

Unterrichtsqualität beim jahrgangsgemischtem Lernen – Ein Blick in „Optimalklassen“

Meike Munser-Kiefer  · Sabine Martschinke  · Alfred Lindl  ·
Andreas Hartinger 

Angenommen: 4. Juni 2025 / Online publiziert: 14. Juli 2025
© The Author(s) 2025

Zusammenfassung Dieser Beitrag berichtet über eine quasi-experimentelle Längsschnittstudie zum jahrgangsgemischtem Lernen in den Schulbesuchsjahren 3 und 4 ($N=665$). Es wird untersucht, inwieweit sich ein Cluster von Schulklassen, in denen die Kinder sowohl gute Lernerfolge (im Lesen) als auch eine positive Entwicklung des lesespezifischen Selbstkonzepts aufweisen (im Folgenden als „Optimalklassen“ bezeichnet), identifizieren lässt und ob sich dieses Cluster in ausgewählten Merkmalen der Unterrichtsqualität (in der Wahrnehmung der Schüler:innen) von anderen Clustern unterscheidet. Unter Berücksichtigung der Klassenstruktur zeigen sich in hierarchischen Regressionsanalysen überzufällige Unterschiede zugunsten der Optimalklassen. Die Schüler:innen fühlen sich besser von den Peers unterstützt, nehmen sich als selbstbestimmter wahr und empfinden den Unterricht als klarer.

✉ Prof. Dr. Meike Munser-Kiefer

Lehrstuhl für Grundschulpädagogik und -didaktik (Diversität), Institut für Bildungswissenschaft, Universität Regensburg, Sedanstraße 1, 93055 Regensburg, Deutschland
E-Mail: meike.munser-kiefer@ur.de

Prof. Dr. Sabine Martschinke

Lehrstuhl für Grundschulpädagogik und -didaktik mit dem Schwerpunkt Umgang mit Heterogenität, Institut für Grundschulforschung, Universität Erlangen-Nürnberg, Regensburger Straße 160, 90478 Nürnberg, Deutschland
E-Mail: sabine.martschinke@fau.de

Dr. Alfred Lindl

Lehrstuhl für Educational Data Science, Institut für Bildungswissenschaft, Universität Regensburg, Sedanstraße 1, 93055 Regensburg, Deutschland
E-Mail: alfred.lindl@ur.de

Prof. Dr. Andreas Hartinger

Lehrstuhl für Grundschulpädagogik und -didaktik, Universität Augsburg, Universitätsstraße 10, 86159 Augsburg, Deutschland
E-Mail: andreas.hartinger@uni-a.de

Schlüsselwörter Jahrgangsgemischtes Lernen · Optimalklassen · Unterrichtsqualität · Lesen · Selbstkonzept

Teaching quality in multi-grade learning—Insights from “best practice classes”

Abstract This article reports on a quasi-experimental longitudinal study on multi-grade learning in grades 3 and 4 ($N=665$). It examines whether a cluster of school classes can be identified in which children demonstrate both good learning success (in reading) and positive self-concept development (referred as “best practice classes”), and whether this cluster differs from other clusters in selected characteristics of teaching quality (as perceived by the pupils). Considering the class structure, multilevel regression analyses show above-average differences in favor of the best practice classes. The students feel better supported by their peers, perceive themselves as more self-determined and perceive the lessons as clearer.

Keywords Multi-grade · Best practice classes · Teaching quality · Reading · Self-concept

1 Einleitung

Jahrgangsgemischter Unterricht hat inzwischen eine lange Tradition. Dabei werden verschiedene Klassenstufen gemeinsam unterrichtet – mit dem pädagogisch-didaktischen Ziel, das Bildungsgefälle zwischen den Jahrgängen produktiv zu nutzen und adaptivere Lernangebote bereitzustellen (Carle und Metzen 2017; Klöver 2014). In Regelschulen werden zumeist die 1. und 2. bzw. die 3. und 4. Jahrgangsstufe zusammengefasst; in Reformschulen oder bei einer sechsjährigen Grundschule finden sich auch die Kombinationen von 1.–3., 4.–6. oder 1.–4.

Insbesondere in der Schuleingangsstufe der 1. und 2. Klasse fußt die Jahrgangsmischung auf umfangreichen Pilotierungen mit wissenschaftlicher Begleitung (z. B. Arbeitskreis Wissenschaftliche Begleitung Schulanfang auf neuen Wegen 2006; BeS-Te 2005–2008, Carle und Metzen 2009; FiS, Eckerth et al. 2015; FLEX, Liebers et al. 2008), Unterrichtsentwicklung und Fortbildungen (Stiftung Bildungspakt Bayern & Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst 2015). Mehrfach wurden die Effekte des jahrgangsgemischten Lernens im Vergleich zu jahrgangshomogen unterrichteten Klassen auch jenseits dieser Modellversuche erhoben (zusammenfassend: Götz und Krenig 2024). Dabei wurden zumeist eine Pattsituation in der Leistung sowie tendenziell bessere Entwicklungen im Persönlichkeitsbereich festgestellt (repräsentativ im Querschnitt: Kuhl et al. 2013; im Längsschnitt: Munser-Kiefer et al. 2022). Für die Grundschule spricht dies für ein leistungsbezogen zumindest gleichwertiges Unterrichtskonzept mit Potenzial zu günstigeren Effekten.

Allerdings finden sich in den Studien deutliche Unterschiede in der Gestaltung des jahrgangsgemischten Unterrichts (z. B. Hyry-Beihammer und Hascher 2015;

Sonnleitner 2021) sowie in Abhängigkeit von Erfahrung und Überzeugung der jahrgangsgemischten unterrichtenden Lehrkräfte, die sich als handlungsleitend für die Gestaltung des Unterrichts erweisen (z. B. Thoren 2017; Munser-Kiefer et al. 2017). Darüber hinaus zeigen Analysen über die jeweilige jahrgangsgemischte Phase einerseits eine eher positive Lernentwicklung zu Beginn, die sich zum Ende wieder nivelliert (z. B. Grittner et al. 2013; Munser-Kiefer et al. 2022), andererseits eine gegenteilige Entwicklung des Selbstkonzepts der Kinder (Munser-Kiefer et al. 2023) sowie zudem eine große Varianz der Lernerfolge zwischen den jahrgangsgemischten Klassen (Thoren 2017; Thoren und Brunner 2019).

In der Weiterführung des Forschungsstrangs zum jahrgangsgemischten Unterricht wird in diesem Beitrag untersucht, ob sich Klassen finden, in denen es gelingt, dass sich sowohl Leistung als auch Selbstkonzept positiv entwickeln; im Anschluss daran wird geprüft, welche Qualitätsmerkmale der Gestaltung jahrgangsgemischten Unterrichts sich in diesen Klassen identifizieren lassen. Der Fokus liegt auf der Jahrgangsmischung in der 3. und 4. Klasse.

Im ersten Teil werden ausgehend von empirisch nachgewiesenen Oberflächenstrukturen des jahrgangsgemischten Unterrichts Merkmale der Unterrichtsqualität identifiziert, die diesen potenziell zugrunde liegen. Im empirischen Teil des Beitrags wird untersucht, ob sich sogenannte „Optimalklassen“ (mit günstiger Entwicklung in Leistung und Selbstkonzept) finden und ob in diesen Klassen wichtige Merkmale der Unterrichtsqualität besonders ausgeprägt sind – mit dem Ziel, einem potenziellen „Erfolgsgeheimnis“ jahrgangsgemischten Unterrichts auf die Spur zu kommen.

2 Unterricht in der Jahrgangsmischung

2.1 Oberflächenstrukturen im Unterricht der Jahrgangsmischung

Studien zur Jahrgangsmischung berichten selten über tiefenstrukturelle Qualitätsmerkmale auf der Prozessebene des Unterrichts, sondern beziehen sich meist auf typische Organisationsformen. Diese zeigen, dass Lehrkräfte im jahrgangsgemischten Unterricht oftmals folgende Gruppierungsformen einsetzen, die in Inhaltsauswahl und -aufbereitung variieren können (Hyry-Beihammer und Hascher 2015; Ronksley-Pavia et al. 2019; Schmidt 1999; Sonnleitner 2021): 1) Abteilungsunterricht, 2) individualisierter Unterricht und 3) Unterricht in jahrgangsgemischten Teams. Einer Studie von Schmidt (1999) zufolge verteilen sich diese zu 57% auf Abteilungsunterricht (ähnlich in Deutschland: Sonnleitner 2021; Finnland: Kalaoja und Pietarinen 2009; Irland: Mulryan-Kyne 2007; im alpinen Raum der Schweiz und Österreich: Raggl 2015; zusammenfassend: Carle und Metzen 2017), zu 10% auf individualisierende Unterrichtsformen und zu 33% auf Lernformen, welche die jahrgangsübergreifende Heterogenität der Klasse nutzen.

Abteilungsunterricht beschreibt einen Unterricht, bei dem die Lehrkraft häufig nur mit einem Teil der Klasse gruppiert nach Jahrgängen arbeitet. Hyry-Beihammer und Hascher (2015) identifizieren in ihrer Interviewstudie verschiedene Formen, die sich in der Anordnung der Inhalte unterscheiden (z. B. Parallelcurriculum, Themenstaffelung). Das Parallelcurriculum scheint dabei die dominante Form zu sein,

obwohl es Lehrkräfte als herausfordernd empfinden, eine Gruppe stillarbeiten zu lassen, während die andere unterrichtet wird, und der Zeitverlust durch Warten als Problem bezeichnet wird. Auch in der Interviewstudie von Sonnleitner (2021) zeigt sich Abteilungsunterricht häufig als frontale Unterrichtsform zur Einführung bzw. Differenzierung im Schriftspracherwerb und in Mathematik (ähnlich im Schweizer Schulversuch EDK-Ost bei Mathematik und sprachlichen Inhalten, Vogt 2009). Sonnleitner (2021) fand jedoch auch „modernen Abteilungsunterricht“ (S. 284), bei dem die Schüler:innen getrennt nach Jahrgängen das gleiche Thema mit jahrgangsspezifisch differenziertem Material auf verschiedenen Niveaus bearbeiteten. Ebenso berichten Hyry-Beihammer und Hascher (2015) davon, dass Lehrkräfte die Lehrplaninhalte so ordnen, dass Schüler:innen im ersten Besuchsjahr auf einem Basisniveau arbeiten und die Lerninhalte im zweiten Besuchsjahr vertiefen.

Individualisiertes Lernen im jahrgangsgemischten Unterricht beschreibt ein an die individuellen Lernvoraussetzungen bzw. Interessen angepasstes Angebot (z. B. Hess und Lipowsky 2017). Dies kann einzelne Aufgaben oder Aufgabenreihen (z. B. als Lernleiter) umfassen oder ein bausteinartiges Angebot sein, das von Schüler:innen in Einzelarbeit gemäß eigenem Lernfortschritt und/oder Tempo bearbeitet wird. Umgesetzt wird dies z. B. mit individuellen Arbeitsplänen (Hyry-Beihammer und Hascher 2015), die z. T. recht kleinschrittig aufgebaut sind (Rathgeb-Schnierer und Rechtsteiner-Merz 2010). Potenzial für eine lernwegorientiertere Didaktik fanden Kucharz und Wagener (2007) für die Leseförderung. Auch die Evaluation des Schulversuchs Flexible Grundschule in Bayern (Klöver 2014) zeigte, dass Lehrkräfte neben Formen geöffneten Unterrichts Portfolios, Lerntagebücher und Lernlandkarten einsetzen. Insgesamt weisen Befunde (außerhalb von Modellversuchen) jedoch darauf hin, dass individualisierter Unterricht in den verschiedenen Praxen von Lehrkräften als sehr herausfordernd empfunden wird (Sonnleitner 2017). Auch in den Ergebnissen zur verpflichtenden Einführung des Jahrgangsübergreifenden Lernens in Berlin befindet sich nach Einschätzung der Schulinspektion innere Differenzierung auf einem eher niedrigen Niveau (Thoren 2017; Thoren und Brunner 2019).

Eine nur im jahrgangsgemischten Setting mögliche Form ist das *Arbeiten in jahrgangsgemischten Teams*. Hier lernen Kinder verschiedener Jahrgangsstufen zu zweit oder in jahrgangsübergreifenden Gruppen zusammen. Dabei handelt es sich entweder um Ad-hoc-Gruppierungen für eine spezielle Aufgabenstellung (spontan durch die Kinder bzw. angeleitet durch die Lehrkraft; Hyry-Beihammer und Hascher 2015) oder um festinstallierte Helfersystems (Wagener 2014). Die Aufgabenstellung kann asymmetrisch unterstützend (z. B. durch Lernen durch Lehren oder Modelllernen von Lernstrategien und Lernverhalten; Feuchtenberger et al. 2019) oder auch zusammenwirkend sein (z. B. durch die Arbeit auf verschiedenen Niveaus bzw. an verschiedenen Aspekten eines gemeinsamen Lerngegenstandes; Feyerer 2012). Gerade die helfende Form scheint in jahrgangsgemischten Klassen prägend zu sein. So fand Wagener (2014) in ihrer Beobachtungsstudie und den begleitenden Kinderinterviews zu jahrgangsübergreifenden Lerngruppen in der Grundschule (JüLiG 3/4) ein intuitiv-verinnerlichtes Rollenverständnis gemäß dem Altersgefälle (älteres Kind als Tutor:in, jüngeres Kind als Tutee). Allerdings scheinen dazu Anleitung bzw. Aufgabenstrukturen erforderlich zu sein, da die Kinder einander eher innerhalb der eigenen Jahrgangsstufe helfen und nur in freien Arbeitsphasen über die

Jahrgangsstufen hinweg kooperieren (Hyry-Beihammer und Hascher 2015). Auch in der Implementation des Jahrgangsübergreifenden Lernens in Berlin fand sich hier noch viel Entwicklungspotenzial (Thoren 2017; Thoren und Brunner 2019).

2.2 Theorie und Forschung zu Qualitätsmerkmalen jahrgangsgemischter Organisationsstrukturen

Zur Wirksamkeit dieser Gruppierungsformen liegen kaum systematische Befunde vor – auch, weil sie häufig in Mischformen umgesetzt werden (Adelmeier und Liebers 2008; Schmidt 1999; Sonnleitner 2021). Ohnehin ist zu erwarten, dass der qualitativen Ausgestaltung des Unterrichts eine größere Bedeutung zukommt als der Gruppierungsform (zusammenfassend: Kleickmann 2024). Die gut erforschten und belegten Basisdimensionen von Unterrichtsqualität (Praetorius et al. 2018) – Klassenführung, kognitive Aktivierung und konstruktive Unterstützung – bieten hierbei einen Referenzrahmen, innerhalb dessen für die Jahrgangsmischung bedeutsame Qualitätskriterien identifiziert werden können. Aufgrund möglicher Abhängigkeiten zwischen Oberflächen- und Tiefenmerkmalen von Unterricht (Hess und Lipowsky 2020), können diese auch in den typischen Organisationsformen angenommen werden:

- Im Abteilungsunterricht ergibt sich das Potenzial aus der kleineren (und nicht jahrgangsgemischten) Lerngruppe: Hier können z. B. durch *konstruktive Unterstützung* durch die Lehrkraft einzelne Schüler:innen stärker eingebunden und aktiviert werden; zudem kann durch die übersichtlichere Lerngruppe der Überblick über individuelle (Lern-)Entwicklungen und dadurch das Anwenden der *individuellen Bezugsnorm* (bei gleichzeitiger Einschränkung der *sozialen Bezugsnorm*) erleichtert werden.
- Auch beim individualisierten Lernen liegt das Potenzial in der *konstruktiven Unterstützung* der einzelnen Schüler:innen, wenn zum Leistungsstand passende Aufgaben, Material und Instruktionenformen (Intended Adaptive Teaching) angeboten werden und Unterstützungsmaßnahmen wie Aufforderungen, Hinweise, Erklärungen und Feedback (Implemented Adaptive Teaching) zu einer besseren Nutzung führen (Hardy et al. 2019). Die Passung des Lernangebots kann auch durch Wahlfreiheiten entstehen, wenn Schüler:innen *Autonomie* eingeräumt wird und sie im Unterricht mitbestimmen. Eine Grundvoraussetzung für die Qualität des Angebots und das Potenzial, dieses zu nutzen, ist darüber hinaus die *Klarheit* in Ziel und Aufgabenstellung.
- Gleiches gilt für die jahrgangsgemischte Lerngruppe, deren Potenzial sich zusätzlich durch die *sozio-emotionale Unterstützung der Lernpartner:innen* entfaltet. Zusätzlich kann die Unterstützung durch die Kinder untereinander weitere Unterstützungskapazitäten der Lehrkraft freisetzen.

Die identifizierten Qualitätsmerkmale lassen sich im Kern dem Basiskriterium der *konstruktiven Unterstützung* zuordnen (Bezugsnormen, Unterstützung durch Lehrkraft und Peers sowie Autonomieempfinden), stets unter der Voraussetzung, dass die Lehrperson insbesondere dafür Sorge trägt, dass individuelle oder kooperative Lernprozesse der Kinder(gruppen) gelingen sowie die Aufgabenstellungen (u. a. durch

die Klarheit des Unterrichts) die nötigen Voraussetzungen für kognitive Aktivierung schaffen.

Bezugsnormen Die Verwendung einer Bezugsnorm ist eine zentrale Voraussetzung zum Messen einer Leistung und ihrer Bewertung (Ertl und Zumhasch 2024), denn sie setzt den Referenzrahmen für die Einschätzung der Leistung. Bei der *individuellen Bezugsnorm* ist die Referenz im Lernprozess und -produkt ebenso das individuelle Kind wie sein Lernzuwachs. Der Gebrauch der individuellen Bezugsnorm ist im Grundschulkontext ein gutes Mittel zur Steigerung von Motivation und Selbstkonzept. Sie erwies sich in Best-Practice-Analysen als wichtiges Kriterium für multikriteriale Zielerreichung (Martschinke und Kammermeyer 2003). Bei der *sozialen Bezugsnorm* bildet die Referenz die soziale Gruppe. Die soziale Bezugsnorm wirkt sich eher negativ auf das Selbstkonzept aus und es finden sich Hinweise, dass sich dies in ungünstigen Effekten auf die Leistungsentwicklung manifestiert (zusammenfassend: Worbach et al. 2019). Gleichzeitig zeigt z. B. die KILIA-Studie (Martschinke und Kammermeyer 2003), dass es durch einen Fokus auf individueller statt sozialer Bezugsnorm zu einer Entkopplung des Zusammenhangs von Selbstkonzept und Leistung kommt und der überproportionale Abwärtstrend im Selbstkonzept vermieden werden kann.

In der Jahrgangsmischung verändert sich der Bezugsrahmen für die Bezugsnormen, da sich dieser entweder an den Schüler:innen des gleichen Jahrgangs bzw. an Kindern aller Jahrgänge orientieren kann. Munser-Kiefer und Kolleg:innen (2023) fanden Hinweise darauf, dass sich die Schüler:innen eher an der gesamten Klasse (und nicht wie angenommen am eigenen Jahrgang) orientieren. Dies spiegelt sich in einem signifikant geringeren Selbstkonzept der Kinder aus dem „unteren Jahrgang“ bei gleicher Leistungsfähigkeit im Vergleich zu jahrgangshomogen unterrichteten Schüler:innen wider. Es erscheint jedoch möglich, dass dies selbstwertdienlich weniger als Bedrohung wahrgenommen wird und eher zu einer realistischen Einschätzung der eigenen Fähigkeiten führt. Denn im zweiten Besuchsjahr der Jahrgangsmischung gleichen sich diese Effekte aus. Die insgesamt stärkere Erschütterung des Selbstkonzepts in der Jahrgangsmischung deutet jedoch darauf hin, dass ein verstärkter Fokus auf die individuelle Bezugsnorm bzw. eine Reduktion der sozialen Bezugsnorm günstig auf die Entwicklung von Selbstkonzept und Leistung wirken könnte.

Unterstützung Als eine von drei Basisdimensionen von Unterricht (Kleickmann 2024) kann konstruktive Unterstützung einerseits durch die Lehrkraft und andererseits durch die Schüler:innen untereinander erfolgen. Sie kann sich dabei auf kognitive, sozio-emotionale, motivationale oder regulative Aspekte beziehen (Praetorius und Charalambous 2018).

In allen Formen jahrgangsgemischten Unterrichts ergibt sich (z. B. durch die kleinere Gruppe im Abteilungsunterricht oder durch Peer-Tutoring im jahrgangsgemischtem Team) im Vergleich zu Jahrgangsklassen mehr Potenzial, aber auch Notwendigkeit zur Unterstützung. Hinsichtlich der direkten Lehrkraft-Schüler:innen-Interaktion fanden Thoren und Brunner (2019) hohe Werte für „systematische Unterstützung, Förderung und Beratung“ (S. 292) für alle beteiligten jahrgangsgemischten

unterrichtenden Lehrkräfte. Dies deckt sich mit der Perspektive der Schüler:innen, bei denen sich die meisten im ersten und zweiten Besuchsjahr gleichwertig unterstützt fühlten (Martschinke et al. 2022). In der Gestaltung des Unterrichts fanden Munser-Kiefer und Kolleg:innen (2017) im Vergleich jahrgangshomogen und jahrgangsgemischt unterrichteter Klassen signifikant mehr Elemente adaptiver Unterrichtsgestaltung (operationalisiert über Diagnose, Differenzierung, individuelle Förderung und Rückmeldungen), wobei sich hier noch Potenzial zur Unterrichtsentwicklung offenbarte. Sonnleitner (2021) zeigte, dass Lehrkräfte in jahrgangsgemischten Klassen Schwierigkeiten haben, das gesamte Leistungsspektrum und damit alle Schüler:innen im Blick zu haben. Darauf verweisen auch die (älteren) Ergebnisse von Gutiérrez und Slavin (1992).

Konstruktive Unterstützung kann im jahrgangsgemischten Setting ergänzend durch die (älteren) Schüler:innen erfolgen und positive Effekte des Arbeitens im jahrgangsgemischten Team ermöglichen. Campana Schleusener (2014) konnte in den Basisstufenklassen (4- bis 8-Jährige) gezielte Hilfen der jeweils älteren Kinder beobachten. Weitere Studien belegen einen Trend zu mehr (Matz und Knauf 2010) und angemessenerer (Malti und Perren 2008) Unterstützung mit steigendem Alter. Voraussetzung dafür ist, dass die Schüler:innen darauf vorbereitet und entsprechend angeleitet werden (Smit et al. 2015); zudem müssen geeignete Rahmenbedingungen geschaffen werden (z. B. durch passende Aufgabenstellungen).

Autonomie Das Autonomieempfinden bezeichnet das subjektive Erleben, selbstbestimmt und eigenständig handeln zu können. Es gilt als wichtige Grundlage für die Entwicklung von lernförderlicher Motivation und Interesse (Ryan und Deci 2020).

In jahrgangsgemischten Klassen ist sowohl beim individualisierten Lernen als auch beim Arbeiten in jahrgangsübergreifenden Lerngruppen zu erwarten, dass die Schüler:innen gewisse Freiheiten und Wahlmöglichkeiten haben. Dies zeigt sich zumindest in Ansätzen auch bei Munser-Kiefer und Kolleg:innen (2017). Hier gaben die Lehrkräfte im jahrgangsgemischten Unterricht an, mindestens einmal pro Woche selbstständiges Arbeiten durch entsprechende Unterrichtsformen (wie Forscher:innenfragen oder inhaltlich offene Arbeitsaufträge) zu ermöglichen – und damit häufiger als Kolleg:innen aus jahrgangshomogenen Klassen. Parallel dazu empfanden sich die Schüler:innen der jahrgangsgemischten Klassen als selbstbestimmter (Munser-Kiefer et al. 2016). Die Befunde zur Implementation des Jahrgangsübergreifenden Lernens in Berlin ergaben dagegen, dass Schüler:innen im jahrgangsgemischten Unterricht wenig Gelegenheit zum selbstständigen Lernen finden. Hier gibt es jedoch signifikante Unterschiede zugunsten von „Pionieren“ im Vergleich zu „allgemein Nichtüberzeugten“ (Thoren und Brunner 2019, S. 81).

Klarheit Klarheit im Unterricht bezieht sich auf die Strukturierung von Inhalten sowie auf die Klarheit von Lernzielen und Arbeitsanweisungen. Damit hat dieses Qualitätsmerkmal zum einen Einfluss auf die effektive Nutzung von Lernzeit und zum anderen auf die kognitive Aktivierung der Lerner:innen (Praetorius et al. 2020).

In jahrgangsgemischten Klassen ist die inhaltliche und strukturelle Klarheit des Unterrichts eine besondere Herausforderung, da durch die unterschiedlichen und zieldifferenten Gruppen immer wieder komplexere und differenziertere Aufgaben-

stellungen und Organisationsformen erforderlich sind (Götz und Krenig 2024). Zum individuellen Lernen in jahrgangsgemischten Klassen konkretisierten Breidenstein und Rademacher (2013) in ihrer ethnographischen Beobachtungsstudie zum individualisierten (auch jahrgangsgemischten) Unterricht drei Dimensionen des Umgangs mit der Zeit, die indirekt für Klarheit in Zeitnutzung und Aufgabe stehen: „Begrenzung und Markierung der Unterrichtszeit“, „Kennzeichnung [...] legitimer Tätigkeiten“ und „interne Regulierung des Verhältnisses zwischen Zeit und Arbeitspensum“ (S. 341). Im gelungenen Zusammenspiel der drei Dimensionen konnte durch Individualisierung eine Steigerung der effektiv genutzten Lernzeit beobachtet werden, weil z.B. Wartezeiten entfielen (Breidenstein und Rademacher 2013). In der Implementation des Jahrgangsübergreifenden Lernens in Berlin fanden sich sowohl in der „Lehr- und Lernzeit“ als auch in „Strukturierung und transparenter Zielausrichtung“ (Thoren und Brunner 2019, S. 288) in allen jahrgangsgemischten Klassen hohe Werte.

3 Konkretisierung der Forschungsfragen

Zusammenfassend können Bezugsnormen, Unterstützung durch Lehrkraft und Peers, Autonomieempfinden und die Klarheit des Lern- und Unterstützungsangebots geeignete Stellschrauben sein, um das besondere Potenzial der Jahrgangsmischung (als Oberflächenmerkmal) zu nutzen, aber auch um die spezifischen Herausforderungen zu meistern. Es besteht jedoch Forschungsbedarf, ob diese Qualitätsmerkmale tatsächlich einen positiven Beitrag zur Leistungs- und Persönlichkeitsentwicklung der Schüler:innen in jahrgangsgemischten Klassen leisten.

Die vorliegende Studie greift dieses Desiderat auf und berichtet Ergebnisse zu folgenden Fragen:

1. Lassen sich unter den jahrgangsgemischten Klassen sogenannte „Optimalklassen“ finden, in denen sich sowohl das Lernen als auch das Selbstkonzept der Kinder (im Lesen) in besonderer Weise positiv entwickeln?
2. Unterscheiden sich diese „Optimalklassen“ – im Vergleich zu den anderen Klassen – bezüglich bestimmter Merkmale von Unterrichtsqualität?

4 Methodik

4.1 Untersuchungsdesign

Die Daten entstammen einer quasi-experimentellen Längsschnittstudie zu jahrgangsgemischten Klassen, in denen das 3. und 4. Schuljahr gemeinsam unterrichtet wurde. Zur Beantwortung unserer Fragen wurden Daten auf Schüler:innenebene im Verlauf der 3. Jahrgangsstufe erhoben. Erfasst wurden Leistungstestdaten aus dem Bereich Lesen, persönlichkeitsbezogene Fragebogendaten zum fachspezifischen Selbstkonzept im Lesen – jeweils zu Beginn (Oktober; MZP1) und am Ende (Juli; MZP3) des Schuljahres. Die Einschätzung zentraler Facetten der Unterrichtsqualität durch

Tab. 1 Stichprobenkennwerte der Lehrkräfte bzw. Schulklassen ($N = 66$)

Geschlecht (m/w/d)	Dienstalter		Klassengröße	
	$M (SD)$		$M (SD)$	
	Insgesamt	Davon in JM 3/4	Insgesamt	In der 3. Jgst.
7/59/0	17,48 (10,07)	2,67 (1,67)	20,35 (2,89)	10,68 (2,63)

m männlich, *w* weiblich, *d* divers; *M* arithmetischer Mittelwert, *SD* Standardabweichung, *JM* Jahrgangsmischung

die Schüler:innen geschah durch Fragebögen in der Mitte des Schuljahrs (März; MZP2). Als Hintergrundvariablen dienten Geschlecht, Alter, Familiensprache sowie die Anzahl der Bücher im Haushalt als Indikator für Bildungsnähe. Zusätzlich liegen Daten auf Lehrkräfteebene zu Geschlecht und Dienstalter bzw. zur Klassengröße und Anzahl der Drittklässler:innen vor.

4.2 Stichprobe

In die Analyse gingen die Daten von 655 jahrgangsgemischt unterrichteten Schüler:innen aus 66 Klassen an 29 Schulen ein. Rekrutiert wurde die Stichprobe in den Städten Augsburg und Nürnberg sowie deren Umland. Es gelang, 90% der jahrgangsgemischt arbeitenden Klassen dieser Regionen für die Teilnahme an der Studie zu gewinnen. Die Absagegründe ließen keine Systematik erkennen, sodass von einer zufälligen Verteilung auszugehen ist und die Stichprobe als repräsentativ für die Gebiete gelten kann. Tab. 1 zeigt Kennwerte der untersuchten Teilstichprobe auf Lehrkraft- und Klassenebene, Tab. 2 auf Schüler:innenebene.

4.3 Erhebungsinstrumente

Lesen Zur Erfassung der Leseleistung wurde VERA 2006 als lehrplanvalider Test zu Beginn der 3. Jahrgangsstufe eingesetzt. Der kontinuierliche Sachtext erfasst im geschlossenen und offenen Antwortformat mit 13 Items ($\alpha_{MZP1} = 0,73$) hierarchie-niedrige und -höhere Verstehensprozesse (Lorenz et al. 2022). Die Leseleistung zum Ende der 3. Jahrgangsstufe wurde mit den bayernweit durchgeführten Vergleichsarbeiten (VERA Bayern) gemessen.

Tab. 2 Stichprobenkennwerte der Schüler:innen ($N = 665$)

Geschlecht (m/w/d)	Alter $M (SD)$	Bildungsnähe Anzahl/Prozent	Familiensprache Anzahl/Prozent	
			Deutsch	Nichtdeutsch
329/326/0	8,57 (0,58)	Bildungsnah: 349/66 Eher bildungsnah: 146/22 Eher bildungsfern: 53/8 Bildungsfern: 20/3	582/89	73/11

m männlich, *w* weiblich, *d* divers; *M* arithmetischer Mittelwert, *SD* Standardabweichung, *JM* Jahrgangsmischung

Familiensprache erfasst über „Welche Sprache sprichst du meistens zu Hause?“; Bildungsnähe über Bücher als Indikator (bildungsnah: mehr als 200 Bücher; drei Regale; eher bildungsnah: ein Regal; eher bildungsfern: ein Regalbrett, bildungsfern: keine Bücher)

Tab. 3 Merkmale von Unterrichtsqualität aus Schüler:innenperspektive mit Operationalisierung, Beispiel und Kennwerten

Merkmal	Operationalisierung	Beispiel	Anz.	M (SD)	Cronbachs α
Individuelle Bezugsnorm	Intrapersonaler Leistungsvergleich mit Fokus auf die Dynamik der Leistungsentwicklung (Indikatoren: Wahrnehmung von Freude/Enttäuschung, Lob)	Wenn ein Schüler sich verbessert, lobt ihn meine Lehrerin sehr.	4	2,60 (0,41)	0,66
Soziale Bezugsnorm	Interpersoneller Leistungsvergleich mit Fokus auf den Leistungsstand (Indikatoren: Öffentlichkeit von Leistungen/Leistungsvergleiche, Reaktion auf beste Leistungen)	Meine Lehrerin vergleicht oft unsere Leistungen miteinander.	4	1,57 (0,70)	0,65
Unterstützung durch Lehrkraft	Unterstützung durch Hilfestellungen (Indikatoren: Hilfe/Unterstützung erkennen bzw. geben)	Unsere Lehrerin kümmert sich um jeden, der Hilfe braucht.	5	2,52 (0,45)	0,79
Unterstützung durch Peers	Individuelle Wahrnehmung der Beziehung der gesamten Klasse (Indikatoren: einander mögen/helfen/freundlicher Umgang, Spaß miteinander)	Wir helfen uns in der Klasse gegenseitig, wenn jemand etwas nicht verstanden hat.	6	2,26 (0,47)	0,74
Autonomie	Entscheidungsfreiheit im Unterricht (Indikatoren: mit-/bestimmen, Auswahlmöglichkeiten)	Bei uns bestimmen die Schüler im Unterricht mit.	5	1,69 (0,58)	0,79
Klarheit	Strukturierung des Unterrichts auf der Makro- (= Struktur der Unterrichtsstunde) und der Mikroebene (= Lernprozess) (Indikatoren: Verständnis von Ziel und Aufgaben; Klarheit von Wichtigem/weniger Wichtigem)	Ich weiß im Unterricht immer, was ich gerade zu tun habe.	5	2,25 (0,46)	0,74

M arithmetischer Mittelwert, SD Standardabweichung

Skalierung: 0 = stimmt überhaupt nicht; 1 = stimmt eher nicht; 2 = stimmt eher; 3 = stimmt genau

Selbstkonzept Das Selbstkonzept wird hier gefasst als kognitive Repräsentation der Einschätzungen und Bewertungen einer Person über ihre Fähigkeiten (Moschner und Dickhäuser 2018), die in Anlehnung an Shavelson et al. (1976) fachspezifisch konzeptualisiert ist. Die Erfassung des Selbstkonzepts erfolgte über die Subskala „Lesen“ des Fragebogens der KILIA-Studie (Martschinke und Kammermeyer 2003) über 5 Items ($\alpha_{MZP1}=0,82$; $\alpha_{MZP3}=0,86$, „Musst du dich im Lesen viel anstrengen, um gut zu sein?“) mit einer vierstufigen Ratingskala (0= gar nicht bis 3= ganz viel).

Unterrichtsqualität Die von den Schüler:innen wahrgenommene Unterrichtsqualität wurde über einen Fragebogen mit einer vierstufigen Skala mit den Subskalen individuelle und soziale Bezugsnorm, Unterstützung durch Lehrkraft und Peers, Autonomie und Klarheit erhoben. Tab. 3 gibt einen Überblick über Operationalisierung, Beispielitems und Kennwerte.

4.4 Datenaufbereitung und -auswertung

Fehlende Werte wurden für jeden Erhebungszeitpunkt mithilfe des Expectation-Maximization-Algorithmus geschätzt (Enders 2010) – zunächst für jeden Messzeitpunkt einzeln, dann über die verschiedenen Messzeitpunkte hinweg, wenn hierfür eine Mindestdatenmenge von 70 % aller Variablen verfügbar war.

Zur Beantwortung von Forschungsfrage 1 wird überprüft, ob unterscheidbare Cluster auf Basis lernerfolgsrelevanter Schüler:innenvariablen identifiziert werden können. In die Clusteranalyse gehen vier jeweils *z*-standardisierte Variablen ein: Lernentwicklung im Lesen und Entwicklung des Selbstkonzepts Lesen (jeweils Differenz von Anfang und Ende der 3. Jgst.) sowie Lernniveau im Lesen und Niveau des Selbstkonzepts Lesen (Ende der 3. Jgst.). Damit werden im Sinne eines multi-kriterialen Ansatzes sowohl das kognitive Lernen als auch persönlichkeitsbezogene Variablen berücksichtigt.

Die Bestimmung der Cluster erfolgte mittels hierarchischer Analysen mit dem Ähnlichkeitsmaß der quadrierten euklidischen Distanz. Mithilfe eines Dendrogrammes mit dem Single-Linkage-Verfahren wurde ein Ausreißer identifiziert und eliminiert. Dies ermöglicht die Nutzung des Ward-Verfahrens, das nur dann zu sinnvollen, „wahren“ Gruppenbildungen führt, wenn solche Ausreißer zuvor entfernt werden (Backhaus et al. 2016). Als Entscheidungshilfe für die Clusteranzahl wurde auf Basis einer clusteranalytischen Berechnung mit der Ward-Methode ein Scree-Plot angelegt und der „Ellbogen“ (Elbow) in der Entwicklung des Heterogenitätsmaße bestimmt (Backhaus et al. 2016). Da diese übliche visuelle Analyse keine klare Entscheidung für eine 3- oder 4-Clusterlösung zuließ, erfolgte eine Bewertung über die standardisierten Fusionskoeffizienten mittels des Mojena-Tests (Mojena 1977). Für die 4-Clusterlösung wird der Wert 2,46 erreicht, die 3-Clusterlösung überschreitet mit einem Wert von 3,31 den von Mojena (1977) empfohlenen Schwellenwert von 2,75 und kann somit als beste Lösung gelten.

Um die Cluster hinsichtlich der angeführten clusterbildenden Variablen zu beschreiben und die Stabilität der Clusterbildung zu überprüfen, wurden zu diesen Kriterien unter Berücksichtigung der Klassenstruktur hierarchische lineare Regressionsmodelle mit der Clusterzugehörigkeit als dummy-codierter Prädiktorvariable

geschätzt (Hilbert et al. 2019; Lindl et al. 2020). Das erste Cluster fungierte dabei als Referenzkategorie. Gemäß Forschungsfrage 1 sollte dadurch überprüft werden, ob eines der Cluster als Optimalcluster beschreibbar ist, da die Schüler:innen im Lernen wie auch Selbstkonzept eine positive Entwicklung zeigen und am Ende der dritten Jahrgangsstufe ein hohes Niveau aufweisen. Im Anschluss daran wurde die Ähnlichkeit der Cluster sowohl mit Blick auf Hintergrundvariablen auf Lehrkraft- und Schüler:innenebene (wie z. B. Geschlecht, Familiensprache) als auch auf Unterschiede in den clusterbildenden Variablen zu Beginn der dritten Jahrgangsstufe untersucht, um mögliche Zusammenhänge jenseits der in diesem Beitrag interessierenden Unterrichtsmerkmale ausschließen zu können.

Inwiefern sich die Unterschiede zwischen den Clustern auch in verschiedenen Merkmalen der Unterrichtsgestaltung widerspiegeln (Forschungsfrage 2), wird ebenfalls mittels hierarchischer linearer Regressionsmodelle analysiert, welche die Klassenstruktur in den Daten berücksichtigen. In diese gehen die fokussierten Unterrichtsqualitätsmerkmale (Tab. 3) jeweils als abhängige Variablen, die Clusterzugehörigkeit als dummy-kodierte unabhängige Variable ein. Um direkte Vergleiche mit dem Optimalcluster zu erhalten, dient dieses als Referenzkategorie. Die Auswertungen erfolgen mittels der Statistiksoftware R (R Core Team 2024) und entsprechenden Zusatzpaketen wie lme4 (Bates et al. 2014), lmerTest (Kuznetsova et al. 2017), multilevel (Bliese 2022), MuMIn (Bartoń 2023) oder psych (Revelle 2024).

5 Ergebnisse

5.1 Clusterstruktur und Optimalklassen

Die durchgeführte Clusteranalyse resultiert in drei substantiell unterscheidbaren und inhaltlich interpretierbaren Clustern, die 19, 35 bzw. 12 Schulklassen umfassen und deren Größe auch für vergleichende Analysen genügt. Einen deskriptiven Überblick über zentrale Kennwerte und Unterschiede der drei Gruppen hinsichtlich der clusterbildenden Variablen Leistungs- und Selbstkonzeptentwicklung im Lesen während der 3. Jahrgangsstufe bzw. Leistungs- und Selbstkonzeptniveau im Lesen an deren Ende gibt Tab. 4.

Die Mittelwerte in den clusterbildenden Variablen und die Effektgrößen bei paarweisen Vergleichen der drei Gruppen ($d \geq |0,53|$) lassen deutliche Unterschiede zwischen den Clustern erkennen (Tab. 4), die in dummy-kodierten hierarchischen Regressionsanalysen unter Kontrolle der Klassenstruktur inferenzstatistisch abgesichert werden können.

In den Hintergrundvariablen (Tab. 1 und 2) finden sich dagegen keine signifikanten Abweichungen zwischen den drei Clustern. Es gibt keine Differenzen in der Zusammensetzung der Klassen und der Dienstserfahrung (allgemein oder im Unterrichten in jahrgangsgemischten Klassen) der Lehrpersonen. Dies gilt mit einer Ausnahme ebenso für die Schüler:innenvariablen wie Alter, Geschlechterverteilung, Familiensprache oder die Bildungsnähe (gemessen über das Indiz der „Anzahl der Bücher zu Hause“). Ein einziger signifikanter Unterschied liegt zwischen den Clustern 1 und 3 vor: Die Lesefähigkeiten der Schüler:innen in Cluster 1 fallen zu Beginn

Tab. 4 Arithmetische Mittelwerte (*M*), Standardabweichungen (*SD*) und Unterschiedseffekte (*d*) der clusterbildenden Variablen zwischen den Clustern

Clusterbildende Variable	Cluster 1 <i>N</i> = 19 <i>n</i> = 190 <i>M</i> (<i>SD</i>)	Cluster 2 <i>N</i> = 35 <i>n</i> = 345 <i>M</i> (<i>SD</i>)	Cluster 3 <i>N</i> = 12 <i>n</i> = 120 <i>M</i> (<i>SD</i>)	<i>d</i> _{Cl.1-Cl.3}	<i>d</i> _{Cl.1-Cl.3}
Lesen Entwicklung	0,39 (0,27)	0,15 (0,25)	-0,30 (0,28)	1,77	2,56
Lesen Niveau	0,50 (0,20)	0,12 (0,26)	-0,26 (0,33)	1,36	2,95
Selbstkonzept Entwicklung	0,25 (0,22)	-0,27 (0,30)	0,13 (0,16)	-1,48	0,63
Selbstkonzept Niveau	0,19 (0,14)	-0,18 (0,29)	-0,03 (0,17)	-0,53	1,47

N Klassenanzahl, *n* Schüler:innenanzahl, *z* standardisierte Werte, *d* Effektstärke nach Cohen (1992)

Tab. 5 Hierarchische lineare Regressionsmodelle für sechs Merkmale von Unterrichtsqualität unter Berücksichtigung der Klassenstruktur ($N_{\text{Klassen}} = 66$, $n_{\text{Schüler:innen}} = 655$)

Unterrichtsmerkmal (ICC)	Individuelle Bezugsnorm (11,7%)				Soziale Bezugsnorm (36,0%)					
	β	SE	df	t	p	β	SE	df	t	p
Feste Effekte										
Konstante	0,03	0,10	54,54	0,32	0,75	-0,19	0,15	62,52	-1,27	0,21
Cluster 1 vs. 2	0,07	0,12	54,95	0,52	0,60	0,26	0,19	62,52	1,38	0,17
Cluster 1 vs. 3	-0,33	0,16	54,75	-2,05	0,04	0,34	0,24	62,12	1,42	0,16
Marg. R^2 kond. R^2	0,02			0,12			0,02			0,37
Unterstützung durch Lehrkraft (18,8%)										
Konstante	-0,03	0,12	54,65	-0,26	0,79	0,30	0,13	63,84	2,24	0,03
Cluster 1 vs. 2	0,08	0,15	54,80	0,55	0,58	-0,30	0,17	63,89	-1,82	0,07
Cluster 1 vs. 3	-0,02	0,20	54,40	-0,11	0,92	-0,71	0,21	63,44	-3,32	<0,01
Marg. R^2 kond. R^2	0,00			0,20			0,06			0,32
Autonomie (23,9%)										
Konstante	0,27	0,13	61,16	2,19	0,03	0,27	0,11	60,04	2,38	0,02
Cluster 1 vs. 2	-0,31	0,16	61,28	-1,98	0,05	-0,33	0,14	60,28	-2,30	0,02
Cluster 1 vs. 3	-0,50	0,20	60,83	-2,47	0,02	-0,45	0,18	59,88	-2,46	0,02
Marg. R^2 kond. R^2	0,03			0,24			0,03			0,19

ICC Intraklassenkorrelation, β (standardisierter) Regressionskoeffizient, SE Standardfehler, t t-Wert, p Wahrscheinlichkeit für einen Fehler 1. Art, (marginales/konditionales) R^2 Anteil aufgekklärter Varianz (feste/feste und zufällige Effekte)

der dritten Jahrgangsstufe etwas höher aus. Damit lassen sich die drei Cluster folgendermaßen beschreiben:

- In **Cluster 1** (19 Klassen, 190 Schüler:innen) fand eine positive Entwicklung sowohl im Lesen als auch im lesebezogenen Selbstkonzept statt, jeweils auf hohem Niveau. Aufgrund der durchgängig positiven Werte wird Cluster 1 als „*Optimalcluster*“ bezeichnet.
- In **Cluster 2** (35 Klassen, 345 Schüler:innen) gab es eine tendenziell günstige Leistungsentwicklung im Lesen auf mittlerem Niveau, allerdings eine *ungünstige Entwicklung des lesebezogenen Selbstkonzepts* auf niedrigem Niveau.
- In **Cluster 3** (12 Klassen, 120 Schüler:innen) ist die Entwicklung umgekehrt: Bei einer *ungünstigen Leistungsentwicklung* sind die Werte des Selbstkonzepts etwas positiver als in Cluster 2.

5.2 Clusterbezogene Merkmale von Unterrichtsqualität

Der Vergleich der durchschnittlichen Einschätzungen verschiedener Unterrichtsmerkmale durch die Schüler:innen zwischen dem Optimalcluster (Cluster 1) und den anderen beiden Clustern zeigt folgende Befunde (Tab. 5 und Abb. 1):

Hinsichtlich der *individuellen Bezugsnorm* weicht die Bewertung der Schüler:innen im Optimalcluster kaum von derjenigen in Cluster 2, aber mit kleiner bis mittlerer Effektstärke und, wie hierarchische Regressionsanalysen unterstreichen, auch signifikant zuungunsten von derjenigen in Cluster 3 ab.

Die *soziale Bezugsnorm* wird nach schüler:innenseitiger Wahrnehmung im Optimalcluster weniger angelegt als in den Clustern 2 und 3, wobei die kleinen bis

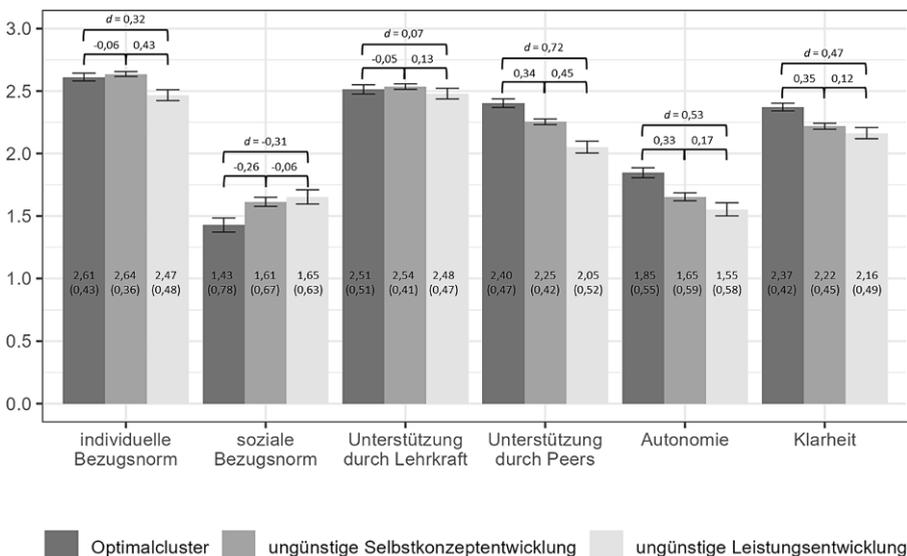


Abb. 1 Durchschnittliche schüler:innenseitige Einschätzung der Unterrichtsqualitätsmerkmale (inklusive Standardfehler der Mittelwerte) pro Cluster mit M (SD) und Effektstärke Cohens d

mittleren Effekte zwischen den Klassen angesichts der relativ hohen Varianz, die auf die Klassenstruktur selbst zurückzuführen ist, nicht signifikant sind.

Die wahrgenommenen Unterschiede bei der *Unterstützung durch die Lehrkraft* zwischen den Clustern sind gering und nicht überzufällig.

In Bezug auf die *Unterstützung durch Peers* erreicht das Optimalcluster tendenziell höhere Einschätzungswerte als Cluster 2 und mit mittlerem bis großem Effekt signifikant bessere Werte als Cluster 3.

Das schüler:innenseitige *Erleben von Autonomie* und die wahrgenommene *Klarheit des Unterrichts* sind im Optimalcluster jeweils mit kleinem bis mittlerem Effekt signifikant höher als in den beiden anderen Clustern.

Insgesamt wird das Optimalcluster damit in fünf der sechs fokussierten Merkmale vergleichsweise günstig bewertet; lediglich bei der Unterstützung durch die Lehrkraft unterscheiden sich die Cluster nicht nennenswert.

6 Diskussion

Auf der Suche nach jahrgangsgemischt unterrichteten „Optimalklassen“ ist festzustellen, dass es in etwa einem Drittel der Klassen gelingt, sowohl Leistungsziele (hier: im Lesen) als auch Ziele im Persönlichkeitsbereich (hier: das lesebezogene Selbstkonzept) zu stärken und auf einem hohen Niveau zu manifestieren. Dies ist insofern bemerkenswert, da es in der Grundschule üblicherweise einen Abwärtstrend des Selbstkonzepts gibt (z.B. Helmke 1998) – und dieser Befund für die jüngeren Kinder in jahrgangsgemischten Klassen sogar noch verstärkt auftritt (Munser-Kiefer et al. 2023). Das verdeutlicht, dass eine gute Leistungsentwicklung bei gleichzeitig günstigem Selbstkonzept auch in jahrgangsgemischten Klassen erreicht werden kann.

Erwartungsgemäß steht diese günstige Entwicklung mit zentralen Merkmalen der Unterrichtsqualität in Zusammenhang: In den Optimalklassen erleben die Schüler:innen den Unterricht als klarer, nehmen sich als selbstbestimmter wahr und stimmen verstärkt zu, Unterstützung durch Peers zu erhalten. Diese Merkmale der Unterrichtsqualität könnten mit Merkmalen der Oberflächenstruktur der Jahrgangsmischung korrespondieren (Hess und Lipowsky 2020): So böte z.B. die Zusammenarbeit mit den älteren Schüler:innen bzw. im erweiterten Leistungsspektrum in jahrgangsgemischten Lerngruppen eine Chance für konstruktive Unterstützung. Die verstärkt wahrgenommene Autonomie könnte ein Indiz für mehr Freiheitsgrade im Unterricht sein, welche wiederum durch den reflektierten Umgang mit der erhöhten Heterogenität entstehen könnten. Die Klarheit des Unterrichts ist eine Voraussetzung für gelingende Unterrichtsprozesse im Allgemeinen, könnte jedoch insbesondere in den individualisierten bzw. kooperativen Lernformen sowie beim zieldifferenten Arbeiten im Abteilungsunterricht von besonderer Bedeutung sein.

Entgegen unseren Erwartungen konnten keine Unterschiede in den wahrgenommenen Bezugsnormen zwischen den Clustern festgestellt werden. Bei der *individuellen Bezugsnorm* lässt sich dies zumindest teilweise durch Deckeneffekte erklären, da sowohl der Skalenmittelwert als auch die Mittelwerte der zugehörigen Einzelitems relativ hoch sind und jeweils mehr als die Hälfte der Schüler:innen die Höchstbewer-

tung vergab ($Md=3$), wie weiterführende Analysen (auch für jedes Cluster) zeigen. Bei der *sozialen Bezugsnorm* sind – aus pädagogischer Sicht erfreulich – die Werte in allen Clustern eher niedrig. Allerdings findet sich auch hier zumindest deskriptiv der geringste Wert im Optimalcluster.

Hinsichtlich der *Unterstützung durch die Lehrkraft* unterscheiden sich die Cluster hingegen nicht, was auch hier möglicherweise auf Deckeneffekte auf Skalen- wie auch Einzelitemebene zurückzuführen ist. Denn vier der fünf Items erhalten relativ hohe Zustimmungswerte und von mehr als der Hälfte aller Schülerinnen sogar die Höchstbewertung ($Md=3$). Nur die Einschätzung des Items „Die Lehrkraft merkt immer, wenn ich Hilfe brauche“ fällt etwas geringer aus, weist aber zugleich auch die höchste Streuung innerhalb dieser Skala auf ($M=2,20$, $Md=2$, $SD=0,77$). Diese insgesamt eher hohen Beurteilungen decken sich mit den Befunden aus der Implementation des Jahrgangübergreifenden Lernens in Berlin (Thoren 2017; Thoren und Brunner 2019).

Die untersuchten Merkmale der Unterrichtsqualität wurden üblicherweise in jahrgangshomogenen Regelklassen nachgewiesen. Unsere Befunde sind Indizien dafür, dass sie weitestgehend auch in der Jahrgangsmischung bedeutsam sind. Zu hinterfragen sind die Merkmale, bei denen es keine Unterschiede zwischen den Clustern gibt, wie z. B. das Ausmaß der wahrgenommenen Unterstützung durch die Lehrkraft. Dieser Befund lässt sich zum einen durch Deckeneffekte erklären (s. oben). Zum anderen besteht jedoch auch die Möglichkeit, dass die lehrkraftseitige Unterstützung in jahrgangsgemischten Klassen für den Lernerfolg der Schüler:innen von geringerer Bedeutsamkeit ist, wenn es zugleich ausreichend Unterstützung durch Peers gibt.

Limitationen und Forschungsdesiderata Unsere Befunde wurden durch Schüler:innenbefragung gewonnen. Dies ist nur eine Perspektive und lässt nur indirekt Rückschlüsse auf den Unterricht und die Klassensituation zu (Fauth et al. 2014). Neuere Befunde stützen jedoch die Wertigkeit von Schüler:innenwahrnehmung (Rieser und Naumann 2024). Folgestudien sollten ebenfalls auf bewährte, aber breitere Skalen zurückgreifen, um Deckeneffekte und Varianzeinschränkungen zu vermeiden, wie auch weitere Perspektiven und Datenquellen einbeziehen (z. B. Lehrkräfteperspektive, Unterrichtsbeobachtungen, Dokumentenanalysen). Darüber hinaus sollte die Bedeutung von Qualitätskriterien in den beiden Organisationsformen – jahrgangshomogen vs. jahrgangsgemischt – auch vergleichend geprüft werden. Dazu wäre auch zu überlegen, inwieweit Qualitätskriterien an die jeweilige Organisationsform anzupassen – und gegebenenfalls zu operationalisieren – sind.

Für das Optimalcluster kann nicht ausgeschlossen werden, dass die günstige Entwicklung in Leistung und Selbstkonzept durch die zu Beginn tendenziell höheren Lesefähigkeiten unterstützt wurden, z. B. dadurch, dass dies den Lehrer:innen erleichtert, verstärkt Wahlmöglichkeiten zu geben oder jahrgangsgemischte Lerngruppen einzusetzen. Dies müsste in Folgestudien geklärt werden. Zudem beschränkt sich unsere Studie auf den Inhaltsbereich Lesen. Hier sollten weitere Inhaltsbereiche in den Blick genommen werden, auch wenn Lesen als zentrale Aufgabe der Grundschule gewertet werden kann. Darüber hinaus könnten differenzielle Effekte für Schüler:innen mit niedrigen Kompetenzen geprüft werden. Da sich hier bereits Hinweise in den Befunden von Munser-Kiefer und Kolleg:innen (2022) fin-

den, werden Unterschiede zwischen Leistungssubgruppen – zusammen mit weiteren Merkmalen (z. B. Unterrichtsgestaltung bzw. -qualität aus Lehrkräfteperspektive) – in Folgeanalysen in den Fokus genommen.

Praktische Implikationen Mit dem Blick auf Optimalklassen ordnen wir uns in die Tradition der Best-Practice-Analysen ein. Diese möchten Lehrkräften – insbesondere bei komplexen Aufgaben wie hier der Jahrgangsmischung – Unterstützung bei der Gestaltung von Unterricht geben. Auf diese Weise liefern Befunde aus diesem Analysetypus Anhaltspunkte für die Aus- und Fortbildung von Lehrkräften.

Mit unserer Studie haben wir versucht, dem „Erfolgsgeheimnis“ der Lehrkräfte auf die Spur zu kommen, deren jahrgangsgemischter Unterricht multikriterial – in leistungs- und persönlichkeitsbezogenen Variablen – zu einer günstigen Entwicklung ihrer Schüler:innen führt. Wir konnten Qualitätskriterien identifizieren, die sich überzufällig (Klarheit, Autonomie, individuelle Bezugsnorm und Unterstützung durch Peers) bzw. deskriptiv (soziale Bezugsnorm) als förderlich erwiesen. Dies gibt Hinweise darauf, wie jahrgangsgemischter Unterricht qualitativ weiterentwickelt werden kann und ist eine Chance, das Denken in Oberflächenstrukturen (wie z. B. Abteilungsunterricht, Individualisierung und Lernen im jahrgangsgemischtem Team) abzulösen durch den Fokus auf qualitätsvolle Gestaltung. Kombiniert mit einer Umstrukturierung der Inhalte, die das Potenzial der Jahrgangsmischung in ihrer durch die unterschiedliche Anzahl an Schulbesuchsjahren zusätzlich gespreizten Heterogenität (Lernerfahrungen, Leistung u. a.) nutzt, könnte dies dazu beitragen, das pädagogisch-didaktische Versprechen dieser Unterrichtsorganisationsform einzulösen.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, 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 in diesem Artikel 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 für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Adelmeier, C., & Liebers, K. (2008). Umsetzung des jahrgangsübergreifenden Unterrichts in den Stundenplänen in FLEX-Klassen. In K. Liebers, A. Prengel & G. Bieber (Hrsg.), *Die flexible Schuleingangsphase. Evaluationen zur Neugestaltung des Anfangsunterrichts* (S. 156–162). Weinheim: Beltz.
- Arbeitskreis Wissenschaftliche Begleitung Schulanfang auf neuen Wegen (2006). *Schulanfang auf neuen Wegen. Abschlussbericht zum Modellprojekt*. Stuttgart: Kohlhammer.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2016). *Multivariate Analysemethoden*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-46076-4>
- Bartoň, K. (2023). MuMin: multi-model inference (R package version 1.47.5). <https://CRAN.R-project.org/package=MuMin>

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bliese, P. (2022). multilevel: Multilevel Functions (R package version 2.7). <https://CRAN.R-project.org/package=multilevel>
- Breidenstein, G., & Rademacher, S. (2013). Vom Nutzen der Zeit. Beobachtungen und Analysen zum individualisierten Unterricht. *Zeitschrift für Pädagogik*, 59(3), 336–356.
- Campana Schleusener, S. (2014). Wenn Lernen und Lehren zusammentreffen: gegenseitiges Helfen in heterogenen Klassen. In B. Kopp, S. Martschinke, M. Munser-Kiefer, M. Haider, E.-M. Kirschhock, G. Ranger & G. Renner (Hrsg.), *Individuelle Förderung und Lernen in der Gemeinschaft* (S. 166–169). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-04479-4_25
- Carle, U., & Metzgen, H. (2009). *Die Schuleingangsphase lohnt sich! Bericht der wissenschaftlichen Begleitung nach zweieinhalb Jahren ‚BeSte‘ (2005–2008)*. Universität Bremen.
- Carle, U., & Metzgen, H. (2017). *Wie wirkt Jahrgangsübergreifendes Lernen? Internationale Literaturübersicht zum Stand der Forschung, der praktischen Expertise und der pädagogischen Theorie*. Frankfurt am Main: Campus.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Eckerth, M., Hanke, P., & Hein, A. K. (2015). Die Gestaltung der Schuleingangsphase in jahrgangsbezogenen und jahrgangsübergreifenden Klassen. Ergebnisse aus dem Projekt FiS („Förderung der Lern- und Bildungsprozesse in der Schuleingangsphase“). In C. Müller, L. Amberg, T. Dütsch, E. Hildebrandt, F. Vogt & E. Wannack (Hrsg.), *Perspektiven und Potentiale in der Schuleingangsphase* (S. 163–172). Münster: Waxmann.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.
- Ertl, S., & Zumhasch, C. (2024). Schulleistungsbeurteilung: Leistungen feststellen, bewerten und rückmelden. In M. Götz, A. Hartinger, F. Heinzel, J. Kahlert, S. Miller & U. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und -didaktik* (5. Aufl.; S. 358–366). Bad Heilbrunn: Klinkhardt.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive. Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für pädagogische Psychologie*, 28(3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>
- Feuchtenberger, S., Martschinke, S., Munser-Kiefer, M., & Hartinger, A. (2019). „Mehr Zeit für einzelne Kinder“ oder „mehr Stress“ – Eine Interviewstudie zu Chancen und Risiken jahrgangsgemischten Lernens in der dritten und vierten Jahrgangsstufe aus der Perspektive von Lehrkräften. In C. Donie, F. Foerster, M. Obermayr, A. Deckwerth, G. Kammermeyer, G. Lenske, M. Leuchter & A. Wilde-mann (Hrsg.), *Grundschulpädagogik zwischen Wissenschaft und Transfer* (S. 263–269). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-26231-0_34
- Feyerer, E. (2012). Der Umgang mit besonderen Bedürfnissen im Bildungswesen. *Zeitschrift für Inklusion*, 4. <https://www.inklusion-online.net/index.php/inklusion-online/article/download/33>. Zugegriffen: 29. Nov. 2024.
- Götz, M., & Krenig, K. (2024). Jahrgangsmischung in der Grundschule. In M. Götz, A. Hartinger, F. Heinzel, J. Kahlert, S. Miller & U. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (5. Aufl.; S. 129–135). Bad Heilbrunn: Klinkhardt. <https://doi.org/10.36198/9783838588360>
- Grittner, F., Hartinger, A., & Rehle, C. (2013). Wer profitiert beim jahrgangsgemischtem Lernen? *Zeitschrift für Grundschulforschung*, 6, 102–113.
- Gutiérrez, R., & Slavin, R. E. (1992). Achievement effects of the nongraded elementary school: a best evidence synthesis. *Review of Educational Research*, 62(4), 333–376. <https://doi.org/10.3102/00346543062004333>
- Hardy, I., Decristan, J., & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for Educational Research Online*, 11(2), 169–191.
- Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitskonzeptes vom Kindergarten bis zur 6. Klassenstufe. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 115–132). Weinheim: Beltz.
- Hess, M., & Lipowsky, F. (2017). Lernen individualisieren und Unterrichtsqualität verbessern. In F. Heinzel & K. Koch (Hrsg.), *Individualisierung im Grundschulunterricht: Anspruch, Realisierung und Risiken* (S. 23–31). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-15565-0_3
- Hess, M., & Lipowsky, F. (2020). Zur (Un-)Abhängigkeit von Oberflächen- und Tiefenmerkmalen im Grundschulunterricht. Fragen von Lehrpersonen im öffentlichen Unterricht und in Schülerarbeitsphasen im Vergleich. *Zeitschrift für Pädagogik*, 66. Beiheft, 117–131. <https://doi.org/10.3262/ZPB2001117>

- Hilbert, S., Stadler, M., Lindl, A., Naumann, F., & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *Testing, Psychometrics, Methodology in Applied Psychology*, 26(1), 101–119.
- Hry-Beihammer, E. K., & Hascher, T. (2015). Multi-grade teaching practices in Austrian and Finnish primary schools. *International Journal of Educational Research*, 74, 104–113. <https://doi.org/10.1016/j.ijer.2015.07.002>.
- Kalaoja, E., & Pietarinen, J. (2009). Small rural primary schools in Finland: a pedagogically valuable part of the school network. *International Journal of Educational Research*, 48, 109–116. <https://doi.org/10.1016/j.ijer.2009.02.003>
- Kleickmann, T. (2024). Unterrichtsqualität. In M. Götz, A. Hartinger, F. Heinzel, J. Kahlert, S. Miller & U. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und -didaktik* (5. Aufl.; S. 406–410). Bad Heilbrunn: Klinkhardt.
- Klöver, B. (2014). *Evaluationsbericht Flexible Grundschule*. München: Staatsinstitut für Schulqualität und Bildungsforschung.
- Kucharz, D., & Wagener, M. (2007). *Jahrgangübergreifendes Lernen. Eine empirische Studie zu Lernen, Leistung und Interaktion von Kindern in der Schuleingangsphase*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Kuhl, P., Felbrich, A., Richter, D., Stanat, P., & Pant, H. A. (2013). Die Jahrgangsmischung auf dem Prüfstand: Effekte jahrgangübergreifenden Lernens auf Kompetenzen und sozio-emotionales Wohlbefinden von Grundschülerinnen und Grundschulern. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte* (S. 299–324). Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-18985-7_11
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Liebers, K., Pregel, A., & Bieber, G. (2008). *Die flexible Schuleingangsphase: Evaluationen zur Neugestaltung des Anfangsunterrichts*. Weinheim Basel: Beltz.
- Lindl, A., Krauss, S., Schilcher, A., & Hilbert, S. (2020). Statistical methods in transdisciplinary educational research. *Frontiers in Education*, 5, 97. <https://doi.org/10.3389/educ.2020.00097>
- Lorenz, R., McElvany, N., Schilcher, A., & Ludewig, U. (2022). Lesekompetenz von Viertklässlerinnen und Viertklässlern im internationalen Vergleich: Testkonzeption und Ergebnisse von IGLU 2021. In N. McElvany, R. Lorenz, A. Frey, F. Goldhammer, A. Schilcher & T. Stubbe (Hrsg.), *IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre* (S. 53–87). Münster: Waxmann. <https://doi.org/10.31244/9783830997009>
- Malti, T., & Perren, S. (2008). *Soziale Kompetenzen bei Kindern und Jugendlichen*. Stuttgart: Kohlhammer. <https://doi.org/10.17433/978-3-17-022724-8>
- Martschinke, S., & Kammermeyer, G. (2003). Schulleistung und Fähigkeitsselbstbild im Anfangsunterricht – Ergebnisse aus dem KILIA-Projekt. *Empirische Pädagogik*, 17, 486–503.
- Martschinke, S., Munser-Kiefer, M., Hartinger, A., & Lindl, A. (2022). „Funktioniert“ die Jahrgangsmischung auch in der 3. und 4. Jahrgangsstufe? Leistungsentwicklung und adaptive Unterstützung im Unterricht. *Grundschule aktuell*, 160, 30–33.
- Matz, S., & Knauf, T. (2010). Altersmischung in der Praxis einer Montessori-Schule – eine Beobachtungsstudie zur Auftretenshäufigkeit ausgewählter Aspekte altersgemischter Lerngruppen. In R. Laging (Hrsg.), *Altersgemischtes Lernen in der Schule* (4. Aufl.; S. 72–79). Baltmannsweiler: Schneider Verlag Hohengehren.
- Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation. *The Computer Journal*, 20(4), 359–363. <https://doi.org/10.1093/comjnl/20.4.359>
- Moschner, B., & Dickhäuser, O. (2018). Selbstkonzept. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 750–756). Weinheim: Beltz.
- Mulryan-Kyne, C. (2007). The preparation of teachers for multigrade teaching. *Teaching and Teacher Education*, 23(4), 501–514. <https://doi.org/10.1016/j.tate.2006.12.003>.
- Munser-Kiefer, M., Martschinke, S., & Hartinger, A. (2016). Selbstbestimmtes Lernen im jahrgangsgemischten Unterricht der 3. und 4. Klasse – erste Ergebnisse aus einer Evaluationsstudie. In F. Heinzel & K. Koch (Hrsg.), *Individualisierung im Grundschulunterricht. Jahrbuch Grundschulforschung 21* (S. 87–91). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-15565-0_13
- Munser-Kiefer, M., Martschinke, S., & Hartinger, A. (2017). Adaptive Unterrichtsgestaltung und Überzeugungen von Lehrpersonen in jahrgangsgemischten und jahrgangshomogenen Klassen. *Zeitschrift für Grundschulforschung*, 10(1), 147–161.
- Munser-Kiefer, M., Martschinke, S., Lindl, A., & Hartinger, A. (2022). Leistungsentwicklung in jahrgangsgemischten und jahrgangshomogenen dritten und vierten Klassen. *Unterrichtswissenschaft*, 50, 453–478. <https://doi.org/10.1007/s42010-021-00132-9>.

- Munser-Kiefer, M., Martschinke, S., Lindl, A., & Hartinger, A. (2023). Development of self-concept in multi-grade 3rd and 4th classes. *International Electronic Journal of Elementary Education*, 15(4), 343–356. <https://doi.org/10.26822/iejee.2023.305>
- Praetorius, A. K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: looking back and looking forward. *ZDM Mathematics Education*, 50, 535–553. <https://doi.org/10.1007/s11858-018-0946-0>
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of three basic dimensions. *ZDM Mathematics Education*, 50, 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Praetorius, A.-K., Rogh, W., & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft*, 48(3), 303–318. <https://doi.org/10.1007/s42010-020-00072-w>
- R Core Team (2024). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raggl, A. (2015). Teaching and learning in small rural primary schools in Austria and Switzerland—Opportunities and challenges from teachers’ and students’ perspectives. *International Journal of Educational Research*, 74, 127–135. <https://doi.org/10.1016/j.ijer.2015.09.007>
- Rathgeb-Schnierer, E., & Rechtsteiner-Merz, C. (2010). *Mathematiklernen in der jahrgangübergreifenden Eingangsstufe: Gemeinsam, aber nicht im Gleichschritt*. Vortrag auf der 43. Tagung für Didaktik der Mathematik. Jahrestagung der Gesellschaft für Didaktik der Mathematik, Oldenbourg, 02.03–06.03.2008.
- Revelle, W. (2024). psych: Procedures for Psychological, Psychometric, and Personality Research (R package version 2.4.3). <https://CRAN.R-project.org/package=psych>
- Rieser, S., & Naumann, A. (2024). Primary school students’ ratings of teaching—do they differentiate between subjects and teachers? *School Effectiveness and School Improvement*, 35(4), 486–505. <https://doi.org/10.1080/09243453.2024.2396942>
- Ronksley-Pavia, M., Barton, G., & Pendergast, D. (2019). Multiage education: an exploration of advantages and disadvantages through a systematic review of the literature. *Australian Journal of Teacher Education*, 44(5), 24–41. <https://doi.org/10.14221/ajte.2018v44n5.2>
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. *Contemporary educational psychology*, 61, 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Schmidt, H. J. (1999). Jahrgangübergreifender Unterricht. In H. Brügelmann, M. Fölling-Albers, S. Richter & A. Speck-Hamdan (Hrsg.), *Jahrbuch Grundschule 1999. Fragen der Praxis – Befunde der Forschung* (S. 120–124). Seelze: Kallmeyer.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: validation of construct interpretations. *Review of Educational Research*, 46(3), 407–441. <https://doi.org/10.3102/00346543046003407>
- Smit, R., Hyry-Beihammer, E. K., & Raggl, A. (2015). Teaching and learning in small, rural schools in four European countries: Introduction and synthesis of mixed-/multi-age approaches. *International Journal of Educational Research*, 74, 97–103. <https://doi.org/10.1016/j.ijer.2015.04.007>
- Sonnleitner, M. (2017). Individualisierung als Entwicklungsziel bei der Einführung von Jahrgangsmischung. In F. Heinzel & K. Koch (Hrsg.), *Individualisierung im Grundschulunterricht. Anspruch, Realisierung und Risiken* (S. 77–81). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-15565-0_11
- Sonnleitner, M. (2021). *Schule entwickeln: Jahrgangsmischung in der Grundschule: Eine empirische Studie zur pragmatisch bedingten Initiierung und Implementierung aus Sicht von Schulleitungen und Lehrkräften*. Bad Heilbrunn: Klinkhardt. <https://doi.org/10.35468/5900>
- Stiftung Bildungspakt Bayern, & Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (2015). *Schulversuch Flexible Grundschule – Dokumentation, Ergebnisse, Empfehlungen für die Praxis*. München.
- Thoren, K. (2017). *Implementationserfolg von Schulreformen in der Berliner Schulanfangsphase*. Freie Universität Berlin.
- Thoren, K., & Brunner, M. (2019). Flächendeckende Implementation des Jahrgangübergreifenden Lernens: Welche Typen gibt es und zeigen diese Unterschiede in der Schul- und Unterrichtsqualität? *Zeitschrift für Erziehungswissenschaft*, 22(2), 279–300. <https://doi.org/10.1007/s11618-018-0841-z>
- Vogt, F. (2009). Altersdurchmischung, Teamteaching und Differenzierung: Evaluationsergebnisse zur Basisstufe in der Schweiz. In C. Röhner, C. Henrichwark & M. Hopf (Hrsg.), *Europäisierung der Bil-*

- dung. Konsequenzen und Herausforderungen für die Grundschulpädagogik* (S. 226–230). Wiesbaden: VS. https://doi.org/10.1007/978-3-531-91721-4_33
- Wagener, M. (2014). *Gegenseitiges Helfen. Soziales Lernen im jahrgangsgemischten Unterricht*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-03402-3>
- Worbach, M., Drechsel, B., & Carstensen, C.H. (2019). Messen und Bewerten von Lernergebnissen. In D. Urhahne, M. Dresel & F. Fischer (Hrsg.), *Psychologie für den Lehrberuf* (S. 493–516). Berlin: Springer. https://doi.org/10.1007/978-3-662-55754-9_25

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.