen mit sonderpädagogischem Förderbedarf (SPF) in Schulleistungsstudien, Auswirkungen des gemeinsamen Unterrichts auf Schüler*innen ohne SPF, Kompositionseffekte, Reproduktion, Replikation und Sekundärdatenanalysen in der Bildungsforschung
Dr. Aleksander Kocaj
Humboldt-Universität zu Berlin
Institut zur Qualitätsentwicklung im Bildungswesen
Unter den Linden 6
10099 Berlin
a.kocaj@iqb.hu-berlin.de

Dr. Maria Krieg, wissenschaftliche Mitarbeiterin am Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ), Hamburg. Forschungs- und Arbeitsschwerpunkte: Bildungsmonitoring, Schul- und Unterrichtsentwicklung, Schul- und Klassenprofilierung, Entwicklung und Förderung motivational-affektiver Merkmale von Lernenden, Bildungsungleichheiten, Methodenpluralität.
Dr. Maria Krieg
Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ)
Beltgens Garten 25
20537 Hamburg
maria.krieg@ifbq.hamburg.de

Prof. Dr. Poldi Kuhl, seit 2021 Professorin für Pädagogische Psychologie an der Leuphana Universität Lüneburg. Forschungs- und Arbeitsschwerpunkte: Empirische Schul- und Unterrichtsforschung zu Fragen der Qualitätssicherung und -entwicklung im Bildungssystem, Schulische Inklusion, Digitales Lehren und Lernen sowie Lehrkräfteprofessionalisierung.
Prof.in Dr.in Poldi Kuhl
Fakultät Bildung – Institut für Psychologie
Leuphana Universität Lüneburg
Universitätsallee 1, C1.226
21335 Lüneburg
poldi.kuhl@leuphana.de

Prof. Dr. Fani Lauermann, Professorin für Empirische Bildungsforschung und Pädagogische Psychologie, Rheinische Friedrich-Wilhelms-Universität Bonn; Assoziiertes Mitglied: Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Bildungsverläufe, Berufswahlentscheidungen, Motivation und Leistung, berufliche Kompetenzen von Lehrkräften.
Prof. Dr. Fani Lauermann
Bonner Zentrum für Lehrerbildung (BZL),
Poppelsdorfer Allee 15
53115 Bonn
fani.lauermann@uni-bonn.de; fani.lauermann@tu-dortmund.de

PD Dr. Ramona Lorenz, seit 2009 wissenschaftliche Mitarbeiterin am Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Empirische Schul- und Unterrichtsforschung, Digitale Medien in Schule und Unterricht, Leseleitungen von Schüler*innen.
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
ramona.lorenz@tu-dortmund.de

Dr. Ulrich Ludewig, wissenschaftlicher Mitarbeiter, Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkt: Lesekompetenz, Large-Scale-Assessment, Testentwicklung.
Dr. Ulrich Ludewig
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
ulrich.ludewig@tu-dortmund.de

Prof. Dr. Oliver Lüdtke, seit 2015 Direktor der Abteilung Pädagogisch-Psychologische Methodenlehre am IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Professor für Pädagogisch-Psychologische Methodenforschung an der Christian-Albrechts-Universität zu Kiel, Zentrum für internationale Bildungsvergleichsstudien (ZIB). Forschungs- und Arbeitsschwerpunkte: Mehrebenenanalyse, Methodik von Schulleistungsstudien, Anwendung Bayesianischer Verfahren, Schätzung von kausalen Effekten mit nicht-experimentellen Daten.
Prof. Dr. Oliver Lüdtke
IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
Abteilung Pädagogisch-Psychologische Methodenlehre
Olshausenstraße 62
24118 Kiel
oluedtke@leibniz-ipn.de

Prof. Dr. Nele McElvany, Geschäftsführende Direktorin des Instituts für Schulentwicklungsforschung (IFS) und Prorektorin Forschung der Technischen Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Lehr-/Lernforschung im schulischen Kontext; Kompetenzen von Lehrkräften und Unterrichtsqualität; Entwicklung und Förderung von Schriftsprachkompetenzen; Bildung und Migration; Pädagogisch-psychologische Diagnostik und Large-Scale Assessments.
Prof. Dr. Nele McElvany
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
nele.mcelvany@tu-dortmund.de

Prof. Dr. Benjamin Nagengast, Professor für Pädagogische Psychologie, Hector-Institut für Empirische Bildungsforschung, Universität Tübingen. Forschungs- und Arbeitsschwerpunkte: Effektivität von Bildungseinrichtungen und Bildungssystemen, Evaluation von Interventionen im Bildungskontext, Entwicklung und Förderung von Motivation und Persönlichkeit, quantitative Forschungsmethoden.
Prof. Dr. Benjamin Nagengast
Hector-Institut für Empirische Bildungsforschung
Universität Tübingen
Europastr. 6
72072 Tübingen
benjamin.nagengast@uni-tuebingen.de

Prof. Dr. Cora Parrisius, Juniorprofessorin, Professorin für Bildungswissenschaftliche Forschungsmethoden, Pädagogische Hochschule Karlsruhe. Forschungs- und Arbeitsschwerpunkte: Förderung von Motivation, Evaluation von Interventionen im Bildungskontext, Unterrichtsqualität und Motivation von Lehrenden und Lernenden, Einsatz intelligenter tutorieller Systeme.
Jun.-Prof. Dr. Cora Parrisius
PH Karlsruhe
Bismarckstr. 10
76133 Karlsruhe
cora.parrisius@ph-karlsruhe.de

Prof. Dr. Charlott Rubach, Assistant Professor at Rostock University. Research interests: Teacher education, teaching with digital media, teaching quality, motivation and emotion.
Prof. Dr. Charlott Rubach
Department of Education, Institute for School Pedagogy and Educational Science
August-Bebel-Straße 28
18055 Rostock
Charlott.rubach@uni-rostock.de

Dr. Ricarda K. Rübben, Akademische Rätin a. Z. am Institut für Erziehungswissenschaft, Otto-Friedrich-Universität Bamberg. Forschungs- und Arbeitsschwerpunkte: Profession und Lehrer:innenbildung, Wissenschafts- und Berufsfeldorientierung in der Lehrer:innenbildung, Schulpädagogik als Disziplin und Inklusion.
Dr. Ricarda K. Rübben
Lehrstuhl für Schulpädagogik
Otto-Friedrich-Universität Bamberg
Markusplatz 3
96047 Bamberg
ricarda.ruebben@uni-bamberg.de

Dr. Rahim Schaufelberger, seit 2014 wissenschaftlicher Mitarbeiter am Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: Bildungsdisparitäten, Bildungsentscheidungen und Berufsorientierung.
Dr. Rahim Schaufelberger
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
rahim.schaufelberger@tu-dortmund.de

Dr. Sina Schürer, Akademische Rätin am Institut für Erziehungswissenschaft der Universität Münster. Forschungs- und Arbeitsschwerpunkte: Soziale Partizipation von Schülerinnen und Schülern in inklusiven schulischen Settings; Professionalität von Lehrkräften mit Fokus auf Einstellung zu Inklusion, Pädagogische Beziehungen sowie Diagnostik im Kontext des Grundschulübergangs.
Dr. Sina Schürer
Institut für Erziehungswissenschaft
Universität Münster
Georgskommende 33
48143 Münster
schuerer@uni-muenster.de

Prof. Dr. Tobias C. Stubbe, Universitätsprofessor für Quantitative Methoden der Sozialwissenschaften und Statistik, Geschäftsführender Direktor des Instituts für Methoden und methodologische Grundlagen der Sozialwissenschaften (IMMS), Georg-August-Universität Göttingen. Forschungs- und Arbeitsschwerpunkte: Quantitative Forschungsmethoden / Statistik, Large-Scale-Assessments, Soziale Disparitäten im Bildungssystem, Soziale Netzwerke in Schulklassen.
Prof. Dr. Tobias C. Stubbe
Institut für Methoden und methodologische Grundlagen der Sozialwissenschaften
Georg-August-Universität Göttingen
Goßlerstr. 19
37073 Göttingen
tstubbe@uni-goettingen.de

Prof. Dr. Stefanie van Ophuysen, Universitätsprofessorin für Methoden der empirischen Bildungsforschung am Institut für Erziehungswissenschaft der Universität Münster. Forschungs- und Arbeitsschwerpunkte: Soziale Beziehungen im (inklusiven) Schulkontext; Pädagogische Diagnostik & diagnostische Kompetenz von Lehrkräften; Grundschulübergang sowie Motivationsförderung in der akademischen Bildung.
Prof.'in Dr. Stefanie van Ophuysen
Institut für Erziehungswissenschaft
Universität Münster
Georgskommende 33
48143 Münster
vanOphuysen@uni-muenster.de

Prof. Dr. Luise von Keyserlingk, until 14.05.2023 Postdoctoral Scholar at TU Dortmund University, Center for Research on Education and School Development. Since 15.05.2023 Assistant Professor at University of Tübingen. Research interests: Motivation and educational choices, self-regulated learning, mental health, higher education research.
Prof. Dr. Luise von Keyserlingk,
University of Tübingen
Hector Research Institute of Education Sciences and Psychology
Europastraße 6
72072 Tübingen
luise.keyserlingk@tu-dortmund.de

Dr. Sittipan Yotyodying, wissenschaftlicher Mitarbeiter am Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund. Forschungs- und Arbeitsschwerpunkte: schulisches Elternengagement, partnerschaftliche Zusammenarbeit von Schule und Eltern, schulische Medienbildungsforschung und digitale Interventionsforschung.
Dr. Sittipan Yotyodying
Institut für Schulentwicklungsforschung
TU Dortmund
Vogelpothsweg 78
44227 Dortmund
sittipan.yotyodying@tu-dortmund.de

Prof. Dr. Renzhe Yu, Assistant Professor at Teachers College, Colombia University. Research interests: Learning analytics, educational data mining.
Renzhe Yu, PhD
Teachers College Colombia University
460 GDodge
525 West 120th Street
New York, NY 10027
renzheyu@tc.columbia.edu

Silke Birgitta Gahleitner
Soziale Arbeit als Beziehungsprofession
Bindung, Beziehung und Einbettung professionell ermöglichen
2017, 388 Seiten, broschiert
ISBN: 978-3-7799-3477-6
Auch als E-BOOK erhältlich

In der Sozialen Arbeit wird zwar nicht bezweifelt, dass die Qualität der jeweiligen Hilfe unmittelbar an das Gelingen einer professionellen Beziehung gekoppelt ist. Wie man sie professionell ermöglicht, darüber bestehen jedoch nach wie vor viele Unklarheiten. Drei Studien der Autorin zeigen überzeugend auf, dass das Gelingen von Hilfe eine authentische, emotional tragfähige, von Nähe geprägte und dennoch reflexiv und fachlich durchdrungene Diagnostik und Beziehungsführung erfordert. Es zeigt sich darüber hinaus, dass Hilfe besonders dann gelingt, wenn beziehungserschütterte KlientInnen die Möglichkeit zu persönlichen Beziehungsdimensionen und -momenten erhalten, die als Alternativerfahrungen zu früheren Beziehungserschütterungen den Weg zurück in soziale Zusammenhänge bahnen. Die Ergebnisse werden in einem Entwurf verknüpft, der die Entwicklung von der ‚schützenden Inselerfahrung' zum ‚persönlich geprägten Netzwerk' prozessorientiert veranschaulicht.

www.beltz.de
Beltz Juventa · Werderstraße 10 · 69469 Weinheim

Milena Feldmann | Markus Rieger-Ladich |
Carlotta Voß | Kai Wortmann (Hrsg.)
Schlüsselbegriffe der
Allgemeinen Erziehungswissenschaft
Pädagogisches Vokabular in Bewegung
2022, 472 Seiten, Hardcover
ISBN: 978-3-7799-6819-1
Auch als E-BOOK erhältlich

Das pädagogische Vokabular wird von Begriffen geprägt, die manche als »einheimische« kennzeichnen. Aber es gibt auch eine Vielzahl von »Neuankömmlingen«, die in Anspruch genommen werden, wenn es darum geht, die Veränderungen des pädagogischen Feldes zu beobachten und auf den Begriff zu bringen.
Das Buch reflektiert diese Entwicklung, verschafft einen lesbaren, prägnanten Überblick über das pädagogische Vokabular und setzt sich kritisch mit der Genese, Bestimmung und Verwendung der Begriffe auseinander. Dabei geht es darum, den pädagogischen Diskurs in seiner Widersprüchlichkeit, Ungleichzeitigkeit und Dynamik möglichst unvoreingenommen und multiperspektivisch zum Gegenstand zu machen.
Auf diese Weise wird das Buch zu einem attraktiven Nachschlagewerk und Ideengeber für Student*innen, Doktorand*innen und Fachwissenschaftler*innen gleichermaßen.

Nkechi Madubuko
Praxishandbuch Empowerment
Rassismuserfahrungen von Kindern
und Jugendlichen begegnen
2021, 228 Seiten, broschiert
ISBN: 978-3-7799-6478-0
Auch als E-BOOK erhältlich

Rassismuserfahrungen von Kindern und Jugendlichen verlangen nach einer Antwort der pädagogischen Professionen und der Sozialen Arbeit. Aus der Betroffenenperspektive heraus analysiert die Autorin, welche Haltung, Reflexion und welches Wissen als Fachkraft unabdingbar ist, um Rassismus zu erkennen und Empowerment mitzudenken. Empowerment-orientierte Handlungskompetenz, Umgang mit Unterschieden, Wirkungen von Rassismuserfahrungen und wie Diskriminierung zu begegnen ist, wird praxisnah vorgestellt und mit Beispielen zur Umsetzung verknüpft.
Das Besondere: Erstmals stellen Empowerment-Trainer_innen im Buch ihre Methoden in geschützten Räumen (Safer Spaces) vor.

www.beltz.de
Beltz Juventa · Werderstraße 10 · 69469 Weinheim