

Kompetenzentwicklung im Schuljahr nach PISA 2012: Effekte von Klassenwiederholungen

Timo Ehmke · Christine Sälzer · Marcus Pietsch · Barbara Drechsel ·
Katharina Müller

Online publiziert: 12. Mai 2017
© Springer Fachmedien Wiesbaden 2017

Zusammenfassung Diese Studie untersucht die Wirkung von Klassenwiederholungen auf die Entwicklung mathematischer Kompetenz und hinsichtlich mathematikbezogener Motivationen und Einstellungen. Eine für Deutschland repräsentative Stichprobe der 9. Jahrgangsstufe (PISA 2012 und Messwiederholung ein Jahr später) dient als Datenbasis. *Same-age*-Vergleiche zwischen Klassenwiederholerinnen und -wiederholern ($N = 89$) und einer mittels *Propensity Score Matching* (PSM) gewonnenen (aber regulär versetzten) Vergleichsgruppe ($N = 89$) zeigen, dass sich hinsichtlich der mathematischen Kompetenz keine Effekte der pädagogischen Maßnahme Klassenwiederholung nachweisen lassen. Die Klassenwiederholerinnen und -wiederholer (KWH) können ihre mathematische Kompetenz in gleichem Umfang verbessern wie die KWH-Vergleichsgruppe. Hinsichtlich der mathematikbezogenen Motivationen und Einstellungen steigerten sich die Klassenwiederholerinnen und -wiederholer in Bezug auf ihre selbsteingeschätzte mathematikbezogene Ar-

Prof. Dr. T. Ehmke (✉) · Dr. M. Pietsch · Dr. K. Müller
Leuphana Universität Lüneburg, Scharnhorststr. 1, 21335 Lüneburg, Deutschland
E-Mail: tehmke@leuphana.de

Dr. M. Pietsch
E-Mail: pietsch@leuphana.de

Dr. K. Müller
E-Mail: katharina.m.mueller@leuphana.de

PD Dr. C. Sälzer
Zentrum f. intern. Vergleichsstudien, TU München, Arcisstr. 21, 80333 München,
Deutschland
E-Mail: christine.saelzer@tum.de

Prof. Dr. B. Drechsel
Otto-Friedrich-Universität Bamberg, Markusplatz 3, 96047 Bamberg, Deutschland
E-Mail: barbara.drechsel@uni-bamberg.de

beitseinstellung. Keine Effekte zeigten sich für das mathematikspezifische Interesse, die Selbstwirksamkeit, die instrumentelle Motivation und das Arbeitsverhalten.

Schlüsselwörter Klassenwiederholung · Propensity Score Matching · Mathematische Kompetenz · PISA · Mathematikbezogene Motivation · Mathematikbezogene Einstellungen

Competence development in the school year after PISA 2012: effects of grade retention

Abstract The study analyzes the effects of grade retention on the development of mathematics competence and on motivation and beliefs in the field of mathematics using data from a representative sample of 9th graders in Germany (PISA 2012 and repeated measurement one year later). Same-age comparisons were applied between a group of retained students ($N = 89$) and a matched comparison group of promoted students ($N = 89$), applying the method of *Propensity Score Matching* (PSM). The results show no different effects in the development of mathematics competence between retained students and promoted students. Concerning motivation and beliefs in the field of mathematics the retained students significantly increased their mathematics work ethics compared to the promoted group. However, no group specific effects could be found for mathematics interest, self-efficacy, instrumental motivation, and behavior.

Keywords Grade retention · Mathematics competence · Propensity Score Matching · PISA · Mathematics related motivation and beliefs

1 Einleitung

Schulklassen sind einander bezüglich ihrer Alterszusammensetzung relativ ähnlich. In Staaten mit gegliederten Schulsystemen, etwa Deutschland, den Niederlanden, Korea oder Belgien, werden Lernende darüber hinaus auch nach ihrer Leistung gruppiert. Da die Schülerinnen und Schüler aufgrund ihrer früheren Leistungen auf unterschiedliche weiterführende Schulformen verteilt werden, sind Schulklassen im gegliederten System vergleichsweise leistungshomogen. Allerdings machen Schülerinnen und Schüler unterschiedlich schnelle Fortschritte, so dass sich die anfängliche Homogenität im Verlauf der Schullaufbahn, auch innerhalb eines Schuljahres, zunehmend diversifiziert. Eine Maßnahme zur Wiederherstellung einer relativen Leistungshomogenität ist die Wiederholung einer Klassenstufe. Betroffen sind Schülerinnen und Schüler, die die grundlegenden Lernziele einer Jahrgangsstufe nicht erreicht haben und bei denen ein erfolgreiches Lernen in der nächsten Jahrgangsstufe nicht erwartet werden kann (Roßbach und Tietze 2010).

Die Frage, inwieweit die Wiederholung eines Schuljahres aufgrund mangelhafter Leistungen lernförderlich ist und welche Nebenwirkungen damit verbunden sein können, ist nach wie vor Gegenstand der Diskussion (vgl. Boller et al. 2013; Ehmke et al. 2008; Hong und Raudenbush 2005; Krohne et al. 2011; OECD 2011;

Klapproth et al. 2016). Aus der Perspektive der unterrichtenden Lehrperson können Klassenwiederholungen insofern von Vorteil sein, als damit in Lerngruppen die Leistungshomogenität gestärkt bzw. wiederhergestellt wird und sie einfacher zu unterrichten sind. Aus Sicht der betroffenen Schülerinnen und Schüler überwiegen nach Analysen von Daten aus PISA 2003 aufgrund des Entzugs von Lerngelegenheiten, des Verlusts relevanter sozialer Einbindung durch Peers oder auch durch die Minderung der Lernmotivation die Nachteile (Ehmke et al. 2010). Das Verhältnis von aufgewendeten Ressourcen zu den Erträgen einer Klassenwiederholung stellt sich oft als unvorteilhaft heraus. Zudem liegt wenig empirische Evidenz für die Wirksamkeit der Wiederholung einer Klasse vor (Alexander et al. 2003; Klemm 2009).

International vergleichende Schulleistungsstudien wie PISA haben wiederholt gezeigt, dass die pädagogische Maßnahme Klassenwiederholung in verschiedenen Staaten unterschiedlich häufig genutzt wird. In einer Reihe von OECD-Staaten gibt es so gut wie keine Klassenwiederholungen (z. B. Japan, Norwegen), in weiteren Staaten besteht diese Möglichkeit auf Wunsch der Schülerinnen und Schüler (beispielsweise in Estland, Finnland oder Griechenland) und in einer dritten Gruppe von Staaten, zu der neben z. B. Frankreich, Chile und den Niederlanden auch Deutschland zählt, sind Klassenwiederholungen ein mehr oder weniger häufig eingesetztes Instrument (z. B. OECD 2010; Sälzer et al. 2016). Bemerkenswert dabei ist, dass sich kein klarer Zusammenhang zwischen der Häufigkeit von Klassenwiederholungen und dem durchschnittlichen Kompetenzniveau der fünfzehnjährigen Schülerinnen und Schüler in verschiedenen Staaten abzeichnet (vgl. etwa OECD 2011, 2013; Sälzer et al. 2013, 2016; Goos et al. 2013). Klassenwiederholungen werden also weder ausschließlich in besonders leistungsstarken, noch vorzugsweise in besonders leistungsschwachen Staaten praktiziert.

In PISA 2003 war die Klassenwiederholungsquote Deutschlands mit 23,1 % eine der höchsten Raten weltweit (Drechsel und Senkbeil 2004) und sorgte damals für Aufsehen. Im Anschluss an PISA 2003 waren in der Erweiterungsstudie PISA Plus die beteiligten Jugendlichen nach einem Schuljahr erneut getestet worden (Prenzel et al. 2006). Im Rahmen dieser Messwiederholungsstudie wurde damals der Lernzuwachs der Klassenwiederholerinnen und -wiederholer im Vergleich zu ihren versetzten Mitschülerinnen und Mitschülern untersucht (Ehmke et al. 2008). Die Effekte der Klassenwiederholung wurden mit Blick auf die mathematische Kompetenz und das Selbstkonzept der beiden Gruppen im Vergleich betrachtet. Die damaligen Befunde weisen auf nicht bedeutsame Effekte bezüglich der mathematischen Kompetenz hin, wenn eine Jahrgangsstufe wiederholt wurde. Die Wiederholerinnen und Wiederholer zeigten aber ein robusteres Selbstkonzept als ihre versetzten Kolleginnen und Kollegen (Ehmke et al. 2010).

Im Nachgang zu PISA 2003 wurden in Deutschland zahlreiche Maßnahmen ergriffen, welche die im internationalen Vergleich hohen Wiederholungsquoten reduzieren sollten. Beispielsweise wurde in Hamburg das Klassenwiederholen als pädagogische Maßnahme bis einschließlich der 10. Klasse ersetzt durch „individuelle Lern- und Fördervereinbarungen“, die zwischen Schülerin bzw. Schüler und der Schule unter Einbezug der Eltern geschlossen werden. Für diese ergänzenden Fördermaßnahmen werden den Schulen auch zusätzliche Ressourcen zur Verfügung

gestellt. In Bremen drohen Klassenwiederholungen erst ab der achten Klasse und in Berlin nur den Gymnasiastinnen und Gymnasiasten. Befunden aus PISA 2015 zufolge liegt der Anteil der fünfzehnjährigen Schülerinnen und Schüler, die in ihrer bisherigen Schullaufbahn bereits mindestens einmal ein Schuljahr wiederholt haben, mit nunmehr 18,1 % etwas niedriger, jedoch nach wie vor deutlich über dem OECD-Durchschnitt (12,0 %; vgl. Sälzer et al. 2016). Dieser leichte Trend zur Verringerung der Klassenwiederholungsquote könnte künftig verstärkt werden, da in einzelnen Ländern Deutschlands seit kurzem auf die Wiederholung von Klassen verzichtet wird. Allerdings sei erwähnt, dass auch Befunde in die entgegengesetzte Richtung weisen, beispielsweise berichten Huebener und Marcus (2015) über einen leichten Anstieg der Klassenwiederholungsquoten vor allem der höheren Klassen im Gymnasium im Zuge der Einführung des achtjährigen Gymnasiums.

Ausgehend von der nach wie vor uneindeutigen Befundlage zur Wirksamkeit von Klassenwiederholungen und der noch immer relativ hohen Klassenwiederholungsquote in Deutschland nutzt der vorliegende Beitrag die Chance einer erneuten Messwiederholungsstudie im Anschluss an PISA 2012 und knüpft unmittelbar an die Befunde der Messwiederholungsstudie in PISA 2003 an (Ehmke et al. 2010). Untersucht wird, inwieweit sich a) die mathematische Kompetenz und b) mathematikbezogene Motivationen und Einstellungen von Schülerinnen und Schülern, die die neunte Klassenstufe wiederholt haben, im Schuljahr nach dem PISA-Test 2012 im Vergleich zu Schülerinnen und Schülern ohne Klassenwiederholung entwickelt haben.

2 Stand der Forschung: Effekte von Klassenwiederholungen

Die Untersuchung möglicher Folgen von Klassenwiederholungen hat eine lange Tradition. Eine Reihe älterer Meta-Analysen (etwa Jackson 1975; Holmes 1989; Jimerson 2001a, 2001b), berichtet ambivalente bis vorwiegend negative Effekte von Klassenwiederholungen. Dies gilt sowohl für die Leistungen als auch für die nicht-kognitive Anpassung der betreffenden Schülerinnen und Schüler, wobei die negativen Effekte sich bei den akademischen Leistungen stärker als bei den sozio-emotionalen Faktoren, beispielsweise dem Selbstkonzept, zeigen. Typisch für diese Befunde ist auch, dass positive Effekte, sofern sie gefunden wurden, lediglich für das wiederholte Schuljahr nachgewiesen werden konnten und die verbesserten Leistungen in den Folgejahren wieder schwächer wurden.

Mit der Methodenkritik von Lorence (2006), der auf z. T. erhebliche Designmängel bisheriger Studien hinwies, beginnt eine neue Phase der wissenschaftlichen Auseinandersetzung mit dem Thema. Das Fehlen adäquater Vergleichsgruppen sei ein Hauptproblem, das gezogene Schlussfolgerungen in Frage stelle. Klassenwiederholungen können nach Lorence (2006) durchaus einen positiven Beitrag zur Verbesserung schulischer Leistungen liefern. Diese neue Perspektive wurde kurze Zeit später von einer Meta-Analyse mit Mehrebenen-Design von Allen et al. (2009) bekräftigt, welche die Qualität des Studiendesigns als Ebene-2-Variable berücksichtigte und als zentralen Befund hervorbrachte, dass ein qualitativ hochwertigeres Studiendesign damit verbunden ist, dass Klassenwiederholung und Schulleistungen weniger stark

zusammenhängen als aufgrund bisheriger Studienergebnisse angenommen. Für die Qualität des Designs scheint es besonders wichtig zu sein, wie die Studien mit der bereits vor der Klassenwiederholung vorhandenen Heterogenität der untersuchten abhängigen Variablen (z. B. Schulleistungen) umgehen. Insgesamt bedarf es gemäß Allen et al. (2009) einer größeren Zahl an Studien mit Kontrollgruppendesign, die eine Form von Matching oder anderer statistischer Kontrolltechniken zur Datenanalyse einsetzen. Im Vergleich zu den früheren Meta-Analysen sprechen die Ergebnisse von Allen et al. (2009) weniger eindeutig für einen Vorzug der Versetzung in das nächste Schuljahr gegenüber Klassenwiederholungen, wenn für die Leistungsheterogenität vor der Klassenwiederholung kontrolliert wird.

Auch wenn in jüngerer Zeit keine weiteren Meta-Analysen zu möglichen Wirkungen von Klassenwiederholungen publiziert wurden, so greifen einige Publikationen in einzelnen Studien diese methodischen Kritikpunkte auf. Ellsworth und Lagacé-Séguin (2009) konnten in einer retrospektiven Matchingstudie zeigen, dass sich die Experimentalgruppe in Bezug auf das Selbstkonzept und die Selbstwirksamkeit nicht von der Kontrollgruppe unterschied. Das Matching erfolgte anhand der Variablen Alter, Geschlecht, Notendurchschnitt im letzten Schuljahr der Sekundarstufe I, Familienkonstellation sowie Familienstressoren. Einschränkend muss jedoch bedacht werden, dass alle 51 an der Studie beteiligten Versuchspersonen Studierende waren und damit eine Positivauswahl eher leistungsstarker Personen vorliegt. Auch Martin (2011) verfolgte in seiner Studie zu Effekten von Klassenwiederholungen auf akademische und sozioemotionale Variablen wie das akademische Selbstkonzept, die Lernmotivation oder die Häufigkeit von Schulabsentismus einen Matching-Ansatz, der neben der kognitiven Entwicklung auch motivationale Variablen untersucht. Die Schülerinnen und Schüler wurden dabei nach Leistungen, Alter und Geschlecht gematcht, so dass Effekte von Klassenwiederholungen sowohl in Bezug auf die Kompetenz der Jugendlichen als auch hinsichtlich ihrer Motivation und ihres Selbstkonzepts gezeigt werden konnten. Martins Ergebnisse (2011) lassen auf negative Effekte von Klassenwiederholungen in Bezug auf das akademische Selbstkonzept, die Hausaufgaben erledigung, die Motivation und die Anwesenheit an der Schule schließen. Diese zunächst in einer Large-Scale Stichprobe ($N = 3261$) gefundenen Effekte wurden in einer Follow-Up-Studie mit 372 Schülerinnen und Schülern (Experimental- und Kontrollgruppe jeweils $n = 186$) bestätigt und unterstrichen. In ihrer Matching-Studie nahmen Lamote et al. (2014) den Effekt von Klassenwiederholungen auf die sprachlichen Leistungen und das akademische Selbstkonzept von Schülerinnen und Schülern in den Blick. Dabei wurden Achtklässlerinnen und Achtklässler mit einem *Propensity-Score*-Verfahren so gematcht, dass Wiederholende und Nicht-Wiederholende anhand latenter Wachstumskurven miteinander verglichen werden konnten. Als unmittelbare Folgen einer Klassenwiederholung, d. h. im Jahr der Wiederholung, fanden Lamote et al. (2014) keinen Zusammenhang mit der Sprachkompetenz und einen positiven Effekt hinsichtlich des akademischen Selbstkonzepts. Mittelfristig, d. h. mit dem nächsten Schulabschluss der untersuchten Jugendlichen, kehren sich die Effekte um: Während die Sprachkompetenz der Klassenwiederholerinnen und -wiederholer deutlich geringer ausfällt als die der Schülerinnen und Schüler ohne Wiederholung, findet sich kein Zusammenhang hinsichtlich des akademischen Selbstkonzepts. Klapproth et al. (2016) konnten anhand einer Längsschnittstudie

in Luxemburg mit Klassenwiederholerinnen und -wiederholern in der siebten und achten Klasse ebenfalls zeigen, dass kurz- und mittelfristige Effekte differenziert betrachtet werden müssen: Während im ersten Jahr nach der Wiederholung bei den Siebt- wie bei den Achtklässlerinnen und Achtklässlern signifikante Notenunterschiede gefunden wurden, haben diese Vorteile gegenüber den versetzten Schülerinnen und Schülern nach einem Jahr ein geringeres Ausmaß und sind nach zwei Jahren nicht mehr nachweisbar. Ähnliche Effekte finden sich für standardisierte Leistungstests. Auch einige nicht-kognitive Variablen wurden untersucht (Fragebogendaten zum Selbstkonzept, zu schulbezogener Angst, eigenen Interessen und Schulzufriedenheit sowie die Einschätzung des Klassenklimas, der Lehrer-Schüler-Beziehung und die Häufigkeit von Störungen in der Klasse), die sich – mit Ausnahme des Selbstkonzepts, das zugunsten der versetzten Schülerinnen und Schüler ausfiel – allesamt nicht signifikant voneinander unterschieden, was die Autoren mit den vergleichsweise sehr hohen Klassenwiederholungsquoten in Luxemburg erklären, durch welche Klassenwiederholungen eher den Status der Normalität als eines Stigmas haben (Klapproth et al. 2016).

Einige Studien betrachteten weitere abhängige Variablen im Zusammenhang mit Klassenwiederholungen: Einen mehrebenenanalytischen Ansatz nutzten Demanet und van Houtte (2013). In ihrer Studie mit knapp 12.000 Schülerinnen und Schülern in 85 Sekundarschulen untersuchten die Autoren den Zusammenhang zwischen Klassenwiederholung und auffallendem Verhalten in der Schule. Analysiert wurde dabei sowohl die Rolle des Anteils von Klassenwiederholerinnen und -wiederholern pro Schule (Ebene-2-Variable) als auch der individuelle Zusammenhang von Klassenwiederholung und Verhaltensauffälligkeiten. Erwartungsgemäß zeigten die Schülerinnen und Schüler an Schulen mit einer relativ hohen Wiederholungsquote eher abweichendes Verhalten als Schülerinnen und Schüler an anderen Schulen. Besonders zentral ist der Befund, dass Jugendliche, die bereits im Grundschulalter eine Klasse wiederholt hatten, seltener verhaltensauffällig waren als Jugendliche, die in der Sekundarstufe sitzen geblieben waren. Zu beachten ist, dass der Zusammenhang zwischen Klassenwiederholung und auffallendem Verhalten durch die Quote der Wiederholer pro Schule moderiert wird. Den Befund, dass der Effekt von Klassenwiederholungen vom Zeitpunkt der Wiederholung abhängt, bestätigen auch die Studien von Fruehwirth et al. (2011) und Garcia-Pérez et al. (2014). In einer Kombination von Kontrollgruppendesign und einer Erweiterung des Fixed-Effects-Ansatzes analysierten Fruehwirth et al. (2011) längsschnittliche, national repräsentative Daten aus den USA und zeigten, dass sowohl der Zeitpunkt der Wiederholung als auch latente Fähigkeitsmaße der untersuchten Schülerinnen und Schüler relevant sind für die gefundenen Kompetenzzuwächse. Garcia-Pérez et al. nutzten die spanischen Daten aus PISA 2009 und untersuchten den Zusammenhang zwischen Klassenwiederholungen und der Kompetenz der Schülerinnen und Schüler. Insgesamt fanden sie einen negativen Effekt von Wiederholungen auf das in PISA gemessene Kompetenzniveau. Allerdings liegen starke Selektionseffekte vor. Anders als in der Studie von Demanet und van Houtte (2013) schlägt sich bei Garcia-Pérez et al. (2014) eine Wiederholung im Grundschulalter stärker negativ nieder als eine Wiederholung in der Sekundarstufe, wobei bei Demanet und van Houtte (2013) Ver-

haltensauffälligkeiten und bei Garcia-Pérez et al. (2014) die Kompetenz im PISA-Test als abhängige Variable betrachtet wurden.

Demski und Liegmann (2014) erweiterten die Forschungsperspektive, indem anhand von Daten des Nationalen Bildungspanels (NEPS) die Zusammenhänge von Klassenwiederholungen, Berufsqualifikation und Arbeitslosigkeit und damit mögliche Spätfolgen von Klassenwiederholungen untersucht wurden. Diese Untersuchung greift also andere abhängige Variablen auf als die meisten weiteren Studien zum Thema. Die Muster, die sich bei der Analyse der Effekte von Klassenwiederholungen auf die weitere Bildungs- und Erwerbsbiographie finden, sind insgesamt uneinheitlich. Während man erwarten würde, dass die Wiederholungsquote unter den beruflich Erfolgreichen niedriger sein müsste als bei weniger Erfolgreichen, zeigt sich dieser Zusammenhang nicht. In Bezug auf Arbeitslosigkeit ist zumindest die Tendenz zu erkennen, dass die Gruppe von Personen mit einer Klassenwiederholung insgesamt häufiger arbeitslos ist oder war als die Gruppe ohne Klassenwiederholung.

Zusammengefasst greifen die einschlägigen wissenschaftlichen Studien auf unterschiedliche Stichproben, Verfahren und Vergleichsperspektiven zurück und legen nahe, dass es insgesamt deutlich mehr negative Effekte gibt als erstrebenswerte Wirkungen.

3 Fragestellungen

Basierend auf den Test- und Fragebogendaten aus PISA 2012 und aus dem gleichzeitig durchgeführten IQB-Ländervergleich im Fach Mathematik soll mit der vorliegenden Messwiederholungsstichprobe in PISA 2012 eine quasi Replikation der Analysen zu Klassenwiederholungen in PISA 2003 (vgl. Ehmke et al. 2010) vorgenommen werden. Im Zentrum stehen die Schülerinnen und Schüler der 9. und 10. Jahrgangsstufe, für die folgende Fragestellung beantwortet werden soll:

Inwieweit können sich Schülerinnen und Schüler, die die 9. Jahrgangsstufe wiederholen, im Vergleich zu einer adäquaten Vergleichsgruppe, die regulär in die 10. Jahrgangsstufe versetzt worden ist, a) in ihrer mathematischen Kompetenz und b) in ihren mathematikbezogenen Motivationen und Einstellungen verbessern?

Mit Bezug auf die Leistungsentwicklung kann dabei vermutet werden, dass ein reguläres Versetzen von leistungsschwachen Schülerinnen und Schülern mit einem geringen Leistungszuwachs in der nächsthöheren Klassenstufe einhergeht. Dahinter steht die Annahme, dass diese Kinder nicht genügend Vorwissen besitzen, um den Anforderungen gewachsen zu sein und Anschluss an den neuen Unterrichtsstoff zu behalten (Plummer und Graziano 1987). Diese Annahme ist gerade in Schulsystemen verbreitet, die auf eine leistungsorientierte Gliederung ausgerichtet sind. Mit Blick auf die gegenüber PISA 2003 geringere Wiederholungsquote in PISA 2012 (23,1 % versus 20,3 %) ist ferner anzunehmen, dass ein nicht unerheblicher Teil sogenannter Grenzfälle, die 2003 noch nicht versetzt worden wären, 2012 hingegen schon, bei jeweils ähnlichen Leistungen nun nicht mehr zur Gruppe der Wiederholenden gehören, sondern zur Gruppe der Versetzten. Ein zweites Argument bezieht sich auf die Verlängerung der Lernzeit durch die Klassenwiederholung. Demnach kann die pädagogische Maßnahme als eine Form der Förderung im Sinne des zieler-

reichenden Lernens gesehen werden, die den Kindern ein Schuljahr mehr Zeit zum Lernen bietet (McCoy und Reynolds 1999; Reynolds 1992). Vor diesem Hintergrund wäre zu erwarten, dass es für Schülerinnen und Schüler mit geringen Schulleistungen adäquater ist eine Klassenstufe zu wiederholen als in die nächsthöhere Stufe versetzt zu werden. Entsprechend sollten sich die Klassenwiederholenden mit niedrigen Mathematikleistungen deutlicher im Fach Mathematik verbessern können als leistungsmäßig vergleichbare Schülerinnen und Schüler, die regulär versetzt worden sind. Für die Mehrheit der regulär versetzten Schülerinnen und Schüler ist anzunehmen, dass die Anforderungen der höheren Klassenstufe nicht zu hoch angesetzt sind. Zudem lernen diese Kinder neuen Unterrichtsstoff kennen, während die Klassenwiederholerinnen und -wiederholer die gleichen Unterrichtsthemen ein zweites Mal durchnehmen. Unterstützende pädagogische Maßnahmen sind an Schulen in Deutschland bislang eher weniger verbreitet (Bellenberg et al. 2005). Entsprechend lässt sich folgende Hypothese formulieren:

(H1) Leistungsschwache Jugendliche, die eine Klassenstufe wiederholen, können sich in ihrer mathematischen Kompetenz deutlicher steigern als vergleichbar leistungsschwache Schülerinnen und Schüler, die regulär versetzt werden.

Im Hinblick auf das sozio-emotionale Befinden der Klassenwiederholerinnen und -wiederholer lassen sich auf Grundlage der Literatur zwei gegenläufige theoretische Annahmen treffen. Zum einen kann vermutet werden, dass durch die Klassenwiederholung der Leistungsdruck zurückgeht und die Kinder nicht mehr zu den Schwächsten in einer Klasse zählen, weil sie ja in eine leistungsschwächere Gruppe versetzt worden sind (Bless et al. 2004). Bei positiven Effekten durch das Wiederholen einer Jahrgangsstufe wird auch die Reife der Wiederholerinnen und Wiederholer ins Feld geführt (z. B. Wu et al. 2010), die ein Jahr älter sind als ihre ebenfalls leistungsschwachen versetzten Mitschülerinnen und Mitschüler. Nach dieser Annahme sollten sich das mathematische Selbstkonzept und auch die Selbstwirksamkeit der Klassenwiederholerinnen und -wiederholer durch die Maßnahme verbessern.

Möglicherweise erhöht sich insbesondere an nicht-gymnasialen Schulformen bei den Klassenwiederholerinnen und -wiederholern im Vergleich zu versetzten Schülerinnen und Schülern mit ähnlichem Leistungsniveau die instrumentelle Motivation im Fach Mathematik. Diese Schülerinnen und Schüler stehen vor dem Einstieg in die Arbeitswelt und bekommen durch die Klassenwiederholung noch einmal eine Chance, ihre Kompetenzen zu verbessern und sich auf den Übergang in den Beruf vorzubereiten. Möglicherweise wird ihnen dadurch bewusster, wie wichtig auch mathematische Kompetenzen bei Bewerbungen in der Berufswelt sind. Es wird zudem davon ausgegangen, dass die Schülerinnen und Schüler sich in ihrem neuen, leistungsschwächeren Klassenkontext mehr zutrauen, weil ihre individuelle Kompetenz im Vergleich zur neuen Bezugsgruppe nicht mehr unbedingt am unteren Ende der Verteilung liegt. Zum anderen lässt sich vermuten, dass ein Wechsel der Peer-Group insbesondere im Jugendalter nicht ohne Probleme verläuft und die „Neuzugänge“ möglicherweise stigmatisiert werden und dementsprechend Schwierigkeiten haben, sich in einen bestehenden festen Klassenverbund zu integrieren. Stigmatisierung kann mit einem Verlust von Selbstwirksamkeitserleben und Motivation einhergehen (Pagani et al. 2001). Wenn dies der Fall ist, wäre zu erwarten, dass sich die

Einstellungen, insbesondere gegenüber schulischen Arbeiten, das Interesse am Fach Mathematik sowie die fachspezifische Selbstwirksamkeit der Klassenwiederholerinnen und -wiederholer negativ entwickeln. Die empirische Befundlage ist hier auch nicht eindeutig. Während Ehmke et al. (2010) im ersten Jahr nach der Klassenwiederholung einen positiven Effekt auf das mathematikbezogene Selbstkonzept durch die Klassenwiederholung fanden, wurden in der Meta-Analyse überwiegend negative Konsequenzen herausgearbeitet (Jimerson 2001a, 2001b). Entsprechend der aktuelleren Befunde aus der PISA-Messwiederholung 2003/2004 vermuten wir:

(H2) Klassenwiederholerinnen und -wiederholer können sich in ihren mathematikbezogenen Motivationen und Einstellungen durch die Klassenwiederholung verbessern, während diese in der leistungsschwachen Vergleichsgruppe, die aber regulär versetzt wurde, auf unverändertem Niveau bleiben.

4 Methode

4.1 Stichprobe und Erhebungsdesign

Für die Analysen in dieser Studie werden die Daten der beiden Studien PISA 2012 (Prenzel et al. 2013) und des IQB-Ländervergleichs 2012 (Pant et al. 2013) sowie deren Messwiederholung ein Jahr später herangezogen. In beiden Studien wurden unter anderem die mathematische Kompetenz und die darauf bezogenen motivationalen Einstellungen der Schülerinnen und Schüler erhoben. Beide Studien basieren jeweils auf repräsentativen, klassenbasierten Stichproben der 9. und 10. Jahrgangsstufe, wobei die Schülerstichprobe in PISA 2012 eine Teilmenge der Schülerstichprobe des IQB-Ländervergleichs 2012 darstellte. Die Schülerinnen und Schüler nahmen also zunächst an PISA teil und einige Tage später am Ländervergleich.

Die realisierte Stichprobe in der PISA 2012-Erhebung umfasst $N = 9998$ Schülerinnen und Schüler aus 222 Schulen. Für die Analysen wurden alle Hauptschulen, Berufsschulen und Förderschulen ausgeschlossen, da diese nicht durchgängig eine 10. Jahrgangsstufe anbieten. Im Datensatz verbleiben nach dieser Auswahl noch $N = 8874$ Schülerinnen und Schüler. In einer dritten und vierten Stufe wurden alle Personen ausgeschlossen, für die keine gültigen Angaben zur Klassenstufe und keine Mathematik-Testergebnisse vorliegen. Der finale Datensatz besteht aus $N = 3287$ Schülerinnen und Schülern, davon $N = 106$ Jugendliche, die die 9. Jahrgangsstufe wiederholt haben (vgl. Tab. 1).

Das Design der Studie besteht aus einem *same-age*-Vergleich, bei dem Klassenwiederholerinnen und -wiederholer mit gleichaltrigen Jugendlichen vor und nach dem Wiederholungsjahr verglichen werden.

Tab. 1 Stichprobenübersicht

Schulart	MWH-Stichprobe	MWH-Stichprobe Ohne Hauptschule und Förderschulen	MWH-Stichprobe Ohne Hauptschule und Förderschulen Angaben für Klassenstufe vorh.	MWH-Stichprobe Ohne Hauptschule und Förderschulen Angaben für Klassenstufe vorh. PVs Mathematik t1/t2 vorhanden	MWH-Stichprobe Ohne Hauptschule und Förderschulen Angaben für Klassenstufe vorh. Regulär versetzt	MWH-Stichprobe Ohne Hauptschule und Förderschulen Angaben für Klassenstufe vorh. PVs Mathematik t1/t2 vorhanden Klassenwiederholer/in
Hauptschule	971	–	–	–	–	–
MBG	1398	1398	911	464	430	34
Realschule	2631	2631	1907	1028	996	32
Gesamtschule (IG)	1020	1020	562	300	297	3
Gymnasium	3825	3825	2823	1495	1458	37
Sonderschule/ Förderschule	153	–	–	–	–	–
Gesamt	9998	8874	6203	3287	3181	106

4.2 Testinstrumente und Fragebogenskalen

4.2.1 Mathematische Kompetenz

Zur Messung der mathematischen Kompetenz im IQB-Ländervergleich (Pant et al. 2013) wurden 300 Testitems eingesetzt und skaliert, die auf der Grundlage der Bildungsstandards für die Sekundarstufe I in Mathematik entwickelt worden sind. Das technische Vorgehen der Skalierung wird ausführlich in Nagy et al. (2017) beschrieben. In der vorliegenden Studie werden die normierten *Plausible Value*-Werte der Gesamtskala Mathematik herangezogen. Ein *Plausible Value*-Wert von 0 entspricht damit der durchschnittlichen Mathematikkompetenz in der Gesamtstichprobe (inklusive der Hauptschulen).

4.2.2 Mathematikbezogene Motivationen und Einstellungen

Die Messung der mathematikbezogenen Motivation und Einstellungen der Schülerinnen und Schüler erfolgte durch fünf Fragebogenskalen aus dem internationalen PISA-Schülerfragebogen, die jeweils mit identischen Items zu t1 und t2 erfasst wurden. Als Index für die Skalen wurde jeweils der Mittelwert aus den Einzelitems gebildet. Tab. 2 zeigt für alle fünf Skalen jeweils ein Beispielitem, das jeweili-

Tab. 2 Fragebogenskalen zu den mathematikbezogenen Motivationen und Einstellungen aus PISA 2012

Merkmal	Beispielitem	Itemanzahl	Antwortformat	Cronbachs α (t ₁ /t ₂)
Interesse an Mathematik	Ich mache Mathematik, weil es mir Spaß macht	4	1 = stimme nicht zu, 2 = stimme eher nicht zu, 3 = stimme eher zu, 4 = stimme voll zu	0,89/0,89
Instrumentelle Motivation in Mathematik	Ich gebe mir in Mathematik Mühe, weil es mir in meinem späteren Job weiterhelfen wird	4	1 = stimme nicht zu, 2 = stimme eher nicht zu, 3 = stimme eher zu, 4 = stimme voll zu	0,86/0,85
Arbeitseinstellung in Mathematik	Im Mathematik-Unterricht passe ich auf	9	1 = stimme nicht zu, 2 = stimme eher nicht zu, 3 = stimme eher zu, 4 = stimme voll zu	0,87/0,88
Mathematikbezogenes Verhalten	Ich spreche mit meinen Freunden über Mathematik	8	1 = nie, 2 = manchmal, 3 = oft, 4 = immer	0,70/0,74
Mathematikbezogene Selbstwirksamkeitserwartung	Ich kann Graphiken in Zeitungen verstehen	8	1 = gar nicht, 2 = eher nicht sicher, 3 = eher sicher, 4 = sehr sicher	0,82/0,84

ge Antwortformat, die Anzahl der Einzelitems und die Skalenreliabilität zu beiden Messzeitpunkten. Für alle Skalen in Tab. 2 wurde von den Autorinnen und Autoren die Messinvarianz zwischen beiden Messzeitpunkten überprüft. Die Ergebnisse (auf Anfrage bei den Autorinnen und Autoren erhältlich) zeigen, dass hier von einer konfiguralen Messinvarianz ausgegangen werden kann.

4.2.3 Soziodemographische Angaben und weitere Schülermerkmale

Soziodemographische Angaben wie Alter, Geschlecht und Informationen zur sozialen Herkunft der Schülerinnen und Schüler wurden ebenfalls mithilfe eines Fragebogens erhoben. Als Indikator für die soziale Herkunft wurde der ESCS-Index gebildet, der Informationen zum sozioökonomischen Status, zum elterlichen Bildungsabschluss und zu häuslichen Besitztümern integriert (Ehmke und Siegle 2005). Zudem wurden die Schulnoten in den Unterrichtsfächern Mathematik, Deutsch, Englisch und den Naturwissenschaften bzw. den naturwissenschaftlichen Fächern Biologie, Chemie und Physik erhoben.

4.3 Umgang mit fehlenden Werten

In empirischen Studien und Schulleistungsuntersuchungen stellen fehlende Werte ein systematisches Problem dar. In Selektivitätsanalysen hat sich gezeigt, dass bei Jugendlichen mit geringen Leistungen häufiger Angaben in Fragebögen fehlen als bei Schülerinnen und Schülern mit höheren Kompetenzen. Dies kann zu metho-

dischen Verzerrungen bei den Ergebnisanalysen führen. Um diesem Problem zu begegnen, wurden die Auswertungen für die Gruppenvergleiche mit dem Programm MPlus 6 (Muthén und Muthén 2010) durchgeführt, das fehlende Werte anhand des FIML-Algorithmus (Full Information Maximum Likelihood) schätzt.

Der Anteil an fehlenden Werten variiert zwischen den Fragebogenskalen zu den mathematikbezogenen Motivationen und Einstellungen (Tab. 2) zwischen 22 und 38 %. Im sozioökonomischen Status gemessen am ESCS-Index fehlen für 10 % der Schülerinnen und Schüler Angaben, bei den Schulnoten schwanken die Anteile zwischen Deutsch und Mathematik (jeweils 2 %), Naturwissenschaften (12 %), Chemie (14 %), Physik (18 %) und Biologie (22 %). Vollständige Datensätze liegen für das Alter, Geschlecht und die mathematische Kompetenz (PVs) vor.

4.4 Statistisches Vorgehen (*Propensity Score Matching*)

Um die Wirksamkeit der pädagogischen Maßnahme Klassenwiederholung zu untersuchen, verbieten sich aus ethischen Gründen experimentelle Untersuchungen. Um aber doch kausale Effekte analysieren zu können, ist es wichtig, relevante Merkmale zu kontrollieren, die zur Vorhersage beitragen, ob eine Schülerin oder ein Schüler eine Jahrgangsstufe wiederholen muss. Um dies zu realisieren, wird im Folgenden auf das Verfahren des *Propensity Score Matching* (PSM) zurückgegriffen (vgl. Rosenbaum und Rubin 1983). Dabei wird für eine Treatmentgruppe eine Kontrollgruppe ermittelt, so dass beide Gruppen einander in relevanten Merkmalen möglichst ähnlich sind. Auf diesem Wege soll sichergestellt werden, dass die zentrale Annahme der kontrafaktischen Kausalanalyse, die konditionale Unabhängigkeit (*conditional independence assumption*, CIA), erfüllt wird. Dies bedeutet, dass nach der Kontrolle von Kovariaten die Verteilung der Analyseeinheiten über Treatment- und Kontrollgruppe möglichst zufällig ist.

Die Schätzung der Propensity Scores sowie das Matching wurden mithilfe des SPSS-Plugins PS Matching (Thoemmes 2014) durchgeführt, das auf das R-Paket „match-it“ zurückgreift. Für den Datensatz, der die Kovariaten für das Matching beinhaltet, wurde im Vorfeld eine Mehrebenen-Imputation in MPlus 6 (Muthén und Muthén 2010) durchgeführt. Die Güte des Matchings lässt sich anhand verschiedener Indikatoren überprüfen: Einerseits sollen sich die beiden Gruppen hinsichtlich der kontrollierten Variablen nach dem Matching möglichst nicht mehr statistisch nachweisbar unterscheiden. Andererseits sollte der *Standardisierte Bias* (SB), der Unterschied in den Mittelwerten der Kovariaten unter Berücksichtigung der Varianz dieser Merkmale, auf einen Wert von deutlich unter 20 bis 25 (vgl. Harder et al. 2010; Rosenbaum und Rubin 1985) reduziert sein. Weiterhin sollte sich der vorhandene Bias nach dem Matching (*Percent Bias Reduction*, PBR) idealerweise um 80 % oder mehr reduziert haben (vgl. Pan und Bai 2015). Darüber hinaus muss die nicht-parametrische Identifikation des Treatmenteffekts durch eine Überlappung der Propensity Scores (*Common Support*) der beiden Gruppen garantiert werden (vgl. Rubin 2001). Letzteres lässt sich u. a. mithilfe eines globalen χ^2 -Tests zur Prüfung der Stichprobenbalance (vgl. Hansen und Bowers 2008) prüfen. Dabei sollte der χ^2 -Test nach dem Matching idealerweise nicht signifikant ausfallen.

5 Ergebnisse

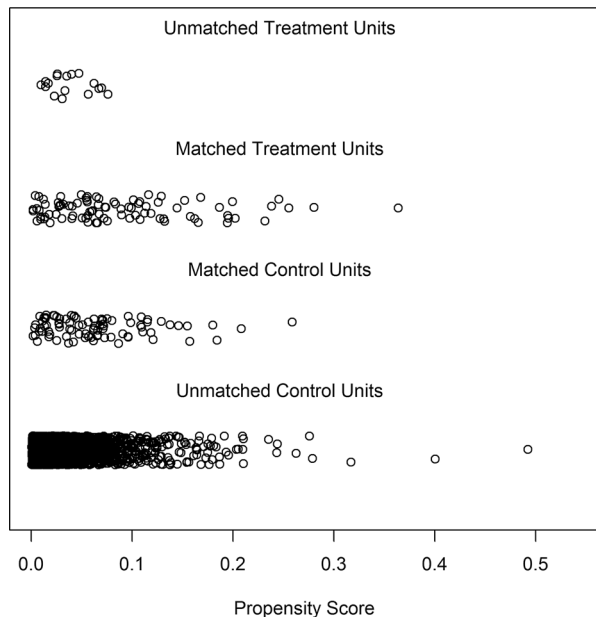
5.1 Ergebnisse des *Propensity Score Matching*

Um den Effekt der Klassenwiederholung bestmöglich zu isolieren, wurde für diejenigen 106 Schülerinnen und Schüler, die im Datensatz als Klassenwiederholerinnen und -wiederholer ausgewiesen waren, mithilfe eines *Propensity Score Matchings* eine Kontrollgruppe synthetisiert. Eingesetzt wurde hierfür ein *Nearest-Neighbor-Matching* mit einer 1-zu-1-Zuordnung ohne Caliper und ohne Zurücklegen. Die binär kodierte Variable *Klassenwiederholer* diente als Selektionsvariable. Als Kovariaten wurden im Matching-Modell das *Alter* sowie das *Geschlecht* der Schülerinnen und Schüler, ihr *ökonomischer, sozialer und kultureller Status* (ESCS), ihr *mathematisches Selbstkonzept*, ihre *Schulnoten* in den Fächern Deutsch, Mathematik, Biologie, Chemie, Physik und Naturwissenschaften sowie ihre *mathematische Ausgangsleistung* zum ersten Messzeitpunkt kontrolliert. Darüber hinaus wurde die Schule, die die einzelnen Schülerinnen und Schüler besuchen als restringierendes, exaktes Matchingkriterium gewählt. So wurden auf diesem Wege mögliche Effekte der spezifischen Einzelschulen im Modell kontrolliert.

Insgesamt lagen vor dem Matching Daten von $n = 3287$ Schülerinnen und Schülern vor. Hiervon waren $n = 106$ als Klassenwiederholend ausgewiesen; $n = 3181$ hatten hingegen zwischen den beiden Messzeitpunkten die Klassen nicht wiederholt.

Wie Tab. 3 zeigt, unterschieden sich die beiden Gruppen vor dem Matching statistisch nachweisbar ($p < 0,05$) mit Blick auf das mathematische Selbstkonzept, einen Großteil der Schulnoten sowie die mathematische Ausgangsleistung zum ersten Messzeitpunkt. Im Rahmen des PSM konnte nun für 89 der Klassenwiederhole-

Abb. 1 Verteilung der Propensity Scores (Logits) für die Klassenwiederholerinnen und -wiederholer (Matched Treatment Units, $N = 89$) und die KWH-Vergleichsgruppe (Matched Control Units, $N = 89$)



Tab. 3 Soziodemographische Merkmale in der Gruppe der Klassenwiederholerinnen und -wiederholer (KWH) und in der Vergleichsgruppe (KWH-Vergleichsgruppe) vor und nach dem Matching

	Vor Matching		Nach Matching		Signifikanz
	KWH (N = 106)	Gesamtstichprobe (N = 3181)	KWH (N = 89)	KWH-Vergleichsgruppe (N = 89)	
Alter (M)	15,38	15,42	15,36	15,41	$p > 0,10$
Geschlecht, weiblich (in %)	52,80	50,80	50,60	51,70	$p > 0,10$
Geschlecht, männlich (in %)	47,20	49,20	49,40	48,30	
ESCS (M)	0,22	0,30	0,19	0,35	$p > 0,10$
Mathematisches Selbstkonzept (M)	-0,22	0,86	-0,25	-0,46	$p > 0,10$
Note Deutsch (M)	2,58	1,95	2,69	2,52	$p > 0,10$
Note Mathematik (M)	2,75	2,11	2,87	2,47	$p < 0,05$
Note Biologie (M)	2,17	1,80	2,27	2,03	$p > 0,10$
Note Chemie (M)	2,42	1,88	2,46	2,26	$p > 0,10$
Note Physik (M)	2,42	1,90	2,43	2,33	$p > 0,10$
Note Naturwissenschaft (M)	1,79	1,78	1,84	1,78	$p > 0,10$
Mathematische Kompetenz t1 (M)	-0,36	0,30	-0,30	-0,13	$p > 0,10$
Mittlerer Standardisierter Bias (in %)	47,5		16,8		

rinnen und -wiederholer ein Matchingpartner gefunden werden, der trotz gleichwertiger, individueller Voraussetzungen in den kontrollierten Kovariaten, in eine zehnte Klasse an derselben Schule versetzt wurde. Der einzig statistisch nachweisbar verbleibende Unterschied betrifft die Schulnote im Fach Mathematik. Selbst nach dem Matching bleibt hier ein leichter Unterschied zu Gunsten der regulär versetzten Schülerinnen und Schüler nachweisbar ($\Delta M = -0,40$, $p < 0,05$).

Dass im vorliegenden Datensatz an den einzelnen Schulen vergleichbare Schülerinnen und Schüler für die Klassenwiederholenden gefunden werden konnten, macht auch der grafische Vergleich deutlich, der in Abb. 1 wiedergegeben ist. Die Verteilungen überlappen sich in weiten Teilen, so dass davon auszugehen ist, dass nahezu die gesamte Gruppe der Klassenwiederholenden in die *Area of Common Support* fällt; es gelingt somit für nahezu alle Fälle der Treatmentgruppe Fälle in der Kontrollgruppe zu finden, die in den Kovariaten vergleichbar sind. Der mittlere standardisierte Bias (SB) konnte dabei um 65 % von 47,5 auf 16,8 und damit deutlich unter die kritische Schwelle von 20 gesenkt werden, was angesichts der Tatsache, dass ein exaktes Matching auf Ebene der Einzelschule durchgeführt wurde, akzeptabel erscheint. Auch der globale χ^2 -Tests, der mit $p > 0,10$ statistisch nicht signifikant ausfällt, unterstützt die Annahme, dass durch das durchgeführte PSM eine gut ausbalancierte Vergleichsgruppe regulär versetzter Schülerinnen und Schüler für die Klassenwiederholerinnen und -wiederholer der neunten Jahrgangsstufe zu finden ist.

5.2 Ergebnisse zur Wirkung von Klassenwiederholung auf die mathematische Kompetenz

Um die Effekte von Klassenwiederholungen auf die mathematische Kompetenz und auf mathematikbezogene Motivationen und Einstellungen zu untersuchen, wurden im Folgenden Regressionanalysen unter Berücksichtigung der geclusterten Datenstruktur mit der Statistiksoftware MPlus 6 (Muthen und Muthen 2010) berechnet. Dabei wurden als Prädiktoren die Gruppenzugehörigkeit (regulär versetzt: KWH = 0 und Klassenwiederholung: KWH = 1), der Messzeitpunkt (t_1 : MZP = 0 und t_2 : MZP = 1) und die Interaktion zwischen Gruppenzugehörigkeit und Messzeitpunkt (KWH \times MZP) einbezogen. Als abhängige Variablen wurden die mathematische Kompetenz und die mathematikbezogenen Motivations- und Einstellungsskalen berücksichtigt. Die Ergebnisse der Regressionsanalysen sind in Tab. 4 zusammengefasst. Um die Gruppenmittelwerte (Tab. 5) anschaulich miteinander vergleichen zu können, sind diese in den Abb. 2, 3, 4, 5, 6 und 7 zusammen mit den Mittelwerten der Gesamtpopulation abgebildet.

Die Ergebnisse hinsichtlich der mathematischen Kompetenzentwicklung sind in Tab. 4 und Abb. 2 dargestellt. Der Regressionskoeffizient für den Messzeitpunkt ist auf dem 5-Prozentniveau signifikant, das heißt, sowohl die Klassenwiederholerinnen und -wiederholer als auch die regulär versetzten Schülerinnen und Schüler in der Vergleichsgruppe können sich zwischen dem ersten und zweiten Messzeitpunkt in ihrer mathematischen Kompetenz signifikant verbessern. Eine signifikante Interaktion zwischen Messzeitpunkt und Gruppenzugehörigkeit lässt sich nicht nachweisen.

Die grafische Darstellung der Mittelwerte in Abb. 2 veranschaulicht, dass beim ersten Messzeitpunkt die Klassenwiederholerinnen und -wiederholer und die Ver-

Tab. 4 Achsenabschnitte und unstandardisierte Regressionskoeffizienten (B) der Regressionsanalysen für mathematische Kompetenz und mathematikbezogene Motivations- und Einstellungsskalen

	Intercept/B	SE	beta	t-Wert	p
<i>Mathematische Kompetenz</i>					
Intercept	-0,18	0,21	-	-	-
KWH	-0,17	0,10	-0,07	-1,66	0,10
MZP	0,46	0,11	0,19	4,20	$p < 0,001$
KWH \times MZP	-0,04	0,17	-0,02	-0,26	0,80
<i>Interesse an Mathematik</i>					
Intercept	1,94	0,09	-	-	-
KWH	-0,17	0,13	-0,12	-1,30	0,20
MZP	-0,04	0,15	-0,03	-0,28	0,78
KWH \times MZP	0,17	0,22	0,10	0,77	0,45
<i>Instrumentelle Motivation in Mathematik</i>					
Intercept	2,66	0,07	-	-	-
KWH	-0,38	0,14	-0,24	-2,75	$p < 0,01$
MZP	-0,13	0,11	-0,08	-1,17	0,24
KWH \times MZP	0,29	0,23	0,16	1,29	0,20
<i>Arbeitseinstellung in Mathematik</i>					
Intercept	2,66	0,07	-	-	-
KWH	-0,28	0,09	-0,22	-2,94	$p < 0,01$
MZP	-0,07	0,10	-0,06	-0,70	0,48
KWH \times MZP	0,39	0,16	0,27	2,41	$p < 0,05$
<i>Mathematikbezogenes Verhalten</i>					
Intercept	1,48	0,05	-	-	-
KWH	0,03	0,07	0,04	0,49	0,62
MZP	-0,08	0,05	-0,09	-1,66	0,10
KWH \times MZP	0,12	0,10	0,12	1,15	0,25
<i>Mathematikbez. Selbstwirksamkeitserw.</i>					
Intercept	3,18	0,07	-	-	-
KWH	-0,18	0,08	-0,17	-2,28	$p < 0,05$
MZP	0,10	0,07	0,09	1,34	0,18
KWH \times MZP	0,03	0,13	0,02	0,20	0,84

In Fettdruck hervorgehobene Koeffizienten sind statistisch signifikant

gleichsgruppe erwartungskonform im Vergleich zur Gesamtpopulation deutlich unterdurchschnittliche Kompetenzwerte in der mathematischen Kompetenz aufweisen. Ein Skalenwert von 0 entspricht dem Mittelwert der Gesamtpopulation (inklusive der Hauptschulen). Der Lernzuwachs im Verlaufe eines Schuljahres fällt dabei für alle drei Gruppen positiv aus. Die Ergebnisse zeigen, dass sich durch die Klassenwiederholung der Abstand zur Gesamtstichprobe der regulär versetzten Schülerinnen und Schüler nicht verringert.

Tab. 5 Mittelwerte und Standardabweichungen für die Gesamtstichprobe, die KWH-Vergleichsgruppe und die Gruppe der Klassenwiederholerinnen und -wiederholer

	t ₁			t ₂		
	<i>M</i>	<i>SE</i>	<i>SD</i>	<i>M</i>	<i>SE</i>	<i>SD</i>
<i>Mathematische Kompetenz</i>						
Gesamtstichprobe	0,24	0,10	10,22	0,59	0,09	10,29
KWH-Vergleichsgruppe (versetzt)	-0,18	0,21	10,02	0,28	0,23	10,12
Klassenwiederholer/innen	-0,35	0,26	10,14	0,06	0,27	10,34
<i>Interesse an Mathematik</i>						
Gesamtstichprobe	2,15	0,02	0,79	2,12	0,02	0,80
KWH-Vergleichsgruppe (versetzt)	1,91	0,09	0,73	1,89	0,12	0,76
Klassenwiederholer/innen	1,75	0,10	0,65	1,94	0,10	0,80
<i>Instrumentelle Motivation in Mathematik</i>						
Gesamtstichprobe	2,78	0,02	0,77	2,75	0,02	0,76
KWH-Vergleichsgruppe (versetzt)	2,64	0,07	0,81	2,52	0,08	0,69
Klassenwiederholer/innen	2,26	0,11	0,84	2,46	0,12	0,88
<i>Arbeitseinstellung in Mathematik</i>						
Gesamtstichprobe	2,84	0,02	0,58	2,83	0,02	0,60
KWH-Vergleichsgruppe (versetzt)	2,65	0,07	0,58	2,58	0,09	0,61
Klassenwiederholer/innen	2,39	0,08	0,59	2,71	0,09	0,68
<i>Mathematikbezogenes Verhalten</i>						
Gesamtstichprobe	1,54	0,01	0,40	1,53	0,01	0,42
KWH-Vergleichsgruppe (versetzt)	1,48	0,05	0,35	1,40	0,04	0,34
Klassenwiederholer/innen	1,52	0,04	0,41	1,57	0,09	0,61
<i>Mathematikbezogene Selbstwirksamkeitserwartung</i>						
Gesamtstichprobe	3,23	0,03	0,53	3,34	0,03	0,54
KWH-Vergleichsgruppe (versetzt)	3,14	0,07	0,51	3,27	0,09	0,48
Klassenwiederholer/innen	2,98	0,09	0,49	3,17	0,14	0,65

Geringfügige Abweichungen zu den Kennwerten in der Tab. 4 entstehen durch die Schätzung der fehlenden Werte in den Regressionsmodellen

5.3 Ergebnisse zur Wirkung von Klassenwiederholung auf mathematikbezogene Motivation und Einstellungen

Die Ergebnisse der Regressionsanalyse für das Interesse an Mathematik zeigen keine signifikanten Vorhersageeffekte. In Abb. 3 sind die gruppenspezifischen Mittelwerte abgetragen. Der Anstieg im mathematischen Interesse für die Gruppe der Klassenwiederholerinnen und -wiederholer lässt sich statistisch jedoch nicht absichern. Das im Vergleich zur Gesamtpopulation etwas geringere Interesse an Mathematik für die Klassenwiederholerinnen und -wiederholer und für die Vergleichsgruppe ist erwartungskonform.

Hinsichtlich der instrumentellen Motivation in Mathematik lässt sich ein signifikanter Unterschied zwischen Klassenwiederholerinnen und -wiederholern und der KWH-Vergleichsgruppe zum ersten Messzeitpunkt, also vor der Klassenwiederholung, feststellen. Die grafische Veranschaulichung der Gruppenmittelwerte in Abb. 3 suggeriert zwar eine gegenläufige Tendenz zwischen den Gruppen in den Entwick-

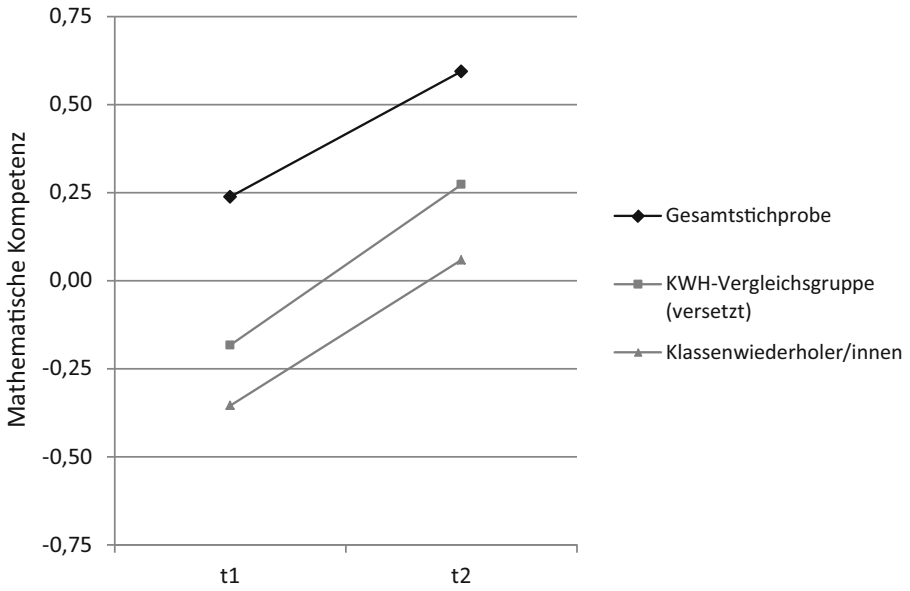


Abb. 2 Same-age-Vergleich der mathematischen Kompetenz (y-Achse) für die Gesamtstichprobe, die Gruppe der Klassenwiederholerinnen und -wiederholer und die Vergleichsgruppe

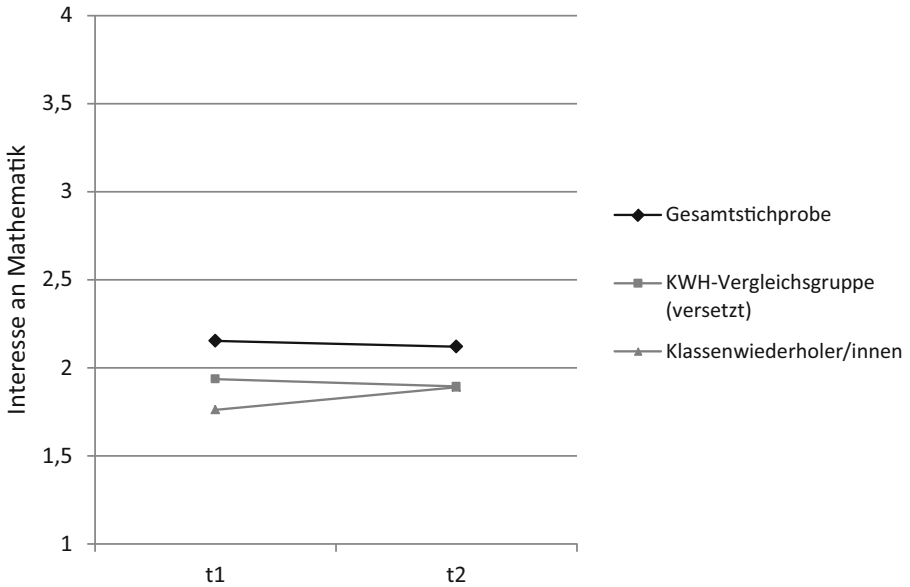


Abb. 3 Same-age-Vergleich des Interesses an Mathematik für die Gesamtstichprobe, die Gruppe der Klassenwiederholerinnen und -wiederholer und die KWH-Vergleichsgruppe

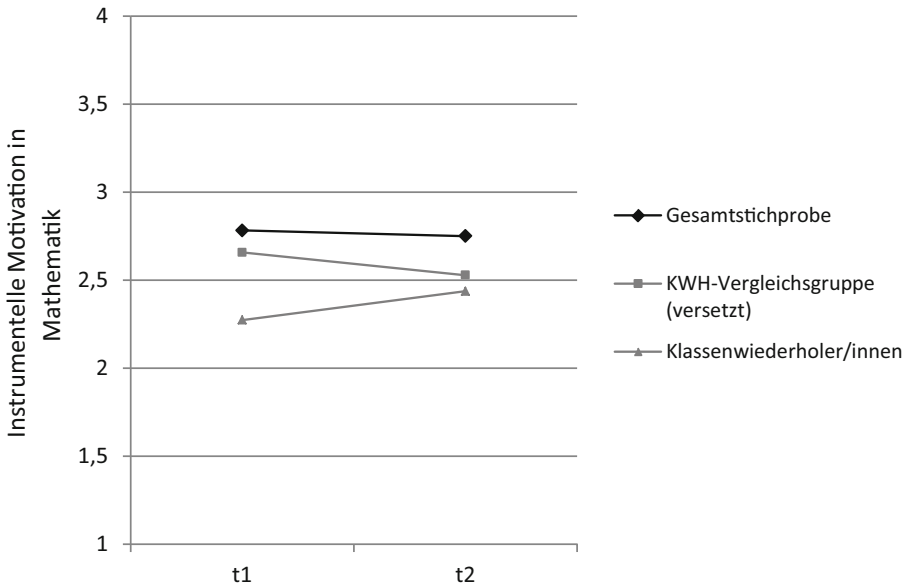


Abb. 4 Same-age-Vergleich der instrumentellen Motivation in Mathematik für die Gesamtstichprobe, die Gruppe der Klassenwiederholerinnen -wiederholer und die KWH-Vergleichsgruppe

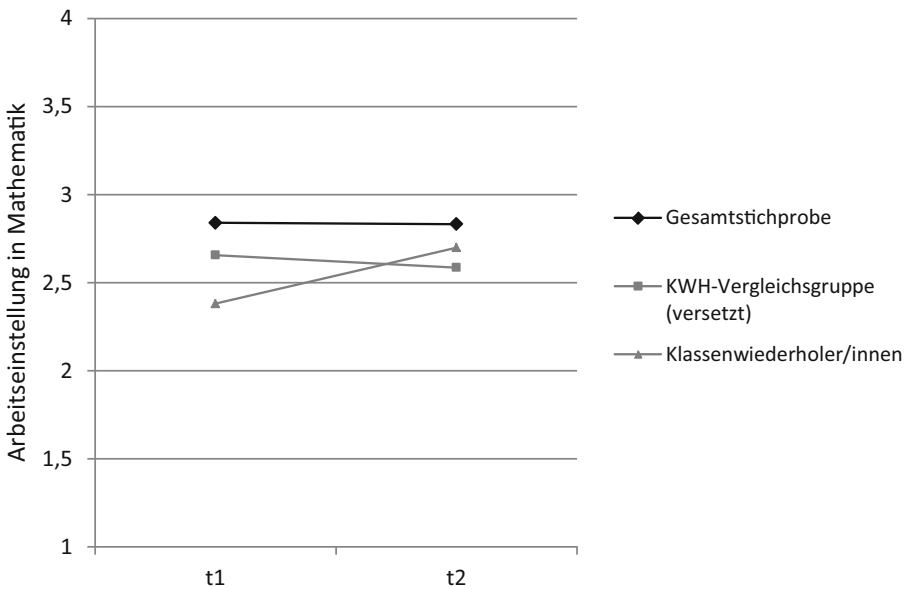


Abb. 5 Same-age-Vergleich der Arbeitseinstellung in Mathematik für die Gesamtstichprobe, die Gruppe der Klassenwiederholerinnen und -wiederholer und die KWH-Vergleichsgruppe

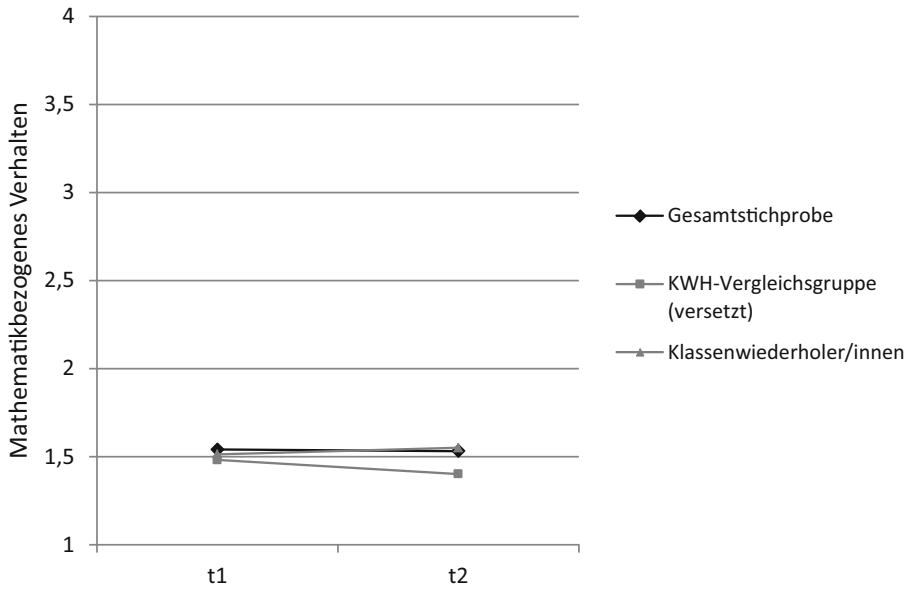


Abb. 6 Same-age-Vergleich des mathematikbezogenen Verhaltens für die Gesamtstichprobe, die Gruppe der Klassenwiederholerinnen und -wiederholer und die KWH-Vergleichsgruppe

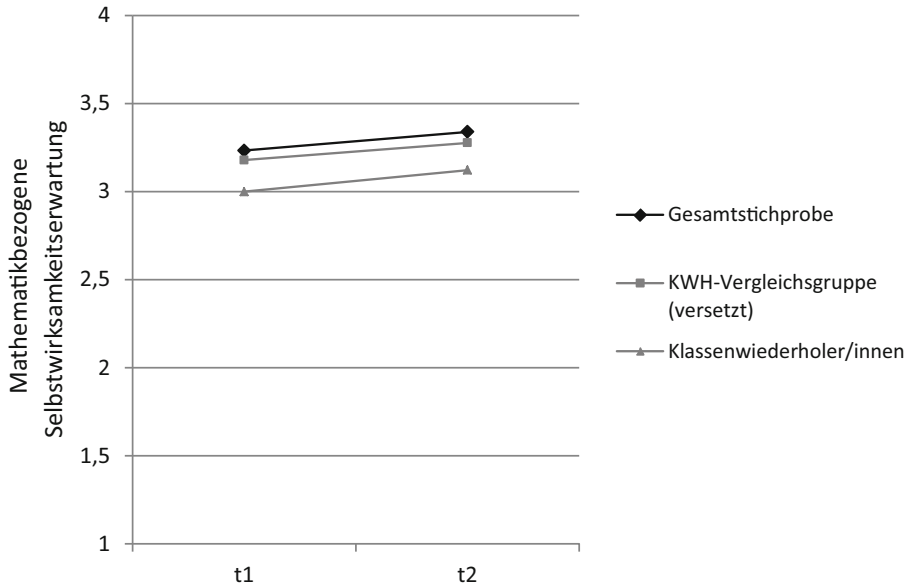


Abb. 7 Same-age-Vergleich der mathematikbezogenen Selbstwirksamkeitserwartung für die Gesamtstichproben, die Gruppe der Klassenwiederholerinnen und -wiederholer und die KWH-Vergleichsgruppe

lungen über die Zeit. Die Regressionskoeffizienten für den Messzeitpunkt und den Interaktionsterm in Tab. 4 sind jedoch nicht statistisch signifikant. Für die instrumentelle Motivation in Mathematik lassen sich demnach in dieser Studie keine Effekte durch die Klassenwiederholung nachweisen.

Hinsichtlich der Arbeitseinstellung in Mathematik bestehen zum ersten Zeitpunkt statistisch signifikante Unterschiede zwischen der KWH-Vergleichsgruppe und den Klassenwiederholerinnen und -wiederholern (Tab. 4 und Abb. 5). Der ebenfalls statistisch signifikante Regressionskoeffizient für den Interaktionsterm zeigt, dass sich zwischen beiden Gruppen eine unterschiedliche Entwicklung über die Zeit nachweisen lässt. Die Arbeitseinstellung in Mathematik verbessert sich für die Klassenwiederholerinnen und -wiederholer im Wiederholungsjahr, während sie in der KWH-Vergleichsgruppe und in der Gesamtstichprobe gleich hoch bleibt.

Hinsichtlich des mathematikbezogenen Verhaltens lassen sich keine statistisch signifikanten Vorhersageeffekte der Gruppen, des Messzeitpunktes oder der Interaktion zwischen beiden feststellen (Tab. 4 und Abb. 6). Auch zur Gesamtstichprobe gibt es keine bedeutsamen Unterschiede. Inhaltlich berichten die Jugendlichen aller Gruppen relativ übereinstimmend, dass sie zwischen selten und nie mit Freunden über Mathematik sprechen oder mehr als zwei Stunden für Mathematik üben.

Für die mathematikbezogene Selbstwirksamkeitserwartung lässt sich ein statistisch signifikanter Effekt für die Gruppe der Klassenwiederholerinnen und -wiederholer feststellen (Tab. 4 und Abb. 7). Diese zeigen eine bedeutsam geringere mathematikbezogene Selbstwirksamkeit beispielweise darin, mathematische Abbildungen in Zeitungen zu verstehen. Für den Messzeitpunkt und den Interaktionsterm lassen sich keine bedeutsamen Effekte nachweisen. Der in Abb. 7 tendenziell erkennbare Anstieg in der Selbstwirksamkeitserwartung bleibt im statistischen Schwankungsbereich.

6 Diskussion

In dieser Studie wurde die Wirksamkeit der pädagogischen Maßnahme Klassenwiederholung hinsichtlich der Entwicklung der mathematischen Kompetenz und der mathematikbezogenen Motivationen und Einstellungen bei Schülerinnen und Schülern untersucht, die die 9. Jahrgangsstufe ein zweites Mal besucht haben. Um eine adäquate Vergleichsgruppe zu bilden, die der Gruppe der Klassenwiederholerinnen und -wiederholer in relevanten Prädiktorvariablen möglichst ähnlich ist, wurde ein *Propensity Score Matching* (PSM) vorgenommen. Die Ergebnisse des PSM zeigen, dass hier erfolgreich ein systematischer Bias zwischen der Gruppe der nicht versetzten Schülerinnen und Schüler und der KWH-Vergleichsgruppe reduziert werden konnte.

Mit unserer ersten Hypothese haben wir vermutet, dass leistungsschwache Jugendliche, die eine Klassenstufe wiederholen, sich in ihrer mathematischen Kompetenz stärker verbessern als vergleichbar leistungsschwache Schülerinnen und Schüler, die regulär versetzt werden. Diese Hypothese konnte durch unsere Analyse nicht bestätigt werden. Die Ergebnisse zeigen, dass sowohl die regulär versetzte KWH-Vergleichsgruppe als auch die Klassenwiederholerinnen und -wiederholer

einen signifikanten Lernzuwachs aufweisen. Die theoretische Annahme, dass leistungsschwache Schülerinnen und Schüler über zu wenig Vorwissen verfügen, um vom Unterricht der nächsthöheren Klassenstufe zu profitieren, konnte demnach nicht bestätigt werden.

Mit Blick auf das *same-age*-Design dieser Studie muss festgehalten werden, dass designbedingt möglicherweise ein positiver Bias zugunsten der regulär versetzten Schülerinnen und Schüler besteht, da diese ja einen curricular anspruchsvolleren Mathematikunterricht besuchen. Allerdings sind die Items im verwendeten IQB-Ländervergleichstest für Mathematik für die 9. Jahrgangsstufe so konzipiert, dass dieser designbedingte Bias vermutlich nicht bedeutsam ins Gewicht fällt. Es ist allerdings davon auszugehen, dass der Lernzuwachs für regulär versetzte Schülerinnen und Schüler höher ausfallen würde, wenn der Mathematiktest für den Mathematikunterricht der 10. Jahrgangsstufe sensitiv wäre.

Die zweite Hypothese bezog sich auf die Veränderung hinsichtlich der mathematikbezogenen Motivation und Einstellungen der Schülerinnen und Schüler. Die Vermutung war, dass sich diese bei den Klassenwiederholerinnen und -wiederholern tendenziell positiv entwickeln und bei den regulär versetzten Schülerinnen und Schülern unverändert bleiben.

Die Ergebnisse der Analysen zeigen, dass sich die Mittelwerte für das Interesse an Mathematik, für die instrumentelle Motivation, für das mathematikbezogene Verhalten sowie für die mathematikbezogene Selbstwirksamkeit nicht signifikant erhöhen. Bezüglich der Arbeitseinstellung in Mathematik kann hingegen festgehalten werden, dass sich diese bei den Klassenwiederholerinnen und -wiederholern im Vergleich zu den regulär versetzten Schülerinnen und Schülern statistisch bedeutsam erhöht hat. Dies spricht zumindest bezüglich der Arbeitseinstellung in Mathematik für die theoretische Annahme, dass durch die Klassenwiederholung und den damit verbundenen Wechsel in eine leistungsschwächere Lerngruppe eine günstige Motivations- und Einstellungsänderung verbunden sein kann (Bless et al. 2004; Ehmke et al. 2010). Für vier der fünf Motivations- und Einstellungsmaße kann jedoch keine Veränderung gegenüber der Vergleichsgruppe festgestellt werden.

Die Befunde der vorliegenden Studie sind hinsichtlich der fachlichen Kompetenzentwicklung vergleichbar mit der ersten Replikationsstudie (Ehmke et al. 2010) und den Ergebnissen von Lamote et al. (2014). In beiden Studien konnte ein Jahr nach der Wiederholung kein Effekt auf die mathematische Kompetenz bzw. auf die Sprachleistung festgestellt werden. Der positive Effekt der Klassenwiederholung auf das mathematikbezogene Selbstkonzept, der in Ehmke et al. (2010) gefunden wurde, konnte in dieser Replikationsstudie nicht mehr festgestellt werden. Dies steht im Widerspruch zu den Befunden von Lamote et al. (2014), die im Schuljahr nach der Wiederholung zumindest kurzfristig ein höheres fachbezogenes Selbstkonzept festgestellt hatten. Teilweise ist die Diskrepanz zwischen unseren beiden Replikationsstudien dadurch zu erklären, dass hierfür unterschiedliche Skalen verwendet wurden. Während in unserer ersten Studie das mathematikbezogene Selbstkonzept erhoben wurde, ist das hier verwendete Maß für mathematikbezogene Selbstwirksamkeit weniger anfällig für Referenzgruppeneffekte. Der Unterschied zur Studie von Lamote et al. ist möglicherweise durch die unterschiedlichen untersuchten Fächer zu erklären.

Einschränkend ist für die vorliegende Studie festzuhalten, dass nur kurzfristige Effekte untersucht werden können. So kann nicht überprüft werden, inwieweit der positive Effekt auf die Arbeitseinstellung auch dauerhaft bestehen bleibt. Zahlreiche Befunde (z. B. Jimerson 2001a, 2001b; Klapproth et al. 2016) belegen, dass Klassenwiederholungen kurzfristig durchaus eine Verbesserung der motivationalen Situation der Betroffenen bewirken können; diese Effekte, die sich im ersten Jahr nach der Wiederholung zeigen, verlieren sich mittelfristig und haben langfristig negative Effekte auf die sozioemotionalen Einstellungen.

Ein interessanter Befund zeigte sich bei der Durchführung des PSM. Für $N = 89$ Klassenwiederholerinnen und -wiederholer konnte jeweils ein Matchingpartner gefunden werden, der trotz gleichartiger, individueller Voraussetzungen in den kontrollierten Kovariaten in eine zehnte Klasse an **derselben** Schule versetzt wurde. Dies wirft die Frage auf, inwieweit es noch weitere Merkmale gibt, die diese Jugendlichen voneinander unterscheiden bzw. welche Entscheidungsprozesse an den Schulen letztlich dazu führen, dass einige Neuntklässlerinnen und Neuntklässler eine Klassenstufe wiederholen müssen und andere nicht, **obwohl** sie sich in versetzungsrelevanten Merkmalen wie Zeugnisnoten, aber auch in Selbstkonzepten und soziodemographischen Hintergrundmerkmalen sehr ähnlich sind.

Weiterhin bleibt einschränkend festzuhalten, dass das emotionale Befinden der Schülerinnen und Schüler nicht explizit untersucht werden konnte. Möglicherweise gibt es neben den fachbezogenen Motivationen und den Einstellungen durch die Klassenwiederholung Auswirkungen auf das emotionale Befinden, die nicht mit der Motivation und den Einstellungen einhergehen.

Auch ist die Stichprobe dieser Studie zwar repräsentativ für die 9. Jahrgangsstufe. Hier stehen etwa die Schülerinnen und Schüler an den Realschulen am Ende ihrer Schulzeit und sind möglicherweise noch einmal „besonders“ motiviert, einen erfolgreichen Übergang in das Berufsleben zu vollziehen. Die Situation für Schülerinnen und Schüler an Gymnasien ist jedoch zu diesem Zeitpunkt weniger von einem anstehenden Übergang geprägt, was die untersuchte Stichprobe in unserer Studie vermutlich ebenfalls prägt. Weitere Studien sollten daher verstärkt auch jüngere Altersgruppen in den Blick nehmen und dabei auch mögliche Langzeitwirkungen an verschiedenen Schulformen überprüfen. Dazu wären mehr als zwei Messzeitpunkte wünschenswert.

Unter Berücksichtigung individueller und gesamtgesellschaftlicher Kosten-Nutzen-Relationen lässt sich auf der Grundlage unserer Befunde, keine positive Empfehlung für die pädagogische Maßnahme Klassenwiederholung aussprechen. Vor diesem Hintergrund und mit Blick auf die individuellen Kosten (Verlängerung der Schulzeit) und gesellschaftlichen Kosten (z. B. mehr Lehrerstunden, die bezahlt werden müssen, vgl. Klemm 2009) ist diese Maßnahme nur in Einzelfällen zu empfehlen. Anstelle der Wiederholung aller Unterrichtsfächer mit gleichen Inhalten in einem Schuljahr, scheint eher eine gezielte, begleitende individuelle Förderung der leistungsschwachen Schülerinnen und Schüler in den betreffenden Fächern angezeigt. Insbesondere der durch das PSM verdeutlichte Befund, dass innerhalb derselben Schule offenbar zahlreichere und andere als die bisher untersuchten Faktoren darüber entscheiden, ob ein Jugendlicher in die nächsthöhere Klasse versetzt wird

oder nicht, setzt am Ende unserer Untersuchung einen wichtigen Anknüpfungspunkt für anschließende Studien.

Literatur

- Alexander, K., Entwisle, D., & Dauber, S. (2003). *On the success of failure: A reassessment of the effects of retention in the early grades*. Cambridge: Cambridge University Press.
- Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, 20(10), 215–229.
- Bellenberg, G., Hovestadt, G., & Klemm, K. (2005). *Selektivität und Durchlässigkeit im allgemein bildenden Schulsystem*. Frankfurt: Arbeitsgruppe Bildungsforschung/Bildungsplanung.
- Bless, G., Schüpbach, M., & Bonvin P. (2004). *Klassenwiederholung – Determinanten, Wirkungen und Konsequenzen*. Bern u.a.: Haupt-Verlag
- Boller, S., Möller, M., & Palowski, M. (2013). Wiederholen in der gymnasialen Oberstufe – wissenschaftliche Befunde und pädagogische Unterstützungsmöglichkeiten. In D. Bosse, F. Eberle & B. Schneider-Taylor (Hrsg.), *Standardisierung in der gymnasialen Oberstufe* (S. 175–188). Wiesbaden: SpringerVS.
- Demanet, J., & van Houtte, M. (2013). Grade retention and its association with school misconduct in adolescence: A multilevel approach. *School Effectiveness and School Improvement*, 24(4), 417–434. doi:10.1080/09243453.2012.727834.
- Demski, D., & Liegmann, A. B. (2014). Klassenwiederholungen im Kontext von Schul- und Berufsbiographien. In A. B. Liegmann (Hrsg.), *Facetten von Übergängen im Bildungssystem. Nationale und internationale Ergebnisse empirischer Forschung*. Münster: Waxmann.
- Drechsel, B., & Senkbeil, M. (2004). Institutionelle und organisatorische Rahmenbedingungen von Schule und Unterricht. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (S. 284–291). Münster: Waxmann.
- Ehmke, T. & Siegle, T. (2005). ISEI, ISCED, HOMEPOS, ESCS. Indikatoren der sozialen Herkunft bei der Quantifizierung von sozialen Disparitäten. *Zeitschrift für Erziehungswissenschaft*, 8(4), 521–540.
- Ehmke, T., Drechsel, B., & Carstensen, C.H. (2008). Klassenwiederholen in PISA-I-Plus: Was lernen Sitzbleiber in Mathematik dazu? *Zeitschrift für Erziehungswissenschaft*, 11(3), 368–387. doi:10.1007/s11618-008-0033-3.
- Ehmke, T., Drechsel, B., & Carstensen, C.H. (2010). Effects of grade retention on achievement and self-concept in science and mathematics. *Studies in Educational Evaluation*, 36, 27–35.
- Ellsworth, A., & Lagacé-Séguin, D.G. (2009). In retrospect: is youth grade retention associated with self-esteem and self-efficacy in early adulthood? *International Journal of Adolescence and Youth*, 15(1–2), 21–38. doi:10.1080/02673843.2009.9748019.
- Fruehwirth, J. C., Navarro, S., & Takahashi, Y. (2011). How The Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects. *Human Capital and Economic Opportunity: A Global Working Group. Working Paper No. 2011–015*.
- García-Pérez, J.I., Hidalgo-Hidalgo, M., & Robles-Zurita, J.A. (2014). Does grade retention affect students' achievement? Some evidence from Spain. *Applied Economics*, 46(12), 1373–1392. doi:10.1080/00036846.2013.872761.
- Goos, M., Schreiber, B., Knipprath, H., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How can cross-country differences in the practice of grade retention be explained? A closer look at national educational policy factors. *Comparative Education Review*, 57(1), 54–84.
- Hansen, B.B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23, 219–236.
- Harder, V.S., Stuart, E.A., & Anthony, J.C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, 15(3), 234–249.
- Holmes, C.T. (1989). Grade level retention effects: A meta-analysis of research studies. In L.A. Shepard & M.L. Smith (Hrsg.), *Flunking grades: research and policies on retention* (S. 16–33). Philadelphia: Taylor & Francis.
- Hong, G., & Raudenbush, S.W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, 27(3), 205–224.

- Huebener, M., & Marcus, J. (2015). Auswirkungen der G8-Schulzeitverkürzung: Erhöhte Zahl von Klassenwiederholungen, aber jüngere und weniger Abiturienten. DIW Wochenbericht Nr. 18. https://www.diw.de/documents/publikationen/73/diw_01.c.502825.de/15-18-1.pdf. Zugegriffen: 03.05.2017.
- Jackson, G. B. (1975). The research evidence on the effects of grade retention. *Review of Educational Research*, 45(4), 613–635.
- Jimerson, S. R. (2001a). Meta-Analysis of Grade Retention Research: Implications for Practice in the 21st Century. *School Psychology Review*, 30(3), 420–437.
- Jimerson, S. R. (2001b). A synthesis of grade retention research: looking backward and moving forward. *The California School Psychologist*, 6, 47–59.
- Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. *Learning and Individual Differences*, 50, 182–194.
- Klemm, K. (2009). *Klassenwiederholungen – teuer und unwirksam. Eine Studie zu den Ausgaben für Klassenwiederholungen in Deutschland*. Gütersloh: Bertelsmann.
- Krohne, J. A., Meier, U., & Tillmann, K.-J. (2011). Sitzenbleiben, Geschlecht und Migration – Klassenwiederholungen im Spiegel der PISA-Daten. *Zeitschrift für Pädagogik*, 50(3), 373–391.
- Lamote, C., Pinxten, M., van Den Noortgate, W., & van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies*, 40(5), 496–514. doi:10.1080/03055698.2014.936828.
- Lorence, J. (2006). Retention and academic achievement research revisited from a united states perspective. *International Education Journal*, 7(5), 731–777.
- Martin, A. J. (2011). Holding back and holding behind: grade retention and students' non-academic and academic outcomes. *British Educational Research Journal*, 37(5), 739–763. doi:10.1080/01411926.2010.490874.
- McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology*, 37(3), 272–298.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus Software (Version 6)*. Los Angeles CA: Muthén & Muthén.
- Nagy, G., Haag, N., Lüdtke, O., Köller, O. (2017). Längsschnittskalierung der Tests zur Überprüfung des Erreichens der Bildungsstandards der Sekundarstufe I im PISA-Längsschnitt 2012/2013. doi:10.1007/s11618-017-0755-1
- OECD (2010). *What makes a school successful? PISA 2009 Results*. Paris: OECD Publishing.
- OECD (2011). *When students repeat grades or are transferred out of school: what does it mean for education systems? PISA in Focus No. 2011/6*. Paris: OECD Publishing.
- OECD (2013). *PISA 2012 results: what makes schools successful? Resources, policies and practices*. Bd. IV. Paris: OECD Publishing.
- Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development and Psychopathology*, 13(2), 297–315.
- Pan, W., & Bai, H. (2015). *Propensity score analysis: fundamentals and developments*. New York: Guilford Press.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T., & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Plummer, D. L., & Graziano, W. G. (1987). Impact of grade retention on the social development of elementary school children. *Developmental Psychology*, 23(2), 267–275.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (Hrsg.). (2006). *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (Hrsg.). (2013). *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.
- Reynolds, A. J. (1992). Grade retention and school adjustment: An exploratory analysis. *Educational Evaluation and Policy Analysis*, 14(2), 101–121.
- Rosenbaum, P., & Rubin, D. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rosenbaum, P., & Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33–38.
- Roßbach, H.-G., & Tietze, W. (2010). Sitzenbleiben. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 781–788). Weinheim: Beltz.

- Rubin, D. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2, 169–188.
- Sälzer, C., Prenzel, M., & Klieme, E. (2013). Schulische Rahmenbedingungen der Kompetenzentwicklung. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012: Fortschritte und Herausforderungen in Deutschland* (S. 155–187). Münster: Waxmann.
- Sälzer, C., Prenzel, M., Schiepe-Tiska, A., & Hammann, M. (2016). Schulische Rahmenbedingungen der Kompetenzentwicklung. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015: Eine Studie zwischen Kontinuität und Innovation* (S. 176–218). Münster: Waxmann.
- Thoemmes, F. (2014). Propensity score matching in SPSS. <http://sourceforge.net/projects/psmspss/>. Zugegriffen: 03.05.2017.
- Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of Grade Retention in First Grade on Psychosocial Outcomes. *Journal of Educational Psychology*, 102(1), 135–152.