

Edition ZfE

Jeanette Ziehm-Eicher
Marcus Hasselhorn
Hans-Günther Roßbach *Hrsg.*

Kinder mit erhöhtem Risiko für Bildungsmisserfolg

OPEN ACCESS

 Springer VS

Edition ZfE

Band 20

Die Reihe ‚Edition ZfE‘ wird von den Herausgeber(inne)n der Zeitschrift für Erziehungswissenschaft verantwortet. In der Reihe werden Originalbeiträge publiziert, die den strengen Qualitätsmaßstäben für die Publikation von Manuskripten in der Zeitschrift standhalten. Veröffentlicht werden von Expert(inn)en begutachtete erstklassige Beiträge zu aktuellen Befunden und Entwicklungen der Erziehungswissenschaft und Bildungsforschung. Die Zahl solcher Beiträge übersteigt die Möglichkeiten der Publikation in der Zeitschrift für Erziehungswissenschaft. Mit der ‚Edition ZfE‘ wird hier Spielraum eröffnet. Durch die Auswahl von Herausgeber(inne)n und die Themenwahl stehen die Beiträge zur Buchreihe ebenso wie die ZfE selbst für den interdisziplinären Charakter einer umfassenden Erziehungswissenschaft, deren Gegenstand der gesamte Lebenslauf des Menschen ist. Die gezielte Aufnahme internationaler Beiträge gewährleistet den Anschluss an erziehungswissenschaftliche Entwicklungen außerhalb Deutschlands. Die Leser(innen) der ‚ZfE-Edition‘ verfügen somit über eine zusätzliche Informationsquelle, die ihnen die für Erziehung und Bildung wichtigen internationalen und interdisziplinären Entwicklungen in weiten Bereichen der Erziehungswissenschaft zuverlässig, nüchtern und nachvollziehbar präsentiert.

Jeanette Ziehm-Eicher ·
Marcus Hasselhorn ·
Hans-Günther Roßbach
(Hrsg.)

Kinder mit erhöhtem Risiko für Bildungsmisserfolg

 Springer VS

Hrsg.

Jeanette Ziehm-Eicher
DIPF | Leibniz-Institut für Bildungsforschung und
Bildungsinformation
Frankfurt am Main, Deutschland

Marcus Hasselhorn
DIPF | Leibniz-Institut für Bildungsforschung und
Bildungsinformation
Frankfurt am Main, Deutschland

IDEa – Center for Individual Development
and Adaptive Education of Children at Risk
Frankfurt am Main, Germany

IDEa – Center for Individual Development
and Adaptive Education of Children at Risk
Frankfurt am Main, Germany

Hans-Günther Roßbach
Universität Bamberg
Bamberg, Deutschland



ISSN 2512-0778

ISSN 2512-0786 (electronic)

Edition ZfE

ISBN 978-3-658-48673-0

ISBN 978-3-658-48674-7 (eBook)

<https://doi.org/10.1007/978-3-658-48674-7>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://portal.dnb.de> abrufbar.

Dieses Werk wurde gefördert durch DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation.

© Der/die Herausgeber bzw. der/die Autor(en) 2025. Dieses Buch ist eine Open-Access-Publikation.

Open Access Dieses Buch wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor*in(nen) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Buch enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des/der betreffenden Rechteinhaber*in einzuholen.

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede kommerzielle Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des/der Autor*in und ggf. des/der Herausgeber*in. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Der Verlag hat eine nicht-exklusive Lizenz zur kommerziellen Nutzung des Werkes erworben.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jede Person benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des/der jeweiligen Zeicheninhaber*in sind zu beachten.

Der Verlag, die Autor*innen und die Herausgeber*innen gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autor*innen oder die Herausgeber*innen übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer VS ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Wenn Sie dieses Produkt entsorgen, geben Sie das Papier bitte zum Recycling.

Vorwort¹

Erfolgreiche Bildung ist eine wesentliche Voraussetzung für gesellschaftliche Teilhabe und individuelle Entwicklung. Unser Bildungssystem versucht, möglichst allen Kindern eine gute Bildung zuteil werden zu lassen. Das gelingt jedoch leider nicht immer. Bildungswege und -erfolge sind oft ungleich verteilt, und zahlreiche Kinder stehen vor Herausforderungen, die ihre schulischen und persönlichen Entwicklungsmöglichkeiten einschränken – manchmal so sehr, dass ihre Bildungswege durch eine nicht überwindbare Kette von Misserfolgen charakterisiert zu sein scheint. Sozioökonomische Benachteiligungen, sprachliche Hürden oder individuelle Entwicklungsbedingungen erhöhen das Risiko für individuellen Bildungsmisserfolg und erfordern gezielte Unterstützung (z. B. Hartmann et al. 2017; Hasselhorn 2021).

Der vorliegende Sammelband vereint aktuelle Forschungsergebnisse und praxisorientierte Perspektiven zu Kindern mit erhöhtem Risiko für Bildungsmisserfolg. Aus der Perspektive unterschiedlicher Disziplinen beleuchten die Beiträge verschiedene Einflussfaktoren, darunter sozioökonomische Bedingungen, sprachliche Heterogenität sowie die Rolle von Lehrkräften und adaptiven Unterrichtskonzepten. In ihrer Zusammenschau können sie dazu beitragen, ein umfassenderes Verständnis von Bildungsrisiken und den ihnen zugrunde liegenden Mechanismen zu ermöglichen und so die Diskussion von Lösungsansätzen für eine chancengerechtere Bildung zu bereichern.

Der Band entstand in Anlehnung an ein Forum des Teams der Herausgebenden der Zeitschrift für Erziehungswissenschaft im Dezember 2023, das in Kooperation mit dem Frankfurter IDeA-Zentrum durchgeführt wurde. Der Schwerpunkt

¹ Teile des Vorworts wurden mit Hilfe von künstlicher Intelligenz generiert und anschließend manuell überarbeitet.

des IDeA-Zentrums (IDeA steht für Individual Development and Adaptive Education of Children at Risk) liegt auf der empirischen Untersuchung von Kindern im Kleinkind-, Kindergarten- und Grundschulalter, die aufgrund verschiedener Faktoren – (neuro-)kognitiver oder sozio-emotionaler Art – ein erhöhtes Risiko für eine beeinträchtigte Entwicklung schulischer Fertigkeiten haben. IDeA initiiert und realisiert Untersuchungen zu den Entwicklungsverläufen dieser Kinder und nähert sich der Forschungsfrage, mit welchen vorschulischen und schulischen Lernarrangements sie bestmöglich in ihrer Entwicklung gefördert werden können.

Der Band gliedert sich in drei Themenbereiche: sozioökonomisch bedingte Risikofaktoren, die Rolle von Lehrkräften sowie Anforderungen an eine adaptive Schule aus Sicht von Schüler*innen. Der erste Themenbereich umfasst vier Kapitel, die sich mit kumulativen Risikofaktoren, der Auswirkung von Bildungsrisiken auf die kindliche Selbstregulation und die Entwicklung des Wortschatzes bei Kindern mit Migrationshintergrund beschäftigen. Außerdem werden Bildungskontexte als Ort für die Ausübung sexueller Gewalt beleuchtet. Der zweite Themenkomplex behandelt in zwei Kapiteln die Kompetenz von Lehrkräften hinsichtlich der Sprachförderung in heterogenen Gruppen sowie der Inklusion von Kindern im Autismus-Spektrum. Der letzte Themenkomplex widmet in sich in drei Kapiteln der Perspektive der Schüler*innen und adressiert das Thema mehrsprachigkeitssensible Unterrichtskonzepte, adaptive Lernmöglichkeiten sowie ein Training zur Steigerung der Leseflüssigkeit. Alle Beiträge haben ein doppelblindes Begutachtungsverfahren (Peer Review) durchlaufen.

Unser Dank gilt den Autor*innen, die mit ihren wissenschaftlichen Beiträgen neue Erkenntnisse und wertvolle Impulse liefern. Ebenso danken wir allen, die an der Entstehung dieses Bandes beteiligt waren.

Wir hoffen, dass dieser Band einen Beitrag zur wissenschaftlichen Diskussion für mehr Bildungsgerechtigkeit leistet und Argumente für Entscheidungsträger*innen in der Bildungspolitik bereitstellt, um die Bedeutung einer gezielten Unterstützung für Kinder mit erhöhtem Risiko für Bildungsmisserfolg weiter in den Fokus zu rücken.

Literatur

- Hartmann, U., Hasselhorn, M., & Gold, A. (Hrsg.). (2017). *Entwicklungsverläufe verstehen – Kinder mit Bildungsrisiken wirksam fördern: Forschungsergebnisse des Frankfurter IDeA-Zentrums*. Kohlhammer.
- Hasselhorn, M. (2021). Lernstörungen: Ein unvermeidbares Schicksal? *Zeitschrift für Pädagogische Psychologie*, 36, 1–17. <https://doi.org/10.1024/1010-0652/a000324>

Inhaltsverzeichnis

1 Cumulative Socioeconomic Risk Factors and Infant Temperament	3
Dave Möwisch, Annika Susann Wienke, Emilija Meier-Faust, Birgit Mathes und Manja Attig	
Teil I Sozioökonomisch bedingte Risikofaktoren	
2 Risk Factors for Academic Underachievement and Young Students' Self-Regulated Learning	27
Bernadette van Berk und Charlotte Dignath	
3 Addressing the Heterogeneity of German Language Competence in Children with Immigrant Background: Patterns of Vocabulary Development from Age Three to Nine	73
Christian Lohmann und Birgit Becker	
4 Sexuelle Gewalt in der Kindheit und Bildung – Wissensgenerierung durch Aufarbeitung und Zeugenschaft	103
Sabine Andresen und Andrea Pohling	
Teil II Die Rolle von Lehrkräften	
5 Preschool Teachers' Language Scaffolding in Early Science Education in Linguistically Heterogeneous Groups	129
Ada Cecil Haen, Ilonca Hardy, Henrik Saalbach, Miriam Leuchter und Mirjam Steffensky	

-
- 6 Die Kompetenz von Lehrkräften im inklusiven Unterricht von Kindern im Autismus-Spektrum: Entwicklung und Validierung eines Assessments** 163
Nick Gerrit Hasche, Mareike Kunter, Ann-Kathrin Krause und Charlotte Dignath
- Teil III Anforderungen an eine adaptive Schule aus Sicht der Schüler*innen**
- 7 Zur Rolle von Überzeugungen zu Mehrsprachigkeit für die Akzeptanz und den Lernerfolg eines mehrsprachigkeitssensiblen Unterrichtskonzepts** 191
Victoria Bertram, Dominique Rauch, Valentina Reitenbach, Katharina M. Schneider und Jasmin Decristan
- 8 Adaptiver Unterricht? Übereinstimmungen und Unterschiede in den Schüler*innenwahrnehmungen von Unterrichtsqualität** 217
Katharina Schnitzler, Nora Fröhlich, Jasmin Decristan, Benjamin Fauth und Hanna Dumont
- 9 FLINK: Evaluation of a Digitized Reading Fluency Training** 255
Fenke Kachisi, Jan-Henning Ehm, Jelena Marković und Telse Nagler

Teil I

Sozioökonomisch bedingte Risikofaktoren



Cumulative Socioeconomic Risk Factors and Infant Temperament

1

Dave Möwisch, Annika Susann Wienke, Emilija Meier-Faust, Birgit Mathes and Manja Attig

Abstract

Numerous studies in recent decades have shown a significant association between familiar socioeconomic status (SES) and child development. Less is known about the cumulative effects of socioeconomic risk factors, which may have a greater impact on child outcomes than single factors, especially in the first years of life. The study utilized latent mixture models to investigate the effects of socioeconomic risk factors on earlier infant temperament, a key predictor of later child development. Data from $N = 271$ families living in socially and culturally disadvantaged neighborhoods in the city of Bremen and participating in the BRISE (Bremen Initiative to Foster Early Childhood Development) study were analyzed. The considered risk factors included parental education, household income, parental employment status, immigrant

D. Möwisch (✉) · M. Attig
Leibniz-Institut für Bildungsverläufe, Bamberg, Germany
E-Mail: dave.moewisch@lifbi.de

M. Attig
E-Mail: manja.attig@lifbi.de

A. S. Wienke · B. Mathes
University of Bremen, Bremen, Germany
E-Mail: annika.wienke@uni-bremen.de

B. Mathes
E-Mail: birgit.mathes@uni-bremen.de

E. Meier-Faust
German Institute for Research, Berlin, Germany
E-Mail: emeierfaust@diw.de

© Der/die Autor(en) 2025

J. Ziehm-Eicher et al. (Hrsg.), *Kinder mit erhöhtem Risiko für Bildungsmisserfolg*,
Edition ZfE 20, https://doi.org/10.1007/978-3-658-48674-7_1

3

background and single parenthood. Infant temperament was measured using four items from the Infant Behavior Questionnaire to assess negative affectivity. The latent mixture models showed three different groups of families: (1) families with high cumulative socioeconomic risk, (2) families with low cumulative socioeconomic risk, and (3) a mixed group of families. As hypothesized, infants from families in the group with increased risk factors exhibited significantly higher levels of negative affectivity compared to those in the lower risk group.

Keywords

Socioeconomic Status • Infant Temperament • Cumulative Risk • Infancy • Latent Mixture Models

Introduction

A growing interest in understanding and addressing social disparities in child development as well as when and how they emerge can be observed among researchers and policymakers. Numerous studies have demonstrated significant associations between socioeconomic status (SES) and cognitive, socioemotional and language development (Bradley & Corwyn, 2002; Letourneau et al., 2013). Previous research has focused primarily on competence development and the association to SES in the pre-school years and less on infancy (see e.g., Halle et al., 2009; Weinert et al., 2016). Mixed results, however, were shown for the association between SES and infant temperament, a crucial predictor in infancy for both further child development (Sanson et al., 2004) and parenting behavior (Paulussen-Hoogeboom et al., 2007). While some studies indicate that infants from lower-SES families tend to have a more difficult temperament (Conger et al., 2012; Jansen et al., 2009), other studies found no or relatively small associations (e.g., Matheny et al., 1987). As most of these studies tend to focus on preschool-aged children, the question arises as to whether socioeconomic disparities in infants' temperament are already evident in the very early phases of temperament development. Thus, we examine the association between socioeconomic risk factors and infant temperament.

Research has further demonstrated that taking cumulative risk into account has more explanatory power with regard to child outcomes than accounting for a single factor (Deater-Deckard et al., 1998). The present study utilizes latent mixture

models to identify different constellations of socioeconomic risk factors and their cumulative effects on infant temperament. Specifically, data on families from the BRISE study (Bremen Initiative to Foster Early Childhood Development; Schütte et al., 2020) is utilized having a particular focus on families from socially and culturally disadvantaged areas of the city of Bremen in Germany, which allows for a more in-depth examination of socioeconomic¹ risk factors.

Theoretical Background

Infant Temperament

The concept of *infant temperament* refers to individual differences in emotional, motor and attentional reactivity as well as self-regulation (Rothbart & Bates, 1998). Infant temperament is a heterogeneous construct with different dimensions (Sanson & Oberklaid, 2013). The sub-concept of *difficult temperament* is particularly important among these different theoretical approaches (Bates, 1980; Thomas & Chess, 1977). A difficult temperament is characterized by a high level of negative affectivity which is also one of the main dimensions of infant temperament (Paulussen-Hoogbeem et al., 2007). Children with a higher level of negative affectivity tend to react to stressors with high degrees of emotionality, including anger, fear or sadness (Rothbart et al., 1994).

Infant temperament, particularly negative affectivity, may impact on various child outcomes and other development-related aspects, such as parenting behavior (for a meta-analysis, see Paulussen-Hoogbeem et al., 2007; Samdan et al., 2020). More specifically, a more difficult infant temperament is associated with lower peer interaction quality (Acar et al., 2015; Sanson et al. 2004), higher psychopathology (Kostyrka-Allchorne et al., 2020), later internalizing and externalizing problem behavior (Eisenberg et al., 2009), and lower academic achievement (Coplan et al., 1999; Schoen & Nagle, 1994). It therefore seems promising to identify relevant factors that can predict or affect temperament – besides genetic aspects (Saudino, 2005). Psychosocial factors including sensitive parental care or maternal social support were shown to be particularly relevant to infant temperament (Wachs & Bates, 2001) and they may also moderate or mediate effects of

¹ In this study, the term “socioeconomic” is used to refer to risk factors. It should be noted that the analyses also consider sociodemographic factors, including immigrant background and single parenthood.

familial socioeconomic factors. Studies indicate that lower parental SES is associated with increased negative affectivity in school-age children (Conger et al., 2012). In a recent study, Gouge et al. (2020) calculated a cumulative risk score from binary SES variables and examined its association to negative affectivity. They could show a moderate correlation ($r = .44$) between the cumulative risk score and negative affectivity.

It thus appears promising to use a cumulative risk approach to investigate the association between socioeconomic risk factors and a child's negative affectivity, particularly in the early years when social disparities are likely to emerge.

Cumulative Risk Factors

To investigate social inequalities, studies often examine different aspects of SES, such as education, household income or occupational status. These studies often include different aspects of SES as separate predictors in one model (Bradley & Corwyn, 2002). This approach is limited in terms of multi-collinearity and effects can be distorted (Cohen et al., 2003). An alternative approach to addressing this limitation is the *cumulative risk factor model* (Evans et al., 2013). This theoretical approach assumes that a composite score of different risk factors is more predictive than using predictors separately since different risk exposures may share the same relevant variance. Those composite scores are not uncommon when examining the various aspects of SES due to their significant intercorrelations (Bornstein et al., 2003; White, 1982).

Cumulative indicators demonstrate good validity and predictability for various child outcomes. In their meta-analysis, Evans et al. (2013) summarized more than 140 studies that showed a significant cumulative risk main effect on children's language, cognitive or socio-emotional outcome for people aged under 19. Among these studies, an increased cumulative risk was found to be associated with delayed cognitive development (e.g., Laucht et al., 1997), poorer mental health (e.g., Gerard & Buehler, 2004), more internalizing and externalizing problem behaviors (e.g., Appleyard et al., 2005), and poorer reading and math skills (Morales & Guerra, 2006). Moreover, these studies examined various domains of risk factors that affect child development, including the health of both the child (e.g., birth weight) and mother (e.g., depression), structural socioeconomic characteristics (e.g., education, family income, occupational status), and psychosocial characteristics of the setting (e.g., social support or parenting style). Ongoing discussions have focused on the domains of risk factors that are particularly relevant for child development and the variables that should be considered for cumulative

composite scores (Evans et al., 2013). In a recent study, Bignardi et al. (2022) used canonical correlation analysis to test which dimensions of risk factors (e.g., SES, child or family characteristics) are particularly predictive for different child outcomes. They identified socioeconomic risk factors as belonging to the main components explaining variance across various cognitive and behavioral child outcomes.

This study therefore aims to examine the relationship between familial socioeconomic risk factors and the child's negative affectivity using a cumulative risk approach.

Our Study

Classical approaches to cumulative risk have primarily integrated dichotomous risk factors into a composite score. There are some crucial weaknesses (for an overview, see Evans et al., 2013). One central limitation is the loss of information resulting from dichotomizing variables, which can lead to variance restriction and a decrease in predictive power. Defining thresholds to create dichotomous variables can be challenging and a theoretical or empirical basis is needed. In addition, this approach assumes an additive model, which does not account for interactions between risk factors. To mitigate these shortcomings, the current study employed latent mixture models (Oberski, 2016) as an alternative approach to identifying latent groups of risk factors and to address the limitations of the classic approach using dichotomous variables. Another advantage of this approach concerns the identification of specific combinations of risk factors that extend beyond a cumulative pattern. For instance, it is conceivable that there are families who immigrated to Germany from affluent European countries and have an immigrant background yet exhibit a high level of family education and income. Conversely, refugee families also have an immigrant background but tend to have a lower level of education or income (e.g., Edele et al., 2021). The use of latent mixture models can be beneficial in identifying such specific constellations of risk factors.

Previous studies have often focused on larger cohort samples (e.g., Bignardi et al. 2022; Jansen et al., 2009), which provide substantial statistical power but are less focused on risk samples where cumulations of risk factors may be more common.

The effects of socioeconomic inequalities might be underestimated (Howe et al., 2013). Risk samples provide the opportunity to investigate the possible impact of socioeconomic circumstances in depth. For example, Appleyard et al. (2005) investigated the impact of cumulative risk factors on behavior outcomes

of at-risk urban children and found a positive association between the number of risk factors and externalizing/internalizing problem behavior. The risk sample of the present study refers to families living in socially and/or culturally disadvantaged suburban areas of the mid-sized city of Bremen in Germany (Schütte et al., 2020). We first aim to investigate whether different cumulative constellations of risk factors in this sample can be identified using latent mixture models.

As dependent variable, we focus on negative affectivity as a core dimension of infant temperament. Previous research has shown that there is a significant link between SES and infant temperament, suggesting that children from lower-SES families have higher levels of negative affectivity (Conger et al., 2012; Jansen et al., 2009, Strickhouser & Sutin, 2020). Generally, studies have focused on different aspects of infants' temperament during school age rather than infancy (Lengua et al., 2007). There is less research on whether socioeconomic disparities in infants' temperament are already evident in infancy as well as when and how they emerge. This is particularly relevant since this potential association can affect further temperament development (Strickhouser & Sutin, 2020) and adolescent behavior problems (Appleyard et al., 2005). Additionally, social disparities in infant temperament would also have implications for early interventions. Thus, we aim to also examine the effects of cumulative SES risk factors at birth on negative affectivity prospectively when the children are seven months old. We hypothesized that a greater cumulative risk of socioeconomic status factors is positively associated with children's negative affectivity. This would suggest that children from families with lower SES may exhibit higher levels of negative affectivity.

Method

Data and Sample

We used data from the longitudinal project BRISE. BRISE aims to investigate the developmental effects of a continuous support chain by systematically linking interventions that support socioeconomically disadvantaged families throughout their child's entire pre-school life period. This chain includes intervention programs to foster early child development starting during pregnancy and ending before the child starts school. One of these BRISE programs is *Tipp Tapp* (Sadowski, 2017), which is a preventive health advice program that provides low-threshold support to parents during their infants' first year of life.

Tipp Tapp offers postnatal early prevention advice by a specialist from the health and care sector as part of home visits. It is offered after birth, in the sixth and twelfth month of the child's life. Depending on their needs, families receive information about nutrition, infant care, vaccinations or early intervention programs, for example. *Tipp Tapp* is an outreach program meaning that it actively contacts all families living in the districts in which it operates. These districts typically are characterized by a high proportion of socio-economically disadvantaged and culturally diverse families.

Families were recruited from urban neighborhoods in the city of Bremen, Germany, with an above average of socioeconomically disadvantaged families (see for detailed information on these criteria Schütte et al., 2020). Participation in BRISE required that families met at least one of three family-related criteria: (1) low educational attainment of the parents – at least one of the parents did not complete a vocational training or at least one of the parents holds a school-leaving certificate after ten years (*Mittlere Reife*), (2), immigrant background – at least one of the parents or grandparents was born outside of Germany, (3) precarious family income – due to at least one parent being non-employed or a low-wage earner or both parents working part-time.

The BRISE sample comprises $N = 550$ children in total. Currently, data for 300 children is available as data preparation is still ongoing. Among the 300 children were eight pairs of siblings. Only the youngest participating child from each family was considered in the analysis to account for the statistical dependence of family characteristics, particularly socio-economic variables. Another 21 families were excluded from the analysis because they dropped out of the study before the questionnaire about the infants' temperament was administered. The final sample included data on $N = 271$ children and their families.

Data were derived from the screening interview (T1) as well as a regular interview appointment (T2). The screening interview was conducted via telephone to determine whether the families met the criteria for inclusion in the BRISE study and provided information regarding the socioeconomic position of the family regarding the described inclusion criteria. At that time, either the mothers were still pregnant ($M = 12.04$ ($SD = 8.67$) weeks before birth, $N = 123$) or the children were no more than 17 weeks old (children's age: $M = 8.41$ ($SD = 3.15$) weeks, $N = 148$).

The second measurement point was conducted when the children were approximately seven months old ($M = 221$ days, $SD = 25.78$). With two exceptions the data was collected before the outbreak of the COVID-19 pandemic and conducted through an interview at the families' homes. The other two families received a

paper–pencil questionnaire that had to be filled out by themselves. The T2 questionnaire included all additional information analyzed in this study, including an instrument assessing the children’s negative affectivity.

Measures

Dependent Variable

Children’s negative affectivity was measured using a short version of the German *Infant Behavior Questionnaire – Revised Version* (Bayer et al., 2015). The following four items were used: “If your child is tired, how often does he/she show signs of stress and discomfort?”, “If you are busy with something else and your child is not able to get your attention, how often does he/she cry?”, “If your child cannot get what it wants, how often does he/she get angry?”, “When it is time for bed and your child does not want to sleep, how often does he/she whine?”. All items were answered on a 7-point scale ranging from 0 (never) to 6 (always). The four items were used to model a single latent factor representing negative affectivity that demonstrated acceptable model fit (CFI = .959; SRMR = .033; RMSEA = .077).

Independent Variables

The effect of cumulative risk factors, measured at T1, on infant temperament was investigated by analyzing seven independent variables as maternal and parental education, household income, maternal and paternal employment status, immigrant background, and single parenthood.

Maternal and Paternal Education Maternal and paternal education were measured using the Standard International Classification of Education developed by UNESCO to compare education systems and educational attainment internationally (ISCED-2011; UNESCO, 2011). The ISCED comprises nine different levels: Level 0 = early childhood education, level 1 = primary education, level 2 = lower secondary education, level 3 = upper secondary education, level 4 = post-secondary non-tertiary education, level 5 = short-cycle tertiary education, level 6 = bachelor or equivalent, level 7 = master or equivalent, level 8 = doctorate or equivalent. Educational qualifications obtained abroad were assessed according to their official recognition in Germany in order to reflect their socioeconomic significance within the country where the families currently live.

Household Income The current net labor income of each household serves as another socioeconomic risk factor.

Maternal and Paternal Employment Status The employment status of both parents was represented by two dichotomous variables (1 = full-time, part-time or marginally employed, 0 = non-employed). Maternal employment status reflected the period prior to maternity leave.

Immigrant Background The immigrant background of the families was quantified using a dichotomous variable: 1 = at least one parent was born abroad and 0 = no parent was born abroad.

Single Parenthood Single parenthood was assessed using a single dichotomous item (1 = yes; 0 = no).

Control Variables

We used the gender and the age of the child (in days) as well as the age of the mother (in years) as control variables (all assessed in T2).

Families' participation in the early intervention program *Tipp Tapp* (Sadowski, 2017) was also used. Participation in the *Tipp Tapp* program was represented using a dichotomous variable: 0 = did not take part in *Tipp Tapp*, 1 = took part in at least one appointment.

Statistical Analyses

We used a two-step approach in order to examine the impact of cumulative SES risk factors on negative affectivity. In the first step, we computed latent mixture models (Oberski, 2016) to identify cumulative patterns of socioeconomic risk factors. For this purpose, various models with different numbers of latent groups were calculated and compared in terms of their model quality (Ferguson et al., 2020). The following model fit indicators were used: Bayesian Information Criterion (BIC), Sample-Adjusted BIC (SABIC), Lo-Mendell-Rubin Test (LMRT) and entropy. The model with the best fit indicators was selected for further analysis. In the second step, categorical variables were calculated based on the latent mixture models. These variables indicate how children are assigned to different latent risk groups. Finally, we conducted regression analyses via structural equation modeling using Mplus 8.4 (Muthén and Muthén 1998–2017) to analyze the

effects of the categorical variables representing children's risk group on children's negative affectivity. The models were estimated using robust maximum likelihood (MLR), which uses parameter estimates with standard errors and chi-squared statistics that are robust to non-normality and non-independence of the observations. Missing data were treated using full information maximum likelihood (FIML).

Results

Descriptive statistics for all study variables and their bivariate correlations are presented in Tables 1.1 and 1.2.

The results and model fit indicators of the latent mixture models are summarized in Table 1.3. The results show that the model fit indicators improve as the number of latent groups increases. The significant LMRT indicates that a model with three latent risk factor groups explains significantly more variance than a model with only two groups. For the model with four latent groups, the LMRT also shows a significant p -value. Therefore, it is not entirely clear whether a model with four latent groups would be more suitable. Since the four-group solution would result in one small group of $N = 31$, we chose to use the model with three latent groups for the next steps in the analysis.

In the following step, we compared the levels of the different SES factors of the assigned families between the three groups (see Table 1.4) to identify significant differences. The results of the pairwise comparative analyses across all groups can be found in the appendix.

The latent mixture models identified three distinct groups: (1) those with *high cumulative socioeconomic risk*, (2) those with *low cumulative socioeconomic risk*, and (3) a *mixed group*. The group with a high cumulative socioeconomic risk ($N = 76$) is characterized by significantly higher values on all risk factors compared to the other groups. In contrast, families from the low cumulative socioeconomic risk group ($N = 95$) have a significantly higher educational level of the parents and higher financial resources in comparison to the other two groups. When contrasted with the mixed group ($N = 100$), however, no significant differences are observed with regard to the parental employment status. Interestingly, the proportion of families with an immigrant background is significantly higher in the low cumulative socioeconomic risk group compared to the mixed group, despite the parents' level of education and household income being higher in the latter.

Table 1.1 Descriptive statistics of all study variables

	Mean/ %	(SD)	Minimum	Maximum	Missings in %
Risk factors (T1)					
Education mother (ISCED)	4.02	2.06	1	8	2.3
Education father (ISCED)	3.92	1.92	1	8	12.5
Household net income (in €)	2665	1399	200	7600	18.8
Employment status mother	50.1% un- or marginally employed mothers				1.5
Employment status father	78.4% un- or marginally employed fathers				4.4
Immigrant background	50.2% families with migration background				0.0
Single parenthood	9.4% families with a single parent				1.8
Temperament (negative affectivity; T2)					
Item 1: If your child is tired, how often does he/she show signs of stress and discomfort?	4.42	1.48	0	6	15.9
Item 2: When it is time for bed and your child does not want to sleep, how often does he/she whine?	3.65	1.97	0	6	18.1
Item 3: If you are busy with something else and your child is not able to get your attention, how often does he/she cry?	3.85	1.73	0	6	17.3
Item 4: If your child cannot get what he (she) wants, how often does he/she get angry?	3.96	1.69	0	6	16.6
Controls (T2)					
Age mother	31.94	5.71	18	48	12.9
Age infant (in days)	221	25.78	177	337	12.9
Participation Tipp Tapp	45.2% took part in <i>Tipp Tapp</i>				0.37
Gender infant	56.1% girls				2.9

Note. $N = 271$. Employment status mother/father: 0 = non-employed or marginally employed, 1 = full-time or part-time. Immigrant background: 0 = no parent was born abroad, 1 = at least one parent was born abroad. Single parenthood: 0 = no single parenthood, 1 = single parenthood. *Tipp Tapp*: 0 = did not take part in the program *Tipp Tapp*, 1 = took part in at least one appointment.

Table 1.2 Correlations (Spearman Rho or Kendall's Tau) of all study variables

Variables	1	2	3	4	5	6	7	8	9	10	11
1 Education mother (ISCED)	–										
2 Education father (ISCED)	.55*	–									
3 Household net income	.42*	.33*	–								
4 Employment status mother	.44*	.23*	.38*	–							
5 Employment status father	.25*	.23*	.34*	.34*	–						
6 Immigrant background	–.12*	–.12*	–.26*	–.32*	–.24*	–					
7 Single parenthood	–.16*	–.15*	–.24*	–.18*	–.17*	.11	–				
8 Negative affectivity ^a	–.08	–.09	–.06	.05	–.07	.03	.05	–			
9 Age mother	.44*	.27*	.44*	.45*	.16*	–.01	–.15*	.03	–		
10 Age children (in days)	–.02	.03	.13*	–.09	.04	–.05	.02	.02	–.02	–	
11 Participation Tipp Tapp	.22*	.17*	.14*	.14*	.11	.05	–.10	.05	.09	.03	–
12 Gender	–.05	.06	.02	.01	.02	–.06	–.03	–.11	–.04	.02	–.02

Note. * $p < .05$, $N = 220-271$. Employment status mother/father: 0 = non-employed or marginally employed, 1 = full-time or part-time. Immigrant background: 0 = no parent was born abroad, 1 = at least one parent was born abroad. *Tipp Tapp*: 0 = did not take part in the program *Tipp Tapp*, 1 = took part in at least one appointment. Single parenthood: 1 = no single parenthood, 2 = single parenthood. Gender: 0 = male, 1 = female.

^aManifest mean of indicators

Table 1.3 Model fit indices of the latent mixture models

Number of groups	LL	BIC	SABIC	LMRT (<i>p</i> -value)	Entropy
1	-2542.234	5140.490	5108.783	-	-
2	-2345.661	4792.161	4735.088	<.001	.989
3	-2260.060	4665.775	4583.337	<.001	.924
4	-2193.711	4577.894	4470.090	.029	.983

Note. $N = 271$. LL = Loglikelihood, BIC = Bayesian Information Criterion, SABIC = Sample-Adjusted BIC, LMRT = Lo-Mendell-Rubin Test.

Table 1.4 Comparisons of means and frequencies of the socioeconomic risk factors between three latent groups

Risk factors	Group 1 High Risk ($N = 76$)	Group 2 Low Risk ($N = 95$)	Group 3 Mixed ($N = 100$)	F/ Chi-square	<i>p</i>
Education mother (ISCED)	1.98	6.63	2.96	1948.00	< .001
Education father (ISCED)	2.52	5.15	3.55	49.37	< .001
Household net income (in €)	1383	3346	2839	37.54	< .001
Employment status mother ^a	1.7%	66.5%	71.4%	102.01	< .001
Employment status father ^a	45.9%	85.8%	93.5%	60.43	< .020
Immigrant background ^a	76.9%	48.5%	32.0%	37.74	< .001
Single parenthood ^a	19.4%	3.2%	7.9%	14.70	< .001

Note. ^aChi-square test was conducted. Group 1: high cumulative socioeconomic risk. Group 2: low cumulative socioeconomic risk. Group 3: mixed group. Employment status mother/father: 0 = non-employed or marginally employed, 1 = full-time or part-time. Immigrant background: 0 = no parent was born abroad, 1 = at least one parent was born abroad. Single parenthood: 0 = no single parenthood, 1 = single parenthood.

Cumulative Socioeconomic Risk Factors and Negative Affectivity

Regression analyses via structural equation modeling were conducted to analyze the effects of cumulative socioeconomic risk on children's negative affectivity (Model 1: high cumulative socioeconomic risk group vs. low cumulative socioeconomic risk group and mixed group; Model 2: low cumulative socioeconomic risk vs. high cumulative socioeconomic risk and mixed group; Model 3: mixed group vs. high and low cumulative socioeconomic risk groups). The results are

presented in Table 1.5. They indicate that mothers from families with a lower cumulative socioeconomic risk rated their children significantly lower on negative affectivity compared to mothers from families with a higher cumulative socioeconomic risk or from the mixed group ($B = -.196$; $SE = 0.08$; $p < 0.05$). In addition, in Model 1, mothers rated boys significantly higher in negative affectivity compared to girls. Finally, no significant association was found between participation in the *Tipp Tapp* intervention program and mother's ratings of children's negative affectivity.

Table 1.5 Standardized regression coefficients for negative affectivity

	Negative affectivity
Model 1^a	
Age of the child	.002 (.09)
Gender	-.181 (.09)*
Age of the mother	.067 (.09)
<i>Tipp Tapp</i>	.064 (.09)
Group 2 (low risk) and 3 (mixed) vs. Group 1 (high risk)	.135 (.11)
Model 2^b	
Age of the child	.009 (.08)
Gender	-.132 (.10)
Age of the mother	.104 (.08)
<i>Tipp Tapp</i>	.044 (.09)
Group 1 (high risk) and 3 (mixed) vs. Group 2 (low risk)	-.196 (.08)*
Model 3^c	
Age of the child	.008 (.08)
Gender	-.129 (.10)
Age of the mother	.034 (.08)
<i>Tipp Tapp</i>	.019 (.09)
Group 1 (high risk) and 2 (low risk) vs. Group 3 (mixed)	.128 (.09)

Note. $N = 271$. Standard errors are shown in brackets. * $p < .05$. Group 1: high cumulative socioeconomic risk. Group 2: low cumulative socioeconomic risk. Group 3: mixed group. Gender: 0 = male, 1 = female. *Tipp Tapp*: 0 = did not take part in the program *Tipp Tapp*, 1 = took part in at least one appointment

^aModel fit indices: CFI = .928, SRMR = .043, RMSEA = .029

^bModel fit indices: CFI = .996, SRMR = .039, RMSEA = .007

^cModel fit indices: CFI = .907, SRMR = .044, RMSEA = .033

Discussion

We examined the associations between cumulative socioeconomic risk factors and negative affectivity when the children were 7 months old. The study showed two main findings.

First, latent mixture models showed that three latent groups of socioeconomic risk factors could be distinguished in the sample. The first group comprised families with a higher risk on average for each of the seven risk indicators used, indicating an increased cumulative socioeconomic risk. These families typically have fewer financial resources, lower levels of education, are more likely to have an immigrant background, and a higher likelihood of unemployment. The second group consisted of families with a significantly lower cumulative risk. In comparison to the other two groups, this group exhibits significantly higher levels of parental education and income. No notable differences were observed with respect to the employment status of the parents and the proportion of families with single parenthood in comparison to a mixed group, which represents the third identified group from the latent mixture models. One potential explanation for this may lie in the fact that the mixed group is made up of different types of families. On the one hand, the mixed group includes families with an immigrant background, and on the other hand, families without immigrant background who also tend to have a lower level of parental education and financial resources. In contrast, the low cumulative socioeconomic risk group appears to include families with an immigrant background who have a higher level of education and income. However, it is important to note that the model with four latent groups also demonstrated acceptable model fit. In extension of the current findings, the analyses will be rerun for the full BRISE sample once data acquisition is completed. In general, the results indicate that there are not only cumulative patterns of risk factors that all occur simultaneously, but also distinct subgroups, as in the mixed group. This emphasizes the necessity of considering different risk factors simultaneously in order to take into account the different interactions and constellations between the risk factors and thus obtain more accurate results.

The second main finding indicates that children from families with a lower cumulative socioeconomic risk exhibit a lower level of negative affectivity. This is in line with previous research that could also demonstrate a positive association between socioeconomic hardship and negative affectivity (Conger et al., 2012; Jansen et al., 2009; Strickhouser & Sutin, 2020). It also extends the study by Gouge et al. (2020), who found a moderately positive correlation between cumulative socioeconomic risk factors and difficult child temperament at school age. The results of the present study demonstrate that this association exists far

earlier in the development of children's temperament. It is important, however, to interpret these findings with caution since the measurement of negative affectivity was based solely on the mothers' assessment through questionnaires and studies could show that parental reports of elementary school students' personalities are associated with familial SES as well as cultural capital (e.g., Meier-Faust & Watermann, 2024). Despite these potential biases, parent reports of child temperament demonstrate good predictive validity in terms of observable child behavior (Freund, 2019; Parade & Leerkes, 2008). Further research suggests that mothers' judgements of children's temperament are an important predictor of parenting behavior (Fields-Olivieri et al., 2017; Paulussen-Hoogbeem et al., 2007) highlighting the relevance of assessments via questionnaires.

Implications

The results indicate some theoretical and methodological implications, one of them being that social differences in children's temperament are already evident in the first year of life. This is particularly important since temperament is linked to further socioemotional development (Appleyard et al., 2005; Strickhouser & Sutin, 2020), academic achievement (Coplan et al., 1999; Schoen & Nagle, 1994), and parenting behavior (Paulussen-Hoogbeem et al., 2007). Regarding parenting behavior, the findings align with Freund et al. (2017), who demonstrated a negative correlation between negative affectivity and the quality of mother-child interaction as the number of risk factors increased. The results suggest that mothers facing socioeconomic risk factors, such as low income or single parenthood, may have fewer resources to meet their child's needs, resulting in less sensitive and stimulating interactions with their child. Children who grow up in challenging conditions may face additional developmental risks if they exhibit a difficult temperament, even at a very early age. Early intervention programs especially for at-risk samples should consider the relevance of a difficult temperament for further child development. This highlights the importance of interventions such as the BRISE project that focus on socioeconomically disadvantaged families. Although the results did not show a significant association between participation in the *Tipp Tapp* intervention program and infant temperament, further data and measurements in later stages of the intervention chain are necessary to gain more insight. It is important to note that *Tipp Tapp*'s main focus is on the health of mother and child, not especially on temperament. Also, the families had only participated in a maximum of two out of three appointments at the time of the analyses. Therefore, it is reasonable to conclude that there were no

significant effects of participating in *Tipp Tapp* on children's negative affectivity at this stage.

One methodological implication concerns the use of the cumulative risk model (Evans et al., 2013) and latent mixture models, which appears to be a useful approach for the sample used. Interestingly, the results of the structural equation models differ from the manifest, bivariate correlations of the individual risk factors. None of the socioeconomic risk factors showed a significant correlation with negative affectivity. The usefulness of the latent mixture models is also demonstrated in additional analyses we conducted, wherein a set of binary variables was calculated for each of the seven risk factors, which were then combined into a sum score with a range of 0 to 7. This procedure aligns with the conventional methodology of the cumulative risk factor model. In these analyses, however, the sum score did not demonstrate any significant associations with the child's negative affectivity. This may be attributed to the necessity of forming binary variables for this approach, which can potentially lead to challenges in determining the cut-off values. Thus, latent mixture models represent a viable alternative that should be considered for future studies. The use of a cumulative approach appears to enhance the predictive power of risk factors, particularly in smaller sample sizes. This highlights the potential value of this approach for future research.

Limitations and Conclusion

This study has some noteworthy limitations. First, one of the main limitations is that the analyses were based on data from only 271 BRISE families, as additional data still needs to be processed (the final sample will consist of 550 children). Regarding the latent mixture models, further analyses with the complete sample could reveal more latent groups as the number of families in each latent group increases. Second, we analyzed socioeconomic risk factors as predictors of infant temperament. Although Bignardi et al. (2022) showed that socioeconomic factors are a key dimension of risk factors for child development, the present study did not consider other potential risk factors, such as the health of the child and mother. Third, only information from the mothers about the children's negative feelings was used. Research suggests that there can be differences in reports of child temperament between parents (Tackett, 2011). Fourth, negative affectivity is only one of the central dimensions of a child's temperament, despite its considerable relevance in terms of further child development. Future studies should

consider other dimensions such as surgency or orienting, which may reveal different relationship patterns. Fifth, it was not possible to include data on household size into the analysis of income, as this information is not yet available.

In sum, we found significant social disparities regarding infant temperaments. This has far-reaching implications for the socioemotional development of children and early intervention programs that focus on socioeconomically disadvantaged families.

Acknowledgments This article is based on a project that was funded by the German Federal Ministry of Education and Research (grant no. 01NV1601A-G). The contents are the responsibility of the authors. Additional support was provided by the City of Bremen and the Jacobs Foundation.

Appendix

Table 1 Comparisons of means and frequencies of the socioeconomic risk factors between Group 1 (high risk) and Group 2 (low risk).

Risk factors	Group 1 High Risk (<i>N</i> = 76)	Group 2 Low Risk (<i>N</i> = 95)	<i>t</i> / Chi-square	<i>p</i>
Education mother (ISCED)	1.98	6.63	46.11	< .001
Education father (ISCED)	2.52	5.15	9.22	< .001
Household net income (in €)	1383	3346	8.53	< .001
Employment status mother ^a	1.7%	66.5%	74.51	< .001
Employment status father ^a	45.9%	85.8%	27.24	< .001
Immigrant background ^a	76.9%	48.5%	13.99	< .001
Single parenthood ^a	19.4%	3.2%	10.68	< .001

Note. ^aChi-square test was conducted. Group 1: high cumulative socioeconomic risk. Group 2: low cumulative socioeconomic risk. Employment status mother/father: 0 = non-employed or marginally employed, 1 = full-time or part-time. Immigrant background: 0 = no parent was born abroad, 1 = at least one parent was born abroad. Single parenthood: 0 = no single parenthood, 1 = single parenthood.

Table 2 Comparisons of means and frequencies of the socioeconomic risk factors between Group 1 (high risk) and Group 3 (mixed)

Risk factors	Group 1 High Risk (<i>N</i> = 76)	Group 3 Mixed (<i>N</i> = 100)	t/ Chi-square	<i>p</i>
Education mother (ISCED)	1.98	2.96	13.70	< .001
Education father (ISCED)	2.52	3.55	4.46	< .001
Household net income (in €)	1383	2839	7.33	< .001
Employment status mother ^a	1.7%	71.4%	85.49	< .001
Employment status father ^a	45.9%	93.5%	47.30	< .001
Immigrant background ^a	76.9%	32.0%	35.74	< .001
Single parenthood ^a	19.4%	7.9%	5.10	< .05

Note. ^aChi-square test was conducted. Group 1: high cumulative socioeconomic risk. Group 3: mixed group. Employment status mother/father: 0 = non-employed or marginally employed, 1 = full-time or part-time. Migration background: 0 = no parent was born abroad, 1 = at least one parent was born abroad. Single parenthood: 0 = no single parenthood, 1 = single parenthood.

Table 3 Comparisons of means and frequencies of the socioeconomic risk factors between Group 2 (low risk) and Group 3 (mixed)

Risk factors	Group 2 Low Risk (<i>N</i> = 95)	Group 3 Mixed (<i>N</i> = 100)	t/ Chi-square	<i>p</i>
Education mother (ISCED)	6.63	2.96	58.07	< .001
Education father (ISCED)	5.15	3.55	6.26	< .001
Household net income (in €)	3346	2839	2.25	< .05
Employment status mother ^a	66.5%	71.4%	0.37	n.s
Employment status father ^a	85.8%	93.5%	2.79	n.s
Immigrant background ^a	48.5%	32.0%	5.48	< .05
Single parenthood ^a	3.2%	7.9%	0.85	n.s

Note. ^aChi-square test was conducted. Group 2: low cumulative socioeconomic risk. Group 3: mixed group. Employment status mother/father: 0 = non-employed or marginally employed, 1 = full-time or part-time. Migration background: 0 = no parent was born abroad, 1 = at least one parent was born abroad. Single parenthood: 0 = no single parenthood, 1 = single parenthood.

References

- Acar, I. H., Rudasill, K. M., Molfese, V., Torquati, J., & Prokasky, A. (2015). Temperament and preschool children's peer interactions. *Early Education and Development, 26*(4), 479–495. <https://doi.org/10.1080/10409289.2015.1000718>
- Appleyard, K., Egeland, B., van Dulmen, M. H. M., & Sroufe, L. A. (2005). When more is not better: The role of cumulative risk in child behavior outcomes. *Journal of Child Psychology and Psychiatry, 46*(3), 235–245. <https://doi.org/10.1111/j.1469-7610.2004.00351.x>
- Bates, J. (1980). The concept of difficult temperament. *Merrill-Palmer Quarterly of Behavior and Development, 26*(4), 299–319.
- Bayer, M., Wohlking, F., Freund, J.-D., Ditton, H., & Weinert, S. (2015). *Temperament bei Kleinkindern: Theoretischer Hintergrund, Operationalisierung im Nationalen Bildungspanel (NEPS) und empirische Befunde aus dem Forschungsprojekt (NEPS Working Paper Nr. 58)*. Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel. https://www.neps-data.de/Portals/0/Working%20Papers/WP_LVIII.pdf
- Bignardi, G., Dalmaijer, E. S., & Astle, D. E. (2022). Testing the specificity of environmental risk factors for developmental outcomes. *Child Development, 93*(3), e282–e298. <https://doi.org/10.1111/cdev.13719>
- Bornstein, M. H., Hahn, C.-S., Suwalsky, J. T. D., & Haynes, O. M. (2003). The Hollingshead four-factor index of social status and the socioeconomic index of occupations. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 29–82). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*(1), 371–399. <https://doi.org/10.1146/annurev.psych.53.100901.135233>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd edn.). Taylor & Francis Ltd. <https://doi.org/10.4324/9780203774441>
- Conger, K. J., Martin, M. J., Reeb, B. T., Little, W. M., Craine, J. L., Shebloski, B., & Conger, R. D. (2012). Economic hardship and its consequences across generations. In V. Maholmes & R. B. King (Eds.), *The Oxford handbook of poverty and child development* (pp. 37–53). New York, NY: Oxford University Press.
- Coplan, R. J., Barber, A. M., & Lagace-Seguin, D. G. (1999). The role of child temperament as a predictor of early literacy and numeracy skills in preschoolers. *Early Childhood Research Quarterly, 14*(4), 537–553. [https://doi.org/10.1016/S0885-2006\(99\)00025-3](https://doi.org/10.1016/S0885-2006(99)00025-3)
- Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Psychopathology, 10*(3), 469–493. <https://doi.org/10.1017/S0954579498001709>
- Edele, A., Kristen, C., Stanat, P., & Will, G. (2021). The education of recently arrived refugees in Germany: Conditions, processes, and outcomes. *Journal of Educational Research Online, 13*(1), 5–15. <https://doi.org/10.25656/01:22064>
- Eisenberg, N., Valiente, C., Spinrad, T. L., Cumberland, A., Liew, J., Reiser, M., Zhou, Q., & Losoya, S. H. (2009). Longitudinal relations of children's effortful control, impulsivity,

- and negative emotionality to their externalizing, internalizing, and co-occurring behavior problems. *Developmental Psychology*, 45(4), 988–1008. <https://doi.org/10.1037/a0016213>
- Evans, G. W., Li, D., & Whipple, S. S. (2013). Cumulative risk and child development. *Psychological Bulletin*, 139(6), 1342–1396. <https://doi.org/10.1037/a0031808>
- Ferguson, S. L., Moore, G., & E. W., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development*, 44(5), 458–468. <https://doi.org/10.1177/0165025419881721>
- Fields-Olivieri, M. A., Cole, P. M., & Maggi, M. C. (2017). Toddler emotional states, temperamental traits, and their action: Associations with mothers' and fathers' parenting. *Journal of Research in Personality*, 67, 106–119. <https://doi.org/10.1016/j.jrp.2016.05.007>
- Freund, J. D. (2019). Early temperament in parental report and scientific observation. *Early Child Development and Care*, 189(14), 2318–2333. <https://doi.org/10.1080/03004430.2018.1450252>
- Freund, J.-D., Linberg, A., & Weinert, S. (2017). Einfluss eines schwierigen frühkindlichen Temperaments auf die Qualität der Mutter-Kind-Interaktion unter psychosozialen Risikolagen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 49(4), 197–209. <https://doi.org/10.1026/0049-8637/a000181>
- Gouge, N., Dixon, W. E., Jr., Driggers-Jones, L. P., & Price, J. S. (2020). Cumulative sociodemographic risk indicators for difficult child temperament. *The Journal of Genetic Psychology*, 181(1), 32–37. <https://doi.org/10.1080/00221325.2019.1699012>
- Gerard, J. M., & Buehler, C. (2004). Cumulative environmental risk and youth maladjustment: The role of youth attributes. *Child Development*, 75(6), 1832–1849. <https://doi.org/10.1111/j.1467-8624.2004.00820.x>
- Halle, T., Forry, N., Hair, E., Perper, K., Wandner, L., Wessel, J., & Vick, J. (2009). *Disparities in early learning and development: Lessons from the early childhood longitudinal study-birth cohort (ECLS-B)*. Child Trends.
- Howe, L. D., Tilling, K., Galobardes, B., & Lawlor, D. A. (2013). Loss to follow-up in cohort studies: Bias in estimates of socioeconomic inequalities. *Epidemiology*, 24(1), 1–9. <https://doi.org/10.1097/EDE.0b013e31827623b1>
- Kostyrka-Allchorne, K., Wass, S. V., & Sonuga-Barke, E. J. (2020). Research review: Do parent ratings of infant negative emotionality and self-regulation predict psychopathology in childhood and adolescence? A systematic review and meta-analysis of prospective longitudinal studies. *Journal of Child Psychology and Psychiatry*, 61(4), 401–416. <https://doi.org/10.1111/jcpp.13144>
- Jansen, P. W., Raat, H., Mackenbach, J. P., Jaddoe, V. W., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2009). Socioeconomic inequalities in infant temperament: The generation R study. *Social Psychiatry and Psychiatric Epidemiology*, 44(2), 87–95. <https://doi.org/10.1007/s00127-008-0416-z>
- Laucht, M., Esser, G., & Schmidt, M. H. (1997). Developmental outcome of infants born with biological and psychosocial risks. *Journal of Child Psychology and Psychiatry*, 38(7), 843–853. <https://doi.org/10.1111/j.1469-7610.1997.tb01602.x>

- Lengua, L. J., Honorado, E., & Bush, N. R. (2007). Contextual risk and parenting as predictors of effortful control and social competence in preschool children. *Journal of Applied Developmental Psychology, 28*(1), 40–55. <https://doi.org/10.1016/j.appdev.2006.10.001>
- Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic status and child development: A meta-analysis. *Journal of Emotional and Behavioral Disorders, 21*(3), 211–224. <https://doi.org/10.1177/1063426611421007>
- Matheny, A., Wilson, R., & Thoben, A. (1987). Home and mother: Relations with infant temperament. *Developmental Psychology, 23*(3), 323–331. <https://doi.org/10.1037/0012-1649.23.3.323>
- Meier-Faust, E., & Watermann, R. (2024). Perceiver effects and socioeconomic background: Contrasting parent-reports against teacher-reports of elementary school students' personality. *Journal of Personality Assessment, 106*(4), 482–495. <https://doi.org/10.1080/00223891.2023.2286449>
- Morales, J. R., & Guerra, N. G. (2006). Effects of multiple context and cumulative stress on urban children's adjustment in elementary school. *Child Development, 77*(4), 907–923. <https://doi.org/10.1111/j.1467-8624.2006.00910.x>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th edn). Muthén & Muthén.
- Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In J. Robertson & M. Kaptein (Eds.), *Modern statistical methods for HCI* (pp. 275–287). Human-Computer Interaction Series. Springer. https://doi.org/10.1007/978-3-319-26633-6_12
- Parade, S. H., & Leerkes, E. M. (2008). The reliability and validity of the infant behavior questionnaire-revised. *Infant Behavior and Development, 31*(4), 637–646. <https://doi.org/10.1016/j.infbeh.2008.07.009>
- Paulussen-Hoogbeem, M. C., Stams, G. J. M., Hermanns, J. M. A., & Peetsma, T. T. D. (2007). Child negative emotionality and parenting from infancy to preschool: A meta-analytic review. *Developmental Psychology, 43*(2), 438–453. <https://doi.org/10.1037/0012-1649.43.2.438>
- Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*(1), 21–39.
- Rothbart, M. K., & Bates, J. E. (1998). Temperament. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional and personality development* (5th edn., vol. 3, pp. 105–176). New York: Wiley.
- Sadowski, G. (2017). TIPP TAPP-Gesund ins Leben wird Teil der Langzeitstudie BRISE in der Stadtgemeinde Bremen. *Das Gesundheitswesen, 79*(4), 299–374. <https://doi.org/10.1055/s-0037-1601937>
- Samdan, G., Kiel, N., Petermann, F., Rothenfuß, S., Zierul, C., & Reinelt, T. (2020). The relationship between parental behavior and infant regulation: A systematic review. *Developmental Review, 57*, Article 100923. <https://doi.org/10.1016/j.dr.2020.100923>
- Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development, 13*(1), 142–170. <https://doi.org/10.1046/j.1467-9507.2004.00261.x>
- Sanson, A., & Oberklaid, F. (2013). Infancy and early childhood. In S. Vassallo & A. Sanson (Eds.), *The Australian temperament project: The first 30 years* (pp. 4–5). Melbourne, Australia: Australian Institute of Family Studies.

- Saudino, K. J. (2005). Behavioral genetics and child temperament. *Developmental and Behavioral Pediatrics*, 26(3), 214–223.
- Schoen, M. J., & Nagle, R. J. (1994). Prediction of school readiness from kindergarten temperament scores. *Journal of School Psychology*, 32(2), 135–147. [https://doi.org/10.1016/0022-4405\(94\)90007-8](https://doi.org/10.1016/0022-4405(94)90007-8)
- Schütte, K., Köller, O., Anders, Y., & Petermann, F. (2020). Kontinuität als Schlüssel zu wirksamer Förderung frühkindlicher Entwicklung benachteiligter Kinder. *Psychologie in Erziehung und Unterricht*, 67, 1–12. <https://doi.org/10.2378/peu2020.art27d>
- Strickhouser, J. E., & Sutin, A. R. (2020). Family and neighborhood socioeconomic status and temperament development from childhood to adolescence. *Journal of Personality*, 88(3), 515–529. <https://doi.org/10.1111/jopy.12507>
- Tackett, J. L. (2011). Parent informants for child personality: Agreement, discrepancies, and clinical utility. *Journal of Personality Assessment*, 93(6), 539–544. <https://doi.org/10.1080/00223891.2011.608763>
- Thomas, A., & Chess, S. (1977). *Temperament and development*. Brunner/Mazel.
- United Nations Educational, Scientific and Cultural Organization (UNESCO). (2011). *International Standard Classification of Education: ISCED 2011*. UNESCO Institute for Statistics.
- Wachs, T. D., & Bates, J. E. (2001). Temperament. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development: Handbooks of developmental psychology* (pp. 465–501). Malden, MA: Blackwell.
- Weinert, S., Linberg, A., Attig, M., Freund, J.-D., & Linberg, T. (2016). Analyzing early child development, influential conditions, and future impacts: Prospects of a German newborn cohort study. *International Journal of Child Care and Education Policy*, 10(1), 1–20. <https://doi.org/10.1186/s40723-016-0022-6>
- White, K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91(3), 461–481. <https://doi.org/10.1037/0033-2909.91.3.461>

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Risk Factors for Academic Underachievement and Young Students' Self-Regulated Learning

2

Bernadette van Berk and Charlotte Dignath

Abstract

Decades of research have repeatedly highlighted the relations between academic achievement and family contexts, such as socioeconomic status and immigration. Children facing such educational risks as well as those with learning and attention disorders encounter several challenges. A broad body of research emphasizes the relevance of self-regulated learning (SRL) for educational success. This study investigates the use of SRL strategies and metacognitive knowledge among students with educational risks. SRL among young learners was assessed using a multimethod approach. Results indicate relations with parents' socioeconomic status and self-reported SRL strategy use. Regarding immigration, results show that children from immigrant backgrounds show higher SRL strategy use in self-reports. Analyses also indicated associations of learning difficulties with lower SRL measured with the parent rating. No relations were found between the various risk factors and students' metacognitive knowledge. These findings offer first insights into the relations between SRL and educational risks, providing a basis for tailored interventions to enhance SRL.

B. van Berk (✉) · C. Dignath
Goethe-Universität Frankfurt am Main, Frankfurt am Main, Germany
E-Mail: vanberk@psych.uni-frankfurt.de

IDeA – Center for Research on Individual Development and Adaptive Education of Children at Risk, Frankfurt am Main, Germany

Keywords

Self-regulated learning • Students at risk • Metacognitive knowledge • Learning difficulties

Introduction

Numerous studies have shown that self-regulated learning (SRL) is a strong predictor of academic performance and motivation to learn (Dent & Koenka, 2016) and is therefore considered an important basis for lifelong learning (Bjork et al., 2013). Accordingly, students who plan their learning, set themselves goals, monitor and control their learning performance and use a variety of strategies tend to show higher motivation and self-efficacy beliefs (Bai & Guo, 2021; Schunk & Ertmer, 2000). On the other hand, deficits in SRL predict diverse future problems in academic contexts and social life (Robson et al., 2020; Sawyer et al., 2015). School closures and distance learning during the Covid-19 pandemic once again emphasized the importance of SRL in education (Berger et al., 2021). However, for many students, SRL presented a major challenge during the pandemic (Berger et al., 2021), and the dramatic learning deficits of the affected student cohorts are currently evident in national and international school performance studies (Betthäuser et al., 2023; OECD, 2023).

Not only during the pandemic but over the last decades, the strong influence of various educational risk factors on students' academic performance has been repeatedly demonstrated (OECD, 2023). These can be family-related risk factors such as socio-economic status (SES) and immigrant background, but also individual risk factors, including learning and attention difficulties as well as learning disorders (LD) and attention-deficit/hyperactivity disorder (ADHD) (Arnold et al., 2020; Büttner & Hasselhorn, 2011). Many students with educationally relevant risks have difficulties with learning that may be partly caused by a deficient SRL. However, to date, the extent to which students' SRL is related to educational risk factors has not been sufficiently explored. Nevertheless, previous research suggests that differences in SRL related to risk factors may exist in young children and that the identification of risk factors associated with SRL may provide crucial insights for overcoming educational barriers and promoting educational success. This study aims to investigate the extent to which students at educational risk are impaired in their SRL competencies. To this end, we investigate whether the use of SRL strategies and students' metacognitive knowledge are associated with

RQ1: socio-economic risks (SES and parental education).

RQ2: immigration-related risks (immigrant background and frequency German is spoken at home).

RQ3: developmental risks (learning and attention disorders and need for homework support).

The study thus provides insights into the interplay between risk factors and SRL in the context of, young students' learning in schools, which are relevant for both further research and for practice.

Theoretical Background

SRL and Metacognitive Knowledge

Self-regulated learners plan their learning approach, set themselves goals, use multiple strategies to advance their learning process and reflect on what could be improved (Zimmerman, 2000). At the heart of SRL is a variety of strategies that learners use to regulate their own learning process. These strategies include various cognitive, metacognitive, resource-oriented, emotion-regulatory and motivational processes that are used to achieve learning goals and enhance learning outcomes (Pintrich, 2000). While cognitive strategies, such as elaboration and organization, are used to regulate information processing, metacognitive strategies are involved in evaluating and adapting the learning process, for example through planning (Boekaerts, 1996). In addition, SRL theories highlight the need for students to maintain or increase their motivation and to regulate their emotions during the learning process, by using diverse strategies like positive self-talk or cognitive reappraisal (Boekaerts, 2007; Efklides, 2011). SRL has been shown to be a dynamic and cyclical process involving different phases. In the forethought phase, learners set themselves goals, plan their strategy use and activate prior knowledge (Zimmerman, 2000). This is followed by the performance phase, where individuals employ the planned strategies, monitor their progress, and adjust their approach as needed. Finally, the self-reflection phase involves evaluating the own performance and reflecting on improvement for future learning scenarios. Within this process, students constantly monitor their learning approach and evaluate it against their self-set goal or standard for their task performance (Pintrich, 2000).

The complex process of SRL can be seen as an overarching framework including different concepts and abilities (Dinsmore et al., 2008; Roebbers, 2017).

Besides the diverse SRL strategies, one key component of SRL is metacognition, which spans metacognitive regulation and metacognitive knowledge (Händel et al., 2013). The regulative aspect of metacognition, often integrated in the set of the various SRL strategies (Boekaerts, 1996), can be further distinguished and described as the interplay of monitoring and control of one's own cognitive and behavioral processes (Nelson & Narens, 1994). Metacognitive knowledge on the other hand describes the understanding of one's own learning strategies, tasks and abilities (Schneider, 2008). In order for learners to know when and how to use which SRL strategy most effectively, they need this metacognitive knowledge, i.e. knowledge about SRL strategies and about the conditions on which SRL strategies are used (Weinstein et al., 2000). Metacognitive strategy knowledge refers to an individual's awareness of strategies that can be employed to enhance learning (declarative knowledge), their understanding about how a strategy can be applied effectively (procedural knowledge) and which strategy is the appropriate approach to solve the task (conditional knowledge) (Flavell, 1979; Händel et al., 2013). Together metacognitive knowledge and SRL strategy use enable students to shape their learning process autonomously and proactively (Pintrich, 2000; Zimmerman, 2000).

Research with young children often focuses on self-regulation beyond the learning context, describing the ability to regulate emotions, behavior and cognition in diverse contexts (Moilanen, 2007; Wesarg-Menzel et al., 2023). Here, studies often focus on behavioral self-regulation or executive functions as a set of higher-order cognitive processes, like working memory, shifting and inhibition abilities (Roebbers, 2017). SRL can be seen as a concept which tries to integrate metacognition, cognitive and behavioral self-regulation, aiming to conjoin the different research traditions in the academic domain (Dinsmore et al., 2008). Despite the extensive research on SRL, the assessment of SRL, especially with young learners, such as primary school students, remains challenging. Especially in large samples, questionnaires remain the most common assessment method in SRL research due to their time-efficient administration, although their validity and reliability have been repeatedly questioned (Gascoine et al., 2017; Veenman, 2011; Veenman & van Cleef, 2019). More valid methods such as observations, interviews, and think-aloud protocols can produce a vast amount of verbal and nonverbal behavior data, which are resource-intensive and time-consuming in their preparation and analysis (Veenman & van Cleef, 2019). Furthermore, some of these approaches can only be applied in controlled laboratory environments, making them less suitable for practical use in school-based or digital research (Marulis et al., 2016; McClelland & Cameron, 2012). Given these challenges,

multi-method approaches to investigate SRL seem necessary (Veenman & van Cleef, 2019; Winne, 2010).

SRL and Academic Achievement

Empirical evidence supports the positive association between SRL and academic achievement across different populations and educational settings (Dent & Koenka, 2016). In line with this, some researchers have shown that self-regulatory abilities in a young age, involving regulation of emotions, attention, behavior and cognition, are associated with school readiness and early academic outcomes (e.g. Howard et al., 2021). School-aged students who employ effective SRL strategies tend to outperform their peers on academic tasks (Dent & Koenka, 2016) and display superior problem-solving abilities (e.g. Hatala et al., 2023; Paquette et al., 2021; Tzohar-Rozen & Kramarski, 2017). Moreover, previous research has highlighted the enduring impact of SRL on individuals' educational attainment and the risk for school drop-out (Meyers et al., 2013). Besides the positive relationship with academic achievement, several studies have emphasized the relation of SRL and motivational aspects, such as academic self-efficacy, task value and task interest (Callan et al., 2021; Lim & Yeo, 2021; Pintrich, 1999; Zimmerman and Schunk 2011).

Children with lower levels of self-regulation in early childhood have been found to be at risk of lower performance on academic tasks (McClelland et al., 2006; McClelland et al., 2000) and future behavior problems (Sawyer et al., 2015).

Based on the research showing that even young learners can apply SRL strategies (Bryce & Whitebread, 2012; Wesarg-Menzel et al., 2023), childhood is considered to be a critical period for the development of SRL. Metacognitive knowledge and regulation as components of SRL emerge in the early school years and become more differentiated with increasing age (Veenman et al., 2006). Metacognitive knowledge of strategies develop with increasing age, as does the ability to regulate behavior (Wigfield et al., 2011). Children as young as seven years are able to monitor and adjust their use of strategies (Schneider, 2008). Some studies indicate gender differences in SRL (Vandavelde et al., 2013), with girls tending to use more SRL strategies and more deep-level SRL strategies (Heirweg et al., 2019).

Despite the recognized relevance and well-documented benefits of SRL, many learners struggle to use SRL strategies effectively (Bjork et al., 2013). Nevertheless, meta-analyses show training effects even for young learners (Dignath et al.,

2008). Moreover, in particular children with various difficulties seem to benefit from these trainings (Berkeley & Larsen, 2018; Dignath & Hasselhorn, 2023). While the benefits of SRL are evident, the effective implementation of SRL in schools and the promotion of SRL pose challenges. Tailored training is needed to address individual differences in learners, such as cognitive abilities, motivational orientations, resources and socio-cultural backgrounds. Additionally, the integration of SRL principles into educational curricula requires collaboration between researchers and educators and an increased awareness of students who might face difficulties in SRL.

Risks for Academic Underachievement

National and international studies have repeatedly highlighted that students' school performance is related to demographic variables such as socio-economic status, parents' education and immigrant background (Mang et al., 2023; OECD, 2023; Reiss et al., 2019; Schwippert et al., 2020). These factors assumingly influence educational trajectories of children, due to, for example, availability of financial resources for educational material or activities and cultural capital, as well as by decisions regarding education (Barone, 2006; Schwippert et al., 2020). Distance learning during the Covid-19 pandemic also increased the influence of the different opportunities for parents to support their children and therefore the associated educational inequalities (Anger & Plünnecke, 2021; Betthäuser et al., 2023).

Considering the high prevalence of LD (Grigorenko et al., 2020; Hasselhorn, 2022) and ADHD (Schlack et al., 2007; Thomas et al., 2015) and the call for inclusive education (Convention on the Rights of Persons with Disabilities, 2007), students with learning and attention disorders are also at risk for educational underachievement and face challenges in education.

Socio-Economic Status as a Risk Factor

The recent international comparative PISA study (Programme for International Student Assessment) (OECD, 2023) once again highlights the relevance of the socio-economic status for various educational outcomes. A variety of indicators are used to measure SES, such as parental occupation, parental education, and family income (Sirin, 2005). In OECD countries, socio-economically disadvantaged students are significantly more likely to perform below the basic competency levels in mathematics and science, with a similar trend observed for literacy (OECD, 2023). A meta-analytic review suggests a medium to strong association

between students' SES and their academic achievement (Sirin, 2005). In addition to the relation between sociodemographic risk factors and actual academic performance (Berkowitz et al., 2017; O'Connor et al., 2019), students from low SES backgrounds often face difficulties in the motivational aspect (Hornstra et al., 2013) and tend to have higher school dropout rates (Winding & Andersen, 2015).

Parents' SES and parents' education are associated with childrens' educational pathways and their academic achievement in Germany (Anger & Plünnecke, 2021). For example, children from academic families are more likely to continue with higher education than children whose parents did not study (Lergetporer et al., 2021). Large-scale international comparative studies have repeatedly shown significant differences in math's, science (Schwippert et al., 2020) and reading skills (Reiss et al., 2019) between, for example, students with at least one parent with a tertiary education compared to students whose parents did not attend tertiary education.

Migration as a Risk Factor

Several studies have shown that immigrant students face significant disadvantages in the school context (Heath et al., 2008). Immigration is a complex phenomenon and the group of learners defined as having an immigrant background is often very heterogeneous and is not universally defined (Gogolin & Maaz, 2019; Heath et al., 2008). In order to take into account not only the migration experience of the students themselves, but also the indirect family migration experience which might influence educational trajectories, the currently widely accepted concept of migration is based on a generational approach (Gogolin & Maaz, 2019; Heath et al., 2008). Research typically considers at least one generation (parents' migration) or even more than one generation (grandparents' migration) to assess migration and in addition asks for further information, such as the time of migration and the countries of origin (Gogolin & Maaz, 2019).

Prior research suggests that students from immigrant backgrounds show lower academic performance in various areas (e.g. Heath et al., 2008; Kao & Thompson, 2003; OECD, 2023). Across different countries and outcome measures, a pattern has emerged showing that in western countries certain minorities whose parents come from less developed non-European countries are at risk of having a significantly lower level of educational attainment than their peers (Heath et al., 2008). For example, the international comparative PISA study has repeatedly shown that children from immigrant families have comparatively low reading skills in the national language across all OECD countries (OECD, 2023; Reiss et al., 2019). Moreover, some studies suggest that immigrant students are significantly more likely to drop out of school (e.g. Kao & Thompson, 2003) or do not

continue in higher education (e.g. Becker & Klein, 2021). Furthermore, there is empirical evidence that these negative educational trajectories already occur at a young age (e.g. Becker & Klein, 2021; Hahn & Schöps, 2019; Sprong & Skopek, 2022).

Immigrant students are often at risk in several ways. In addition to their immigrant background, their families often have a low SES as well and they rarely speak the national language at home, which can lead to several challenges in educational systems (OECD, 2023). The PISA study suggests that students with an immigrant background often speak a different language at home than the one used in schools, especially first-generation students born abroad (OECD, 2023). In many countries, the number of students speaking a different language at home has risen in recent years, posing a challenge for integration in the educational systems due to language barriers (OECD, 2023). In line with this, the PISA results for Germany suggest that children who speak German less frequently at home have significantly lower reading skills in German (OECD, 2023; Reiss et al., 2019). Along with these findings, there is a strong link between immigration and language and its influence on educational trajectories (Gogolin & Maaz, 2019; Hahn & Schöps, 2019; Washbrook et al., 2012).

However, these findings are not conclusive and empirical studies show that the educational disadvantages and challenges do not apply equally to all groups of immigrants (Washbrook et al., 2012).

Learning and Attention Disorder as a Risk Factor

Learning disorder (LD), whether unspecified as unexpected poor academic performance that cannot be explained by cognitive potential or by external factors, as well as specific learning disorders, like dyslexia and dyscalculia (Büttner & Hasselhorn, 2011; World Health Organization [WHO] 2021), both cause low academic achievement and various difficulties in school performance (WHO, 2021). Central to the definition of LD is that children or adolescents have deficits in essential learning processes (Kavale & Forness, 2000). A large body of research has demonstrated the disadvantages and challenges faced by students with LD in schools (Bender & Smith, 1990). Consistent with the core symptom of LD, students with LD such as dyslexia and dyscalculia show lower performance in various academic domains (e.g. Landerl et al., 2009; Walker & Nabuzoka, 2007). Moreover, studies indicate that students with LD have low academic self-efficacy beliefs (e.g. Bear et al., 2002; Hampton & Mason, 2003), more negative academic emotions, such as anxiety towards school related tasks (e.g. Sainio et al., 2019), and more negative behaviors, such as being disruptive or bullying (e.g. Walker &

Nabuzoka, 2007). Meta-analyses show the negative impact of various learning difficulties on school drop-out (Gubbels et al., 2019).

Not only students with learning disorders but also students diagnosed with attention deficit/hyperactivity disorder (ADHD) show difficulties in school and in academic performance (Arnold et al., 2020). ADHD is associated with the core symptoms of inattention, hyperactivity and impulsivity, or a combination of these (WHO 2021). Students with ADHD often face barriers in their academic careers, they are at a higher risk of academic underachievement (e.g. Ehm et al., 2014; McConaughy et al., 2011) and face emotional and behavioral problems (Al-Yagon, 2016). Some studies have highlighted the relevance of inattention in particular and the moderation of classroom behavior and homework performance on the academic achievement of students with ADHD (Langberg et al., 2011; Massetti et al., 2008). In general, meta-analyses indicate a moderate to large discrepancy in academic achievement between students with ADHD and students without the diagnosis (Frazier et al., 2007). Longitudinal studies have shown that ADHD predicts academic performance in mathematics, reading and spelling even when controlling for IQ (e.g. Iines et al. 2023; Massetti et al., 2008). As a recent systematic review has shown, ADHD has long-term effects on academic outcomes (Arnold et al., 2020). ADHD often co-occurs with LD, putting the children at even greater risk of academic underachievement (Barnard-Brak et al., 2011).

However, there are inconsistencies in the operationalization of LD and common definitions of unspecific LD as a substantial IQ–achievement discrepancy are repeatedly criticized (Büttner & Hasselhorn, 2011; Kavale & Forness, 2000), and a large number of children with various learning difficulties is neglected. Therefore, some authors point out that in addition to a formal diagnosis, students with symptoms but without diagnosis should also be included and compared when investigating academic underachievement (e.g. Büttner & Hasselhorn, 2011; Loe & Feldman, 2007). As students with ADHD and LD face significantly more homework difficulties than their peers, this could be one first indicator for learning difficulties, which should be observed and targeted to support students who have not (yet) been diagnosed with a clinical learning or attention deficit disorder (Bryan et al., 2001; Langberg et al., 2010).

Findings About SRL and Risk Factors

Students differ greatly in their use of SRL strategies (Barnard-Brak et al., 2010; Wigfield et al., 2011) and some face great difficulty regulating their learning

(Winne, 2005). While previous research has explored the different facets of SRL and highlighted the relevance of SRL in the context of academic achievement (Dent & Koenka, 2016; Robson et al., 2020; Weinstein et al., 2011) and positive developmental outcomes (Robson et al., 2020), the factors that explain the variation in students' use of SRL strategies and their metacognitive knowledge, especially among young learners, have remained relatively unexplored.

Based on the various findings supporting that family background, such as SES, parental education but also immigrant background and the frequency the national language is spoken at home, have a strong influence on students' academic achievement, we can assume that these factors also influence students' SRL. A recent meta-review demonstrates that children's development of self-regulation (in learning) is related to various parental behaviors (Wesarg-Menzel et al., 2023). A systematic review has highlighted the relevance of parental behavior related to SRL in particular (Pino-Pasternak & Whitebread, 2010). Students' use of SRL strategies and also their development of metacognitive knowledge may therefore be related to resources such as the time and money parents invest in education, parental support for SRL and also their own educational experiences.

Socio-Economic Risks and SRL

As well as influencing academic performance, SES also has an influence on the development of children's SRL. Families with a low SES often face diverse challenges, like poverty or higher stress (Blair & Raver, 2015). The liabilities families with low SES experience can be related to fewer learning opportunities, less educational material or scarce interactions that support education and cognitive development (Ermisch, 2008). In addition, parents with a lower SES may not be able to draw sufficiently on their own experience in the educational system and pass on skills and knowledge to their children. They may therefore provide only a limited direct or indirect support for SRL, for example through modeling SRL, which in turn is an important aspect in the early promotion of SRL (Pino-Pasternak & Whitebread, 2010; Wesarg-Menzel et al., 2023).

In line, a meta-analysis has indicated that the environment has a strong influence on children's development of self-regulation (Wesarg-Menzel et al., 2023). So far, there has predominantly been research on the relation between socio-economic status and behavioral regulation, especially in pre-school aged children. For example, it has been shown that children from low-socio-economic families or families with a low household income show less behavioral regulation (e.g. Sektan et al., 2010; Størksen et al., 2015). Poverty has been found to affect the quality of parents' caregiving practices, and therefore may also be related to pre-school aged childrens' reduced self-regulation (e.g. Blair & Raver, 2015;

Li-Grining 2012). In addition, some studies have indicated that preschoolers' behavioral regulation is also related to their parents' educational background, especially maternal education (Miech et al., 2001; Montroy et al., 2016).

While there is comprehensive research investigating the relationship between pre-school aged children's behavior regulation and socio-economic risk factors, there is a lack of research focusing on the possible impact on learners' SRL. The limited research to date suggests that low-SES students use fewer SRL strategies (Vandavelde et al., 2017) and show lower levels of metacognition (Pappas et al., 2003). However, intervention studies highlight that children from low-SES households in particular benefit from training and that this not only improves the use of SRL strategies but also academic performance (Azevedo et al., 2023; Rosário et al., 2016).

Immigration-Related Risks and SRL

As described above, immigrant background and the various specific challenges this poses can impact students' educational trajectories. This also appears to be of interest with regard to the development and differences in students' SRL.

Related to immigration, potential cultural differences in the relevance of behavioral regulation in culture and society, but also students' underlying motivational beliefs influenced by their background, may have an impact on the development of SRL abilities of learners (Blom & Severiens, 2008; Wanless et al., 2011). So far findings on differences in SRL based on students' immigrant background are inconclusive. The few existing studies show on the one hand that immigrant students tend to use fewer SRL strategies (Alivernini et al., 2019a, 2019b; Vandavelde et al., 2017), while on the other hand if taking into account interaction effects between gender and immigrant background, some studies show that immigrant girls report using the most deep-level SRL strategies compared to their classmates (e.g. Blom & Severiens, 2008). Insight from studies focusing on pre-school aged children shows that there might be differences in behavior regulation in different countries (Wanless et al., 2011).

In the context of migration and language, for example dual language learners who speak primarily a minority language at home, is also considered to have an impact on children's self-regulation (Hanno & Surrain, 2019). In a learning process, students use words and private speech to organize and evaluate thoughts and behavior, making language essential for cognitive development (Vygotsky, 1986). Research therefore suggests that the development of SRL, especially metacognition, is bi-directionally related to children's level of expressive language (e.g. Bohlmann et al., 2015). In the case of bilingual students, if the used languages were balanced students use private speech in two languages (Sawyer, 2016).

However, there is no empirical evidence showing the impact of bilingual private speech or language use on SRL so far. Nevertheless, it can be argued that students' proficiency in the language used in school can be seen as a key factor enabling students to take part in classroom learning and interactions.

While the little existing research focuses on the relationship between language and behavioral regulation in young childhood, there is a lack of research investigating the impact of language specifically on SRL strategy use and metacognitive knowledge. However, when considering a relation between language and SRL, context seems to be important. Coping with regulatory challenges at school might therefore also be more closely related to the language spoken at school (Hanno & Surrain, 2019) rather the general language abilities. So far, there has been limited research of the relation with young students' SRL.

Learning and Attention Difficulties as Risk Factors and SRL

Experiencing difficulties in regulating cognitive processes, emotions and behavior is often a challenge for students with learning difficulties, in particular LD and ADHD (Mason & Reid, 2018). These barriers to learning are strongly associated with deficits in SRL. Studies have shown that students with LD have difficulties in self-monitoring (Crane et al., 2017), use less appropriate metacognitive and SRL strategies (Mason & Reid, 2018), and show deficits in working memory and executive function (Johnson et al., 2010). Some studies suggest that these difficulties with SRL in relation to LD already appear early (e.g. Tzohar-Rozen et al., 2021). Children aged five to six years who are at risk of specific LD have been found to use less monitoring and control during task performance and to have difficulties with metacognitive regulation (Tzohar-Rozen et al., 2021) and show poorer metacognitive knowledge (Desoete et al., 2006). Even when students have not yet been diagnosed with LD, but only indicate difficulties in learning, studies have found a negative association with SRL strategy use (Bergey et al., 2017). Difficulties while executing homework and the support needed from parents during learning at home can be one indicator for learning difficulties (Bryan et al., 2001; Langberg et al., 2010). For example, previous research on students' learning independence and need for parental help showed significant correlations with scores on the hyperactivity subscale of the German version of the Strengths and Difficulties Questionnaire (SDQ-Deu) (Blume et al., 2021; Woerner et al., 2002), which may be related to SRL, but is also a strong indicator for attention difficulties.

Besides LD, ADHD is also associated with difficulties in executive functions and impulse control and results in reduced self-regulation (Barkley, 2014; Willcutt et al., 2005). Students with ADHD show problems in applying SRL strategies and

show low persistence in learning (Bakracevic Vukman et al., 2013). Moreover, students with ADHD also show lower self-efficacy beliefs for SRL in comparison to students without (Bakracevic Vukman et al., 2013; Major et al., 2013). Overall, students with LD or ADHD often face challenges with SRL-related activities such as inhibiting behavior, delaying gratification, maintaining persistence and focus during task performance, and organizing and directing actions towards goals (Reid et al., 2012). Although there is evidence that at-risk students benefit from SRL interventions (Berkeley & Larsen, 2018; Burke et al., 2020; Reddy et al., 2018), few studies have examined and documented differences in young students' SRL in the context of LD.

Current Study

Overall, previous research provides initial indications that family-related risk factors such as SES and immigrant background, but also individual risk factors including learning and attention difficulties have not only a direct influence on learning outcomes but also on SRL of young learners and thus have an impact on potential learning trajectories. However, apart from evidence for the association between the various risk factors and behavioral regulation, there is limited research on the relationship between these risks and students' underachievement and SRL in particular. Nevertheless, studies have shown that especially students with these risk factors benefit from SRL training (Azevedo et al., 2023; Rosário et al., 2016). SRL can therefore also be seen as an important ability that could also compensate for potential risk factors for educational underachievement. Identifying which groups of learners need special support due to family or individual risk factors is therefore even more important, to develop and provide targeted interventions. Accordingly, our study aims to offer a first step towards closing this research gap and investigate

RQ1. To what extent do primary school students differ in SRL strategy use and metacognitive knowledge due to family background (parents' profession and occupation and parental education)? Hypothesizing that:

H1: Students from families with a lower SES, combining parent's profession and occupation and parental education as indicators, show lower SRL strategy use and lower metacognitive knowledge.

RQ2. To what extent do primary school students differ in their use of SRL strategies and metacognitive knowledge due to their immigrant background (parents and language spoken at home)? The following hypothesis was formulated:

H2: Students with a higher migration-related risk, combining migration of parents and language is spoken at home as indicators, show lower strategy use and lower metacognitive knowledge.

RQ3. The extent to which primary school students differ in SRL strategy use and metacognitive knowledge due to learning difficulties (learning and attention disorders and need for homework support) was investigated. The following hypothesis was tested:

H3: Students with higher learning difficulties, combining learning and attention disorders and the need for homework support as indicators, show lower SRL strategy use and lower metacognitive knowledge.

Participants and Procedure

A total of 141 students from 2nd to 5th grade (age $M = 9.9$ years ($SD = 1.25$, $Min = 7.3$, $Max = 13.3$ ($N = 133$); 53 % female ($N = 137$)) and their parents participated in the study. Participants were recruited throughout Germany via social media, newspapers and in addition via schools in the federal states of North-Rhine Westphalia and Hesse. Informed parental consent was obtained for parents' and their children's participation in the study. Families received a toy and book voucher for their participation. The data was collected completely online due to school closures and social distancing requirements during the Covid-19 pandemic. Children participated in the study with their personal computers at home, with guidance provided by trained test administrators via telephone. The study was approved by an ethics committee prior to being conducted.

Instruments

In response to the substantial criticism regarding the validity of SRL assessment, particularly in the context of young students, we adopted a multi-method approach to comprehensively assess SRL and its components.

Self-Regulated Learning Measures

SRL self-report The students completed a self-report questionnaire (adapted from Blasiman et al., 2017) to indicate how often they used a set of ten different SRL strategies from the areas of metacognition, motivation, emotion-regulation and resource

strategies. Short examples of two fictional characters who use a certain SRL strategy while learning were presented for each item as an introduction. Following these short scenarios, students were asked to rate how often they used each strategy themselves on a 4-point-likert scale from 1 = never to 4 = always (e.g. 'Pauli and Toni make a plan before doing their homework: they think about which tasks they have to do and when, how long they need to complete them and which task they start with. How often do you plan your homework?'). The internal consistency of the scale was moderate, McDonald's Omega Total (ω) = 0.75.

SRL parent rating Parents were asked to rate their children's use of SRL strategies during homework and learning at home (e.g. 'My child makes a schedule to help organize his or her study time') on a 5-point-likert scale questionnaire (1 = never to 5 = always). For this purpose the Self-Regulation Strategy Inventory: Parent Rating Scale (SRSI-PRS) (Chen et al., 2015) was adapted to primary school students. Items related to learning for math were rephrased to the context of learning at home. To align with the structure of the other instruments utilized in this study, we incorporated three items from a questionnaire for assessing emotion regulation in children and adolescents (Greuel et al., 2018). These emotion regulation items were adjusted by including the context of 'while learning.' Additionally, three items that relate to SRL during the self-reflection phase of learning (Otto, 2007) were included to assess the whole SRL process as described earlier (Zimmerman, 2000). In total, the questionnaire consisted of 29 items. The questionnaire showed good reliability, McDonald's Omega Total (ω) = .9.

Metacognitive knowledge In addition, the NEPS declarative metacognition test for 3rd grade was used to assess students' strategy knowledge (Lockl, 2017). Students were presented a series of ten scenarios wherein a child faces realistic challenges in different school and out-of-school contexts (see Lockl, 2017 for an example). For each scenario, the students had to rate the usefulness of three approaches to deal with the situation, which differed in their quality on a three-point Likert scale. The possible approaches presented included the use of cognitive, metacognitive and resource-orientated strategies. To enhance understanding, each scenario and the three possible strategic approaches were supplemented with illustrations. The scenario test is scored using paired comparisons based on expert judgment of the relative usefulness of the approaches presented. The scenario tests showed good reliability, McDonald's Omega Total (ω) = .87.

Educational Risks

Socio-economic status To measure the parental socio-economic status, parents' profession and occupation were assessed. These variables were first coded according to ISCO 08 and afterwards transformed using the standard International Socio-Economic Index of Occupational Status (ISEI) scale (Ganzeboom et al., 1992). The ISEI score serves as a measure that reflects an individual's position within an occupational hierarchy, while indirectly reflecting their educational attainment and income level (Connelly et al. 2016) . The ISEI scale encompasses a broad range of occupations, spanning from agricultural laborers, such as farmers, to professionals like lawyers and scientists, and provides scores within the range of ten to 90 (Ganzeboom et al., 1992).

In addition to the ISEI, the highest level of parental education was assessed to provide more specific information on educational background. Educational background was then divided into parents with a tertiary education and parents without a tertiary education.

Based on these indicators, a combined score with scores ranging from 0 to 2 was created for further analysis (similar to Favre et al., 2023; Kassiss et al., 2022). Therefore, the ISEI was first divided into quintiles and children in the lowest quintile were coded as having a low family ISEI, while children in the median three quintiles were coded as having a medium family ISEI and children in the highest quintile were coded as having a high family ISEI. Students with parents who had a low ISEI and whose parents have no tertiary education were assigned a 2 in the combined SES score, which indicates a high SES risk factor. Students whose parents have tertiary education and also a high or medium ISEI score were coded to a 0 in the combined SES score, representing a low SES risk factor. The other children were assigned a 1 in the created SES score, indicating a SES medium risk factor.

Migration and language To investigate migration and language-related disparities in young students' SRL, parents were asked whether the mother and the father were born in Germany. Based on this information, the participating children were divided into two groups (with and without immigrant background), regardless of the child's place of birth. Children with at least one parent not born in Germany were assigned to the group with an immigrant background.

Another indicator of migration and language-related risks was the frequency with which the test language (German) was spoken at home, as an indicator for bilingualism. Parents were asked how often German was spoken at home (1 = always

to 4 = never). The participants were then divided into two groups, children who always or almost always speak German at home and children who sometimes or never speak German at home.

Based on a previous discussion about the operationalization of migration status (Kemper et al., 2021) and since prior research has demonstrated that the impact of immigration on students' educational trajectories is complex and that the parents' immigrant status is not necessarily the influential factor, and in order to include overlaps in immigration status and the frequency with which German is spoken at home in the sample in the analyses, a new variable was created as a joint score, ranging from 0 to 2. Students who had an immigrant background as described above and who also never or rarely spoke German at home were coded with a 2 for a high migration and language-related risk factor. Students for whom only one of the two variables applied (either immigrant or never spoke German at home) were assigned a 1 in the new combine migration variable, indicating a medium risk factor. Students who often or always spoke German at home and whose parents were not migrants were assigned a 0, representing a low risk in this area.

Learning and attention disorder/learning difficulties Some parents indicated that their children had been diagnosed with an LD or ADHD. As not all children were formally tested for learning disorders, parents were additionally asked with three items how much support their child needed for homework (1 = never to 5 = always) (e.g. "How often does your child need help while completing their homework?" or "How often does your child need help with checking homework?") (1 = never to 5 = always) as an indicator for difficulties in the context of learning. The questionnaire was adapted based on questions from studies on the COVID-19 pandemic and the involvement of parents in their childrens' learning process during that time (Blume et al., 2021; Schmidt et al., 2021). The questionnaire showed good reliability, with McDonald's Omega Total (ω) = .87.

As with the other two risk areas, a variable was created that combines the two variables and thus reflects overlaps and particular learning difficulties, ranging from 0 to 2. To this end, needed homework support was recoded based on quantiles, and children were coded into low (lowest quantile), medium (middle three quantiles), and high (highest quantile) need for homework support. Students who showed both a diagnosed learning or attention disorder and a high need for support with their homework were coded as 2 for a high learning difficulty risk factor, while children without a diagnosis and with low or moderate need for homework support were assigned a 0, indicating a low learning difficulty risk

factor. The remaining students were coded with 1, as a medium learning difficulty risk factor. Although previous studies have often used several indicators (mostly IQ, performance and diagnoses) to form categories related to learning difficulties and disorders (Fischbach et al., 2013), the approach of including a subjective assessment of difficulties with homework and forming a score based on these indicators is new.

Analyses

All statistical analyses were created in *R*, version 4.2.2. As a result of technical challenges during data collection, the data exhibited missing values that occurred completely at random. To address missing values in the SRL variables, multiple imputations were performed using multiple imputation by chained equations (MICE) algorithm with the *mice* package in *R* (van Buuren & Groothuis-Oudshoorn, 2011), generating 30 datasets with 10 iterations each to ensure stable estimates. Missing values in the variables for the educational risk factors were not imputed. Imputing dichotomous variables, like Germany as country of birth, could have led to distortions affecting data authenticity. Since these predictor variables were potentially not subject to random missingness, but rather a result of intentional non-response, we decided imputation would not be appropriate. The cases with missing values in these variables were not included in the analyses of the corresponding risk factor, which leads to different numbers of cases in the following analyses. Cases which included missings for all items of a variable were not included in imputation process. To ensure the robustness of the results, we conducted sensitivity analyses comparing the results from the analysis and descriptive statistics of the scales from the imputed datasets with those from the complete cases, inspecting convergence diagnostics, and evaluating potential variations across imputations. In addition, the data was tested for outliers. Realistic outliers in the children's or parents' data were not excluded. Realistic outliers were defined as values exceeding two standard deviations from the mean, without indicating systematic response patterns (e.g., consistently selecting the same response) or implausible data entries (e.g., an age of 100), representing natural variance, particularly in SRL measures."

Given prior studies, weak relations resp. small effects were assumed for the SES (Miech et al., 2001; Montroy et al., 2016; Sektan et al., 2010), immigration and parental education (Alivernini et al., 2019a, 2019b; Blom & Severiens, 2008), learning difficulties (Bergey et al., 2017) on self-regulation (in learning) and metacognitive knowledge. The most appropriate type of analysis would have been

conclusive multiple regression models or moderation analysis including analysis of interaction effects.

However, post-hoc power analyses showed that the sample was too small for multiple regression models that include all different indicators for risk for underachievement as predictors. As a result, the combined risk scores were defined as described above and an approach was chosen in which several analyses were carried out for the three risk areas. This approach has implications and limitations which will be addressed in the discussion, but we intend to provide initial insights despite the small sample size. To take into account the potential of false discoveries among the results of the several tested models Benjamini-Hochberg procedure (Benjamini et al., 2009), as false discovery rate (FDR) approach was used to adjust the p-values in all analyses.

Results

Descriptive statistics for the different variables divided across the groups based on the single indicators for risks are presented in Table 2.1 and also grouped based on the created combined risk scores in Table 2.2. The parents came from a total of 22 different nations (including Germany), of which 13 were non-European.

Relationship Between Students' SRL and Risk Factors

Table 1 in Appendix reports the correlations among all assessed variables. Positive significant correlations were found between all three of the SRL-related measures: SRL strategy use rated by the parents, self-reported SRL strategy use and metacognitive knowledge ($r = .19 - .35; p < .05$). In addition, there were significant correlations between the different assessments for each risk factor. For example, we found a significant correlation between parental education and SES ($r = .45; p < .01$), indicating that parents without a tertiary degree tend to have a lower SES. The correlation between the different indicators from one risk area support the approach to create combined scores for further analysis to investigate the combined influences. However, no significant correlations were found between the different risk factors (e.g. immigration and SES).

Moreover, correlation analyses indicated significant relations between some of the risk factors and SRL strategy use and metacognitive knowledge, which were further analysed by means of regression.

Table 2.1 Descriptive statistics for SRL measures and demographics grouped by risk factors

	Parental education		Immigrant background		German language at home			Diagnosed learning or attention disorder	
	Tertiary education <i>N</i> = 90	No tertiary education <i>N</i> = 44	No <i>N</i> = 107	Yes <i>N</i> = 28	Always / Nearly always <i>N</i> = 122	Sometimes/ Never <i>N</i> = 15	No <i>N</i> = 82	Yes <i>N</i> = 8	
<i>N</i> (%)									
Gender									
Female	43 (47.8 %)	26 (59.1 %)	51 (47.7 %)	18 (64.3 %)	59 (48.4 %)	11 (73.3 %)	46 (56.1 %)	3 (37.5 %)	
Male	47 (52.2 %)	18 (40.9 %)	56 (52.3 %)	10 (35.7 %)	63 (51.6 %)	4 (26.7 %)	36 (43.9 %)	5 (62.5 %)	
Age	9.82 (1.27)	10.0 (1.13)	9.97 (1.29)	9.74 (1.11)	9.94 (1.27)	9.76 (1.02)	10.1 (1.75)	10.1 (1.75)	
<i>M</i> (<i>SD</i>)									
SRL Parent rating	2.64 (0.55)	2.72 (0.57)	2.61 (0.54)	2.92 (0.53)	2.63 (0.55)	2.90 (0.53)	2.72 (0.55)	2.38 (0.47)	
SRL Self report	2.54 (0.43)	2.41 (0.41)	2.45 (0.41)	2.72 (0.45)	2.46 (0.42)	2.77 (0.42)	2.47 (0.40)	2.10 (0.16)	
Metacognitive Knowledge	0.68 (0.20)	0.65 (0.22)	0.69 (0.20)	0.58 (0.21)	0.68 (0.21)	0.56 (0.17)	0.70 (0.19)	0.65 (0.15)	

(Fortsetzung)

Table 2.1 (Fortsetzung)

	Parental education		Immigrant background		German language at home		Diagnosed learning or attention disorder	
	Tertiary education <i>N</i> = 90	No tertiary education <i>N</i> = 44	No <i>N</i> = 107	Yes <i>N</i> = 28	Always / Nearly always <i>N</i> = 122	Sometimes/ Never <i>N</i> = 15	No <i>N</i> = 82	Yes <i>N</i> = 8
Needed Homework support	2.87 (1.03)	2.76 (0.96)	2.79 (1.00)	3.00 (0.99)	2.81 (1.01)	3.16 (0.91)	2.49 (0.92)	3.79 (0.78)
SES (ISEI 08)	60.6 (13.2)	47.8 (8.76)	55.6 (12.8)	60.7 (15.8)	56.7 (13.3)	55.2 (17.0)	54.4 (13.7)	62.6 (9.68)

Note. Table display *N* and percentage (%) in round brackets, retrospectively Means = *M*, and standard deviation = *SD* in round brackets

Table 2.2 Descriptive statistics for SRL measures and demographics grouped by risk indices

	Combined SES score			Combined migration-related risk score			Combined learning difficulties score		
	0	1	2	0	1	2	0	1	2
	<i>N</i> = 70	<i>N</i> = 33	<i>N</i> = 21	<i>N</i> = 106	<i>N</i> = 14	<i>N</i> = 14	<i>N</i> = 78	<i>N</i> = 13	<i>N</i> = 3
Age	9.69 (1.31)	9.95 (1.19)	10.4 (1.03)	9.97 (1.29)	9.68 (1.20)	9.81 (1.04)	10.1 (1.16)	10.2 (1.31)	9.11 (2.46)
SRL strategy use	2.66 (0.56)	2.62 (0.50)	2.59 (0.58)	2.61 (0.54)	2.90 (0.54)	2.94 (0.53)	2.75 (0.55)	2.53 (0.46)	1.93 (0.16)
Parent rating									
SRL strategy use	2.53 (0.44)	2.46 (0.40)	2.34 (0.39)	2.45 (0.41)	2.59 (0.51)	2.83 (0.36)	2.47 (0.40)	2.52 (0.45)	2.30 (.)
Self-report									
Metacognitive Knowledge	0.68 (0.20)	0.64 (0.19)	0.68 (0.22)	0.69 (0.20)	0.62 (0.25)	0.55 (0.17)	0.70 (0.19)	0.69 (0.17)	0.50 (.)

Note. Table display *N* and percentage (%) in round brackets, retrospectively Means = *M*, and standard deviation = *SD* in round brackets

Differences in Young Students' Use of SRL Strategies

We aimed to investigate how young students' SRL strategy use is related to three categories of risk factors for educational underachievement RQ1: socio-economic risks (SES and parental education); RQ2: migration-related risks (immigrant background and language); RQ3: developmental risks (learning and attention disorder and need for homework support). Despite the broad body of research showing that socio-economic variables, migration, frequency the national language is spoken at home and learning and attention difficulties have an impact on learning outcomes and academic achievement, the potential influence on young students' SRL, as a crucial basis for lifelong learning, has so far only been investigated to a limited extent. Consistent with prior research that has supported the age-related nature of SRL (Wesarg-Menzel et al., 2023) and previous findings indicating gender differences in SRL (Heirweg et al., 2019; Vandeveldel et al., 2013), age and gender were added as covariates to all regression models. In order to capture potential non-linear relationships and group-specific differences that might not be evident in a linear approach, the predictors were modeled as factors (low, medium or high risk factor) in the following regression analysis.

Differences in Students' SRL Related to Socioeconomic Status

To investigate the potential relationship of the SES and students' SRL strategy use and metacognitive knowledge, linear regression analyses were conducted for each SRL outcome variable. In these analyses, the combined SES score based on the ISEI and parental education, as described above, was used as a predictor. Age and gender were included as covariates.

The results revealed a small significant negative relationship between SES and SRL strategy use measured with the self-report. Compared to the reference group of students with a low SES risk factor, students in the high SES risk factor group showed significantly less use of SRL strategies ($B = -0.26$, 95 % CI $[-0.49, -0.03]$, $t(96) = -2.21$, $adj. B-H p = .05$), while no significant effect was found in the group of students with a medium SES risk factor compared to the reference group. However, SES did not show an effect on the parents' rating of students' SRL strategy use or students metacognitive knowledge ($p > .05$) when age and gender were included as covariates. Table 2 in the Appendix provides an overview of the results.

Differences in Students' SRL Related to Migration and Language

Based on parents' immigrant background and the frequency German is spoken at home as reported by the parents as indicators for migration-related risks for students underachievement, a combined score variable was created as described above. Regression analyses were used to investigate the potential impact of this migration score on SRL strategy use and metacognitive knowledge. For SRL strategy use measured with students' self-report, linear regression analysis with age and gender as covariates indicated a statistically significant positive association with the combined migration-related risk score as factor variable. Contrary to our hypothesis, children with a higher migration-related risk score scored significantly higher than the reference group of children with a lower migration risk score ($B = 0.41$, 95 % CI [0.14, 0.68], $t(105) = 2.98$, *adj. B-H* $p = .01$). There was no significant effect found for the medium migration related risk factor. However, the regression analysis with students' metacognitive knowledge showed an opposite tendency but did not reveal a significant effect of the migration related risk ($B = -0.13$, 95 % CI [-0.26, -3.25e-03], $t(105) = -2.03$, *adj. B-H* $p = .10$) (see Appendix Table 3 for detailed results).

Differences in Students' SRL Related to Learning Difficulties

To explore whether learning difficulties are associated with young learners' SRL, we investigated the relationship between the combined learning difficulties score as factor variable and SRL strategy use and metacognitive knowledge assessed with the different measurements. There was no significant association found for learning difficulties with students' self-reported SRL strategy use ($p > .05$). Nevertheless, the regression analysis revealed a statistically significant negative association with SRL strategy use measured with the parent rating. Compared to the reference group of children with a low learning difficulty risk factor, parents of children with a high learning difficulty risk factor rated their children's SRL strategy use significantly lower ($B = -0.74$, 95 % CI [-1.37, -0.12], $t(80) = -2.36$, *adj. B-H* $p = .05$). There was no effect related to the medium learning difficulty risk factor. However, we found no statistical significant association between students' metacognitive knowledge and learning difficulties ($p > .05$) (See Appendix Table 4 for detailed results).

Discussion

There are several educational risk factors that can present barriers to students' learning and to their academic achievement. Whereas researchers investigating these risk factors have mainly tested in how far students' learning outcomes (such as their academic achievement) are related to being at risk, we need to know more about the relationship between such risk factors and students' learning *behavior* (such as their SRL). We aimed to shed light on the relationship between young students' SRL and various risk factors for educational underachievement. Our multi-method approach enabled us to consider various facets of SRL.

The first research question addressed differences in students' SRL based on SES, with parents' occupation and parental education as indicators. The analyses showed a significant relationship between students' SRL measured with the self-report and their families' SES. Students with a lower combined SES score rated their SRL strategy use lower in the self-report, while there was no relationship between SES and students' SRL strategy use rated by parents or metacognitive knowledge. These findings therefore only partially confirm our hypothesis and agree to a limited degree with the previous research on the relationship of SES, maternal education and SRL (e.g. Vandavelde et al., 2017) or pre-schoolers behavior regulation (e.g. Sektan et al., 2010). The absence of a significant relationship between SES and two of the three SRL measurements suggests that other factors, not captured in this study, may play a more prominent role in influencing young learners' SRL. This raises the question of how relevant common measures such as SES are to students' SRL, and whether more specific measures might be more appropriate. For example, besides typical measures like income and SES, families' background seems to be linked to their educational expenses as well as to parenting styles (Anger & Plünnecke, 2021). These in turn could also have an influence on the development of learners' SRL. Previous research investigating parental behavior related to the self-regulation of young children, and in particular to behavior-regulation and SRL in primary school children (e.g. Pino-Pasternak & Whitebread, 2010), provides first insights into the potential to explore relations specifically in the context of SRL as well.

Related to the second research question regarding differences based on migration-related risks, contrary to our hypothesis, students with a higher migration-related risk, with immigrant background and low frequency German is spoken at home, showed higher scores in the self-report SRL strategy use. These results are in line with previous research demonstrating both lower and higher SRL in students with an immigrant background compared to their peers (e.g. Blom & Severiens, 2008; Neubauer et al., 2014; Vandavelde et al., 2017).

The differences found in SRL strategy use and metacognitive knowledge in this study and previous research related to immigration, may be traced back to the value of behavioral regulation in societies, culture specific parenting practices, children's general role in society or immigrant students' underlying motivational beliefs (Blom & Severiens, 2008; Wanless et al., 2011), which have to be investigated further.

Another potential reason for the inconsistent findings in this study could relate to SRL measurement. Overall, SRL research in recent years has increasingly criticized the validity of both self-reports and adult rating (McCoy, 2019; Veenman, 2011), and the question arises as to what extent these methods are dependent on culture and language and if these instruments possibly measure different cultural-dependent concepts. In addition, previous research suggests that learners tend to overestimate their SRL in self-reports (e.g. Boekaerts & Corno, 2005; Cromley and Azevedo 2007). Even if several studies highlight the academic disadvantages of students with immigrant background, some studies suggest that these students are often optimistic but incoherent when they assess their own performance, self-efficacy and self-expectations (Hornstra et al., 2013). However, the contradictory findings can also be due to ambitious educational goals and a special optimism of immigrant families (Heath & Brinbaum, 2007), that may have a positive effect on SRL of learners with an immigrant background. Moreover, the contradictory findings indicate that the mere operationalization of immigrant background based on the parents' origin is no longer an adequate measure. Transmigration is becoming increasingly common and the origins are diverse, involving different cultural influences that could also have an impact on the relevance and the support of SRL within families. Studies investigating differences in academic achievement have indicated that the challenges that education poses to learners differ between immigrant groups and that not all groups are equally affected (Brinbaum & Heath, 2014).

Related to the third research question, differences in students' SRL were investigated based on developmental risks, with learning and attention disorders and needed homework support as indicators. The analyses showed that students with higher learning difficulties had significant lower SRL in the parent rating. These findings are in line with previous research showing a negative association of learning and attention disorders with SRL strategy use (e.g. Mason & Reid, 2018), but only confirm our hypothesis with one of the SRL measures. However, this finding also supports previous findings from studies assessing students SRL deficits related to other indicators for learning difficulties besides clinical diagnoses (e.g. Bergey et al., 2017) and take into account the critique on current LD definitions (Büttner & Hasselhorn, 2011).

The regression analyses using factor variables for the risk factors revealed that particularly children with high risk scores in the three areas (SES, migration and language, learning difficulties), who are affected by both indicators showed significant associations with SRL. In contrast, children with medium risk scores, affected by only one of the indicators, did not show significant effects, highlighting the need for further investigation of these groups in future studies.

Limitations and Future Directions

The findings of this study should be interpreted within the context of several limitations. Firstly, the sample size was relatively small, which may limit the generalizability of the discussed relationships between risk factors for academic underachievement and young students SRL. Based on the sample size and the assumption that the effects tend to be relatively small, a multitude of models were analyzed which could have led to alpha inflation.

Secondly, this study recruited families online, leading to a convenience sample and the possibility of selection bias. Additionally, the investigated groups varied in size, with only a few children exhibiting the specific risk factors, especially being diagnosed with LD or ADHD. The small sample size of participants, particularly those with a diagnosis of LD/ADHD, may have limited the power to detect notable effects of learning disorders on SRL. These factors should be considered when interpreting the implications of this research, highlighting the need for larger, more diverse samples in future investigations. In addition, the various series of risks for educational underachievement were assessed using established but very simplified measures. Therefore, future studies with a larger sample size should examine, for example, more precise socio-economic and cultural background variables, values and educational styles in addition to the simplified measure used in this study.

Additionally, the cross-sectional design of this study limits the ability to draw causal inferences, especially the development of SRL related to risk factors. Moreover, the study was conducted during the covid-19 pandemic, which may have influenced students' SRL and thus limits the generalizability of the findings. Further research on inter-individual differences in students' SRL is needed, given that the various benefits of this important skill have been established throughout years of research.

Since SES and parental education had limited explanatory value, research would be of interest that is more closely orientated to children's environments

and examines the connections between learners' SRL and, for example, parenting style, given autonomy and agency, hobbies or parents' time resources as underlying mechanisms influencing academic achievement and also SRL. This study did not assess these indicators or the socio-cultural status as assessed, for example, in international school performance studies (Lewalter et al., 2023). These background variables would be of particular interest for future research.

In order to continue researching the above-mentioned differences in relation to students' immigrant background, special attention should be paid to the assessment of SRL. In addition to a precise examination of cultural dependency in SRL measurements, such as questionnaires, methods that assess students' actual SRL strategy use during task processing in a cultural and language independent way should be developed, validated and included in future research investigating inter-individual differences in SRL. Moreover, to visualize the actual underlying mechanism related to immigration background, future research should investigate the relevance of SRL in culture, students' motivational beliefs related to diverse cultures or also parents' perceived value of SRL in that context. Additionally, exploring the interplay between language, SRL and metacognitive knowledge – including the language used in students' private speech could provide valuable insights into inter-individual differences.

In particular, SRL strategy use and metacognitive knowledge has been insufficiently investigated in children with learning disorders and ADHD. Our study gives first insights of SRL deficits of students with learning and attention disorders and students who need more support during homework. However, there is a lack of meta-analyses addressing the specific deficits of learners with learning and attention disorders in for example SRL strategy use, self-efficacy and utility beliefs regarding SRL and metacognitive knowledge. Especially in the context of students with learning difficulties, empirical research on SRL with task-related assessment methods such as observations, think aloud or trace data would be useful to avoid bias. Moreover, the multifaceted relationship between required homework support, learning difficulties, and students' SRL should be further investigated, along with additional indicators of learning difficulties and their impact on SRL.

Furthermore, based on previous research on the relationship between risk factors, SRL strategy use and metacognitive knowledge, it could be of interest whether there are certain SRL aspects in where children at risk show deficits. Research on the different strategy areas (metacognition, cognition, motivation- and emotion-regulation) could be informative. A limitation of this study is that we did not assess the adaptive use or quality of SRL strategy use. While these aspects are crucial for a comprehensive understanding of SRL, research in this

area remains limited, highlighting the need for further studies to explore these aspects and potential differences in students in greater depth.

Based on these results, specific trainings could be developed for students with learning difficulties. In particular, it is necessary to recognize young learners in order to equalize possible differences in educational opportunities based on risk factors for underachievement and to empower students' at risk to become self-regulated learners as early as possible.

Conclusion

The findings underline the complexity of the diverse influences on students' SRL and call for further investigation to uncover additional contributing factors that may explain differences in young learners' SRL strategy use. With the multi-method approach to assess the complex construct of SRL from different perspectives, this study shows that various risk factors for educational underachievement are related to the SRL. It was found that students at risk show SRL deficits in some areas whereas some risk factors for academic underachievement are not related to young students SRL or, in the case of immigrant background, are even related to a greater SRL strategy use. In the context of the widely researched links between SRL and academic achievement, SRL can also be seen as a protective factor that could reduce existing disadvantages through early promotion of SRL among students at risk. Therefore, the study once again reinforces the need to promote SRL in the **classroom**.

Acknowledgments Funding: This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation [Grant 429624323]).

Appendix

Table 1 Means, standard deviations and correlations with confidence intervals

Variable	M	SD	1	2	3	4	5	6	7	8	9
1. SRL parent rating	2.67	0.55									
2. SRL self-report	2.50	0.43	.35**								
			[.18, .50]								
3. Metacognitive knowledge	0.67	0.20	.23*	.19*							
			[.05, .40]	[.00, .35]							
4. ISEI 08	56.56	13.34	.01	.05	.10						
			[-.17, .19]	[-.14, .24]	[-.09, .29]						
5. Parents Education (0 = tertiary education)			.07	-.15	-.05	-.45**					
			[-.10, .24]	[-.33, .03]	[-.24, .13]	[-.58, -.29]					
6. Immigration Background (0 = no immigrant background)			.23**	.25**	-.21*	.14	-.10				
			[.06, .38]	[.07, .41]	[-.38, -.03]	[-.04, .31]	[-.27, .07]				
7. German Language at Home (0 = always/nearly always)			.15	.23*	-.19*	-.03	-.01	.66**			
			[-.01, .31]	[.05, .39]	[-.36, -.01]	[-.20, .15]	[-.18, .16]	[.56, .75]			

(Fortsetzung)

Table 1 (Fortsetzung)

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9
8. Diagnosed Learning Difficulty (0 = no diagnose)			-0.18 [-.37, .03]	-0.21 [-.42, .02]	-0.06 [-.28, .18]	.17 [-.05, .37]	-0.04 [-.25, .16]	-0.12 [-.32, .09]	-0.10 [-.30, .11]		
9. Needed Homework support	2.85	1.00	-.49**	-.13	-.24**	.13	-.05	.09	.11	.38**	
10. Gender (0 = female)			[-.61, -.36]	[-.30, .06]	[-.41, -.06]	[-.05, .30]	[-.22, .12]	[-.08, .25]	[-.06, .27]	[-.19, .54]	
			-.21*	-.02	-.11	.19*	-.11	-.13	-.16	.11	.21*
			[-.36, -.04]	[-.20, .17]	[-.28, .08]	[.01, .35]	[-.27, .06]	[-.30, .03]	[-.32, .01]	[-.10, .31]	[.05, .37]
11. Age	9.92	1.25	.13	.23*	.30**	-.16	.09	-.07	-.04	.01	-.19*
			[-.04, .29]	[.04, .40]	[.12, .46]	[-.33, .02]	[-.08, .26]	[-.24, .10]	[-.21, .13]	[-.20, .22]	[-.05, .28]

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95 % confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$

Table 4 Separate linear regression analysis with the combined learning difficulties score as factor variable as Predictor and different SRL measures as outcome

Outcome variable	SRL parent rating				SRL self-rating				Metacognitive Knowledge			
	B	Std. Error	t-value	adj. B-H P	B	Std. Error	t-value	adj. B-H P	B	Std. Error	t-value	adj. B-H P
Intercept	2.09	0.477	4.37	.00	1.75	0.440	3.97	.00	0.21	0.195	1.05	.61
Learning Difficulties (medium risk factor)	-.027	0.173	-1.55	.17	0.003	0.164	0.02	.98	0.05	0.072	0.70	.61
Learning Difficulties (high risk factor)	-.074	0.316	-2.36	.05*	0.04	0.434	0.10	.98	-0.07	0.192	-0.39	.70
Age	0.07	0.047	1.51	.17	0.07	0.042	1.62	.27	0.05	0.019	2.65	.05*
Gender (0 = female)	-0.0	0.116	-0.78	.43	0.01	0.100	0.14	.98	-0.03	0.044	-0.77	.61
				Adj. R ²				Adj. R ²				Adj. R ²
				.09				-0.01				.07

Note. SRL = self-regulation of learning. * p < .05. ** p < .01. *** p < .001

References

- Alivernini, F., Cavicchiolo, E., Girelli, L., Lucidi, F., Biasi, V., Leone, L., Cozzolino, M., & Manganelli, S. (2019a). Relationships between sociocultural factors (gender, immigrant and socioeconomic background), peer relatedness and positive affect in adolescents. *Journal of Adolescence*, *76*, 99–108. <https://doi.org/10.1016/j.adolescence.2019.08.011>
- Alivernini, F., Manganelli, S., Cavicchiolo, E., Chirico, A., & Lucidi, F. (2019b). The use of self-regulated cognitive strategies across students with different immigrant backgrounds and gender. *Journal of Psychoeducational Assessment*, *37*(5), 652–664. <https://doi.org/10.1177/0734282918785072>
- Al-Yagon, M. (2016). Perceived close relationships with parents, teachers, and peers: Predictors of social, emotional, and behavioral features in adolescents with LD or comorbid LD and ADHD. *Journal of Learning Disabilities*, *49*(6), 597–615. <https://doi.org/10.1177/0022219415620569>
- Anger, C., & Plünnecke, A. (2021). *Bildungsgerechtigkeit: Herausforderung für das deutsche Bildungssystem. IW-Analysen* (vol. 140). IW Medien.
- Arnold, L. E., Hodgkins, P., Kahle, J., Madhoo, M., & Kewley, G. (2020). Long-term outcomes of ADHD: Academic achievement and performance. *Journal of Attention Disorders*, *24*(1), 73–85. <https://doi.org/10.1177/1087054714566076>
- Azevedo, R. [Roger], Rosário, P [P.], Magalhães, P., Núñez, J. C [J. C.], Pereira, B., & Pereira, A. (2023). A tool-kit to help students from low socioeconomic status background: A school-based self-regulated learning intervention. *European Journal of Psychology of Education*, *38*(2), 495–518. <https://doi.org/10.1007/s10212-022-00607-y>
- Bai, B., & Guo, W. (2021). Motivation and self-regulated strategy use: Relationships to primary school students' English writing in Hong Kong. *Language Teaching Research*, *25*(3), 378–399. <https://doi.org/10.1177/1362168819859921>
- Bakracevic Vukman, K., Funcic Masic, T., & Schmidt, M. (2013). Self-regulation of learning in secondary school students with special educational needs and other students of vocational and technical schools. *The New Educational Review*, *33*(3), 295–305. <https://czasopisma.marszalek.com.pl/en/10-15804/tner/11548-tner2013325>
- Barkley, R. A. (Ed.). (2014). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (4th edn). The Guilford Press.
- Barnard-Brak, L., Paton, V. O., & Lan, W. Y. (2010). Profiles in self-regulated learning in the online learning environment. *The International Review of Research in Open and Distributed Learning*, *11*(1), 61. <https://doi.org/10.19173/irrodl.v11i1.769>
- Barnard-Brak, L., Sulak, T. N., & Fearon, D. D. (2011). Coexisting disorders and academic achievement among children with ADHD. *Journal of Attention Disorders*, *15*(6), 506–515. <https://doi.org/10.1177/1087054710369667>
- Barone, C. (2006). Cultural capital, ambition and the explanation of inequalities in learning outcomes: A comparative analysis. *Sociology*, *40*(6), 1039–1058. <https://doi.org/10.1177/0038038506069843>
- Bear, G. G., Minke, K. M., & Manning, M. A. (2002). Self-concept of students with learning disabilities: A meta-analysis. *School Psychology Review*, *31*(3), 405–427. <https://doi.org/10.1080/02796015.2002.12086165>

- Becker, B., & Klein, O. (2021). The primary effect of ethnic origin – rooted in early childhood? An analysis of the educational disadvantages of Turkish-origin children during the transition to secondary education in Germany. *Research in Social Stratification and Mobility*, 75, Article 100639. <https://doi.org/10.1016/j.rssm.2021.100639>
- Bender, W. N., & Smith, J. K. (1990). Classroom behavior of children and adolescents with learning disabilities: A meta-analysis. *Journal of Learning Disabilities*, 23(5), 298–305.
- Benjamini, Y., Heller, R., & Yekutieli, D. (2009). Selective inference in complex research. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 367(1906), 4255–4271. <https://doi.org/10.1098/rsta.2009.0127>
- 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, Article 701255. <https://doi.org/10.3389/fpsyg.2021.701255>
- Bergey, B. W., Deacon, S. H., & Parrila, R. K. (2017). Metacognitive reading and study strategies and academic achievement of university students with and without a history of reading difficulties. *Journal of Learning Disabilities*, 50(1), 81–94. <https://doi.org/10.1177/0022219415597020>
- Berkeley, S., & Larsen, A. (2018). Fostering self-regulation of students with learning disabilities: Insights from 30 years of reading comprehension intervention research. *Learning Disabilities Research & Practice*, 33(2), 75–86. <https://doi.org/10.1111/ldrp.12165>
- Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research*, 87(2), 425–469. <https://doi.org/10.3102/0034654316669821>
- Bethhäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, 7(3), 375–385. <https://doi.org/10.1038/s41562-022-01506-4>
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444. <https://doi.org/10.1146/annurev-psych-113011-143823>
- Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology*, 66, 711–731. <https://doi.org/10.1146/annurev-psych-010814-015221>
- Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behaviour. *Memory (Hove, England)*, 25(6), 784–792. <https://doi.org/10.1080/09658211.2016.1221974>
- Blom, S., & Severiens, S. (2008). Engagement in self-regulated deep learning of successful immigrant and non-immigrant students in inner city schools in inner city schools. *European Journal of Psychology of Education*, 13(1), 41–58.
- Blume, F., Schmidt, A., Kramer, A. C., Schmiedek, F., & Neubauer, A. B. (2021). Home-schooling during the SARS-CoV-2 pandemic: The role of students' trait self-regulation and task attributes of daily learning tasks for students' daily self-regulation. *Zeitschrift Für Erziehungswissenschaft : ZfE*, 24(2), 367–391. <https://doi.org/10.1007/s11618-021-01011-w>
- Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *European Psychologist*, 1(2), 100–112. <https://doi.org/10.1027/1016-9040.1.2.100>

- Boekaerts, M. (2007). Understanding students' affective processes in the classroom. In R. Pekrun & P. A. Schutz (Eds.), *Emotion in education* (pp. 37–56). Elsevier. <https://doi.org/10.1016/B978-012372545-5/50004-6>
- Boekaerts, M., & Corno, L. (2005). Self-Regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review*, 54(2), 199–231.
- Bohlmann, N. L., Maier, M. F., & Palacios, N. (2015). Bidirectionality in self-regulation and expressive vocabulary: Comparisons Between monolingual and dual language learners in preschool. *Child Development*, 86(4), 1094–1111.
- Brinbaum, L.L.-P.Y., & Heath, A. (2014). Academic and vocational tracking in upper secondary education. *British Academy*. <https://doi.org/10.5871/bacad/9780197265741.001.0001>
- Bryan, T., Burstein, K., & Bryan, J. (2001). Students with learning disabilities: homework problems and promising practices. *Educational Psychologist*, 36(3), 167–180. https://doi.org/10.1207/S15326985EP3603_3
- Bryce, D., & Whitebread, D. (2012). The development of metacognitive skills: Evidence from observational analysis of young children's behavior during problem-solving. *Metacognition and Learning*, 7(3), 197–217. <https://doi.org/10.1007/s11409-012-9091-2>
- Burke, K. M., Raley, S. K., Shogren, K. A., Hagiwara, M., Mumbardó-Adam, C., Uyanik, H., & Behrens, S. (2020). A meta-analysis of interventions to promote self-determination for students with disabilities. *Remedial and Special Education*, 41(3), 176–188. <https://doi.org/10.1177/0741932518802274>
- Büttner, G., & Hasselhorn, M. (2011). Learning disabilities: Debates on definitions, causes, subtypes, and responses. *International Journal of Disability, Development and Education*, 58(1), 75–87. <https://doi.org/10.1080/1034912X.2011.548476>
- Callan, G. L., DaVia Rubenstein, L., Barton, T., & Halterman, A. (2021). Enhancing motivation by developing cyclical self-regulated learning skills. *Theory into Practice*, 1–13. <https://doi.org/10.1080/00405841.2021.1932153>
- Chen, P. P., Cleary, T. J., & Lui, A. M. (2015). Examining parents' ratings of middle-school students' academic self-regulation using principal axis factoring analysis. *School Psychology Quarterly : The Official Journal of the Division of School Psychology, American Psychological Association*, 30(3), 385–397. <https://doi.org/10.1037/spq0000098>
- Connelly, R., Gayle, V., & Lambert, P. S. (2016). A review of occupation-based social classifications for social survey research. *Methodological innovations*, 9, 2059799116638003.
- Crane, N., Zusho, A., Ding, Y., & Cancelli, A. (2017). Domain-specific metacognitive calibration in children with learning disabilities. *Contemporary Educational Psychology*, 50, 72–79. <https://doi.org/10.1016/j.cedpsych.2016.09.006>
- Cromley, J. G., & Azevedo, R. [Roger]. (2007). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, 1(3), 229–247. <https://doi.org/10.1007/s11409-006-9002-5>
- 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>
- Desoete, A., Roeyers, H., & Huylebroeck, A. (2006). Metacognitive skills in Belgian third grade children (age 8 to 9) with and without mathematical learning disabilities. *Metacognition and Learning*, 1(2), 119–135. <https://doi.org/10.1007/s11409-006-8152-9>

- 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>
- Dignath, C., & Hasselhorn, M. (2023). Selbstregulationsförderung: Gelingensbedingungen und Herausforderungen bei Lern- und Verhaltensschwierigkeiten. In M. Börner-Ringleb, G. Casale, M. Balt, & M. Herzog (Eds.), *Lern- und Verhaltensschwierigkeiten in der Schule: Erscheinungsformen – Entwicklungsmodelle – Implikationen für die Praxis* (1. Aufl., pp. 177–185). Verlag W. Kohlhammer.
- Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, 20(4), 391–409. <https://doi.org/10.1007/s10648-008-9083-6>
- 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>
- Ehm, J.-H., Merkt, J., Gawrilow, C., & Hasselhorn, M. (2014). Selbstkonzept und Schulleistungen von Grundschulern mit ADHS-Symptomen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46. <https://doi.org/10.25656/01:12908> (Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie 46 (2014) 2, pp. 79–88).
- Ermisch, J. (2008). Origins of Social immobility and inequality: Parenting and early child development. *National Institute Economic Review*, 205, 62–71.
- Favre, C. A [Céline A.], Kassiss, W., Müller, C., & Aksoy, D. (2023). Sozio-emotionale Resilienzentwicklung im Jugendalter trotz Miterleben verbaler elterlicher Paargewalt – ein transitionsanalytischer Zugang über zwei Wellen. In S. Gingelmaier, B. Herz, D. Hövel, L. Dietrich, J. Langer, P.-C. Link, & X. Müller (Eds.), *ESE Emotionale und Soziale Entwicklung in der Pädagogik der Erziehungshilfe und bei Verhaltensstörungen. Heft 5 – Reichweite einer Pädagogik sozio-emotionaler Entwicklungsförderung* (pp. 30–49). Verlag Julius Klinkhardt. <https://doi.org/10.35468/6021-03>
- Fischbach, A., Schuchardt, K., Brandenburg, J., Kleczewski, J., Balke-Melcher, C., Schmidt, C., Büttner, G., Grube, D., Mähler, C., & Hasselhorn, M. (2013). Prävalenz von Lernschwächen und Lernstörungen: Zur Bedeutung der Diagnosekriterien. *Lernen Und Lernstörungen*, 2(2), 65–76. <https://doi.org/10.1024/2235-0977/a000035>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive – developmental inquiry. *American Psychologist*, 34(10), 906–911.
- Frazier, T. W., Youngstrom, E. A., Glutting, J. J., & Watkins, M. W. (2007). ADHD and achievement: meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *Journal of Learning Disabilities*, 40(1), 49–65.
- Ganzeboom, H. B., de Graaf, P. M., & 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](https://doi.org/10.1016/0049-089X(92)90017-B)
- Gascoine, L., Higgins, S., & Wall, K. (2017). The assessment of metacognition in children aged 4–16 years: A systematic review. *Review of Education*, 5(1), 3–57. <https://doi.org/10.1002/rev3.3077>
- Convention on the Rights of Persons with Disabilities (2007).
- Gogolin, I., & Maaz, K. (2019). Editorial „Migration und Bildungserfolg“. *Zeitschrift Für Erziehungswissenschaft*, 22(S1), 1–14. <https://doi.org/10.1007/s11618-019-00886-0>

- Greuel, J. F., Briegel, W., & Heinrichs, N. (2018). Die Eltern-Kurzversion des Fragebogens zur Erhebung der Emotionsregulation bei Kindern und Jugendlichen (FEEL-KJ). *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, 47(1), 48–58. <https://doi.org/10.1026/1616-3443/a000463>
- Grigorenko, E. L., Compton, D. L., Fuchs, L. S., Wagner, R. K., Willcutt, E. G., & Fletcher, J. M. (2020). Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. *The American Psychologist*, 75(1), 37–51. <https://doi.org/10.1037/amp0000452>
- Gubbels, J., van der Put, C. E., & Assink, M. (2019). Risk factors for school absenteeism and dropout: A meta-analytic review. *Journal of Youth and Adolescence*, 48(9), 1637–1667. <https://doi.org/10.1007/s10964-019-01072-5>
- Hahn, I., & Schöps, K. (2019). Bildungsunterschiede von Anfang an? *Frühe Bildung*, 8(1), 3–12. <https://doi.org/10.1026/2191-9186/a000405>
- Hampton, N. Z., & Mason, E. (2003). Learning disabilities, gender, sources of efficacy, self-efficacy beliefs, and academic achievement in high school students. *Journal of School Psychology*, 41(2), 101–112. [https://doi.org/10.1016/S0022-4405\(03\)00028-1](https://doi.org/10.1016/S0022-4405(03)00028-1)
- Händel, M., Artelt, C., & Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument. *Journal for Educational Research Online*, 5(2), 162–188.
- Hanno, E., & Surrain, S. (2019). The direct and indirect relations between self-regulation and language development among monolinguals and dual language learners. *Clinical Child and Family Psychology Review*, 22(1), 75–89. <https://doi.org/10.1007/s10567-019-00283-3>
- Hasselhorn, M. (2022). Lernstörungen: Ein unvermeidbares Schicksal? *Zeitschrift Für Pädagogische Psychologie*, 36(1–2), 1–17. <https://doi.org/10.1024/1010-0652/a000324>
- Hatala, M., Nazeri, S., & Salehian Kia, F. (2023). Progression of students' SRL processes in subsequent programming problem-solving tasks and its association with tasks outcomes. *The Internet and Higher Education*, 56, Article 100881. <https://doi.org/10.1016/j.iheduc.2022.100881>
- Heath, A., & Brinbaum, Y. (2007). Guest editorial: Explaining ethnic inequalities in educational attainment. *Ethnicities*, 7(3), 291–304. <https://doi.org/10.1177/1468796807080230>
- Heath, A., Ronthon, C., & Kilpi, E. (2008). The second generation in western europe: education, unemployment, and occupational attainment. *Annual Review of Sociology*, 34(1), 211–235. <https://doi.org/10.1146/annurev.soc.34.040507.134728>
- Heirweg, S., de Smul, M., Devos, G., & van Keer, H. (2019). Profiling upper primary school students' self-regulated learning through self-report questionnaires and think-aloud protocol analysis. *Learning and Individual Differences*, 70, 155–168. <https://doi.org/10.1016/j.lindif.2019.02.001>
- Hornstra, L., van der Veen, I., Peetsma, T., & Volman, M. (2013). Developments in motivation and achievement during primary school: A longitudinal study on group-specific differences. *Learning and Individual Differences*, 23, 195–204. <https://doi.org/10.1016/j.lindif.2012.09.004>
- Howard, S. J., Vasseleu, E., Neilsen-Hewett, C., de Rosnay, M., Chan, A. Y. C., Johnstone, S., Mavilidi, M., Paas, F., & Melhuish, E. C. (2021). Executive function and self-regulation:

- bi-directional longitudinal associations and prediction of early academic skills. *Frontiers in Psychology*, 12, Article 733328. <https://doi.org/10.3389/fpsyg.2021.733328>
- Iines, R., & P., Sami J., M., Vesa M., N., & Hannu K., S. (2023). ADHD symptoms and maladaptive achievement strategies: The reciprocal prediction of academic performance beyond the transition to middle school. *Emotional and Behavioural Difficulties*, 28(1), 3–17. <https://doi.org/10.1080/13632752.2023.2189404>
- Johnson, E. S., Humphrey, M., Mellard, D. F., Woods, K., & Swanson, H. L. (2010). Cognitive processing deficits and students with specific learning disabilities: A selective meta-analysis of the literature. *Learning Disabilities Quarterly*, 33(1), 3–18.
- Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual Review of Sociology*, 29(1), 417–442. <https://doi.org/10.1146/annurev.soc.29.010202.100019>
- Kassis, W., Aksoy, D., Favre, C. A [Céline Anne], Janousch, C., & Artz, S. T.-G. (2022). Thriving despite parental physical abuse in adolescence: A two-wave latent transition analysis on hedonic and eudaimonic violence-resilience outcome indicators. *Children (Basel, Switzerland)*, 9(4). <https://doi.org/10.3390/children9040553>
- Kavale, K. A., & Forness, S. R. (2000). What definitions of learning disability say and don't say. *Journal of Learning Disabilities*, 33(3), 239–256.
- Kemper, T., Supik, L., & Colbasevici, L. (2021). Klassifikationen von Migration und Sprache in Datensätzen und Publikationen der Bildungsforschung und der amtlichen (Schul-) Statistik. *Gesellschaft unter Spannung. Verhandlungen des 40. Kongresses der Deutschen Gesellschaft für Soziologie 2020*. Advance online publication. <https://doi.org/10.25656/01:26121>
- Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology*, 103(3), 309–324. <https://doi.org/10.1016/j.jecp.2009.03.006>
- Langberg, J. M., Arnold, L. E., Flowers, A. M., Altaye, M., Epstein, J. N. [Jeff N.], & Molina, B. S. G. (2010). Assessing homework problems in children with ADHD: Validation of a parent-report measure and evaluation of homework performance patterns. *School Mental Health*, 2(1), 3–12. <https://doi.org/10.1007/s12310-009-9021-x>
- Langberg, J. M., Molina, B. S. G., Arnold, L. E., Epstein, J. N [Jeffery N.], Altaye, M., Hinshaw, S. P., Swanson, J. M., Wigal, T., & Hechtman, L. (2011). Patterns and predictors of adolescent academic achievement and performance in a sample of children with attention-deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, 40(4), 519–531. <https://doi.org/10.1080/15374416.2011.581620>
- Lergetporer, P., Werner, K., & Woessmann, L. (2021). Does ignorance of economic returns and costs explain the educational aspiration gap? Representative evidence from adults and adolescents. *Economica*, 88(351), 624–670. <https://doi.org/10.1111/ecca.12371>
- Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O., & Reiss, K. (2023). *PISA 2022*. Waxmann Verlag GmbH. <https://doi.org/10.31244/9783830998488>
- Li-Grining, C. P. (2012). The role of cultural factors in the development of latino preschoolers' self-regulation. *Child Development Perspectives*, 6(3), 210–217. <https://doi.org/10.1111/j.1750-8606.2012.00255.x>

- Lim, S. L., & Yeo, K. J. (2021). A systematic review of the relationship between motivational constructs and self-regulated learning. *International Journal of Evaluation and Research in Education (IJERE)*, 10(1), 330. <https://doi.org/10.11591/ijere.v10i1.21006>
- Lockl, K. (2017). Assessment of declarative metacognition: Starting Cohort 2 – Grade 3. *Leibniz Institute for Educational Trajectories, National Educational Panel Study*.
- Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology*, 32(6), 643–654. <https://doi.org/10.1093/jpepsy/jsl054>
- Major, A., Martinussen, R., & Wiener, J. (2013). Self-efficacy for self-regulated learning in adolescents with and without attention deficit hyperactivity disorder (ADHD). *Learning and Individual Differences*, 27, 149–156. <https://doi.org/10.1016/j.lindif.2013.06.009>
- Mang, J., Müller, K., Lewalter, D., Kastorff, T., Müller, M., Ziernwald, L., Tupac-Yupanqui, A., Heine, J.-H., & Kölle, O. (2023). Herkunftsbezogene Ungleichheiten im Kompetenzerwerb. In D. Lewalter, J. Diedrich, F. Goldhammer, O. Köller, & K. Reiss (Hrsg.), *PISA 2022* (S. 163–197). Waxmann Verlag GmbH.
- Marulis, L. M., Palincsar, A. S., Berhenke, A. L., & Whitebread, D. (2016). Assessing metacognitive knowledge in 3–5 year olds: The development of a metacognitive knowledge interview (McKI). *Metacognition and Learning*, 11(3), 339–368. <https://doi.org/10.1007/s11409-016-9157-7>
- Mason, L. H., & Reid, R. (2018). Self-regulation: Implications for individuals with special needs. In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (2nd edn). Routledge; Taylor & Francis.
- Massetti, G. M., Lahey, B. B., Pelham, W. E., Loney, J., Ehrhardt, A., Lee, S. S., & Kipp, H. (2008). Academic achievement over 8 years among children who met modified criteria for attention-deficit/hyperactivity disorder at 4–6 years of age. *Journal of Abnormal Child Psychology*, 36(3), 399–410. <https://doi.org/10.1007/s10802-007-9186-4>
- McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, 21(4), 471–490. <https://doi.org/10.1016/j.ecresq.2006.09.003>
- McClelland, M. M., & Cameron, C. E. (2012). Self-Regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, 6(2), 136–142. <https://doi.org/10.1111/j.1750-8606.2011.00191.x>
- McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly*, 15(3), 307–329.
- McConaughy, S. H., Volpe, R. J., Antshel, K. M., Gordon, M., & Eiraldi, R. B. (2011). Academic and social impairments of elementary school children with attention deficit hyperactivity disorder. *School Psychology Review*, 40(2), 200–225.
- McCoy, D. C. (2019). Measuring young children’s executive function and self-regulation in classrooms and other real-world settings. *Clinical Child and Family Psychology Review*, 22(1), 63–74. <https://doi.org/10.1007/s10567-019-00285-1>
- Meyers, R., Pignault, A., & Houssemand, C. (2013). The role of motivation and self-regulation in dropping out of school. *Procedia – Social and Behavioral Sciences*, 89, 270–275. <https://doi.org/10.1016/j.sbspro.2013.08.845>

- Miech, R., Essex, M. J., & Goldsmith, H. H. (2001). Socioeconomic status and the adjustment to school: The role of self-regulation during early childhood. *Sociology of Education, 74*(2), 102–120.
- Moilanen, K. L. (2007). The adolescent self-regulatory inventory: The development and validation of a questionnaire of short-term and long-term self-regulation. *Journal of Youth and Adolescence, 36*(6), 835–848. <https://doi.org/10.1007/s10964-006-9107-9>
- Montroy, J. J., Bowles, R. P., Skibbec, L. E., McClelland, M. M., & Morrison, F. J. (2016). The development of self-regulation across early childhood. *Developmental Psychology, 52*(11), 1744–1762.
- Nelson, T., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25) [publisher not identified].
- Neubauer, A., Gawrilow, C., & Hasselhorn, M. (2014). Selbstkontrolle bei Vorschulkindern unterschiedlicher kultureller Herkunft. *Frühe Bildung : Interdisziplinäre Zeitschrift für Forschung, Ausbildung und Praxis, 3*. <https://doi.org/10.25656/01:12342> (Frühe Bildung : Interdisziplinäre Zeitschrift für Forschung, Ausbildung und Praxis 3 (2014) 3, pp. 146–154).
- O'Connor, M., Chong, S., Hutchinson, D., Sanson, A., McIntosh, J., Olsson, C. A., & Goldfeld, S. (2019). Socioeconomic disadvantage in infancy and academic and self-regulation outcomes. *Pediatrics, 143*(5). <https://doi.org/10.1542/peds.2018-2640>
- OECD. (2023). *PISA 2022 Ergebnisse: Lernstände und Bildungsgerechtigkeit*. wbv Media GmbH & Co. KG. <https://doi.org/10.3278/6004956w>
- Otto, B. (2007). SELVES: Schüler-, Eltern- und Lehrertraining zur Vermittlung effektiver Selbstregulation.
- Pappas, S., Ginsburg, H. P., & Jiang, M. (2003). SES differences in young children's metacognition in the context of mathematical problem solving. *Cognitive Development, 18*(3), 431–450. [https://doi.org/10.1016/S0885-2014\(03\)00043-1](https://doi.org/10.1016/S0885-2014(03)00043-1)
- Paquette, L., Grant, T., Zhang, Y., Biswas, G., & Baker, R. (2021). Using epistemic networks to analyze self-regulated learning in an open-ended problem-solving environment. In A. R. Ruis & S. B. Lee (Hrsg.), *Advances in quantitative ethnography – second international conference, ICQE 2020, proceedings* (S. 185–201). Springer International Publishing.
- Pino-Pasternak, D., & Whitebread, D. (2010). The role of parenting in children's self-regulated learning. *Educational Research Review, 5*(3), 220–242. <https://doi.org/10.1016/j.edurev.2010.07.001>
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*(6), 459–470. [https://doi.org/10.1016/S0883-0355\(99\)00015-4](https://doi.org/10.1016/S0883-0355(99)00015-4)
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50043-3>
- Reddy, L. A., Cleary, T. J., Alperin, A., & Verdesco, A. (2018). A critical review of self-regulated learning interventions for children with attention-deficit hyperactivity disorder. *Psychology in the Schools, 55*(6), 609–628. <https://doi.org/10.1002/pits.22142>
- Reid, R. R., Harris, K. R., Graham, S., & Rock, M. (2012). Self-Regulation among students with LD and ADHD. In *Learning about learning disabilities* (pp. 141–173). Elsevier. <https://doi.org/10.1016/B978-0-12-388409-1.00005-9>

- Reiss, K., Weis, M., Klieme, E., & Köller, O. (2019). *PISA 2018. Grundbildung im internationalen Vergleich*. Waxmann. <https://doi.org/10.25656/01:18315>
- Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*, *146*(4), 324–354. <https://doi.org/10.1037/bul0000227>
- Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45*, 31–51. <https://doi.org/10.1016/j.dr.2017.04.001>
- Rosário, P. [Pedro], Núñez, J. C [José Carlos], Vallejo, G., Cunha, J., Azevedo, R [Raquel], Pereira, R., Nunes, A. R., Fuentes, S., & Moreira, T. (2016). Promoting Gypsy children school engagement: A story-tool project to enhance self-regulated learning. *Contemporary Educational Psychology*, *47*, 84–94. <https://doi.org/10.1016/j.cedpsych.2015.11.005>
- Sainio, P. J., Eklund, K. M., Ahonen, T. P. S., & Kiuru, N. H. (2019). The role of learning difficulties in adolescents' academic emotions and academic achievement. *Journal of Learning Disabilities*, *52*(4), 287–298. <https://doi.org/10.1177/0022219419841567>
- Sawyer, A., Miller-Lewis, L. R., Searle, A. K., Sawyer, M. G., & Lynch, J. W. (2015). Is greater improvement in early self-regulation associated with fewer behavioral problems later in childhood? *Developmental Psychology*, *51*(12), 1740–1755. <https://doi.org/10.1037/a0039829>
- Sawyer, J. (2016). In what language do you speak to yourself? A review of private speech and bilingualism. *Early Childhood Research Quarterly*, *36*, 489–505. <https://doi.org/10.1016/j.ecresq.2016.01.010>
- Schlack, R., Hölling, H., Kurth, B.-M., & Huss, M. (2007). Die Prävalenz der Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) bei Kindern und Jugendlichen in Deutschland. Erste Ergebnisse aus dem Kinder- und Jugendgesundheitsurvey (KiGGS) [The prevalence of attention-deficit/hyperactivity disorder (ADHD) among children and adolescents in Germany. Initial results from the German Health Interview and Examination Survey for Children and Adolescents (KiGGS)]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, *50*(5–6), 827–835. <https://doi.org/10.1007/s00103-007-0246-2>
- Schmidt, A., Kramer, A. C., Brose, A., Schmiedek, F., & Neubauer, A. B. (2021). Distance learning, parent-child interactions, and affective well-being of parents and children during the COVID-19 pandemic: A daily diary study. *Developmental Psychology*, *57*(10), 1719–1734. <https://doi.org/10.1037/dev0001232>
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain and Education*, 114–121.
- Schunk, D. H., & Ertmer, P. A. (2000). Self-Regulation and Academic Learning: Self-Efficacy Enhancing Interventions. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Hrsg.), *Handbook of self-regulation* (S. 631–649). Academic.
- Schwippert, K. [., Kasper, D. [., Köller, O. [., McElvany, N. [., Selter, C. [., Steffensky, M. [., & Wendt, H. [. (2020). *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann. <https://doi.org/10.25656/01:21353>

- Sektnan, M., McClelland, M. M., Acock, A., & Morrison, F. J. (2010). Relations between early family risk, children's behavioral regulation, and academic achievement. *Early Childhood Research Quarterly*, 25(4), 464–479. <https://doi.org/10.1016/j.ecresq.2010.02.005>
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453.
- Sprong, S., & Skopek, J. (2022). Academic achievement gaps by migration background at school starting age in Ireland. *European Societies*, 24(5), 580–604. <https://doi.org/10.1080/14616696.2022.2084558>
- Størksen, I., Ellingsen, I. T., Wanless, S. B., & McClelland, M. M. (2015). The influence of parental socioeconomic background and gender on self-regulation among 5-year-old children in Norway. *Early Education and Development*, 26(5–6), 663–684. <https://doi.org/10.1080/10409289.2014.932238>
- Thomas, R., Sanders, S., Doust, J., Beller, E., & Glasziou, P. (2015). Prevalence of attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Pediatrics*, 135(4), e994–1001. <https://doi.org/10.1542/peds.2014-3482>
- Tzohar-Rozen, M., & Kramarski, B. (2017). Metacognition and meta-affect in young students: Does it make a difference in mathematical problem solving? *Teachers College Record: The Voice of Scholarship in Education*, 119(13), 1–26. <https://doi.org/10.1177/016146811711901308>
- Tzohar-Rozen, M., Waisel, T., & Shamir, A. (2021). Self-regulated learning skills and the quality of execution of construction tasks: A comparison between young children at risk for specific learning disorders and typically developing peers. *Metacognition and Learning*, 16(3), 833–853. <https://doi.org/10.1007/s11409-021-09266-8>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3). <https://doi.org/10.18637/jss.v045.i03>
- Vandevelde, S., van Keer, H., & Merchie, E. (2017). The challenge of promoting self-regulated learning among primary school children with a low socioeconomic and immigrant background. *The Journal of Educational Research*, 110(2), 113–139. <https://doi.org/10.1080/00220671.2014.999363>
- Vandevelde, S., van Keer, H., & Rosseel, Y. (2013). Measuring the complexity of upper primary school children's self-regulated learning: A multi-component approach. *Contemporary Educational Psychology*, 38(4), 407–425. <https://doi.org/10.1016/j.cedpsych.2013.09.002>
- Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning*, 6(2), 205–211. <https://doi.org/10.1007/s11409-011-9080-x>
- Veenman, M. V. J., & van Cleef, D. (2019). Measuring metacognitive skills for mathematics: Students' self-reports versus on-line assessment methods. *ZDM Mathematics Education*, 51(4), 691–701. <https://doi.org/10.1007/s11858-018-1006-5>
- Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1(1), 3–14. <https://doi.org/10.1007/s11409-006-6893-0>
- Vygotsky, L. S. (1986). *Thought and language – revised*. The MIT Press.

- Walker, A., & Nabuzoka, D. (2007). Academic Achievement and Social Functioning of Children With and Without Learning Difficulties. *Educational Psychology, 27*(5), 635–654. <https://doi.org/10.1080/01443410701309175>
- Wanless, S. B., McClelland, M. M., Acock, A. C., Ponitz, C. C., Son, S.-H., Lan, X., Morrison, F. J., Chen, J.-L., Chen, F.-M., Lee, K., Sung, M., & Li, S. (2011). Measuring behavioral regulation in four societies. *Psychological Assessment, 23*(2), 364–378. <https://doi.org/10.1037/a0021768>
- Washbrook, E., Waldfogel, J., Bradbury, B., Corak, M., & Ghanghro, A. A. (2012). The development of young children of immigrants in Australia, Canada, the United Kingdom, and the United States. *Child Development, 83*(5), 1591–1607. <https://doi.org/10.1111/j.1467-8624.2012.01796.x>
- Weinstein, C. E., Acee, T. W., & Jung, J. (2011). Self-regulation and learning strategies. *New Directions for Teaching and Learning, 2011*(126), 45–53. <https://doi.org/10.1002/tl.443>
- Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-Regulation Interventions with a Focus on Learning Strategies. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Hrsg.), *Handbook of self-regulation* (S. 727–747). Academic.
- Wesarg-Menzel, C., Ebbes, R., Hensums, M., Wagemaker, E., Zaharieva, M. S., Staaks, J. P., van den Akker, A. L., Visser, I., Hoeve, M., Brummelman, E., Dekkers, T. J., Schuitema, J. A., Larsen, H., Colonesi, C., Jansen, B. R., Overbeek, G., Huizenga, H. M., & Wiers, R. W. (2023). Development and socialization of self-regulation from infancy to adolescence: A meta-review differentiating between self-regulatory abilities, goals, and motivation. *Developmental Review, 69*, Article 101090. <https://doi.org/10.1016/j.dr.2023.101090>
- Wigfield, A., Klauda, S. L., & Cambria, J. (2011). Influences on the Development of Academic Self-Regulatory Processes. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of Self-Regulation of Learning and Performance*. Routledge. <https://doi.org/10.4324/9780203839010.ch3>
- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 57*(11), 1336–1346. <https://doi.org/10.1016/j.biopsych.2005.02.006>
- Winding, T. N., & Andersen, J. H. (2015). Socioeconomic differences in school dropout among young adults: The role of social relations. *BMC Public Health, 15*, 1054. <https://doi.org/10.1186/s12889-015-2391-0>
- Winne, P. H. (2005). A Perspective on State-of-the-art Research on Self-regulated Learning. *Instructional Science, 33*(5–6), 559–565. <https://doi.org/10.1007/s11251-005-1280-9>
- Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267–276. <https://doi.org/10.1080/00461520.2010.517150>
- Woerner, W., Becker, A., Friedrich, C., Klasen, H., Goodman, R., & Rothenberger, A. (2002). Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung [Normal values and evaluation of the German parents' version of Strengths and Difficulties Questionnaire (SDQ): Results of a representative field study]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 30*(2), 105–112. <https://doi.org/10.1024/1422-4917.30.2.105>

- World Health Organization. (2021). *International statistical classification of diseases and related health problems (11th ed.)*. World Health Organization.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Hrsg.), *Handbook of self-regulation* (S. 13–39). Academic.
- Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. Routledge. <https://doi.org/10.4324/9780203839010>

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Addressing the Heterogeneity of German Language Competence in Children with Immigrant Background: Patterns of Vocabulary Development from Age Three to Nine

3

Christian Lohmann and Birgit Becker

Abstract

Children with an immigrant background are often regarded as a risk group for poor educational outcomes due to their on average lower levels of German language skills. However, the heterogeneity within this group is often neglected when only averages are considered. We address this heterogeneity by analyzing patterns of German vocabulary development in early childhood among children with and without immigrant background with an explorative approach. Using the data from the newborn cohort of the National Educational Panel Study (NEPS), we examine children's German vocabulary trajectories from age three to nine by applying Growth Mixture Modeling. Our results

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann https://doi.org/10.1007/978-3-658-48674-7_3.

C. Lohmann (✉)

Leibniz-Institut für Bildungsverläufe, Bamberg, Germany

E-Mail: Christian.Lohmann@lifbi.de

B. Becker

Goethe-Universität Frankfurt am Main, Frankfurt am Main, Germany

E-Mail: Bi.Becker@em.uni-frankfurt.de

IDeA – Center for Research on Individual Development and Adaptive Education of Children at Risk, Frankfurt am Main, Germany

© Der/die Autor(en) 2025

J. Ziehm-Eicher et al. (Hrsg.), *Kinder mit erhöhtem Risiko für Bildungsmisserfolg*,
Edition ZfE 20, https://doi.org/10.1007/978-3-658-48674-7_3

73

show that a large share of children with an immigrant background exhibit an average or even above-average vocabulary trajectory. However, almost a fifth can be considered as having a “risk profile” since they show very below-average vocabulary levels at all measurement points. Such a risk profile also exists in the sample of non-immigrant children, but the share is smaller. The number of books in the family, the family language and children’s early cognitive skills are the strongest predictors for being in this risk profile.

Keywords

Immigrant background • Vocabulary development • Language skills • Early childhood

Introduction

Many studies have found that children with immigrant background show lower levels of educational achievement in the German education system than their non-immigrant peers (e.g., Hunkler & Schotte, 2023; for an overview see Diehl et al., 2016; OECD, 2023). Lacking German language skills are often regarded as a key explanatory factor for this finding (Esser, 2006; Kempert et al., 2016), as the German education system with its *monolingual habitus* requires all children to have a good command of German (Gogolin, 2008). In fact, children with immigrant background, on average, start school with a lower level of German language skills (Linberg & Wenz, 2017; Relikowski et al., 2015), which seems to put them at risk for unfavorable educational trajectories right from the start. This is the case because children’s majority language skills in early childhood are predictive of their later school performance (Dale et al., 2023; Durham et al., 2007), which also applies to children with an immigrant background (Becker & Klein, 2021). Thus, children of immigrants are frequently perceived as a risk group for poorer educational outcomes, which is often formulated as a blanket statement for this entire group.

However, children with an immigrant background are very heterogeneous regarding their German language skills (e.g., many of them grow up bilingually, others monolingually with German or mainly with another language), which is often neglected when only averages are considered. Not all children with an immigrant background have low German language skills in early childhood. Many of them probably show a German language development that is very similar

to the average of non-immigrant children, and others may catch up until school starts. Thus, children with an immigrant background probably exhibit a great variety of German language trajectories in early childhood, which is not well understood yet. Thus, when it comes to identifying and understanding possible risks for the further educational career, it seems important to get a more differentiated picture on children's early language development and to take a closer look at the heterogeneity *within* the group of children with an immigrant background.

In this article, we address this heterogeneity and analyze patterns of German language development in (early) childhood among children with an immigrant background with a special focus on patterns indicating significant risks for unfavorable educational trajectories. More concretely, we investigate the following research questions:

1. Which patterns of German vocabulary development from age three to nine years can be identified in children with an immigrant background? Is there a *risk profile*?
2. To what extent do profiles differ from the profiles of non-immigrant children? Are there similar *risk profiles* in both groups of children?
3. How can these profiles be described in terms of child and family characteristics? Which children are found in the *risk profile*?

This article adds to the literature by using an exploratory approach in order to identify patterns of language trajectories. Previous literature about the early language development of children with an immigrant background has mostly focused either on average scores of children with and without immigrant origin or on differences between predefined subgroups (e.g., by family language or country of origin). In contrast, we reverse the order of analysis and examine which patterns of language trajectories can be identified among children with and without an immigrant background as a starting point. Although we have some hypotheses about emerging patterns, our exploratory approach also allows the detection of other (unexpected) profiles. The profiles will be compared with a normative sample. Finally, we analyze which child and family characteristics are associated with profile group membership. This also allows the identification of characteristics that are associated with a *risk profile*. For the empirical analyses, we use data from the cohort of newborns of the National Educational Panel Study (NEPS).

State of Research

A comprehensive literature review of the language development of children with an immigrant background is beyond the scope of this article (see Esser, 2006; Hoff, 2014; Paradis, 2019). We will briefly report some key findings of the literature regarding 1.) the majority language skills of children with an immigrant background, and 2.) influencing factors on their vocabulary development in the majority language more specifically. The complex topic of bilingualism exceeds the scope of this study and cannot be addressed in more detail (see deSouza et al., 2023; Hoff, 2014, 2018 for a comprehensive review).

Majority Language Skills of Children with an Immigrant Background

Compared to their non-immigrant peers and due to factors like the shorter time of residence, most first-generation immigrants lag behind in terms of majority language proficiency. In the case of Germany, this does not only hold for the classic migrant worker generations who arrived from 1954 to 1970s (Dustmann, 1994), but also for recently arrived immigrants (Kristen et al., 2016). These comparably lower German language proficiencies often carryover into the second immigrant generation (for an overview see Kristen, 2019) putting them at a disadvantageous educational trajectory (Kempert et al., 2016). The most recent IQB-Bildungstrend from 2022, which is a nationwide study in Germany with assessments of different competencies of ninth graders, shows clear gaps between children with and without an immigrant background in the domains of listening and reading in German (Henschel et al., 2023). Similar cross-sectional results for children at school age were found by other researchers (for an overview see Olczyk et al., 2016a, b). Various international studies suggest that these gaps are emerging before the school enrolment at the age of six years (Højten et al., 2019; Sprong & Skopek, 2023; Washbrook et al., 2012).

In the German context, similar early differences are reported in several studies ranging from the age of five (Linberg & Wenz, 2017; Relikowski et al., 2015) back to the age of three (Becker et al., 2013). However, longitudinal studies differ in their assessment of how such early gaps develop over time, i.e. whether they tend to decrease (Becker et al., 2013), remain stable (Sprong & Skopek, 2023), or increase (Kigel et al., 2015), which might be due to different immigrant samples in the respective studies.

Vocabulary Development in Children with an Immigrant Background: Influencing Factors and Potential Heterogeneity

In our data, we find nearly one third of children with an immigrant background in the third generation. Their parents were born in Germany and usually went through the German schooling system, giving them exposure to the German language and letting them grow up bilingually. Furthermore, some second-generation immigrants married individuals from their country of origin, giving the spouse the opportunity to migrate to Germany. Their children will then have one parent from the first and one parent from the second immigrant generation with different skill levels in the German language. This leads to variations in the home language use as well as parental proficiency in German, resulting in a variety of possible different language outcomes for the child, depending on the constellation (Becker, 2011).

In many European countries including Germany, families with immigrant background tend to be socioeconomically disadvantaged (Schnepf, 2007). Besides the direct effect of SES on the language proficiency of the child (Hart & Risley, 1995; Bradley & Corwyn, 2002), there are also indirect effects on the entry age at early childhood education and care (ECEC) centers as well as the quality of institutions (Fuchs-Rechlin & Bergmann, 2014), leading to later entry for children from families with a low SES into facilities with often comparably lower quality (Alexandersen et al., 2021). Additionally, the usage of ECEC varies especially until the age of three between families with and without a migration history (Fuchs-Rechlin & Bergmann, 2014).

Many children with an immigrant background grow up bilingually, which is (on average) associated with a better working memory compared to their monolingual peers (Grundy & Timmer, 2017). In turn, the phonological working memory is an important factor for language learning (Mitchell et al., 2019, Sect. “SRL and Academic Achievement”). However, while children in monolingual environments are fully exposed to one language, their peers in bilingual environments are less exposed to each language, therefore their possibilities to broaden their vocabularies is limited compared to monolingual children (Hoff, 2018). Depending on the parents’ majority language competencies, the quality of parental input may also be lower for children with immigrant background when compared to their non-immigrant peers (Hart & Risley, 1995; Hoff, 2018). If the family language is not the majority language, the amount of exposure to the majority language may vary with the child’s age of entering an ECEC center. However, even though bilingual children may score lower than monolinguals in vocabulary test measuring the

majority language, their combined vocabulary of minority and majority language excels the vocabulary of monolinguals in addition to advantages in executive function and cognitive flexibility (Grundy & Timmer, 2017).

In summary, the development of majority language vocabulary of children with an immigrant background depends on more factors than the vocabulary development of their native peers, suggesting more heterogeneous vocabulary trajectories. However, most studies have only compared children with vs. without an immigrant background or monolingual vs. bilingual children in their vocabulary, ignoring potential heterogeneity *within* these (sub)groups. Thus, the aim of this study is to examine this heterogeneity and take a closer look at the different language trajectories of children with an immigrant background.

Theoretical Framework and Hypotheses

For our theoretical framework, we draw on Chiswick and Miller's (1998, 2001) model of language acquisition in the context of migration. This model has an economic research tradition but has been generalized and applied to explain differences in first and second language development in general (Esser, 2006; Seuring & Will, 2022; van Tubergen & Mentjox, 2014). The model assumes that three main factors affect the process of language development: exposure, efficiency, and incentives/motivation (Chiswick & Miller, 1998, p. 886).

Exposure: In order to learn a language, the learner needs to be in contact with that language. Different terms are used by different scholars to describe this basic requirement: The learner needs language *input* (Klein & Dimroth, 2003; Paradis, 2019), *access* (Ahrenholz, 2010; Esser, 2006) or *exposure* to the language (Bialystok, 1978; Chiswick & Miller, 1998), respectively.

In empirical studies, exposure is often measured by the frequency of language input in a specific context, the home context being the most crucial one for the language development of children (Hart & Risley, 1995; Hoff, 2018). Especially for children's vocabulary acquisition, the home learning environment is particularly important (Linberg et al., 2020; Rodriguez & Tamis-LeMonda, 2011). Besides the home, language input in other contexts such as ECEC institutions is also relevant—especially for children with less majority language input at home (Klein & Becker, 2017; Seuring & Will, 2022). For the majority language acquisition of children with an immigrant background, migration-specific exposure variables are often considered such as the length of exposure to this language or the share of majority language use (Hoff, 2018; Paradis, 2019; Seuring & Will,

2022), as well as social contacts to members of the host society and the ethnic composition of the neighborhood (Esser, 2006; Rydland et al., 2013).

Efficiency: Chiswick and Miller define efficiency quite technically as “the extent of improvement in destination-language skills per unit of exposure” (Chiswick & Miller, 2001, p. 393). This refers to individuals’ ability for language learning (Esser, 2006; Klein & Dimroth, 2003; Paradis, 2019; Spolsky, 1989). Individual differences in language learning exist and can be—besides other factors—traced back to differences in language aptitude, working memory and learning strategies (Mitchell et al., 2019, Sect. “SRL and Academic Achievement”).

In empirical studies, general cognitive abilities are often used as operationalizations of efficiency (Seuring & Will, 2022; van Tubergen & Mentjox, 2014). There is extensive literature on the (reciprocal) relationship between children’ cognitive and language skills (for an overview see Deák, 2014). The working memory in particular is regarded a central determinant for language development (see Mitchell et al., 2019, Sect. “SRL and Academic Achievement”).

Incentives/motivation: In their economic framework, Chiswick and Miller (2001) regarded immigrants’ *economic incentives* for language learning in terms of expected labor market returns. In other disciplines, this factor is termed as *motivation* of the learner and is conceptualized more broadly beyond an economic utilization perspective (Esser, 2006; Spolsky, 1989). Klein and Dimroth (2003) mention social integration and communicative needs as main reasons and incentives for (second) language learning. Ahrenholz (2010, p. 25) argues that children usually have an “uninhibited desire for integration into the social community” (own translation), so that the motivation of young children to learn the language surrounding them should be generally high. We follow this reasoning and will not consider the factor of motivation in the remainder of this paper.

From this language acquisition model, we derive our hypotheses about the German vocabulary trajectories of children with an immigrant background in early childhood.

1. High-achieving profile: We expect that those children who have a high amount of exposure to the German language *and* who have a high efficiency will consistently show an above-average German vocabulary.
2. Average profile: We expect an average trajectory of German vocabulary development for children with average levels of exposure and efficiency (or with a high level in one factor, but an average level in the other).
3. Risk profile: We expect a consistently below-average vocabulary for those children who have a low level of exposure *and/or* efficiency.

4. Catching up profile: We expect that children with a high level of efficiency who initially start with a low but rising level of exposure (e.g., due to visiting ECEC) will start at a low level of German vocabulary but show an especially steep increase (*catching up*).

The assessments as (*above/below*) *average* in vocabulary is made with reference to a normative sample. These hypotheses describe the patterns that we regard as most likely based on the theory. However, given our exploratory approach other patterns may emerge empirically.

Data and Methods

Data

For our study, we use the Newborn Cohort of the German National Educational Panel Study (NEPS; Blossfeld & Roßbach, 2019; NEPS Network, 2023). In this nationally representative cohort, 3481 infants at the age of six to eight months and their respective caretakers (mostly mothers) were surveyed in the first wave in 2012. Afterwards, the families were surveyed every year and assessments for various competency domains of the child as well as interviews with the parents were conducted (for an overview of the NEPS Newborn Cohort see Attig et al., 2023). When the first measurement (T1) of the receptive vocabulary was administered, the children in our sample were approximately 39 months old. At the second (T2), third (T3) and fourth (T4) measurement time point, they were on average 61, 85 and 111 months old, respectively.

Children with a vocabulary score below five were set to missing because the result could be due to pure randomness as the test ends prematurely (see Lenhard et al., 2015, p. 38). All children with a valid value in at least one of the four measurement time points of the receptive vocabulary were included in our analysis sample which leaves us with an overall sample size of $N = 2422$, of which $N = 953$ children have an immigrant background. An immigrant background is defined as having at least one grandparent who was not born in Germany (Olczyk et al., 2016a, b).

As other panel studies, the NEPS suffers from selective panel attrition. Between the waves 1 and 4 (at child age three), 1059 cases dropped out of the panel and this dropout is selective with regard to social origin and family language (see Table 3.A4 in the Appendix). In order to control for this selective panel dropout,

we use the weighting variable provided by NEPS in our descriptives and regressions. This weighting variable adjusts for the sample design and selective panel attrition, ensuring that the weighted sample is representative for children born in Germany in 2012 (Würbach, 2017). By drawing on a cohort of newborns, this study does not consider recently arrived migrants (i.e. first generation).

Measures and Operationalization

German Vocabulary

The German receptive vocabulary is measured with the German version of the Peabody Picture Vocabulary Test 4 (PPVT, Lenhard et al., 2015). In the digital version of the NEPS, children listen to a voice on the tablet computer while seeing four tiles with different pictures. Then they hear a specific word and are asked to touch the tile with the corresponding picture. The test consists of different sets that are ordered by difficulty, making the test age sensitive and the scores longitudinally comparable.

To contextualize our results, we compare them to a nationally representative normative sample of the German version of the PPVT (Lenhard et al., 2015). All assessments of children's vocabulary as being *above* or *below average* refer to a comparison with this normative sample. This clustered sample was drawn between September 2013 and April 2014 via schools and ECEC centers in several German states, utilizing quotas for characteristics such as sex, school type and immigrant background of the child. The sample consists of a total of 3555 children, 29,3% of whom have at least one immigrant parent.

Given the cross-sectional design of the normative study, we are comparing within-scores of the longitudinal NEPS assessment with the between-scores of the normative sample: For every year of age between three and 16 years, there are roughly 100 to 400 children in the normative sample for age-adequate evaluation. When comparing the mean vocabulary scores of our profiles with the cross-sectional reference groups of the PPVT sample, it is crucial to keep in mind that the mean of the normative sample includes the scores of highly proficient as well as very low scoring children. Finally, it is important to note that the normative sample may suffer from problems of selective participation as well: Because of the selective lower share of children with immigrant background in pre-school facilities, the normative sample cannot achieve representativeness for children until the mandatory school age of six years.

Exposure

Exposure is measured via a set of variables. First and foremost, we use the family household language coded as 0 = “only German” 1 = “both languages” and 2 = “only other language” as an indicator for exposure (Relikowski et al., 2015; Strobel & Seuring, 2016). Here, we use the first wave of the Newborn Cohort when the children are about seven months old to have the same point of measurement for all children and avoid losing information due to selective panel attrition (Zinn et al., 2020).

Furthermore, to account for exposure in the home learning environment, we add the frequency of reading to the child measured at the same time points as the first PPVT measurement. Reading to the child is coded as 1 = “once a week or less”, 2 = “multiple times a week”, 3 = “once daily” and 4 = “multiple times a day”.

The used language of siblings does not always match the language used by the parent when multiple languages are spoken in a family (Strobel & Seuring, 2016), therefore we add a dummy for the presence of an older sibling to control for additional potential German language exposure within the family.

For exposure outside of the family, we use the age of the child when first entering any kind of ECEC center (Becker, 2019; Melhuish et al., 2008). Since the spoken language in most of the ECEC centers is German, the child is exposed to more German the earlier ECEC attendance starts.

Efficiency

In the domain of efficiency, we include very early measurements for cognitive and sensorimotor development that were assessed at the age of six to eight months with an adaption of the internationally recognized Bayley Scales (Bayley, 2006) as a z-standardized variable. Moreover, the phonological working memory was measured when the children were three and six years old via the repetition of digit spans. The phonological working memory functions as an ability to process and store phonological information. The more spoken words children are able to process and store, the stronger their vocabulary will grow (Ebert et al., 2013; Gathercole, 2006). We add the variable of the phonological working memory as the number of correctly repeated digit spans.

SES Correlates

We use some indicators for the socio-economic status (SES) that influence exposure and efficiency at the same time and therefore constitute a separate category. These include the highest value on the *International Socio-Economic Index of Occupational Status* (ISEI) (Ganzeboom et al., 1992) and the education of the

mother measured via the *Comparative Analysis of Social Mobility in Industrial Nations* (CASMIN) classification (Brauns et al., 2003). While the former captures primarily occupation and income that influence for example the child's home learning environment (Becker, 2019), the latter adds the education of the mother (who is mostly the primary caregiver) as a proxy for the quality of verbal input (Hoff et al., 2018). We also include the number of books in the household as a proxy for cultural capital and opportunity for language exposure. It should be noted that these SES correlates have various direct and indirect impacts on children's language development (Becker, 2019; Bradley & Corwyn, 2002), also via their correlation with other relevant constructs such as immigrant parents' integration (Portes et al., 2009). Besides the inclusion of the other exposure and efficiency indicators, it is not feasible to consider the possible mechanisms.

Migration-Related Exposure Variables

To analyze the sub-group of the children with an immigrant background in more detail, we additionally include migration-related exposure variables. As a further proxy for the quality of language input, we included a binary variable for whether at least one parent is a native German speaker. The share of the child's friends with immigrant background at age five (1 = "none" to 7 = "all") (Kristen et al., 2016) as well as the residential segregation (Friedrichs & Triemer, 2009) as potential exposure reducing factors are also considered. The residential segregation was measured via the share of foreigners (defined by a name analysis of household leaders) and coded into a scale ranging from 1 "lowest" to 9 "highest". For the sake of brevity, we recoded this variable into 1 = "low", 2 = "medium" and 3 = "high". We also complement the frequency of shared book reading by adding a variable for the language of this activity at age two. Here, we are comparing the items for the frequencies of shared book reading in German and in the respective language of origin, recoding it into a new variable indicating 1 = "same amount of time for both languages", 2 = "more often in language of origin" and 3 = "more often in German".

Controls

We include migrants' generation status, which might capture further unmeasured aspects of exposure. If both parents and all grandparents were born in Germany, children have no immigrant background. Children are from the second immigrant generation if both parents were born abroad and migrated to Germany. We defined the group of *interethnics* if one parent was born abroad and the other parent and the associated grandparents were born in Germany. Finally, we categorized children as third immigrant generation if both parents were born in Germany but

at least one of their grandparents was not (Olczyk et al., 2016a, b). In addition, we control for the sex of the child.

The distribution of all variables is presented in Table 3.1, their correlations in Table 3.A6 in the Appendix.

Methods

To examine the different vocabulary trajectories, we apply a Growth Mixture Modeling (GMM) approach, which is a “model for identifying multiple unobserved sub-populations, describing longitudinal change within each unobserved sub-population and examining differences in change among unobserved sub-populations” (Ram & Grimm, 2009, p. 565). The GMM can be regarded as an extension of a Latent Growth Curve by additionally identifying latent classes, but instead of a priori dividing the sample into sub-groups, the classes will be chosen post hoc. However, the decision of the best class solution should be based on theory, state of research and careful descriptive examination of the data beforehand (Ram & Grimm, 2009; Wickrama et al., 2016). Every person in the sample is assigned a probability of belonging to a certain class. Membership to a latent class is then determined by the highest class probability of a person. To extract the latent classes of vocabulary trajectories, an unconditional GMM was calculated with Mplus 8.4 (Muthén & Muthén, 2017, pp. 1998–2019) using only the sum scores of the PPVT. Missing values (see Table 3.A5 for shares of missings) were accounted for by using multiple imputation with chained equations ($N = 50$ data sets) and combining them via Rubin’s rules (Rubin, 1987). Weights were applied to the extracted classes (*profiles*), in order to correct their proportions for selective dropout.

In the next step, we describe the profiles using weighted descriptive statistics (means or proportions of the explanatory variables). Finally, membership in the *risk profile* is examined with a weighted logistic regression using Stata 17.

Results

Table 3.1 shows the average German vocabulary from age three to nine among children with and without an immigrant background. As in previous studies, we observe a gap at the age of three that remains significant until the age of nine. On average, children with an immigrant background exhibit a score of 10 points

Table 3.1 Descriptive Statistics by Immigration Background and Profile

	Full sample	Non-Immigrants	Children with IB	At Risk Profile	Medium-Low Profile	Medium Profile	High Profile
Number of Cases ¹	2422	1469	953	129	659	1181	453
Intercept ²	—	—	—	25 ^{d,e,f}	41 ^{c,e,f}	57 ^{c,d,f}	72 ^{c,d,e}
Slope ²	—	—	—	34 ^{d,e,f}	53 ^{c,f}	58 ^{c,f}	67 ^{c,d,e}
Slope squared ²	—	—	—	1 ^{d,e,f}	—5 ^{c,f}	—6 ^{c,f}	—9 ^{c,d,e}
Vocabulary Age 3 Ø	49 (20)	53 ^a (20)	43 ^b (19)	23 ^{d,e,f} (11)	38 ^{c,e,f} (14)	56 ^{c,d,f} (15)	72 ^{c,d,e} (15)
Vocabulary Age 5 Ø	100 (24)	106 ^a (21)	91 ^b (25)	59 ^{d,e,f} (13)	86 ^{c,e,f} (11)	110 ^{c,d,f} (13)	133 ^{c,d,e} (13)
Vocabulary Age 7 Ø	138 (22)	143 ^a (19)	129 ^b (24)	94 ^{d,e,f} (16)	126 ^{c,e,f} (9)	147 ^{c,d,f} (9)	169 ^{c,d,e} (9)
Vocabulary Age 9 Ø	165 (19)	170 ^a (16)	159 ^b (21)	130 ^{d,e,f} (11)	154 ^{c,e,f} (9)	175 ^{c,d,f} (8)	190 ^{c,d,e} (6)
Baseline							
<i>Immigrant Background</i>		a	b	d,e,f	c,e,f	c,d	c,d
Non-Immigrants	59%	100%	—	17%	51%	69%	75%

(Fortsetzung)

Table 3.1 (Fortsetzung)

	Full sample	Non-Immigrants	Children with IB	At Risk Profile	Medium-Low Profile	Medium Profile	High Profile
2 nd Generation	20%	—	49%	72%	27%	7%	6%
Interethnics	9%	—	22%	8%	10%	10%	5%
3 rd Generation	12%	—	29%	3%	12%	14%	14%
Female	48%	47%	50%	43%	54%	48%	34 ^{c,d,e}
SES -Correlates							
HISEI Age 7	57	59 ^a	52 ^b	38 ^{d,e,f}	50 ^{c,e,f}	62 ^{c,d,f}	70 ^{e,d,e}
Months Ø	(21)	(20)	(22)	(17)	(21)	(18)	(16)
<i>Number of Books</i>		^a	^b	^{d,e,f}	^{c,e,f}	^{c,d,f}	^{c,d,e}
Age 2							
0-25	17%	10%	26%	64%	23%	6%	2%
26-200	52%	52%	52%	34%	61%	52%	40%
>200	31%	38%	22%	2%	16%	42%	58%
<i>Mother's Education Age 7</i>		^a	^b	^{d,e,f}	^{c,e,f}	^{c,d,f}	^{c,d,e}
<i>Months</i>							
Lower	15%	13%	18%	34%	22%	9%	1%
Secondary or None							
Intermediate (Realschule)	32%	31%	32%	43%	37%	29%	18%
Higher Secondary (Abitur)	28%	24%	33%	22%	25%	31%	29%

(Fortsetzung)

Table 3.1 (Fortsetzung)

	Full sample	Non-Immigrants	Children with IB	At Risk Profile	Medium-Low Profile	Medium Profile	High Profile
Tertiary Degree	26%	32%	18%	1%	16%	32%	52%
Exposure							
<i>Age of first ECEC Entry</i>		a	b	d,e,f	c,e	c,d	c
Younger than 2 Years	42%	47%	34%	15%	39%	47%	52%
Age 2 Years	34%	34%	33%	28%	37%	35%	25%
Age 3 Years or Older	24%	19%	33%	57%	25%	18%	23%
Older Sibling in Household	48%	47%	50%	54% ^f	53% ^f	47% ^f	37% ^{c,d,e}
<i>Family Language</i>		a	b	d,e,f	c,e,f	c,d	c,d
Only German	80%	100% ^a	51%	27%	75%	91%	94%
Both Languages	7%	0%	16%	5%	8%	7%	4%
Only other Language	13%	0%	33%	68%	17%	2%	2%
<i>Reading Aloud Age 3</i>		a	b	e,f	e,f	c,d,f	c,d,e
Once a Week or Less	5%	4%	7%	12%	8%	2%	0%
Multiple Times a Week	19%	14%	26%	35%	25%	13%	9%

(Fortsetzung)

Table 3.1 (Fortsetzung)

	Full sample	Non-Immigrants	Children with IB	At Risk Profile	Medium-Low Profile	Medium Profile	High Profile
Once Daily	34%	36%	33%	25%	35%	38%	27%
Multiple Times Daily	42%	47%	35%	28%	32%	47%	64%
Efficiency							
Sensomotoric Dev. Age 7 Months $\bar{\emptyset}$	-.10 (1.32)	-.12 (1.25)	-.09 (1.41)	-.50 ^{e,f} (1.50)	-.21 ^{e,f} (1.32)	-.01 ^{c,d} (1.23)	.17 ^{c,d} (1.37)
Phon. Working Memory Age 3 $\bar{\emptyset}$	3.09 (1.73)	3.02 (1.78)	3.20 (1.64)	2.54 ^{e,f} (1.23)	2.71 ^{e,f} (1.46)	3.24 ^{c,d,f} (1.77)	4.05 ^{c,d,e} (2.05)
Phon. Working Memory Age 6 $\bar{\emptyset}$	7.84 (1.77)	7.83 (1.78)	7.85 (1.76)	6.64 ^{d,e,f} (1.82)	7.31 ^{c,e,f} (1.60)	8.22 ^{c,d,f} (1.66)	8.83 ^{c,d,e} (1.64)
Migration-related Exposure							
Native German Speaker in Household	86%	100% ^a	66% ^b	38% ^{d,e,f}	82% ^{c,e,f}	96% ^{c,d}	98% ^{c,d}
Share of Foreigners in Postal District		a	b	d,e,f	c,e	c,d	c
Low	29%	40%	13%	12%	24%	36%	29%
Medium	30%	31%	27%	15%	29%	33%	32%

(Fortsetzung)

Table 3.1 (Fortsetzung)

	Full sample	Non-Immigrants	Children with IB	At Risk Profile	Medium-Low Profile	Medium Profile	High Profile
High	42%	29%	60%	74%	47%	31%	39%
Share of Child's Friends with IB Age 5 Ø	2.60	2.06 ^a	3.37 ^b	3.82 ^{d,e,f}	2.76 ^{c,e,f}	2.29 ^{c,d,f}	2.33 ^{c,d,e}
<i>Reading Language Age 2</i>	(1.47)	(1.06)	(1.61)	(1.72)	(1.56)	(1.24)	(1.16)
Same Amounts		^a	^b	^{d,e,f}	^{c,e,f}	^{c,d}	^{c,d}
More in other Language	10%	0%	25%	23%	10%	9%	8%
More in German	79%	100%	48%	41%	14%	5%	2%
				36%	76%	86%	91%

Source National Educational Panel Study (NEPS) Newborn Cohort, own calculations
Notes Weighted data was used. ¹Unweighted number of cases are reported. ²Weighting in Growth Mixture Model not available, therefore values of Intercept and Slopes are unweighted. ^{a-f}Significant difference ($p \leq 0.05$) to the group (a) children with immigrant background, (b) children without immigrant background, (c) Risk-Profile, (d) Medium-Low Profile, (e) Medium Profile, (f) High Profile. Tests for significant differences: Weighted logistic regressions were used for binary variables, weighted ordinal logistic regressions for categorical variables and weighted linear regressions for metric variables. Standard deviations in parentheses. Imputed values from $N = 50$ data sets were used. Percentages are rounded

or more below their non-immigrant peers, which corresponds to half of a standard deviation or to the learning progress of nearly six months. Differences by immigrant background are also evident in most exposure-related variables and in the SES correlates (see Table 3.1).

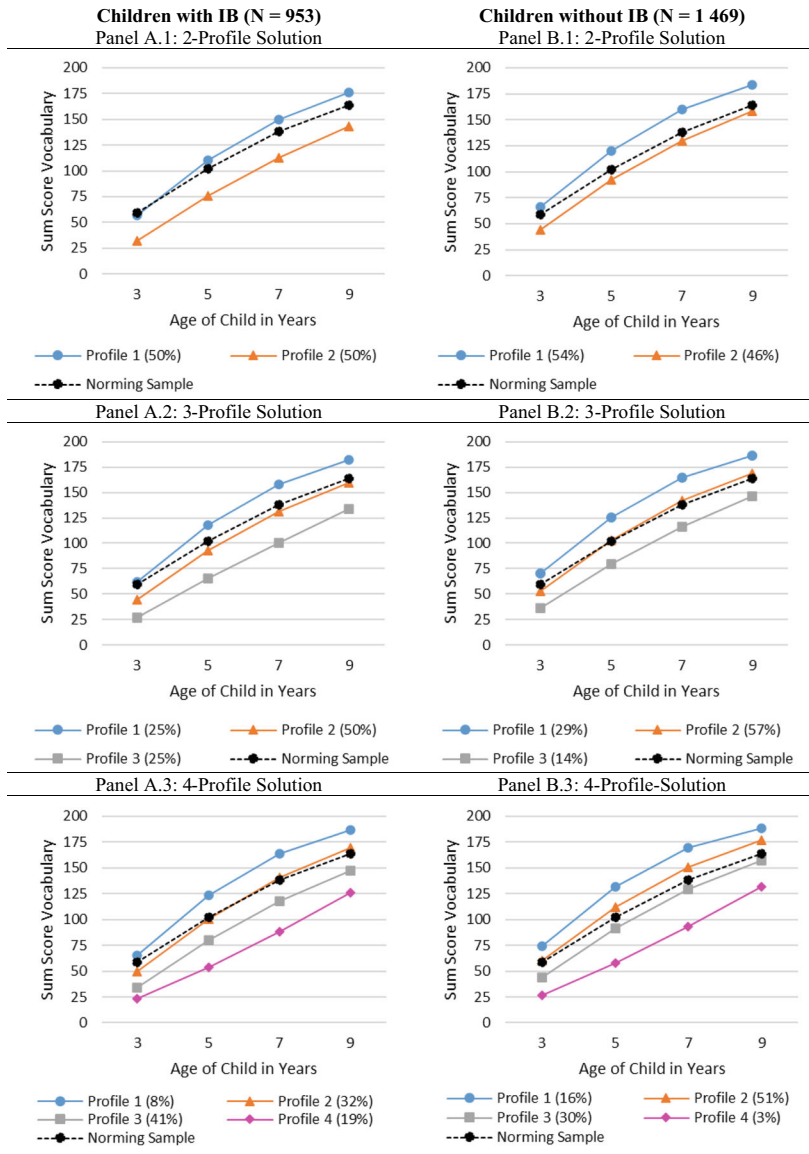
In order to identify different patterns of vocabulary development in children with an immigrant background (research question 1), we conducted GMM for this sample. The results are illustrated in Fig. 3.1, panel A. We present the solutions with two, three and four profiles. Each graph also includes the norm-referenced test scores (dashed black line). We do not present solutions with more than four profiles, as these solutions contained profiles with a very low number of cases.

We start with the *solution with two profiles* in the sample of children with immigrant background (Fig. 3.1, panel A.1). This solution differentiates between an average/above-average trajectory and a below-average trajectory, as compared to the normative sample. In the *solution with three profiles* (panel A.2), all three profiles show parallel trajectories but differ in their intercepts.

Finally, the *solution with four profiles* (panel A.3) differentiates between a high-achieving profile (profile 1: 8% of the sample), an average or medium profile (profile 2: 32% of the sample), a medium–low profile (profile 3: 41% of the sample), and an extremely below-average profile (profile 4: 19% of the sample). We regard profile 4 as a *risk profile* because the children in this profile have a German vocabulary that is about 35–40 points below the norm. This gap corresponds roughly to three years of learning, which puts these children at a very high risk for unfavorable educational outcomes.

The fit statistics (see Table 3.A2 in the Appendix) indicate good confidence of classification (entropy at around .80, see Ram & Grimm, 2009, p.571), but do not vary much between the presented solutions. Therefore, we used theoretical arguments in selecting the preferred solution, and opted for the solution with four profiles, as a “*risk profile*” is most obvious in this solution and the trajectories correspond well to our theoretical expectations. However, we also replicated the following detailed analysis of the risk profile using a 3-profile-solution, combining the “*medium–low*” profile with the “*risk profile*” since both profiles show constantly below-average vocabulary scores (see Appendix, Table 3.A3).

In order to answer our second research question, we compared the results from the GMM between the sample of children with immigrant background (Fig. 3.1, panel A) and the sample of non-immigrant children (Fig. 3.1, panel B). Most strikingly, the general patterns look very similar in both samples. Two main differences are notable: First, the vocabulary level of non-immigrant children within each profile is usually slightly higher than that of children with an immigrant background. Second, the distribution of children across the profiles differs by



Source: National Educational Panel Study (NEPS) Newborn Cohort, own calculations.
Notes: Share of each profile (with weighting) is shown in parentheses (rounded).

Fig. 3.1 Profile Solutions of Children with and without Immigrant Background

immigrant background. This especially applies to the “*risk profile*” (profile 4 in the 4-profile-solution) which includes 19% of the immigrant sample but only 3% of the non-immigrant sample.

Because the identified trajectories are highly similar in the sample of immigrant and non-immigrant children, we continue our analysis using the solution with four profiles from the full sample (see Appendix, Figure 3.A1, Panel C.3). The vocabulary scores of the four profiles in the full sample are presented in Table 3.1. As a next step, we describe these four profile with regard to child and family characteristics (research question three, Table 3.1). Overall, 41% of the children in the full sample have an immigrant background. However, this share is only 25% in the *high profile*, but 49% in the *medium–low* and even 83% in the *risk profile*. With the exception of the presence of older siblings, the exposure and efficiency indicators tend to be above-average in the *high profile*, mostly average in the *medium profile* and slightly below-average in the *medium–low profile* and markedly below-average in the *risk profile*.

As a final step, we analyze membership in the risk profile using weighted logistic regressions. The results for the full sample and the different subsamples are presented in Table 3.2. In the full sample (Model 1) the number of books, a late entry into ECEC, using only another family language than German as well as phonological working memory at age six are predictive for membership in the risk profile. The latter is also the only significant effect for the non-immigrant subsample (Model 2). The absence of other significant differences might be due to the small number of children in the risk profile in this subsample ($N = 26$).

The results for the group of children with an immigrant background mostly resemble those of the full sample with the addition of a more pronounced influence of an early ECEC entry and no effect of the phonological working memory. Furthermore, an intermediate educational certificate of the mother raises the probability of being in the risk profile compared to having a lower secondary or even no educational certificate, which is a rather contradictory effect and should be interpreted with caution. Finally, Model 4 shows the additional results for migration-related exposure variables in the subsample of children with immigrant background. Controlling for all other variables, having at least one native German speaking parent reduces the probability of being in the risk profile by about 11%, while the aforementioned effect of the maternal education disappears and the phonological working memory becomes significant.

Table 3.2 Avg. Marginal Effects from the Logistic Regression of Membership in Risk Profile

	Full Sample	Without IB	With IB	With IB
	Model 1	Model 2	Model 3	Model 4
Immigrant Background				
Non-Immigrant Student	.0140			
2 nd Generation	.0461		.0699	.0576
Interethnics	.0493		.0746	.0821
3 rd Generation (<i>reference category</i>)				
Female	-.0040	.0041	-.0098	-.0200
Highest ISEI in Family at Age 7 Months	-.0005	.0000	-.0011	-.0012
Mother's Education at Age 7 Months				
Lower Sec. or None (<i>reference category</i>)				
Intermediate ("Realschule")	.0251	-.0315	.0841*	.0760
Higher Secondary ("Abitur")	-.0071	-.0342	.0131	-.0054
Tertiary Degree	-.0485 ⁺	-.0427	-.0916	-.0964 ⁺
Number of Books in Household				
Less than 25 (<i>reference category</i>)				
26–200	-.0581*	-.0416	-.0854*	-.1029*
More than 200	-.1023***	-.0507	-.1839**	-.1844**
Age of first ECEC Entry				
Younger than 2 Years (<i>reference category</i>)				
Age 2 Years	.0238	-.0052	.0750*	.0771*
Age 3 Years or Older	.0478*	.0005	.1192**	.0978*
Older Sibling in Household	.0035	.0123	.0085	-.0016
Family Language				
Only German (<i>reference category</i>)				
Both Languages	.0035		.0229	-.0049
Only other Language	.2003***		.2651***	.1584**

(Fortsetzung)

Table 3.2 (Fortsetzung)

	Full Sample	Without IB	With IB	With IB
	Model 1	Model 2	Model 3	Model 4
Reading Aloud Age 3				
Once a Week or Less (<i>reference category</i>)				
Multiple Times a Week	-.0116	.0048	-.0399	-.0613
Once Daily	-.0165	.0376 ⁺	-.0713	-.0841
Multiple Times Daily	-.0132	.0243	-.0611	-.0530
Sensomotoric Development at 7 Months	-.0109 ⁺	-.0108 ⁺	-.0134	-.0185 ⁺
Phonological Working Memory Age 3	-.0047	-.0038	-.0095	-.0107
Phonological Working Memory Age 6	-.0142 ^{**}	-.0156 ^{***}	-.0140	-.0155 [*]
Native German Speaker in Household				-.1121 ^{**}
Share of Foreigners in Postal District				
Low (<i>reference category</i>)				
Medium				-.0506
High				-.0334
Share of Child's Friends with IB Age 5				.0028
Reading Language at Age 2				
Same amounts (<i>reference category</i>)				
More in German				-.0465
More in other Language				-.0816 ⁺
Average Pseudo-R ²	.32	.45	.42	.50
N	2422	1469	953	953

Source National Educational Panel Study (NEPS) Newborn Cohort, own calculations

Notes Weighted data was used. Dependent variable coded as 0 = Not at Risk / 1 = At Risk
 N of Risk Profile in full Sample = 129. N of Risk Profile in Subsample without Immigration Background (IB) = 26. N of Risk Profile in Subsample with IB = 103

⁺p ≤ .1, ^{*}p ≤ 0.05, ^{**}p ≤ 0.01, ^{***}p ≤ 0.001

Discussion

We analyzed patterns of German vocabulary development from age three to nine among children with an immigrant background in Germany with a focus on identifying children who are at risk for unfavorable educational trajectories. However, instead of only considering the average vocabulary among children with and

without immigrant background, we used an explorative approach to investigate the heterogeneity in children's early vocabulary trajectories within both groups in more detail.

Our research has yielded four main findings. *First*, we have found that a large share of children with an immigrant background show an average or even above-average German vocabulary trajectory, as compared to the normative sample. *Second*, there are striking similarities between the vocabulary trajectories of children with an immigrant background and their non-immigrant peers. The Growth Mixture Models have provided very similar patterns in both samples. *Third*, we identified a *risk profile* in both samples. However, the share of children in this risk profile is considerably larger in the sample of immigrant children (19%) than in the sample of non-immigrant children (3%). *Fourth*, children with a non-German home language, no native German speaking parent, less cultural capital at home, a later ECEC entry and lower (language relevant) cognitive skills are more likely to be in the risk profile. For non-immigrant children, only cognitive skills are predictive for membership in the risk profile.

While the empirical results match our theoretical expectations about a *high achieving profile*, an average profile and a *risk profile*, it is notable that we could not identify a *catching up profile*. One reason for this could be the fact that most children have already started attending ECEC before the first measurement occasion, with a result that a catching up process may have occurred earlier but was not observed in our data. It is generally remarkable how stable the differences between the profiles remained over time. This finding is in line with other research showing that early gaps in language skills tend to be very persistent (Skopek & Passaretta, 2021; Sprong & Skopek, 2023).

Our study has several limitations. *First*, like most longitudinal studies, the NEPS also suffers from selective panel attrition, i.e., children with lower SES origins and non-German family languages are more likely to drop out (see Table 3.A4 in the Appendix). Hence, it is likely that we underestimate the share of children in the risk profile and its negative trajectory (this applies to children with and without immigrant background). Thus, one should be cautious about generalizing our results to the whole population of children in the age range of three to nine years (with and without immigrant background) of Germany and further studies are needed to substantiate our results. The selective panel attrition may also be a reason why we did not detect a *catching-up profile*: Children in such a profile (with low vocabulary scores in the beginning) would most likely come from a low-SES and non-German family background, who are exactly those with the highest dropout rates. However, we tried to account for this problem and minimize the bias by using the weights provided by NEPS after conducting the

GMM. *Second*, we only considered children's receptive vocabulary in German, which should not be regarded as a measure of children's overall language skills. Neither did we consider other aspects of children's language skills (e.g., productive vocabulary or grammar), nor language of origin skills in the case of bilingual children. While this is a clear limitation, we can justify this restricted focus on the basis of the general importance of majority language vocabulary for children's later educational success (Durham et al., 2007; Prevoo et al., 2016). *Third*, we were not able to include any characteristics of children's ECEC. This might also be the reason why we only observed a weak to moderate association between ECEC attendance and the probability of being in the risk profile. Other studies have demonstrated that ECEC quality rather than mere attendance of *any* ECEC is a crucial determinant for children's development (for an overview see Roßbach et al., 2008). In particular, (language-related) ECEC quality is also crucial for children's language development, especially among immigrant children (Klein & Becker, 2017; Schmerse et al., 2018). Literature has also shown that the choice of high-quality ECEC centers is socially selective (Lehrl et al., 2014; Stahl et al., 2018), which we could not address in our study. *Fourth*, while we analyzed children's vocabulary trajectories longitudinally from age three to nine, the change in most exposure and efficiency indicators could not be captured at all observation times. Thus, the analysis of belonging to the risk profile is cross-sectional. *Fifth*, we could not control for the country of origin in immigrant families due to low number of cases per category. However, we considered the generation status as more crucial and included this in all analyses. *Sixth*, it should be noted that the NEPS Newborn sample was drawn in 2012. Therefore, our analyses did not capture children from recent immigration waves like refugees from Syria or Ukraine. It should also be noted that, by design (Newborn cohort), our study did not include first generation immigrant children and we cannot draw any conclusions about this group. *Last*, we have to mention the frequent limitation of survey data that we cannot rule out the influence of unobserved characteristics and cannot make any causal claims.

However, we conducted several robustness checks such as excluding children with speech impairments or therapy and re-running the analysis. The results were overall very similar with some very minor differences in the intercept and slope of the medium–low and risk profile. These two profiles were also combined into one category to re-run the logistic regression and examine differences of being below average vs. being in the medium or high profile (Table 3.A3). Since children were assigned to a profile based on the highest probability of membership, we additionally conducted a linear regression using the probability of being in the risk profile as dependent variable (Table 3.A7). The results were nearly identical

to those of the logistic regression with profile membership based on the highest probability as the dependent variable. Finally, because half of the children with immigrant background speak only German at home, we replicated the GMM without them and the results revealed similar trajectories with minor differences on the intercept.

Furthermore, our findings bear important implications. Most importantly, we demonstrated heterogeneous trajectories of German vocabulary development among children with an immigrant background. Thus, *children with an immigrant background* should not be regarded as a homogeneous group—also not with respect to their early German language skills. Children with an immigrant background are often considered as a risk group for poor educational outcomes, which is often formulated as a blanket statement for the whole group. We can contradict this statement now as many of them exhibit an average or even above-average German vocabulary trajectory in early childhood. However, we identified a noteworthy share of children with extremely below-average German vocabulary levels at all measurement points who must be regarded as at risk for unfavorable educational trajectories in the future. Such *children at risk* can be found among children with and without immigrant background alike—although the shares differ substantially. Since this risk group already shows very low levels of vocabulary and cognitive skills very early in childhood, early diagnostic measures as well as appropriate interventions are pivotal tools for proper language trajectories.

References

- Ahrenholz, B. (2010). Bedingungen des Zweitspracherwerbs in unterschiedlichen Altersstufen. In Friedrich-Ebert-Stiftung (Hrsg.), *Sprache ist der Schlüssel zur Integration. Bedingungen des Sprachlernens von Menschen mit Migrationshintergrund* (S. 19–42). Friedrich-Ebert-Stiftung.
- Alexandersen, N., Zachrisson, H. D., Wilhelmsen, T., Wang, M. V., & Brandlistuen, R. E. (2021). Predicting selection into ECEC of higher quality in a universal context: The role of parental education and income. *Early Childhood Research Quarterly*, 55, 336–348.
- Attig, M., Vogelbacher, M., & Weinert, S. (2023). Education from the crib on: The potential of the newborn cohort of the German national educational panel study. *Journal of Open Psychology Data*, 11(1), 13.
- Bayley, N. (2006). *Scales of infant and toddler development* (3. Aufl.). Harcourt Assessment.
- Becker, B. (2011). Cognitive and language skills of Turkish children in Germany: A comparison of the second and third generation and mixed generational groups. *International Migration Review*, 45(2), 426–459.

- Becker, B. (2019). Early educational inequality—Growing up in Different Learning Environments. In R. Becker (Hrsg.), *Research handbook on the sociology of education* (S. 233–252). Elgar.
- Becker, B., & Klein, O. (2021). The primary effect of ethnic origin—Rooted in early childhood? An analysis of the educational disadvantages of Turkish-origin children during the transition to Secondary Education in Germany. *Research in Social Stratification and Mobility*, 75, Article 100639.
- Becker, B., Klein, O., & Biedinger, N. (2013). The development of cognitive, language, and cultural skills from age 3 to 6: A comparison between children of Turkish origin and children of native-born German parents and the role of immigrant parents' acculturation to the receiving society. *American Educational Research Journal*, 50(3), 616–649.
- Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning*, 28(1), 69–83.
- Blossfeld, H.-P., & Roßbach, H.-G. (Hrsg.). (2019). *Education as a lifelong process. The German National Educational Panel Study (NEPS)*. Springer VS.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53(1), 371–399.
- Brauns, H., Scherer, S., & Steinmann, S. (2003). The CASMIN educational classification in international comparative research. 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. 221–244). Springer, US.
- Chiswick, B. R., & Miller, P. W. (1998). Language skill definition: A study of legalized aliens. *International Migration Review*, 32(4), 877–900.
- Chiswick, B. R., & Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, 38(3), 391–409.
- Dale, P. S., Paul, A., Rosholm, M., & Bleses, D. (2023). Prediction from early childhood vocabulary to academic achievement at the end of compulsory schooling in Denmark. *International Journal of Behavioral Development*, 47(2), 123–134.
- Deák, G. O. (2014). Interrelationship of language and cognitive development (Overview). In P. J. Brooks & V. Kempe (Hrsg.), *Encyclopedia of language development* (S. 284–291). Sage.
- deSouza, D. K., Lin, H., & Cox, R. B. (2023). Immigrant parents and children navigating two languages: A scoping review. *Journal of Family Theory & Review*, 15(1), 133–161.
- Diehl, C., Hunkler, C., & Kristen, C. (Hrsg.). (2016). *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*. Springer VS.
- Durham, R. E., Farkas, G., Hammer, C. S., Tomblin, J. B., & Catts, H. W. (2007). Kindergarten oral language Sskill: A key variable in the intergenerational transmission of socioeconomic status. *Research in Social Stratification and Mobility*, 25(4), 294–305.
- Dustmann, C. (1994). Speaking fluency, writing fluency and earnings of migrants. *Journal of Population Economics*, 7(2), 133–156.
- Ebert, S., Lockl, K., Weinert, S., Anders, Y., Kluczniok, K., & Rossbach, H.-G. (2013). Internal and external influences on vocabulary development in preschool children. *School Effectiveness and School Improvement*, 24(2), 138–154.
- Esser, H. (2006). *Migration, language and integration*. AKI research review 4. Social Science Research Center Berlin (WZB).

- Friedrichs, J., & Triemer, S. (2009). *Gespaltene Städte? Soziale und ethnische Segregation in deutschen Großstädten* (2. Aufl.). VS Verlag.
- Fuchs-Rechlin, K., & Bergmann, C. (2014). Der Abbau von Bildungsbenachteiligung durch Kindertagesbetreuung für unter 3-Jährige—zwischen Wunsch und Wirklichkeit. *Zeitschrift für Erziehungswissenschaft*, 17(2), 95–118.
- Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56.
- Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27(4), 513–543.
- Gogolin, I. (2008). *Der monolinguale Habitus der multilingualen Schule* (2. Aufl.). Waxmann.
- Grundy, J. G., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, 33(3), 325–340.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Brookes Publishing Company.
- Henschel, S., Heppt, B., & Weirich, S. (2023). Zuwanderungsbezogene Disparitäten. In P. Stanat, S. Schipolowski, R. Schneider, S. Weirich, S. Henschel & K. A. Sachse (Hrsg.), *IQB-Bildungstrend 2022. Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im dritten Ländervergleich* (S. 299–344). Waxmann.
- Hoff, E. (2014). *Language development* (5. Aufl.). Cengage Learning.
- Hoff, E. (2018). Bilingual development in children of immigrant families. *Child Development*, 12(2), 80–86.
- Hoff, E., Burrige, A. B., Ribot, K. M., & Giguere, D. (2018). Language specificity in the relation of maternal education to bilingual children's vocabulary growth. *Developmental Psychology*, 54(6), 1011–1019.
- Højen, A., Bleses, D., Jensen, P., & Dale, P. S. (2019). Patterns of educational achievement among groups of immigrant children in Denmark emerge already in preschool second-language and preliteracy skills. *Applied Psycholinguistics*, 40(4), 853–875.
- Hunkler, C., & Schotte, K. (2023). Educational integration by the third generation? Placement and academic achievement of students with immigrant background in Germany. *Zeitschrift für Erziehungswissenschaft*, 26(2), 373–401.
- Kempert, S., Edele, A., Rauch, D., Wolf, K. M., Paetsch, J., Darsow, A., & Stanat, P. (2016). Die Rolle der Sprache für zugewanderungsbezogene Ungleichheiten im Bildungserfolg. In C. Diehl, C. Hunkler, & C. Kristen (Hrsg.), *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten* (S. 157–241). Springer VS.
- Kigel, R. M., McElvany, N., & Becker, M. (2015). Effects of immigrant background on text comprehension, vocabulary, and reading motivation: A longitudinal study. *Learning and Instruction*, 35, 73–84.
- Klein, O., & Becker, B. (2017). Preschools as language learning environments for children of immigrants. Differential effects by familial language use across different preschool contexts. *Research in Social Stratification and Mobility*, 48, 20–31.
- Klein, W., & Dimroth, C. (2003). Der ungesteuerte Zweitspracherwerb Erwachsener: Ein Überblick über den Forschungsstand. In Vorstand des IMIS (Hrsg.), *IMIS-Beiträge 21. Themenheft Qualitätsanforderungen für die Sprachförderung im Rahmen der Integration von Zuwanderern* (S. 127–161). Institut für Migrationsforschung und Interkulturelle Studien (IMIS).

- Kristen, C. (2019). Language assimilation and the education of immigrant students. In R. Becker (Hrsg.), *Research handbook on sociology of education* (S. 519–534). Elgar.
- Kristen, C., Mühlau, P., & Schacht, D. (2016). Language acquisition of recently arrived immigrants in England, Germany, Ireland, and the Netherlands. *Ethnicities, 16*(2), 180–212.
- Lehrl, S., Kuger, S., & Anders, Y. (2014). Soziale Disparitäten beim Zugang zu Kindergartenqualität und differenzielle Konsequenzen für die vorschulische mathematische Entwicklung. *Unterrichtswissenschaft, 42*(2), 132–151.
- Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. P. (2015). *Peabody picture vocabulary test—revision IV (Deutsche adaption)*. Pearson Assessment.
- Linberg, A., Lehrl, S., & Weinert, S. (2020). The early years home learning environment—associations with parent-child-course attendance and children’s vocabulary at age 3. *Frontiers in Psychology, 11*, Article 1425.
- Linberg, T., & Wenz, S. E. (2017). Ausmaß und Verteilung sozioökonomischer und migrationspezifischer Ungleichheiten im Sprachstand fünfjähriger Kindergartenkinder. *Journal for Educational Research Online (JERO), 9*(1), 77–98.
- Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues, 64*(1), 95–114.
- Mitchell, R., Myles, F., & Marsden, E. (2019). *Second language learning theories* (4. Aufl.). Routledge.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user’s guide*. Muthén & Muthén.
- NEPS Network. (2023). National educational panel study, scientific use file of starting cohort newborns. Leibniz Institute for Educational Trajectories (LifBi). <https://doi.org/10.5157/NEPS:SC1:10.1.0>.
- Olczyk, M., Seuring, J., Will, G., & Zinn, S. (2016a). Migranten und ihre Nachkommen im deutschen Bildungssystem: Ein aktueller Überblick. In C. Diehl, C. Hunkler, & C. Kristen (Hrsg.), *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten* (S. 33–70). Springer VS.
- Olczyk, M., Will, G., & Kristen, C. (2016). *Immigrants in the NEPS: identifying generation status and group of origin*. NEPS Survey Paper No. 4. LifBi.
- Organisation for Economic Co-operation and Development (OECD). (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing.
- Paradis, J. (2019). *English second language acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors*. Paper presented at the Proceedings of the 43rd Boston University Conference on Language Development, Somerville, MA.
- Portes, A., Fernandez-Kelly, P., & Haller, W. (2009). The adaptation of the immigrant second generation in America: A theoretical overview and recent evidence. *Journal of Ethnic and Migration Studies, 35*(7), 1077–1104.
- Prevo, M. J. L., Malda, M., Mesman, J., & van IJzendoorn, M. H. (2016). Within- and cross-language relations between oral language proficiency and school outcomes in bilingual children with an immigrant background: A meta-analytical study. *Review of Educational Research, 86*(1), 237–276.

- Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, 33(6), 565–576.
- Relikowski, I., Schneider, T., & Linberg, T. (2015). Rezeptive Wortschatz- und Grammatikkompetenzen von Fünfjährigen mit und ohne Migrationshintergrund. Eine empirische Untersuchung aus bildungssoziologischer Perspektive. *Frühe Bildung*, 4(3), 135–143.
- Rodriguez, E. T., & Tamis-LeMonda, C. S. (2011). Trajectories of the home learning environment across the first 5 years: Associations with children's vocabulary and literacy skills at prekindergarten. *Child Development*, 82(4), 1058–1075.
- Roßbach, H.-G., Kluczniok, K., & Kuger, S. (2008). Auswirkungen eines Kindergartenbesuchs auf den kognitiv-leistungsbezogenen Entwicklungsstand von Kindern. *Zeitschrift für Erziehungswissenschaft*, 10(Sonderheft 11), 139–158.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.
- Rydland, V., Aukrust, V. G., & Fulland, H. (2013). Living in neighborhoods with high or low co-ethnic concentration: Turkish–Norwegian-speaking students' vocabulary skills and reading comprehension. *International Journal of Bilingual Education and Bilingualism*, 16(6), 657–674.
- Schmerse, D., Anders, Y., Flöter, M., Wieduwilt, N., Roßbach, H.-G., & Tietze, W. (2018). Differential effects of home and preschool learning environments on early language development. *British Educational Research Journal*, 44(2), 338–357.
- Schnepf, S. V. (2007). Immigrants' educational disadvantage: An examination across ten countries and three surveys. *Journal of Population Economics*, 20(3), 527–545.
- Seuring, J., & Will, G. (2022). German language acquisition of refugee children—The role of preschools and language instruction. *Frontiers in Sociology*, 7, 1–14.
- 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.
- Spolsky, B. (1989). *Conditions for second language learning. Introduction to a general theory*. Oxford University Press.
- Sprong, S., & Skopek, J. (2023). The development of host language achievement gaps by migration background during primary school: A path model of educational inequalities. *British Educational Research Journal*, 49(6), 1273–1290.
- Stahl, J. F., Schober, P. S., & Spieß, C. K. (2018). Parental socio-economic status and childcare quality: Early inequalities in educational opportunity? *Early Childhood Research Quarterly*, 44, 304–317.
- Strobel, B., & Seuring, J. (2016). Spracherhalt oder Sprachverlagerung? Erstsprachgebrauch und Zweitsprachkompetenzen bei Jugendlichen mit Migrationshintergrund. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(2), 309–339.
- van Tubergen, F., & Mentjox, T. (2014). Minority language proficiency of adolescent immigrant children in England, Germany, the Netherlands, and Sweden. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66, 241–262.
- Washbrook, E., Waldfogel, J., Bradbury, B., Corak, M., & Ghangro, A. A. (2012). The development of young children of immigrants in Australia, Canada, the United Kingdom, and the United States. *Child Development*, 83(5), 1591–1607.
- Wickrama, K. A. S., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher-order growth curves and mixture modeling with Mplus: A practical guide*. Routledge/Taylor & Francis.

- Würbach, A. (2017). *Samples, weights and nonresponse: The early childhood cohort of the national educational panel study (Wave 4) (NEPS Technical Report)*. Leibniz Institute for Educational Trajectories (LIfBi).
- Zinn, S., Würbach, A., Steinhauer, H. W., & Hammon, A. (2020). Attrition and selectivity of the NEPS starting cohorts: An overview of the past 8 years. *ASTA Wirtschafts- und Sozialstatistisches Archiv*, 14(2), 163–206.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Sexuelle Gewalt in der Kindheit und Bildung – Wissensgenerierung durch Aufarbeitung und Zeugenschaft

4

Sabine Andresen und Andrea Pohling

Zusammenfassung

Der Beitrag fokussiert sexuelle Gewalt an Kindern und Jugendlichen als internationales und institutionell gerahmtes *pädagogisches* Phänomen aus einer kindheitstheoretischen Perspektive. Methodisch basiert er auf einem vergleichsweise neuen Weg, um Wissen über Unrecht an Kindern und Jugendlichen und sexuelle Gewalt zu generieren: Seit den 1990er Jahren liegen Erfahrungen mit und Berichte von Aufarbeitungskommissionen weltweit vor, die mit dem Konzept der Zeugenschaft arbeiten. Überwiegend wird dabei die Aufarbeitung sexuellen Kindesmissbrauchs und der Misshandlung von Kindern und Jugendlichen in institutionellen Erziehungs- und Bildungssettings aller Art adressiert. Für den vorliegenden Beitrag wurden Berichte, die Aufarbeitungskommissionen aus Deutschland, Australien, England und Kanada (Eine Auswahl von internationalen Kommissionen findet sich auf der Homepage der Unabhängigen Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs in Deutschland (<https://www.aufarbeitungskommission.de/kommission/aufarb>

S. Andresen (✉) · A. Pohling
Goethe-Universität Frankfurt am Main, Frankfurt am Main, Deutschland
E-Mail: S.Andresen@em.uni-frankfurt.de

A. Pohling
E-Mail: Pohling@em.uni-frankfurt.de

S. Andresen
IDeA – Center for Research on Individual Development and Adaptive Education of Children at Risk, Frankfurt am Main, Germany

itung/aufarbeitung-international, abgerufen am 2. März 2023). Eine weitere Übersicht über Kommissionen und Untersuchungen international bietet die Website <https://www.lib.latrobe.edu.au/research/ageofinquiry/index.html> des Department of Social Inquiry der La Trobe University in Melbourne Australien unter der Überschrift: „The age of inquiry. A global mapping of institutional abuse inquiries“ (abgerufen am 17.03.2023) veröffentlicht haben, herangezogen und der Frage nachgegangen, auf welche Weise betroffene Menschen hierin das Verhältnis von Bildung, Erziehung und Gewalterfahrungen beschreiben, also welche Bezüge zu pädagogischen institutionellen Settings und Bildungsprozessen hergestellt werden. Eine grundlegende Fragestellung, die nicht nur, aber vor allem auch für im Bildungsbereich tätige pädagogische Fachkräfte relevant ist. Dabei zeigt sich als ein grundlegendes Ergebnis, dass in allen untersuchten Berichten der Aufarbeitungskommissionen pädagogische Settings bzw. institutionalisierte Bildungskontexte als Ermöglichungsräume für sexuellen Kindesmissbrauch beschrieben werden. Zudem wird gezeigt, dass auf drei Ebenen Erfahrungen sexueller Gewalt in einer längerfristigen biografischen, aber auch gesamtgesellschaftlichen Dimension in Bezug auf Bildung und Erziehung thematisiert werden: 1. auf der Ebene der Bildungsbiographie Betroffener, 2. auf der Ebene des erzieherischen Verhältnisses Betroffener gegenüber eigenen Kindern sowie 3. auf der Ebene von Bildung als Schutzfaktor vor zukünftiger Gewalt.

Schlüsselwörter

Sexuelle Gewalt • Folgen von Gewalt • Bildungsinstitutionen • Aufarbeitung • Zeugenschaft

Gewalt als Teil von Bildung und Erziehung – einleitende Überlegungen

Es gibt gute Gründe, das Recht von Kindern und Jugendlichen auf gewaltfreies Aufwachsen auch als Ausgangspunkt für die Gestaltung ihrer Bildung zu betrachten. Die Geschichte von Bildung und Erziehung ließe sich als Geschichte der Gewalt in allen Konstellationen, in denen Kindern und Jugendlichen Erziehungs- und Bildungsangebote gemacht werden, erzählen (Andresen, 2018). Seit dem Jahr 2000 heißt es im Bürgerlichen Gesetzbuch (BGB) in Deutschland: „Kinder haben

ein Recht auf gewaltfreie Erziehung. Körperliche Bestrafungen, seelische Verletzungen und andere entwürdigende Maßnahmen sind unzulässig“ (BGB § 1631, 2). Diese Gesetzesnovelle zielte zunächst primär auf die Familie und schränkt seither das Recht der Eltern ein bzw. ruft diese dazu auf, auf Gewalt in jeder Form zu verzichten und sich gegebenenfalls in der Erziehungsberatung Unterstützung zu suchen. Inzwischen hat sich der Blick geweitet und ist über die Familie hinaus auf alle Bereiche des Aufwachsens, also auch auf Bildungsinstitutionen wie Kindergarten und Schule, gerichtet. Der Begriff des Aufwachsens korrespondiert mit den erziehungswissenschaftlichen Grundbegriffen Erziehung, Bildung, Sorge, Hilfe und Lernen, die ihren Ort im Generationenverhältnis haben. Generationenverhältnisse und die soziale Position, die Kinder und Jugendliche als Aufwachsende darin haben, sind auch leitend für die hier im Folgenden vorgenommene Analyse von Aufarbeitungsberichten. In den Aufarbeitungsberichten findet sich Wissen über Gewalt an Kindern und Jugendlichen in Bildungsinstitutionen oder Einrichtungen der Kinder- und Jugendhilfe. Der Fokus liegt also auf denjenigen Formen von Gewalt, die ebenfalls in generationalen Ordnungen verortet sind. Somit stellt sich systematisch und empirisch die Frage, was es bedeutet, als Kind oder jugendliche Person, die von wohlmeinenden Erwachsenen in Familie oder Schule abhängig ist, Gewalt zu erleben.

Mit diesem Blick auf Gewalt, verortet in generationalen Ordnungen und damit auch als möglicher Teil von Bildungskonstellationen, sind folgende Überlegungen verbunden: Geht es um Gewalt, so lassen sich Erziehung und Bildung nicht strikt voneinander trennen, sie sind vielmehr eng miteinander verflochten. Aus diesem Grund wird im Folgenden von Bildung **und** Erziehung die Rede sein. Zu berücksichtigen ist, dass Gewalt nicht das ganz andere von Erziehungs- und Bildungsverhältnissen ist, sondern diesen inhärent sein kann. Damit ist gerade für Pädagog*innen Wissen darüber relevant, mit welchen Gewaltrisiken der Anspruch und die Notwendigkeit, Kindern Bildung und Erziehung angedeihen zu lassen, einhergehen kann.

Die historische Forschung hat zu struktureller Gewalt verbunden mit sadistischen Praktiken von Eltern, Lehrkräften oder Sozialpädagog*innen und Erzieher*innen, Geistlichen mit Bildungsaufgaben oder Trainer*innen in sportpädagogischen Settings, eindrucksvolle Beispiele vorgelegt. So ist die Geschichte der Heimerziehung in der Bundesrepublik der Nachkriegszeit – also die 1950er und 1960er Jahre – ein Beispiel dafür, wie neben brutalen Handlungsweisen gegenüber den *Zöglingen* auch der Entzug von Bildungsangeboten, das Vorenthalten von Qualifikationen, gezieltes Verhindern von Schulabschlüssen zu einer spezifischen Form von Gewalt wurde (Runder Tisch Heimerziehung in den 50er und 60er Jahren 2010).

Bereits das bislang vorliegende Wissen über Gewalt im Kontext von Bildung und Erziehung – auf das in diesem Beitrag eingegangen wird – fordert die empirische, historische und systematische Bildungsforschung heraus. Weder die Thematisierung sexueller Gewalt gegen Kinder und Jugendliche noch gut durchdachte und gesteuerte Konzepte gegen diese sind selbstverständliche Bestandteile der gesellschaftlichen und pädagogischen Auseinandersetzung. Ein in seiner Kindheit von sexueller Gewalt innerhalb der Familie Betroffener äußert sich in einer vertraulichen Anhörung bei der „Unabhängigen Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs“ dazu folgendermaßen:¹ „Und diese Kommission hatte ich ... also ich habe selbst eine große Hoffnung, dass da mal was ins Rollen kommt. Weil das Thema sexueller Missbrauch ist nach wie vor ein Tabuthema, es wird wenig darüber gesprochen, wenige wissen darüber. Eigentlich hat auch keiner irgendwie eine Ahnung, was für eine Dunkelziffer da vorliegt.“²

Hier legt ein betroffener Mensch Zeugnis ab und teilt nicht nur seine Hoffnung und Einschätzung, sondern auch sein spezifisches Wissen. Geteilte Erfahrungen mit der Tabuisierung sexueller Gewalt stehen vor großen Hürden, gesehen, gehört, anerkannt und einbezogen zu werden. Ähnliches gilt für das Wissen über die Ohnmachtsgefühle von Kindern und Jugendlichen, über die vielen gesellschaftlich etablierten und akzeptierten Schweigepraktiken, über die Ahnungslosigkeit oder Blindheit vieler Menschen angesichts des Ausmaßes sexueller Gewalt und der Folgen, die in vielen Fällen eine Biographie bis ins hohe Alter prägen. Das hier zitierte Zeugnis steht im Kontext von gesellschaftlicher Aufarbeitung, doch der Wunsch, dass „mal was ins Rollen“ komme, kann auch auf Bildungsforschung und –praxis bezogen werden. Bei sexueller Gewalt handelt es sich um ein über lange Zeit tabuisiertes und marginalisiertes Phänomen. Auch Forschung als Teilbereich von Gesellschaft hat sich immer nur in zeitlichen Schüben damit befasst. In Deutschland erhielt die Forschung dazu erst mit der BMBF Förderlinie „Forschung zu sexueller Gewalt gegen Kinder und Jugendliche in pädagogischen Kontexten“ im Handlungsfeld Bildungsforschung größere Aufmerksamkeit. Auch wenn die Erklärbarkeit des Phänomens bisweilen an Grenzen stoßen kann (Reemtsma, 2017), wird davon ausgegangen, dass Prozessen des Anvertrauens von Gewaltbetroffenen (engl. Disclosure) und ihren gesellschaftlichen Dimensionen eine essenzielle Bedeutung zukommt: das Wissen und die Erfahrungen betroffener Menschen sind die Grundlage dafür, dass Betroffenen Möglichkeiten

¹ Siehe Unabhängige Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs, <https://www.aufarbeitungskommission.de/>.

² Sabine Andresen/Marie Demant/Anna Galliker/Luzia Rott, Sexuelle Gewalt in der Familie, Berlin 2021, S. 29.

der Intervention und Hilfe zur Verfügung gestellt und sexuelle Gewalterfahrungen individuell und gesellschaftlich sichtbar und bearbeitbar gemacht werden können. Im vorliegenden Beitrag wird vor diesem Hintergrund im folgenden Kapitel eine kindheitstheoretische Perspektive auf Gewalt in Bildungs- und Erziehungsverhältnissen dargelegt und begründet. Im dritten Teil werden einige grundlegende Hinweise zu Aufarbeitungskommissionen und deren Arbeitsweisen in Deutschland und international gegeben sowie darauf aufbauend im vierten Kapitel die methodische Herangehensweise an das Datenmaterial der (Abschluss-)Berichte unterschiedlicher Aufarbeitungskommissionen besprochen. Im fünften Kapitel werden die Ergebnisse der Analyse vorgestellt und im abschließenden sechsten Kapitel diskutiert.

Kindheitstheoretische Perspektive auf Gewalt und Bildung

Der Beitrag und die darin vorgestellte Analyse basieren auf einer kindheitstheoretischen Annäherung an Gewalt als potenzieller Teil von Bildung und Erziehung. Wie man dem Allgemeinen des Kindes als Mensch und dem Besonderen als Kind gerecht wird, gehört zu den grundlegenden Fragen auch an die Gestaltung von Bildung und Erziehung. Auf die Bedeutung von Gewalt für Bildung, Bildungsprozesse, -institutionen und -konstellationen soll deshalb aus einer kindheitstheoretischen Perspektive geblickt werden. Das Potenzial einer so angelegten Analyse von Gewalt in Erziehungs- und Bildungsverhältnissen auf Basis von Zeugenschaft über eine gewaltvolle Kindheit und Jugend sowie über Erfahrungen mit einer prekären sozialen Positionierung von Kindern soll anhand der systematischen Auswertung von Aufarbeitungsberichten weiter ausgeleuchtet werden (Andresen, 2024).

Zwar ist dem im Jahr 2000 eingeführten Recht auf gewaltfreie Erziehung in Deutschland eine langjährige und kontroverse Debatte vorausgegangen, doch erst seit 2010 zeigt sich eine veränderte öffentliche Aufmerksamkeit für Gewalt in Erziehungs- und Bildungsverhältnissen. Fokussiert wird seither sexueller Kindesmissbrauch, um den strafrechtlichen Begriff hier einzuführen, bzw. sexuelle Gewalt, der Begriff, der im Folgenden verwendet wird. Diese begriffliche Festlegung erfolgt in Anlehnung an Hagemann-White (2016), die betont, dass das Sexuelle der Taten nicht beiseitegelassen werden dürfe, sowie an Baader (2016), die das Zusammenwirken von physischer, psychischer, sexueller, struktureller, sozialer, verbaler und symbolischer Gewalt betont (s.a. Andresen et al., 2021).

Die Aufmerksamkeit für systematische Gewalt und unterlassene Hilfe als Teile von Kindheit und Jugend ist international keineswegs ein neues Phänomen und auch nicht erst seit 2010 erkennbar. Betroffene werden allerdings deutlicher als in den Jahren zuvor gehört und ihre Erfahrungen anerkannt. Aleida Assmann (2014) hat in ihrer erinnerungskulturellen Rahmung dafür eine weiterführende Perspektive eröffnet. Sie diagnostiziert in ihren Arbeiten zu gesellschaftlicher Aufarbeitung und Erinnerung eine ethische Wende. Diese trage dazu bei, den lange Zeit Ohnmächtigen im Rahmen einer Politik der Menschenrechte Macht zu verleihen und ihre Stimmen zu hören (Andresen 2018). Stärker als in der Vergangenheit zeige sich gegenwärtig eine Anerkennung von Betroffenen und Empathie für Unrecht. Geschichten von Betroffenen zuzuhören, trage mit dazu bei, das beharrliche Vergessen zum Schutz von Täter*innen und ihren Netzwerken zu unterwandern. Ein neueres Beispiel dafür ist die wissenschaftliche Aufarbeitung des Wirkens von Akteur*innen der Sozial- und Sexualpädagogik wie Helmut Kentler (Baader et al., 2024).

Die kindheitstheoretische Ausrichtung auf Gewalt als potenziellen Teil von Bildung beruht auf folgenden Annahmen:

- Gewalt ist Bestandteil von Erziehung und Bildung, aber Strategien der Legitimation von Gewalt sind abhängig von Zeit und Raum (Andresen, 2018; 2024).
- Gewalt in Erziehungs- und Bildungsverhältnissen unterliegt unterschiedlichen Modi der Tabuisierung in der Gesellschaft. Davon sind auch die Wissenschaften betroffen (Wright, 2017).
- Anerkennung von Leid durch Gewalt und damit von Unrecht gegenüber Kindern als Kinder in Erziehungs- und Bildungsverhältnissen ist auf Prozesse der Aufarbeitung angewiesen und bedarf des öffentlichen Diskurses (Andresen & Demant, 2017). Ein wichtiger Baustein hierfür sind die veröffentlichten Berichte von Aufarbeitungskommissionen.
- Aufarbeitung von Gewalt in der Kindheit bietet systematisch das Potenzial, die Reichweite von Zeugenschaft, Schweigen brechen und Gehör finden auszuloten (Andresen et al., 2016; Hayner, 2001; Ross, 2003).

Kinder und Jugendliche insgesamt, aber besonders von sexueller Gewalt Betroffene,³ verbindet auch noch als Erwachsene die Erfahrung, dass ihre Erlebnisse,

³ Hier erfolgt keine normative Festlegung auf einen einzigen Begriff. Menschen, denen als Kinder oder Jugendliche sexuelle Gewalt widerfahren ist, verwenden verschiedene Selbstbezeichnungen: Opfer, Überlebende, Betroffene oder auch erfahrungsbetroffene Expert*innen.

ihr Wissen, ihre Deutungen marginalisiert werden. Dieser Sachverhalt, dass man in zentralen gesellschaftlichen Bereichen wie Politik, Arbeitswelt, Forschung, Medien oder Bildung nur bestimmte Ausschnitte zur Kenntnis nimmt, wird als epistemische Ungerechtigkeit thematisiert (Fricker, 2007; Hänel, 2022). Die Benennung von und Kritik an epistemischer Ungerechtigkeit bildet eine Hintergrundfolie auch der Arbeit von Aufarbeitungskommissionen, denn es geht um die Klärung der Fragen, wer eine Stimme hat und gehört wird, wer anerkannt und mit wem verhandelt wird. Es geht rückblickend und vorausschauend aber auch um Erkenntnisse darüber, wem geglaubt wurde und wird und wer über Deutungsmacht verfügt. Diese Aspekte haben eine grundlegende Bedeutung für den Schutz von Kindern und Jugendlichen vor (sexueller) Gewalt. Sie sind relevant für die Einordnung von Alltagswissen, öffentlichen Diskursen, von wissenschaftlichen Erkenntnissen und Praxiserfahrungen gerade auch bezogen auf Erziehung und Bildung.

Barbara Kavemann et al. (2022) haben aufzeigen können, welche Möglichkeiten Betroffene sehen, wenn es um Wege zu Anerkennung von Unrecht geht. Die Vorschläge wie die Einrichtung eines Anerkennungsforums oder die Schaffung eines Gedenkort zielen letztlich auch auf einen veränderten gesellschaftlichen Diskurs, in dem über eine Umverteilung von Macht über die Erzählungen, über die Hervorbringungen von Wissen und Wahrheit, über Handlungsoptionen verhandelt wird.

Zu Ansatz und Auftrag von Aufarbeitung und Zeugenschaft

International hat Aufarbeitung mit den Berichten über massive Gewalt und Ausbeutung in Heimeinrichtungen der katholischen Kirche, durch Geistliche und Ordensschwestern etwa in Irland Bedeutung erlangt. In Deutschland beginnt die fachliche Diskussion über die Notwendigkeit und das Potenzial von Aufarbeitung mit dem Jahr 2010 und der medialen Thematisierung von sexueller Gewalt in zwei Schulen: dem katholischen Canisius-Kolleg und der reformpädagogischen Odenwaldschule. Die Unabhängige Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs (UKASK) in Deutschland nimmt im Jahr 2016 ihre Arbeit auf. Dem voraus gehen mehrere Jahre der Vorbereitung, intensiver Diskussionen zwischen Betroffenen, Politik, Fachleuten unterschiedlicher Professionen und Wissenschaftler*innen im Lichte internationaler Erfahrungen. Das Parlament beauftragt schließlich den damaligen Unabhängigen Beauftragten für Fragen des sexuellen Kindesmissbrauchs (UBSKM) Johannes-Wilhelm Rörig, damit, eine

Aufarbeitungskommission einzusetzen. Der Auftrag an die sieben ehrenamtlich tätigen Mitglieder war und ist anspruchsvoll: Die UKASK soll Strukturen, die sexuellen Kindesmissbrauch ermöglicht und Vertuschung sowie unterlassene Hilfe nicht sanktioniert haben, in allen Bereichen der Gesellschaft aufarbeiten.⁴ Dementsprechend befasst sich die Kommission mit der Familie als Tatkontext ebenso wie mit Anforderungen der Aufarbeitung in pädagogischen Institutionen wie Schulen, mit Aufarbeitung in für Kinder und Jugendliche zuständigen Behörden wie Jugendämtern, mit der Klärung von sexueller Gewalt und dem Umgang damit im Spitzen- und Breitensport oder mit der notwendigen Aufarbeitung in der evangelischen und katholischen Kirche. Darüber hinaus erhält sie angesichts der deutschen Geschichte auf Wunsch des Bundestages den expliziten Auftrag, sich mit sexuellem Kindesmissbrauch in der DDR zu befassen. In einer Studie über sexuellen Kindesmissbrauch in der Familie wurde zudem herausgearbeitet, wie die nationalsozialistische Ideologie Familien in der Nachkriegszeit und insbesondere den Umgang mit Kindern geprägt hat (Andresen et al., 2021; Andresen & Pohling 2023). Angesichts des historischen Zeitraums und der damit verbundenen Komplexität hat die UKASK Fragen der Anerkennung von Leid und Unrecht sowie Formen der Erinnerung und Erinnerungskultur im Blick. Ebenso wird thematisiert, dass sexueller Kindesmissbrauch häufig mit weiteren Gewaltformen einhergeht. Über Recherchen, Werkstattgespräche mit Expert*innen aus unterschiedlichen Bereichen, über Fallstudien und vor allem mittels schriftlicher Berichte wird transparent gemacht, wie die Kommission arbeitet, welche Ergebnisse erzielt werden und welche Fragen offenbleiben (UKASK, 2019).

Die Grundlage dieser, auch die internationalen Untersuchungen begleitenden, zusammenfassenden und/oder abschließenden Berichte sind demnach die Dokumentationen und Auswertungen von Aussagen, Erzählungen und Geschichten von dieser Gewaltform betroffener Menschen weltweit. Im Zentrum der Kommissionsarbeit steht damit das Konzept der Zeugenschaft, durchgeführt werden vertrauliche Anhörungen mit betroffenen Menschen, die sich bei der Kommission melden. Seit 2016 haben sich in Deutschland weit mehr als 2000 Betroffene, aber auch andere Zeitzeug*innen, etwa Lehrkräfte oder Familienangehörige betroffener Menschen, an die deutsche Kommission gewandt. Im Kern basiert die Arbeit der Kommissionen darauf, Betroffenen die Möglichkeit zu geben, ihre Erfahrungen zu artikulieren. Deren Berichte und Erzählungen bilden die Grundlage für Dokumentation, Analyse und Empfehlungen von Kommissionen (Andresen, 2019). Jenseits der Frage nach einer eventuellen Verjährung im Strafrecht, also abseits der Logik von Strafverfahren, wird mit dem Ansatz der Zeugenschaft

⁴ <https://www.aufarbeitungskommission.de/> [Zugriff: 31.8.2023].

eine Perspektive auf sexuelle Gewalt eröffnet, die sowohl ein Verständnis für Komplexität und Tragweite dieser Gewaltform beinhaltet, als auch die Bedeutung von Aufarbeitung über strafrechtliche Aufklärung und individuelle therapeutische Bearbeitung hinaus herausstellt (Andresen & Demant, 2017).

In ihrem Vorgehen schließt die UKASK an die internationalen Kommissionen an⁵ und orientiert sich gerade zu Beginn an den entsprechenden Berichten.⁶ Diese Art einer weltweiten sozialen Bewegung der Aufarbeitung von Unrecht in Kindheit und Jugend hat die Sozialwissenschaftlerin Katie Wright als einzigartige Konstellation beschrieben und mit der Formel “The Age of Inquiry”⁷ charakterisiert. Die internationalen Aufarbeitungskommissionen haben auf unterschiedlichen Wegen und mit unterschiedlichen Befugnissen ausgestattet, Möglichkeiten für Betroffene geschaffen, um Zeugnis abzulegen, ihre Expertise einzubringen und Forderungen für den gesellschaftspolitischen Umgang mit sexueller Gewalt im jeweiligen nationalen Kontext zu artikulieren. Die der Stichprobe zugrundeliegenden internationalen Kommissionen werden im folgenden Kapitel skizziert.

Methodische Herangehensweise und Datenmaterial

Die überwiegende Mehrzahl der internationalen Auseinandersetzungen fokussiert die Aufarbeitung von sexuellem Kindesmissbrauch und der Misshandlung von Kindern und Jugendlichen in institutionellen Einrichtungen aller Art. In den Blick

⁵ Eine Auswahl von internationalen Kommissionen findet sich auf der Homepage der Unabhängigen Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs in Deutschland (<https://www.aufarbeitungskommission.de/kommission/aufarbeitung/aufarbeitung-international>, abgerufen am 2. März 2023). Eine umfassendere Übersicht über Kommissionen und Untersuchungen international bietet die Website <https://www.lib.latrobe.edu.au/research/ageofinquiry/index.html> des Department of Social Inquiry der La Trobe University in Melbourne Australien unter der Überschrift: „The age of inquiry. A global mapping of institutional abuse inquiries“ (abgerufen am 17.3.2023).

⁶ Viele dieser Berichte sind online und in englischer Sprache abrufbar. Allerdings bei weitem nicht alle, der niederländische Bericht ist nicht in englischer Sprache verfügbar, ebenso wie der schwedische Bericht der Commission to Inquire into Child Abuse in Institutions and Foster Homes (2006–2011). Während einige Berichte, hier ist allem der australische zu nennen online problemlos zugänglich sind, sind andere – wenn überhaupt – nur nach intensiver Online Recherche auffindbar.

⁷ Auf der gleichlautenden Homepage:

https://www.lib.latrobe.edu.au/research/ageofinquiry/browse_inquiries.html (abgerufen am 24.3.2023).

kommen insofern vor allem Kinder und Jugendliche in Organisationen wie Schulen, Internaten und Heimen in staatlichen, aber auch religiösen Trägerschaften sowie Pflegefamilien. Vor dem skizzierten Hintergrund ist es das Ziel der für den Beitrag vorgenommenen Analyse, auf der Basis einer Stichprobe von internationalen Berichten, die die Erfahrungen von sexueller Gewalt betroffener Menschen zum Thema machen, einen Einblick in die Frage zu ermöglichen, wie die Bezüge zwischen Bildungs- und Gewaltverhältnissen dargestellt und thematisiert werden. Es werden im Sinne einer diskursiven Skizze Thematisierungsweisen und Deutungsmuster herausgearbeitet, die mit den Schlüsselbegriffen *Bildung*, *Erziehung* und *Gewalt* in einem inhaltlichen Zusammenhang stehen. Herausfordernd an dieser Fragestellung ist dabei nicht nur die Heterogenität und der Umfang der zum Zeitpunkt der Erstellung dieses Beitrags veröffentlichten internationalen Aufarbeitungsberichte der offiziellen Kommissionen der verschiedenen Länder, sondern darüber hinaus die Definitionen und Bedeutungszuschreibungen der zentralen Begriffe *Bildung*, *Erziehung* und *Gewalt* im jeweiligen Land bzw. im englischen Sprachraum.

Menschen, die als Kinder und Jugendliche sexuelle Gewalt und deren Tabuisierung erlebt haben, verfügen über vielfältiges Wissen: sie geben Auskunft über Ermöglichungsräume sexueller Gewalt, über Strategien von Täter*innen ebenso wie über strukturelle Mängel der Versorgung nach erlebter Gewalt, über Diskriminierungserfahrungen, aber auch über erfolgreiche Unterstützungsangebote. Ähnliches gilt für das Wissen über die Ohnmachtsgefühle von Kindern und Jugendlichen, über die vielen gesellschaftlich etablierten und akzeptierten Schweigepraktiken, über die Rolle von mitwissenden Dritten in der Dynamik und im Beziehungsgefüge von sexueller Gewalt, über die Ahnungslosigkeit oder Blindheit vieler Menschen angesichts des Ausmaßes sexualisierter Gewalt und der Folgen, die in vielen Fällen eine Biografie bis ins hohe Alter prägen. „Menschen, die in ihrer Kindheit sexuell missbraucht wurden, berichten nicht nur von Täter*innen, sondern auch von anderen Menschen, die etwas sahen oder wussten, aber nicht eingeschritten sind und den Missbrauch nicht beendet haben. Sie erzählen von Erfahrungen mit Ämtern und Fachkräften, deren Hilfsangebote nicht ausgereicht haben. Sie bezeugen nicht zuletzt das gesellschaftliche Verschweigen von Verbrechen und Vergehen und damit die Aufrechterhaltung eines Tabus.“ (UKASK, 2019, S. 6). Vor diesem Hintergrund stellen die umfangreichen veröffentlichten Berichte und Fallstudien der verschiedenen Kommissionen eine vielschichtige Datenbasis und Wissensressource dar, die für weitere Forschung genutzt werden kann, ohne dass Betroffene dafür erneut befragt werden müssen.

Die methodische Umsetzung der Studie erfolgte im Sinne einer explorativen *Schlüsselwortsuche* und orientiert sich an der Logik einer qualitativen Inhaltsanalyse bzw. an dem Vorgehen der Inhaltlichen Strukturierung nach Philipp Mayring

(1988). Dabei werden anhand von Schlüsselworten, die fortlaufend ergänzt und je nach Text erweitert werden, „Fundstellen“ herausgearbeitet und auf der Grundlage dieser Fundstellen „bestimmte Themen, Inhalte, Aspekte aus dem Material“ (Mayring, 2015, S. 103) extrahiert. Die primären Schlüsselworte, die im Kontext der Fragestellung dieses Beitrags im Fokus der Analyse standen, waren die Begriffe “education“ bzw. “educational”, “school” sowie “violence” bzw. “violence” sowie deren deutsche Entsprechungen „Bildung“, „Schule“ und „Gewalt“. Um diese Schlüsselworte bzw. zentralen Begriffe herum entstand nach und nach eine thematisch fokussierte Skizze, in der auf Grundlage der ersten Fundstellen die Suche auf weiterführende, aber thematisch synonyme Begriffe ausgeweitet wurde. Die so aus dem Material herausgearbeiteten Inhalte konnten dann im breiteren Rahmen der Forschungsfrage systematisiert und diskutiert (ebd.) werden. Das dem Beitrag zugrundeliegende Datenmaterial umfasst Stichproben aus online verfügbaren internationalen Berichten der Aufarbeitungskommissionen d. h. einzelne Bände bzw. Zusammenfassungen der veröffentlichten Berichte aus vier Ländern bzw. Kontinenten: 1. Deutschland, 2. England/Wales, 3. Australien, 4. Kanada. In allen Berichten geht es auf Basis der Zeugnisse von Überlebenden oder Betroffenen auch um sexuelle Gewalterfahrungen.

1. Die Unabhängige Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs (UKASK) hat im Jahr 2016 den gesetzlichen Auftrag angenommen, umfassend in allen gesellschaftlichen Bereichen sämtliche Formen sexueller Gewalt gegen Kinder und Jugendliche in Deutschland zu untersuchen. Dazu zählt z. B. sexuelle Gewalt in Institutionen, in Familien, im sozialen Umfeld, durch Fremdtäter*innen oder im Rahmen von organisierter sexueller Ausbeutung. Mit Institutionen sind alle Einrichtungen gemeint, in denen sich Kinder und Jugendliche während ihres Aufwachsens aufhalten können, z. B. Schule, Sportverein oder Einrichtungen der Kirchen. Sie ist insofern einzigartig, weil sie sich als erste Kommission nicht nur mit institutionell gerahmten sexuellen Kindesmissbrauch befasst, sondern auch mit sexueller Gewalt, die im familiären Kontext stattfand. Der UKASK gehören aktuell sieben ehrenamtlich arbeitende Personen an, die sich seit vielen Jahren für die Aufarbeitung sexuellen Kindesmissbrauchs einsetzen. Sie sind Expert*innen aus unterschiedlichen Fachgebieten wie den Rechts-, und Sozialwissenschaften, der Psychologie, der Medizin und der Politik. Die UKASK veröffentlichte ab 2017 regelmäßig Berichte über ihre Arbeit und veröffentlicht eigens in Auftrag gegebene Fallstudien. Das für den Beitrag herangezogene Datenmaterial umfasst die beiden Bilanzberichte des Jahres 2019 sowie den Tätigkeitsbericht der UKASK für den Zeitraum von 2019 bis 2023.

2. Die englische Kommission “Independent Inquiry into Child Sexual Abuse” (IICSA) mit Sitz in London hat im Jahr 2015 ihre Arbeit aufgenommen. Sie ist eine gesetzlich bestimmte und eingerichtete Kommission, die unabhängig von der Regierung und nicht angegliedert an eine Regierungsverwaltung agiert. Ihr Schwerpunkt ist die Aufarbeitung von sexuellem Missbrauch in Institutionen und Organisationen in England und Wales. Ihre Arbeit gliedert sich in drei Bereiche: das Forschungsprojekt, das Wahrheitsprojekt und das Öffentliche-Hearings-Projekt mit 13 Untersuchungsbereichen. Sie gab im Jahr 2022 den finalen Bericht “The report of the independent inquiry into child sexual abuse” heraus.
3. Die australische “Royal Commission” wurde von der Regierung ins Leben gerufen und untersuchte im Zeitraum von 2013 bis 2017 alle privaten, öffentlichen oder nichtstaatlichen Organisationen, die in der Vergangenheit oder bis heute mit Kindern zu tun hatten bzw. haben, einschließlich Behörden, Schulen, Sportvereine, Kinderheime, Pflegefamilien und religiöse Organisationen. Ihr gehörten sechs Mitglieder an. Sie führte 8013 vertrauliche Anhörungen und 57 öffentliche Anhörungen durch. Im Dezember 2017 legte sie ihren Abschlussbericht in 17 Bänden vor. Der Titel des Berichts lautet “The Final Report of the Royal Commission into Institutional Responses to Child Sexual Abuse”. Die hier vorgenommene Analyse beschränkt sich auf die Bände 1 und 5.
4. Die “Truth and Reconciliation Commission of Canada” (TRC) hatte im Zeitraum von 2009 bis 2015 den Auftrag, die Auswirkungen des staatlichen Internatsschulwesens auf die Kinder der Ureinwohner*innen Kanadas zu untersuchen. Die TRC hörte im Rahmen landesweiter Hearings mehr als 6500 Zeug*innen an und führte sieben nationale Veranstaltungen zur Versöhnung Kanadas mit seiner indigenen Bevölkerung durch. An der Universität von Manitoba in Winnipeg wurde das Nationale Zentrum für Wahrheit und Versöhnung (National Centre for Truth and Reconciliation, NCTR) mit einem Archiv eingerichtet, um die während ihrer Tätigkeit gesammelten Forschungen, Dokumente und Zeugenaussagen aufzubewahren. Im Jahr 2015 veröffentlichte sie u. a. ihren Bericht “Summary of the Final Report of the Truth and Reconciliation Commission of Canada”.

Zum Verhältnis von sexueller Gewalt und Bildung in den Aufarbeitungsberichten

Die Berichte fokussieren unterschiedliche Zeiträume und politisch-gesellschaftliche Strategien des Umgangs mit sexuellem Missbrauch und der Misshandlung von Kindern, gleichwohl kommen Länder in den Blick, die im religiösen Kontext überwiegend christlich geprägt sind. Insbesondere die Berichte aus Kanada und Australien verweisen in diesem Rahmen auf massive Gewalt gegen Kinder und Jugendliche indigener Familien. Die Umerziehung in Schulen bzw. Internaten und Heimeinrichtungen diente der Zerstörung familiärer Bande, kultureller Identität und der jeweiligen Geschichte. Kinder und Jugendliche als kollektiv waren ausgeliefert und es benötigte einen langen gesellschaftlichen Prozess, bevor hierfür Verantwortung übernommen wurde. In der folgenden Analyse der Berichte wurde die Frage bearbeitet, wie sexuelle Gewalt mit den Bildungserfahrungen verbunden thematisiert wird, also welche Bezüge zu pädagogischen Settings in den jeweiligen Aufarbeitungsberichten hergestellt werden.

Sexuelle Gewalt als Teil gewaltvoller Umerziehung in Schulen und der Heimerziehung

Die für die Fragestellung relevanten empirischen Befunde werden in den folgenden Unterkapiteln dargestellt. Im Einzelnen werden dabei Bedeutungsdimensionen sowie thematische Bezüge der zentralen Begriffe „Bildung“ und „Gewalt“ herausgearbeitet und beschrieben, die im Sinne der heuristischen Herangehensweise im Zentrum des analytischen Interesses standen.

Als ein grundlegendes Ergebnis kann zunächst festgehalten werden, dass in allen untersuchten Berichten internationaler Aufarbeitungskommissionen pädagogische Settings bzw. institutionalisierte Bildungskontexte als Ermöglichungsräume für sexuellen Kindesmissbrauch diskutiert und beschrieben werden. Die deutsche Aufarbeitungskommission gliedert die Ergebnisse ihrer Arbeit in verschiedene Themenbereiche, wobei die meisten davon, wie z. B. Schule, Sport, Kirchen auf organisierte bzw. institutionalisierte Rahmungen der berichteten Erfahrungen sexueller Gewalt verweisen (Homepage UKASK). Sie veröffentlichte 2019 einen Bericht, der Empfehlungen für den Umgang mit sexueller Gewalt in institutionellen Settings umfasst (UKASK, 2019). „Institution bedeutet jede private, öffentliche oder nichtstaatliche Organisation, die in der Vergangenheit oder bis heute mit Kindern zu tun hat, zum Beispiel Schulen, Internate,

Kinderheime, Sportvereine, Freizeiteinrichtungen, religiöse Organisationen, Kindertagesstätten, medizinische Einrichtungen oder Behörden.“ (ebd. S. 7). Erziehung und Bildung in ihren institutionalisierten Formen werden als Risikofaktoren für Macht- und Vertrauensmissbrauch und sexuelle Gewalt gegen Kinder thematisiert, was im australischen Bericht in Bezug auf Kinder mit Behinderungen mit dem Begriff der “high-risk institutional settings” (Vol. 1, S. 75) zusammengefasst wird. Im englischen Bericht heißt es dazu: “The Inquiry found evidence of teachers and others exploiting their positions of trust to abuse children in all the various educational settings it considered.” (The report of the independent inquiry into child sexual abuse, S. 13). Im kanadischen Bericht wird die Rolle des Erziehungs- und Bildungssystems als Teil von und Instrument für eine rassistische, repressive und paternalistische sozialpolitische Praxis herausgearbeitet, die als “cultural genocide” charakterisiert wird: “Cultural genocide is the destruction of those structures and practices that allow the group to continue as a group” (Summary of the Final Report of the Truth and Reconciliation Commission of Canada S. 8). “Residential schooling quickly became a central element in the federal government’s Aboriginal policy.” (ebd. S. 3). Das erklärte politische Ziel, das im Jahr 1969 in einem sogenannten “White paper” erneut bekräftigt wurde, war, indigene Kinder in die kanadische Mehrheitsgesellschaft „einzugliedern“. „Third, the schools were seen as engines of cultural and spiritual change: ‘savages’ were to emerge as Christian ‘white men.’ (ebd. S. 58). Vor dem Hintergrund des ideologischen Ziels, Kinder der First Nation zu „christlichen Weißen“ (ebd.) umzuerziehen, kommt die Kommission zu dem Schluss, dass Erziehung und Bildung im eigentlichen Sinn keine Rolle in den untersuchten Einrichtungen spielte: “As educational institutions, the residential schools were failures, and regularly judged as such.” (ebd. S. 71). Zusammenfassend heißt es: “It was, at best, institutionalized child neglect” (ebd. S. 43). Hier wird u. a. die Anfälligkeit von Erziehungs- und Bildungssystemen für ideologischen (Macht)Missbrauch deutlich, indem menschenverachtende und rassistische Denksysteme die Ziele und Werte vorgeben, zu denen *erzogen* werden soll. Unter der Überschrift *Schule* werden im deutschen Bericht die Erzählungen Betroffener in diesem Themenbereich zusammengefasst. Hier wird zunächst darauf verwiesen, dass „Schule (...) für alle Kinder ein zentraler Ort ihres Lebens (ist)“ (UKASK Tätigkeitsbericht, S. 12) und dieser Ort von Betroffenen als einer thematisiert wird, an dessen Rahmen sexuelle Gewalt stattgefunden hat (ebd.). Durch viele Berichte und Anhörungen ziehen sich Hinweise, dass betroffene Kinder und Jugendliche auffälliges Verhalten in der Schule gezeigt haben, diese Zeichen aber von den dort tätigen Pädagog*innen selten richtig verstanden und eingeordnet wurden (ebd.).

In allen untersuchten Berichten werden in diesem Kontext institutionelle Strukturen und Aspekte, wie bspw. die finanzielle und personelle Ausstattung

sowie Fragen nach der Professionalität bzw. Ausbildung von pädagogischen Fachkräften, aber auch von Pflegeeltern, sowie die Frage nach einer staatlichen Kontrolle „von außen“ und der „Kultur“ innerhalb einer Einrichtung als Ermöglichungsstrukturen für Missbrauch, Vernachlässigung und Gewalt diskutiert und dargestellt. Als institutionelle und/oder strukturelle Rahmenbedingungen für Gewalterfahrungen insgesamt, aber sexueller Gewalt im Besonderen, beschreiben Betroffene aus allen drei Kontinenten Erfahrungen absoluter Macht- bzw. Rechtlosigkeit, einen Mangel an Wissen über Sexualität und sexuellen Missbrauch, aber auch eine nicht ausreichende Ernährung und gesundheitliche Versorgung ebenso wie einen Mangel an Aufmerksamkeit, Liebe und Fürsorge in den Einrichtungen, denen sie als Kinder und Jugendliche anvertraut wurden. „Others said that adults responsible for their care had “free rein” in institutions and appeared to be able to do as they pleased with the children“ (ebd. S. 31). Damit einhergehend erschien sexueller Kindesmissbrauch vielen Überlebenden als „normale“ Form des Umgangs, bzw. als Bestandteil der Kultur einer Einrichtung, als wiederkehrender Modus der Bestrafung und/oder auch als Ersatz für Liebe und Zuneigung (ebd. S. 41–42, 55). Im Bericht der australischen Kommission heißt es dazu beispielsweise: „Many survivors described institutions where violence was normalised. These survivors often said that sexual abuse was part of a broader institutional culture of abuse in which sexual, physical and emotional violence between staff and children, and among children, was either condoned or ignored.“ (ebd. Vol. 5, S. 145). „Others felt that as the abuse occurred on a regular basis, it must have been ‘normal’“. (ebd. S. 243).

Übereinstimmend damit wird die gelebte „Kultur“ vieler Einrichtungen als „kalt“ und „beängstigend“ (ebd. S. 33) erinnert. Synonyme, die in den Berichten häufig für die Orte des Aufwachsens der Betroffenen verwendet werden, sind beispielsweise „violent environment“, „violent homes“, „violent places“, „violent nature of the discipline at the schools“ (Summary of the final report, S. 103). In allen Berichten werden demzufolge individuelle Aspekte hervorgehoben, aber vor allem strukturelle Defizite in den Blick genommen, die die Vulnerabilität von Kindern und ihre Wahrscheinlichkeit, Opfer von (sexueller) Gewalt in Institutionen zu werden, zusätzlich erhöhen oder erhöht haben. Diese sind vor allem körperliche und/oder geistige Beeinträchtigungen: „It also noted that children with disabilities were at more risk of child sexual abuse than children who were not disabled.“ (ebd. S. 384). Auch hier werden mangelndes Wissen über Sexualität und über Formen sexuellen Missbrauchs als zusätzliche Risikofaktoren benannt: „Limited knowledge and understanding of child sexual abuse had

made them vulnerable to the abuse“ (Vol 5., S. 103). Und weiterhin: “Many survivors believed their lack of understanding about sex and sexuality as a child contributed to them being vulnerable to sexual abuse.” (ebd. S. 12).

Erfahrungen sexueller Gewalt in Kindheit und Jugend und ihre Auswirkungen im Bildungs- und Erziehungskontext

Der thematische Bezug, in den der Begriff der Bildung hier gestellt wird, ist der, der Folgen bzw. Auswirkungen von sexuellem Kindesmissbrauch in einer längerfristigen biografischen, aber auch gesamtgesellschaftlichen Perspektive. Dieser Bedeutungszusammenhang lässt sich nochmals in zwei thematische Aspekte unterteilen, die sich auf die unterschiedlichen Definitionen von Bildung und Erziehung beziehen.

In einem ersten Bezugspunkt thematisieren Betroffene im internationalen Kontext ihre subjektive Wahrnehmung von sich selbst als Erzieher*in eigener Kinder. Im kanadischen Bericht heißt es dazu: “By removing children from their communities and by subjecting them to strict discipline, religious indoctrination, and a regimented life more akin to life in a prison than a family, residential schools often harmed the subsequent ability of the students to be caring parents.” (Summary of the Final Report of the Truth and Reconciliation Commission of Canada, S. 138). In diesem Sinne leidet das Vertrauen betroffener Menschen in die Fähigkeit, eigene Kinder (gewaltfrei) zu erziehen. In einem anderen Deutungsmuster im Rahmen des thematischen Bezugs zwischen Erziehung und der Erfahrung von (sexueller) Gewalt grenzen sich Betroffene dagegen dezidiert von einer möglichen Wiederholung von in der Kindheit in institutionellem Rahmen gelernten gewaltsamen Erziehungsmustern positiv ab. Das folgende Zitat verdeutlicht diese beiden spezifischen Auslegungen des Themas der Beeinflussung der Erziehung eigener Kinder durch (sexuellen) Kindesmissbrauch: “Many parents told Commissioners that the sexual abuse of their child had affected their confidence in parenting others said they had not repeated the abusive parenting styles learnt in institutions” (The Final Report of the Royal Commission into Institutional Responses to Child Sexual Abuse, 2017, S. 117, 172). Unabhängig von der Frage nach einer positiven oder negativen biografischen Auslegung bleibt also festzuhalten, dass Erfahrungen sexueller Gewalt von Betroffenen als ein Einflussfaktor auf späteres eigenes Erziehungsverhalten, d. h. auf die soziale Interaktion mit eigenen Kindern beschrieben wird (hierzu auch Noll, 2013).

In einem zweiten thematischen Rahmen der Thematisierung von *education* im Sinne des deutschen Bildungsbegriffs werden in den untersuchten internationalen

Berichten die biografischen Bildungserfolge Betroffener als etwas besprochen, das von Erfahrungen sexueller Gewalt in Kindheit und Jugend in Mitleidenschaft gezogen bzw. beeinflusst wird. Im deutschen Bericht heißt es dazu: „Ein großes und sehr bedrückendes Thema ist Armut als eine Folge sexuellen Kindesmissbrauchs und des gesellschaftlichen Umgangs damit. Es besteht längst noch kein Bewusstsein darüber in der Gesellschaft, in welchem Ausmaß sexueller Kindesmissbrauch das spätere Erwerbsleben beeinträchtigt und welche erheblichen sozioökonomischen Einschränkungen damit verbunden sein können.“ (UKASK, 2017, S. 79). Im englischen Bericht wird ebenfalls u. a. auf Bildungsbenachteiligung als eine Folge von Kindesmissbrauch eingegangen, hier heißt es dazu: “The findings indicate that the impacts of child sexual abuse are life long and can severely undermine the mental and physical health of survivors as well as their relationships and educational attainment throughout their adult lives.” (The Final Report of the Royal Commission into Institutional Responses to Child Sexual Abuse, 2017, Vol. 1, S. 236). Hierin wird diesem Aspekt sexueller Gewalterfahrungen ein eigener Abschnitt gewidmet, der unter der Überschrift: “My education suffered” (The report of the independent inquiry into child sexual abuse, 2022, S. 76) zusammengefasst ist. Betroffene gaben demnach häufig an, dass ihre Ausbildung infolge des sexuellen Missbrauchs in der Kindheit erheblich beeinträchtigt wurde: Insgesamt 32 % Prozent der Betroffenen, die im Rahmen der englischen Kommission zu Wort kamen, sprachen über die Auswirkungen des sexuellen Kindesmissbrauchs auf ihre Ausbildung (ebd.). Unabhängig von der Frage, ob die erfahrene sexuelle Gewalt im Kontext der Schule selbst stattfand oder in einem anderen Rahmen, schildern Betroffene Einflussfaktoren auf den Schulbesuch, die unmittelbar mit den Gewalterfahrungen zusammenhängen. So wurden einige von Täter*innen daran gehindert, eine Schule zu besuchen oder eine Ausbildung zu machen. Andere Betroffene, die in der Schule sexuell missbraucht wurden, weigerten sich, dort hinzugehen, begannen zu schwänzen, zu stehlen und/oder exzessiv zu trinken. Viele betroffene Kinder und Jugendliche fühlten sich aufgrund des sexuellen Missbrauchs, den sie erlebten, nicht in der Lage, in der Schule „voll mitzuarbeiten“. 16 % der berichtenden Betroffenen der englischen Kommission berichteten über akademische Schwierigkeiten, und 8 % gaben an, die Schule vorzeitig verlassen zu haben. Betroffene berichten von Konzentrationsschwierigkeiten sowie Prüfungsängsten, die den Schulerfolg maßgeblich beeinträchtigten (ebd. sowie Andresen et al., 2021). Für diejenigen, die in der Schule sexuell missbraucht wurden, scheinen die Auswirkungen besonders akut zu sein (ebd. S. 76). Auch die langfristigen Folgen von sexuellem Kindesmissbrauch im Kontext des Bildungssystems werden im englischen Bericht zum

Thema gemacht: “My education suffered and I didn’t get any qualifications. This then affected my life afterwards and being able to get employment.” (ebd. S. 77).

Dieser Zusammenhang zwischen (sexuellen) Gewalterfahrungen und Schulerfolg und dessen weitere lebenslange biografische Folgen wird auch im Bericht der kanadischen Aufarbeitungskommission betont und herausgearbeitet. Hier sind vor allem Internate im Fokus, in denen Aboriginal Kinder von ihren Eltern und ihrer Kultur getrennt untergebracht und zu christlichen „Weißen“ erzogen werden sollten: “An educational system that degraded Aboriginal culture and subjected students to humiliating discipline must bear a portion of responsibility for the current gap between the educational success of Aboriginal and non-Aboriginal Canadians.” (Summary of the Final Report of the Truth and Reconciliation Commission of Canada, S. 132). In diesen Kontext des mangelnden Bildungserfolges in gewaltförmigen institutionellen Einrichtungen werden auch eine Vielzahl weiterer, daraus resultierender Folgen der erlebten Gewalt gestellt, wie eine hohe Arbeitslosigkeit, Armut, schlechte Wohnverhältnisse, Drogenmissbrauch, Gewalt in der Familie sowie einen schlechten gesundheitlichen Zustand. Diese vieldimensionalen biografischen Auswirkungen sexuellen Kindesmissbrauchs und -misshandlung werden hier nicht nur in einen langfristigen individuellen, sondern auch in einen aktuellen gesamtgesellschaftlichen und sozialen Kontext gestellt: “Poor educational achievement has led to the chronic unemployment or underemployment, poverty, poor housing, substance abuse, family violence, and ill health that many former students of the schools have suffered as adults. Although educational success rates are slowly improving, Aboriginal Canadians still have dramatically lower educational and economic achievements than other Canadians.” (ebd. S. 145).

Gleichwohl die Thematisierung negativer Folgen von sexuellem Kindesmissbrauch in Bildungskontexten überwiegt, wird u. a. im englischen Bericht auch auf quasi positive Auswirkungen eingegangen, indem thematisiert wird, dass Betroffene ihre Ausbildung erfolgreich absolvierten. Dabei verwiesen Betroffene auf ihre Erfahrungen mit sexuellem Missbrauch als Motivation dafür, hart zu arbeiten und etwas erreichen zu wollen, „um dem sexuellen Missbrauch zu entkommen“. So beschrieb eine Betroffene der englischen Aufarbeitungskommission Bildung als „meine rettende Gnade“. (ebd. S. 77). Ähnliche thematischen Bezüge finden sich auch in den Berichten der deutschen Kommission, z. B. hier: „Viele Betroffene beschreiben auch, dass sie ihre persönlichen Kräfte über einen langen Zeitraum abrufen können. Sie schließen mit großer Anstrengung Ausbildungen oder Studium ab und durchschreiten einen erfolgreichen Berufsweg.“ (UKASK 2017, S. 28).

Bildung als Schutzfaktor gegen (sexuelle) Gewalt gegen Kinder und Jugendliche

Auf dieser Bedeutungsebene wird Bildung nicht im Kontext der Ermöglichung oder Folgen von sexuellem Kindesmissbrauch thematisiert, sondern im Gegenteil als Teil der „Lösung“ im Sinne von Prävention und Verhinderung von zukünftiger oder aktueller Gewalt gegen Kinder und als wichtiger Bestandteil von Aufarbeitung. Die deutsche Kommission fasst auf ihrer Homepage den Themenschwerpunkt Schule wie folgt zusammen: „Von sexueller Gewalt betroffene Menschen thematisieren Schule auf unterschiedliche Weise: als möglicher Schutzraum, als Bildungsort aber auch als Tatort.“ (Homepage UKASK).

Im kanadischen Bericht wird dezidiert auf Bildung als Schlüsselfaktor für Aufarbeitungsprozesse eingegangen: “Much of the current state of troubled relations between Aboriginal and non-Aboriginal Canadians is attributable to educational institutions and what they have taught, or failed to teach, over many generations. Despite that history, or, perhaps more correctly, because of its potential, the Commission believes that education is also the key to reconciliation.” (Summary of the Final Report of the Truth and Reconciliation Commission of Canada, 2015, S. 234). Die zum Zeitpunkt der Berichterstellungen bereits existierenden, präventiven Angebote werden vor allem im australischen Bericht der Kommission in durchaus kritischer Art und Weise als Angebote, die zwar vorhanden sind, deren Zugang und Nutzung durch Fachkräfte und Eltern aber als problematisch beschrieben werden, reflektiert: “There are various primary prevention education, training and information resources available within Australia. However, access to and use of these resources is problematic for professionals, parents and community members.” (The Final Report of the Royal Commission into Institutional Responses to Child Sexual Abuse, Vol. 1, S. 242).

In diesem Zusammenhang wird auch angemerkt, dass vor allem Primärprävention gesellschaftlich umstritten bzw. unklar erscheint: “While Australia has various child and parent education programs about child sexual abuse, access to prevention programs for perpetrators appears to focus more on those who have already offended, and there appears to be some confusion and a lack of consensus in Australia about what constitutes the primary prevention of child sexual abuse.” (ebd. Vol. 1, S. 242). In diesem Sinne erhält die Bedeutungsebene der Bildung als Prävention eine paradoxe Dimension, indem Bildung einerseits als zentraler Baustein zur Aufarbeitung von sexueller Gewalt gegen Kinder dargestellt wird, andererseits aber unklar erscheint, ob und wie präventive Angebote diese Funktion bzw. diesen Auftrag erfüllen können.

Vor allem im kanadischen Bericht werden Bildungsangeboten ein umfassender gesellschaftlicher Auftrag zur Veränderung kollektiven Bewusstseins bzw. insgesamt ein hohes Veränderungspotenzial im Kontext von Aufarbeitung zugeschrieben: “But education for reconciliation must do even more. Survivors told us that Canadians must learn about the history and legacy of residential schools in ways that change both minds and hearts.” (Summary of the Final Report of the Truth and Reconciliation Commission of Canada, 2015, S. 234). Um diesen Auftrag zur Veränderung von Gesellschaft und zur Umsetzung des politischen Bildungsprozesses der Aufarbeitung zu erfüllen werden vielfältige Synonyme wie “educational tool”, “educational materials”, “educational role”, “educational reform” oder auch “educational activities” verwendet. Im Tätigkeitsbericht der deutschen Aufarbeitungskommission wird schließlich Schule besonders in einer Rolle als Schutzort thematisiert: Kinder und Jugendliche fanden dort in Lehrkräfte Vertraute, denen sie von sexueller Gewalt berichten konnten, die sie in anderen Zusammenhängen erlebten, zum Beispiel in der Familie.

Diskussion und Fazit

Die untersuchten Berichte von Aufarbeitungskommissionen verdeutlichen das Potenzial, Wissen über sexuelle Gewalt in pädagogischen Institutionen durch Zeugenschaft als Methode und Aufarbeitung als Rahmung hervorzubringen. Eine besonders große Hürde ist für die Überlebenden der Schulen und Heimeinrichtungen in Kanada, England und Australien, auch über die sexuelle Gewalt zu sprechen. Dies bestätigt einen Befund aus der deutschen Aufarbeitung, denn auch Ehemalige der DDR-Heimerziehung haben zunächst nicht über sexuelle Gewalt durch Pädagog*innen gesprochen (Andresen et al., 2021).

Insgesamt werden im Hinblick auf das soziale Phänomen der sexuellen Gewalt Bildung und Bildungseinrichtungen in Aufarbeitungsberichten, die auf Erzählungen betroffener Menschen basieren, vielschichtig und zum Teil ambivalent diskutiert. So wurde deutlich, dass Machtverhältnisse, die Erziehungsprozesse einerseits und Bildungseinrichtungen andererseits prägen, ein großes Potenzial für die Ausübungen von (sexueller) Gewalt gegen Kinder und Jugendliche in sich bergen. So sind pädagogische Räume immer wieder daraufhin zu reflektieren, welche Möglichkeiten des Macht- und Vertrauensmissbrauchs sie bieten und wie Kinder und Jugendliche wirksam davor geschützt werden können. Kinder und Jugendliche vor sexueller Gewalt in Institutionen bzw. an Orten des Aufwachsens zu schützen stellt eine gesamtgesellschaftliche Herausforderung und Aufgabe dar,

die nach wie vor aktuell ist. Die Berichte machen sichtbar, dass erfahrene sexuelle Gewalt den Lebenslauf von Betroffenen auch und gerade im Bildungskontext erheblich geprägt hat und prägt. Zu den Folgen zählen erhebliche Einbußen bei den individuellen Bildungsmöglichkeiten, und zwar über den gesamten Lebenslauf. Auch die Sorge, eigene Kinder nicht erziehen und zu ihrer Bildung beitragen zu können, ist eine auf Bildung bezogene Folge, von der Erwachsene berichten. Zu dieser spezifischen Form transgenerationaler Weitergabe liegen auch unabhängig von den Aufarbeitungskommissionen inzwischen vereinzelt Studien vor (Noll, 2013).

Dies muss in größerem Umfang als bisher von im Bildungsbereich tätigen pädagogischen Fachkräften reflektiert und zur Kenntnis genommen werden. Zwar durchlaufen pädagogische Fachkräfte heute einen auch gesetzlich geregelten Ausbildungsturnus, sodass von einem deutlichen Professionalisierungsschub gesprochen werden kann. Jedoch sind gerade Inhalte zum Thema sexueller Gewalt gegen Kinder und Jugendliche in Aus- und Fortbildung des pädagogischen Personals nach wie vor unterrepräsentiert und/oder nicht verpflichtend zu belegen (Studie Sexuelle Bildung für das Lehramt – SeBiLe, 2020). Auch in großen Studien, die die Bildungserfolge von Schüler*innen messen, sollte in größerem Umfang als bisher der Einfluss von möglichen (sexuellen) Gewalterfahrungen systematisch mitgedacht und mit erhoben werden.

Literatur

- Andresen, S., & Demant, M. (2017). Worin liegt die Verantwortung der Erziehungswissenschaft? Ein Diskussionsbeitrag zur Aufarbeitung sexueller Gewalt in der Erziehungswissenschaft. *Erziehungswissenschaft*, 54(28), 39–49.
- Andresen, S. (2018). Gewalt in der Erziehung als Unrecht thematisieren. Perspektiven aus der Aufarbeitung sexuellen Kindesmissbrauchs. *Zeitschrift für Pädagogik*, 64(1), 6–14.
- Andresen, S. (2019). *Erklärbarkeit von Gewalt im Lichte von gesellschaftlicher Aufarbeitung. Erfahrungen aus der Arbeit der Unabhängigen Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs..* Manuskript zum Vortrag 01.10.2019. Institut für Sozialforschung.
- Andresen, S. (2024). Time as a category in survivors' reports about child sexual abuse: An explorative approach to lifetime abuse. *Journal of Interpersonal Violence*, 39(19–20), 4065–4086.
- Andresen, S., Demant, M., Galliker, A., & Rott, L. (2021). Sexuelle Gewalt in der Familie. Gesellschaftliche Aufarbeitung sexueller Gewalt gegen Kinder und Jugendliche von 1945 bis in die Gegenwart. <https://www.aufarbeitungskommission.de/mediathek/studie-sexuelle-gewalt-in-der-familie/>
- Andresen, S., & Künstler, S. (2015). Vulnerabilität und sexuelle Gewalt in der Kindheit. Herausforderungen der Kindheitsforschung. *Zeitschrift für Sexualforschung*, 28, 318–334.

- Andresen, S., König, J., & Künstler, S. (2016). Anhörungen von Zeitzug_innen und ihre Bedeutung für die Aufarbeitung sexueller Gewalt. Erziehungs- und kindheitstheoretische Perspektiven. Hearings and their meaning for the process of coming to terms with sexual violence. Perspectives from educational and childhood theory. *Zeitschrift für Pädagogik* 62 (2016) 5, 624–637.
- Andresen S., & Pohling, A. (2023). Perspektiven einer erziehungswissenschaftlichen Biographieforschung auf sexuelle Gewalt gegen Kinder in Familien. In D. Nittel, H. von Felten, & M. Mendel (Hrsg.), *Handbuch Erziehungswissenschaftliche Biographieforschung und Biographiearbeit* (S. 390–403). Beltz Juventa.
- Assmann, A. (2014). *Der lange Schatten der Vergangenheit: Erinnerungskultur und Geschichtspolitik*. Beck.
- Baader, M. S. (2016). History and gender matters. Erziehung – Gewalt – Sexualität in der Moderne in geschlechtergeschichtlicher Perspektive. In C. Mahs, B. Rendtorff, & T. V. Riese (Hrsg.), *Erziehung, Gewalt, Sexualität. Zum Verhältnis von Geschlecht und Gewalt in Erziehung und Bildung. Schriftenreihe der Sektion Frauen- und Geschlechterforschung in der Deutschen Gesellschaft für Erziehungswissenschaft* (S. 13–36). Barbara Budrich.
- Baader, M. S., Böttcher, N., Ehлке, C., Oppermann, C., Schröder, J., & Schröder, W. (Hrsg.). (2024). *Ergebnisbericht. Helmut Kentlers Wirken in der Berliner Kinder- und Jugendhilfe – Aufarbeitung der organisationalen Verfahren und Verantwortung des Berliner Landesjugendamtes*. Universitätsverlag Hildesheim.
- Bange, D. (2016). Geschichte der Erforschung von sexueller Gewalt im deutschsprachigen Raum unter methodischer Perspektive. In: C. Helfferich, B. Kavemann, & H. Kindler (Hrsg.), *Forschungsmanual Gewalt. Grundlagen der empirischen Erhebung von Gewalt in Paarbeziehungen und sexueller Gewalt* (S. 33–46). Springer VS.
- Fricker, M. (Hrsg.). (2007). *Epistemic Injustice: Power and the Ethics of Knowledge*. Oxford University Press.
- Hagemann-White, C. (2016). Grundbegriffe und Fragen der Ethik bei der Forschung über Gewalt im Geschlechterverhältnis. In: C. Helfferich, B. Kavemann, & H. Kindler (Hrsg.), *Forschungsmanual Gewalt. Grundlagen der empirischen Erhebung von Gewalt in Paarbeziehungen und sexueller Gewalt* (S. 13–30). Springer VS.
- Hänel, C. H. (2022). Einleitung: Epistemische Ungerechtigkeit. *Zeitschrift für Praktische Philosophie*, 9(1), 141–154.
- Hayner, P. B. (2001). *Unspeakable truths: Confronting state terror and atrocity. How truth commissions around the world are challenging the past and shaping the future*. Routledge.
- Independent Inquiry Child Sexual Abuse (2022). *The Report of the Independent Inquiry into Child Sexual Abuse*. Executive Summary. <https://www.iicsa.org.uk/reports-recommendations/publications/inquiry/final-report.html>. Zugegriffen: 2. Apr. 2023.
- Kavemann, B., Nagel, B., Etsel, A., & Helfferich, C. (2022). *Erwartungen Betroffener von sexueller Gewalt in Kindheit und Jugend an gesellschaftliche Aufarbeitung*. Abschlussbericht. <https://www.aufarbeitungskommission.de/mediathek/wege-zu-mehr-gerechtigkeit-nach-sexueller-gewalt-in-kindheit-und-jugend/>. Zugegriffen: 27. Nov. 2022.
- Krüger, A. K. (2014). *Wahrheitskommissionen: Die globale Verbreitung eines kulturellen Modells*. Campus-Verlag.
- Mayring, P. (Hrsg.) (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Beltz.

- Noll, M. (Hrsg.). (2013). *Sexuelle Gewalt und Erziehung. Auswirkungen familialer Erfahrungen auf die Mutter-Kind-Beziehungen*. Budrich UniPress.
- Reemtma, J. P. (2017). Erklärungsbegehren. *Mittelweg* 36, 26(3), 74–103.
- Ross, F. C. (2003). *Bearing witness: Women and the truth and reconciliation commission in South Africa*. London: Pluto Press.
- Royal Commission into Institutional Responses to Child Sexual Abuse (2017). *Final Report. Volume 1. Our Inquiry*. <https://www.childabuseroyalcommission.gov.au/our-inquiry>. Zugegriffen: 2. Apr. 2023.
- Runder Tisch Heimerziehung in den 50er und 60er Jahren. (2010). Abschlussbericht. https://www.agj.de/fileadmin/files/publikationen/RTH_Abschlussbericht.pdf. Zugegriffen: 7. Okt. 2024.
- The Truth and Reconciliation Commission of Canada (2015). Honouring the Truth, Reconciling for the Future. Summary of the Final Report of the Truth and Reconciliation Commission of Canada. <https://nctr.ca/records/reports/#trc-reports>. Zugegriffen: 2. Apr. 2023.
- Unabhängige Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs (2023). *Tätigkeitsbericht*. <https://taetigkeitsbericht-2023.aufarbeitungskommission.de/ihre-themen/schule>.
- Unabhängige Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs (2019). *Rechte und Pflichten: Aufarbeitungsprozesse in Institutionen. Empfehlungen zur Aufarbeitung sexuellen Kindesmissbrauchs*. https://www.aufarbeitungskommission.de/wp-content/uploads/Empfehlungen-Aufarbeitung-sexuellen-Kindesmissbrauchs_Aufarbeitungskommission-2020.pdf.
- Unabhängige Kommission zur Aufarbeitung sexuellen Kindesmissbrauchs (2017). *Geschichten, die zählen. Zwischenbericht*. https://www.aufarbeitungskommission.de/wp-content/uploads/2017/06/Zwischenbericht_Aufarbeitungskommission_Juni_2017.pdf
- Wright, K. (2017). Remaking collective knowledge: An analysis of the complex and multiple effects of inquiries into historical institutional child abuse. *Child Abuse & Neglect*, 74, 10–22.
- Wright, K., Swain, S., & Sköld, J. (2020). *The Age of Inquiry: A global mapping of institutional abuse inquiries* (2. Aufl.). Melbourne: La Trobe University.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Teil II
Die Rolle von Lehrkräften



Preschool Teachers' Language Scaffolding in Early Science Education in Linguistically Heterogeneous Groups

5

Ada Cecil Haen, Ilonca Hardy, Henrik Saalbach,
Miriam Leuchter and Mirjam Steffensky

Abstract

Research suggests that domain-specific learning environments with language scaffolding support young children's development of both language and conceptual knowledge. Specifically, the types of questions asked by preschool

A. C. Haen (✉)

Fakultät für Erziehungswissenschaft, Universität Hamburg, Hamburg, Deutschland

E-Mail: ada.haen@uni-hamburg.de

I. Hardy

Goethe-Universität Frankfurt am Main, Frankfurt am Main, Deutschland

E-Mail: hardy@em.uni-frankfurt.de

IDeA – Center for Research on Individual Development and Adaptive Education of Children at Risk, Frankfurt am Main, Germany

H. Saalbach

Universität Leipzig, Leipzig, Deutschland

E-Mail: henrik.saalbach@uni-leipzig.de

M. Leuchter

RPTU Kaiserslautern-Landau, Landau, Deutschland

E-Mail: miriam.leuchter@rptu.de

M. Steffensky

Universität Hamburg, Hamburg, Deutschland

E-Mail: mirjam.steffensky@uni-hamburg.de

© Der/die Autor(en) 2025

J. Ziehm-Eicher et al. (Hrsg.), *Kinder mit erhöhtem Risiko für Bildungsmisserfolg*,
Edition ZfE 20, https://doi.org/10.1007/978-3-658-48674-7_5

129

teachers are considered an important aspect of language scaffolding, especially for children at risk. This study adapted a scoring system for preschool teachers' questions in early science education and gathered evidence of its validity. Additionally, it analyzed differences in teachers' language scaffolding among groups with high or low proportions of children at risk of delayed language development in 64 video-recorded preschool science activities ($M = 21.34$ min). The groups were compared based on preschool teachers' use of questions and quality ratings using the CLASS Pre-K scoring system. Furthermore, the verbal participation of both the preschool teacher and the children was analyzed. The results show significant group differences in teachers' use of different question types, particularly in the number of language-stimulating explanation and justification questions. However, the latter types of questions were rarely asked by the teachers. The quality of the CLASS dimension Language Modeling and the preschool teacher's dominance of verbal participation in the interactions did not differ. Still, a relation was found between the extent of the children's verbal participation and their risk of delayed language development. The analyses aim to approach the complex topic of assessing the adaptivity of strategies to the language prerequisites at the group level. They provide insights into language scaffolding within early science education.

Keywords

Early childhood education • Language scaffolding • Classroom Assessment Scoring System • German preschool setting • Children at risk

Introduction

Linguistic diversity of children in (pre-)schools is increasing in many countries (OECD, 2019). Children with low language skills are likely to be at risk of subsequent academic problems (Mullis et al., 2017; National Academies of Sciences, Engineering, and Medicine 2017). Supporting children's opportunities to gain (majority) language skills is therefore a primary focus of early childhood education (ECE). To support early language development, a linguistically rich environment where preschool teachers engage children in frequent and meaningful verbal interactions plays a crucial role in ECE. In these types of learning environments, language takes on a threefold role: It is a prerequisite for learning, an object of learning, and a medium of instruction, as noted by Prediger et al. (2018). This multifaceted nature of language underlines its integral position in

developmental and educational processes, as demonstrated by a body of research (e.g., Hoff, 2006; Jurkic et al., 2023; Justice et al., 2008; Kempert et al., 2019). The crucial role of language can also be seen in domain-specific studies examining the relationships between the development of language skills and, for example, science concepts (Haug & Ødegaard, 2014; Henrichs & Leseman, 2014). We use the term *language scaffolding* to refer to the instructional strategy of supporting young children's domain-specific learning through enriched language use by teachers within children's zone of proximal development (Gibbons, 2015; Vygotsky, 1978). Language scaffolding strategies, such as asking questions, have rarely been considered across different domains of ECE, including science education. Given that preschools predominantly involve group interactions rather than one-to-one interactions, it is pivotal to examine preschool teachers' scaffolding strategies at a group level, thus taking into account group heterogeneity. In line with this, Hormann and Skowronek (2019) propose that assessment of the adaptivity of language scaffolding strategies is achieved by considering differences in the use of strategies at the group level.

Against this background, this study pursues two central aims. First, to adapt and validate a scoring system that assesses the frequency and types of questions implemented by preschool teachers in early science education. Second, to investigate whether differences in preschool teachers' language scaffolding can be identified in relation to groups with a high or low proportion of children at risk of delayed language development.

Language Scaffolding

Learning and development, according to a co-constructivist understanding, are embedded in social interaction and can be understood as a consequence of it (Callanan & Valle, 2008; Vygotsky, 1978). Wood et al. (1976) highlight the role of competent conversation partners in problem-solving activities within social interaction. The more competent individual can provide *scaffolding* to support the less competent person, usually a child, in completing tasks that they would not be able to complete on their own. In institutional contexts, this role is typically taken by teachers or other adult experts, whereas in the home environment, it is typically taken by parents. In recent work, the construct of contingent support in scaffolding is highlighted, referring to support that fits the individual learning prerequisites of a child (van de Pol et al., 2010). While scaffolding may serve different functions (van de Pol et al., 2010), it is especially the function of supporting cognitive development that has been highlighted. Scaffolding may be used

to provide structure during problem-solving, thus reducing the complexity of the task, but it can also serve to stimulate a child's thinking (Saalbach & Kempert, 2022). Studhalter et al. (2021, p. 3) list three scaffolding strategies teachers can use to stimulate children's thinking in an early science context: "activating prior knowledge, asking open-ended questions, or causing cognitive conflicts".

Based on this initial concept of scaffolding, the construct has been extended to language scaffolding in the context of multilingual students' content and language learning (Gibbons, 2015). Following this tradition, we use the term *language scaffolding* when referring to teacher strategies intended to support children's acquisition of language within contexts of domain-specific learning. As posited by Gibbons (2015), language scaffolding can be conceptualized as an approach that integrates language- and content-related learning objectives throughout the planning and implementation phases of instruction (Gabler et al., 2024). Studies show that learning environments with language scaffolding contribute to the acquisition of vocabulary but also to the development of children's conceptual knowledge (Hardy et al., 2019a; Henrichs & Leseman, 2014; Herrmann et al., 2021; Siraj-Blatchford et al., 2002; for an overview, see Kempert et al., 2019). Within the Vygotskian tradition, language scaffolding refers to the application of a wide range of strategies to support children's development of language and conceptual knowledge (e.g., Justice et al., 2018; Whorral & Cabell, 2016). For example, Beckerle (2017) systematizes the various terms and definitions and distinguishes four central strategies of language scaffolding: *modeling techniques*, *corrective feedback*, *stimulating techniques*, and *parallel talk*. These strategies can be used by preschool teachers either in reaction to a child statement or as an initiation to elicit children's verbal participation (Beckerle et al., 2020). For example, teachers may act as language models by taking up a child's statement and extending it linguistically and in terms of content. They may also use corrective feedback by repeating children's statements in a linguistically corrected way. Furthermore, they may employ strategies for promoting language development, including asking questions, making comments, using parallel talk, or providing guidance for actions, concept development, or reasoning (Beckerle et al., 2018). In the preschool context, stimulation techniques, including asking questions, are prevalent (Beckerle et al., 2020).

Previous studies on the implementation of language scaffolding strategies in preschool and primary school settings have shown inconsistent results both in terms of the frequency with which these strategies are used and their adaptation to children's language skills (e.g., Beckerle & Mackowiak, 2019; Elstrodt et al., 2019; Geyer & Müller, 2021; Hormann & Skowronek, 2019; Müller et al., 2016; Schmidt et al., 2019). Nevertheless, there is no consensus on which strategies are

most appropriate for children with different prerequisites. For instance, Hormann and Skowronek (2019) describe the adaptive use of language scaffolding strategies as follows: while strategies such as parallel talk and closed-ended questions may prove beneficial for children with limited language skills, open-ended questions are more suited to children with advanced skills. Other strategies, such as corrective feedback or modeling techniques, were classified in a middle category. Beckerle and Mackowiak (2019) observed that preschool teachers prioritized the clarification of incomprehensible child statements and the reinforcement of appropriate vocabulary understanding and usage for children with lower language skills. For children with higher language skills, teachers employ labeling questions, other extensions, and corrective feedback.

Research on teachers' scaffolding behavior has been conducted at both the group level (e.g., Hormann & Skowronek, 2019) and the individual level (e.g., Beckerle & Mackowiak, 2019). While pioneering studies on scaffolding, such as the one by Wood et al. (1976), focused on contingency and adaptivity in interactions between a more competent adult (e.g., a parent) and a child, the assessment of scaffolding strategies by teachers is often conducted at the group level (e.g., Smit et al., 2013), as this approach is considered especially relevant in institutional settings. For example, Herrmann et al. (2021) examined the impact of teacher scaffolding in primary school science on academic performance in classes with high versus low proportions of multilingual students. Similarly, Studhalter et al. (2021, p. 1) investigated selected "scaffolding utterances" in preschool classes to explore the influence of teacher talk on children's learning in the context of science.

Building on these studies, the present study examines differences in preschool teachers' language scaffolding depending on group composition in terms of children at risk of delayed language development. Particularly, we focus on the language scaffolding strategy of asking questions.

Asking Questions

Preschool teachers' questions, as a strategy of language scaffolding, are intended to stimulate language production and are conceived to have high potential for supporting the development of conceptual knowledge (Lee & Kinzie, 2012). With questions, children are prompted to use and extend their vocabulary, phrases, and grammatical structures in verbalizations of their ideas (Kucharz, 2018; Lee & Kinzie, 2012; Whorrall & Cabell, 2016). Recent findings from the German and international context show that preschool teachers ask questions more frequently

in structured, domain-specific learning environments than during play situations or meals (e.g., Whorrall & Cabell, 2016; Wildgruber et al., 2016). An explanation for this may be the planning of activities in domain-specific, structured learning environments, leading teachers to also plan their questions in advance (Kucharz, 2018). In early science contexts, in particular, a variety of opportunities arise for teachers to stimulate children's use of language by questioning. Within scientific inquiry, children may be stimulated to describe their observations of phenomena, verbalize their ideas, and formulate hypotheses, explanations, and conclusions (Peterson & French, 2008).

Different types of questions have been distinguished. A central distinction is made between closed- and open-ended questions (e.g., Lee & Kinzie, 2012). Closed-ended questions include decision questions, i.e. yes/no questions or questions with alternatives. Some authors also categorize questions as closed-ended if they can be answered with one word or if they are highly constrained, allowing only a limited range of answers (e.g., Hamel et al., 2021; Lee & Kinzie, 2012). Questions that allow for a wider range of language activity are considered open-ended (Walsh & Hodge, 2018). Schmidt et al. (2019) classify questions according to their intention: 1) decision questions (i.e., simple yes/no answers or questions with alternative choices), 2) labeling questions (i.e., labeling of objects or concepts), 3) questions for description (i.e., detailed answers about actions or feelings), and 4) questions for explanation (i.e., complex answers about concepts or problem-solving). In this categorization, the linguistic and cognitive demands on children increase with the complexity of the questions. Some of these questions are called partially open-ended questions (e.g., Beckerle et al., 2020; Schmidt et al., 2019). In a further categorization based on Bloom's Taxonomy of Educational Objectives (Bloom et al., 1956), Gabler et al. (2024) describe lower-order and higher-order thinking questions that relate to the cognitive demand level of the questions. Lower-order questions focus on recalling facts, whereas higher-order questions are oriented towards application and possess an analytical, inferential, and evaluative nature. This distinction is also addressed by Tompkins et al. (2013). Gabler et al. (2024) developed a coding scheme for teacher questioning in primary school science education. It comprises yes/no and alternative questions as closed-ended questions. The differentiation of open-ended questions concerns the degree to which these questions elicit the use of domain-specific language, e.g., when the teacher explicitly asks for the term of an object or activity. Similar to Schmidt et al. (2019), a distinction is made between questions for descriptions and complex explanation and reasoning questions. In contrast to the aforementioned approaches, Voet Cornelli (2022) employs a linguistic perspective, classifying questions that stimulate children's language production based on the syntactic

structure of verb position in the German language. For instance, questions in which the verb occurs in the first position, such as those concerning decisions or alternatives, typically elicit brief answers or yes/no answers. In contrast, questions with the verb in second position allow for a range of potential answers, often taking the form of *wh*-questions. Voet Cornelli further describes questions that elicit verb-containing answers from children, which are particularly complex and well-suited for language diagnostic purposes.

The use of questions as prompts to elicit children's talk is based on the Vygotskian notion of an interplay of language and thought. Labeling questions, for instance, can activate children's prior knowledge and promote understanding of domain-specific vocabulary. They may be important for language acquisition of children with limited language skills, for example to elaborate the meaning of a word or to request a specific expression and vocabulary item (Gabler et al., 2024). Questions that elicit explanations, comparisons, and justifications can stimulate more detailed and complex answers and are also important scaffolding strategies for cognitive engagement (e.g., Saalbach & Kempert, 2022; Studhalter et al., 2021). Questioning is also an integral part of the concept of Sustained Shared Thinking. The concept focuses on cognitive activating strategies, such as questioning and demonstrating in dialogue between adults and children and is considered an indicator of high-quality preschool learning environments (Siraj-Blatchford, 2009). Although in this regard, open-ended questions are especially stimulating, studies indicate that closed-ended questions are prevalent in the (pre-)school context (e.g., Schulte, 2020; Wirts & Cordes, 2021). For example, the REPEY project found that only 5.5 % of the questions posed by preschool teachers had the potential to cognitively engage (Siraj-Blatchford & Manni, 2008). Studies in the context of science education have reached similar conclusions (e.g., Lee & Kinzie, 2012). Hamel et al. (2021) analyzed the questions asked by 14 preschool teachers in a science context. The questions were categorized as closed- or open-ended, and scientific or non-scientific. Only 20 % of the questions were open-ended, meaning they did not restrict children's answers or lead to the reproduction of already known answers.

Different types of questions can influence both the quantity, such as the length of children's answers, and the quality, such as the linguistic complexity or the use of domain-specific vocabulary. Research shows that open-ended questions tend to elicit longer and more detailed answers, as well as more complex and varied vocabulary and sentence structures (Lee & Kinzie, 2012; Röhner et al., 2010; Wirts & Cordes, 2021). In a video study with 31 preschool teachers, Hormann and Skowronek (2019) found a negative correlation between the use of closed-ended question types (including yes/no and alternative questions) and the length (word

count) of children's answers. The importance of language-stimulating questions can also be seen in terms of children's participation in discourse and its relation to learning gains (Sedova et al., 2019). A study by Mannel et al. (2016) found a correlation between active language production and the development of conceptual knowledge of children at risk during preschool science activities. Furthermore, Bürgermeister et al. (2019) showed that scaffolding strategies are especially effective when children participate in verbal interactions. In this regard, closed-ended questions may also contribute to maintaining participation as they are associated with a higher number of turns (Hormann & Skowronek, 2019). However, descriptive analyses examining verbal participation in the preschool context indicate that teachers dominate discourse during activities such as picture book observations (e.g., Wirts & Cordes, 2021), science activities (Bürgermeister et al., 2019), or playtime (e.g., Tompkins et al., 2013).

Assessment of Language Scaffolding Strategies in ECE

Different methodological approaches to the assessment of language scaffolding strategies can be identified in the literature. A common approach is the use of high-inference ratings focusing on the quality of teacher-student interactions, such as the SSTEW (Howard et al., 2018) or the Classroom Assessment Scoring System (CLASS Pre-K, Pianta et al., 2008). The latter is used in the current study and described in more detail below.

The CLASS Pre-K scoring system includes three domains of interaction quality: Emotional Support, Classroom Management, and Instructional Support. Language Modeling is one aspect of interaction quality within the Instructional Support domain. Mashburn et al. (2008) discovered a significant positive correlation between the Instructional Support domain, which includes (language) scaffolding strategies, and the development of language as well as conceptual knowledge in a sample of approximately 2000 children. Several studies show that the CLASS is also applicable in the German preschool context, especially in structured educational settings (Oppermann et al., 2024; Stuck et al., 2016; Sucho-doletz et al., 2014). In terms of assessment, Instructional Support is divided into three dimensions: Concept Development, Quality of Feedback, and Language Modeling. Concept Development involves promoting higher-order thinking and can be seen in the context of scaffolding. Quality of Feedback aims to expand learning processes and participation. Language Modeling includes several language scaffolding strategies such as the frequency of conversation, the repetition and extension of children's statements, parallel talk, the use of a variety of

vocabulary and the linking of new words with familiar ones. However, the high-inference rating of Language Modeling does not provide further information on which specific strategies were used. Moreover, Language Modeling involves the use of open-ended questions, which differ from closed-ended questions in eliciting more than one-word answers from children. Thus, the CLASS scoring also considers the length of the children's answers (Pianta et al., 2008).

A different approach to assessing language scaffolding is a turn-by-turn analysis (Hardy et al., 2019a). The discourse can thus be differentiated at the level of teachers' individual language scaffolding strategies. Beckerle et al. (2020), for example, developed a scoring system for assessing language scaffolding strategies in pre- and primary schools to analyze teachers' corrective feedback, modeling, and stimulation strategies. They distinguish between twelve subcategories of strategies and reactive and initiation language scaffolding strategies. In light of the potential of questions to stimulate language production, a focus on asking questions as a specific scaffolding strategy appears to be promising, especially for children with low language skills. From a linguistic perspective, an examination of syntactic structures offers insight into the language-stimulating potential of questions, which can be utilized for the purpose of language diagnostics (Voet Cornelli, 2022). A cognitive perspective is emphasized when distinguishing between lower- and higher-order questions. The latter aims to stimulate thinking processes in children, which is crucial for the development of domain-specific knowledge. Lower-order or closed-ended questions that elicit short answers help to maintain discourse and support children with limited language skills by expanding their understanding of (domain-specific) vocabulary. Nevertheless, they may limit children's language production, both in terms of quantity and complexity, and promote less complex thinking processes (Gabler et al., 2024; Siraj-Blatchford, 2009). In contrast, open-ended questions promote longer, more complex answers, encouraging children to verbalize their thoughts and expand their language skills. However, they can be cognitively and linguistically challenging, especially for children with limited language abilities.

Rationale and Aims of the Present Study

To summarize, teachers' questions have great potential for stimulating children's verbal participation and the development of domain-specific conceptual knowledge. For an adaptive use of questions, the characteristics of the children must be considered. While there is evidence that science activities offer numerous opportunities for (language) scaffolding, little research has been conducted

on preschool teachers' language scaffolding and their use of questions in early science settings.

In order to analyze preschool teachers' questions, it is necessary to use assessment tools that precisely describe the different types of questions, which vary in complexity in terms of language and content. Therefore, we adapted a scoring system for questions (SSQ) in early science education in preschool (Gabler et al., 2024; Schmidt et al., 2019) which goes beyond dichotomous distinctions between closed-ended and open-ended questions or between scientific and non-scientific questions (Hamel et al., 2021). The first aim of the present study is to analyze the relationship between the SSQ and the Language Modeling subscale from the well-established CLASS assessment system. The objective is to assess whether teachers with higher ratings in the dimension of Language Modeling also tend to ask more questions, thus indicating consistency between the high- and low-inference systems and providing initial evidence for the validity of the SSQ. In addition, the length (operationalized by the word count) of children's answers to different types of questions is computed in order to validate the different potential of questions for stimulating children's verbal participation. The focus is therefore on the quantitative aspects of children's answers (i.e., length), rather than on the quality of the statements (e.g., linguistic complexity or domain-specific vocabulary).

The second aim of this study is to investigate to what degree preschool teachers adapt their language scaffolding in science activities to the linguistic composition of the group of children. This is done by comparing the quality of teachers' Language Modeling and their use of questions between groups with high and low proportions of children with low language skills and therefore at risk of delayed language development. Given the crucial role of children's active verbal participation for the effectiveness of scaffolding (Bürgermeister et al., 2019), and in alignment with the approach proposed by Herrmann et al. (2021), we assess the proportion of discourse between children and teachers at the group level in different group compositions, using the index of word count (e.g., Wirts & Cordes, 2021).

Following these two aims, the present study may provide further insights into teachers' language scaffolding in early science settings with linguistically heterogeneous groups.

The following research questions and hypotheses will be addressed:

1. Validation of the Scoring System for Questions (SSQ):
 - a) What is the relation between ratings of teachers' questions (SSQ) and Language Modeling (CLASS)?

Since asking questions is an indicator in the CLASS dimension of Language Modeling, we assume that teachers in groups with high-quality Language Modeling will ask more questions in the different question types.

- b) What is the relation between different types of questions (SSQ) and the word count in children's answers?

We expect that especially questions asking for descriptions and explanations will elicit a significantly higher word count in answers than other question types.

2. Do preschool teachers in groups with high proportions of children at risk of delayed language development differ from teachers in groups with low proportions in the degree of Language Modeling and teacher questions? What are the differences in teachers' and children's degrees of verbal participation?

Findings on preschool teachers' adapted language scaffolding are inconsistent (Elstrodt et al., 2019; Geyer & Müller, 2021; Hormann & Skowronek, 2019; Müller et al., 2016; Schmidt et al., 2019). However, we expect differences between the two groups in the quality of Language Modeling and teacher questions. Based on initial findings (e.g., Bürgermeister et al., 2019; Herrmann et al., 2021), we further expect teachers in groups with a high proportion of children at risk of delayed language development to talk more and thus be more verbally dominant than teachers in groups with a low proportion of children at risk.

Methods

Sample

64 video sequences and resulting transcripts from the study *ProfinK*¹ were analyzed. The duration of each analyzed sequence was approximately 21 minutes ($M = 21.34$ minutes, $SD = 4.00$). The videos showed a preschool teacher interacting with a group of children ($N = 2$ to 7 children; $M = 4.30$; $SD = 0.93$) on the topic of 'materials and their properties'. The 64 preschool teachers ($M = 41.15$ years, $SD = 10.49$; female: 76.6 %) had worked in preschool settings for an average of 12.40 years ($SD = 10.48$). They took part in a professional development program on formative assessment in early science education as part of the ProfinK study. All videos were recorded before the start of the professional development.

¹ ProfinK—Professional learning communities for preschool teachers to support their ability to use diagnostic information for instructional decisions.

The study participants, referred to as preschool teachers for consistency, came from educational institutions in four medium to large German cities. In this sample, 41 % of participants were trained as preschool teachers, and 19 % have a degree in social/educational studies, 8 % have other educational or social training, 31 % either have other training or no information was given, and 2 % have no vocational training.

The total number of children in the sample was 271 ($M = 4.53$ years, $SD = 0.58$; female: 46 %). Language background was assessed using parental questionnaires. 16 children in the sample were dual-language learners, acquiring German as a second language; 162 were monolingual. For 93 children, information on their language background was not available. For 74 of the 271 children, we identified a risk of delayed language development using the Wechsler Preschool and Primary Scale of Intelligence-IV (Wechsler, 2018) vocabulary acquisition index score (see section *Assessment of children's language skills*).

Instruments

Classroom Assessment Scoring System—CLASS Pre-K

Interaction quality was assessed using the CLASS Pre-K (Pianta et al., 2008). The rating is on a scale of 1 to 7, with 1 and 2 corresponding to low quality, 3, 4, and 5 to mid-range of quality, and 6 and 7 to high quality. The videos in this study refer to exactly one cycle according to CLASS (approximately 20 minutes). They were rated by two raters with a valid certification; 38 percent of the videos in the sample were double-coded for interrater agreement. The remaining videos were rated by the first author. The interrater agreement was acceptable for all dimensions (Intraclass Correlation Coefficient, $ICC = .89$, two-way random, single measure, absolute agreement). The focus of this study is on the domain of Instructional Support, specifically on the dimension of Language Modeling ($ICC = .65$).

Scoring System for Questions (SSQ)

Although CLASS is a validated system for high-inference ratings on Language Modeling, among other domains, it does not allow for a differentiation between various types of language scaffolding strategies, such as asking questions in teacher-child interactions. Therefore, in this study, the scoring systems by Schmidt et al. (2019) for preschool and by Gabler et al. (2024) for primary school were adapted to categorize the types of questions asked by preschool teachers in an early science context (see Tab. 5.1). Gabler et al. (2024) argue

Tab. 5.1 Scoring System for Questions (SSQ).

Types of questions		Short description	Examples	κ	
Closed-ended questions	Code 0 Question particle	The question may be aimed at confirmation/rejection, more rhetorical in nature.	'okay?', 'right?'	0.77	
	Code 1 Simple decision questions	Teacher takes a child's statement and formulates it as a question. The question is more rhetorical in nature.	Child: 'It feels rough.' Teacher: 'Does it feel rough?'	0.60	
	Code 2 Complex decision questions	Code 2a Closed-ended yes/no	The question aims a simple 'yes' or 'no' answer.	'Does the stick float?', 'Is it smooth?', 'Do you know that?'	0.71
		Code 2b Requests for action in the form of questions	The question elicits an action by the child or the group.	'Can you try if the stick floats?'	0.73
Code 2c Alternative questions		The question provides explicit alternative answers.	'Does it feel smooth or rough?'	0.74	
Open-ended question	Code 3 Simple labeling questions	It is not possible to answer 'yes' or 'no'. The question is aimed at a specific answer, e.g., labeling a term.	'What is this?', 'What shape is this?'	0.77	

(Fortsetzung)

Tab. 5.1 (Fortsetzung)

Types of questions		Short description	Examples	κ
	Code 4 Simple descriptive questions	The question asks for a simple description or allows a range of short answers, e.g., to list the properties of objects.	‘How does it feel?’, ‘And what’s different?’, ‘What else is made of wood here?’	0.65
Particularly language-stimulating questions	Code 5 Describing and process-related questions	The question asks for a description of a process, an explanation, a memory, an observation, a report, etc. (drawing on prior knowledge).	‘What have you done with it?’, ‘And what happens?’, ‘What have you sorted by?’	0.51
	Code 6 Explanation and justification questions	The question is aimed at developing own predictions, justifications, reasoned (non-descriptive) comparisons, conclusions, etc. and eliciting own exploration.	‘Why do you think the magnet is holding the coins there?’	0.69

Note. κ : Cohen’s Kappa

that existing scoring systems either under- or over-differentiate the language-stimulating quality of questions and therefore propose six question types in the context of science education. In our system, we adapted these for the context of early science education. Additionally, the distinction by Schmidt et al. (2019) between labeling and description questions was incorporated. We adapt these to distinguish between the description (labeling) of properties and the description of science-related processes. While Gabler et al. (2024) focus on the elicitation of

academic language in primary school, we emphasized domain-specific vocabulary as an age-appropriate measure of advanced language use (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2022).

The following section provides a detailed description of the questions included in the adapted SSQ. Similar to the CLASS rating, this system distinguishes between closed- and open-ended questions. These closed- and open-ended questions were divided into specific types of questions that differ in their potential to stimulate children's language production and in terms of content complexity. The different types of closed-ended questions (Codes 0–2) encompass those that may require only a brief confirmation or rejection from the children, or in some cases, no verbal answer or other reaction at all (Code 0). Questions classified as simple decision questions (Code 1) are also rhetorical in nature: the teacher picks up the child's previous statement and phrases it as a question to elicit confirmation or rejection from the child. More complex decision questions, classified as Code 2a, require a yes/no answer. Although the linguistic demands of these questions are low, they can be cognitively demanding, for example when children are asked to confirm or reject a prediction about whether an object will float in water. Code 2b questions do not necessarily require a verbal reaction from the child but rather elicit an action. Code 2c includes alternative questions in which children are given explicit choices to facilitate their answer, such as 'Does the stick float or sink?'

Open-ended questions distinguish between labeling questions (Code 3), simple description questions (Code 4), and questions that allow for descriptions (Code 5) or explanation of content or inquiry activities (Code 6) and thus the use of domain-specific vocabulary. The latter questions, Code 5 and 6, require more detailed answers regarding content and language than questions of Codes 3 and 4, which can be answered with one word.

We used MAXQDA 2022 (Release 22.8.0; VERBI Software, 2021) to code the guideline-based transcribed videos according to the SSQ. The analysis included all sentences marked with a question mark as well as indirect questions. Each question was given exactly one code. A subset of 20 % of the questions was double coded, and the results indicated good agreement between the two raters (Cohen's Kappa = 0.76). The first author coded the remaining transcripts. Given the length disparity among the videos, we calculated the number of questions per minute.

Word Count

In the videos, the number of words spoken by the children and teachers was analyzed using MAXQDA 2022 (Release 22.8.0; VERBI Software, 2021) by counting each word and part of a word, excluding filler words (Gabler et al., 2024).

Wechsler Preschool and Primary Scale of Intelligence—IV

Children's language skills were measured using the subtest vocabulary acquisition index score of the *Wechsler Preschool and Primary Scale of Intelligence-IV* (WPPSI, Wechsler, 2018) ($M = 90.15$, $SD = 16.23$, $\text{min} = 45$, $\text{max} = 124$). The vocabulary acquisition index subscale evaluates children's (2;6 to 7;7 years) active vocabulary (picture naming; 35 items) and receptive vocabulary (31 items). The children are presented with the various stimuli, and their answers are evaluated. The subtest ends when three consecutive items are not answered or are answered incorrectly. For reasons of test economy, we only implemented one subtest. We chose this subtest because it is a relevant indicator of language skills. However, empirical evidence shows high correlations between subtests that measure different aspects of language skills (Stephan et al., 2022). An index score of 84 or lower, one standard deviation below the norm sample's mean, indicates a risk of delayed language development, which is referred to as relative 'weaknesses' in the WPPSI manual (Wechsler, 2018, p. 149). The German norm sample was stratified in accordance with data from the Federal Statistical Office. Demographic indicators such as migration background and parents' educational background were taken into account. It should be noted that the index score of the vocabulary acquisition subscale assesses only a selected facet of language skills. Therefore, a risk of delayed language development (referred to as children at risk/CaR) indicated by below-average results in this subscale does not imply a diagnosis of a language disorder. Conversely, an average result does not eliminate the possibility of a language disorder.

Based on the individual scores, we calculated the proportion of children at risk within each group of children. To illustrate, in a group of seven children, two exhibited an index score one standard deviation below the mean, thus indicating a risk for 29 % of the children in the group. The mean proportion of children at risk within groups is 33.04 % ($SD = 28.60$, $\text{Mdn} = 33.3 \%$, $\text{min} = 0 \%$, $\text{max} = 100 \%$).

Statistical Analyses

To explore research question 1a, teachers' language scaffolding behavior was analyzed. Groups were categorized based on the quality of the CLASS dimension Language Modeling. Due to the small sample size, low variance, and an overall low score according to CLASS, we decided to divide the sample into two groups. Videos with scores 1, 2 (CLASS low quality) and 3 (CLASS lower mid-range quality) were assigned the label of Low LM, while videos with scores 4, 5 (CLASS higher mid-range quality) and 6 (CLASS high quality) were assigned the label of High LM. Videos with ratings of 7 were not present in our sample. A multivariate analysis of variance (MANOVA) was conducted to evaluate the effects of Low LM vs. High LM groups on several dependent variables representing the seven different question types. Question types Code 2a, b and c were analyzed together as 'Code 2-Complex decision questions'. The assumption of normal distribution was not met for all dependent variables in both groups (Codes 1, 3, 4, 5, 6), as indicated by the Shapiro-Wilk test ($p < .05$). The correlations between the dependent variables were low ($r < .80$) suggesting that multicollinearity was not a significant concern (Pituch & Stevens, 2015). However, the assumption of linearity was only partially satisfied, which may impact the interpretation of the results. The homogeneity of error variances was confirmed for the dependent variables using Levene's test ($p > .05$), and the homogeneity of covariances was verified with Box's test ($p > .05$). Wilks' Lambda was used as the test statistic for the MANOVA. Post-hoc univariate ANOVAs were conducted for each of the different question types as dependent variables to further explore the specific group differences identified by the MANOVA. These analyses were used to identify which particular question types contributed to any overall multivariate differences observed. A Bonferroni adjusted alpha level of 0.007 was used to control for Type 1 errors.

Research question 1b aimed to analyze the length of the children's answers to different question types, operationalized by the word count of the answer. A comparison was made between the answer lengths of different question types. For the analysis, 'Code 0-Question particles' and 'Code 2b-Requests for action in the form of questions' were excluded, as no verbal reaction from the children is expected for these types of questions. Due to violations of the assumptions of homogeneous variance and normally distributed data (Shapiro-Wilk test, $p < .05$), as well as unequal sample sizes, a Welch ANOVA with Games-Howell post-hoc test was performed to compare the different question types. Every child who participated in a video-recording was included in the analysis.

The second research question investigated whether teachers' language scaffolding behavior in the videos differed according to the proportion of children at risk of delayed language development. The proportion of children at risk was calculated for each group, based on the children's individual language outcomes. Based on the proportions, we used a median-split for group comparisons. This resulted in a low-risk subgroup (Low CaR; <33.3 % of children at risk, $N = 28$) and a high-risk subgroup (High CaR; ≥ 33.3 % of children at risk, $N = 34$). A MANOVA was conducted to evaluate the effects of groups on the seven different question types. The assumption of normal distribution was not met for all dependent variables for both groups (Codes 0, 1, 3, 4, 5, 6), as indicated by the Shapiro-Wilk test ($p < .05$). There was no evidence of multicollinearity between the dependent variables ($r < .80$; Pituch & Stevens, 2015). Linearity was absent or partial. There was homogeneity of error variances for the variables ($p > .05$), but not for 'Code 6-Explanation and justification questions' (Levene's test, $p = .006$), and there was no homogeneity of covariances (Box's test, $p = .005$). Wilks' Lambda was used as the test statistic for the MANOVA. Post-hoc univariate ANOVAs were performed for each of the different question types. A Bonferroni adjusted alpha level of 0.007 was used to control for Type 1 errors. Due to the partial violation of assumptions, especially regarding normal distribution and linearity, the statistical power of the analysis may be limited. Furthermore, Mann-Whitney U-tests were employed for the statistical analysis of differences in verbal participation of teachers and children in the groups. The distributions did not differ between the two groups (Kolmogorov—Smirnov, $p > .05$). Two videos were excluded for the analysis of the second research question due to missing information on children's language skills.

All analyses were carried out using the software SPSS (IBM Corp., Released 2021).

Results

Validation of the Scoring System for Questions (SSQ)

Language Modeling (CLASS) and Teacher Questions (SSQ)

To validate the SSQ, this section analyzed the relationship between ratings on Language Modeling in CLASS Pre-K and the frequency of different question types. The descriptive results presented in Tab. 5.2 indicate that, with the exception of 'Code 3-Simple labeling questions', teachers in high LM groups tended to have higher mean values for question types per minute. A one-way MANOVA

Tab. 5.2 Descriptive Statistics and Post-Hoc Univariate ANOVA Results for different Question Types in Low LM and High LM Groups.

Variables	Low LM	High LM					
	<i>M (SD)</i>	<i>M (SD)</i>	F	df ₁	df ₂	<i>p</i>	partial η^2 ara>
Code 0—Question particle	0.62 (0.48)	0.86 (0.53)	3.488	1	62	.067	.053
Code 1—Simple decision questions	0.29 (0.27)	0.32 (0.28)	0.324	1	62	.572	.005
Code 2—Complex decision questions	1.76 (0.78)	2.12 (0.91)	2.937	1	62	.092	.045
Code 3—Simple labeling questions	1.65 (1.15)	1.64 (0.78)	0.000	1	62	.988	.000
Code 4—Simple descriptive questions	0.44 (0.33)	0.75 (0.40)	11.649	1	62	.001	.158
Code 5—Describing and process-related questions	0.20 (0.25)	0.21 (0.31)	0.036	1	62	.849	.001
Code 6—Explanation and justification questions	0.20 (0.22)	0.30 (0.26)	2.532	1	62	.117	.039

Note. Low LM: low quality rating in the dimension of Language Modeling (score 1 to 3), high LM: mid-range to high quality rating (score 4 to 6); $N_{\text{lowLM/highLM}} = 35/29$; M : mean value of question types per minute; $\alpha_{\text{adj}} = 0.007$.

showed a statistically significant difference between groups with low (Low LM) and high (High LM) quality of Language Modeling on the combined dependent variables of the different question types, $F(7, 54) = 2.347$, $p = .036$, partial $\eta^2 = .227$, Wilk's $\Lambda = .773$. The results of the post-hoc univariate ANOVAs showed a statistically significant difference between the two groups with Low LM and High LM for 'Code 4-Simple descriptive questions', $F(1, 62) = 11.649$, $p < .001$, partial $\eta^2 = .158$.

Tab. 5.3 Descriptive Statistics of the Length of Answers to different Question Types.

Variables	N	<i>M</i>	<i>SD</i>	min	max
Code 1—Simple decision questions	54	2.55	1.68	1.00	9.50
Code 2a and c –Closed-ended yes/no and Alternative questions	64	2.31	1.55	1.00	13.50
Code 3—Simple labeling questions	64	2.27	0.83	1.00	6.60
Code 4—Simple descriptive questions	63	3.42	2.01	1.00	14.00
Code 5—Describing and process-related questions	48	3.89	2.63	1.00	12.00
Code 6—Explanation and justification questions	50	6.11	2.91	2.00	21.00

Note. N: Number of videos in which the question type was asked; *M*: mean value of the word count of the answers to different question types.

Types of Questions and Children's Word Count

Tab. 5.3 presents the descriptive statistics of language production following different question types. The mean value of the word count of children's answers to different question types was highest for questions Code 6. Additionally, the answers to Code 5 and 4 questions were longer than the answers to the closed-ended questions.

Results of the Welch ANOVA indicated that there was a significant difference in the word count of the answers, $F(5, 152.00) = 21.42$; $p < .001$, $\eta^2 = .298$. The Games-Howell post-hoc test showed a significant difference in the word count of children's answers to 'Code 6-Explanation and justification questions' compared to all other question types (see Tab. 5.4). In addition, there were significant differences between answers to 'Code 5-Describing and process-related questions' and all other question types except Code 4 questions. The test also showed a significant difference between answers to questions of Codes 2 and 3, 2 and 4, and 3 and 4. However, there were no significant differences between answers to questions of Code 1 and 2, 1 and 3, and 1 and 4.

Differences between Groups of High versus Low Proportions of Children at Risk of Delayed Language Development

Language Modeling (CLASS)

The mean quality ratings in Language Modeling for the groups with a low proportion of children at risk of delayed language development (Low CaR; N = 28;

Tab. 5.4 Games-Howell Post-Hoc Test Results of Differences in the Length of Answers to different Question Types.

Variables		Mean Difference	SE	<i>p</i>	95 % Confidence Interval	
					Lower Bound	Upper Bound
Code 1	Code 2	0.242	0.27	.948	-0.548	1.032
	Code 3	0.287	0.25	.860	-0.446	1.021
	Code 4	-0.867	0.34	.121	-1.857	0.122
	Code 5	-1.339	0.44	.038	-2.633	-0.044
	Code 6	-3.557	0.47	<.001	-4.933	-2.181
Code 2	Code 3	0.045	0.18	<.001	-0.474	0.565
	Code 4	-1.109	0.29	.003	-1.961	-0.258
	Code 5	-1.580	0.41	.003	-2.778	-0.383
	Code 6	-3.799	0.43	<.001	-5.085	-2.513
Code 3	Code 4	-1.155	0.27	<.001	-1.954	-0.355
	Code 5	-1.626	0.39	.002	-2.789	-0.463
	Code 6	-3.845	0.42	<.001	-5.098	-2.591
Code 4	Code 5	-0.471	0.46	.906	-1.802	0.860
	Code 6	-2.690	0.48	<.001	-4.101	-1.280
Code 5	Code 6	-2.219	0.56	.002	-3.848	-0.590

Note. Code 1: Simple decision questions; Code 2: Complex decision questions; Code 3: Simple labeling questions; Code 4: Simple descriptive questions; Code 5: Describing and process-related questions; Code 6: Explanation and justification questions.

$M = 3.54$; $SD = 0.96$) and those with a high proportion (High CaR; $N = 34$; $M = 3.24$; $SD = 0.92$) were comparable and both fell within the lower medium quality range. The Mann-Whitney U test showed no significant difference between the Low CaR ($Mdn = 4.0$) and the High CaR group ($Mdn = 3.0$), $U = 373.0$; $Z = -1.542$; $p = 0.125$; $r = 0.20$.

Teacher Questions (SSQ)

Figure 5.1 shows the distribution of the types of questions asked in the two groups. It is noteworthy that Code 5 and Code 6 questions, which have a high potential to stimulate language production (as indicated by the results of research question 1b), were relatively rare. According to the SSQ, slightly more than half of the questions in both groups were classified as closed-ended. The one-way MANOVA revealed a statistically significant difference between the groups of

Low CaR and High CaR on the combined dependent variables of question types, $F(7, 54) = 2.973$, $p = .010$, partial $\eta^2 = .278$, Wilk's $\Lambda = .722$.

Post-hoc univariate ANOVAs (see Tab. 5.5) indicated a marginally statistically significant difference between the Low CaR and High CaR groups for 'Code 6-Explanation and justification questions', $F(1, 60) = 7.187$, $p = .009$, partial $\eta^2 = .107$. This result could be replicated using non-parametric tests (Mann-Whitney U test).

Teachers' and Children's Word Count

Unexpectedly, we found no significant differences between Low CaR and High CaR groups in the word count of teachers and the word count of children at the group level (Tab. 5.6). An additional exploratory analysis of children's individual word count, depending on whether a child belonged to a Low CaR ($N = 122$, $Mdn = 6.14$) or a High CaR group ($N = 140$, $Mdn = 5.13$), also revealed no significant difference in this groups comparison, $U = 7866.5$; $Z = -1.101$; $p = 0.272$, $r = 0.07$. Only comparing individual word counts per minute of children with or without risk of delayed language development showed differences. Children without risk of delayed language development. ($N = 159$ children; $Mdn = 7.47$) spoke significantly more words than children at risk ($N = 74$ children; $Mdn = 3.84$), $U = 4098.5$; $Z = -3.726$; $p < 0.001$; $r = 0.24$. However, the effect is small (Cohen, 1969).

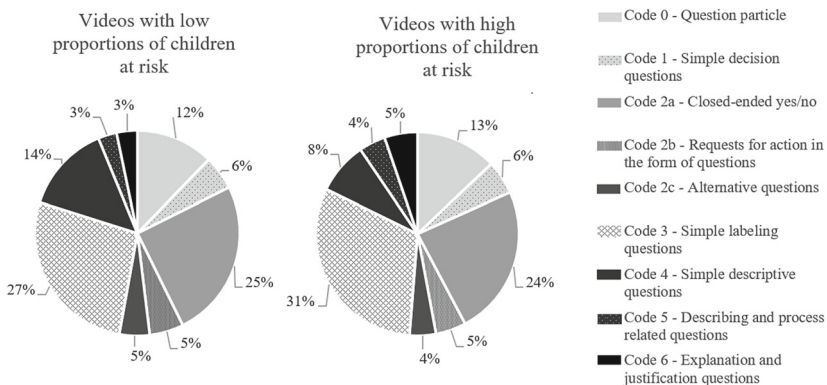


Fig. 5.1 Distribution of Question Types in Low CaR and High CaR Groups.

Tab. 5.5 Descriptive Statistics and Post-hoc Univariate ANOVA Results for different Question Types in Low CaR and High CaR Groups.

Variables	Low CaR	High CaR					
	<i>M (SD)</i>	<i>M (SD)</i>	F	df ₁	df ₂	<i>p</i>	partial η^2
Code 0—Question particle	0.63 (0.40)	0.79 (0.59)	1.509	1	60	.224	.025
Code 1—Simple decision questions	0.27 (0.28)	0.33 (0.27)	0.665	1	60	.418	.011
Code 2—Complex decision questions:	1.81 (0.84)	1.94 (0.82)	0.403	1	60	.528	.007
Code 3—Simple labeling questions	1.39 (0.77)	1.81 (1.13)	2.901	1	60	.094	.046
Code 4—Simple descriptive questions	0.70 (0.44)	0.50 (0.33)	3.908	1	60	.053	.061
Code 5—Describing and process-related questions	0.15 (0.17)	0.25 (0.34)	2.136	1	60	.149	.034
Code 6—Explanation and justification questions	0.16 (0.15)	0.32 (0.28)	7.187	1	60	.009	.107

Note. Low CaR: low proportion of children at risk of delayed language development (<33.30 %), High CaR: high proportion of children at risk of delayed language development (≥ 33.30 %); $\alpha_{adj} = 0.007$.

Discussion

This study investigated language scaffolding in early science education, focusing on different measures of the quality of language support. It builds on previous research that underscores high-quality language scaffolding in ECE as a key factor influencing both children's language development and their domain-specific learning (Hardy et al., 2019b; Henrichs & Leseman, 2014). This is particularly

Tab. 5.6 Descriptive Statistics, Mean Ranks, and Test Statistics of Word Count of Preschool Teachers and Children.

Variables	Low CaR		High CaR		U	p	Z	r
	M (SD)	Mdn	M (SD)	Mdn				
Word count per minute preschool teachers	88.53 (18.41)	93.15	89.71 (24.41)	94.15	443.0	.648	-0.467	.06
Word count per minute children (group level)	34.05 (9.72)	32.49	31.83 (16.00)	25.98	381.0	.183	-1.344	.17

Note. Low CaR: low proportion of children at risk of delayed language development (<33.30 %), High CaR: high proportion of children at risk of delayed language development (≥ 33.30 %); U: Mann-Whitney U test statistic, r: Effect size ($r < 0.30$ small, $0.30 \leq r \leq 0.50$ medium, $r > 0.50$ strong; Cohen, 1969); $N_{\text{low-risk/high-risk}} = 28/34$ groups.

important given the increasing linguistic diversity among children (OECD, 2019). This study specifically highlights the role of different types of preschool teacher questioning as a stimulating strategy in science activities in preschool (e.g., Lee & Kinzie, 2012). Teachers' questions were coded using an adapted Scoring System for Questions (SSQ; Gabler et al., 2024; Schmidt et al., 2019). The first research question focused on the validity of the adapted SSQ. The second research question examines differences in language scaffolding practices and verbal participation between preschool groups with high and low proportions of children at risk for delayed language development.

Research question 1: Validation of the Scoring System for Questions (SSQ)

The SSQ distinguishes seven different question types, ranging from closed-ended to open-ended questions. We examined whether there were differences in the use of question types between groups with high- and low-quality Language Modeling, as assessed by the validated CLASS scoring system (Pianta et al., 2008). We assumed that differences in the use of question types between the two groups indicate consistency between the high-inference CLASS ratings and the low-inference SSQ measure of teacher questioning. Our results indicate notable differences, with higher mean values for almost all question types in the high-quality Language Modeling groups. In particular, 'Code 4—Simple descriptive questions,' which encourage short descriptions and allow for a range of answers, were used significantly more

frequently in the high-quality groups. To further validate the classification of question types with a higher potential to stimulate language production (Codes 5 and 6) within the SSQ, we analyzed the number of words in children's answers to different question types. Children's answers varied significantly based on question type; they produced more words when responding to questions that required explanations, justifications (Code 6), and descriptions of scientific processes (Code 5), compared to closed-ended questions that merely prompted selections, agreements, or disagreements (Code 2).

Both results provide initial evidence of the validity of the SSQ: The SSQ enables the identification of differences in how teachers use questioning as a language scaffolding strategy. The consistency between the two measurement systems (SSQ and CLASS) suggests that the frequency of the different question types used by teachers provides valuable insights into the quality of language scaffolding. In addition, the SSQ was validated by showing that different types of questions have different potential to stimulate language production, with open-ended, process-related descriptive questions and explanation and justification questions eliciting longer answers from children. The results indicate that the SSQ is an effective tool for distinguishing between different types of questions and providing insights into the quality of language scaffolding. Furthermore, it enables the assessment of how preschool teachers use stimulating questions to promote children's active verbal participation, thereby initiating co-constructive learning processes that arise through interaction (Vygotsky, 1978). The SSQ extends previous methods of assessing teachers' questions by focusing on stimulating children's verbal participation, supporting process-related skills and including important aspects of language development, particularly concerning vocabulary for labeling and describing in the preschool context. This focus on science-specific vocabulary is assumed to be crucial for domain-specific learning (Heppt et al., 2020). The use of the SSQ helps to identify needs and provides concrete guidance for professional development aimed at supporting both language development and the acquisition of science knowledge. Language scaffolding in the context of early science can lay the foundation for later acquisition of science-specific language.

Research question 2: Differences between Groups of High versus Low Proportions of Children at Risk of Delayed Language Development

The study also aimed to explore whether differences exist in teacher questioning between groups with high and low proportions of children at risk of delayed language development. Given the previously inconsistent findings on the adaptive use of scaffolding strategies (e.g., Beckerle & Mackowiak, 2019; Hormann & Skowronek,

2019), this analysis of preschool language scaffolding which combines the high-inference CLASS ratings with the low-inference SSQ, provides new insights on teacher language scaffolding in early science activities.

Our findings indicated a low to medium quality of interaction in the CLASS dimension of Language Modeling, with no significant differences between groups with high and low proportions of children at risk. In the SSQ assessments, the proportion of closed- and open-ended questions was found to be relatively similar in both groups. Approximately half of the questions could be classified into one of the two categories. In both groups, approximately one-quarter of the questions were either closed yes/no questions or simple labeling questions. It is noteworthy that there was a significant difference in the total number of 'Explanation and justification questions' (Code 6) in favor of teachers in groups with high proportions of children at risk. This finding is surprising, as these questions are assumed to be particularly demanding both cognitively and linguistically (e.g., Beckerle & Mackowiak, 2019). However, as already mentioned, the Code 6 questions accounted for only 3 % of all questions in Low CaR groups and 5 % in High CaR groups. This infrequent use of open-ended questions with a high potential for stimulating language production is consistent with the findings of other studies that demonstrate a general preference for closed-ended questions in ECE (e.g., Siraj-Blatchford and Manni, 2008). The limited use of these strategies indicates that the potential to enhance children's verbal participation was not fully exploited, a finding also supported by previous research (e.g., Lee & Kinzie, 2012). These results are problematic since active verbal participation by children impacts domain-specific learning (e.g., Gabler et al., 2024; Mannel et al., 2016; Sedova et al., 2019). It can also be assumed that children's verbal participation can positively influence the effectiveness of teachers' scaffolding. (Bürgermeister et al., 2019; Herrmann et al., 2021).

Our analysis further indicates that a high proportion of children at risk for delayed language development in a group does not necessarily result in preschool teachers dominating the interaction or reducing children's verbal participation at group level. However, when examining the word count of children at the individual level, the results show that children at risk spoke significantly less than their peers. This suggests that children at risk may experience fewer opportunities to engage in meaningful verbal interactions. Given the pivotal role of active language participation in language and conceptual knowledge development (e.g., Mannel et al., 2016), this indicates that children at risk may have fewer opportunities for verbal participation than children with higher language skills, which may limit their learning opportunities.

The results do not provide conclusive evidence about the adaptivity of language scaffolding in this sample and should be interpreted with caution. While our findings

suggest that teachers may use a variety of questioning strategies, they do not provide clear evidence that these strategies are adaptively tailored to meet the linguistic needs of their group of children. This reflects a broader challenge in assessing the adaptive use of scaffolding strategies in ECE, particularly at group level (Beckerle & Mackowiak, 2019; Hormann & Skowronek, 2019). These findings highlight the importance of future studies focusing on the adaptive use of scaffolding strategies in groups with diverse language skills. By integrating both high- and low-inference assessment methods, future research might better capture the nuanced ways in which language scaffolding is used and adapted across different learning settings.

Moreover, from a practice perspective, the adaptive use of scaffolding is a major challenge for preschool teachers. Understanding how teachers can effectively adapt their questioning strategies to support both language and cognitive development is therefore also crucial for developing effective professional development programs (Gabler et al., 2024).

Limitations and Future Research

The current research has several limitations that should be considered when interpreting its results. First, the cross-sectional design restricts the ability to examine longitudinal changes in children's development related to language scaffolding. Second, the sample may not be representative, as 19 % of the participating preschool teachers have an academic background, which is more than twice the average of the German preschool sector (Autorengruppe Bildungsberichterstattung, 2020). Furthermore, there was limited information about the linguistic backgrounds of the participating children. From a methodological perspective, it should be noted that both the variance in and the implementation of high-quality Language Modeling were relatively low in our sample, which impacted the statistical analyses.

Moreover, it is important to note that the CLASS rating and the focus on questions in the SSQ do not encompass the full range of language scaffolding strategies crucial for assessing the adaptive behavior of preschool teachers (Beckerle & Mackowiak, 2019). Therefore, future research should relate this specific focus to a broader framework of language scaffolding. Our study also points to the need to examine individual engagement in group discussions in more detail, especially in relation to science content and science-specific vocabulary. It is also important to consider that not all questions asked by teachers are answered by children, especially when multiple questions are asked in quick succession or

when children are not given time to react (Wirts & Cordes, 2021). Additionally, future studies could benefit from examining whether teachers address specific questions to individual children or whether certain types of questions are more likely to elicit answers from children with high or low language skills. In addition to considering the length of children's answers, qualitative analyses also offer promising approaches for gaining deeper insights into the language-stimulating potential of different types of questions, for example, by examining word types or the use of domain-specific vocabulary in answers.

Conclusion and Implications

This study provides insights into language scaffolding in early science education in groups with varying compositions of children at risk of delayed language development. A central component of this study is the Scoring System for Questions (SSQ), adapted from a system originally designed for the primary school context (Gabler et al., 2024) and integrated with elements from a preschool framework (Schmidt et al., 2019). Our findings provide initial evidence supporting the validity of the SSQ. As the SSQ distinguishes between closed- and open-ended questions and emphasizes the language-stimulating potential of the different question types (in terms of quantity) asked by preschool teachers, it can be used for further analyses of language scaffolding on a low-inference level.

Furthermore, the SSQ has practical applications for professional development programs for preschool teachers. By providing examples and differentiating between question types, the SSQ enables preschool teachers to reflect critically on and enhance their language behavior and language scaffolding skills (Gabler et al., 2024). Particularly in the context of professional development in early science education, the specific question types differentiated in the SSQ provide targeted strategies to support both language development and domain-specific learning.

Acknowledgements The project ProfinK (Förderdiagnostische Professionalisierung in der inklusiven naturwissenschaftlichen Bildung in der KiTa [Professional learning communities for preschool teachers to support their ability to use diagnostic information for instructional decisions]) was funded by the Federal Ministry of Education and Research (BMBF) (FKZ: 01NV2110).

We want to thank the preschool teachers, their institutions, the children and their parents who participated in the project. We would also like to thank the student assistants who contributed to the completion of the study.

References

- Autorengruppe Bildungsberichterstattung. (2020). Bildung in Deutschland 2020. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt. <https://www.bildungsbericht.de/de/bildungsberichte-seit-2006/bildungsbericht-2020/pdf-dateien-2020/bildungsbericht-2020-barrierefrei.pdf>. Accessed: 20. Sept. 2024.
- Beckerle, C. (2017). *Alltagsintegrierte Sprachförderung im Kindergarten und in der Grundschule. Evaluation des »Fellbach-Konzepts«*. Beltz.
- Beckerle, C., & Mackowiak, K. (2019). Adaptivität von Sprachförderung im Kita-Alltag. *Lernen und Lernstörungen*, 8(4), 203–211. <https://doi.org/10.1024/2235-0977/a000270>
- Beckerle, C., Mackowiak, K., Koch, K., Löffler, C., Heil, J., Pauer, I., et al. (2018). Der Einsatz von Sprachfördertechniken in unterschiedlichen Settings in Kindertageseinrichtungen. *Frühe Bildung*, 7(4), 215–222. <https://doi.org/10.1026/2191-9186/a000396>
- Beckerle, C., Mackowiak, K., & Kuchart, D. (2020). B-SFT: Beobachtungssystem zur Erfassung von Sprachfördertechniken im Kita- und Grundschulalltag. In C. Beckerle, S. Gentrup, C. Titz, & K. Mackowiak (Hrsg.), *Forschungsinstrumente im Kontext institutioneller (schrift-)sprachlicher Bildung* (S. 79–101). Klinkhardt.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive Domain. David McKay.
- Bürgermeister, A., Große, G., Leuchter, M., Studhalter, U., & Saalbach, H. (2019). Interaktion von pädagogischen Fachkräften und Kindern in naturwissenschaftlichen Lerngelegenheiten im Kindergarten. *Frühe Bildung*, 8(1), 13–21. <https://doi.org/10.1026/2191-9186/a000406>
- Callanan, M., & Valle, A. (2008). Co-Constructing Conceptual Domains Through Family Conversations and Activities. In B. H. Ross (Ed.), *Advances in Research and Theory. Advances in Research and Theory* (Psychology of Learning and Motivation Ser, v. Volume 49, vol. 49, S. 147–165). Elsevier Science & Technology.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.
- Elstrodt, N., Graf, J., Sahn, M., Starke, A., & Ritterfeld, U. (2019). Interaktionen im inklusiven Unterricht der Grundschule – Analysen zum kontingenten Unterrichten. In K. Verrière & L. Schäfer (Hrsg.), *Interaktion im Klassenzimmer* (S. 31–50). Springer Fachmedien Wiesbaden.
- Gabler, K., Henschel, S., Hardy, I., Sontag, C., & Heppt, B. (2024). Sprachförderliches Unterrichten im Sachunterricht der Grundschule. *Zeitschrift für Grundschulforschung*, 17(1), 87–110. <https://doi.org/10.1007/s42278-024-00191-8>
- Geyer, S., & Müller, A. (2021). Wie adaptiv ist Sprachförderung für Kinder mit Deutsch als Zweitsprache? *Zeitschrift für Angewandte Linguistik*, 2021, (74), 31–60. <https://doi.org/10.1515/zfal-2021-2052>
- Gibbons, P. (2015). *Scaffolding language, scaffolding learning. Teaching English language learners in the mainstream classroom* (2. Aufl.). Portsmouth NH: Heinemann.
- Hamel, E., Joo, Y., Hong, S.-Y., & Burton, A. (2021). Teacher Questioning Practices in Early Childhood Science Activities. *Early Childhood Education Journal*, 49(3), 375–384. <https://doi.org/10.1007/s10643-020-01075-z>

- Hardy, I., Decristan, J., & Klieme, E. (2019a). Adaptive teaching in research on learning and instruction. *Journal for Educational Research Online*, 11(2), 169–191. <https://doi.org/10.25656/01:18004>
- Hardy, I., Sauer, S., & Saalbach, H. (2019b). Frühe sprachliche Bildung im Kontext Naturwissenschaften: Effekte einer Intervention im Kindergarten. *Psychologie in Erziehung und Unterricht*, 66, 196–216.
- Haug, B. S., & Ødegaard, M. (2014). From words to concepts: Focusing on word knowledge when teaching for conceptual understanding within an inquiry-based science setting. *Research in Science Education*, 44(5), 777–800. <https://doi.org/10.1007/s11165-014-9402-5>
- Henrichs, L. F., & Leseman, P. P. (2014). Early Science instruction and academic language development can go hand in hand. The promising effects of a low-intensity teacher-focused intervention. *International Journal of Science Education*, 36(17), 2978–2995. <https://doi.org/10.1080/09500693.2014.948944>
- Heppt, B., Henschel, S., Hettmannsperger-Lippolt, R., Sontag, C., Gabler, K., Hardy, I., et al. (2020). Erfassung und Bedeutung des Fachwortschatzes im Sachunterricht der Grundschule. In C. Titz, S. Weber, H. Wagner, A. Ropeter, S. Geyer, & M. Hasselhorn (Eds.), *Sprach- und Schriftsprachförderung wirksam gestalten: Innovative Konzepte und Forschungsimpulse (Bildung durch Sprache und Schrift)*. Verlag W. Kohlhammer.
- Herrmann, A., Bürgermeister, A., Lange-Schubert, K., & Saalbach, H. (2021). Die Bedeutung von Partizipation und Scaffolding für die Leistung im naturwissenschaftlichen Sachunterricht in Klassen mit hohem und niedrigem Anteil mehrsprachiger Schüler*innen. *Zeitschrift für Grundschulforschung*, 14(2), 305–323. <https://doi.org/10.1007/s42278-021-00112-z>
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1), 55–88. <https://doi.org/10.1016/j.dr.2005.11.002>
- Hormann, O., & Skowronek, M. (2019). Wie adaptiv ist der Einsatz von Sprachlehrstrategien in KiTas? Ergebnisse einer Videoanalyse. *Frühe Bildung*, 8(4), 194–199. <https://doi.org/10.1026/2191-9186/a000444>
- Howard, S. J., Siraj, I., Melhuish, E. C., Kingston, D., Neilsen-Hewett, C., Rosnay, M. de, et al. (2018). Measuring interactional quality in pre-school settings: Introduction and validation of the Sustained Shared Thinking and Emotional Wellbeing (SSTEW) scale. *Early Child Development and Care*, 190(7), 1017–1030. <https://doi.org/10.1080/03004430.2018.1511549>
- IBM Corp. (Released 2021). IBM SPSS Statistics for Windows [Computer software]: IBM Corp.
- Jurkic, A., Halliday, S. E., & Hascher, T. (2023). The relationship of language and social competence of preschool- and kindergarten-age single and dual language learners in Switzerland and Germany. *Early Childhood Research Quarterly*, 64, 72–83. <https://doi.org/10.1016/j.ecresq.2023.02.003>
- Justice, L. M., Mashburn, A., Hamre, B., & Pianta, R. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, 23(1), 51–68. <https://doi.org/10.1016/j.ecresq.2007.09.004>
- Justice, L. M., Jiang, H., & Strasser, K. (2018). Linguistic environment of preschool classrooms: What dimensions support children’s language growth? *Early Childhood Research Quarterly*, 42, 79–92. <https://doi.org/10.1016/j.ecresq.2017.09.003>

- Kempert, S., Schalk, L., & Saalbach, H. (2019). Übersichtsartikel: Sprache als Werkzeug des Lernens: Ein Überblick zu den kommunikativen und kognitiven Funktionen der Sprache und deren Bedeutung für den fachlichen Wissenserwerb. *Psychologie in Erziehung und Unterricht*, (0). <https://doi.org/10.2378/PEU2018.art19d>
- Kucharz, D. (2018). Alltagsintegrierte sprachliche Bildung und Förderung im Elementarbereich. In C. Titz, S. Geyer, A. Ropeter, H. Wagner, S. Weber, M. Hasselhorn, et al. (Eds.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln*. W. Kohlhammer GmbH.
- Lee, Y., & Kinzie, M. B. (2012). Teacher question and student response with regard to cognition and language use. *Instructional Science*, 40(6), 857–874. <https://doi.org/10.1007/s11251-011-9193-2>
- Mannel, S., Hardy, I., Sauer, S., & Saalbach, H. (2016). Sprachliches Scaffolding zur Unterstützung naturwissenschaftlichen Lernens im Kindergartenalter. In E. Tschirner, O. Bärenfänger, & J. Möhring (Hrsg.), *Sammelband zum Symposium „Kompetenzprofile Deutsch als fremde Bildungssprache“* (S. 97–113). Tübingen: Stauffenburg Verlag.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child development*, 79(3), 732–749. <https://doi.org/10.1111/j.1467-8624.2008.01154.x>
- Müller, A., Geyer, S., & Smits, K. (2016). Die Sprache der Sprachförderung—Ist das sprachliche Handeln an die Förderbedarfe von DaZ-Kindern angepasst? *Diskurs Kindheits- und Jugendforschung*, 11(1), 81–98. <https://doi.org/10.3224/diskurs.v11i1.22250>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 International Results in Reading. Chestnut Hill, Mass.: TIMSS & PIRLS International Study Center Lynch School of Education Boston College.
- National Academies of Sciences, Engineering, and Medicine, 2017 National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the Educational Success of Children and Youth Learning English*. National Academies Press.
- OECD. (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*. PISA, OECD Publishing.
- Peterson, S. M., & French, L. (2008). Supporting young children's explanations through inquiry science in preschool. *Early Childhood Research Quarterly*, 23(3), 395–408. <https://doi.org/10.1016/j.ecresq.2008.01.003>
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System Manual Pre-K: Teachstone.
- Pituch, K. A., & Stevens, J. P. (2015). *Applied Multivariate Statistics for the Social Sciences*: Routledge.
- Prediger, S., Erath, K., & Moser Opitz, E. (2018). Language challenges for students with mathematical difficulties—An overview on research results and instructional approaches.
- Röhner, C., Li, M., & Hövelbrinks, B. (2010). Fragestrategien im fachbezogenen Sprachförderunterricht. In K.-H. Arnold, K. Hauenschild, B. Schmidt, & B. Ziegenmeyer (Eds.), *Zwischen Fachdidaktik und Stufendidaktik. Perspektiven für die Grundschulpädagogik—Jahrestagung 2009 der Kommission Grundschulforschung und Pädagogik der Primarstufe* (Jahrbuch Grundschulforschung, vol. 14, S. 89–92). VS Verl. für Sozialwiss.
- Saalbach, H., & Kempert, S. (2022). Sprache als wichtiges Werkzeug der Lehrenden. Verbale Interaktion zur Unterstützung des frühen MINT-Lernens. In M. Schneider, R. Grabner,

- H. Saalbach, & L. Schalk (Eds.), *Wie guter Unterricht intelligentes Wissen schafft. Handlungswissen aus der Lehr-Lernforschung: Eine Festschrift für Elsbeth Stern* (S. 172–185). Verlag W. Kohlhammer.
- Schmidt, F. A., Risse, L. S., Beckerle, C., & Mackowiak, K. (2019). Der Einsatz von Fragearten in unterschiedlich stark strukturierten Situationen im Kita-Alltag. *Empirische Sonderpädagogik*, *11*(4), 310–317. <https://doi.org/10.25656/01:18337>
- Schulte, M. (2020). *Bildungssprachliche Anregungsqualität Im Fachunterricht. Videoanalysen Zur Sprachlich-Kognitiven Aktivierung Durch Lehrkraftfragen*. J. B. Metzler'sche Verlagsbuchhandlung & Carl Ernst Poeschel GmbH.
- Sedova, K., Sedlacek, M., Svaricek, R., Majcik, M., Navratilova, J., Drexlerova, A., et al. (2019). Do those who talk more learn more? The relationship between student classroom talk and student achievement. *Learning and Instruction*, *63*, 101217. <https://doi.org/10.1016/j.learninstruc.2019.101217>
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2022). *Gemeinsamer Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen*.
- Siraj-Blatchford, I. (2009). Conceptualising progression in the pedagogy of play and sustained shared thinking in early childhood education: A Vygotskian perspective. *Educational and Child Psychology*, *26*(2), 77–89. <https://doi.org/10.53841/bpsecp.2009.26.2.77>
- Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R., & Bell, D. (2002). *Researching Effective Pedagogy in the Early Years*. University of London, Institute of Education, Department of Educational Studies.
- Siraj-Blatchford, I., & Manni, L. (2008). 'Would you like to tidy up now?' An analysis of adult questioning in the English Foundation Stage. *Early Years*, *28*(1), 5–22. <https://doi.org/10.1080/09575140701842213>
- Smit, J., van A. A. Eerde, H., & Bakker, A. (2013). A conceptualisation of whole-class scaffolding. *British Educational Research Journal*, *39*(5), 817–834. <https://doi.org/10.1002/berj.3007>
- Stephan, F., Gunzenhauser, C., & Saalbach, H. (2022). Function of language skills in preschooler's problem-solving performance: The role of self-directed speech. *Journal of Applied Developmental Psychology*, *81*, 101431. <https://doi.org/10.1016/j.appdev.2022.101431>
- Stuck, A., Kammermeyer, G., & Roux, S. (2016). The reliability and structure of the Classroom Assessment Scoring System in German pre-schools. *European Early Childhood Education Research Journal*, *24*(6), 873–894. <https://doi.org/10.1080/1350293X.2016.1239324>
- Studhalter, U. T., Leuchter, M., Tettenborn, A., Elmer, A., Edelsbrunner, P. A., & Saalbach, H. (2021). Early science learning: The effects of teacher talk. *Learning and Instruction*, *71*, 101371. <https://doi.org/10.1016/j.learninstruc.2020.101371>
- Suchodoletz, A. von, Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher–child interactions in German preschools. *Early Childhood Research Quarterly*, *29*(4), 509–519. <https://doi.org/10.1016/j.ecresq.2014.05.010>
- Tompkins, V., Zucker, T. A., Justice, L. M., & Binici, S. (2013). Inferential talk during teacher–child interactions in small-group play. *Early Childhood Research Quarterly*, *28*(2), 424–436. <https://doi.org/10.1016/j.ecresq.2012.11.001>

- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in Teacher—Student Interaction: A Decade of Research. *Educational Psychology Review*, 22(3), 271–296. <https://doi.org/10.1007/s10648-010-9127-6>
- VERBI Software. (2021) MAXQDA 2022 [Computer software]. VERBI Software. Available from maxqda.com.
- Voet Cornelli, B. (2022). Die Qualität der Sprachentwicklungsdiagnostik bei mehrsprachigen Kindern: Pädiatrische Früherkennungsuntersuchungen auf dem Prüfstand.
- Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Harvard University Press.
- Walsh, R. L., & Hodge, K. A. (2018). Are we asking the right questions? An analysis of research on the effect of teachers' questioning on children's language during shared book reading with young children. *Journal of Early Childhood Literacy*, 18(2), 264–294. <https://doi.org/10.1177/1468798416659124>
- Wechsler, D. (2018). Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition. Deutsche Bearbeitung herausgegeben von Franz Petermann und Monika Daseking. Pearson.
- Whorrall, J., & Cabell, S. Q. (2016). Supporting Children's Oral Language Development in the Preschool Classroom. *Early Childhood Education Journal*, 44(4), 335–341. <https://doi.org/10.1007/s10643-015-0719-0>
- Wildgruber, A., Wertfein, M., Wirts, C., Kammermeier, M., & Danay, E. (2016). Situative Unterschiede der Interaktionsqualität im Verlauf des Kindergartenalltags. *Frühe Bildung*, 5(4), 206–213. <https://doi.org/10.1026/2191-9186/a000283>
- Wirts, C., & Cordes, A.-K. (2021). Wer spricht denn da? Redeanteile und das sprachanregende Potenzial verschiedener Fragetypen. *Forschung Sprache*, (2), 151–159.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Oppermann, E., Barentien, J., Burghardt, L., Steffensky, M., & Anders, Y. (2024). Beobachtungsverfahren zur Erfassung der Interaktionsqualität in Kindertageseinrichtungen. *Frühe Bildung*, 13, (1, 29–35). <https://doi.org/10.1026/2191-9186/a000647>

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Die Kompetenz von Lehrkräften im inklusiven Unterricht von Kindern im Autismus-Spektrum: Entwicklung und Validierung eines Assessments

Nick Gerrit Hasche, Mareike Kunter, Ann-Kathrin Krause und Charlotte Dignath

Zusammenfassung

Autistische Kinder in Regelklassen erreichen oftmals einen geringeren schulischen Erfolg als zu vermuten wäre. Lehrkräfte wiederum, als zentrale Akteur*innen beim Unterrichten von autistischen Kindern an Regelschulen, fühlen sich oft nicht ausreichend vorbereitet, um angemessen mit autistischen Kindern im Unterricht umzugehen. Nach dem Modell der professionellen Kompetenz lassen sich unterschiedliche Aspekte ableiten, die für einen autismussensiblen Unterricht relevant sein dürften. In diesem Artikel wird

N. G. Hasche (✉) · M. Kunter · A.-K. Krause
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt am Main, Deutschland
E-Mail: n.hasche@dipf.de

M. Kunter
E-Mail: m.kunter@dipf.de

A.-K. Krause
E-Mail: a.krause@dipf.de

C. Dignath
Goethe-Universität Frankfurt am Main, Frankfurt am Main, Deutschland
E-Mail: dignath@psych.uni-frankfurt.de

N. G. Hasche · M. Kunter · A.-K. Krause · C. Dignath
IDeA – Center for Research on Individual Development and Adaptive Education of Children at Risk, Frankfurt am Main, Germany

die Entwicklung und Validierung eines Assessments beschrieben, welches Wissen, Motivation und Überzeugung von angehenden und aktiven Lehrkräften in Bezug auf Autismus und autismussensiblen Unterricht erheben soll. Daten aus einer Expert*innenbefragung sowie einer Validierungsstudie mit Lehramtsstudierenden belegen, dass das Instrument verschiedene Kompetenzaspekte reliabel erfassen und Unterschiede zwischen Gruppen sowie instruktionsbedingte Veränderungen valide abbilden kann.

Schlüsselwörter

Autismus • Inklusiver Unterricht • Professionelle Kompetenz von Lehrkräften • Assessment • Validierungsstudie

Einleitung

Durch die zunehmende Prävalenz an Autismus-Diagnosen in Deutschland (Gómez-Marí et al., 2022; Lindmeier, 2018) und der damit steigenden Anzahl an autistischen Schüler*innen¹ in Regelschulen (Wittwer, 2022) erhöht sich auch die Wahrscheinlichkeit für Lehrkräfte, im Laufe ihrer Karriere mindestens ein autistisches Kind in ihrer Klasse zu unterrichten. Dass die Inklusion von Kindern im Autismus-Spektrum derzeit nicht zufriedenstellend umgesetzt wird, zeigen mehrere Studien (Jordan et al., 2019; van der Steen et al., 2020; Wittwer, 2022). Diese belegen, dass sich autistische Kinder in Regelklassen oftmals weniger gut entwickeln als neurotypische Kinder und schlechtere schulische Erfolge erzielen, obwohl sie über vergleichbare kognitive Fähigkeiten verfügen (Wittwer, 2022). Nach Jordan et al. (2019) hängt die Entwicklung vor allem von den sozio-emotionalen Erfahrungen der Kinder ab. Lehrkräfte sehen sich häufig jedoch nicht in der Lage, den Bedürfnissen aller Schüler*innen gerecht zu werden (van der Steen et al., 2020). So kommt es z. B. durch Ablenkungen und individuelle Barrieren dazu, dass autistische Schüler*innen nur ca. die Hälfte der Schulzeit kognitiv aktiviert sind (Sparapani et al., 2016). Sparapani und Kolleg*innen (2016) haben die kognitive Beteiligung von 196 autistischen Kindern anhand von

¹ Im aktuellen Diskurs finden sich verschiedene begriffliche Varianten bei der Benennung von autistischen Personen bzw. Personen mit Autismus oder Personen im Autismus-Spektrum, die unterschiedlich bewertet und akzeptiert werden. Aus diesem Grund bemühen wir uns, alle Varianten gleichermaßen zu verwenden. Gleichzeitig distanzieren wir uns von jeglicher Form pathologisierender Begrifflichkeiten.

beobachtbaren Verhaltensweisen in Form von Anwesenheit, Reaktion, Arbeitsbewältigung und Ausdauer bei schwierigen Aufgaben bewertet. Vor allem fehlende Selbstregulations- und Kommunikationsfähigkeiten und eine geringe Flexibilität konnten dabei als Barrieren im Unterricht festgestellt werden.

Dem aktuellen Bildungsbarometer nach fühlen sich Lehrkräfte auf die Herausforderungen eines inklusiven Unterrichts nicht ausreichend vorbereitet, was zu einer eher ablehnenden Überzeugung demgegenüber führt (Dignath et al., 2022; Robert Bosch Stiftung, 2023). So geben nur knapp über 50 % der befragten Lehrkräfte an, dass sie das Konzept der inklusiven Beschulung grundsätzlich richtig finden (Robert Bosch Stiftung, 2023). Doch gerade die Motivation und die Überzeugung der Lehrkräfte können neben dem Wissen von Lehrkräften als entscheidend für gelingende Inklusion gesehen werden (Haimour & Obaidat, 2013a; Gómez-Marí et al., 2022; Rackles, 2023). Diese Aspekte professioneller Kompetenz von Lehrkräften haben sich empirisch auch über die Frage inklusiven Unterrichts hinaus als relevant für das Unterrichtshandeln gezeigt.

Da noch wenig über das Zusammenspiel verschiedener Aspekte professioneller Lehrkräftekompetenz im Kontext von Autismus bekannt ist, wird dieser Zusammenhang in der folgenden Studie untersucht. Außerdem wird die Konstruktion und Validierung eines Assessment-Tools beschrieben, welches zum einen verschiedene Kompetenzaspekte von Lehrkräften messen und zum anderen diese dabei unterstützen soll, ihre eigene professionelle Kompetenz beim inklusiven Unterrichten von Kindern im Autismus-Spektrum einzuschätzen. Dazu wurde zunächst mittels eines systematischen Reviews ein Katalog an Fragebogen- und Testitems zusammengestellt, welcher inhaltlich durch Expert*innen ($N = 20$) bewertet und anschließend in einer Piloterhebung mit Lehramtsstudierenden ($N = 128$) getestet wurde. Dieser Beitrag stellt ein auf dieser Übersicht basierendes Testinstrument vor und prüft, ob dieses die professionelle Kompetenz im Bereich des autismussensiblen Unterrichtens auch für den deutschen Schulkontext präzise erfasst.

Theoretischer Rahmen

Die Perspektive von Lehrkräften auf die Inklusion von Kindern im Autismus-Spektrum

Lehrkräfte fühlen sich oftmals nicht in der Lage, Kinder im Autismus-Spektrum angemessen zu beschulen. In einer Untersuchung geben sie an, dass sie nicht über ausreichendes Wissen und Fähigkeiten verfügen, um bei internalisierenden

und externalisierenden Problemen angemessen reagieren zu können, was negative Auswirkungen für das autistische Kind haben kann (van der Steen et al., 2020). Vor allem junge Lehrkräfte geben demnach an, in ihrer Ausbildung nicht ausreichend auf diese Aufgabe vorbereitet worden zu sein (van der Steen et al., 2020). In vielen Lehramtsstudiengängen fehlen sonderpädagogische Lehranteile, die sich explizit mit Heterogenität im Klassenzimmer und der Inklusion auseinandersetzen (Berdemann et al., 2021) und auch im Bereich der Förderpädagogik wird das Thema Autismus nur vereinzelt behandelt, was dazu führt, dass viele angehende Lehrkräfte zu dieser Thematik nicht hinlänglich ausgebildet werden.

Das Modell der professionellen Kompetenz

Als zentrale Akteur*innen im inklusiven Unterricht stehen Lehrkräfte in der Verantwortung, diesen so zu gestalten, dass möglichst alle Kinder einen erfolgreichen Lernprozess durchlaufen. Um dieses Ziel zu erreichen gibt es allerdings keine standardisierten Prozesse. Vielmehr bedarf es unterschiedlicher Voraussetzungen der Lehrkraft, um die Kinder selbst in unvorhersehbaren Situationen kognitiv aktivieren und unterstützen zu können. Bei diesen Voraussetzungen handelt es sich um grundlegende Konzepte und Prinzipien, die sich nach Baumert und Kunter (2011) vermitteln und aneignen lassen und zusammenfassend als professionelle Kompetenz bezeichnet werden. Ihrem Modell nach besteht professionelle Kompetenz von Lehrkräften aus einem Zusammenspiel von Wissen, Überzeugung, Motivation und Selbstregulation. Der Aspekt Wissen unterteilt sich in dem ursprünglichen Modell dabei in unterschiedliche Facetten, wie Fachwissen, fachdidaktisches Wissen oder pädagogisches/psychologisches Wissen. Unter dem Kompetenzaspekt der Überzeugung werden die professionellen Werte der Lehrkraft bzw. ihre persönlichen Einstellungen verstanden. Bei der Motivation wird unter anderem zwischen der eigenen Selbstwirksamkeit und der intrinsischen Motivation unterschieden, während sich die Selbstregulation auf die individuellen Fähigkeiten bezieht, das berufliche Wohlbefinden zu regulieren.

Zusammenfassend skizziert das Modell der professionellen Kompetenz die notwendigen Grundlagen für eine erfolgreiche Ausübung des Lehrberufs. Dabei wird auf einer mehrdimensionalen Ebene das „Zusammenspiel kognitiver und motivational-selbstregulativer Merkmale bezogen auf die Bewältigung beruflicher Anforderungen“ beschrieben (Baumert & Kunter, 2011, S. 46 f.). Aufgrund seiner multidimensionalen Auffassung von professioneller Kompetenz lässt sich dieses Modell auch gut für den Kontext der autismsensiblen Beschulung spezifizieren.

Spezifikation für einen autismussensiblen Unterricht

Das Wissen einer Lehrkraft für autismussensiblen Unterricht umfasst spezielles Wissen über Autismus und Wissen um spezifische Unterrichtsstrategien und Handlungsweisen im Umgang mit autistischen Kindern. Es handelt sich also um fachunabhängiges Wissen und könnte somit in den breiten Bereich des pädagogischen Wissens eingeordnet werden (Voss et al., 2015). Das jeweilige Wissen über Autismus dürfte einen „fundamentalen Einfluss“ (Berdelmann, 2023, S. 30) darauf haben, wie Autist*innen überhaupt von der Lehrkraft verstanden und adressiert werden – und dies hat sich in den letzten Jahrzehnten stark gewandelt. Fehlendes Wissen oder Fehlkonzepte über Autismus können so dazu führen, dass Kinder im Autismus-Spektrum nicht in dem Maße gefördert werden, wie es notwendig wäre. Gleichzeitig konnten bei Lehrkräften Zusammenhänge zwischen einem höheren Wissen über Autismus, einer positiveren Überzeugung gegenüber Autist*innen und einer geringeren Stigmatisierung festgestellt werden (Wittwer et al., 2023).

Eine positive Überzeugung der Lehrkraft gilt nach Haimour und Obaidat (2013a) mit als eine der wichtigsten Voraussetzungen für erfolgreichen inklusiven Unterricht. So unterstützt eine positive Einstellung gegenüber Autismus bspw. die Anwendung von inklusiven Praktiken im Unterricht (Gómez-Marí et al., 2022), während negative Überzeugungen dazu führen können, dass sich die Lehrkraft weniger um eine individuelle Zuwendung bemüht (Haimour & Obaidat, 2013a). Negative Überzeugungen können sogar als eine Art Filter wirken und neue Informationen, die der eigenen Vorstellung widersprechen, abblocken und stattdessen nur solche zulassen, welche die bisherige Meinung bestätigen (Dignath et al., 2020).

Insgesamt gelten Überzeugungen als recht stabile Konstrukte, doch können sich auch diese durch eine direkte Adressierung ändern (Dignath et al., 2020). Neben einem besseren Wissen über Autismus hat auch die Motivation, vor allem die Selbstwirksamkeit der Lehrkraft, einen Einfluss auf ihre Überzeugung. Je mehr eine Lehrkraft davon ausgeht, dass sie in der Lage ist, Herausforderungen der Inklusion zu bewältigen, desto offener steht sie auch neuen Informationen und Ideen gegenüber (Gómez-Marí et al., 2022). Eine positive Selbstwirksamkeit führt zu einem höheren Engagement der Lehrkraft und außerdem zu einer positiveren Lernatmosphäre im Klassenzimmer, was einen besseren Schulerfolg autistischer Kinder zur Folge hat (Catalano et al., 2020; Love et al., 2019). Neben der Selbstwirksamkeit ist auch der Enthusiasmus von Lehrkräften, also die Begeisterung, mit der die Lehrkraft ihren Beruf bzw. bestimmte Aspekte davon ausübt, ein wichtiger Faktor für erfolgreiches Unterrichten (Kunter et al., 2011).

Wie sich zeigt, spielen die einzelnen Aspekte der professionellen Kompetenz auch speziell für die erfolgreiche Inklusion von Kindern aus dem Autismus-Spektrum eine zentrale Rolle. Die einzelnen Kompetenzaspekte lassen sich zwar separat voneinander betrachten, allerdings scheinen sie sich auch gegenseitig zu bedingen. So haben sowohl das Wissen als auch die Motivation einer Lehrkraft Einfluss auf ihre Überzeugung, während die Überzeugung die Aufnahme von Informationen und somit das Wissen beeinflusst. Die Motivation von Lehrkräften ist wiederum abhängig von dem Wissen, das man über Autismus und im besten Fall auch um autismussensible Unterrichtsstrategien und Handlungsweisen hat. Nach Petersson-Bloom und Holmqvist (2022) ist nämlich das Verständnis darüber, wie inklusiver Unterricht für Autist*innen gestaltet werden kann, ein weiterer Schlüsselaspekt autismussensiblen Unterrichts.

In dem Assessment sollen daher alle Teilaspekte der professionellen Kompetenz erhoben werden. Aus diesen Gründen wird Wissen in dem autismusspezifischem Assessment über die Facetten *Wissen über Autismus* und das spezifische *Wissen um autismussensible Unterrichtsstrategien und Handlungsweisen* erhoben. Beim ersten Teil handelt es sich dabei um einen allgemeinen Wissenstest, der sich nicht berufsspezifisch an Lehrkräfte richtet, während der zweite Teil konkrete unterrichtsbezogene Situationen und Handlungsweisen beschreibt. Außerdem wird die *Überzeugung der Lehrkraft zum Thema Inklusion speziell bei Kindern aus dem Autismus-Spektrum* erhoben. Die Motivation der Lehrkräfte, den Unterricht autismussensibel zu gestalten, wird mit einer Skala zur Erhebung der *Selbstwirksamkeit im Umgang mit autistischen Kindern* und mit einer Skala zur Erfassung des *Enthusiasmus beim inklusiven Unterrichten von autistischen Kindern* untersucht. Der Kompetenzaspekt der Selbstregulation, welcher zwar auch ein bedeutender Faktor bei der professionellen inklusiven Beschulung sein dürfte, wird an dieser Stelle aufgrund seiner Komplexität und einer fehlenden Grundlagenforschung vorerst ausgeklammert.

Welche Kompetenz weisen Lehrkräfte auf?

Die meisten Studien zur Kompetenz von Lehrkräften im Umgang mit autistischen Kindern stammen aus internationalen Kontexten, während für den deutschen Bildungskontext bisher nur wenige empirische Erkenntnisse vorliegen. Eine der wenigen empirischen Befunde liefert hier eine Studie von Wittwer, Hans und Voss (2023). Diese Studie weist darauf hin, dass Lehrkräfte in Deutschland insgesamt nur ein moderates Wissen über Autismus besitzen. Obwohl die Abfrage evidenzbasierter Unterrichtspraktiken für Autist*innen in dieser Studie nur rudimentär

erfolgte, offenbarte sich ein deutliches Defizit an Wissen über diese Methoden. Dafür zeigten die Lehrkräfte eine relativ positive Einstellung gegenüber der Inklusion von Kindern im Autismus-Spektrum (Wittwer et al., 2023). Im Umgang mit Autist*innen und vor allem, wenn diese herausforderndes Verhalten im Unterricht aufweisen, schätzten Lehrkräfte in Deutschland laut dieser Studie ihre Selbstwirksamkeit allerdings nicht besonders hoch ein.

Gerade die Ergebnisse zur Überzeugung von Lehrkräften gegenüber der Inklusion von Kindern im Autismus-Spektrum scheinen im internationalen Vergleich jedoch stark zu variieren. So kommen andere Studien zu dem Schluss, dass bspw. ein Teil der Regelschullehrkräfte das Konzept der Inklusion vollständig ablehnt (Haimour & Obaidat, 2013a).

Bei der Untersuchung der Selbstwirksamkeit von Lehrkräften kommt ein internationaler Vergleich zwischen Lehrkräften aus den USA und Australien von Love und Kolleg*innen (2019) zu einem ähnlichen Ergebnis wie Wittwer und Kolleg*innen (2023). Dort konnte zwar ein Zusammenhang zwischen der allgemeinen Selbstwirksamkeit von Lehrkräften beim Unterrichten in heterogenen Klassen und der Selbstwirksamkeit beim inklusiven Unterrichten von Kindern aus dem Autismus-Spektrum festgestellt werden, insgesamt fühlten sich die Lehrkräfte für diese Aufgabe aber nicht ausreichend vorbereitet (Love et al., 2019).

Bedeutung für unsere Forschung

Auf internationaler Ebene gibt es deutlich mehr erziehungswissenschaftliche und psychologische Studien, die das Wissen von Lehrkräften über Autismus oder die anderen vorgestellten Kompetenzaspekte bei der Inklusion von Kindern im Autismus-Spektrum untersuchen und nahelegen, dass diese Facetten zu einem erfolgreichen inklusiven Unterricht beitragen (Haimour & Obaidat, 2013b; Harrison et al., 2016). Viele Studien weisen zwar auch auf die anderen Aspekte professioneller Kompetenz als Voraussetzung für autismussensiblen Unterricht hin (Gómez-Marí et al., 2022), allerdings gibt es kaum Studien, die diese zusammen erfassen (Barned et al., 2011). Es erweist sich außerdem als schwierig, die gefundenen Studien miteinander zu vergleichen und die Ergebnisse für eine Einschätzung der Situation in deutschen Klassenzimmern heranzuziehen. Hinzu kommt, dass die Studien teilweise stark hinsichtlich der Methodik und der psychometrischen Qualität divergieren. Harrison und Kolleg*innen (2016) fanden in einer Review-Studie bspw. allein 44 verschiedene internationale Instrumente, die das Wissen von Lehrkräften über Autismus erheben. Unsere Review-Studie

(Hasche et al., i. V.), die nachfolgend näher vorgestellt wird, findet auch für die anderen Kompetenzbereiche jeweils diverse Tests und Surveys, die sich vom Aufbau und Inhalt stark unterscheiden.

An dieser Stelle soll diese Arbeit aus der Perspektive der pädagogischen Psychologie ansetzen, indem ein Assessment entwickelt wird, das eine umfangreiche Erfassung aller vorgestellten Aspekte der professionellen Kompetenz, die zu einem autismussensiblen Unterricht beitragen, bietet und diese in Beziehung setzt. Dadurch entsteht ein international anschlussfähiges Instrument, das zur Schließung der dargelegten Forschungslücke in Deutschland beitragen kann. Dieses Instrument kann zum einen in Forschungskontexten eingesetzt oder zur Evaluierung von Interventionen genutzt werden. Bei guter psychometrischer Bewährung ist auch ein Einsatz als Feedbacktool in der Schulpraxis denkbar.

Entwicklung des Assessments

Grundlage für die Konzeption des Instruments ist eine systematische Literaturrecherche. Für den Review-Prozess wurde in 12 internationalen Datenbanken mittels deutscher und englischer Stichwörter nach bestehenden Studien gesucht. Im Fokus standen Studien, die einen oder mehrere Kompetenzaspekte beim Unterrichten von Autist*innen untersuchen. Die über 14.000 Ergebnisse haben wir anhand der Abstracts untersucht und selektiert, wobei nur Studien weiter berücksichtigt wurden, die Autismus bereits als ein Spektrum verstehen und die die Kompetenz von Lehrkräften in inklusiven Schulsettings erheben. Nach diesem Prozess verblieben 52 internationale Studien, die wir mittels Volltextanalyse untersucht haben. Anschließend wurden die Tests und Fragebögen aus den verbleibenden Studien nach dem jeweils untersuchten Kompetenzaspekt geordnet und dann miteinander verglichen.

Wissen

Für das *Wissen über Autismus* ist eines der am häufigsten verwendeten Instrumente der *Autism Stigma and Knowledge Questionnaire* (ASK-Q) von Harrison und Kolleg*innen (2016), welcher ebenfalls auf einem Review basiert. Wittwer, Hans und Voss legten 2023 eine angepasste deutschsprachige Version vor, die bereits eine gute Reliabilität aufweist (Cronbach's $\alpha = 0,77$) und die wir für unser Assessment nutzen konnten (Wittwer et al., 2023). Der ursprüngliche Test unterteilt das Wissen über Autismus in die Subfaktoren Wissen über (1) die Diagnostik von Autismus, (2) die Ätiologie, (3) den Umgang mit Autist*innen und (4) bestehende Stereotype über Autismus und erhebt diese auf einer Antwortskala

mit drei Ausprägungen (*wahr*, *falsch* und *weiß nicht*), wobei diese zur Auswertung dichotomisiert wurden. Da es bei der Erhebung darum ging fehlendes Wissen zu erfassen, wurden die falschen Antworten und die *weiß nicht*-Angaben zusammengefasst. In Ergänzung zu dem Wissenstest von Wittwer und Kolleg*innen (2023) haben wir zusätzlich die Stereotype über Autismus aus dem ASK-Q übersetzt und übernommen. Der von uns auf dieser Basis zusammengestellte Wissenstest setzt sich aus 46 Items zusammen, wobei wir die Antwortskala beibehalten haben. Für das *Wissen um autismussensible Unterrichtsstrategien und Handlungsweisen* haben wir keine bereits validierten Tests bzw. Fragebögen gefunden; daher haben wir hierfür eine eigene Skala entwickelt. Da es derzeit nur begrenzt wissenschaftlichen Konsens darüber gibt, welche Handlungsweisen und Unterrichtsstrategien als effektiv bezeichnet werden können (Pettersson-Bloom & Holmqvist, 2022), haben wir uns hauptsächlich an den *Evidence-Based Practices* von Hume und Kolleg*innen (2021) orientiert, welche eine Liste an Handlungsempfehlungen systematisch aus einem wiederholt durchgeführten Review zusammengetragen haben. Diese wurden mit den Empfehlungen weiterer Studien abgeglichen und ergänzt (u. a. Pettersson-Bloom & Holmqvist, 2022; Theunissen & Sagrauske, 2019). Dabei merken vor allem Theunissen und Sagrauske (2019) an, dass sich keine bisher bekannte Unterrichtsstrategie oder Handlungsempfehlung allgemein auf alle Autist*innen anwenden lässt, sondern diese von Person zu Person individuell erfolgen sollten. Stark pathologisierende Maßnahmen oder Konzepte wie die *Applied Behavior Analysis*, die im deutschen Diskurs mittlerweile weitestgehend abgelehnt werden (Theunissen & Sagrauske, 2019), haben wir bei der Zusammensetzung der Skala außen vor gelassen. Insgesamt haben wir 45 Items erstellt, die verschiedene Strategien und Handlungsweisen vorstellen und auf einer vierstufigen Likert-Skala nach einer Bewertung der eingeschätzten Effizienz in inklusiven Unterrichtssettings fragen (1 = *überhaupt nicht effektiv*, bis 4 = *sehr effektiv*; – 1 = *Ich kenne diese Strategie nicht*). Unter den Strategien wurden auch Fehlkonzepte über effektives Unterrichtshandeln im Kontext von Autismus vorgestellt, die bei der Auswertung entsprechend umkodiert wurden.

Überzeugung

Bei der Auswahl einer geeigneten Skala zur Erhebung der Überzeugung der Lehrkräfte gegenüber der Inklusion von Kindern im Autismus-Spektrum haben wir uns an der *Opinions Relative to Integration of Students with Disabilities* Skala (ORI – modified; Adhabi, 2018) und der *Attitudes toward Inclusion of Student with Autism Scale* (ATISAS; Haimour & Obaidat, 2013a) orientiert. Dabei handelt es sich um modifizierte Abwandlungen von bereits etablierten Fragebögen, die in mehreren Studien verwendet und dadurch wiederholt validiert wurden.

Auch die Reliabilitäten der Studien wiesen auf sehr gute interne Konsistenzen hin (ORI – modified: Cronbach's $\alpha = 0,85$; ATISAS: Cronbach's $\alpha = 0,90$; Adhabi, 2018; Haimour & Obaidat, 2013a). Ein weiteres Auswahlkriterium war eine trennscharfe Erhebung zum Konstrukt der Selbstwirksamkeit. Diese war in vielen untersuchten Studien nämlich nicht gegeben. Insgesamt haben wir 30 Items aus diesen Surveys für unser Assessment adaptiert. Inhaltlich gab es zwischen beiden Skalen viele Überschneidungen, weshalb wir den ATISAS mit seinen 24 Items komplett übersetzt und übernommen haben. Dieser fragt die allgemeine Überzeugung von Lehrkräften bereits sehr umfangreich ab, ergänzend haben wir aber noch sechs weitere Items aus dem ORI – modified Fragebogen adaptiert, die konkret auf unterrichtsinterne Situationen eingehen. Bei allen Items wurde die Meinung zu einer Aussage abgefragt, wobei die Antwortskala an eine vierstufige Likert-Skala angepasst (1 = *Ich stimme nicht zu*; 4 = *Ich stimme voll zu*) wurde.

Motivation

Bei der Zusammensetzung einer motivationalen Skala zur Selbstwirksamkeit bei der inklusiven Beschulung von Autist*innen konnten wir bei der Übersetzung der *The Self-Efficacy for Autism Scale* (TSEAS) von Love (2016) bei neun Items wieder auf die Vorarbeit von Wittwer und Kolleg*innen (2023) zurückgreifen; die restlichen fünf Items haben wir eigenständig übersetzt und übernommen. Um ein breiteres Spektrum an Unterrichtssituationen abzufragen und auch die eingeschätzte Selbstwirksamkeit bei der Kooperation mit den Eltern und weiteren pädagogischen Fachkräften mit zu erfassen, haben wir außerdem acht Items aus der *Teacher Self-Efficacy for Teaching Students With Autism Spectrum Disorder in Inclusive Classrooms Scale* (TSE-ASDI) von Catalano (2018) herangezogen. Insgesamt umfasst unser Fragebogen hier 22 Items, wobei die Befragten gebeten wurden, auf einer vierstufigen Likert-Skala zu bewerten, ob sie der Meinung sind, dass sie eine beschriebene Situation bewältigen könnten (1 = *Ich stimme nicht zu*; 4 = *Ich stimme voll zu*). Zur Erhebung der intrinsischen Motivation haben wir die Enthusiasmus-Skala aus dem COACTIV-Fragebogen von Baumert und Kunter (2011) für das Unterrichten im Kontext von Autismus adaptiert. Hierbei handelt es sich um sechs Items, die auf einer vierstufigen Likert-Skala Einschätzungen erfragen, mit welcher Begeisterung die Lehrkraft inklusiven Unterricht für autistische Kinder vorbereitet und durchführt. Auch hier konnten die Befragten wieder auf einer vierstufigen Likert-Skala Aussagen zustimmen oder ablehnen (1 = *Ich stimme nicht zu*; 4 = *Ich stimme voll zu*).

Zusammen mit demographischen Fragen zur Person und zur bisherigen Erfahrung im Umgang und Unterrichten von Kindern im Autismus-Spektrum umfasst das Assessment in seiner ersten Form 164 Items.

Ziele und Hypothesen der vorliegenden Studie

Ziel der vorliegenden Untersuchung ist es, das Instrument auf seine messtheoretische Güte hin zu überprüfen. Dazu wurden zwei Studien durchgeführt, um das Assessment weiterzuentwickeln und zu validieren. Zunächst fand eine Expert*innenbefragung statt und anschließend wurde eine Pilotstudie mit Lehramtsstudierenden durchgeführt. Folgende Fragen haben den Entwicklungs- und Validierungsprozess dabei geleitet:

- Weisen die Skalen eine zufriedenstellende interne Konsistenz auf und funktionieren sie auch bei Messwiederholungen?
- Ist das Self-Assessment inhaltlich korrekt und verfügt es über einen ausreichenden Bezug zur Praxis?
- Ist das Self-Assessment in der Lage, angenommene signifikante Gruppenunterschiede auch auszuweisen?
- Ist das Instrument in der Lage, adäquat Veränderungen im Wissen über Autismus, der Überzeugung zur inklusiven Beschulung und der wahrgenommenen Selbstwirksamkeit und dem Enthusiasmus abzubilden?

Aus den genannten Fragestellungen haben wir eine Validierungsstrategie mit den folgenden Hypothesen abgeleitet.

Inhaltliche Validierung

H1. Wir erwarten, dass Expert*innen unser Assessment als verständlich und praxisrelevant einschätzen und die Items inhaltlich korrekt sind.

Das Assessment soll später von aktiven Lehrer*innen und pädagogischen Fachkräften verwendet werden und daher gut verständlich sein und einen relevanten Praxisbezug aufweisen. Aus diesem Grund haben wir das Assessment sowohl von Expert*innen aus der Zielgruppe (Lehrkräfte und pädagogische Fachkräfte mit mehrjähriger Erfahrung im inklusiven Unterrichten von autistischen Kindern) als auch von Wissenschaftler*innen mit entsprechendem Forschungsschwerpunkt bewerten lassen.

Reliabilität

H2. Wir gehen davon aus, dass alle Skalen eine interne Konsistenz aufweisen und auch zu zwei verschiedenen Messzeitpunkten stark miteinander korrelieren.

Anhand der Daten des ersten Erhebungszeitpunkts wurde die interne Konsistenz mittels McDonalds Omega (ω) berechnet. Für das Wissen über Autismus wurde aufgrund seiner dichotomen Items zusätzlich die Kuder-Richardson-Formel 20 bestimmt. Für die Kontrollgruppe konnte zudem eine Test–Retest Analyse anhand der zu zwei Messzeitpunkten vorliegenden Daten durchgeführt werden. Dabei gehen wir davon aus, dass die professionelle Kompetenz der Kontrollgruppe im Bereich des autismussensiblen Unterrichtens in der Zwischenzeit nicht systematisch beeinflusst wurde, sodass die Ergebnisse zu beiden Messzeitpunkten hoch miteinander korrelieren sollten.

Validierung anhand anderer Variablen

H3. Angehende Lehrkräfte mit einem förderpädagogischen Schwerpunkt haben sowohl ein besseres Wissen über Autismus und autismussensible Unterrichtsstrategien als auch eine positivere Überzeugung und eine höhere Motivation gegenüber der Inklusion von autistischen Kindern als Regelschul-lehramtsstudierende.

Die Kriteriumsvalidität überprüfen wir mittels der Technik der bekannten Gruppen. Die Annahme von Strauß und König (2017), dass sich die Überzeugung von Förderschullehrkräften zum Unterrichten von heterogenen Klassen signifikant von der Überzeugung aller anderen Lehrkräfte unterscheidet, wurde bereits von Dignath und Kolleg*innen (2020) bei Lehramtsstudierenden nachgewiesen. Wir nehmen an, dass sich dieser Unterschied auch mit Blick auf den inklusiven Unterricht im Kontext von Autismus zeigt und sich nicht nur auf die Überzeugung beschränkt, sondern auch auf andere Kompetenzaspekte überträgt.

Veränderungssensitivität

H4. Wir nehmen an, dass das Instrument in der Lage ist, adäquat Veränderungen in allen Kompetenzaspekten abzubilden.

Zur Überprüfung der Veränderungssensitivität testen wir mittels Mixed-ANOVA, ob unser Assessment in der Lage ist, Veränderungen der professionellen Kompetenz zwischen der Interventions- und der Kontrollgruppe zu beiden Messzeitpunkten zu erfassen. Wir gehen davon aus, dass sich einzelne Kompetenzaspekte in der Interventionsgruppe zwischen den Messzeitpunkten verbessern, während diese in der Kontrollgruppe über die Zeit gleichbleiben.

Methode

Studie 1: Inhaltsvalidierung durch Expert*innenbefragung

Bei der Inhaltsvalidierung via Expert*innenbefragung haben wir überprüft, ob die angepassten Skalen und ihre Items für die Praxis im deutschen Schulkontext verständlich und relevant sind. Dazu haben wir das Assessment von 15 Lehrkräften und pädagogischen Fachkräften einer Schule, die auf die Inklusion von autistischen Kindern spezialisiert ist, ausfüllen und kommentieren lassen. Dabei sollten die Expert*innen besonders auf die Verständlichkeit der Fragen achten und darauf, ob der Umfang der Items die Vielseitigkeit der Praxis gut abdeckt. Die pädagogischen Fachkräfte hatten dabei die Möglichkeit, jedes Item frei zu kommentieren.

Beim zweiten Durchgang wurde das Assessment von fünf Wissenschaftler*innen, die zum Thema Inklusion von Autist*innen aus einer psychologischen oder erziehungswissenschaftlichen Perspektive forschen, zusätzlich hinsichtlich der inhaltlichen Korrektheit unserer Items bewertet, da sich das Wissen über Autismus und die Perspektive im Fachdiskurs in den letzten Jahren stark gewandelt haben (Berdemann, 2023). Die Angaben wurden genutzt, um das Assessment entsprechend zu überarbeiten.

Studie 2: Validierung mit Lehramtsstudierenden

Vorgehen

Für die Testung der Reliabilität und der Validität haben wir in einer Pilotstudie mit Lehramtsstudierenden ($N = 129$) zu zwei Messzeitpunkten empirische Daten erhoben, wobei alle Studierenden zum ersten Messzeitpunkt das angepasste Assessment bearbeitet haben. Eine Teilstichprobe ($n = 75$) nahm eine Woche nach der Durchführung des ersten Messzeitpunkts (t_1) im Rahmen eines

Seminare zusätzlich an einer Intervention in Form eines Lernmoduls über autis-musspezifische Wahrnehmungsbesonderheiten teil (Interventionsgruppe). Eine weitere Woche später bearbeitete diese Interventionsgruppe das Assessment erneut. Eine Kontrollgruppe ($n = 46$), die im gleichen Studiengang, aber nicht an dem besagten Seminar eingeschrieben war, hat nach zwei Wochen ebenfalls am zweiten Messzeitpunkt (t_2) teilgenommen, ohne in der Zwischenzeit an einer Intervention teilgenommen zu haben.

Stichprobe

Die Rekrutierung der Stichprobe für die Interventionsgruppe erfolgte im Rahmen zweier Seminare für das Lehramtsstudium an einer mitteldeutschen Universität. Für die Kontrollgruppe wurde im weiteren Studiengang für die Teilnahme geworben. Insgesamt nahmen 140 Lehramtsstudierende am ersten Messzeitpunkt der Studie teil, wobei 11 Fälle aufgrund fehlender Werte oder auffälligem Antwortverhalten ausgeschlossen wurden. Der finale Datensatz umfasste demnach 129 Personen (82,54 % weiblich; $M_{\text{Alter}} = 22,96$, $SD_{\text{Alter}} = 4,68$). 28,57 % der Teilnehmenden studierten Grundschullehramt, 15,87 % Haupt- und Realschullehramt, 23,02 % Gymnasiallehramt und 32,54 % Förderschullehramt. Bei der Zusammensetzung der Interventions- und Kontrollgruppen (IG und KG) zeigten sich teilweise jedoch starke Unterschiede: So wies die Kontrollgruppe einen deutlich höheren Anteil weiblicher Studentinnen (98 %) im Vergleich zur Interventionsgruppe (73 %) auf, war im Schnitt ein Semester weiter im Studium, und auch der Anteil an Förderschullehramtsstudierenden war mit 50 % deutlich höher als in der IG (21 %). Zudem gaben 82 % der Teilnehmenden aus der KG an, bereits Erfahrungen im Umgang mit Autist*innen zu haben (61 % in der IG). Knapp 46 % haben Autist*innen bereits im Schulalltag begleitet (28 % in der IG). Diese Unterschiede werden bei der Interpretation von den Untersuchungsergebnissen, bei denen Vergleiche zwischen diesen beiden Gruppen interpretiert werden, relevant. Am zweiten Messzeitpunkt haben 120 Personen teilgenommen.

Ergebnisse

Studie 1: Inhaltliche Bewertung

Die Rückmeldungen der insgesamt 20 Expert*innen zu unserem Fragebogen wurden zunächst untereinander abgeglichen und ausgewertet. Ein zentraler Punkt, der hier vor allem von den Expert*innen aus der Praxis angemerkt wurde, ist der der Verallgemeinerung. Bei der ursprünglichen Formulierung einiger Items wurde

nicht deutlich, dass sich die jeweilige Aussage nur auf einen Teil und nicht auf alle Autist*innen bezieht. Somit war eine präzise Beantwortung mancher Fragen in den Wissensskalen nicht möglich. Ein ähnliches Problem wurde auch im Einleitungstext zu den Überzeugungsskalen angemerkt.

Neben den sprachlichen Formulierungen gab es nur wenige Anpassungsvorschläge seitens der Lehrkräfte und der pädagogischen Fachkräfte. Diese wurden größtenteils überdacht und entsprechend geändert. Von den Expert*innen aus der Wissenschaft gab es zudem noch Anmerkungen zur faktischen Korrektheit einiger Items. So wurden acht Items aufgrund neuer wissenschaftlicher Erkenntnisse angepasst. Bspw. erweist es sich inzwischen als Fehlannahme, dass Kinder im Autismus-Spektrum grundsätzlich keine Gesellschaft mögen. Soziale Situationen können zwar einschüchternd auf Autist*innen wirken und abwehrende Reaktionen auslösen, daraus lässt sich aber keine generelle Ablehnung schließen (Theunissen & Sagrauske, 2019). Dieses Item haben wir daher den Items, die Fehlkonzepte erfassen, zugeordnet. Bei anderen Items ist die empirische Evidenz nicht ganz eindeutig, bspw. bei der Aussage, dass die meisten Autist*innen auch eine intellektuelle Entwicklungsstörung aufweisen. Hier liegt das Verhältnis nach einer aktuellen Metaanalyse nur noch bei knapp 50 % (Russell et al., 2019), daher haben wir dieses Item und fünf weitere aus dem Assessment entfernt.

Das auf diese Weise überarbeitete Instrument umfasste 41 Items für die Skala *Wissen über Autismus*, 45 Items für die Skala *Wissen um autismussensible Unterrichtsstrategien und Handlungsweisen* und 29 Items für die *Überzeugung*. Die Anzahl der anderen Skalen hat sich nicht verändert.

Studie 2: Reliabilität und Validierung

Auch wenn es sich bei den einzelnen Kompetenzaspekten um eigenständige Konstrukte handelt, bestanden zwischen den Mittelwerten der einzelnen Skalen gewisse Abhängigkeiten. Dabei zeigte sich, dass alle Skalen signifikant moderat bis stark miteinander korrelieren, ohne auf eine zu hohe Abhängigkeit hinzuweisen (Tab. 6.1).

Reliabilitäten

Die Skala *Wissen über Autismus* wies eine zufriedenstellende interne Konsistenz und Retest-Reliabilität auf (Tab. 6.2). Auch die Reliabilität für die selbst erstellte Skala *Wissen um autismussensible Unterrichtsstrategien und Handlungsweisen* lag im zufriedenstellenden Bereich. Die Motivationsskalen *Selbstwirksamkeit* und *Enthusiasmus* wiesen beide eine hohe interne Konsistenz aus, während

Tab. 6.1 Mittelwerte, Standardabweichungen und Korrelationen zwischen den Skalen

Variable	<i>M</i>	<i>SD</i>	1	2	3	4
1. Wissen über Autismus	0,65	0,13	–			
2. Wissen um a. s. U	2,81	0,46	0,41*	–		
3. Selbstwirksamkeit	2,88	0,51	0,29*	0,58*	–	
4. Enthusiasmus	3,26	0,63	0,27*	0,30*	0,46*	–
5. Überzeugung	2,95	0,41	0,51*	0,39*	0,27*	0,37*

* $p < 0,05$; a. s. U. = autismussensible Unterrichtsstrategien und Handlungsweisen

Tab. 6.2 Reliabilitätsanalyse der Skalen: McDonald's Omega (ω), Kuder Richardson-Formula 20 (K-R-Formula 20) und Retest-Reliabilität (r)

Skala	McDonald's ω	K-R Formula 20	r
1. Wissen über Autismus	0,77	0,74	0,75
2. Wissen um a. s. U	0,93		0,76
3. Selbstwirksamkeit	0,92		0,68
4. Enthusiasmus	0,93		0,71
5. Überzeugung	0,91		0,61

die Retest-Korrelation etwas niedriger war. Auch die *Überzeugung*-Skala hatte eine sehr hohe interne Konsistenz, wobei der Retest-Korrelationskoeffizient mit $r = 0,61$ deutlich unter dem Zielwert von 0,70 lag (vgl. Danner, 2015). Die Ergebnisse zeigen, dass das Instrument insgesamt zufriedenstellende interne Konsistenzen sowie Test–Retest-Reliabilitäten aufweist, hierbei mit Ausnahme der Überzeugungsskala.

Validierung anhand anderer Variablen

Um zu prüfen, ob das Instrument bekannte Gruppen unterscheiden kann, haben wir die durchschnittlichen Skalenwerte (Tab. 6.3) zwischen allen Förderschullehr-
amtsstudierenden (FS) und den Studierenden der Regelschullehramtsstudiengänge (RS) zum ersten Erhebungszeitpunkt mittels t -Test verglichen. Bei der dichotomen Skala *Wissen über Autismus* zeigte sich für die Gruppe der RS ein signifikant niedrigerer Wert als bei der Gruppe der FS, $F[1,124] = 15,35$; $p < 0,001$; $\eta^2 = 0,11$. Auch bei der richtigen Einschätzung von autismussensiblen Handlungsweisen und Unterrichtsstrategien erzielte die Gruppe der FS signifikant höhere Werte, $F[1,124] = 10,33$; $p < .005$; $\eta^2 = 0,08$. Bei den Motivationsskalen fielen

Tab. 6.3 Mittelwerte und Standardabweichungen von Regelschullehramtsstudierenden (RS) und Förderschullehramtsstudierenden (FS)

Skala	RS (N = 85)	FS (N = 41)
	M (SD)	M (SD)
1. Wissen über Autismus	0,62 (0,12)	0,71 (0,12)
2. Wissen um a. s. U	2,72 (0,45)	3,00 (0,42)
3. Selbstwirksamkeit	2,77 (0,51)	3,11 (0,44)
4. Enthusiasmus	3,09 (0,61)	3,61 (0,54)
5. Überzeugung	2,90 (0,43)	3,06 (0,37)

die Unterschiede wieder deutlicher aus. Auch bei der Selbstwirksamkeit wiesen die RS einen signifikant niedrigeren Wert auf als die FS, $F[1,124] = 21,90$; $p < 0,001$; $\eta^2 = 0,09$. Auch gingen die FS mit einem deutlich höheren Enthusiasmus an diese Aufgabe heran als die RS, $F[1,124] = 21,27$; $p < 0,001$; $\eta^2 = 0,15$. Auch wenn die FS hinsichtlich der Überzeugung deskriptiv höhere Werte aufwiesen als die RS, verfehlte dieser Unterschied knapp das Signifikanzniveau, $F[1,124] = 3,93$; $p = 0,05$; $\eta^2 = 0,03$. Insgesamt bestätigen die Ergebnisse unsere Annahme, dass die FS über höhere Kompetenzen bzgl. des Unterrichtens von Kindern im Autismus-Spektrum verfügen als die RS und liefern damit nach der Technik der bekannten Gruppen einen Indikator für die Kriteriumsvalidität unseres Instruments.

Veränderungssensitivität

Bei der Testung der Veränderungssensitivität wurden Interventions- und Kontrollgruppe gegenübergestellt, sodass hier bei der Interpretation die unterschiedliche Zusammensetzung der Gruppen berücksichtigt werden muss. Bei allen Skalen wies die KG zum ersten Messzeitpunkt für alle gemessenen Kompetenzaspekte signifikant höhere Werte auf als die IG (siehe Tab. 6.4).

Im Gegensatz zur IG fand in der KG zu t2 allerdings kein signifikanter Wissenszuwachs statt, während sich das Wissen in der IG über die Zeitpunkte signifikant erhöht hat. Ähnlich fiel auch der Interaktionseffekt bei der Skala zum *Wissen um autismussensible Unterrichtsstrategien und Handlungsweisen* aus, wo bei der KG zwischen den Zeitpunkten keine Veränderung stattgefunden hat, während das Wissen bei der IG zunahm. Auch bei den Ergebnissen der *Selbstwirksamkeits*-Skala zeigte sich eine ähnliche Entwicklung.

Bei der Skala *Enthusiasmus* fand sich kein Interaktionseffekt. Hier ließ sich zwar auch ein Haupteffekt zwischen den Gruppen feststellen, $F(1, 118) = 12,62$, $p < 0,05$; $\eta^2 = 0,1$, allerdings fand weder bei der KG noch bei der IG ein

Tab. 6.4 Veränderungssensitivität: Mittelwerte, Standardabweichungen und Mixed-ANOVA von KG und IG zu beiden Messzeitpunkten

Skala	KG (N = 48)		IG (N = 72)		Effekt	ANOVA		
	M	SD	M	SD		F	df	η^2
1. Wissen über Autismus								
t1	0,70	0,10	0,62	0,13	G x Z	4,859*	1,118	0,40
t2	0,71	0,12	0,67	0,13	KG	0,567	1,47	0,01
					IG	14,274*	1,71	0,17
2. Wissen um a. s. U.								
t1	2,98	0,39	2,71	0,47	G x Z	23,919*	1,118	0,17
t2	3,05	0,40	3,09	0,34	KG	3,235	1,47	0,06
					IG	76,402	1,71	0,52
3. Selbstwirksamkeit								
t1	3,10	0,40	2,77	0,52	G x Z	15,016*	1,118	0,11
t2	3,16	0,37	3,10	0,52	KG	1,911	1,47	0,01
					IG	45,925*	1,71	0,10
4. Enthusiasmus								
t1	3,46	0,61	3,12	0,62	G x Z	0,184	1,118	0,00
t2	3,54	0,38	3,17	0,57	KG	1,796	1,47	0,01
					IG	1,813	1,71	0,00
5. Überzeugung								
t1	3,11	0,30	2,85	0,42	G x Z	2,453	1,117	0,02
t2	3,18	0,26	2,99	0,41	KG	3,915	1,46	0,01
					IG	21,237*	1,71	0,23

Anmerkung. a.s.U.= autismussensible Unterrichtsstrategien und Handlungsweisen. G = Gruppe; Z = Zeit.
* $p < 0,05$.

signifikanter Kompetenzzuwachs über die Zeit statt. Bei der *Überzeugungs*-Skala konnte ebenfalls kein Interaktionseffekt festgestellt werden. Hier zeigte sich neben den Haupteffekten zwischen den Gruppen zusätzlich auch noch ein signifikanter Kompetenzzuwachs innerhalb der IG, $F(1, 117) = 12,45$, $p < 0,001$; $\eta^2 = 0,1$.

Diskussion

Aufgrund steigender Diagnosezahlen und der Umsetzung von Inklusion kann mittlerweile jede Lehrkraft in Deutschland damit rechnen, mindestens einmal in ihrer Laufbahn ein Kind mit Autismus zu unterrichten. Viele Lehrkräfte

fühlen sich darauf allerdings nicht gut vorbereitet. Wie erfolgreich eine Lehrkraft Kinder im Autismus-Spektrum unterrichtet, hängt dabei von verschiedenen Aspekten ihrer professionellen Kompetenz, wie ihrem Wissen, ihrer Überzeugung und ihren motivationalen Orientierungen ab. Diese Kompetenzen gilt es, reliabel und valide zu erfassen, um in wissenschaftlichen Studien Aussagen über die Fähigkeiten von (angehenden) Lehrkräften im Umgang mit Autismus und ihre Kompetenzentwicklung zu treffen.

In der vorliegenden Studie wurden die Konstruktion und die Validierung eines Assessments zur ganzheitlichen Erfassung der professionellen Kompetenz von angehenden Lehrkräften für den inklusiven Unterricht von autistischen Kindern beschrieben. In einer ersten Expert*innenbefragung wurden dem Assessment dabei keine Mängel nachgewiesen, aber einige Items an den aktuellen Wissensstand angepasst. Das abgefragte Wissen basiert dabei auf pädagogisch/psychologischen Quellen, wobei wir davon ausgehen, dass dort medizinisch/neurologische Evidenzen über Autismus berücksichtigt wurden. In der zweiten Studie wiesen alle Skalen eine sehr gute interne Konsistenz auf. Allerdings konnten aufgrund der geringen Fallzahl keine Faktorenanalysen gerechnet werden, die vor allem für die selbst erstellten Skalen nachgeholt werden müssen, um die Voraussetzungen für die Interpretation des McDonalds ω -Wertes zu überprüfen. Die Test-Retest-Korrelationsanalyse mit der KG weist für die meisten Skalen, mit Ausnahme der Überzeugung, ebenfalls recht gute Werte aus. Eine mögliche Erklärung dafür, dass ausgerechnet die Skala *Überzeugung* eine geringere Test-Retest-Reliabilität aufwies als die anderen Skalen liegt evtl. darin begründet, dass die eigenen Überzeugungen überdacht werden, wenn man auf diese angesprochen wird (Grube et al., 1994).

Des Weiteren konnten wir mithilfe unseres Assessments bekannte Gruppenunterschiede in den Kompetenzen zwischen Regelschullehrer- und Förderschullehrerstudierenden (Haimour & Obaidat, 2013a; Strauß & König, 2017) auch mit unserem Instrument replizieren. Für die Kompetenzaspekte *Wissen über Autismus*, *Wissen um autismsensible Unterrichtsstrategien und Handlungsweisen* als auch die *Selbstwirksamkeit* beim inklusiven Unterrichten von Kindern im Autismus-Spektrum zeigte sich eine hohe Veränderungssensitivität des Assessments. Bei der Motivations-Skala, die den *Enthusiasmus* der angehenden Lehrkräfte zum inklusiven Beschulen von autistischen Kindern abfragt, konnte hingegen keine Veränderung gemessen werden. Bei diesem Kompetenzaspekt wiesen allerdings sowohl die Kontrollgruppe als auch die Interventionsgruppe bereits zum ersten Messzeitpunkt einen recht hohen Enthusiasmus auf. Eine Langzeitstudie von Hartl und Holzberger (2022) zeigt allerdings, dass

die Motivation von Lehramtsstudierenden insgesamt deutlich höher ist als bei aktiven Lehrkräften.

Die Überzeugung der beiden Gruppen hat sich ebenfalls nicht signifikant verändert. Davon war bei der durchgeführten Intervention auch nicht zwingend auszugehen, da die Überzeugung als ein recht stabiles Konstrukt gilt (Dignath et al., 2020), das sich eher langfristig durch zunehmende Erfahrungen verändert (Haimour & Obaidat, 2013a). Gómez-Marí und Kolleg*innen (2022) kommen zu dem Ergebnis, dass die Überzeugung stark mit dem Wissen korrelieren sollte und begründen dies damit, dass durch ein besseres Verständnis von Autismus auch die Vorurteile und Stereotype abgebaut würden. Eine Untersuchung von diesen Fehlkonzepten steht allerdings noch aus; daher wurden die Stereotype der Lehrkräfte bei unserer Überzeugungsskala bisher noch nicht berücksichtigt. Zwischen dem Wissen und der Überzeugung von Lehrkräften scheint also ein ambivalenter Zusammenhang zu bestehen, den man ebenfalls noch genauer untersuchen könnte. Insgesamt sehen wir unsere Hypothese, dass das Instrument in der Lage ist, adäquat Veränderungen im Wissen über Autismus, den Überzeugungen zur inklusiven Beschulung und der wahrgenommenen Selbstwirksamkeit abzubilden, an dieser Stelle aber als bestätigt an.

Limitationen und Ausblick

Eine allgemeine Limitation, die bei quantitativen Untersuchungen zum Wissensstand oder der Einschätzung über Autismus zu tragen kommt, ist die der Verallgemeinerung. Fragen zu Eigenschaften von Autist*innen oder zu Handlungsweisen im Umgang mit ihnen treffen nicht auf alle Personen im Spektrum auch zu. Nach dem Vorbild von Harrison et al. (2017) baten wir die Befragten, bei der Beantwortung entsprechender Fragen an einen spezifischen Fall zu denken. An welche Person dabei gedacht wurde und ob ggf. auch eine Komorbidität vorliegt wird nicht erfasst, sodass die Bewertung der Antwort schwierig sein kann. Insgesamt setzt sich die Forschung hier noch nicht kritisch genug mit dem Problem der potenziellen Verallgemeinerung auseinander.

Die hier gefundenen Ergebnisse können nicht ohne weiteres auf Lehrkräfte im Allgemeinen generalisiert werden, da es sich bei den Teilnehmenden der Erhebung für die zweite Studie um Lehramtsstudierende mit zum Teil unterschiedlichen Erfahrungen und Kompetenzen handelt. Insgesamt kann davon ausgegangen werden, dass im Studium ein eher idealisiertes Bild der Praxis entsteht. Die Herausforderungen, die in einer heterogenen Klassengemeinschaft entstehen, können also in der Praxis von der Theorie aus dem Studium recht stark

abweichen (Hartl & Holzberger, 2022). Daher nehmen wir an, dass die Lehramtsstudierenden aus unserer Stichprobe eine verzerrt positivere professionelle Kompetenz aufweisen im Vergleich zu aktiven Lehrkräften, die ggf. auch schon in Klassen mit autistischen Kindern unterrichtet haben. Eine zweite Pilotierung mit aktiven Lehrkräften soll das Assessment darum auch für diese Zielgruppe validieren. Dabei soll zusätzlich noch eine Überzeugungsskala validiert werden, die Fehlkonzepte wie falsches Wissen und Vorurteile mitefasst. Entsprechende Items sind in dem Assessment bereits vorhanden, wurden jedoch in den bisherigen Analysen noch nicht mitberücksichtigt. Zu Beginn wurden außerdem vier professionelle Kompetenzaspekte vorgestellt. In unserem spezifizierten Modell wurde der Aspekt der Selbstregulation von Lehrkräften aufgrund der fehlenden Grundlagenforschung zu diesem spezifischen Thema bisher allerdings ausgeklammert. In nachfolgenden Studien wird dieser Aspekt ebenfalls aufgenommen und untersucht.

Ein weiterer wichtiger Validierungsschritt, der noch ansteht, ist eine Überprüfung der selbst erstellten Skalen mittels explorativer und konfirmatorischer Faktorenanalysen, welche aufgrund der zu geringen Stichprobengröße hier nicht durchgeführt werden konnten. Diese könnten auch zur weiteren Reduktion des bisher noch recht umfangreichen Assessments genutzt werden.

Die Ergebnisse der Studie weisen darauf hin, dass das Assessment das Potenzial besitzt, die eingangs beschriebenen Ziele zu erreichen und damit ein valides Instrument zur Untersuchung der professionellen Kompetenz von angehenden Lehrkräften darstellt. Durch das Assessment lassen sich nicht nur der aktuelle Stand, sondern auch Veränderungen in der professionellen Kompetenz erfassen, was sowohl auf der individuellen Ebene als Self-Assessment zur Überprüfung des persönlichen Kompetenzzuwachses, als auch aufseiten der Wissenschaft zur Evaluation von Interventionen, die auf eine Professionalisierung von Lehrkräften abzielen, genutzt werden kann.

Auch für die Lehramtsausbildung kann das Assessment nützlich sein, da durch die Erhebung der professionellen Kompetenz bei den Studierenden Lücken im Lehrplan identifiziert werden können. Wissen um spezielle Unterrichtsstrategien bei Autist*innen, aber auch bei anderen neurodivergenten Kindern, sollte ein zentrales Element in der Ausbildung, gerade bei Studierenden des Förderschullehramts, sein. Die aktuellen Ergebnisse weisen jedoch darauf hin, dass dieser Aspekt in der Ausbildung scheinbar noch deutlich zu kurz kommt. Das Assessment stellt damit einen ersten und wichtigen Schritt dar, um solche Defizite in den Kompetenzen von Lehramtsstudierenden zu identifizieren und daraus Konsequenzen für die Lehramtsausbildung ableiten zu können.

Danksagung Wir danken Sumeja Mustafi für die Unterstützung bei der Durchführung des Reviews und bei der Erstellung des Assessments, Jessica Witoszek für ihr Mitwirken bei der Auswertung der ersten Pilotstudie sowie Professor Dr. Marcus Hasselhorn, Johanna Stumpf, Lena Wurm und Alisa Börgmann für ihre Unterstützung bei der Durchführung der Datenerhebung zur zweiten Studie! Außerdem danken wir den Lehrkräften und pädagogischen Fachkräften unserer Kooperationsschule für ihre Teilnahme an der Expert*innenbefragung. Das Forschungsprojekt wird durch die Leibniz-Gemeinschaft gefördert.

Literatur

- Adhabi, E. (2018). *The perceptions of elementary school special education and general education teachers on full inclusion of students with Autism Spectrum Disorder (ASD) in Saudi Arabia* (Dissertation, Saint Louis University). ProQuest Dissertations and Theses Global.
- Hasche, N. G., Dignath, C., Mustafi, S., & Kunter, M. (i.V.). An international review of teachers' professional competence to teach autistic children.
- Barned, N. E., Flanagan Knapp, N., & Neuharth-Pritchett, S. (2011). Knowledge and attitudes of early childhood preservice teachers regarding the inclusion of children with autism spectrum disorder. *Journal of Early Childhood Teacher Education*, 32(4), 302–321. <https://doi.org/10.1080/10901027.2011.622235>
- Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–54). Waxmann.
- Berdelmann, K. (2023). Neurodiversität und Wissen über Autismus im pädagogischen Fachdiskurs. Eine historisch vergleichende Perspektive. In C. Lindmeier, M. Grummt, & M. Richter (Hrsg.), *Neurodiversität und Autismus* (S. 29–45). Kohlhammer.
- Berdelmann, K., Schmiedek, F., & Hasselhorn, M. (2021). Inklusion von Kindern mit Autismus-Spektrum-Störung. Inklusion. Chancen und Herausforderungen. In C. Mähler & M. Hasselhorn (Hrsg.), *Tests und Trends der pädagogisch-psychologischen Diagnostik* (Bd. 18, S. 185–201). Hogrefe.
- Catalano, C. G., Fives, H., Barnes, N., & McKeating, E. (2020). Preservice early childhood teachers' sense of efficacy for teaching children with autism spectrum disorder. *Journal of Early Childhood Education*, 43(2), 167–193. <https://doi.org/10.1080/10901027.2020.1806152>
- Catalano, C. G. (2018). *Developing and validating the Teacher Self-Efficacy for Teaching Students with Autism Spectrum Disorder in Inclusive Classrooms (TSE-ASDI) Scale* (Dissertation, Montclair State University). ProQuest Dissertation & Theses Global.
- Danner, D. (2015). Reliabilität – die Genauigkeit einer Messung. GESIS Survey Guidelines. *GESIS – Leibniz-Institut für Sozialwissenschaften*. https://doi.org/10.15465/gesis-sg_011
- Dignath, C., Meschede, N., Kunter, M., & Hardy, I. (2020). Empirische Arbeit: Ein Fragebogen zur Erfassung von Überzeugungen Lehramtsstudierender zum Unterrichten in

- heterogenen Klassen: Befunde zur Kriteriumsvalidität und Veränderungssensitivität. *Psychologie in Erziehung und Unterricht*, 67(3), 194–211. <https://doi.org/10.2378/peu2020.art16d>
- Dignath, C., Rimm-Kaufman, S., van Ewijk, R., & Kunter, M. (2022). Teachers' beliefs about inclusive education and insights on what contributes to those beliefs: A meta-analytical study. *Educational Psychology Review*, 34(4), 2609–2660. <https://doi.org/10.1007/s10648-022-09695-0>
- Gómez-Marí, I., Sanz-Cervera, P., & Tárraga-Mínguez, R. (2022). Teachers' attitudes toward autism spectrum disorder: A systematic review. *Education Sciences*, 12(2), 1–21. <https://doi.org/10.3390/educsci12020138>
- Grube, J. W., Mayton, D. M., & Ball-Rokeach, S. J. (1994). Inducing change in values, attitudes, and behaviors: Belief system theory and the method of value self-confrontation. *Journal of Social Issues*, 50(4), 153–173. <https://doi.org/10.1111/j.1540-4560.1994.tb01202.x>
- Haimour, A. I., & Obaidat, Y. F. (2013a). School teachers' attitude toward inclusion of students with autism in Saudi Arabia. *International Journal of Current Research*, 5(10), 2970–2977.
- Haimour, A. I., & Obaidat, Y. F. (2013b). School teachers' knowledge about autism in Saudi Arabia. *World Journal of Education*, 3(5), 45–56. <https://doi.org/10.5430/wje.v3n5p45>
- Harrison, A. J., Bradshaw, L. P., Naqvi, N. C., Paff, M. L., & Campbell, J. M. (2017). Development and psychometric evaluation of the Autism Stigma and Knowledge Questionnaire (ASK-Q). *Journal of Autism and Developmental Disorders*, 47(10), 3281–3295. <https://doi.org/10.1007/s10803-017-3242-x>
- Harrison, A. J., Slane, M. M., Hoang, L., & Campbell, J. M. (2016). An international review of autism knowledge assessment measures. *Autism*, 21(3), 262–275. <https://doi.org/10.1177/1362361316638786>
- Hartl, A., & Holzberger, D. (2022). Identifying teachers' motivational profiles and their changes from teacher education into practice: A longitudinal study. *Zeitschrift für Erziehungswissenschaft*, 25(2), 427–451. <https://doi.org/10.1007/s11618-022-01093-0>
- Hume, K., Steinbrenner, J. R., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendrey, S., McIntyre, N. S., Yücesoy-Özkan, S., & Savage, M. N. (2021). Evidence-based practices for children, youth, and young adults with autism: Third generation review. *Journal of Autism and Developmental Disorders*, 51(11), 4013–4032. <https://doi.org/10.1007/s10803-020-04844-2>
- Jordan, R., Roberts, J. M., & Hume, K. (Hrsg.). (2019). *The SAGE handbook of autism and education*. SAGE Publications.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Waxmann.
- Lindmeier, C. (2018). Kinder und Jugendliche aus dem Autismus-Spektrum in der Schule – Forschungsfelder und Forschungsdesiderate. *Zeitschrift für Heilpädagogik*, 69(9).
- Love, A. M. A. (2016). *Development of a self-efficacy scale for teachers who teach kids with Autism Spectrum Disorder* (Masterarbeit, University of Kentucky). Theses and Dissertations – Educational, School, and Counseling Psychology.
- Love, A. M. A., Toland, M. D., Usher, E. L., Campbell, J. M., & Spriggs, A. D. (2019). Can I teach students with Autism Spectrum Disorder? Investigating teacher self-efficacy with

- an emerging population of students. *Research in Developmental Disabilities*, 89, 41–50. <https://doi.org/10.1016/j.ridd.2019.02.005>
- Petersson-Bloom, L., & Holmqvist, M. (2022). Strategies in supporting inclusive education for autistic students: A systematic review of qualitative research results. *Autism & Developmental Language Impairments*, 7, 1–15. <https://doi.org/10.1177/23969415221123429>
- Rackles, M. (2023). *Inklusionspraxis in Deutschland. Ein Armutszeugnis für die Bildungspolitik*. Deutsches Schulportal. <https://deutsches-schulportal.de/expertenstimmen/inklusionspraxis-in-deutschland-ein-armutszeugnis-fuer-die-bildungspolitik-deutsches-schulbarometer-juni-2023/>
- Robert Bosch Stiftung (2023). *Das Deutsche Schulbarometer. Aktuelle Herausforderungen aus Sicht der Lehrkräfte. Ergebnisse einer Befragung von Lehrkräften allgemein- und berufsbildender Schulen*. Robert Bosch Stiftung.
- Russell, G., Mandy, W., Elliott, D., White, R., Pittwood, T., & Ford, T. (2019). Selection bias on intellectual ability in autism research: A cross-sectional review and meta-analysis. *Molecular Autism*, 10(1). <https://doi.org/10.1186/s13229-019-0260-x>
- Sparapani, N., Morgan, L., Reinhardt, V. P., Schatschneider, C., & Wetherby, A. M. (2016). Evaluation of classroom active engagement in elementary students with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 46(3), 782–796. <https://doi.org/10.1007/s10803-015-2615-2>
- Strauß, S., & König, J. (2017). Berufsbezogene Überzeugungen von angehenden Lehrkräften zur inklusiven Bildung. *Unterrichtswissenschaft*, 45(4), 243–261.
- Theunissen, G., & Sagrauske, M. (2019). *Pädagogik bei Autismus. Eine Einführung*. Kohlhammer.
- van der Steen, S., Geveke, C. H., Steenbakkens, A. T., & Steenbeek, H. W. (2020). Teaching students with autism spectrum disorders: What are the needs of educational professionals? *Teaching and Teacher Education*, 90. <https://doi.org/10.1016/j.tate.2020.103036>
- Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften. Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18, 187–223. <https://doi.org/10.1007/s11618-015-0626-6>
- Wittwer, J. (2022). Autismus. Ein Überblick über Lernvoraussetzungen und Fördermaßnahmen. *Zeitschrift für Pädagogische Psychologie*, 36(1–2), 19–36. <https://doi.org/10.1024/1010-0652/a000326>
- Wittwer, J., Hans, S., & Voss, T. (2023). Inclusion of autistic students in schools. Knowledge, self-efficacy, and attitudes of teachers in Germany. *Autism*. <https://doi.org/10.1177/13623613231220210>

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Teil III

Anforderungen an eine adaptive Schule aus Sicht der Schüler*innen



Zur Rolle von Überzeugungen zu Mehrsprachigkeit für die Akzeptanz und den Lernerfolg eines mehrsprachigkeitssensiblen Unterrichtskonzepts

Victoria Bertram, Dominique Rauch, Valentina Reitenbach, Katharina M. Schneider und Jasmin Decristan

Zusammenfassung

Positive Schüler*innen-Überzeugungen zu Mehrsprachigkeit werden als wichtige Voraussetzung für den Erfolg mehrsprachigkeitssensibler Unterrichtskonzepte angesehen. Bisher wurde jedoch nicht empirisch untersucht, wie

V. Bertram (✉)

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt am Main, Deutschland

E-Mail: v.bertram@dipf.de

D. Rauch · K. M. Schneider

Pädagogische Hochschule Ludwigsburg, Ludwigsburg, Deutschland

E-Mail: dominique.rauch@ph-ludwigsburg.de

K. M. Schneider

E-Mail: katharina.m.schneider@ph-ludwigsburg.de

V. Reitenbach · J. Decristan

Bergische Universität Wuppertal, Wuppertal, Deutschland

E-Mail: reitenbach@uni-wuppertal.de

J. Decristan

E-Mail: decristan@uni-wuppertal.de

V. Bertram · D. Rauch · V. Reitenbach · J. Decristan

IDeA – Center for Research on Individual Development and Adaptive Education of Children at Risk, Frankfurt am Main, Germany

diese Überzeugungen den Erfolg solcher Konzepte beeinflussen. Der vorliegende Beitrag schließt an diese Forschungslücke an, indem untersucht wird, ob und wie Schüler*innen-Überzeugungen zu Mehrsprachigkeit die Akzeptanz und den Lernerfolg (Leseverstehen und Selbstwirksamkeitserwartungen) einer mehrsprachigkeitssensiblen Intervention vorhersagen. Hierzu wurden Daten von 785 Grundschüler*innen, die an einer Interventionsstudie teilgenommen haben, analysiert. Die Ergebnisse zeigen, dass positivere Überzeugungen zu Mehrsprachigkeit bei allen Schüler*innen mit einer höheren Akzeptanz der Maßnahme einhergehen. Die Akzeptanz vermittelt zudem den Zusammenhang zwischen den Überzeugungen zu Mehrsprachigkeit und Selbstwirksamkeitserwartungen der Schüler*innen. Die Ergebnisse unterstreichen die Bedeutung positiver Überzeugungen zu Mehrsprachigkeit für eine hohe Akzeptanz mehrsprachigkeitssensibler Unterrichtskonzepte.

Schlüsselwörter

Mehrsprachigkeit • Implementation • Akzeptanz • Grundschule

Einleitung

Im Jahr 2022 wiesen fast die Hälfte (42,6 %) der Grundschüler*innen in Deutschland einen Migrationshintergrund auf (Destatis, 2022). Zudem gab der Großteil (66,4 %) der Schüler*innen mit Migrationshintergrund zwischen 6 und 15 Jahren an, zu Hause auch andere Sprachen als Deutsch zu sprechen (Destatis, 2022). Dies zeigt, dass im Alltag vieler Schüler*innen unterschiedliche Familiensprachen präsent sind. In deutschen Schulklassen kommt dementsprechend eine Vielzahl unterschiedlicher Familiensprachen und Kombinationen migrationsbedingter Mehrsprachigkeit zusammen (Henschel et al., 2023). Im Unterricht wird dieser *superdiverse* Sprachkontext (Prediger et al., 2021) jedoch nur selten aufgegriffen. Der sogenannte *monolinguale Habitus* (Gogolin, 2008) ist auch 2024 in Deutschland prägend für die Sprachpolitik der meisten Schulen – Deutsch ist die alleinige Unterrichts- und Alltagssprache und die Familiensprachen werden aus dem Regelunterricht ausgeschlossen. Lehrkräfte befürchten häufig negative Auswirkungen auf das Lernen, wenn Familiensprachen im Unterricht genutzt werden (Alisaari et al., 2019; Ticheloven et al., 2021). Umgekehrt wird die Aktivierung mehrsprachiger Ressourcen in der Theorie jedoch als lernförderlich konzeptualisiert, unter der Annahme, dass sie Kommunikationsmöglichkeiten erweitert,

sprachliche Barrieren abbaut und individuelle Potenziale der Schüler*innen anerkennt (Grosjean, 2013; Krumm, 2016).

Bildungsstudien zeigen, dass Schüler*innen, die im familiären Sprachgebrauch vorwiegend (eine) andere Sprache(n) als Deutsch sprechen, schwächere Leistungen im Lesen erzielen als Schüler*innen, die zuhause vorwiegend Deutsch sprechen (vgl. Wendt & Schwippert, 2017). Da schwache Leseleistungen als Risikofaktor für den weiteren Bildungsverlauf gelten und das Leseverstehen eine Schlüsselkompetenz am Ende der Grundschulzeit darstellt, eignet sich die Lesekompetenz besonders zur Förderung durch mehrsprachigkeitssensibles Unterrichten. Neben kognitiven Prozessen des Leseverstehens und der Anwendung (meta)kognitiver Lesestrategien (Artelt et al., 2002), gelten auch erfolgreiche motivationale Prozesse, wie die Lesemotivation und Selbstwirksamkeitserwartungen, als Facette kompetenten Lesens (Guthrie et al., 2004).

Mehrsprachigkeitssensible Konzepte, die es ermöglichen, Familiensprachen in den Unterricht zu integrieren, sind selten und wurden, wenn sie umgesetzt wurden, oft nicht evaluiert (García & Wei, 2014; Roth et al., 2021). Eine wichtige Voraussetzung zur Umsetzung solcher Konzepte wird in positiven Überzeugungen zu Mehrsprachigkeit der Akteur*innen gesehen (Lorenz et al., 2021). Bisher ist jedoch nur wenig darüber bekannt, wie Schüler*innen die Integration von Familiensprachen im Unterricht wahrnehmen (Reitenbach et al., 2023; Schastak, 2020) und die Perspektive der Schüler*innen wird in Implementationsstudien häufig vernachlässigt (vgl. Schrader et al., 2020). Vor dem Hintergrund konstruktivistischer Lerntheorien ist diese jedoch besonders relevant, da der Lernerfolg im Unterricht maßgeblich von individuellen Wahrnehmungs- und Nutzungsprozessen der Schüler*innen abhängt (Vieluf et al., 2020). Der vorliegende Beitrag beschäftigt sich daher mit der Bedeutung von Schüler*innen-Überzeugungen zu Mehrsprachigkeit als Voraussetzung für das Gelingen eines mehrsprachigkeitssensiblen Unterrichtskonzepts zur Förderung des Leseverstehens deutscher Texte. Schüler*innen-Überzeugungen zu Mehrsprachigkeit werden als Prädiktor für die Akzeptanz des Konzepts durch die Lernenden sowie den Lernerfolg (operationalisiert über das Leseverstehen und die Selbstwirksamkeit in der Lesestrategieanwendung) untersucht.

Theoretischer Hintergrund

Mehrsprachigkeitssensibles Unterrichten

Mehrsprachige Personen kommunizieren im Alltag häufig unter Nutzung ihres gesamten sprachlichen Repertoires und passen die sprachlichen Handlungen sowohl an eigene Präferenzen als auch an die Erfordernisse der Situation an. Nach Grosjean (2020) verfügen mehrsprachige Sprecher*innen über verschiedene Sprachmodi auf einem Kontinuum zwischen Monolingualität und Mehrsprachigkeit, die nach individueller Kompetenz, Interaktionspartner*innen und sozialem Kontext gewählt werden. Diese Faktoren für eine (Nicht-)Nutzung des vollständigen Sprachrepertoires zeigen sich auch in den Selbstberichten von Schüler*innen, die mit einem mehrsprachigkeitssensiblen Ansatz unterrichtet wurden (Reitenbach et al., 2023). Im Unterricht kann ein besseres Verständnis der Lerninhalte bspw. durch kurze Übersetzungen oder mehrsprachige Verhandlung fachlicher Inhalte erzielt werden (Moschkovic, 2007; Planas & Setati, 2009; Schüler-Meyer et al., 2019). Mehrsprachige Kommunikation kann im Klassenzimmer in nach sprachlichen Voraussetzungen gebildeten Tandems (Decristan et al., 2021; Schastak, 2020) oder Kleingruppen (Decristan et al., 2022) ermöglicht werden.

Konzepte zum Einbezug lebensweltlicher Mehrsprachigkeit im Unterricht fußen häufig auf Strategien pädagogischen Translanguagings (vgl. Lorenz et al., 2021). Das Konzept des *Translanguaging* fokussiert unter einer ganzheitlichen Perspektive von Mehrsprachigkeit die natürliche Sprachpraxis mehrsprachiger Sprecher*innen, die nicht an einer monolingualen Norm ausgerichtet ist. Sprachen werden nicht als getrennte Systeme angenommen, sondern bilden ein gemeinsames Repertoire (García & Wei, 2014). Dabei können Sprachen miteinander interagieren, co-existieren und simultan verwendet werden, sodass Kommunikationssituationen je nach Interaktionspartner*innen und sozialer Situation so effektiv wie möglich gestaltet werden können. Translanguaging soll dabei auch Vorteile für das Lernen mit sich bringen, indem ein tieferes Verständnis der Unterrichtsinhalte und die Entwicklung aller Sprachen gefördert wird (García & Wei, 2014). Die Etablierung einer wertschätzenden Atmosphäre für Translanguaging wird als Kernelement einer erfolgreichen Umsetzung gesehen (García et al., 2017).

Pädagogische Konzepte des Translanguagings bieten konkrete Impulse zur Integration von Familiensprachen beim Lernen und greifen die natürliche Sprachpraxis der Schüler*innen auf, ohne getrennte Lerngelegenheiten (z. B. in herkunftssprachlichem Unterricht) schaffen zu müssen. Kernelemente pädagogischen Translanguagings umfassen u. a. einen ganzheitlichen Blick auf Mehrsprachigkeit

entgegen einer monolingualen Norm, die Wertschätzung aller sprachlichen Ressourcen und die Förderung positiver Überzeugungen zu Mehrsprachigkeit (Lorenz et al., 2021). Diesen Kernelementen kommt eine wichtige Funktion bei mehrsprachigkeitssensiblen Unterrichten zu, da die Sensibilisierung für sprachliche Vielfalt dem monolingualen Habitus (Gogolin, 2008) entgegenwirken und die Schüler*innen zur Nutzung ihrer Familiensprachen einladen soll. Schüler*innen sollen dadurch in ihrer Identitätsentwicklung gestärkt werden und weniger Angst vor Ausgrenzung oder Stigmatisierung durch die Nutzung ihrer Familiensprachen haben (Lucas et al., 2008; Sierens & van Avermaet, 2014). Positive Effekte mehrsprachigkeitssensiblen Unterrichts wurden für sprachliche Kompetenzen, wie die Förderung des Wortschatzes (Bosma et al., 2020) und metalinguistisches Bewusstsein (Leonet et al., 2020), sowie auf motivationaler Ebene im Selbstvertrauen (Cenoz et al., 2022) und der Identitätsentwicklung mehrsprachiger Schüler*innen (Ebe & Chapman-Santiago, 2016) beobachtet.

Überzeugungen zu Mehrsprachigkeit

Positive Überzeugungen zu Mehrsprachigkeit stellen in Anlehnung an pädagogisches Translanguaging ein konzeptuelles Kernelement mehrsprachigkeitssensiblen Unterrichts dar (García et al., 2017; Lorenz et al., 2021). Dies zeigt sich empirisch auch darin, dass negative Überzeugungen zu Mehrsprachigkeit die Nutzung von Familiensprachen im Unterricht hemmen können (Prediger et al., 2021; Reitenbach et al., 2023; Sierens & van Avermaet, 2014) und somit einer gelungenen Implementation mehrsprachigkeitssensibler Konzepte entgegenstehen. Sowohl Schüler*innen als auch Lehrkräfte argumentieren in Deutschland häufig, dass Familiensprachen zugunsten eines höheren Lernerfolgs im Unterricht ausgeschlossen werden sollten (Wamhoff et al., 2022). Da aus wissenschaftlicher (vgl. Fürstenau, 2011) und politischer Perspektive (KMK, 2015) jedoch die Integration von Familiensprachen in den *Regelunterricht* gefordert wird, sind sowohl die Überzeugungen mehrsprachiger als auch monolingualer Schüler*innen von Relevanz. Haukas et al. (2022) beobachten bei norwegischen Schüler*innen, dass sich die Überzeugungen zum Nutzen von Mehrsprachigkeit für das Lernen nicht signifikant zwischen monolingualen und mehrsprachigen Schüler*innen unterscheiden. Beim Vergleich zwischen mehrsprachigen Schüler*innen, deren Erstsprache Norwegisch ist und solchen mit einer anderen Erstsprache, zeigen sich jedoch relevante Unterschiede. Schüler*innen mit einer anderen Erstsprache bewerten Mehrsprachigkeit dabei signifikant positiver.

Die Forschung zur Wahrnehmung mehrsprachigkeitssensibler Unterrichtskonzepte ist noch begrenzt. Studien sind in diesem Kontext meist qualitativ ausgelegt und umfassen kleine Stichproben, die Ergebnisse sind dabei gemischt. Aus positiven Ergebnissen geht hervor, dass sowohl mehrsprachige als auch monolinguale Schüler*innen großes Interesse am mehrsprachigkeitssensiblen Unterricht zeigen (Schnitzer, 2020) und die Nutzung von Familiensprachen im Unterricht von allen Schüler*innen positiv bewertet wird (Hilbe et al., 2017). Am Beispiel bilingualen Unterrichts stellt Mulder (2019) jedoch fest, dass monolingual aufwachsende Schüler*innen bilingualen Unterricht als anstrengender bewerten als bilingual aufwachsende Schüler*innen.

Auch der Zusammenhang zwischen Schüler*innen-Überzeugungen zu Mehrsprachigkeit und Lernerfolg in mehrsprachigkeitssensiblen Unterrichtskonzepten wurde bislang kaum untersucht. Im schulischen Kontext können Überzeugungen von Schüler*innen jedoch einen bedeutsamen Einfluss auf den Lernerfolg haben. Als individuelle Merkmale beeinflussen sie nach der Logik von Angebot-Nutzungsmodellen der Unterrichtsforschung (vgl. Vieluf et al., 2020) Prozesse der Wahrnehmung und Informationsverarbeitung. Höhere Motivation und höheres Engagement in der Bearbeitung der Aufgaben können so den Lernerfolg der Schüler*innen positiv beeinflussen (vgl. Wigfield & Eccles, 2000). Positive Lernerfahrungen können wiederum das Erleben von Selbstwirksamkeit fördern (Bandura, 1977; Phan & Ngu, 2016). Ebenso kann jedoch angenommen werden, dass das Leistungsniveau als individuelle Lernvoraussetzung Verarbeitungsprozesse beeinflusst und in Interaktion mit dem Unterrichtsangebot Überzeugungen geformt werden. Theoretisch sind daher von reziproken Zusammenhängen zwischen Überzeugungen und Lernerfolg auszugehen und in bisherigen Studien werden auch unterschiedliche Wirkrichtungen untersucht. Metaanalysen zeigen bspw., dass Überzeugungen zum Unterrichtsgegenstand in positivem Zusammenhang mit dem Leistungsniveau im naturwissenschaftlichen Unterricht (Mao et al., 2021) und im Lesen (Petscher, 2010) stehen. In einer kleinen quantitativen Untersuchung beobachten Öztürk und Çubukçu (2022) jedoch, dass das Leistungsniveau türkischer Studierender ihre Überzeugungen zum Translanguaging im Englischunterricht nicht vorhersagt.

Implementationserfolg und Akzeptanz didaktischer Konzepte

Damit innovative didaktische Konzepte, wie z. B. zum mehrsprachigkeitssensiblen Unterrichten, in der Unterrichtspraxis erfolgreich und langfristig angewendet

werden, müssen diese nicht nur wirksam, sondern auch gut umsetzbar und praxiskompatibel sein. Im schulischen Kontext wird der Erfolg von Implementationsprozessen häufig aus der Sicht der Lehrkräfte beurteilt, die eine Innovation im Unterricht umsetzen. Konstruktivistische Lerntheorien legen jedoch nahe, dass auch die Schüler*innen den Unterricht aktiv mitgestalten und die Einführung einer Innovation nicht allein durch das Handeln der Lehrkraft bestimmt wird (Schrader et al., 2020). Daher ist es von Bedeutung, dass die Implementationsforschung auch Prozesse auf Schüler*innenebene berücksichtigt.

Zahlreiche Studien zeigen, dass Interventionen besonders erfolgreich sind, wenn sie möglichst genau nach ihrer Konzeption umgesetzt und von den Teilnehmer*innen akzeptiert werden (Durlak & DuPre, 2008; O'Donnell, 2008). Die Akzeptanz einer Intervention stellt damit eine wichtige Facette für Implementationserfolg dar, die auf Schüler*innenebene jedoch bislang vernachlässigt wird (Schrader et al., 2020). Akzeptanz wird im Rahmen der Implementationsforschung definiert als „die Auffassung der an der Implementation Beteiligten, dass, oder in welchem Maße, eine bestimmte Intervention, Behandlung etc. zufriedenstellend ist“ (Petermann, 2014, S. 123). Angelehnt an das Konstrukt *Treatment Acceptability* kann Akzeptanz detaillierter als einstellungsähnliches Konstrukt definiert werden, das als Gesamturteil einer Intervention auf affektiver und kognitiver Ebene zu verstehen ist (Eckert et al., 2021; Kazdin, 1980). Vor dem Hintergrund psychologischer Definitionen von Einstellungen spiegelt die affektive Dimension Gefühle und Emotionen wider, während auf kognitiver Ebene Überzeugungen und Wissen über das Einstellungsobjekt zusammengefasst werden (Haddock & Maio, 2014). Eckert et al. (2017) untersuchen die Akzeptanz von Grundschüler*innen gegenüber einer Intervention zur Förderung der Schreibfähigkeiten und identifizieren in ihrem Instrument zwei Faktoren: die affektive (Freude an der Intervention) und kognitive Akzeptanz (subjektive Wirksamkeitserwartung). Entsprechend wird im vorliegenden Beitrag eine affektive und kognitive Dimension der Akzeptanz differenziert.

Die Akzeptanz gegenüber einer Intervention hängt nachweislich mit der Wirksamkeit der Intervention zusammen (vgl. Durlak & DuPre, 2008). Auch im *Model of Treatment Acceptability* (Witt & Elliott, 1985), das als Heuristik zur Untersuchung von Implementationsprozessen gilt, werden Akzeptanz und Wirksamkeit in positivem und reziprokem Zusammenhang beschrieben. Eine Intervention, die von den Teilnehmer*innen akzeptiert wird, ist demnach eher wirksam. Ebenfalls kann eine hohe Wirksamkeit der Intervention zu hoher Akzeptanz der Teilnehmer*innen beitragen. Aus Perspektive der Unterrichtsforschung ist anzunehmen, dass die Akzeptanz von Unterrichtsinterventionen als Wahrnehmungsprozess auf Ebene der Schüler*innen durch individuelle Lernvoraussetzungen beeinflusst

wird. Die Wahrnehmung und Interpretation von Unterricht kann weiterhin kognitive und motivational-affektive Lernprozesse beeinflussen (Vieluf et al., 2020), sodass die Akzeptanz auch auf den Lernerfolg einwirken kann (Schrader et al., 2020). Es gibt jedoch nur wenige Untersuchungen dieser Zusammenhänge, die mit Interventionen im unterrichtlichen Kontext durchgeführt wurden. In einer Studie von Eckert et al. (2021) zeigen sich differenzielle Zusammenhänge der Akzeptanzdimensionen zum Lernerfolg. Die affektive Akzeptanz hängt positiv mit dem Lernerfolg zusammen, während die kognitive Akzeptanz den Lernerfolg nicht vorhersagt.

Studien aus unterschiedlichen Forschungsfeldern zeigen, dass die Merkmale von Teilnehmer*innen, wie z. B. Einstellungen, Erfahrungen und Kompetenzen, die Akzeptanz beeinflussen (Allinder & Oats, 1997; Venkatesh et al., 2003). Als förderliche Faktoren für eine positive Wahrnehmung einer Intervention werden in einer Systematisierung von Forschungsergebnissen das Geschlecht sowie motivationale Merkmale der Schüler*innen zusammengefasst (Schrader et al., 2020). Es ist demnach anzunehmen, dass Schüler*innen, die positive Überzeugungen zu Mehrsprachigkeit mitbringen, eine Unterrichtseinheit, die explizit darauf ausgerichtet ist, Mehrsprachigkeit in der Schüler*innen-Interaktion zuzulassen, wertzuschätzen und zu fördern, diese besonders positiv wahrnehmen. Weiterhin ist anzunehmen, dass ein positives Erleben der Intervention auf affektiver und kognitiver Ebene mit lernförderlichen Nutzungsprozessen, z. B. einem höheren Engagement bei der Bearbeitung der Aufgaben, in Zusammenhang steht und die Akzeptanz als Wahrnehmungsprozess auf diese Weise zwischen individuellen Merkmalen der Schüler*innen und dem Lernerfolg vermittelt (Schrader et al., 2020; Vieluf et al., 2020).

Fragestellungen und Hypothesen

Positive Überzeugungen zu Mehrsprachigkeit werden als Kernelement mehrsprachigkeitssensibler Unterrichtskonzepte beschrieben (Lorenz et al., 2021). Bislang gibt es jedoch nur wenige Studien dazu, wie Schüler*innen-Überzeugungen zu Mehrsprachigkeit mit der Wahrnehmung und dem Lernerfolg mehrsprachigkeitssensiblen Unterrichtens zusammenhängt. Der vorliegende Beitrag knüpft hieran an und untersucht die Bedeutung von Überzeugungen zu Mehrsprachigkeit für den Implementations- und Lernerfolg eines mehrsprachigkeitssensiblen Unterrichtskonzepts. Als Facette von Implementationserfolg wird die Akzeptanz der Schüler*innen untersucht. Der Lernerfolg wird am Leseverstehen und der Selbstwirksamkeit der Lesestrategianwendung gemessen.

Im vorliegenden Beitrag wird angenommen, dass positive Überzeugungen zu Mehrsprachigkeit dazu beitragen, dass eine mehrsprachigkeitssensible Unterrichtseinheit positiv wahrgenommen wird und sich dies in einer höheren Ausprägung an affektiver und kognitiver Akzeptanz zeigt. Weiterhin wird angenommen, dass die Akzeptanz als Wahrnehmungsprozess zwischen Überzeugungen zu Mehrsprachigkeit und dem Lernerfolg vermittelt (Schrader et al., 2020). Da mehrsprachigkeitssensible Konzepte in den Regelunterricht integriert werden sollen, ist nicht nur die Wahrnehmung mehrsprachiger Schüler*innen relevant, sondern auch die der monolingualen Schüler*innen. Bisherige Forschung deutet darauf hin, dass sich Überzeugungen zu Mehrsprachigkeit und die Wahrnehmung mehrsprachigkeitssensiblen Unterrichts zwischen monolingualen und mehrsprachigen Schüler*innen unterscheiden könnte.

Vor diesem Hintergrund werden folgende Forschungsfragen und Hypothesen formuliert:

- 1) Sagen Überzeugungen zu Mehrsprachigkeit die Akzeptanz der mehrsprachigkeitssensiblen Intervention vorher?

(H1) Es wird angenommen, dass Überzeugungen zu Mehrsprachigkeit zu Beginn der Intervention einen positiven Effekt auf die Akzeptanz haben.

- 2) Sagen Überzeugungen zu Mehrsprachigkeit den Lernerfolg der Schüler*innen, gemessen am Leseverstehen und der Selbstwirksamkeit der Lesestrategieanwendung, im Kontext einer mehrsprachigkeitssensiblen Intervention vorher?

(H2) Es wird angenommen, dass Überzeugungen zu Mehrsprachigkeit zu Beginn der Intervention einen positiven Effekt auf das (a) Leseverstehen und (b) die Selbstwirksamkeit der Lesestrategieanwendung am Ende der Intervention haben.

- 3) Vermittelt die Akzeptanz der Intervention den Zusammenhang zwischen Überzeugungen zu Mehrsprachigkeit und Lernerfolg?

(H3) Es wird angenommen, dass die Akzeptanz den Zusammenhang zwischen Überzeugungen zu Mehrsprachigkeit und dem Lernerfolg im (a) Leseverstehen und (b) der Selbstwirksamkeit zur Lesestrategieanwendung vermitteln.

Zusätzlich soll explorativ untersucht werden, ob sich die Zusammenhänge aus den Forschungsfragen 1–3 zwischen mehrsprachigen und monolingualen Schüler*innen unterscheiden.

- 4) Unterscheiden sich die Zusammenhänge zwischen Überzeugungen zu Mehrsprachigkeit, der Akzeptanz und dem Lernerfolg zwischen mehrsprachigen und monolingualen Schüler*innen?

Methodisches Vorgehen

Intervention

Im vorliegenden Beitrag wird eine mehrsprachigkeitssensible Unterrichtsintervention zur Förderung der Deutsch-Lesekompetenz in vierten Klassen untersucht, in der gemäß des Reziproken Lehrens (vgl. Rosenshine & Meister, 1994) in Kleingruppen und reziproker Rollenverteilung die Lesestrategien *Fragen stellen* und *Zusammenfassen* eingeübt wurden (für eine detaillierte Beschreibung siehe Decristan et al., 2022). Die Intervention umfasst zwölf Unterrichtsstunden von jeweils 45 Minuten und verknüpft die Leseförderung mit dem innovativen Element der Mehrsprachigkeit. In die Unterrichtsstunden wurden mehrsprachigkeitssensible Elemente integriert, um die Familiensprachen der Schüler*innen produktiv im Lernprozess aufzugreifen. Hierzu wurde zu Beginn der Intervention ein *wertschätzendes Klima* für die sprachliche Vielfalt in den Klassen etabliert. Dazu wurden Sprachenportraits (Gogolin & Neumann, 1991) von den Schüler*innen angefertigt, bei denen sie ihre Sprachen mit unterschiedlichen Farben in eine menschliche Silhouette malen und ihre Mehrsprachigkeit damit im Klassenraum sichtbar machen. Um *mehrsprachige Kommunikation* im Lernprozess zu ermöglichen und zu unterstützen, wurden in den Kleingruppen nach Möglichkeit Tandems mit Sprachpartnerschaften gebildet, die neben dem Deutschen eine weitere Familiensprache teilen. Zusätzlich wurde die Mehrsprachigkeit über *bilinguales Unterrichtsmaterial* gefördert. Auf sogenannten Promptkarten waren die wichtigsten Schritte zum Ausführen der Lesestrategien beschrieben. Die Promptkarten waren bilingual gestaltet, sodass neben der deutschen Beschreibung der Schritte auch jeweils die Übersetzung in einer Familiensprache angegeben war. Die Promptkarten konnten zusätzlich mit einem Sprachlernstift verwendet werden, der die Strategieschritte in den Familiensprachen vorliest. Die Intervention wurde im Rahmen des Deutsch-Unterrichts von den Lehrkräften der Klassen durchgeführt, die zuvor in einer Fortbildung an drei Nachmittagen (je vier Stunden) zu den Themen Mehrsprachigkeit, Lese- und Lernstrategien und Reziprokes Lehren fortgebildet wurden. Die Umsetzung im Unterricht war für einen Zeitraum von sechs Wochen mit zwei Unterrichtsstunden pro Woche vorgesehen. Die Angaben der Lehrkräfte zeigen, dass die Intervention durchschnittlich innerhalb von 5,2 Wochen ($SD = 1,8$) mit 2,7 Unterrichtsstunden pro Woche umgesetzt wurde ($SD = 0,7$).

Stichprobe

Die Intervention wurde 2019 in 15 Klassen in Hessen und 23 Klassen in Nordrhein-Westfalen durchgeführt. 2021 wurde sie zusätzlich in acht Klassen in Baden-Württemberg umgesetzt. Aufgrund von mangelnder Implementation der Inhalte durch die Lehrkräfte wurden zwei Klassen ausgeschlossen, sodass die Stichprobe für die vorliegende Untersuchung 44 Klassen mit 785 Schüler*innen umfasst. Die Schüler*innen der Stichprobe waren im Durchschnitt 9,7 Jahre ($SD = 0,7$) alt. 50,4 % waren weiblich ($n = 390$) und 68,5 % ($n = 533$) wurden als mehrsprachig kodiert (vgl. Decristan et al. 2022). Monolinguale und mehrsprachige Schüler*innen unterschieden sich nicht signifikant hinsichtlich des Geschlechts ($\chi^2(1) = 0,07, p = 0,79$), mehrsprachige Schüler*innen waren jedoch signifikant älter als monolinguale ($t(718) = -2,22, p < 0,05$), im Durchschnitt 1,4 Monate.

Instrumente

Die *Überzeugungen zu Mehrsprachigkeit* wurden auf einer selbstkonstruierten vierstufigen Skala (1 = stimmt gar nicht – 4 = stimmt genau) mit vier Items erfasst (z. B. „Ich bewundere es, wenn jemand noch eine andere Sprache als Deutsch sprechen kann.“) und zu Beginn der Intervention erhoben ($\alpha = 0,73$). Die *Akzeptanz* der Intervention wurde mit zwei vierstufigen Subskalen (1 = stimmt gar nicht – 4 = stimmt genau) zur affektiven und kognitiven Akzeptanz erhoben. Die Skalen wurden in Anlehnung an bestehende Instrumente (Eckert et al., 2017; Hebbeker & Souvignier, 2018) für den spezifischen Kontext der Intervention adaptiert und direkt nach Abschluss dieser erhoben. Die affektive Akzeptanz ($\alpha = 0,87$) wurde mit fünf Items abgebildet, die das affektive Erleben während der Intervention beschreiben (z. B. „Das Üben der Lese- und Lernstrategien hat mir gut gefallen.“). Die kognitive Akzeptanz ($\alpha = 0,86$) wurde mit drei Items erhoben, mit denen die Wirksamkeitserwartung und Nützlichkeit der Intervention zur Verbesserung des Leseverstehens abgebildet wird (z. B. „Die Weltreise mit Merle hat mir geholfen, das Lesen zu üben.“). Die theoretisch angenommene Faktorenstruktur von zwei getrennten Akzeptanzdimensionen wurde in vorbereitenden Analysen geprüft. Das Zweifaktorenmodell zeigte dabei eine signifikant bessere Modellanpassung als ein Einfaktorenmodell ($\Delta\chi^2(df) = 118,57(1), p < 0,001$). Da die Akzeptanzdimensionen jedoch hoch miteinander korrelieren, wurden zusätzliche Analysen zur Kollinearitätsdiagnostik durchgeführt. Der *Variance Inflation*

Factor lag dabei mit Werten bis 3.7 unter dem gängigen Cut-Off-Wert (Field, 2009) und zeigt entsprechend keine Multikollinearität an.

Das *Leseverstehen* zu Beginn und nach Abschluss der Intervention wurde mit dem standardisierten Lesetest *VSL* (Walter, 2013) erfasst. Der Test richtet sich an Schüler*innen der 2. bis 6. Klasse und ist als Speedtest konzipiert (Range: 0 bis 50 Punkte). Die *Selbstwirksamkeit* zur Anwendung der Lesestrategien wurde mittels einer vierstufigen Skala (1 = stimmt gar nicht – 4 = stimmt genau) und vier Items zu Beginn ($\alpha = 0,80$) und nach Abschluss der Intervention ($\alpha = 0,85$) erhoben. Die Skala wurde nach bestehenden Instrumenten (Koch & Spörer, 2016) adaptiert und auf die Lesestrategien *Fragen stellen* und *Zusammenfassen* angepasst (z. B. „Wenn ich einen Text zusammenfassen soll, gelingt mir das gut.“).

Statistische Analysen

Zur Beantwortung der Forschungsfragen wurden manifeste lineare Regressions- und Mediationsmodelle in MPlus 8.4 (Muthén & Muthén, 2012–2021) berechnet. Anstelle der Verwendung von Skalenmittelwerten wurden Factorscores aus konfirmatorischen Faktorenanalysen der Konstrukte extrahiert. Dies führt zu einer präziseren Schätzung der Merkmale, da die Faktorladungen gewichtet in die Bildung der Factorscores eingehen (Morris & Guertin, 1977). Die konfirmatorischen Faktorenanalysen wurden anhand des *Weighted-Least-Square-Mean-Variance*-Verfahrens durchgeführt. Dieses gilt als besonders präzise Schätzmethode zur Modellierung ordinalskalierten Daten mit wenigen Stufen (Brown, 2006). Die hierarchische Struktur der Daten wurde über die Korrektur der Standardfehler berücksichtigt. Zur Überprüfung der Signifikanz der indirekten Zusammenhänge wurde das 95 %-Konfidenzintervall unter dem Bootstrapping-Verfahren mit der Ziehung von 1000 Bootstrapping-Stichproben interpretiert (Lai, 2018). Fehlende Werte wurden mithilfe des *Full-Information Maximum-Likelihood*-Verfahrens geschätzt.

Ergebnisse

Die deskriptiven Werte der hier einbezogenen Konstrukte sind in Tab. 7.1 angegeben. Aus den Mittelwerten geht hervor, dass die Überzeugungen zu Mehrsprachigkeit insgesamt positiv ausfallen und die Intervention in hohem Maße

akzeptiert wird. Deskriptiv ist ein Zuwachs im Leseverstehen und in der Selbstwirksamkeit der Lesestrategieanwendung über den Verlauf der Intervention zu verzeichnen. Mehrsprachige Schüler*innen weisen deskriptiv höhere Werte in den Überzeugungen zu Mehrsprachigkeit und den Akzeptanzdimensionen auf. Für monolinguale Schüler*innen sind dagegen höhere Werte im Leseverstehen und der Selbstwirksamkeit der Lesestrategieanwendung zu verzeichnen.

In *Fragestellung 1* wurde die Bedeutung der Überzeugungen zu Mehrsprachigkeit für die Akzeptanz der Intervention untersucht. Hierbei wurde eine Regressionsanalyse gerechnet, in der die Akzeptanzdimensionen als Kriterien durch die Überzeugungen vorhergesagt werden. Aus der Analyse geht hervor, dass Überzeugungen zu Mehrsprachigkeit zu Beginn der Intervention beide Akzeptanzdimensionen positiv vorhersagten (affektive Akzeptanz: $\beta = 0,21$, $SE = 0,05$, $p < 0,001$; kognitive Akzeptanz: $\beta = 0,23$, $SE = 0,04$, $p < 0,001$). *Hypothese 1* kann damit bestätigt werden.

Zur Beantwortung der *Fragestellungen 2* und *3* wurden zwei Mediationsmodelle spezifiziert, in denen das Leseverstehen (Abb. 7.1) und die Selbstwirksamkeit der Lesestrategieanwendung (Abb. 7.2) am Ende der Intervention

Tab. 7.1 Deskriptive Daten der Gesamtstichprobe sowie von monolingualen und mehrsprachigen Schüler*innen

Skala	Gesamtstichprobe ($N = 785$)		Monolinguale Schüler*innen ($n = 245$)		Mehrsprachige Schüler*innen ($n = 533$)	
	M	SD	M	SD	M	SD
Überzeugungen zu Mehrsprachigkeit (Prä)	2,98	0,80	2,78	0,84	3,08	0,77
Affektive Akzeptanz (Post)	3,04	0,84	2,88	0,89	3,11	0,81
Kognitive Akzeptanz (Post)	2,79	1,00	2,52	1,03	2,91	0,97
Leseverstehen (Prä)	15,44	7,13	18,23	7,43	14,10	6,55
Leseverstehen (Post)	16,20	7,77	19,70	8,74	14,59	6,67
Selbstwirksamkeit Lesestrategieanwendung (Prä)	3,05	0,67	3,08	0,63	3,04	0,69
Selbstwirksamkeit Lesestrategieanwendung (Post)	3,24	0,68	3,34	0,63	3,19	0,69

jeweils durch die Überzeugungen zu Mehrsprachigkeit und die Akzeptanzdimensionen vorhergesagt wurden, kontrolliert durch das jeweilige Eingangsniveau der abhängigen Variable.

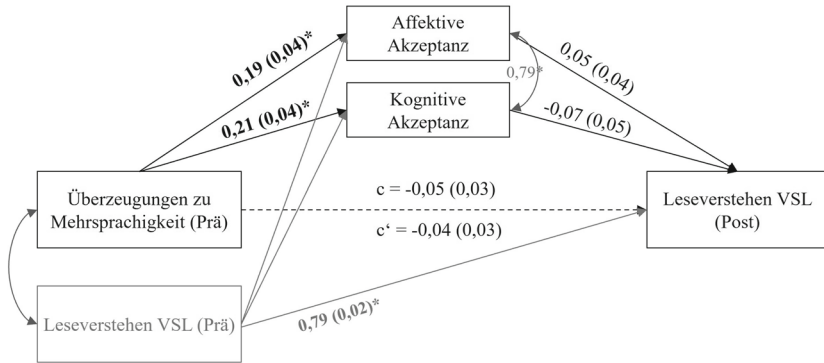


Abb. 7.1 Mediationsmodell zur vermittelnden Rolle der Akzeptanz für den Zusammenhang von Überzeugungen zu Mehrsprachigkeit und Leseverstehen

Anmerkung. * $p < 0,001$.

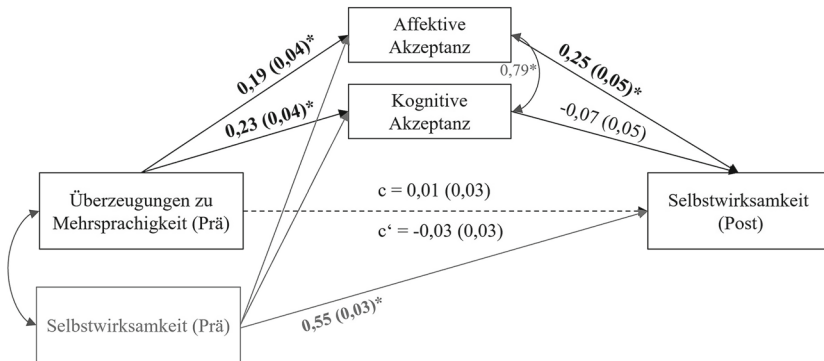


Abb. 7.2 Mediationsmodell zur vermittelnden Rolle der Akzeptanz für den Zusammenhang von Überzeugungen zu Mehrsprachigkeit und Selbstwirksamkeit der Lesestrategieanwendung

Anmerkung. * $p < 0,001$.

Die Überzeugungen zu Mehrsprachigkeit sagten in beiden Modellen den Lernerfolg nicht signifikant vorher, weswegen *Hypothesen 2a und 2b* verworfen wurden. Abb. 7.1 zeigt, dass das Leseverstehen nicht signifikant durch die Akzeptanz vorhergesagt wurde, weshalb für keine der Akzeptanzdimensionen ein vermittelnder Zusammenhang bestand. *Hypothese 3a* konnte somit nicht bestätigt werden. Für die Selbstwirksamkeit der Lesestrategieanwendung (Abb. 7.2) zeigte sich jedoch, dass die affektive Akzeptanz die Selbstwirksamkeit positiv vorher sagte ($\beta = 0,25$, $SE = 0,05$, $p < 0,001$). Auch der vermittelnde Zusammenhang der affektiven Akzeptanz war signifikant ($\beta = 0,05$, $SE = 0,02$, 95 %-CI = $0,022 / 0,098$), *Hypothese 3b* wurde damit bestätigt.

Fragestellung 4 untersuchte schließlich explorativ, ob sich die Zusammenhänge von Überzeugungen, Akzeptanz und Lernerfolg zwischen monolingualen und mehrsprachigen Schüler*innen unterscheiden. Um die Zusammenhänge valide vergleichen zu können, wurden die Skalen auf Messinvarianz zwischen den Gruppen geprüft. Die nächsthöhere Invarianzstufe wurde als zutreffend angenommen, wenn sich mit zunehmender Gleichheitsrestriktion der CFI um weniger als 0,01 verringerte und der RMSEA um höchstens 0,015 zunahm (Chen, 2007; Cheung & Rensvold, 1999). Für alle Skalen konnte auf Basis dieser Schwellenwerte die für die Zusammenhangsanalysen nötige metrische Invarianz angenommen werden (Weiber & Mühlhaus, 2014). In einem Mehrgruppenmodell sagten die Überzeugungen zu Mehrsprachigkeit die Akzeptanzdimensionen in beiden Gruppen positiv vorher (monolingual: affektive Akzeptanz: $\beta = 0,25$, $SE = 0,08$, $p = 0,001$; kognitive Akzeptanz: $\beta = 0,24$, $SE = 0,08$, $p = 0,002$; mehrsprachig: affektive Akzeptanz: $\beta = 0,15$, $SE = 0,05$, $p = 0,005$; kognitive Akzeptanz: $\beta = 0,18$, $SE = 0,05$, $p < 0,001$). Ein χ^2 -Differenztest zwischen dem frei geschätzten Mehrgruppenmodell und einem restringierten Modell, in dem die Pfade zwischen den Gruppen gleichgesetzt wurden, zeigte keinen signifikanten Unterschied der Zusammenhänge zwischen den Gruppen ($\Delta\chi^2(df) = 4,13(3)$, $p = 0,248$). Weiterhin wurden zwei Mehrgruppen-Mediationsmodelle für die Vorhersage des Leseverstehens (Abb. 7.3) und der Selbstwirksamkeit (Abb. 7.4) spezifiziert.

Das Mehrgruppen-Mediationsmodell zur Vorhersage des Leseverstehens (Abb. 7.3) zeigt, dass die Überzeugungen bei mehrsprachigen Schüler*innen das Leseverstehen signifikant negativ vorhersagten. Bei monolingualen Schüler*innen zeigte sich kein direkter Zusammenhang. Die Selbstwirksamkeit (Abb. 7.4) wurde weder bei monolingualen noch bei mehrsprachigen Schüler*innen durch die Überzeugungen vorhergesagt.

Schließlich wurde die vermittelnde Rolle der Akzeptanz zwischen beiden Gruppen verglichen. Bei monolingualen Schüler*innen vermittelten beide

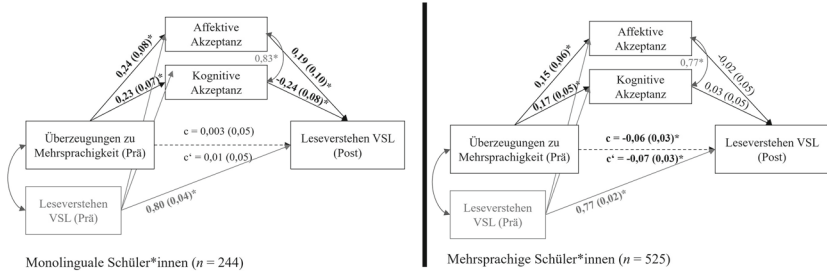


Abb. 7.3 Mehrgruppen-Mediationsmodell zur vermittelnden Rolle der Akzeptanz für den Zusammenhang von Überzeugungen zu Mehrsprachigkeit und Leseverstehen
Anmerkung. * p 0,001.

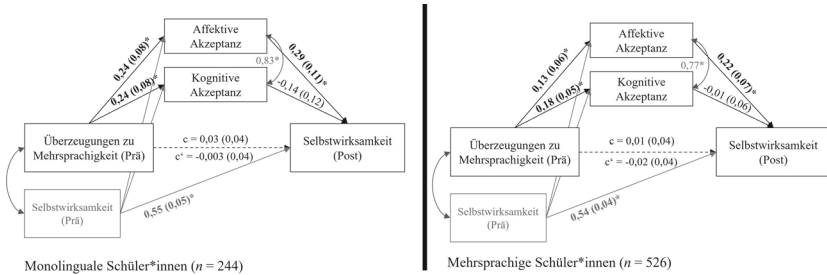


Abb. 7.4 Mehrgruppen-Mediationsmodell zur vermittelnden Rolle der Akzeptanz für den Zusammenhang von Überzeugungen zu Mehrsprachigkeit und Selbstwirksamkeit
Anmerkung. * p 0,001.

Akzeptanzdimensionen den Zusammenhang zwischen Überzeugungen zu Mehrsprachigkeit und dem Leseverstehen (Abb. 7.3). Es zeigte sich ein positiver indirekter Zusammenhang, vermittelt über die affektive Akzeptanz ($\beta = 0,05$, $SE = 0,03$, $95\% \text{-CI} = 0,003/0,103$). Der indirekte Zusammenhang der kognitiven Akzeptanz fiel, entgegen theoretischer Annahmen, signifikant negativ aus ($\beta = -0,06$, $SE = 0,03$, $95\% \text{-CI} = -0,122/-0,010$). Der Zusammenhang zwischen Überzeugungen zu Mehrsprachigkeit und der Selbstwirksamkeit (Abb. 7.4) wurde in beiden Gruppen durch die affektive Akzeptanz vermittelt (monolingual: $\beta = 0,07$, $SE = 0,04$, $95\% \text{-CI} = 0,005/0,164$; mehrsprachig: $\beta = 0,03$, $SE = 0,02$, $95\% \text{-CI} = 0,003/0,068$).

Diskussion

Zusammenfassung und Interpretation

Der Beitrag untersuchte die Bedeutung von Überzeugungen zu Mehrsprachigkeit für die Akzeptanz und den Lernerfolg von Schüler*innen im Rahmen einer mehrsprachigkeitssensiblen Intervention zur Leseförderung in der Grundschule. Ein zentrales Ergebnis der Untersuchung in der Gesamtstichprobe ist, dass positivere Überzeugungen zu Mehrsprachigkeit zu Beginn der Intervention mit einem höheren Maß an Akzeptanz einhergehen. Vermittelt über die affektive Akzeptanz konnte zudem ein Zusammenhang zwischen den Überzeugungen und der Selbstwirksamkeit der Lesestrategianwendung festgestellt werden. Die Akzeptanz hängt damit zum einen mit individuellen Voraussetzungen der Schüler*innen (Überzeugungen) und mit dem Lernerfolg zusammen. Dieses Ergebnis ist anschlussfähig an bestehende Forschungsbefunde der Akzeptanz- und Implementationsforschung, die die Passung zu individuellen Voraussetzungen der Adressat*innen als zentral für den Implementationserfolg einer Innovation ansehen (vgl. Rogers, 2003; Schrader et al., 2020; Venkatesh et al., 2003). Im Kontext mehrsprachigkeitssensibler Konzepte, deren Umsetzung noch wenig untersucht ist (Roth et al., 2021), können die Ergebnisse zudem Überzeugungen zu Mehrsprachigkeit als wichtiges Element für den Implementationserfolg und schließlich auch den Lernerfolg (gemessen an der Selbstwirksamkeit) empirisch bestätigen. Dies bestätigt, dass positive Überzeugungen zu Mehrsprachigkeit ein Kernelement und wichtige Voraussetzung von Strategien pädagogischen Translanguagings sind (García et al., 2017; Lorenz et al., 2021). Positive Überzeugungen zu Mehrsprachigkeit scheinen in der mehrsprachigkeitssensiblen Unterrichtseinheit das Erleben positiver Lernerfahrungen zu begünstigen, die sich in einer höheren Selbstwirksamkeit widerspiegeln (Phan & Ngu, 2016). In der vorliegenden Untersuchung zeigte sich zwar kein vermittelnder Zusammenhang der Akzeptanz für den Zusammenhang von Überzeugungen zu Mehrsprachigkeit und Leseverstehen, jedoch kann angenommen werden, dass positive Selbstwirksamkeitserwartungen das zukünftige Leseverhalten positiv beeinflussen (vgl. McElvany et al., 2008) und so eine höhere Akzeptanz langfristig zur Förderung des Leseverstehens beitragen könnte. Ein direkter Zusammenhang zwischen den Überzeugungen und dem Lernerfolg konnte in der Gesamtstichprobe dagegen nicht bestätigt werden. Dies fällt entgegen theoretischer Annahmen aus, nach denen positivere Überzeugungen zum Unterrichtsgegenstand mit einem höheren Leistungszuwachs assoziiert sein sollten (vgl. Vieluf et al., 2020; Wigfield & Eccles, 2000). Die *indirect-only-mediation* (Zhao et al., 2010) könnte hier

mit erklärend sein, wonach sich Zusammenhänge unterschiedlicher Vorzeichen wechselseitig aufheben und ein direkter Effekt ausbleibt. Aus den vorliegenden Analysen resultierten positive Effekte von Überzeugungen auf die affektive Akzeptanz, aber negative (teils aber nicht signifikante) Effekte der kognitiven Akzeptanz auf den Lernerfolg.

Mit der vierten Fragestellung wurden Unterschiede in den Zusammenhängen zwischen monolingualen und mehrsprachigen Schüler*innen untersucht. Die Zusammenhänge von Überzeugungen und Akzeptanz unterschieden sich nicht signifikant zwischen beiden Gruppen. Ein spezifischer Befund der Studie ist damit, dass positive Überzeugungen zu Mehrsprachigkeit für alle Schüler*innen gleichermaßen für eine Akzeptanz der Intervention relevant sind. Während die Zusammenhänge zur Vorhersage der Selbstwirksamkeit zwischen den Gruppen das gleiche Muster aufwiesen wie in der Gesamtstichprobe (d. h. die affektive Akzeptanz medierte den Zusammenhang von Überzeugungen und Lernerfolg), zeigten sich Unterschiede in den Modellen zur Vorhersage des Leseverstehens. Positivere Überzeugungen hingen bei mehrsprachigen Schüler*innen mit einem geringeren Zuwachs im Leseverstehen zusammen (negativer direkter Effekt). Eine mögliche Erklärung dieses Befundmusters ist, dass positivere Überzeugungen zu Mehrsprachigkeit insbesondere bei Schüler*innen mit einer anderen Erstsprache als Deutsch zu beobachten sein könnten (Haukas et al., 2022). Diese verfügen wiederum häufig über geringere sprachliche Kompetenzen im Deutschen und schwächere Leseleistungen (Edele et al., 2020). Ein weiterer unerwarteter Befund ist, dass die Akzeptanzdimensionen zudem bei mehrsprachigen Schüler*innen nicht als Mediatoren wirkten, während sie bei monolingualen Schüler*innen den Zusammenhang zwischen Überzeugungen zu Mehrsprachigkeit und dem Leseverstehen am Ende der Intervention vermitteln. Hier scheinen unterschiedliche Wirkmechanismen zwischen den Gruppen vorzuliegen. Eine mögliche Erklärung könnte mit dem mehrsprachigkeitssensiblen Charakter der Intervention zusammenhängen. Mehrsprachige Kinder haben durch die Intervention Wertschätzung für ihre sprachlichen Kompetenzen erfahren und konnten im Vergleich zu monolingualen Schüler*innen zusätzlich mehrsprachigkeitssensible Impulse, die die Aktivierung des gesamtsprachlichen Repertoires zum Ziel hatten, nutzen. Es wäre denkbar, dass für mehrsprachige Schüler*innen nicht die Akzeptanz als Mediator fungiert, sondern die Nutzung der mehrsprachigkeitssensiblen Impulse mit dem Lernerfolg in Zusammenhang steht. In Decristan et al., (2022) hat sich der Lernerfolg der Intervention im Leseverstehen beispielsweise unterschieden, je nachdem ob die Schüler*innen eine Nutzung des mehrsprachigen Repertoires angegeben hatten oder nicht. Forschungsergebnisse zu mehrsprachigkeitssensiblen Konzepten zeigen jedoch häufig, dass die tatsächliche Aktivierung

mehrsprachiger Interaktionen oft nur begrenzt gelingt und Schüler*innen ihre Familiensprachen eher selten für das Lernen nutzen (Decristan et al., 2021; Meyer & Prediger, 2011). Der vorherrschende monolinguale Habitus, geringere Kompetenzen in den Familiensprachen sowie negative Überzeugungen von Lehrkräften und Schüler*innen erschweren die aktive Nutzung mehrsprachiger Ressourcen (vgl. Prediger et al., 2021).

Schließlich ist der negative Zusammenhang zwischen der kognitiven Akzeptanz und dem Leseverstehen bei monolingualen Schüler*innen zu diskutieren. Dies fällt ebenfalls entgegen theoretischer Annahmen aus, nach denen ein höheres Maß an Akzeptanz mit einem höheren Leistungszuwachs assoziiert sein sollte (Eckert et al., 2021). Monolinguale Schüler*innen, die die Intervention als besonders hilfreich zur Leseförderung bewerteten (kognitive Akzeptanz), wiesen demnach einen geringeren Lernzuwachs auf. Dies könnte darauf zurückzuführen sein, dass insbesondere Schüler*innen mit geringerem Leseverstehen die Intervention als hilfreich wahrgenommen haben. Eine mögliche Erklärung dafür, dass dieser negative Zusammenhang nur in der Gruppe monolingualer Schüler*innen zu beobachten war, könnte auch hier darin gesehen werden, dass mehrsprachige Schüler*innen im Vergleich zu monolingualen Schüler*innen zusätzlich die mehrsprachigkeitssensiblen Impulse der Intervention nutzen konnten. Durch den Einbezug des gesamtsprachlichen Repertoires bei mehrsprachigen Schüler*innen könnten diese die Intervention unabhängiger von ihren Lernvoraussetzungen als wirksam erachten.

Limitationen

Die Interpretation der Ergebnisse muss vor dem Hintergrund folgender Limitationen erfolgen. Die Akzeptanz einer Intervention ist von ihrer konkreten Umsetzung abhängig (vgl. Fogarty et al., 2014). Zwar wurden die Lehrkräfte in einer gemeinsamen Fortbildung auf die Umsetzung der Unterrichtseinheit vorbereitet, dennoch kann sich trotz erfolgtem Implementationscheck die Umsetzung in den Klassen qualitativ unterschieden haben. Die Unterschiede in der vermittelnden Rolle der Akzeptanz zwischen monolingualen und mehrsprachigen Schüler*innen könnte durch mehr Informationen zur Umsetzung, insbesondere zur Nutzung der mehrsprachigkeitssensiblen Impulse, besser erklärt werden. Aus diesem Grund sollte auch die Umsetzung und Implementationsqualität durch die Schüler*innen in zukünftigen Studien berücksichtigt werden und nach Möglichkeit auf Beobachtungsdaten basieren. Zudem ist zu berücksichtigen, dass sich

die Wirkrichtung der Zusammenhänge in den Mediationsmodellen nicht feststellen lässt. Theoretisch wurde angenommen, dass die Überzeugungen, vermittelt über die Akzeptanz als Wahrnehmungs- und Interpretationsprozess, den Lernerfolg beeinflussen. Dies kann empirisch jedoch nicht endgültig geprüft werden und vor dem Hintergrund konzeptueller Modelle der Implementations- (Witt und Elliott, 1985) und Unterrichtsforschung (Vieluf et al., 2020) ist von reziproken Zusammenhängen zwischen den Konstrukten auszugehen, die in zukünftigen Studien näher beleuchtet werden sollten. Als weitere Limitation ist zu nennen, dass die Gruppe mehrsprachiger Schüler*innen eine Vielzahl an unterschiedlichen Sprachbiografien umfasst, die hier zusammengefasst und nicht weiter differenziert wurden. Mehrsprachige Schüler*innen könnten sich in ihren Überzeugungen zu Mehrsprachigkeit, z. B. aufgrund ihrer Erstsprache (Haukas et al., 2022), unterscheiden oder die mehrsprachigkeitssensiblen Impulse aufgrund ihrer sprachlichen Kompetenz unterschiedlich ausgeprägt nutzen. Schließlich ist zu berücksichtigen, dass die Stichprobe aus Lehrkräften bestand, die freiwillig und aus Interesse zum Thema Mehrsprachigkeit an der Intervention teilgenommen haben. Die Schüler*innen der Klassen könnten durch das Interesse der Lehrkräfte bereits Anknüpfungspunkte zum Thema Mehrsprachigkeit im Unterricht aufweisen und entsprechend auch besonders positive Überzeugungen mitbringen.

Implikationen

Die Ergebnisse bestätigen die Bedeutung positiver Überzeugungen zu Mehrsprachigkeit für das Gelingen mehrsprachigkeitssensibler Unterrichtskonzepte. Als Hürden für den Einbezug von Mehrsprachigkeit werden häufig negative Überzeugungen von Lehrkräften gesehen. Die Ergebnisse zeigen jedoch, dass auch negative Überzeugungen bei Schüler*innen eine Hürde darstellen können. Dies zeigte sich für monolinguale und mehrsprachige Kinder gleichermaßen, sodass ein wertschätzendes Klima für sprachliche Vielfalt für alle Schüler*innen gefördert werden sollte.

Literatur

- 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.

- Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based measurement and student achievement in mathematics computation. *Remedial and Special Education, 18*(2), 113–120. <https://doi.org/10.1177/074193259701800205>.
- Artelt, C., Schiefele, U., Schneider, W., & Stanat, P. (2002). Leseleistungen deutscher Schülerinnen und Schüler im internationalen Vergleich (PISA): Ergebnisse und Erklärungen. *Zeitschrift für Erziehungswissenschaft, 5*(1), 6–27.
- Bandura, A. (1977). Self-Efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. <https://doi.org/10.1037/0033-295x.84.2.191>.
- Bosma, E., Bakker, A., Zenger, L., & Blom, E. (2022). Supporting the development of the bilingual lexicon through translanguaging: A realist review integrating psycholinguistics with educational sciences. *European Journal of Psychology of Education, 38*, 1–23. <https://doi.org/10.1007/s10212-021-00586-6>.
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Cenoz, J., Santos, A., & Gorter, D. (2022). Pedagogical translanguaging and teachers' perceptions of anxiety. *International Journal of Bilingual Education and Bilingualism, 1–12*. <https://doi.org/10.1080/13670050.2021.2021387>
- 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. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*(1), 1–27. <https://doi.org/10.1177/014920639902500101>.
- Decristan, J., Bertram, V., Reitenbach, V., Schneider, K. M., Köhler, C., & Rauch, D. (2022). Linguistically responsive reciprocal teaching in primary school: Effectiveness of an intervention study on students' reading competence. *Journal of Multilingual and Multicultural Development, 1–19*. <https://doi.org/10.1080/01434632.2022.2141757>.
- Decristan, J., Schastak, M., Reitenbach, V., & Rauch, D. (2021). Außerunterrichtliches Peer Tutoring mit deutsch-türkischsprachigen Grundschulkindern: Umsetzungsgenauigkeit und Umfang von bilingualer Kommunikation. *Zeitschrift für Erziehungswissenschaft, 24*, 841–860. <https://doi.org/10.1007/s1618-021-01023-6>.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3), 327–350.
- Ebe, A. E., & Chapman-Santiago, C. (2016). Student voices shining through: exploring translanguaging as a literary device. In O. García & T. Kleyn (Hrsg.), *Translanguaging with multilingual students: learning from classroom moments* (S. 57–82). Routledge.
- Edele, A., Kempert, S., & Stanat, P. (2020). Mehrsprachigkeit und Bildungserfolg. In I. Gogolin, A. Hansen, S. McMonagle, & D. Rauch (Hrsg.), *Handbuch Mehrsprachigkeit und Bildung* (S. 151–155). Springer.
- Eckert, T. L., Eggleston, B. N., Hamsho, N. F., & Williams, N. M. (2021). Assessing children's perceptions of intervention acceptability ratings. *Psychology in the Schools, 58*(10), 1962–1979. <https://doi.org/10.1002/pits.22564>.
- Eckert, T. L., Hier, B. O., Hamsho, N. F., & Malandrino, R. D. (2017). Assessing children's perceptions of academic interventions: The kids intervention profile. *School Psychology Quarterly, 32*(2), 268–281. <https://doi.org/10.1037/spq0000200>.

- Field, A. (2009). *Discovering statistics using SPSS*. Sage.
- Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., Clemens, N., & Robers, G. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: A focus on treatment fidelity. *Educational Psychology Review*, 26(3), 425–449. <https://doi.org/10.1007/s10648-014-9270-6>.
- Fürstenau, S. (2011). Mehrsprachigkeit als Voraussetzung und Ziel schulischer Bildung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Mehrsprachigkeit* (S. 25–50). Springer.
- García, O., Johnson, S. I., & Seltzer, K. (2017). *The Translanguaging classroom: Leveraging student bilingualism for learning*. Caslon.
- García, O., & Wei, L. (2014). *Translanguaging: Language, Bilingualism and education*. Palgrave Macmillan.
- Gogolin, I. (2008). *Der monolinguale Habitus der multilingualen Schule*. Waxmann.
- Gogolin, I. & Neumann, U. (1991). Sprachliches Handeln in der Grundschule. *Die Grundschulzeitschrift*, 43(5), 6–13.
- Grosjean, F. (2013). Bilingualism: A short introduction. In F. Grosjean & P. Li (Hrsg.), *The Psycholinguistics of Bilingualism* (S. 5–26). Wiley-Blackwell.
- Grosjean, F. (2020). Individuelle Zwei- und Mehrsprachigkeit. In I. Gogolin, A. Hansen, S. McMonagle, & D. Rauch (Hrsg.), *Handbuch Mehrsprachigkeit und Bildung* (S. 13–22). Springer.
- Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Toboada, A., Davis, M. H., Scaffidi, N. T., & Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, 96(3), 403–423. <https://doi.org/10.1037/0022-0663.96.3.403>.
- Haddock, G., & Maio, G. R. (2014). Einstellungen. In K. Jonas, W. Stroebe, & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 197–230). Springer.
- Haukas, A., Storto, A., & Tiurikova, I. (2022). School Students' beliefs about the benefits of multilingualism. *Journal of Multilingual and Multicultural Development*. <https://doi.org/10.1080/01434632.2022.2075001>.
- Hebbecke, K., & Souvignier, E. (2018). Formatives Assessment im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts. *Zeitschrift für Erziehungswissenschaft*, 21(4), 735–765. <https://doi.org/10.1007/s11618-018-0834-y>.
- Henschel, S., Heppt, B., & Weirich, S. (2023). Zuwanderungsbezogene Disparitäten. In P. Stanat, S. Schipolowski, R. Schneider, S. Weirich, S. Henschel, & K. A. Sachse (Hrsg.), *IQB Bildungstrend 2022: Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im dritten Ländervergleich*. (S. 299–344). Waxmann.
- Hilbe, R., Kutzelmann, S., Massler, U., & Peter, K. (Hrsg.). (2017). *Mehrsprachiges Vorlesen durch die Lehrperson: Handbuch zu Theorie und Praxis*. Barbara Budrich.
- Kazdin, A. E. (1980). Acceptability of alternative treatments for deviant child behavior. *Journal of Applied Behavior Analysis*, 13(2), 259–273.
- KMK. (2015). *Empfehlungen zur Arbeit in der Grundschule. Beschluss der Kultusministerkonferenz vom 02.07.1970 i. d. F. vom 11.06.2015*. https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2015/Empfehlung_350_KMK_Arbeit_Grundschule_01.pdf

- Koch, H., & Spörer, N. (2016). Förderung der Lesekompetenz mittels reziproken Lehrens: Implementation und Wirksamkeit im Regelunterricht. *Zeitschrift für Pädagogische Psychologie*, 30(4), 213–225. <https://doi.org/10.1024/1010-0652/a000176>.
- Krumm, H.-J. (2016). Mehrsprachigkeit als Kapital – eine Einsicht und ihre Konsequenzen für ein bildungsgerechtes Schulwesen. In A. Wegener & I. Dirim (Hrsg.), *Mehrsprachigkeit und Bildungsgerechtigkeit. Erkundungen einer didaktischen Perspektive* (S. 59–68). Barbara Budrich.
- Lai, K. (2018). Estimating Standardized SEM Parameters given nonnormal data and incorrect model: Methods and comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 600–620.
- Leonet, O., Cenoz, J., & Gorter, D. (2020). Developing morphological awareness across languages: Translanguaging pedagogies in third language acquisition. *Language Awareness*, 29(1), 41–59. <https://doi.org/10.1080/09658416.2019.1688338>.
- Lorenz, E., Krulatz, A., & Torgersen, E. N. (2021). Embracing linguistic and cultural diversity in multilingual EAL classrooms: The impact of professional development on teacher beliefs and practice. *Teaching and Teacher Education*, 105, 103428f. <https://doi.org/10.1016/j.tate.2021.103428>.
- Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education: preparing classroom teachers to teach English language learners. *Journal of Teacher Education*, 59(4), 361–373. <https://doi.org/10.1177/0022487108322110>.
- Mao, P., Cai, Z., He, J., Chen, X., & Fan, X. (2021). The Relationship between attitude toward science and academic achievement in science: A three-level meta-analysis. *Frontiers in Psychology*, 12, Article 784068. <https://doi.org/10.3389/fpsyg.2021.784068>.
- McElvany, N., Kortenbruck, M., & Becker, M. (2008). Lesekompetenz und Lesemotivation: Entwicklung und Mediation des Zusammenhangs durch Leseverhalten. *Zeitschrift für Pädagogische Psychologie*, 22(3–4), 207–219. <https://doi.org/10.1024/1010-0652.22.34.207>.
- Meyer, M., & Prediger, S. (2011). the use of first language turkish as a resource. A German case study on chances and limits for building conceptual understanding. In M. Setati, T. Nkambule & L. Goosen (Eds.), *Proceedings of the ICMI Study 21 Mathematics and Language Diversity* (S. 225–234). University Press.
- Morris, J. D., & Guertin, W. H. (1977). The superiority of factor scores as predictors. *The Journal of Experimental Education*, 45(4), 41–44.
- Moschkovic, J. (2007). Using two languages when learning mathematics. *Educational Studies in Mathematics*, 64, 121–144. <https://doi.org/10.1007/s10649-005-9005-1>.
- Mulder, F. G. W. (2019). *Ein Vergleich der Einstellungen gegenüber bilingualem Unterricht und Mehrsprachigkeit von monolingual und bilingual aufgewachsenen Schülern* innen an einer deutsch-niederländischen Euregio Realschule in Kranenburg*. Dissertation. Radboud University.
- Muthén, L. K., & Muthén, B. O. (2012–2021). *Mplus* (Version 8) [Computer software]. Muthén & Muthén.
- O’Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in k–12 curriculum intervention research. *Review of Educational Research*, 78(1), 33–84.

- Öztürk, S., & Çubukçu, F. (2022). Exploring Student-directed Translanguaging in the English-centric Classroom. *Language Teaching Research Quarterly*, 31, 159–173. <https://doi.org/10.32038/ltrq.2022.31.11>.
- Petermann, F. (2014). Implementationsforschung: Grundbegriffe und Konzepte. *Psychologische Rundschau*, 65(3), 122–128. <https://doi.org/10.1026/0033-3042/a000214>.
- Petscher, Y. (2010). A meta analysis of the relationship between student attitudes towards reading and achievement in reading. *Journal of Research in Reading*, 33(4), 335–355. <https://doi.org/10.1111/j.1467-9817.2009.01418.x>.
- Phan, H. P., & Ngu, B. H. (2016). Sources of self-efficacy in academic contexts: a longitudinal perspective. *School Psychology Quarterly*, 31(4), 548. <https://doi.org/10.1037/spq0000151>.
- Planas, N., & Setati, M. (2009). Bilingual students using their languages in the learning of mathematics. *Mathematics Education Research Journal*, 21(3), 36–59. <https://doi.org/10.1007/BF03217552>.
- Prediger, S., Uribe, Á., Wagner, J., Krause, A., & Redder, A. (2021). Sieben Sprachen im Sachfachunterricht – Ansätze zum Einbezug nicht-geteilter mehrsprachiger Ressourcen zum Aufbau von Konzeptverständnis. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 26(2), 165–194.
- Reitenbach, V., Decristan, J., Rauch, D., Bertram, V., & Schneider, K. M. (2023). Selbstberichtete Gründe für die (Nicht-)Nutzung von Familiensprachen beim mehrsprachigkeits-sensiblen Reziproken Lehren. *Unterrichtswissenschaft*, 51, 221–243. <https://doi.org/10.1007/s42010-023-00171-4>.
- Rogers, E. M. (2003). *Diffusion of innovations*. Free Press.
- Roth, H.-J., Uçan, Y., Sieger, S., & Gollan, C. (2021). Stichwort: Implementationsforschung zwischen Intervention und Transfer im Kontext von Mehrsprachigkeit und sprachlicher Bildung. *Zeitschrift für Erziehungswissenschaft*, 24(4), 775–818. <https://doi.org/10.1007/s11618-021-01024-5>.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479–530. <https://doi.org/10.3102/00346543064004479>.
- Schastak, M. (2020). *Bilinguale Interaktion beim Peer-Learning: Eine Mixed-Methods Studie mit bilingual türkisch-deutschsprachig aufwachsenden Schülerinnen*. Barbara Budrich.
- Schnitzer, K. (2020). *Mehrsprachigkeit als Ressource: Zur Praxis des Sprachunterrichts in der Sekundarstufe I*. Waxmann.
- Schrader, J., Hasselhorn, M., Hetfleisch, P., & Goeze, A. (2020). Stichwortbeitrag Implementationsforschung: Wie Wissenschaft zu Verbesserungen im Bildungssystem beitragen kann. *Zeitschrift für Erziehungswissenschaft*, 23(1), 9–59. <https://doi.org/10.1007/s11618-020-00927-z>.
- Schüler-Meyer, A., Prediger, S., Kuzu, T., Wessel, L., & Redder, A. (2019). Is formal language proficiency in the home language required to profit from a bilingual teaching fintervention in mathematics? a mixed methods study on fostering multilingual students' conceptual understanding. *International Journal of Science and Mathematics Education*, 17, 317–339. <https://doi.org/10.1007/s10763-017-9857-8>.
- Sierens, S., & van Avermaet, P. (2014). Language Diversity in Education: Evolving from multilingual education to functional multilingual learning. In D. Little, C. Leung & P.

- van Avermaet (Hrsg.), *New Perspectives on Language and Education: Bd. 33. Managing Diversity in Education: Languages, Policies, Pedagogies* (S. 204–222). Multilingual Matters.
- Statistisches Bundesamt (Destatis). (2022). *Mikrozensus: Bevölkerung nach Migrationshintergrund, Wiesbaden*. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/Downloads-Migration/statistischer-bericht-migrationshintergrund-erst-2010220227005.html>. Zugegriffen: 18. Mar. 2024.
- Ticheloven, A., Blom, E., Leseman, P., & McMonagle, S. (2021). Translanguaging challenges in multilingual classrooms: Scholar, teacher and student perspectives. *International Journal of Multilingualism*, 18(3), 491–514. <https://doi.org/10.1080/14790718.2019.1686002>.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M., & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts: Ein kritischer Vergleich verschiedener Modellvarianten. In A.-K. Praetorius, J. Grünkern, & E. Klieme (Hrsg.), *Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen* (Zeitschrift für Pädagogik: Beiheft 66, S. 63–80). Beltz Juventa. <https://doi.org/10.3262/ZPB2001063>.
- Walter, J. (2013). *Verlaufsdagnostik sinnerfassenden Lesens*. Hogrefe.
- Wamhoff, D., Maahs, I.-M., & von Dewitz, N. (2022). „Wichtiger find’ ich jetzt nicht, aber manche sind halt einfach nützlicher“ – Sprachliche Hierarchisierungen aus Sicht mehrsprachig aufwachsender Schüler*innen. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 27(1), 345–368.
- Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-35012-2>.
- Wendt, H., & Schwippert, K. (2017). Lesekompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe, & V. Renate (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 219–234). Waxmann.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81.
- Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Hrsg.), *Advances in School Psychology* (Bd. 4, S. 251–288). Routledge.
- Zhao, X., Lync, J. G., & JR., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206. <https://doi.org/10.1086/651257>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Adaptiver Unterricht? Übereinstimmungen und Unterschiede in den Schüler*innenwahrnehmungen von Unterrichtsqualität

Katharina Schnitzler, Nora Fröhlich, Jasmin Decristan,
Benjamin Fauth und Hanna Dumont

Zusammenfassung

Adaptiver Unterricht ist entscheidend für den Umgang mit der Heterogenität von Schüler*innen. Dafür gilt es, Lernangebote kontinuierlich an die individuellen Lernvoraussetzungen anzupassen. Adaptivität wird jedoch bislang

K. Schnitzler (✉) · H. Dumont
Universität Potsdam, Potsdam, Deutschland
E-Mail: katharina.schnitzler@uni-potsdam.de

H. Dumont
E-Mail: hanna.dumont@uni-potsdam.de

N. Fröhlich · B. Fauth
Institut für Bildungsanalysen Baden-Württemberg, Stuttgart, Deutschland
E-Mail: Nora.Froehlich@ibbw.kv.bwl.de

B. Fauth
E-Mail: Benjamin.Fauth@ibbw.kv.bwl.de

J. Decristan
Bergische Universität Wuppertal, Wuppertal, Deutschland
E-Mail: decristan@uni-wuppertal.de

B. Fauth
Universität Tübingen, Tübingen, Deutschland

K. Schnitzler · J. Decristan · B. Fauth · H. Dumont
IDeA – Center for Research on Individual Development and Adaptive Education of
Children at Risk, Frankfurt am Main, Germany

selten empirisch erforscht. Dieser Beitrag untersucht daher Schulpreisträgerschulen des Deutschen Schulpreises, deren Unterrichtspraxis ein hohes Potenzial für Adaptivität aufweist. Ziel ist es, Übereinstimmungen und Unterschiede in den Wahrnehmungen der Unterrichtsqualität in Abhängigkeit von soziodemographischen Merkmalen, kognitiven und motivational-emotionalen Lernvoraussetzungen sowie von Differenzierungsmaßnahmen zu explorieren. Mit den Daten von 552 Grundschüler*innen aus 50 Klassen zeigen sich höherer Konsens für die Klassenführung und konstruktive Unterstützung als für die kognitive Aktivierung. Mehrebenenregressionsanalysen zeigen kein einheitliches Befundmuster zu interindividuellen Wahrnehmungsunterschieden abhängig von den Lernvoraussetzungen. Der Grad an Unterrichtsdifferenzierung steht in keinem Zusammenhang mit interindividuellen Unterschieden der Schüler*innen.

Schlüsselwörter

Adaptiver Unterricht • Unterrichtsqualität • Lernvoraussetzungen • Heterogenität • Grundschule

Einleitung

Lehrpersonen stehen täglich vor der Herausforderung, die Kompetenzentwicklung aller Schüler*innen ihrer Klasse durch ihren Unterricht zu unterstützen. Dabei begegnen sie einer großen Vielfalt an Lernvoraussetzungen, was die Frage aufwirft, wie ein produktiver Umgang damit gelingen kann. Dies betrifft insbesondere Lehrpersonen an Grundschulen, da sich die Lernausgangslagen von Schüler*innen bereits zu Beginn der Schullaufbahn stark unterscheiden (Decristan et al., 2014). Diesbezüglich alarmieren Befunde von aktuellen Schulleistungsstudien: Bisher gelingt es nicht, den unterschiedlichen Merkmalen und Lernvoraussetzungen der Schüler*innen gerecht zu werden. Niedriger sozioökonomischer Status und Migrationshintergrund bleiben entscheidende Risikomerkmale für Bildungsmisserfolge (McElvany et al., 2023; Stanat et al., 2022).

Adaptiver Unterricht – als eine Gesamtunterrichtsstrategie, die auf die kontinuierliche Anpassung der Lernangebote an die individuellen Lernvoraussetzungen der Schüler*innen abzielt – wird als entscheidend für den produktiven Umgang mit heterogenen Lernvoraussetzungen von Schüler*innen gesehen (Corno, 2008).

Um die Anpassung von Lernangeboten zu ermöglichen, kommen dabei Formen der Individualisierung und Differenzierung zum Tragen (Dumont, 2019). Allerdings scheint eine solche Art der Unterrichtsgestaltung „gleichermaßen eine Herausforderung und eine Seltenheit im Schulalltag darzustellen“ (Gräsel et al., 2022, S. 509).

Preisträgerschulen des Deutschen Schulpreises haben sich der Herausforderung eines produktiven Umgangs mit Heterogenität angenommen und werden unter anderem für den erfolgreichen Umgang mit Vielfalt ausgezeichnet (Beutel et al., 2017). Sie setzen auf eine Unterrichtspraxis, die nicht nur an die individuellen Lernvoraussetzungen angepasst ist (Schratz, 2013), sondern diese als Ressource nutzt (Dumont et al., 2024). Bisher ist allerdings ungeklärt, inwieweit es hierdurch gelingt, eine Passung zwischen Lernangeboten und individuellen Lernvoraussetzungen herzustellen. Um empirische Hinweise zur Adaptivität von Unterricht zu gewinnen, exploriert die vorliegende Studie daher die Wahrnehmungen der Unterrichtsqualität durch Schüler*innen in Abhängigkeit von ihren Lernvoraussetzungen und dem Grad der Differenzierung im Unterricht. Dabei wird davon ausgegangen, dass eine adaptive Unterrichtsgestaltung, die eine Passung zwischen Unterrichtsangeboten und Lernvoraussetzungen für alle Schüler*innen einer Klasse herstellt, sowohl zu einer hohen Übereinstimmung in den Wahrnehmungen der Unterrichtsqualität (Becker & Keller, 2022; Rieser & Decristan, 2025; Schweig & Martinez, 2021) als auch zu einer Entkopplung von Wahrnehmungen und Lernvoraussetzungen führen sollte.

Theoretischer Rahmen

Adaptiver Unterricht

Es gibt mittlerweile zahlreiche Konzepte, deren gemeinsame Grundidee darin besteht, die heterogenen Lernvoraussetzungen von Schüler*innen im Unterricht stärker zu berücksichtigen. Ein vielversprechendes Konzept ist der adaptive Unterricht, der das Ziel verfolgt, alle Schüler*innen einer Klasse im Unterricht so zu adressieren, dass eine Passung zwischen Lernangeboten und individuellen Lernvoraussetzungen entsteht (Corno, 2008; Randi, 2022). Adaptiver Unterricht wird als eine Gesamtunterrichtsstrategie verstanden (Dumont, 2019), die sich in ihren grundlegenden Elementen von einem klassischen „One-size-fits-all“-Ansatz unterscheidet (Bondie et al., 2019). Basis ist eine (formative) Diagnostik der Lernvoraussetzungen, Lernverläufe und der sich daraus ergebenden Bedürfnisse

von Schüler*innen (Corno & Snow, 1986; Hardy et al., 2019). Darauf aufbauend wird die Unterrichtsgestaltung kontinuierlich und systematisch an die Lernvoraussetzungen von Schüler*innen angepasst, um diese in ihrer individuellen Zone der proximalen Entwicklung zu adressieren (Dumont, 2019; Klieme & Warwas, 2011; Parsons et al., 2018). Dies gelingt, indem Lehrpersonen das Maß an Unterstützung und Herausforderung für die einzelnen Schüler*innen im Sinne des Scaffolding anpassen (Randi, 2022; van de Pol et al., 2010). So benötigen Schüler*innen mit ungünstigeren Lernausgangslagen mehr Unterstützung und Strukturierung durch die Lehrperson als Schüler*innen mit günstigen Lernvoraussetzungen, die von mehr Freiheiten im Unterricht profitieren (Corno, 2008). Dabei wird zwischen geplanten, längerfristig angelegten Anpassungen (Makroadaptationen) und spontanen, situativen Anpassungen (Mikroadaptationen) unterschieden (Corno, 2008; Klieme & Warwas, 2011). Mit zunehmenden Kompetenzen übernehmen die Schüler*innen sukzessive mehr Verantwortung für ihre Lernprozesse. Neben fachlichen Kompetenzen entwickelt sich so auch das selbstregulierte Lernen weiter (Dumont, 2019; Randi, 2022).

Für die konkrete Umsetzung von adaptivem Unterricht in der Praxis ist die Herstellung der Passung zwischen Lernvoraussetzungen und Lernangeboten handlungsleitend und weniger der Einsatz bestimmter Methoden (Corno & Snow, 1986; Hess & Lipowsky, 2017). Gleichwohl wird angenommen, dass im adaptiven Unterricht Methoden der Differenzierung und Individualisierung als Formen der Makroadaption vermehrt zum Einsatz kommen (Dumont, 2019; Hardy et al., 2019).¹ Damit ist die gezielte Variation von Unterrichtsangeboten für Gruppen von Schüler*innen mit unterschiedlichen Lernvoraussetzungen oder für einzelne Schüler*innen gemeint (Bohl et al., 2012; Jager et al., 2022). Oftmals erfolgt dies in Form von methodischen, medialen, quantitativen, qualitativen oder inhaltlichen Variationen der im Unterricht eingesetzten Aufgaben und Materialien, was zur Folge hat, dass die Schüler*innen einer Klasse an Aufgaben auf unterschiedlichen Schwierigkeitsniveaus arbeiten, unterschiedliche Aufgabenmengen bewältigen oder bei gleicher Aufgabenstellung auf unterschiedliche Materialien zurückgreifen können. Solch eine Differenzierung kann zwar hilfreich sein,

¹ Die Begriffe Differenzierung und Individualisierung werden in der Literatur nicht trennscharf verwendet. In der Regel bezieht sich der Begriff Differenzierung auf die Variation des Unterrichtsangebots für Gruppen von Schüler*innen, während der Begriff Individualisierung für die Variation des Unterrichtsangebots für einzelne Schüler*innen oder als Bezeichnung für einen „dezentralen“ Unterricht, in dem die Schüler*innen zur gleichen Zeit Unterschiedliches tun, verwendet wird (im Überblick siehe Dumont, 2019).

um eine Passung zwischen Lernangeboten und Lernvoraussetzungen herzustellen, gewährleistet diese jedoch nicht automatisch (Gräsel et al., 2022; Hess & Lipowsky, 2017; Jäger et al., 2022).

Da sich adaptiver Unterricht nicht durch den Einsatz bestimmter Methoden und Formen der Unterrichtsorganisation definiert, sondern darauf bezieht, ob die Unterrichtsangebote wirklich zu den Lernvoraussetzungen der Schüler*innen passen, verwundert nicht, dass es bisher weder eine abschließende Antwort darauf gibt, ab wann ein Unterricht als adaptiv gilt, noch Konsens darüber herrscht, auf Basis welcher Kriterien und methodischen Zugänge eine solche Beurteilung erfolgen sollte (Hardy et al., 2019). Derzeit werden verschiedene Möglichkeiten exploriert. In Unterrichtsbeobachtungen wird Lehrkraftverhalten in den Blick genommen, das als adaptiv gilt (Ankrum et al., 2020; Schipper et al., 2020; Vaughn, 2019). So schätzten externe Rater*innen beispielsweise die Adaptivität unter anderem anhand diagnostischer Strategien und variierender Unterstützungsmaßnahmen ein (Hardy et al., 2022). Mit Fragebögen wird eine adaptive Gestaltung des Unterrichts über Selbstberichte von Lehrpersonen erfasst (Vaughn et al., 2022). Zudem werden auch Schüler*innen zur Wahrnehmung der Adaptivität ihres Unterrichts befragt. Dabei berichteten diese, inwiefern sie die Unterstützung durch ihre Lehrkraft als passend im Verhältnis zu ihrem Verständnisniveau empfanden (van de Pol et al., 2022). In Ergänzung zu den bisherigen Methoden exploriert der vorliegende Beitrag die Wahrnehmung der Unterrichtsqualität als eine weitere Möglichkeit, Hinweise auf die Adaptivität von Unterricht zu generieren.

Adaptivität und Unterrichtsqualität

Im deutschsprachigen Raum hat es sich in der pädagogisch-psychologischen Forschung etabliert, Unterricht anhand von Oberflächen- und Tiefenmerkmalen zu beschreiben. Unter Oberflächenmerkmalen (auch als Sichtstrukturen bekannt) werden Unterrichtseigenschaften, wie Organisationsformen, Methoden und Sozialformen gefasst, die von außen gut beobachtbar sind und vielfältig ausgestaltet werden können (Decristan et al., 2020). Im Vergleich dazu sind Tiefenmerkmale auf Ebene der Unterrichtsprozesse verortet und auf die Qualität der Interaktionen zwischen Lehrpersonen, Schüler*innen und Lerngegenständen bezogen, welche vor allem mit der Wirksamkeit von Unterricht assoziiert sind (Klieme, 2022; Kunter & Voss, 2011). Das Verhältnis von Oberflächen- zu Tiefenmerkmalen des

Unterrichts ist weitestgehend ungeklärt (Decristan et al., 2020), wobei sich empirische Hinweise darauf finden, dass Oberflächenmerkmale als Moderatoren die Wirksamkeit von Tiefenmerkmalen verstärken können (Decristan et al., 2015).

Auf Ebene der Tiefenmerkmale werden oftmals drei generische Basisdimensionen – Klassenführung, konstruktive Unterstützung und kognitive Aktivierung – unterschieden (Baumert et al., 2010; Klieme et al., 2001; Kunter & Voss, 2011; Praetorius et al., 2020). Eine effektive Klassenführung zielt auf die bestmögliche Nutzung der Unterrichtszeit zur Auseinandersetzung mit dem Lerngegenstand ab. Darunter fallen der Umgang mit Störungen, flüssige Übergänge zwischen Lernaktivitäten, Regeln und Routinen sowie das Monitoring-Verhalten der Lehrperson (Decristan et al., 2015; Lipowsky et al., 2009; Kunter & Voss, 2011; Praetorius et al., 2018). Die konstruktive Unterstützung bezieht sich auf Aspekte des Unterrichtsklimas (Praetorius et al., 2018), wie einen respektvollen Umgang, eine positive Lehrpersonen-Schüler*innen-Beziehung, einen konstruktiven Umgang mit Fehlern und eine positive Feedbackkultur (Decristan et al., 2015; Klieme, 2022; Lipowsky et al., 2009). Kognitive Aktivierung zielt auf eine vertiefte Auseinandersetzung mit den Lerngegenständen (Klieme, 2022; Kunter & Voss, 2011) durch Aktivierung von Vorwissen, herausfordernde Aufgabenstellungen, das Auslösen von kognitiven Konflikten und die Förderung des problemlösenden Denkens (Lipowsky et al., 2009).

In jüngerer Zeit wird Adaptivität als weiterer, integraler Bestandteil von Unterrichtsqualität diskutiert (Charalambos & Praetorius, 2020; Parsons et al., 2018; Praetorius & Charalambos, 2018). Adaptivität wird dabei ebenso wie die drei Basisdimensionen auf Ebene der Tiefenmerkmale verortet, da es darum geht, ob eine Passung zwischen Lernvoraussetzungen und Unterrichtsangeboten tatsächlich erzielt wird. Dies entscheidet sich nicht auf Ebene der Oberflächenmerkmale, d. h. durch die methodische Organisation des Unterrichts, sondern in den Interaktionen zwischen Lehrkräften, Schüler*innen und Lerngegenständen. Dabei wird angenommen, dass Adaptivität quer zu den drei Basisdimensionen liegt und mit diesen verzahnt ist, d. h., dass diese in Passung zu den individuellen Lernvoraussetzungen ausgestaltet werden müssen (Fauth et al., 2024). So kann eine tiefergehende Auseinandersetzung mit den Lerngegenständen, „ohne zu über- oder unterfordern, (...) eigentlich nur gelingen, wenn die individuellen Lernvoraussetzungen der Schülerinnen und Schüler berücksichtigt werden“ (Rieser & Decristan, 2025, S. 166). Ähnliches gilt für die konstruktive Unterstützung. Eine positive Beziehung zur Lehrperson sollte sich vor allem dann einstellen, wenn die Lernvoraussetzungen und damit verbundene Bedürfnisse der einzelnen Schüler*innen in Interaktionen und Rückmeldungen berücksichtigt werden.

Auch die Klassenführung sollte beispielsweise bei der Optimierung der Lernzeit den Lernvoraussetzungen der Schüler*innen entsprechen. Primär steht bei dieser Dimension jedoch die Steuerung der gesamten Klasse und weniger der individuellen Schüler*innen im Fokus. Angesichts der konzeptuellen Verzahnung von Adaptivität und Unterrichtsqualität stellt sich die Frage, wie sich daraus empirische Hinweise auf Adaptivität ableiten lassen.

Schüler*innenwahrnehmungen der Unterrichtsqualität als Hinweis auf Adaptivität

Zur Operationalisierung von Unterrichtsqualität werden neben hochinferenten Ratings, die durch externe Beobachter*innen für die gesamte Klasse vergeben werden, häufig Urteile von Schüler*innen über den Unterricht eingesetzt (Schweig, 2016). Bisherige Studien im Grundschulbereich zeigen, dass Schüler*innenwahrnehmungen der Unterrichtsqualität von soziodemographischen Merkmalen sowie kognitiven und motivational-emotionalen Lernvoraussetzungen abhängen (Fauth et al., 2016; Iglar et al., 2019; Ruelmann et al., 2021; Stang & McElvany, 2021). Insbesondere die Wahrnehmung der kognitiven Aktivierung scheint mit verschiedenen Lernvoraussetzungen in Zusammenhang zu stehen. So konnten beispielsweise Iglar et al. (2019) zeigen, dass Vorwissen, kognitive Grundfähigkeiten, Motivation und sozialer Hintergrund prädiktiv dafür waren, wie herausfordernd Schüler*innen Aufgabenstellungen im Unterricht wahrnahmen. Unter der Prämisse, dass alle Schüler*innen einer Klasse das gleiche Unterrichtsangebot erhalten und Unterrichtsqualität auf der Klassen-ebene zu verorten ist (Vieluf et al., 2020; Wagner et al., 2013), werden solche interindividuellen Wahrnehmungsunterschiede oft als fehlerhafte Aussagen der Schüler*innen und „Rauschen“ interpretiert (Ruelmann et al., 2021; Schweig, 2016). Es konnte jedoch gezeigt werden, dass Divergenzen innerhalb von Klassen als valide Varianz und inhaltlich bedeutsame Information über die individuellen Lernprozesse von Schüler*innen interpretiert werden können (Göllner et al., 2018). So könnte die Ursache für interindividuelle Wahrnehmungsunterschiede des Unterrichts auch in der Gestaltung desselben liegen (Atlay et al., 2019; Bardach et al., 2018; Göllner et al., 2018; Iglar et al., 2019; Schenke et al., 2018; Schweig, 2016; Schweig & Martinez, 2021; Stang & McElvany, 2021). Vor diesem Hintergrund ist davon auszugehen, dass auch eine adaptive Unterrichtsgestaltung mit den Wahrnehmungen von Schüler*innen in Zusammenhang steht (Rieser & Decristan, 2025).

Dem oben skizzierten Verhältnis von Adaptivität und Unterrichtsqualität folgend gehen wir in der vorliegenden Studie davon aus, dass sich Adaptivität in einer als durchschnittlich hoch wahrgenommenen Unterrichtsqualität widerspiegelt. Da Adaptivität die Passung des Unterrichtsangebots zu den Lernvoraussetzungen der einzelnen Schüler*innen sicherstellen soll, gilt es jedoch auch, die interindividuellen Unterschiede in der Wahrnehmung von Unterrichtsqualität eingehender zu berücksichtigen: Zwar erhalten im adaptiven Unterricht die Schüler*innen häufig unterschiedliche Lernangebote, diese sind aber auf die individuellen Lernvoraussetzungen der Schüler*innen abgestimmt. Dadurch sollten – sofern tatsächlich eine Passung zwischen Lernangeboten und Lernvoraussetzung hergestellt wird – Unterschiede in der Wahrnehmung der Unterrichtsqualität nicht verstärkt, sondern verringert werden (Becker & Keller, 2022; Rieser & Decristan, 2025; Schweig & Martinez, 2021). Folglich könnten sowohl ein hoher Konsens innerhalb von Klassen in Bezug auf die wahrgenommenen Unterrichtsqualität als auch eine Unabhängigkeit der Wahrnehmungen der Unterrichtsqualität von den Lernvoraussetzungen der Schüler*innen als empirische Hinweise auf Adaptivität gedeutet werden. Dies gilt insbesondere mit Blick auf die kognitive Aktivierung und die konstruktive Unterstützung, die sich definitionsgemäß stärker auf die Schüler*innen als Individuen als auf die Klasse als Ganzes beziehen.

Die vorliegende Studie

Adaptiver Unterricht ist für den Umgang mit Heterogenität relevant und verfolgt eine Passung zwischen Lernangeboten und individuellen Lernvoraussetzungen. Dennoch wurde genau diese Passung bislang kaum empirisch erforscht. Die vorliegende Studie nimmt sich diesem Desiderat an und exploriert dafür die Wahrnehmungen der Unterrichtsqualität an Grundschulen, die mit dem deutschen Schulpreis ausgezeichnet wurden und deren Unterrichtspraxis ein hohes Potenzial für Adaptivität aufweist. Da angenommen wird, dass die Differenzierung von Aufgaben als ein unterrichtsorganisatorisches Oberflächenmerkmal des Unterrichts zwar hilfreich für das Erreichen von Adaptivität auf Ebene der Tiefenmerkmale sein kann, nicht aber zwangsläufig ausreichend ist, wird ihre Rolle zusätzlich genauer beleuchtet. Es wird folgenden Forschungsfragen nachgegangen:

Fragestellung 1: Wie hoch ist die Übereinstimmung der Schüler*innen innerhalb von Klassen bezüglich der wahrgenommenen Unterrichtsqualität (1a) und hängt diese mit der Differenzierung im Unterricht zusammen (1b)?

Fragestellung 2: Stehen innerhalb von Klassen soziodemographische Merkmale sowie kognitive und motivational-emotionale Lernvoraussetzungen mit der wahrgenommenen Unterrichtsqualität in Zusammenhang (2a) und moderiert der Grad an Differenzierung im Unterricht diesen Zusammenhang (2b)?

Methode

Studiendesign und Stichprobe

In diesem Beitrag wurden Daten aus dem Ada*Q-Projekts genutzt (Dumont et al., 2024), welches im Schuljahr 2019/20 nach Genehmigung durch die jeweiligen Kultusministerien und mit Zustimmung der Schulleitungen und Erziehungsberechtigten durchgeführt wurde.

An der Studie nahmen neun Grundschulen teil, die mit dem Deutschen Schulpreis ausgezeichnet wurden. Obwohl sich die konkreten Konzepte voneinander unterschieden, lassen sich die teilnehmenden Schulen durch gemeinsame Merkmale charakterisieren: Die Heterogenität von Schüler*innen wurde als Ressource gesehen und mitunter absichtlich gesteigert (z. B. Jahrgangsmischung in Klassen). Methoden der Differenzierung und Individualisierung (z. B. individualisierte Lernpläne) und das selbstregulierte Lernen (z. B. freie Lernzeiten) spielten eine große Rolle. Ebenso wurden gemeinschaftliche und kooperative Lernformen (z. B. Helfersysteme) eingesetzt. Dies ging mit einer formativen Diagnostik und hohen Flexibilisierung der Unterrichtsorganisation einher (Dumont et al., 2024).

Von ursprünglich 54 Klassen wurden vier Klassen mit weniger als fünf teilnehmenden Schüler*innen ausgeschlossen, um eine reliable Schätzung der Analysen zu gewährleisten. Daraufhin konnten Daten von 552 Schüler*innen der dritten und vierten Jahrgangsstufe ($M_{\text{Alter}} = 9,19$ Jahre; $SD = 0,68$) aus 50 Klassen analysiert werden. Je Klasse lagen durchschnittlich Daten von 11 Schüler*innen ($SD = 5,2$) vor. Von den Klassen waren 24 jahrgangsübergreifend zusammengesetzt, während 18 im Team-Teaching durch feste Teams von zwei Lehrpersonen unterrichtet wurden. In den Team-Teaching-Klassen wurde jeweils eine der beiden Lehrpersonen zufällig ausgewählt und in die Studie aufgenommen. Von den 50 Lehrpersonen gaben 81.3% ein weibliches Geschlecht an. Sie wiesen ein durchschnittliches Alter von 44 Jahren ($SD = 8,82$) und eine mittlere Berufserfahrung von 16 Jahren ($SD = 8,53$) auf. Zu Beginn des Schuljahres wurden die Lernvoraussetzungen der Schüler*innen sowie die Wahrnehmung von Unterrichtsqualität und Differenzierung im Unterricht erfasst.

Erhebungsinstrumente

Die Items der eingesetzten Skalen werden in Tab. 8.1 vollständig berichtet.

Soziodemographische Merkmale

Das *Geschlecht* der Schüler*innen wurde binär unterteilt (0 = männlich, 1 = weiblich). Der *Migrationshintergrund* wurde ebenfalls binär kodiert. Ein Wert von eins wurde vergeben, wenn die Schüler*innen und/oder die Lehrperson angab, dass mindestens ein Elternteil außerhalb von Deutschland geboren wurde. Um den *sozioökonomischen Status* zu bestimmen, wurden die von den Schüler*innen angegebenen Berufe der Eltern gemäß den ISEI-08 Vorgaben kodiert und der höchste Beruf ausgewählt (Ganzeboom, 2010).

Kognitive Lernvoraussetzungen

Zur Erfassung der kognitiven Lernvoraussetzungen wurden standardisierte Tests durchgeführt. Die *kognitiven Grundfähigkeiten* wurden mithilfe des Matrizen-Subtests des CFT 20-R (Weiß, 2006) gemessen (0–15 Punkte). Die Leistung im *sinnerfassenden Lesen* wurde durch den VSL (Walter, 2013) ermittelt (0–50 Punkte).

Motivational-emotionale Lernvoraussetzungen

Die motivational-emotionalen Lernvoraussetzungen wurden mit Fragebögen erfasst. Die Schüler*innen gaben ihre Antworten auf einer vierstufigen Likert-Skala (1 = „stimmt gar nicht“ bis 4 = „stimmt genau“) an. Zur Erfassung der *Anstrengungsbereitschaft im Unterricht* wurde eine Skala der TRAIN-Studie (Jonkmann et al., 2013) mit vier Items ($\alpha = ,81$) eingesetzt. Die *Lernfreude im Unterricht* wurde mit drei Items ($\alpha = ,80$) erhoben (Jerusalem et al., 2009).

Wahrnehmungen der Unterrichtsqualität

Zur Erfassung der Wahrnehmungen der Unterrichtsqualität wurden Fragebögen eingesetzt. Die Schüler*innen beantworteten die Items jeweils auf einer vierstufigen Likert-Skala (1 = „stimmt gar nicht“ bis 4 = „stimmt genau“). Die *Klassenführung* wurde mit vier Items aus der TRAIN-Studie (Jonkmann et al., 2013) erhoben, die das Monitoring-Verhalten der Lehrperson fokussierten ($\alpha = ,81$). Für die *konstruktive Unterstützung* wurden 9 Items ($\alpha = ,89$) der IGEL-Studie verwendet (Fauth et al., 2014). Die *kognitive Aktivierung* wurde mit sechs Items ($\alpha = ,75$) gemessen, die ebenfalls aus der IGEL-Studie stammten (Fauth et al., 2014).

Tab. 8.1 Skalen zur Erfassung der motivational-emotionalen Lernvoraussetzungen, der wahrgenommenen Unterrichtsqualität und der Differenzierung**Anstrengungsbereitschaft****Wie sehr stimmst du den folgenden Aussagen zu?**

Ich bin im Unterricht wirklich fleißig.

Ich tue im Unterricht mein Bestes.

Ich strengte mich im Unterricht wirklich an.

Im Unterricht bearbeite ich alle Aufgaben sehr ordentlich.

Lernfreude**Wie sehr stimmst du den folgenden Aussagen zu?**

Manche Themen in der Schule machen mir so viel Spaß, dass ich mich schon vorher darauf freue.

Ich habe Spaß daran, im Unterricht Neues zu lernen.

Ich freue mich oft schon richtig auf den Unterricht.

Klassenführung**Lehrerin A/B**

weiß immer genau, was in der Klasse vor sich geht.

sieht sofort, wenn wir abgelenkt sind.

achtet sehr darauf, dass wir aufpassen.

merkt sofort, wenn Schüler/innen beginnen, etwas anderes zu treiben.

Konstruktive Unterstützung**Lehrerin A/B**

ist freundlich zu mir, auch wenn ich einen Fehler mache.

kümmert sich um mich.

macht mir Mut, wenn ich eine Aufgabe schwierig finde.

sagt mir, wie ich es besser machen kann, wenn ich einen Fehler gemacht habe.

mag mich gern.

sagt mir, was ich schon kann und was ich noch lernen muss.

ist freundlich zu mir.

lobt mich, wenn ich etwas gut gemacht habe.

denkt, dass ich auch schwierige Aufgaben lösen kann.

Kognitive Aktivierung**Lehrerin A/B**

(Fortsetzung)

Tab. 8.1 (Fortsetzung)

Anstrengungsbereitschaft
fragt genau nach, was ich verstanden habe und was noch nicht.
stellt Fragen, über die ich ganz genau nachdenken muss.
stellt mir Aufgaben, die mir am Anfang schwierig vorkommen.
fragt mich, was ich zu einem neuen Thema schon weiß.
stellt mir Aufgaben, über die ich gerne nachdenke.
möchte, dass ich meine Antworten auch erklären kann.
Differenzierung
<i>Welche Aussagen treffen auf euren Unterricht zu?</i>
Alle Schüler/innen arbeiten mit den gleichen Schulbüchern. (rekodiert)
Alle Schüler/innen machen die gleichen Aufgaben. (rekodiert)
Alle Schüler/innen machen zur selben Zeit das Gleiche. (rekodiert)
Die einzelnen Schüler/innen arbeiten an verschiedenen Aufgaben.
Die einzelnen Schüler/innen arbeiten mit verschiedenen Materialien.
Alle Schüler/innen sollen am Ende einer Stunde die gleichen Aufgaben gelöst haben. (rekodiert)
Alle Schüler/innen haben zur selben Zeit das gleiche Fach. (rekodiert)
Die einzelnen Schüler/innen haben unterschiedlich schwere Aufgaben.

Differenzierung

Die Differenzierung wurde von den Schüler*innen über einen Fragebogen mit acht Items beurteilt, die auf die Aufgabendifferenzierung im Unterricht abzielten (Dumont, 2016). Die Bewertung erfolgte auf der gleichen vierstufigen Likert-Skala wie die anderen Fragebogenitems ($\alpha = ,79$).

Analytisches Vorgehen

Fehlende Werte

Der jeweilige Anteil fehlender Werte ist in Tab. 2 dargestellt. Insbesondere für den sozioökonomischen Status und den CFT fehlten verhältnismäßig viele Werte, da die entsprechenden Messinstrumente aus organisatorischen Gründen nicht in allen Klassen eingesetzt wurden. Die fehlenden Werte können daher als zufällig fehlend (Missing at random (MAR)) klassifiziert werden (Graham, 2009; Lüdtke &

Robitzsch, 2011). Es wurde daher eine multiple Imputation von 50 vollständigen Datensätzen in R Version 4.3.2 (R Core Team, 2023) mit dem mice-Paket (van Buuren & Groothuis-Oudshoorn, 2011) durchgeführt. Alle weiteren Analysen wurden separat für die imputierten Datensätze durchgeführt und anschließend gemäß Rubin (1987) zusammengefasst.

Voranalysen

Da es sich um eine gezielt selektive Stichprobe handelt und die Annahme bestehen könnte, dass die Schüler*innen an Preisträgerschulen z. B. überwiegend aus Familien mit hohem sozioökonomischen Status stammen oder nur vereinzelt einen Migrationshintergrund aufweisen, wurden in einem ersten Analyseschritt die Zusammensetzungen der einzelnen Klassen hinsichtlich der soziodemographischen Merkmale und Lernvoraussetzungen deskriptiv betrachtet. Hierzu wurden die jeweiligen Standardabweichungen der Klassenmittelwerte herangezogen. Darüber hinaus wurden die Intraklassenkorrelationen ICC(1) und ICC(2) als Reliabilitätsmaße für genestete Daten berechnet (Gärtner, 2010; Lüdtke et al., 2006, 2009).

Übereinstimmungsindizes (Fragestellung 1a und 1b)

Um zu beurteilen, inwieweit die Schüler*innen innerhalb einer Klasse in ihren Wahrnehmungen der Unterrichtsqualität übereinstimmen, wurden der AD_M -Index (average absolut deviation; Burke et al., 1999; Burke & Dunlap, 2002) und der $r^*_{wg(J)}$ -Index (Lindell et al., 1999) mit dem multilevel-Paket in R (Bliese et al., 2022) berechnet. Der AD_M -Index gibt die mittlere betragliche Abweichung der Angaben der Schüler*innen einer Klasse vom Klassenmittelwert in der Metrik der Variablen an (Gärtner, 2010), sodass kleinere Werte für eine höhere Übereinstimmung stehen. Bei den vorliegenden Skalen mit 4 Antwortkategorien wird ab einem Wert von ,67 von einer hohen Übereinstimmung ausgegangen (Burke et al., 1999; Burke & Dunlap, 2002). Der $r^*_{wg(J)}$ -Index relativiert die beobachtete Varianz zwischen den Schüler*innen einer Klasse an der Varianz einer Nullverteilung, wobei höhere Werte für eine höhere Übereinstimmung stehen. Einer Daumenregel folgend werden Werte ab ,70 als ausreichend für eine hohe Übereinstimmung gesehen (James et al., 1984, 1993), wobei in der Unterrichtsforschung bisher lediglich Werte zwischen ,21 und ,68 berichtet wurden (Gärtner, 2010; Bardach et al., 2018). Um zu überprüfen, ob die Übereinstimmungen in den Wahrnehmungen mit der Differenzierung im Unterricht zusammenhängen, wurden Korrelationen zwischen den Übereinstimmungsindizes und den Klassenmittelwerten der Differenzierung berechnet.

Mehrebenen-Regressionsanalysen (Fragestellung 2a und 2b)

Es wurden Mehrebenen-Regressionsanalysen durchgeführt (Lütke et al., 2009; Raudenbush & Bryk, 2010), um die Frage zu beantworten, inwiefern Zusammenhänge zwischen den Lernvoraussetzungen mit den Wahrnehmungen der Unterrichtsqualität innerhalb von Klassen bestehen. Prädiktoren auf Level-1 wurden standardisiert und am Klassenmittelwert zentriert; Prädiktoren auf Level-2 wurden standardisiert und am Grand-Mean zentriert (Enders & Tofghi, 2007). Die Modelle wurden schrittweise aufgebaut und in Mplus 8.10 (Muthén und Muthén 2012–2022) geschätzt. Im ersten Schritt wurden für jede Dimension der Unterrichtsqualität Modelle geschätzt, in welchen die Merkmale bzw. Lernvoraussetzungen der Schüler*innen jeweils einzeln auf beiden Analyseebenen als Prädiktoren aufgenommen wurden. Um die moderierende Rolle der Unterrichtsdifferenzierung zu prüfen, wurde in einem zweiten Schritt zusätzlich die Differenzierung im Unterricht als Prädiktor auf beiden Ebenen aufgenommen sowie die Cross-Level-Interaktion zwischen der jeweiligen Lernvoraussetzung auf Level-1 und der Differenzierung auf Level-2 geschätzt. Als dritter Schritt wurde je Basisdimension zunächst ein Modell geschätzt, in dem die soziodemographischen Merkmale sowie die kognitiven und motivational-affektiven Lernvoraussetzungen simultan als Prädiktoren aufgenommen wurden. Für diese Modelle wurde dann in einem vierten Schritt jeweils einzeln die Cross-Level-Interaktionen zwischen den Prädiktoren auf Level-1 und der Differenzierung auf Level-2 aufgenommen.

Ergebnisse

Deskriptive Befunde

Die deskriptiven Ergebnisse der Voranalysen sind in Tab. 8.2 dargestellt.

Insgesamt waren 53,1 % der Kinder weiblich und 25,7 % hatten einen Migrationshintergrund. Die breiten Streubereiche für den sozioökonomischen Status sowie für die kognitiven und motivational-emotionalen Lernvoraussetzungen zeigen, dass die Gesamtstichprobe hinsichtlich der unterschiedlichen Lernvoraussetzungen heterogen zusammengesetzt war. Dies spiegelt sich auch in der Zusammensetzung der einzelnen Klassen wider: Die Werte für die mittleren Standardabweichungen innerhalb von Klassen zeigten deutliche interindividuelle Unterschiede zwischen den Schüler*innen hinsichtlich ihrer Lernvoraussetzungen.

Tab. 8.2 Deskriptive Statistiken der Stichprobe auf Individual- und Klassenebene

		Individualebene										Klassenebene							
Fehlende Werte		Skala		ICC(1)	ICC(2)	Min	Max	M	SD	ICC(1)	ICC(2)	Min	Max	M	SD	Min	Max	M	SD
n	(%)																		
Lernvoraussetzung																			
Weiblich	29	(5,3)						53,1%				27,3%	83,3%	53,9%	12,56				
Migrationshintergrund	7	(1,3)						25,7%				0%	88,9%	27,6%	22,90				
Sozioökonomischer Status	213	(38,6)	11,56-88,96			14,21	88,96	58,26	22,31			45,35	77,81	59,19	7,44	10,38	27,90	21,17	3,09
Kognitive Grundfähigkeiten	132	(23,9)	0-15			0	14	8,18	2,96			6,11	11,40	8,33	1,11	1,09	3,88	2,82	0,58
Sinnerfassendes Lesen	14	(2,5)	0-50			0	42	9,31	7,44			2,96	20,00	9,55	3,86	1,75	16,68	6,59	2,46
Anstrengungsbereitschaft ^a	123-127	(22,3-23,0)	1-4	,04	,33	1	4	3,28	0,64			2,67	3,72	3,24	0,24	0,27	1,00	0,60	0,17
Lernfreude ^a	123-125	(22,3-22,6)	1-4	,06	,39	1	4	3,21	0,78			2,17	3,78	3,17	0,32	0,18	1,41	0,73	0,22
Unterrichtsqualität																			
Klassentführung ^a	23-29	(4,2-5,3)	1-4	,10	,55	1	4	3,47	0,64			2,80	3,95	3,43	0,30	0,14	1,37	0,56	0,26
Konstruktive Unterstützung ^a	21-41	(3,8-7,4)	1-4	,09	,53	1	4	3,52	0,60			2,32	3,96	3,48	0,29	0,09	1,06	0,49	0,25
Kognitive Aktivierung ^a	24-30	(4,4-5,4)	1-4	,15	,66	1	4	3,15	0,61			2,07	3,88	3,08	0,34	0,15	1,32	0,54	0,20
Moderator																			
Differenzierung ^a	26-57	(4,7-10,3)	1-4	,54	,93	1	4	2,79	0,68			1,78	3,79	2,90	0,52	0,23	0,71	0,42	0,11

N = 552 Schüler*innen aus 50 Klassen mit durchschnittlich 11,04 Schüler*innen.

^a Für die Skalen ist die Spanne der fehlenden Werte auf Item-Ebene angegeben.

Die Qualität des Unterrichts nahmen die Schüler*innen über alle drei Dimensionen hinweg insgesamt als hoch wahr. Die Werte der ICC(1) bestätigen die Mehrebenenstruktur der Daten und machen Unterschiede zwischen den Klassen deutlich. Die Werte der ICC(2) liegen (knapp) unterhalb des Grenzwerts von 0,7 für eine akzeptable Reliabilität. Die Schüler*innen berichteten im Mittel von einer Unterrichtsdifferenzierung etwas oberhalb des Skalenmittelpunkts. Der ICC(1) zeigt, dass sich die Klassen in der Differenzierung deutlich voneinander unterscheiden. Gleichzeitig konnten die Klassenmittelwerte der Differenzierung laut ICC(2) reliabel erfasst werden.

Tab. 8.3 zeigt die bivariaten Korrelationen zwischen den Lernvoraussetzungen, den Wahrnehmungen der Unterrichtsqualität und der Differenzierung im Unterricht auf Individual- und auf Klassenebene.

Auf der Individualebene zeigte sich, dass Mädchen die Unterrichtsqualität positiver wahrnahmen. Außerdem nahmen Schüler*innen die Unterrichtsqualität positiver wahr, je höher ihre Anstrengungsbereitschaft und Lernfreude ausgeprägt war. Im Gegensatz dazu nahmen die Schüler*innen die Unterrichtsqualität negativer wahr, je höher ihr SES und die kognitiven Grundfähigkeiten ausfielen. Der Migrationshintergrund und die Leseleistung korrelierten hingegen nicht signifikant mit der Unterrichtsqualität.

Auf der Klassenebene zeigte sich ein in weiten Teilen ähnliches Muster. Je höher die Anstrengungsbereitschaft und Lernfreude war, desto positiver wurde die Unterrichtsqualität wahrgenommen. Je höher der SES, die kognitiven Grundfähigkeiten und die Leseleistung auf Klassenebene ausfielen, desto negativer wurde die Unterrichtsqualität wahrgenommen. Es ergaben sich weder für das Geschlecht noch für den Migrationshintergrund signifikante Korrelationen mit der Unterrichtsqualität.

Frage 1: Übereinstimmungen in den Wahrnehmungen der Unterrichtsqualität und der Zusammenhang mit der Differenzierung

Tab. 8.4 zeigt, inwiefern die Schüler*innen innerhalb von Klassen in den Wahrnehmungen der Unterrichtsqualität übereinstimmen (Frage 1a).

Die mittleren Ausprägungen beider Übereinstimmungsindizes legen nahe, dass der Konsens für die wahrgenommene Klassenführung und für die konstruktive Unterstützung höher ausfiel als für die kognitive Aktivierung. Wird die Anzahl der Klassen betrachtet, deren AD_M -Index unterhalb des Grenzwerts von ,67 beziehungsweise deren $r^*wg(J)$ -Index oberhalb des Grenzwerts von ,70 liegt, wird

Tab. 8.3 Bivariate Korrelationen auf Individual- und Klassenebene

	1	2	3	4	5	6	7	8	9	10	11
Lernvoraussetzungen											
1 Weiblich		-0,15	-0,01	-0,13	-0,21**	-0,07	0,22*	0,06	0,04	0,10	0,26**
2 Migrationshintergrund	-0,03		-0,14	0,18*	0,13*	0,21*	-0,04	-0,01	0,13	-0,01	0,00
3 Sozioökonomischer Status	0,03	-0,14*		0,51**	0,34**	-0,22**	-0,27**	-0,39**	-0,37**	-0,48**	0,03
4 Kognitive Grundfähigkeiten	-0,01	-0,01	0,23**		0,61**	-0,03	-0,06	-0,23*	-0,22*	-0,39**	0,07
5 Sinnerfassendes Lesen	0,04	-0,08	0,22**	0,31**		0,18*	0,13*	-0,01	-0,09*	-0,20**	0,18**
6 Anstrengungsbereitschaft	0,13*	0,08	-0,08	0,01	0,05		0,68**	0,28**	0,25**	0,29**	-0,15*
7 Lernfreude	0,28**	0,01	-0,10	0,00	0,03	0,66**		0,50**	0,45**	0,45**	0,01
Unterrichtsqualität											
8 Klassenführung	0,14*	-0,03	-0,15*	-0,13*	0,02	0,36**	0,41**		0,62**	0,46**	-0,21**
9 Konstruktive Unterstützung	0,15*	-0,01	-0,08	-0,05	0,06	0,46**	0,48**	0,67**		0,61**	-0,05
10 Kognitive Aktivierung	0,07	-0,03	-0,13*	-0,18**	-0,05	0,40**	0,42**	0,52**	0,61**		-0,02
Moderator											
11 Differenzierung	0,01	-0,05	0,07	0,05	0,08	-0,07	-0,04	-0,11*	-0,04	-0,06	

N = 552 Schüler*innen aus 50 Klassen

Koeffizienten für Individualebene unterhalb der Diagonalen. Koeffizienten auf Klassenebene oberhalb der Diagonalen * p < 0,05. ** p < ,001

Tab. 8.4 Deskriptive Statistiken der Übereinstimmungs-Indizes

	AD _M					r* _{wg(J)}				
	Min	Max	M	SD	Klassen mit	Min	Max	M	SD	Klassen mit
					AD _M < ,67					r* _{wg(J)} > ,70
Klassenführung	0,09	1,16	0,53	0,22	38 (76,0%)	0,00	0,96	0,56	0,27	19 (38,0%)
Konstruktive Unterstützung	0,06	0,86	0,47	0,17	45 (90,0%)	0,00	0,97	0,56	0,26	19 (38,0%)
Kognitive Aktivierung	0,19	1,15	0,65	0,15	29 (58,0%)	0,00	0,92	0,42	0,20	3 (6,0%)

N = 50 Klassen.

ebenfalls deutlich, dass für die wahrgenommene Klassenführung und konstruktive Unterstützung wesentlich mehr Klassen den Grenzwert erreichten als für die kognitive Aktivierung. Gleichzeitig zeigen die Spannweiten und Standardabweichungen, dass sich die Klassen in Bezug auf die Übereinstimmung der Klassenführung und konstruktiven Unterstützung tendenziell stärker voneinander unterscheiden als hinsichtlich der kognitiven Aktivierung.

Zur Frage (1b), ob ein Zusammenhang zwischen der Differenzierung im Unterricht und den Übereinstimmungen besteht, waren weder die Korrelationen zwischen der Differenzierung und den AD_M-Indizes (-,18 bis ,15) noch die Korrelationen zwischen der Differenzierung und den r*_{wg(J)}-Indizes (-,17 bis ,11) signifikant (alle $p > ,05$). Der wahrgenommene Grad an Differenzierung stand somit in keinem Zusammenhang mit der Übereinstimmung zur Unterrichtsqualität innerhalb von Klassen.

Frage 2: Unterschiede in den Wahrnehmungen der Unterrichtsqualität in Abhängigkeit der individuellen Lernvoraussetzungen und der Differenzierung

Für die Fragestellung (2a), inwiefern soziodemographische Merkmale sowie kognitive und motivational-emotionalen Lernvoraussetzungen mit den Wahrnehmungen der Unterrichtsqualität innerhalb von Klassen zusammenhängen, zeigen Tab. 8.5 bis 8.7 die Ergebnisse der Mehrebenen-Regressionsanalysen auf Level 1 (jeweils in Modell a).

In Bezug auf die soziodemographischen Merkmale nahmen Mädchen die Klassenführung und die konstruktive Unterstützung signifikant positiver wahr als Jungen. Außerdem nahmen Schüler*innen mit höherem sozioökonomischen Status die Klassenführung signifikant negativer wahr. Der Migrationshintergrund wies keinen signifikanten Zusammenhang mit den Wahrnehmungen der Unterrichtsqualität auf. Für die kognitiven Lernvoraussetzungen berichteten Schüler*innen mit höheren kognitiven Grundfähigkeiten von einer signifikant geringeren Klassenführung und kognitiven Aktivierung. Alle anderen Zusammenhänge zwischen den kognitiven Lernvoraussetzungen und den Wahrnehmungen der Unterrichtsqualität waren nicht signifikant. Hinsichtlich der motivational-emotionalen Lernvoraussetzungen zeigte sich ein einheitliches Bild: Je höher die Anstrengungsbereitschaft und Lernfreude, desto positiver war die Wahrnehmung von Klassenführung, konstruktiver Unterstützung und kognitiver Aktivierung.

Bei simultaner Aufnahme von allen Prädiktoren blieben die meisten Zusammenhänge bestehen. Je höher die kognitiven Grundfähigkeiten, desto niedriger nahmen Schüler*innen die Klassenführung ($\beta = -12$, $SE = 0,05$, $p = 0,02$) und kognitive Aktivierung ($\beta = -0,15$, $SE = 0,05$, $p = ,001$) wahr. Je stärker die Lernfreude der Schüler*innen ausgeprägt war, desto positiver nahmen sie die Klassenführung ($\beta = 0,27$, $SE = 0,07$, $p < ,001$), konstruktive Unterstützung ($\beta = 0,30$, $SE = 0,08$, $p < ,001$) und kognitive Aktivierung wahr ($\beta = 0,25$, $SE = 0,06$, $p < ,001$). Zudem zeigte sich, dass die Schüler*innen die konstruktive Unterstützung ($\beta = 0,24$, $SE = 0,07$, $p < ,001$) und kognitive Aktivierung ($\beta = 0,20$, $SE = 0,05$, $p < ,001$) ebenfalls positiver wahrnahmen, je höher ihre Anstrengungsbereitschaft ausfiel. Allerdings blieb der zuvor signifikante Zusammenhang der Anstrengungsbereitschaft mit der Klassenführung nicht bestehen. Die soziodemographischen Merkmale standen ebenfalls in keinem signifikanten Zusammenhang mehr mit der wahrgenommenen Unterrichtsqualität.

Eine moderierende Rolle der Differenzierung im Unterricht (Frage 2b) konnte anhand der Cross-Level-Interaktionen nicht bestätigt werden (siehe Tab. 8.5 bis 8.7 jeweils Modell b). Einzig der Zusammenhang zwischen der Leseleistung und der wahrgenommenen kognitiven Aktivierung wurde durch die Differenzierung moderiert. In Klassen mit geringerer Differenzierung nahm die Wahrnehmung der kognitiven Aktivierung mit steigender Leseleistung ab, während sie in Klassen mit stärkerer Differenzierung mit zunehmender Leistung anstieg. Diese Interaktion war allerdings nicht mehr signifikant, sobald alle Prädiktoren gemeinsam in das Modell aufgenommen wurden. Auch alle anderen Interaktionseffekte wurden nicht signifikant.

Tab. 8.6 Wahrnehmungen der Unterrichtsqualität in Abhängigkeit von kognitiven Lernvoraussetzungen und Differenzierung

	Klassenführung			Konstruktive Unterstützung			Kognitive Aktivierung					
	Modell 10a	Modell 10b	Modell 11a	Modell 11b	Modell 12a	Modell 12b	Modell 13a	Modell 13b	Modell 14a	Modell 14b	Modell 15a	Modell 15b
	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)
Level-1												
Kognitive Grundfähigkeiten	-0,11* (0,05)	-0,10 (0,06)			-0,02 (0,06)	-0,03 (0,07)				-0,13* (0,05)	-0,14* (0,05)	
Sinnerfassendes Lesen			0,04 (0,05)	0,04 (0,06)			0,10 (0,06)	0,11* (0,06)			-0,01 (0,06)	-0,01 (0,05)
Differenzierung		-0,10 (0,08)		-0,12 (0,08)		-0,07 (0,09)		-0,08 (0,09)			-0,07 (0,07)	0,09 (0,07)
Level-2												
Intercept	-0,12 (0,07)	-0,13 (0,07)	-0,11 (0,07)	-0,12 (0,07)	-0,16* (0,07)	-0,17* (0,08)	-0,16* (0,08)	-0,16* (0,08)	-0,13 (0,07)	-0,13 (0,07)	-0,12 (0,08)	-0,12 (0,08)
GM Kognitive Grundfähigkeiten	-0,12 (0,07)	-0,11 (0,08)			-0,08 (0,08)	-0,07 (0,08)			-0,17* (0,07)	-0,17* (0,08)		
GM Sinnerfassendes Lesen			-0,05 (0,06)	-0,02 (0,06)			-0,04 (0,07)	-0,02 (0,07)			-0,08 (0,06)	-0,07 (0,07)
GM Differenzierung		-0,09 (0,07)		-0,10 (0,07)		-0,03 (0,07)		-0,04 (0,07)		-0,05 (0,07)		-0,06 (0,07)

(Fortsetzung)

Tab. 8.6 (Fortsetzung)

	Klassenführung			Konstruktive Unterstützung			Kognitive Aktivierung					
	Modell 10a	Modell 10b	Modell 11a	Modell 11b	Modell 12a	Modell 12b	Modell 13a	Modell 13b	Modell 14a	Modell 14b	Modell 15a	Modell 15b
Cross-Level Interaktion		0,00 (0,05)		0,03 (0,04)		-0,05 (0,06)		0,05 (0,06)		-0,05 (0,05)		0,12* (0,06)
Zufällige Effekte												
Level-1 Residualvarianz	1,05** (0,13)	1,04** (0,13)	1,06** (0,13)	1,04** (0,13)	1,15** (0,16)	1,13** (0,17)	1,14** (0,16)	1,13** (0,17)	0,92** (0,08)	0,91 (0,09)	0,93** (0,08)	0,91** (0,08)
Level-2 Residualvarianz	0,11* (0,03)	0,10 (0,03)	0,12* (0,04)	0,12 (0,07)	0,13 (0,09)	0,13 (0,10)	0,14 (0,09)	0,15 (0,09)	0,14 (0,07)	0,14 (0,08)	0,17* (0,08)	0,17* (0,08)
Steigung Lernvoraussetzung		0,00 (0,04)		0,01 (0,11)		0,02 (0,03)		0,01 (0,06)		0,01 (0,03)		0,01 (0,02)
Modelfit												
AIC	1640,50	1645,51	1648,79	1651,10	1693,38	1698,33	1691,61	1697,19	1576,53	1583,29	1591,39	1592,11
BIC	1662,07	1688,65	1670,36	1694,23	1714,95	1741,47	1713,17	1740,33	1598,10	1626,42	1612,95	1635,24

N = 552 Schüler*innen aus 50 Klassen.

GM = Klassenmittelwert.

* p < 0,05. ** p < ,001.

Tab. 8.7 Wahrnehmungen der Unterrichtsqualität in Abhängigkeit von motivational-emotionalen Lernvoraussetzungen und Differenzierung

	Klassenführung				Konstruktive Unterstützung				Kognitive Aktivierung			
	Modell 16a	Modell 16b	Modell 17a	Modell 17b	Modell 18a	Modell 18b	Modell 19a	Modell 19b	Modell 20a	Modell 20b	Modell 21a	Modell 21b
	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)
Level –1												
Anstrengungsbereitschaft	0,30** (0,07)	0,32** (0,07)			0,44** (0,06)	0,43** (0,06)			0,34** (0,05)	0,36** (0,06)		
Lernfreude		0,37** (0,07)	0,37** (0,07)	0,37** (0,07)		0,48** (0,07)	0,47** (0,08)				0,37** (0,05)	0,38** (0,06)
Differenzierung		-0,07 (0,07)	-0,09 (0,08)	-0,09 (0,08)		-0,04 (0,08)	-0,03 (0,08)			-0,05 (0,07)		-0,04 (0,07)
Level –2												
Intercept	-0,09 (0,06)	-0,09 (0,06)	-0,09 (0,06)	-0,09 (0,06)	-0,13* (0,07)	-0,13 (0,07)	-0,13 (0,06)	-0,12 (0,09)	-0,10 (0,07)	-0,10 (0,07)	-0,09 (0,06)	-0,09 (0,07)
GM	0,15 (0,10)	0,14 (0,10)			0,21 (0,13)	0,22 (0,13)			0,20 (0,11)	0,21 (0,06)		
Anstrengungsbereitschaft			0,24** (0,06)	0,24** (0,05)			0,30** (0,08)	0,32* (0,10)			0,27** (0,07)	0,29** (0,08)
GM Lernfreude												
GM Differenzierung		-0,06 (0,06)	-0,09 (0,06)	-0,09 (0,06)		0,00 (0,06)	-0,01 (0,07)			-0,03 (0,06)		-0,07 (0,06)

(Fortsetzung)

Tab. 8.7 (Fortsetzung)

	Klassenführung				Konstruktive Unterstützung				Kognitive Aktivierung			
	Modell 16a	Modell 16b	Modell 17a	Modell 17b	Modell 18a	Modell 18b	Modell 19a	Modell 19b	Modell 20a	Modell 20b	Modell 21a	Modell 21b
	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)	Koeff. (SE)
Cross-Level Interaktion		0,07 (0,07)	0,07 (0,06)			0,00 (0,06)		-0,01 (0,07)		0,01 (0,06)		0,01 (0,06)
Zufällige Effekte												
Level-1 Residualvarianz	0,97** (0,11)	0,88** (0,12)	0,92** (0,10)	0,83** (0,10)	0,05** (0,13)	0,90 (0,14)	0,92** (0,12)	0,84** (0,19)	0,81** (0,06)	0,77** (0,06)	0,79** (0,06)	0,76** (0,06)
Level-2 Residualvarianz	0,11* (0,03)	0,08 (0,04)	0,08* (0,03)	0,08* (0,02)	0,13 (0,07)	0,13 (0,07)	0,08* (0,04)	0,08 (0,34)	0,15* (0,07)	0,16* (0,07)	0,12* (0,05)	0,11* (0,04)
Steigung Lernvoraussetzung		0,08 (0,04)		0,09 (0,05)		0,04 (0,03)		0,07 (0,20)		0,03 (0,03)		0,04 (0,03)
Modellfit												
AIC	1596,43	1581,18	1563,92	1539,09	1590,32	1590,79	1563,81	1543,35	1512,53	1515,67	1493,68	1494,27
BIC	1618,00	1624,32	1585,49	1582,22	1611,88	1633,92	1585,38	1586,49	1534,10	1558,81	1515,25	1537,40

N = 552 Schüler*innen aus 50 Klassen.

GM = Klassenmittelwert.

* p < 0,05. ** p < ,001.

Diskussion

Adaptiver Unterricht gilt als entscheidend für einen erfolgreichen Umgang mit der Heterogenität von Schüler*innen. Bislang bestehen jedoch kaum konsensuelle Zugänge, wie Adaptivität erfasst werden kann. Vor diesem Hintergrund wurden in der vorliegenden Studie gezielt Preisträgerschulen des Deutschen Schulpreises untersucht, deren Unterrichtspraxis ein hohes Potenzial für Adaptivität aufweist. Die Wahrnehmungen der Unterrichtsqualität durch Schüler*innen wurde als möglicher empirischer Hinweis für Adaptivität exploriert. Hierfür wurden einerseits die Übereinstimmungen der Schüler*innen hinsichtlich der Wahrnehmungen der Klassenführung, konstruktiven Unterstützung und kognitiven Aktivierung analysiert. Andererseits wurde geprüft, ob sich spezifische interindividuelle Unterschiede in den Wahrnehmungen der Unterrichtsqualität in Abhängigkeit von soziodemographischen Merkmalen sowie kognitiven und motivational-emotionalen Lernvoraussetzungen zeigen. Dabei wurde die Rolle der Differenzierung sowohl für die Übereinstimmungen als auch für die Unterschiede berücksichtigt.

Übereinstimmungen in den Unterrichtswahrnehmungen

In der Gesamtbetrachtung zeigten die Übereinstimmungsindizes für die vorliegende Stichprobe von Dritt- und Viertklässler*innen an Grundschulen des Deutschen Schulpreises, dass die Schüler*innen einer Klasse in der Wahrnehmung der Klassenführung und konstruktiven Unterstützung stärkere Übereinstimmungen aufwiesen als in der Wahrnehmung der kognitiven Aktivierung. Dies wurde nicht nur anhand der Mittelwerte deutlich, sondern auch an der Anzahl der Klassen, die die Grenzwerte für eine hohe Übereinstimmung erreichten. Ein Vergleich mit bisherigen Studien gestaltet sich als herausfordernd, da Übereinstimmungsindizes bisher nur vereinzelt und meist bezüglich anderer Aspekte als den drei Basisdimensionen berichtet wurden. Die Ergebnisse dieser Studie zeigen, dass der Anteil der Klassen mit einem hohen Konsens in effektiver Klassenführung und konstruktiver Unterstützung im Vergleich zu vorherigen Befunden je nach genutztem Index ähnlich oder sogar höher ausfällt (Gärtner, 2010; Lütke et al., 2006). Auch die mittleren Übereinstimmungen für Klassenführung und konstruktive Unterstützung sind größer als bisher berichtete Werte (Bardach et al., 2018; Lütke et al., 2006). Für die kognitive Aktivierung fallen der Anteil an Klassen mit einer hohen Übereinstimmung und mittlere Übereinstimmungswerte in

der Stichprobe der Preisträgerschulen jedoch geringer aus als in anderen Studien (Rieser & Decristan, 2025).

Unter der Annahme, dass adaptiver Unterricht mit einem hohen Konsens hinsichtlich der Unterrichtswahrnehmungen einhergehen sollte (Rieser & Decristan, 2025; Schweig & Martinez, 2021), scheint es durch das unterrichtliche Handeln an Preisträgerschulen somit gelungen zu sein, die individuellen Bedürfnisse bei der Klassenführung und konstruktiven Unterstützung zu berücksichtigen. In Bezug auf die kognitive Aktivierung zeigt der niedrigere Konsens, dass es auch an Preisträgerschulen eine Herausforderung bleibt, Unterrichtsangebote so zu gestalten, dass alle Schüler*innen einer Klasse sie als Gelegenheit zur vertieften Auseinandersetzung mit den Lerngegenständen wahrnehmen.

Unterschiede in den Unterrichtswahrnehmungen

Die Ergebnisse der Mehrebenen-Regressionsanalysen verweisen teils auf interindividuelle Unterschiede in den Wahrnehmungen der Unterrichtsqualität innerhalb von Klassen, die an bisherige Studienergebnisse anschlussfähig sind. Schüler*innen mit höheren kognitiven Grundfähigkeiten empfanden die Klassenführung und kognitive Aktivierung signifikant negativer als ihre Mitschüler*innen (Iglar et al., 2019). Insbesondere motivational-emotionale Lernvoraussetzungen prägten die Wahrnehmungen der Unterrichtsqualität. Je höher die Lernfreude, desto positiver waren die berichtete Klassenführung, konstruktive Unterstützung und kognitive Aktivierung und je höher die Anstrengungsbereitschaft, desto positiver wurde die konstruktive Unterstützung und kognitive Aktivierung wahrgenommen. Dieser Befund stimmt mit bisherigen Studien überein, die die große Bedeutung der Motivation von Schüler*innen für die Wahrnehmungen von Unterricht herausstellen (Fauth et al., 2020; Iglar et al., 2019; Rühlmann et al., 2021). Im Hinblick auf eine adaptive Unterrichtsgestaltung ist inhaltlich vor allem relevant, welche Zusammenhänge zwischen Wahrnehmungen und Lernvoraussetzungen nicht signifikant sind. So zeigten sich in dieser Studie keine Unterschiede in der wahrgenommenen Unterrichtsqualität abhängig vom Geschlecht, Migrationshintergrund, SES und der Leseleistung. Insgesamt ist das Befundmuster allerdings uneinheitlich.

Die Rolle der Differenzierung im Unterricht

Die mittlere Ausprägung und die Streuung der Differenzierung über die Klassen zeigten, dass sich auch der Unterricht an Schulpreisträgergrundschulen im jeweiligen Grad an Differenzierung deutlich unterscheidet. Der Grad an Differenzierung stand wiederum in keinem systematischen Zusammenhang mit der von Schüler*innen wahrgenommenen Unterrichtsqualität. Die Ergebnisse der Cross-Level-Interaktionen im Rahmen der Mehrebenen-Regressionsanalysen demonstrierten ebenfalls, dass die Differenzierung im Unterricht anscheinend keine Rolle dafür spielt, in welcher Art sich interindividuelle Wahrnehmungsunterschiede in Klassen in Abhängigkeit von soziodemographischen Merkmalen sowie kognitiven und motivational-emotionalen Lernvoraussetzungen zeigen. Diese Befunde könnten durch zwei konträre Mechanismen erklärt werden, die möglicherweise zeitgleich zum Tragen kommen und sich dadurch gegenseitig aufheben. Einerseits könnten Differenzierungen, die eine Passung realisieren, zu einer höheren Übereinstimmung in den Wahrnehmungen führen. Andererseits besteht die Gefahr, dass Differenzierungsmaßnahmen auf inakkuraten Diagnosen beruhen und so keinen Beitrag zu einer Passung leisten können, entsprechend sollten sich die Wahrnehmungen dann stärker unterscheiden (Rieser & Decristan, 2025; Schweig & Martinez, 2021). Die Befunde bestätigen daher, dass die Differenzierung als eine unterrichtsorganisatorische Maßnahme nicht automatisch mit einer Passung der Lernangebote und Lernvoraussetzungen einhergeht (Gräsel et al., 2022; Hess & Lipowsky, 2017). Die Umsetzung von Oberflächenmerkmalen hat somit nicht zwangsläufig Adaptivität auf Ebene der Tiefenmerkmale zur Folge (Decristan et al., 2020) – von dem Grad an Differenzierung allein darf somit nicht auf das Ausmaß an Passung rückgeschlossen werden.

Limitationen

Bei der Interpretation der Studie ist zu berücksichtigen, dass die Stichprobengröße mit 50 Klassen und durchschnittlich 11 teilnehmenden Schüler*innen je Klasse recht klein ausfällt und an der Grenze zu der empfohlenen Stichprobengröße für Mehrebenenanalysen liegt. Dies kann die Power der Analysen einschränken und die Schätzung der Standardfehler beeinflussen (Maas & Hox, 2005; Snijders, 2005). Darüber hinaus fielen die ICC(2)-Werte für die wahrgenommene Unterrichtsqualität aufgrund der relativ geringen Anzahl an Schüler*innen pro Klasse eher niedrig aus. Stichproben mit mehr teilnehmenden Schüler*innen je Klasse wären daher für weitere Studien wünschenswert.

Bei der Messung der Variablen ist außerdem zu beachten, dass alle Variablen auf Angaben der Schüler*innen beruhen. Mit Blick auf den SES ist anzumerken, dass es für Schüler*innen der dritten und vierten Klasse herausfordernd sein könnte, die Berufe ihrer Eltern korrekt zu benennen. Für zukünftige Untersuchungen wäre es daher wünschenswert, Angaben der Eltern zu nutzen. Für die wahrgenommene Unterrichtsqualität und die Differenzierung ist darüber hinaus zu beachten, dass diese nicht über externe Angaben oder Ratings von außenstehenden Beobachter*innen validiert wurden. Auch wenn argumentiert werden kann, dass Schüler*innen am besten beurteilen können, ob sie beispielsweise im Unterricht kognitiv aktiviert sind und externe Beobachter*innen lediglich das Potenzial zur kognitiven Aktivierung erfassen können, könnte die (Aufgaben-) Differenzierung als ein Oberflächenmerkmal in zukünftigen Studien sehr wohl von externen Beobachter*innen und mit Angaben der Lehrpersonen validiert werden.

Für die Untersuchung wurden ausschließlich Schüler*innendaten im Querschnitt genutzt. Dabei lässt sich vor allem für proximale Lernvoraussetzungen auch von reziproken Zusammenhängen mit der wahrgenommenen Unterrichtsqualität ausgehen. So könnte der vorgefundene Zusammenhang mit Motivation auch damit erklärbar sein, dass eine höhere Unterrichtsqualität zu einer höheren Anstrengungsbereitschaft und Lernfreude beigetragen hat. Im Rahmen der hier vorliegenden Daten kann diese Annahme jedoch nicht überprüft werden.

Zudem berücksichtigen die Analysen bislang nicht hinreichend die Interkorrelation der Lernvoraussetzungen. In zukünftigen Studien sollte daher anstelle von bivariaten Zusammenhangsanalysen auch personenzentrierte Ansätze wie z. B. Profilanalysen eingesetzt werden (Ruelmann et al., 2022; Tetzlaff et al., 2023; Willems, 2022), um weitere empirische Erkenntnisse zu gewinnen.

In der Literatur zum adaptiven Unterricht findet sich im Rückgriff auf den Expertise-Reversal-Effekt häufig die Annahme, dass Schüler*innen mit niedrigen Lernvoraussetzungen mehr Unterstützung durch die Lehrkraft bedürfen als Schüler*innen mit günstigen Lernvoraussetzungen. Gleichzeitig nehmen Schüler*innen im Unterricht möglicherweise ganz andere, individuelle Bedürfnisse zum Beispiel nach sozialem Austausch mit Freund*innen wahr. Diese konnte in der vorliegenden Studie nicht berücksichtigt werden.

Der vorliegende Beitrag verfolgt zudem die Argumentation, dass Unterschiede bzw. Übereinstimmungen in den Wahrnehmungen auf den Unterricht zurückzuführen sind. Zusätzlich kann die Übereinstimmung aber auch von generellen Herausforderungen bei der Messung der Basisdimensionen geprägt sein. So könnte es im Vergleich zur konstruktiven Unterstützung und Klassenführung für Schüler*innen schlicht schwieriger sein, die Items zur kognitiven Aktivierung

valide zu beantworten. Darüber hinaus könnten jedoch auch systematisch mit bestimmten Lernvoraussetzungen gekoppelte Beurteilungstendenzen eine alternative Erklärungsmöglichkeit sein (Göllner et al., 2018). Vor diesem Hintergrund gilt es in zukünftigen Analysen, beispielsweise unter Einbezug weiterer Perspektiven auf den Unterricht, zusätzliche empirische Hinweise zur Fundierung dieser Aussagen zu erlangen.

Implikationen

In der vorliegenden Studie wurde argumentiert, dass eine adaptive Unterrichtsgestaltung zu einem höheren Konsens innerhalb von Klassen hinsichtlich der Unterrichtswahrnehmungen durch Schüler*innen führen sollte. In der Literatur finden sich unterschiedliche Indizes, um den Grad an Übereinstimmungen zu bestimmen. Wie schon in vorherigen Studien zeigte sich auch bei den hier berichteten Analysen, dass die Ergebnisse der Indizes nicht einheitlich ausfallen (Gärtner, 2010; Lüdtke et al., 2006). Zudem beruht die Beurteilung einer hohen Übereinstimmung bisher auf Daumenregeln und die vorgeschlagenen Werte wurden in der Unterrichtsforschung kaum erreicht (Gärtner, 2010; Bardach et al., 2018). Unklar ist daher, welcher Index für die Beurteilung von Übereinstimmungen am besten geeignet ist und ab welchem Grad einer Übereinstimmung von einer gelungenen Adaptivität ausgegangen werden kann.

Die Ergebnisse dieser Studie zeigen deutliche Unterschiede in der von Schüler*innen wahrgenommenen kognitiven Aktivierung. Dieser Befund ist anschlussfähig an die Argumentation, dass vor allem für eine kognitive Aktivierung Lernvoraussetzungen eine entscheidende Rolle spielen und hier interindividuelle Wahrnehmungsunterschiede zwischen Schüler*innen im Sinne einer kognitiven Aktiviertheit systematischer in den Blick genommen werden sollten (Rieser & Decristan, 2025). In diesem Zuge deuten die Befunde dieser Studie darauf hin, dass auch in der Unterrichtspraxis der Preisträgerschulen die kognitive Aktivierung aller Schüler*innen einer Klasse eine Herausforderung darstellt. Gleichwohl weisen die Übereinstimmungsindizes darauf hin, dass es deutliche Unterschiede zwischen den Klassen darin gibt, wie gut es gelingt die verschiedenen Schüler*innen zur vertieften Auseinandersetzung mit den Lerngegenständen anzuregen. Weitergehende Analysen hinsichtlich der konkreten Unterrichtspraxis in solchen Klassen mittels Videoanalysen oder Interviews in Kombination mit einer situativen Erfassung der Unterrichtswahrnehmung beispielsweise über Experience Sampling könnten Aufschluss darüber geben, welche spezifischen

Unterrichtsmerkmale für die kognitive Aktivierung der gesamten Klasse hilfreich sind und zur Identifizierung von Good-Practice-Beispielen beitragen.

Die Ergebnisse dieser Studie demonstrieren, dass die Beforschung von Schulen, die den Deutschen Schulpreis erhalten haben, lohnenswert ist. Diese Herangehensweise bietet eine Möglichkeit, den aktuellen Bestrebungen nach einer zukunftsorientierten Unterrichtsforschung gerecht zu werden und die Herausforderung der zunehmenden Heterogenität in den Fokus zu rücken, um robuste Implikationen für die Praxis zu generieren (Dumont et al., 2024). Die vorliegende Studie liefert in diesem Rahmen einen explorativen empirischen Beitrag zur Erfassung und zum Ausmaß von Adaptivität anhand von Unterschieden und Übereinstimmungen hinsichtlich der von Schüler*innen wahrgenommenen Unterrichtsqualität.

Danksagung Das Projekt Ada*Q wurde im Rahmen des Forschungsprogramms „Wie geht gute Schule?“ der Robert Bosch Stiftung und der Deutschen Schulakademie gefördert.

Literatur

- Ankrum, J. W., Morewood, A. L., Parsons, S. A., Vaughn, M., Parsons, A. W., & Hawkins, P. M. (2020). Documenting adaptive literacy instruction: The adaptive teaching observation protocol (ATOP). *Reading Psychology, 41*(2), 71–86.
- Atlay, C., Tieben, N., Fauth, B., & Hillmert, S. (2019). The role of socioeconomic background and prior achievement for students' perception of teacher support. *British Journal of Sociology of Education, 40*(7), 970–991.
- Bardach, L., Yanagida, T., Schober, B., & Lüftenegger, M. (2018). Within-class consensus on classroom goal structures: Relations to achievement and achievement goals in mathematics and language classes. *Learning and Individual Differences, 67*, 78–90.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180.
- Becker, E. S., & Keller, M. M. (2022). Ich fand die Unterrichtsstunde interessant – du etwa nicht? *Zeitschrift für Erziehungswissenschaft, 25*(2), 405–425.

- Beutel, S.-I., Höhmann, K., Schratz, M., & Pant, H. A. (Hrsg.). (2017). *Handbuch gute Schule. Sechs Qualitätsbereiche für eine zukunftsweisende Praxis* (2. Aufl). Klett/Kallmeyer.
- Bliese, P., Chen, G., Dowens, P., Schepker, D., & Lang, J. (2022). Package 'multilevel' [Computer software]. <https://cran.r-project.org/web/packages/multilevel/multilevel.pdf>.
- Bondie, R. S., Dahnke, C., & Zusho, A. (2019). How does changing "One-Size-Fits-All" to differentiated instruction affect teaching? *Review of Research in Education*, 43(1), 336–362.
- Bohl, T., Batzel, A., & Richey, P. (2012). Öffnung – Differenzierung – Individualisierung – Adaptivität. Charakteristika, didaktische Implikationen und Forschungsbefunde verwandter Unterrichtskonzepte zum Umgang mit Heterogenität. In T. Bohl, M. Bönsch, M. Trautmann, & B. Wischer (Hrsg.), *Binnendifferenzierung: Didaktische Grundlagen und Forschungsergebnisse zur Binnendifferenzierung im Unterricht* (S. 40–69). Prolog-Verl.
- Burke, M. J., & Dunlap, W. P. (2002). Estimating interrater agreement with the average deviation index: A user's guide. *Organizational Research Methods*, 5(2), 159–172.
- Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods*, 2(1), 49–68.
- Charalambos, Y., & Praetorius, A.-P. (2020). Creating a forum for researching teaching and its quality more synergistically. *Studies in Educational Evaluation*, 67, Article 100894.
- Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, 43(3), 161–173.
- Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching. A project of the American Educational Research Association* (S. 608–629). Macmillan.
- Decristan, J., Hess, M., Holzberger, D., & Praetorius, A.-K. (2020). Oberflächen- und Tiefenmerkmale: Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. *Zeitschrift für Pädagogik*, 66, 102–116.
- Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., Hondrich, A. L., Rieser, S., Hertel, S., & Hardy, I. (2015). Embedded formative assessment and classroom process quality. *American Educational Research Journal*, 52(6), 1133–1159.
- Decristan, J., Naumann, A., Fauth, B., Rieser, S., Büttner, G., & Klieme, E. (2014). Heterogenität von Schülerleistungen in der Grundschule. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46(4), 181–190.
- Dumont, H. (2016). Die empirische Untersuchung von individueller Förderung als Perspektive für die Unterrichtsqualitätsforschung. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (S. 107–116). Waxmann.
- Dumont, H. (2019). Neuer Schlauch für alten Wein? Eine konzeptuelle Betrachtung von individueller Förderung im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 22, 249–277.
- Dumont, H., Decristan, J., & Fauth, B. (2024). Adaptiver Unterricht erfordert eine adaptive Unterrichtsforschung. *Unterrichtswissenschaft*, 52, 199–211.
- Enders, C. K., & Tofghi, D. (2007). Centering predictor variables in cross-sectional multi-level models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138.
- Fauth, B., Decristan, J. & Dumont, H. (2024). Kognitive Aktivierung und adaptiver Unterricht: Konzeptionelle Verbindungen und Weiterentwicklungen. In A.-K. Praetorius, W. Wemmer-Rogh, P. Schreyer & M. Brinkmann (Hrsg.), *Kognitive Aktivierung unter der*

- Lupe Möglichkeiten und Herausforderungen der Weiterentwicklung eines prominenten Konstrukts* (S.248–260). Waxmann.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2016). Urteile zur Unterrichtsqualität von Grundschüler/innen: Mehr als “Ich mag meine Lehrerin”? Facetten grundschulpädagogischer und -didaktischer Forschung. In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler, & R. Schmidt (Hrsg.), *Jahrbuch Grundschulforschung* (Bd. 20, S. 157–162). Springer VS.
- Fauth, B., Wagner, W., Bertram, C., Göllner, R., Roloff, J., Lüdtke, O., Polikoff, M. S., Klusmann, U., & Trautwein, U. (2020). Don't blame the teacher? The need to account for classroom characteristics in evaluations of teaching quality. *Journal of Educational Psychology, 112*(6), 1284–1302.
- Ganzeboom, H. (2010). *Questions and answers about ISEI-08*. <http://www.harryganzeboom.nl/isco08/qa-isei-08.htm>. Zugegriffen: 8. März 2024.
- Gärtner, H. (2010). Wie Schülerinnen und Schüler ihre Lernumwelt wahrnehmen. *Zeitschrift für Pädagogische Psychologie, 24*(2), 111–122.
- Göllner, R., Wagner, W., Eccles, J. S., & Trautwein, U. (2018). Students' idiosyncratic perceptions of teaching quality in mathematics: A result of rater tendency alone or an expression of dyadic effects between students and teachers? *Journal of Educational Psychology, 110*(5), 709–725.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.
- Gräsel, C., Hertel, S., Rohlf, C., & Fingerle, M. (2022). Adaptive Lernumgebungen. In M. Harring, C. Rohlf, & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (S. 509–517). Waxmann.
- Hardy, I., Decristan, J., & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for Educational Research Online, 11*, 169–191.
- Hardy, I., Meschede, N., & Mannel, S. (2022). Measuring adaptive teaching in classroom discourse: Effects on student learning in elementary science education. *Frontiers in Education, 7*.
- Hess, M., & Lipowsky, F. (2017). Lernen individualisieren und Unterrichtsqualität verbessern. In F. Heinzel & K. Koch (Hrsg.), *Individualisierung im Grundschulunterricht* (S. 23–31). Springer VS.
- Igler, J., Ohle-Peters, A., & McElvany, N. (2019). Mit den Augen eines Grundschulkindes. Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität. *Zeitschrift für Pädagogische Psychologie, 33*(3–4), 191–205.
- Jager, L., Denessen, E., Cillessen, A., & Meijer, P. C. (2022). Capturing instructional differentiation in educational research: Investigating opportunities and challenges. *Educational Research, 64*(2), 224–241.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*(1), 85–98.
- James, L. R., Demaree, R. G., & Wolf, G. (1993). Rwg: An assessment of within-group interrater agreement. *Journal of Applied Psychology, 78*(2), 306–309.

- Jerusalem, M., Drössler, S., Kleine, D., Klein-Heßling, J., Mittag, W., & Röder, B. (2009). *Förderung von Selbstwirksamkeit und Selbstbestimmung im Unterricht: Skalen zur Erfassung von Lehrer- und Schülermerkmalen*. Zugegriffen: 3. Jan. 2024.
- Jonkmann, K., Rose, N., & Trautwein, U. (Hrsg.) (2013). *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen – Abschlussbericht für die Länder Baden-Württemberg und Sachsen*. Projektbericht an die Kultusministerien der Länder.
- Klieme, E. (2022). Unterrichtsqualität. In M. Harring, C. Rohlfis & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (S. 411–426). Waxmann.
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur“ und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (Hrsg.), *TIMMS – Impulse für Schule und Unterricht: Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (S. 43–57). BMBF Publik.
- Klieme, E., & Warwas, J. (2011). Konzepte der Individuellen Förderung. *Zeitschrift für Pädagogik*, 57(6), 805–818.
- Kunter, M., & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85–113). Waxmann.
- Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement*, 23(2), 127–135.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19(6), 527–537.
- Lüdtke, O., & Robitzsch, A. (2011). Umgang mit fehlenden Daten in der empirischen Bildungsforschung. In S. Maschke & L. Stecher (Hrsg.), *Enzyklopädie Erziehungswissenschaft online: Methoden der empirischen erziehungswissenschaftlichen Forschung*. Beltz Juventa.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, 34(2), 120–131.
- Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Analyse von Lernumwelten. *Zeitschrift für Pädagogische Psychologie*, 20(1/2), 85–96.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86–92.
- McElvany, N., Lorenz, R., Frey, A., Goldhammer, F., Schilcher, A., & Stubbe, T. C. (Hrsg.) (2023). *IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre*. Waxmann.
- Muthén, L. K., & Muthén, B. (2012–2022). *Mplus user's guide*. Muthén & Muthén.
- Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., Pierczynski, M., & Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, 88(2), 205–242.
- Praetorius, A.-K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM*, 50(3), 535–553.

- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM*, 50(3), 407–426.
- Praetorius, A.-K., Klieme, E., Kleickmann, T., Brunner, E., Lindmeier, A., Taut, S., & Charalambous, C. (2020). Towards developing a theory of generic teaching quality: Origin, current status, and necessary next steps regarding the three basic dimensions model. *Zeitschrift für Pädagogik*, 66(Beiheft 66), 15–35.
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Randi, J. (2022). Adaptive teaching. In T. L. Good & M. McCaslin (Hrsg.), *Routledge encyclopedia of education, educational psychology*. Routledge.
- Raudenbush, S. W., & Bryk, A. S. (2010). *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Rieser, S., & Decristan, J. (2025). Kognitive Aktivierung in Befragungen von Schülerinnen und Schülern: Unterscheidung zwischen dem Potenzial zur kognitiven Aktivierung und der individuellen kognitiven Aktivierung. *Zeitschrift für Pädagogische Psychologie*, 39(3), 155–169.
- Rubin, D. B. (1987). *Wiley series in probability and statistics: Multiple imputation for nonresponse in surveys*. Wiley.
- Ruelmann, M., Torchetti, L., Zulliger, S., Buholzer, A., & Praetorius, A.-K. (2021). Kognitiv-motivationale Schüler*innenprofile und ihre Bedeutung für die Schüler*innenwahrnehmung der Lernunterstützung durch die Lehrperson. *Unterrichtswissenschaft*, 49(3), 395–422.
- Schenke, K., Ruzek, E., Lam, A. C., Karabenick, S. A., & Eccles, J. S. (2018). To the means and beyond: Understanding variation in students' perceptions of teacher emotional support. *Learning and Instruction*, 55, 13–21.
- Schipper, T. M., van der Lans, R. M., de Vries, S., Goei, S. L., & van Veen, K. (2020). Becoming a more adaptive teacher through collaborating in Lesson Study? Examining the influence of Lesson Study on teachers' adaptive teaching practices in mainstream secondary education. *Teaching and Teacher Education*, 88, Article 102961.
- Schratz, M. (Hrsg.). (2013). *Was für Schulen! Der Deutsche Schulpreis 2013*. Klett/Kallmeyer.
- Schweig, J. D. (2016). Moving beyond means: Revealing features of the learning environment by investigating the consensus among student ratings. *Learning Environments Research*, 19(3), 441–462.
- Schweig, J. D., & Martinez, J. F. (2021). Understanding (dis)agreement in student ratings of teaching and quality of the learning environment. In W. Rollett, H. Bijlsma & S. Röhl (Hrsg.), *Student feedback on teaching in schools. Using student perceptions for the development of teaching and teachers* (S. 91–110). Springer International Publishing.
- Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Hrsg.), *Encyclopedia of statistics in behavioral science* (S. 1570–1573). Wiley.
- 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.

- Stang, J., & McElvany, N. (2021). Unterschiede in der Wahrnehmung der Qualität des Deutschunterrichts zwischen Grundschülerinnen und Grundschülern. *Zeitschrift für Pädagogische Psychologie*, 35(4), 261–273.
- Tetzlaff, L., Edelsbrunner, P., Schmitterer, A., Hartmann, U., & Brod, G. (2023). Modeling interactions between multivariate learner characteristics and interventions: A person-centered approach. *Educational Psychology Review*, 35, 112.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67.
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271–296.
- van de Pol, J., Vries, N. de, Poorthuis, A. M., & Mainhard, T. (2022). The questionnaire on teacher support adaptivity (QTSA): Reliability and validity of student perceptions. *The Journal of Experimental Education*, 1–33.
- Vaughn, M. (2019). Adaptive teaching during reading instruction: A multi-case study. *Reading Psychology*, 40(1), 1–33.
- Vaughn, M., Parsons, S. A., & Gallagher, M. A. (2022). Challenging scripted curricula with adaptive teaching. *Educational Researcher*, 51(3), 186–196.
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M., & Pietsch, M. (2020). Angebot-Nutzungs-Modelle der Wirkweise des Unterrichts: Ein kritischer Vergleich verschiedener Modellvarianten. *Zeitschrift für Pädagogik*, 66, 63–80.
- Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction*, 28, 1–11.
- Walter, J. (2013). *VSL: Verlaufsdiagnostik sinnerfassenden Lesens*. Hogrefe.
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 (CFT 20-R)*. Hogrefe.
- Willems, A. S. (2022). Individuelle Schüler*innenprofile des situationalen und dispositionalen Interesses und ihre Bedeutung für die Wahrnehmung der Unterrichtsqualität im Fach Mathematik. *Zeitschrift für Erziehungswissenschaft ZfE*, 25(2), 377–404.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





FLINK: Evaluation of a Digitized Reading Fluency Training

9

Fenke Kachisi, Jan-Henning Ehm, Jelena Marković
and Telse Nagler

Abstract

Purpose: This study investigates whether FLINK, a newly developed digital reading training addressing silent reading fluency, is an effective way to increase the reading speed in an external standardized reading test (*quop*) of participating children. **Method:** 725 German-speaking third graders participated in FLINK as well as in the learning progress assessment platform *quop*, in which their reading speed was continuously monitored throughout the school year. A latent growth curve model was used to evaluate training gains with group (high vs. low training intensity) as predictor. **Results:** The analyses

F. Kachisi (✉) · J. Marković · T. Nagler

DIPF Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

E-Mail: f.kachisi@dipf.de

J. Marković

E-Mail: j.markovic@dipf.de

T. Nagler

E-Mail: t.nagler@dipf.de

J.-H. Ehm

Pädagogische Hochschule Heidelberg, Heidelberg, Germany

E-Mail: jan.ehm@ph-heidelberg.de

T. Nagler

Universität Bremen, Bremen, Germany

F. Kachisi · J.-H. Ehm · J. Marković · T. Nagler

IDeA – Center for Research on Individual Development and Adaptive Education of Children at Risk, Frankfurt am Main, Germany

© Der/die Autor(en) 2025

J. Ziehm-Eicher et al. (Hrsg.), *Kinder mit erhöhtem Risiko für Bildungsmisserfolg*,
Edition ZfE 20, https://doi.org/10.1007/978-3-658-48674-7_9

255

revealed a significant linear and quadratic growth of reading speed over the school year. The group with high training intensity showed a higher growth in the external reading speed test, confirming the effectiveness of the training approach regarding the improvement of reading speed. Conclusion: The digital reading fluency training FLINK, implemented in the learning progress assessment platform *quop*, can be considered as an effective way to foster reading speed in German-speaking children in third grade. Further analyses should focus on long-term and time-specific training effects simultaneously in order to obtain insights into the training's mechanisms of action.

Keywords

Reading fluency • Digital training • Reading difficulties • Text-fading

Introduction

The acquisition of literacy is considered to be the most far-reaching factor influencing school achievement. Reading skills have a particular influence on school success, as fluent and meaningful reading is required in most school subjects in order to understand instructions and complete tasks (Schneider, 2017). Since there are unfortunately many children who have difficulties with reading acquisition, effective training programs are needed to help these children and, ideally, to enable them to catch up with typically reading children. For this reason, the present study examines the effectiveness of a newly developed digital reading training aimed at improving reading speed in an external standardized reading test.

Theoretical and Empirical Framework

Successful reading acquisition is reflected in skilled and automatized reading. The degree of accuracy and speed of reading determines how well words, sentences or texts can be understood (Ehm et al., 2023; Huemer et al., 2008; Nagler et al., 2018). From a theoretical perspective, two-route models, such as the dual-route model (Coltheart, 1978; Coltheart et al., 1993, 2001) can be used to explain the difference between automatized and non-automatized reading of words. Coltheart et al. (1993) distinguish between two separate processing routes that can be

used during reading: a lexical and a non-lexical processing route. If an unknown word is read via the non-lexical (phonological) route, single letters are converted into sounds according to the learned grapheme-phoneme correspondence rules (phonological recoding). Hence, the word is read letter by letter and processed phonologically to determine the meaning of the word (decoding, Coltheart et al., 1993). The use of this non-lexical route is a laborious process. It requires the recognition of letters and the correct allocation with corresponding sounds as well as the retention of sounds for successful consolidation of the word's sound sequence (Ennemoser et al., 2012). The non-lexical route strains the working memory capacities and limits the resources available for higher processes, such as reading comprehension (LaBerge & Samuels, 1974; Perfetti, 1986). When using the lexical route for word recognition, the presented letter clusters activate word representations that are stored in the mental lexicon and can be recognized as an entity. This way of processing enables the direct retrieval of a word's meaning from memory (Coltheart et al., 1993), thereby releasing cognitive resources that can be used for reading comprehension.

Whether the lexical or non-lexical route is activated depends to a large extent on individual reading skills. Although Coltheart's model was not designed as a developmental model of reading, it can add to the understanding of processes involved in learning how to read. To describe reading acquisition, Frith's theoretical approach, which suggests successive developmental stages of the reading acquisition process (Frith, 1985, 1986), can be referred to. In the proposed *logographic stage*, children are able to recognize familiar words by their graphic features alone. With the beginning of reading instruction, children advance to the *alphabetic stage*, where basic grapheme-phoneme correspondences are applied to decode words via the non-lexical route. With increasing reading experience, children build up their mental lexicon by storing representations of word parts and whole words in their memory (cf. Grainger et al., 2012). The non-lexical route is therefore necessary to build up sufficient representations in the mental lexicon in order to eventually retrieve needed information directly. When high-quality representations have accumulated in the mental lexicon, children can progress to the *orthographic stage* where they are able to retrieve larger orthographic units or whole words directly from the mental lexicon (Frith, 1985, 1986), which corresponds to the lexical route in Coltheart's model.

If the reading process is sufficiently automatized and orthographic units (such as parts of words or whole words) can be retrieved directly from memory (Wolf & Katzir-Cohen, 2001), the reading process can be considered fluent. Reading fluency is defined as a sufficient degree of accuracy and speed in reading (Nagler et al., 2018) that enables reading comprehension (Kim et al., 2012; Lee & Chen,

2019). Accordingly, reading fluency is considered a crucial bridge between decoding and comprehension (Lervåg & Melby-Lervåg, 2022). Most children learn to read fluently through practice and schooling during their primary school years (Huemer et al., 2008). Unfortunately, there are also children who show problematic reading acquisition. Children with reading difficulties show various reading errors, e.g., omitting or adding letters, as well as poor reading speed and reading comprehension (Thomas et al., 2015).

From a pedagogical point of view, it is of particular interest to identify problems in reading acquisition as early as possible, as unrecognized reading problems are usually persistent (Kohn et al., 2013) and early intervention essential. Interventions for children with reading difficulties that show the best efficacy (Ise et al., 2012; Schulte-Körne, 2017) are symptom-oriented, i.e., they focus directly on the reading process, and involve a high number of sessions (Galuschka et al., 2014; Ise et al., 2012). The most prominent symptom of children with reading difficulties in transparent languages (such as German) is poor reading speed (Landerl et al., 1997), resulting in poor reading fluency and comprehension (LaBerge & Samuels, 1974). In opaque languages (such as English) impaired reading accuracy is additionally characteristic (Parrila et al., 2005; Ziegler et al., 2003). One example of symptom-specific training that focuses on reading fluency is repeated reading (originally developed by Dahl (1979) and Samuels (1979)), which typically consists of reading words or passages of text repeatedly out loud in order to increase the automatization of reading and thus reading fluency (Hintikka et al., 2008). Repeated reading can improve the reading speed and accuracy of participating children, but training effects are mainly found for trained words and transfer effects are rather small (Berends & Reitsma, 2006). Furthermore, the often necessary one-to-one assistance during training is time consuming and cost intensive. Further development of reading fluency trainings is therefore needed, especially focusing on the generalizability of effects in an ideally economically manner (i.e., several children can work simultaneously in one room, one-to-one supervision is not necessary so that costs and personnel resources can be reduced). Furthermore, the focus should be on silent reading, as children mainly engage in silent reading in their daily routine (cf. Hiebert and Reutzel, 2010).

In the literature, a manipulation in which text presented on a screen is faded out letter by letter in the direction of reading is described as an effective pedagogical-psychological procedure to increase silent reading fluency and reading comprehension (Bar-Kochva & Hasselhorn, 2015; Breznitz, 1987a, b; Nagler et al., 2014, 2016). The concept of this approach is based on the so-called *Acceleration Phenomenon*. This phenomenon describes that people who are artificially encouraged to read faster (by continuously fading out computerized

presented text letter by letter, i.e., text-fading) show better reading performance in terms of reading accuracy, reading speed and reading comprehension (Breznitz, 1987b). Especially children with reading difficulties seem to benefit from the text-fading manipulation (Breznitz, 1987b, 1997a, b). In comparison to typically reading samples, children with reading difficulties showed greater reading gains as a result of the text-fading manipulation.

Several attempts have been made to transfer the text-fading approach into a training intervention to improve silent reading fluency and reading comprehension. The effectiveness of such a training has been demonstrated, for example, for unimpaired reading children (e.g., Nagler et al., 2015), for adult readers with reading difficulties (Breznitz et al., 2013) as well as for readers in different orthographies (e.g., Chinese, Dai et al., 2016; Hebrew, English, Horowitz-Kraus et al., 2014; Spanish, López-Escribano, 2016; German, Korinth et al., 2016). In most intervention studies focusing on the text-fading approach, an increased reading speed with constant reading comprehension could be observed (for an overview see Korinth & Nagler, 2021). In order to examine whether this text-fading manipulation can also foster German children with reading difficulties, a training program for elementary school children was developed. The conducted text-fading training study revealed lasting effects on reading comprehension for children with reading difficulties (Nagler et al., 2021). These results are encouraging to transfer this text-fading training approach into a more efficient format that can reach a large number of children and extend the results from previous training studies.

For this purpose, a digital, browser-based training program addressing reading fluency based on the text-fading approach was developed. The training program *FLINK* (Flüssig lesen lernen – Individuelle Förderung der Lesekompetenz [*How to Read Fluently: Individual Intervention of Reading Competencies*]) was implemented in the *quop* learning progress assessments (LPA) platform (Förster et al., 2023; Souvignier et al., 2021). *Quop* was developed in Germany to provide teachers with an opportunity to monitor their students' reading and mathematics development over the course of the school year from grades one to six (Souvignier et al., 2021). Offering information on children's achievement and learning progress to teachers can help to improve children's learning (Souvignier et al., 2021). For reading and mathematics, ten parallel tests are available for each grade level. The tests are administered approximately every three weeks and have a duration of ten to fifteen minutes. The reading tests for grade three examine reading accuracy, reading speed, as well as text- and knowledge-based reading comprehension. The tests include fictional and non-fictional texts. The children have to complete varying task formats such as multiple-choice questions

or verification tasks (Souvignier et al., 2021). As *quop* has proven to be effective (Förster & Souvignier, 2014, 2015), this particular LPA seems promising for monitoring and improving children's learning success.

The implementation of FLINK in *quop* allows individual reading performance to be tracked and, if necessary, to react quickly to difficulties or performance drops. By digitizing the text-fading training and implementing it in *quop*, it is possible to reach a large number of children who participate in the training program. The evaluation of FLINK's effectiveness on the basis of a large sample generates valuable information and enables further development.

Persistent reading difficulties and the necessity to start reading intervention early highlight the need for efficient individual support that can reach many affected children at the lowest possible educational barriers. The aim of this study is therefore to evaluate the effectiveness of the newly developed FLINK reading fluency training. The influence of FLINK on *quop* reading speed measures is examined. Based on the previous training studies with text-fading showing training effects in external reading measures that were maintained over a period of six months (cf. Korinth & Nagler, 2021; Nagler et al., 2021), we hypothesize that gains in reading speed can be observed, especially in a group of children who have participated in many training sessions.

Method

A two-group design with eight measurement points was used to investigate training intensity effects of FLINK on the external *quop* measure of reading speed. The sample was divided into children who had completed numerous training sessions (high training intensity; experimental group) and children who had completed only a few training sessions (low training intensity; control group).

Participants and Procedure

$N = 1437$ unselected third grade children participated in a first implementation attempt of FLINK in the 2022/2023 school year. Recruitment efforts continued throughout the school year. Thus, teachers were able to register their classes or selected children during the whole course of the school year. This led to a variety of different training starts. As a result, many children were not able to complete the maximum number of training sessions. Only children who started FLINK or *quop* before May 15th, 2023, were included for data analyses,

to ensure that an informative amount of *quop* tests were realized, leading to a final sample of $N = 725$ children (47.45 % female, age $M = 8.75$, $SD = 0.53$) from mostly Baden-Württemberg, Hesse and North Rhine-Westphalia. The low training intensity group (participation in one to six training sessions, $M = 3.06$ sessions) included $N = 474$ children; the high training intensity group (participation in seven to 40 sessions, $M = 13.2$ sessions) comprised $N = 251$ children. As shown in Fig. 9.1, the low training intensity group completed a great proportion of sessions in April/May 2023, so that their low training participation can be explained by a late start in the school year.

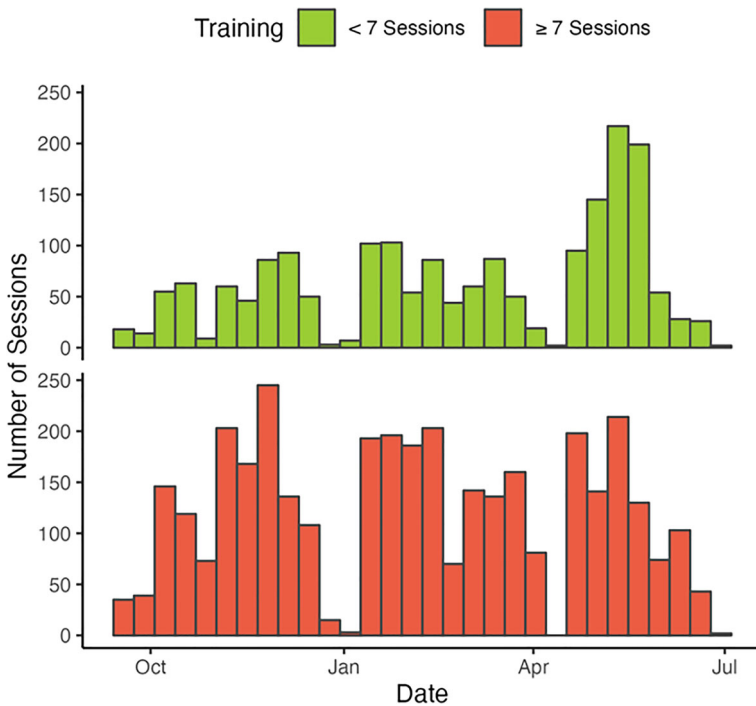


Fig. 9.1 Overview of the Timing of Completed Training Sessions by all Children in the Respective Month Divided by Group

Note. Green = low intensity group; less than 7 sessions; red = high intensity group; 7 and more sessions.

Materials

FLINK Reading Fluency Training

FLINK is designed to enable teachers who participate in *quop* to independently activate the reading fluency training for their pupils. Teachers were advised to particularly register children in the middle and lower reading performance range and base their selection on the children's *quop* tests results. However, which children were selected for participation in FLINK was at the teacher's discretion. It was recommended to carry out the training twice a week, which corresponds to six FLINK sessions between two *quop* tests, leading to a maximum of 54 possible sessions per school year. The suggestions were deliberately left open, as a preset design may restrict the everyday school life of teachers.

Each of the possible 54 training sessions contained 30 items. The items were self-developed and extended based on the reading material used in previous studies (Nagler et al., 2015, 2021). Each item consisted of a sentence (4 to 22 words, mean: 13.398 words) and a respective multiple-choice question containing four possible answer options. To ensure attentive reading of the sentences, questions and multiple-choice answers used different wording and meaningful distractor items (cf. Rost & Sparfeldt, 2007). In the first training session, all children started with the lowest fading rate at level 1 (i.e., 500 ms per letter, which can be accelerated up to level 10, i.e., 50 ms per letter) that was determined on the basis of the previous training study (see Nagler et al., 2021). This was designed to avoid a potential overload of children's initial capacities. Within a FLINK session, the rate at which the text is faded out is adapted depending on the child's comprehension performance. If a child continuously shows a high level of comprehension, i.e., the questions related to the sentences are answered correctly (in more than 75% to 90% of the items), the child advances to the next fading rate level and the text is faded out slightly faster. If the questions are answered incorrectly several times (i.e., lower than 30 % to 45 % correctly answered items), the fading rate is reduced and the text is faded out at the precedent level.

A single reading item is processed as follows: After a child presses "Go!", the sentence is visible and fades out after a short delay. If the sentence is completely faded out, or if the child presses "Next" because they finished reading earlier, a comprehension question appears. This question must be answered before moving on to the next item. In this procedure, several measures are recorded: comprehension performance (correct answer to multiple-choice questions), the fading level achieved (depending on comprehension performance), and the actual reading speed (how quickly the children press "Next" before the sentence is entirely faded out).

Reading Speed

To measure reading speed via *quop* in third grade, children are presented with a maze task in which every seventh word is missing. Children are instructed to select the correct missing word from three options (the correct target and two distractors) as quickly and as accurately as possible. The maze tasks comprised a total of 130 to 138 words and 17 gaps. Reading speed is assessed by the time needed to complete the entire maze task (cf. Förster et al., 2018). The reading speed tests were administered every three weeks throughout the school year. The results from measurement points 1 to 8 were included in the present analyses.

Statistical Analyses

In order to test our hypothesis, we estimated the development of reading speed as a latent growth curve model (LGCM) predicted by group membership. All statistical analyses were conducted using R version 4.1.1, and models were fitted using the lavaan package (Rosseel, 2012). Model fit was evaluated with several commonly recommended descriptive measures of model fit (Hu & Bentler, 1998): the standardized root mean square residual (SRMR) and the root mean square error of approximation (RMSEA), the Tucker-Lewis Index (TLI), as well as the comparative fit index (CFI). Models were evaluated as fitting satisfactorily when the baseline fit indices (SRMR and RMSEA) were <0.08 and the incremental fit index (CFI) was .90 or larger.

Results

A precondition for examining the influence of the training on reading development is to model the growth in reading speed. To fit these growth curves, we estimated two LGCMs, a linear growth model (Model 1) and a quadratic growth model (Model 2). As can be seen in Table 9.1, both models fit the data well (linear growth model: $\chi^2(38) = 120.565$, $p < .001$, CFI = .949, TLI = .963, RMSEA = .055, SRMR = .059; quadratic growth model: $\chi^2(33) = 98.092$, $p < .001$, CFI = .960, TLI = .966, RMSEA = .052, SRMR = .054).

According to the Chi-Square Difference Test, the quadratic growth model had a slightly better fit ($\chi^2_{\text{diff}} = 22.473$, $df = 5$, $p < .001$). Thus, this model was used to examine the research question. In this model, significant variance was found in the intercept ($\sigma^2 = 84.238$) but not in the linear ($\sigma^2 = 2.412$) or quadratic slope ($\sigma^2 = 0.027$). That is, children differ in their initial values but show comparable

Table 9.1 *The Goodness of Fit of the Models*

Model	$\chi^2(df)$	CFI	TLI	RMSEA	SRMR	AIC
(1) Linear Growth	120.565 (38)	.949	.963	.055	.059	30616.234
(2) Quadratic Growth	98.092 (33)	.960	.966	.052	.054	30603.760

Note. CFI = comparative fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean squared residual; AIC = Akaike Information Criterion

patterns of growth. The model estimates indicate that children showed a mean linear of $\mu_s = 1.477$ ($z = 6.049$, $p < .001$) and quadratic growth of $\mu_s = 0.070$ ($z = 2.213$, $p < .05$) from one time point to the next. Furthermore, negative but not significant associations were found between the intercept and the growth factors ($\text{cov}\sigma^2_{\text{intercept/linear slope}} = -3.81$; $r = -0.086$; $\text{cov}\sigma^2_{\text{intercept/quadratic slope}} = -0.403$; $r = -0.268$), and between the growth factors ($\text{cov}\sigma^2_{\text{linear slope/quadratic slope}} = -0.140$; $r = -0.548$).

To test whether the training intensity influences reading speed development, a dichotomous variable (Group: 0 = less than seven sessions; 1 = greater than or equal to seven sessions) was included in the model. We chose this option in order to come as close as possible to a classic experimental procedure. The formation of the two groups based on the cut-off of at least seven sessions is thereby justified by a certain pragmatism. In a previous training study (Nagler et al., 2021), we were able to demonstrate training effects with an average number of 12 sessions. In order to have a comparable number of sessions in the “high intensity group” ($M = 13.2$) and at the same time to have groups of nearly similar size, we decided to use the cut-off value of seven.

The results of this model are depicted in Fig. 9.2 and show that the coefficient of group on the intercept was $\beta = -0.074$ ($SE = 1.077$) and not statistically significant ($z = -1.325$, $p = .185$), indicating no group difference in the level of reading speed. However, of central interest, the coefficient of group on the linear slope was $\beta = 0.297$ ($SE = 0.490$) and statistically significant ($z = 1.973$, $p = .048$). The effect of the group membership on the quadratic growth was $\beta = -0.294$ ($SE = 0.068$) and not statistically significant ($z = -1.519$, $p = .129$). The different patterns of growth for the two groups are shown in Fig. 9.3.

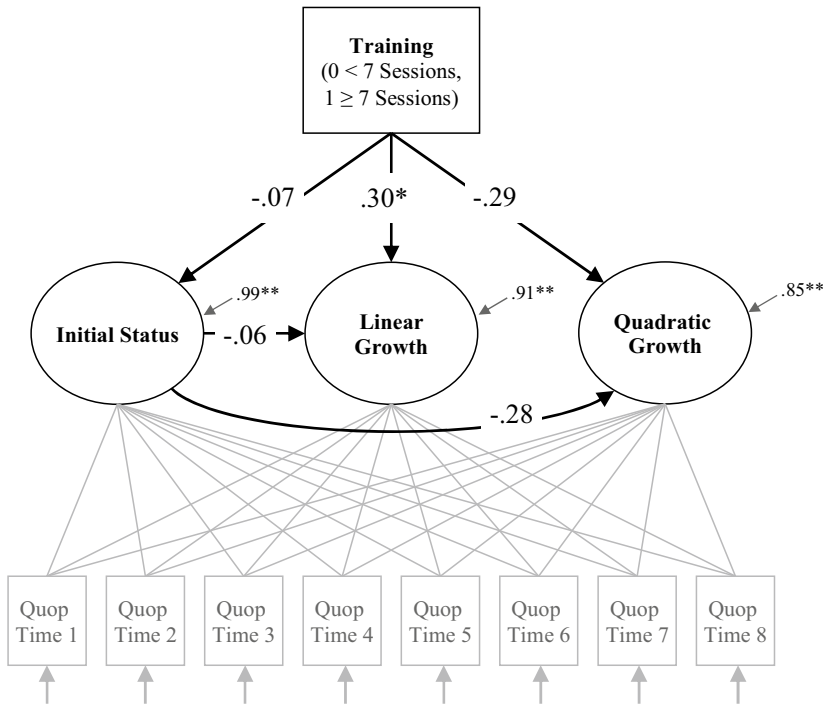


Fig. 9.2 The Quadratic Latent Growth Curve Model of Reading Speed Predicted by Group *Note*. For the sake of clarity, the correlation between the residuals of the linear and quadratic growth factors ($r = -.56, p .05$) and the loadings of the indicators are not shown. Parameters shown are standardized. ** $p .01$. * $p .05$.

Discussion

The present study investigated training effects through FLINK on reading speed gains in *quop* of German third graders. Regarding our hypothesis on training effects, we were able to confirm the effectiveness of the FLINK training on reading speed gains in the external standardized *quop* reading speed test with our latent growth curve model. Children who accomplished more training sessions had a stronger growth of reading speed than children who accomplished less training sessions. Controlling for interindividual differences regarding the intercept, children with a high training intensity reached an average speed gain of 2.154

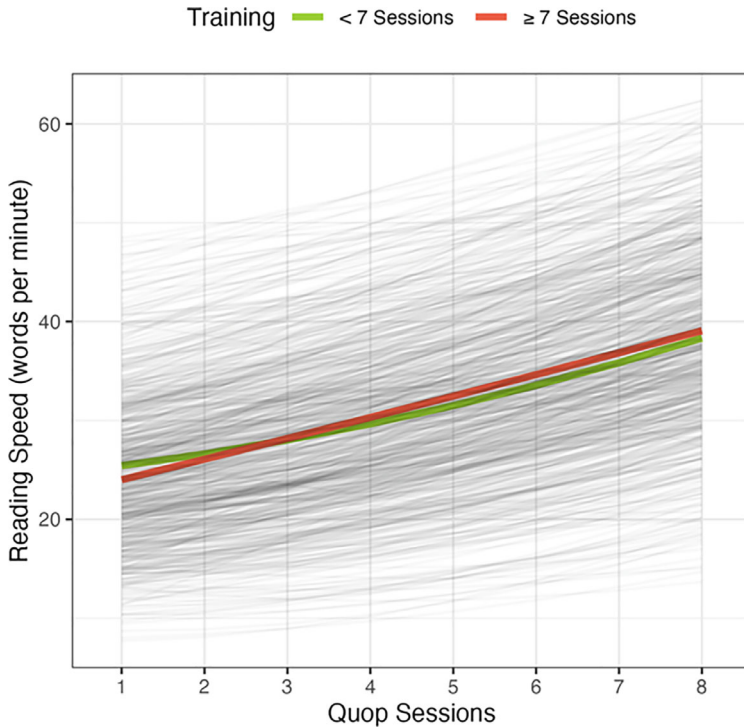


Fig. 9.3 *Note.* Visualization of the reading speed development over the 8 measurement time points, divided by group (green = less than 7 sessions; red = 7 or more sessions)

words per minute, whereas children with a low training intensity reached an average speed gain of 1.844 words per minute from one measurement point to the next. These findings support previous research, demonstrating reading improvements through text-fading training in typically reading children (Nagler et al., 2015) as well as children with reading difficulties (Nagler et al., 2021). As previously proposed (e.g., Breznitz, 1987a, 1997b; Nagler et al., 2014, 2021), the training effects may be explained by a switch to more efficient reading processes due to the perceived time constraint during the training. To explain this in terms of the dual route model (Coltheart et al., 1993), children might reduce their use of phonological recoding through training and engage more often in lexical processing instead. The perceived time constraint may have caused children to focus

less on reliable but slow phonological recoding (in a transparent language like German, recoding phonologically leads quite consistently to the correct word meaning), but rather focus on the fast retrieval of whole words or parts of words. This potentially increased use of direct retrieval may have led to greater automatization and a faster reading routine. In addition, through training the switch between the lexical and non-lexical route may have become faster and children may have carried out non-lexical processes more efficiently (cf. Siegler, 2006). It is possible that children may progress from less efficient reading processes within the non-lexical route (i.e., the laborious sequencing of the word's individual letters) to more efficient processes (e.g., the sequencing of larger letter units such as syllables or other parts of words (cf. Lindberg et al., 2011). This may especially be the case for children who still mostly rely on the phonological route. Empirical results imply that word reading skills must be developed to a certain degree before an increase in these skills has an effect on reading speed. Such threshold values have already been found for the non-transparent Danish language (Juul et al., 2014) as well as for the transparent German language (Karageorgos et al., 2020), showing that a certain threshold of word reading accuracy must be exceeded before word reading speed gains can emerge. Therefore, it may be that FLINK has no beneficial effect for children whose word reading skills are not yet sufficiently developed. For these children, other support measures to strengthen the phonological route first, such as syllable-based measures (e.g., Heß et al., 2024), may be more appropriate to begin with. Future studies should therefore consider possible threshold values in order to individualize the training and fit the individual needs of participating children.

Children in the high-intensity group showed greater reading speed gains over the course of the training than children in the low-intensity group. This finding indicates the effectiveness of the training. However, whether the effects of the training are short-term or long-term still needs to be clarified. To investigate long-term training effects, follow-up tests should be included in future studies. Furthermore, detailed time-specific effects (i.e., influences of the time interval between training and performance measurement) should also be taken into account in future studies. Since the low intensity training group completed only very few sessions, it may be that these sessions were distributed unevenly over the course of the school year. It is possible that many of the completed FLINK sessions were only carried out towards the end of the *quop* test period in this group. This could explain the reduced differences in reading speed at the end. It should also be noted that although the number of FLINK sessions between two *quop* tests was limited to six, the exact intervals at which these should be carried out were not predetermined. This may, for example, have resulted in all possible

FLINK sessions being carried out right at the beginning of the period before the next *quop* test or, conversely, only at the very end. It could be that the training has a stronger effect on *quop* results, if it is carried out directly before the reading test. However, the flexibility of the training implementation, which we consider to be particularly important for teachers' compliance in the implementation of the training, leads to uncertainties in the interpretation of the results. We expect these irregular intervals between FLINK training sessions and *quop* tests to equalize on average over the course of the school year, but we cannot be certain of this. These implications illustrate the importance to particularly model time-specific effects in further studies in order to draw more profound conclusions about the effects of the training on reading speed gains.

Of course, it should also be considered that the effects of the training may depend on the point in the school year at which the training was conducted. In addition to possible temporal effects, it can also make a difference how far the students have progressed with reading instruction in the school year as well as how much reading experience they have. The training may require certain prerequisites (e.g., high-quality knowledge entries in the mental lexicon (cf. Perfetti & Hart, 2002) for the usage of efficient reading strategies) in order to be effective. If these prerequisites are not met, the training may not be effective or even have detrimental effects (cf. Nagler et al., 2014, 2021). The availability of the necessary prerequisites may depend on the progress of the school year (cf. Crone & Whitehurst, 1999; Skibbe et al., 2012). Hence, when interpreting the training effects, it should be considered that children started the training at different time points during the school year, resulting in a possible confounding of training intensity and timing. For this reason, further studies should focus on the individual progressions of training. This makes it possible to identify children for whom training effects are particularly pronounced. A closer look at these children can provide valuable information about particular relevant prerequisites for training success as well as for recommendations regarding the indication or contraindication of training participation.

In summary, the presented data implies that the newly developed reading fluency training FLINK is an effective way to foster reading speed in an external standardized reading speed test. Nevertheless, the question of necessary prerequisites for the success of the text-fading manipulation in particular has not yet been answered and needs further investigation.

Limitations

The present results are limited to German-speaking third graders of an unselected sample. Although teachers were advised to specifically register children in the middle and lower performance range in reading based on *quop* test results, the actual selection process was up to the teachers. Hence, we cannot ensure that teachers acted according to our advice. Further investigation and analyses are needed to clarify whether the training is most effective for children with reading difficulties (as studies on the acceleration phenomenon suggest, e.g., Breznitz, 1987b, 1997a, b), or whether the effects are comparable for reading impaired and typically reading children.

Furthermore, we compared two groups of children who both participated in the training, albeit with varying intensity. In future studies, matched control groups (i.e., same age and grade level with comparable reading proficiency at the beginning of the school year), participating in the *quop* tests but not in the FLINK training, should be included. By including such a control group, it can be explored whether stronger increases in reading speed could be detectable for children participating in the training compared to a control group of children who have not undergone any training at all. At the same time, a control group participating in the identical training and reading the same training material, but without the text-fading, is essential. With such a control group, further possible explanations for reading gains, such as increased reading motivation due to the digital presentation or reading gains due to additional reading practice, can be controlled for (cf. Nagler et al., 2021).

Even though we observed training effects over a period of approximately ten months in this training study, statements about the long-term training effects are not yet possible and should be addressed in future studies.

Conclusions

It was shown that FLINK is an effective training method designed to increase the reading speed of German-speaking elementary school children in third grade. However, only reading speed gains in an external standardized reading test were investigated in a group comparison of children with high training intensity vs. children with low training intensity. In future studies, possible reading accuracy and reading comprehension gains should also be addressed, ideally including a control group not participating in the training. However, the demonstrated effectiveness of the training in terms of reading speed is promising. The training

should be further developed and provided to children (with reading difficulties) to individually foster reading speed.

Acknowledgments We would like to thank the entire *quop* team from hfp and the University of Münster for their support and the implementation of FLINK in the *quop* platform. We would like to thank Thomas Hauser and Johannes Huthmacher from hfp for the technical conception and digital realization of FLINK as well as Elmar Souvignier, Natalie Förster and Ophelia Urbach from the University of Münster for their support with all arising questions.

Funding This study was conducted as part of the joint project “Schule macht stark—SchuMaS [School makes you strong]” and funded by the Federal Ministry of Education and Research (BMBF). Funding reference 01PR2101A.

References

- Bar-Kochva, I., & Hasselhorn, M. (2015). In search of methods enhancing fluency in reading: An examination of the relations between time constraints and processes of reading in readers of German. *Journal of Experimental Child Psychology*, *140*, 140–157. <https://doi.org/10.1016/j.jecp.2015.06.012>.
- Berends, I. E., & Reitsma, P. (2006). Remediation of fluency: Word specific or generalised training effects? *Reading and Writing*, *19*(2), 221–234. <https://doi.org/10.1007/s11145-005-5259-3>.
- Breznitz, Z. (1987a). Increasing first graders’ reading accuracy and comprehension by accelerating their reading rates. *Journal of Educational Psychology*, *79*(3), 236–242. <https://doi.org/10.1037/0022-0663.79.3.236>.
- Breznitz, Z. (1987b). Reducing the gap in reading performance between Israeli lower- and middle-class first-grade pupils. *The Journal of Psychology*, *121*(5), 491–501. <https://doi.org/10.1080/00223980.1987.9915503>.
- Breznitz, Z. (1997a). Effects of accelerated reading rate on memory for text among dyslexic readers. *Journal of Educational Psychology*, *89*(2), 289–297. <https://doi.org/10.1037/0022-0663.89.2.289>.
- Breznitz, Z. (1997b). Enhancing the reading of dyslexic children by reading acceleration and auditory masking. *Journal of Educational Psychology*, *89*(1), 103–113. <https://doi.org/10.1037/0022-0663.89.1.103>.
- Breznitz, Z., Shaul, S., Horowitz-Kraus, T., Sela, I., Nevat, M., & Karni, A. (2013). Enhanced reading by training with imposed time constraint in typical and dyslexic adults. *Nature Communications*, *4*(1), 1486. <https://doi.org/10.1038/ncomms2488>.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Hrsg.), *Strategies of information processing* (S. 151–216). Academic.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589–608. <https://doi.org/10.1037/0033-295X.100.4.589>.

- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. <https://doi.org/10.1037//0033-295x.108.1.204>.
- Crone, D. A., & Whitehurst, G. J. (1999). Age and schooling effects on emergent literacy and early reading skills. *Journal of Educational Psychology*, *91*(4), 604–614. <https://doi.org/10.1037//0022-0663.91.4.604>.
- Dahl, P. J. R. (1979). An experimental program for teaching high speed word recognition and comprehension skills. In J. E. Button, T. Lovitt, & T. Rowland (Eds.), *Communications research in learning disabilities and mental retardation* (S. 33–65). University Park Press. <https://files.eric.ed.gov/fulltext/ED099812.pdf>.
- Dai, L., Zhang, C., & Liu, X. (2016). A special Chinese reading acceleration training paradigm: To enhance the reading fluency and comprehension of Chinese children with reading disabilities. *Frontiers in Psychology*, *7*, 1–7. <https://doi.org/10.3389/fpsyg.2016.01937>.
- Ehm, J. H., Schmitterer, A. M. A., Nagler, T., & Lervåg, A. (2023). The underlying components of growth in decoding and reading comprehension: Findings from a 5-year longitudinal study of German-speaking children. *Scientific Studies of Reading*, *27*(4), 311–333. <https://doi.org/10.1080/10888438.2022.2164199>.
- Ennemoser, M., Marx, P., Weber, J., & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens. Evidenz aus zwei Längsschnittstudien vom Kindergarten bis zur 4. Klasse. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *44*(2), 53–67. <https://doi.org/10.1026/0049-8637/a000057>.
- Förster, N., Erichsen, M., & Forthmann, B. (2023). Measuring reading progress in second grade: Psychometric properties of the quop-L2 test series. *European Journal of Psychological Assessment*, *39*(1), 5–13. <https://doi.org/10.1027/1015-5759/a000688>.
- Förster, N., Kawohl, E., & Souvignier, E. (2018). Short- and long-term effects of assessment-based differentiated reading instruction in general education on reading fluency and reading comprehension. *Learning and Instruction*, *56*, 98–109. <https://doi.org/10.1016/j.learninstruc.2018.04.009>.
- Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction*, *32*, 91–100. <https://doi.org/10.1016/j.learninstruc.2014.02.002>.
- Förster, N., & Souvignier, E. (2015). Effects of providing teachers with information about their students' reading progress. *School Psychology Review*, *44*(1), 60–75. <https://doi.org/10.17105/SPR44-1.60-75>.
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. E. Marshal, & M. Coltheart (Eds.), *Surface dyslexia: Neurological and cognitive studies of phonological reading* (S. 301–330). Lawrence Erlbaum. <https://doi.org/10.1016/j.jcnyhd.2010.08.009>.
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, *36*(1), 69–81. <https://doi.org/10.1007/BF02648022>. 10.1007/BF02648022.
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE*, *9*(2), 1–12. <https://doi.org/10.1371/journal.pone.0089900>.

- Grainger, J., L  t  , B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition*, *123*(2), 280–292. <https://doi.org/10.1016/j.cognition.2012.01.003>.
- He  , J., Riedmann, A., Karageorgos, P., Schaper, P., Lugin, B., Richter, T., & M  ller, B. (2024). MobiLe: Konzeption einer digitalen silbenbasierten Lesef  rderung f  r die Grundschule. *Psychologie in Erziehung Und Unterricht*, *71*(1), 41–51. <https://doi.org/10.2378/peu2024.art05d>.
- Hiebert, E. H., & Reutzel, D. R. (2010). *Revisiting silent seading: New directions for teachers and researchers*. International Reading Association. <https://doi.org/10.1598/0833>
- Hintikka, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency: Is it important to practice reading aloud and is generalization possible? *Annals of Dyslexia*, *58*(1), 59–79. <https://doi.org/10.1007/s11881-008-0012-7>.
- Horowitz-Kraus, T., Cicchino, N., Amiel, M., Holland, S. K., & Breznitz, Z. (2014). Reading improvement in English- and Hebrew-speaking children with reading difficulties after reading acceleration training. *Annals of Dyslexia*, *64*(3), 183–201. <https://doi.org/10.1007/s11881-014-0093-4>.
- Hu, L., & Bentler, P. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. <https://doi.org/10.1037/1082-989X.3.4.424>.
- Huemer, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency among poor readers of German: Many ways to the goal. *Annals of Dyslexia*, *58*(2), 115–137. <https://doi.org/10.1007/s11881-008-0017-2>.
- Ise, E., Engel, R. R., & Schulte-K  rne, G. (2012). Ergebnisse einer Metaanalyse zur Wirksamkeit deutschsprachiger F  rderans  tze. *Kindheit Und Entwicklung*, *21*(2), 122–136. <https://doi.org/10.1026/0942-5403/a000077>.
- Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology*, *106*(4), 1096–1106. <https://doi.org/10.1037/a0037100>.
- Karageorgos, P., Richter, T., Haffmans, M. B., Schindler, J., & Naumann, J. (2020). The role of word-recognition accuracy in the development of word-recognition speed and reading comprehension in primary school: A longitudinal examination. *Cognitive Development*, *56*, Article 100949. <https://doi.org/10.1016/j.cogdev.2020.100949>.
- Kim, Y. S., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from Grade 1 to Grade 2. *Journal of Experimental Child Psychology*, *113*(1), 93–111. <https://doi.org/10.1016/j.jecp.2012.03.002>.
- Kohn, J., Wyschkon, A., Ballaschk, K., Ihle, W., & Esser, G. (2013). Verlauf von Umschriebenen Entwicklungsst  rungen: Eine 30-Monats-Follow-up-Studie. *Lernen Und Lernst  rungen*, *2*(2), 77–89. <https://doi.org/10.1024/2235-0977/a000032>.
- Korinth, S. P., Dimigen, O., Sommer, W., & Breznitz, Z. (2016). Reading training by means of disappearing text: Effects on reading performance and eye movements. *Reading and Writing*, *29*(6), 1245–1268. <https://doi.org/10.1007/s11145-016-9635-y>.
- Korinth, S. P., & Nagler, T. (2021). Improving reading rates and comprehension? Benefits and limitations of the reading acceleration approach. *Language and Linguistics Compass*, *15*(3), Article e12408. <https://doi.org/10.1111/lnc3.12408>

- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6(2), 293–323. [https://doi.org/10.1016/0010-0285\(74\)90015-2](https://doi.org/10.1016/0010-0285(74)90015-2).
- Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, 63(3), 315–334. [https://doi.org/10.1016/s0010-0277\(97\)00005-x](https://doi.org/10.1016/s0010-0277(97)00005-x).
- Lee, K., & Chen, X. (2019). An emergent interaction between reading fluency and vocabulary in the prediction of reading comprehension among French immersion elementary students. *Reading and Writing*, 32(7), 1657–1679. <https://doi.org/10.1007/s11145-018-9920-z>.
- Lervåg, A., & Melby-Lervåg, M. (2022). Modeling the development of reading comprehension. In M. J. Snowling, C. Hulme, & K. Nation (Eds.), *The Science of Reading: A Handbook* (2nd ed., S. 280–297). Wiley Blackwell.
- Lindberg, S., Lonnemann, J., Linkersdörfer, J., Biermeyer, E., Mähler, C., Hasselhorn, M., & Lehmann, M. (2011). Early strategies of elementary school children’s single word reading. *Journal of Neurolinguistics*, 24(5), 556–570. <https://doi.org/10.1016/j.jneuroling.2011.02.003>.
- López-Escribano, C. (2016). Training reading fluency and comprehension of Spanish children with dyslexia. In A. Khateb & I. Bar-Kochva (Eds.), *Reading Fluency* (S. 141–161). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-30478-6_9.
- Nagler, T., Korinth, S. P., Linkersdörfer, J., Lonnemann, J., Rump, B., Hasselhorn, M., & Lindberg, S. (2015). Text-fading based training leads to transfer effects on children’s sentence reading fluency. *Frontiers in Psychology*, 6, 119. <https://doi.org/10.3389/fpsyg.2015.00119>.
- Nagler, T., Lindberg, S., & Hasselhorn, M. (2018). Leseentwicklung im Grundschulalter: Kognitive Grundlagen und Risikofaktoren. *Lernen Und Lernstörungen*, 7(1), 33–44. <https://doi.org/10.1024/2235-0977/a000185>.
- Nagler, T., Linkersdörfer, J., Lonnemann, J., Hasselhorn, M., & Lindberg, S. (2016). The impact of text fading on reading in children with reading difficulties. *Journal for Educational Research Online*, 8(1), 26–41. <https://doi.org/10.25656/01:12028>.
- Nagler, T., Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2014). The impact of reading material’s lexical accessibility on text fading effects in children’s reading performance. *Reading and Writing*, 27(5), 841–853. <https://doi.org/10.1007/s11145-013-9468-x>.
- Nagler, T., Zarić, J., Kachisi, F., Lindberg, S., & Ehm, J.-H. (2021). Reading-impaired children improve through text-fading training: Analyses of comprehension, orthographic knowledge, and RAN. *Annals of Dyslexia*, 71(3), 458–482. <https://doi.org/10.1007/s11881-021-00229-x>.
- Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology*, 97(3), 299–319. <https://doi.org/10.1037/0022-0663.97.3.299>.
- Perfetti, C. A. (1986). Continuities in reading acquisition, reading skill, and reading disability. *Remedial and Special Education*, 7(1), 11–21. <https://doi.org/10.1177/074193258600700105>.

- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of Functional Literacy* (pp. 189–213). John Benjamins Publishing. <https://doi.org/10.1075/swll.11.14per>.
- RosseeL, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Rost, D. H., & Sparfeldt, J. R. (2007). Leseverständnis ohne Lesen? Zur Konstruktvalidität von Multiple-choice-Leseverständnistestaufgaben. *Zeitschrift Für Pädagogische Psychologie*, 21(3–4), 305–314. <https://doi.org/10.1024/1010-0652.21.3.305>.
- Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher*, 32(4), 403–408. <http://www.jstor.org/stable/20194790>.
- Schneider, W. (2017). *Lesen und Schreiben lernen – Wie erobern Kinder die Schriftsprache?* Springer-Verlag. https://doi.org/10.1007/978-3-662-50319-5_8.
- Schulte-Körne, G. (2017). Lese- und/oder Rechtschreibstörung: Symptomatik. *Diagnostik und Behandlung. Monatsschrift Für Kinderheilkunde*, 165(6), 476–481. <https://doi.org/10.1007/s00112-017-0290-4>.
- Siegler, R. S. (2006). Microgenetic analyses of learning. In D. Kuhn & R. S. Siegler (Hrsg.), *Handbook of child psychology: Cognition, perception, and language* (6. Aufl., S. 464–510). Wiley.
- Skibbe, L., Grimm, K., & Morrison, F. (2012). Literacy growth in the academic year versus summer from preschool through second grade: Differential effects of schooling across four skills. *Scientific Studies of Reading*, 16(2), 141–165. <https://doi.org/10.1080/10888438.2010.543446>.
- Souvignier, E., Förster, N., Hebbecker, K., & Schütze, B. (2021). Using digital data to support teaching practice – quop: An effective web-based approach to monitor student learning progress in reading and mathematics in entire classrooms. In S. Jormitz & A. Wilmers (Eds.), *International perspectives on school settings, education policy and digital strategies: A transatlantic discourse in education research* (S. 283–298). Budrich. <https://doi.org/10.2307/j.ctv1gbrzf4.20>.
- Thomas, K., Schulte-Körne, G., & Hasselhorn, M. (2015). Stichwort – Entwicklungsstörungen schulischer Fertigkeiten. *Zeitschrift Für Erziehungswissenschaft*, 18(3), 431–451. <https://doi.org/10.1007/s11618-015-0642-6>.
- Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading*, 5(3), 211–239. <https://doi.org/10.1207/S1532799XSSR0503>.
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology*, 86(3), 169–193. [https://doi.org/10.1016/S0022-0965\(03\)00139-5](https://doi.org/10.1016/S0022-0965(03)00139-5).

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>) veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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 auch für die oben aufgeführten nicht-kommerziellen Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

