

Veränderungen der Lesekompetenz von der 9. zur 10. Klasse: Differenzielle Entwicklungen in Abhängigkeit der Schulform, des Geschlechts und des soziodemografischen Hintergrunds?

Gabriel Nagy · Jan Retelsdorf · Frank Goldhammer ·
Anja Schiepe-Tiska · Oliver Lüdtke

Online publiziert: 9. Mai 2017
© Springer Fachmedien Wiesbaden 2017

Zusammenfassung Im vorliegenden Beitrag wurde die Entwicklung der Lesekompetenz im letzten Abschnitt der Sekundarstufe I (Klassenstufen 9 bis 10) untersucht. Neben der Veränderung der Testleistungen in der Gesamtpopulation wurden die Assoziationen ausgewählter institutioneller (Schulform), familiärer (Zuwanderungshintergrund und sozioökonomischer familiärer Status) und individueller Merkmale (Geschlecht) mit der Leistungsentwicklung erfasst. In Übereinstimmung mit aktuellen Studien, die eine abflachende Entwicklung der Leseleistung in späteren Phasen der Beschulung zeigen, konnten wir keinen Leistungszuwachs in der Gesamtstichprobe feststellen. Ebenso fanden sich keine belastbaren Hinweise dafür, dass die betrachteten Erklärungsvariablen mit der Kompetenzentwicklung assoziiert sind. Die Auswertungen lieferten jedoch deutliche Indizien dafür, dass die schülerseitige

Prof. Dr. G. Nagy (✉) · Prof. Dr. O. Lüdtke
Pädagogisch-Psychologische Methodenlehre, Leibniz-Institut für die Pädagogik der
Naturwissenschaften und Mathematik, Olshausenstraße 62, 24118 Kiel, Deutschland
E-Mail: nagy@ipn.uni-kiel.de

Prof. Dr. O. Lüdtke
E-Mail: oluedtke@ipn.uni-kiel.de

Prof. Dr. J. Retelsdorf
Erziehungswissenschaft, Leibniz-Institut für die Pädagogik der Naturwissenschaften und
Mathematik, Olshausenstraße 62, 24118 Kiel, Deutschland
E-Mail: jretelsdorf@ipn.uni-kiel.de

Prof. Dr. F. Goldhammer
Bildungsqualität und Evaluation, Deutsches Institut für Internationale Pädagogische Forschung,
Solmsstraße 73–75, 60486 Frankfurt am Main, Deutschland
E-Mail: goldhammer@dipf.de

Dr. A. Schiepe-Tiska
TUM School of Education, Zentrum für internationale Vergleichsstudien, Technische
Universität München, Arcisstraße 21, 80333 München, Deutschland
E-Mail: schiepe-tiska@tum.de

Persistenz der Testbearbeitung, die mittels Positionseffekten erfasst wurde, sich systematisch in Abhängigkeit der Hintergrundvariablen veränderte, wobei stärkere Abnahmen der Bearbeitungspersistenz an nichtgymnasialen Schulformen, bei Jungen, und Schülerinnen und Schülern aus sozioökonomisch schlechter gestellten Familien festgestellt wurden. Die Nichtberücksichtigung des Testbearbeitungsverhaltens führte zu „Scheineffekten“ der Erklärungsvariablen Schulform und Geschlecht, die sich bei einer genaueren Betrachtung jedoch nicht auf den realen Kompetenzzuwachs bezogen.

Schlüsselwörter Lesekompetenz · Kompetenzentwicklung · Sekundarstufe I · Positionseffekte · Testmotivation

Changes in reading skills from 9th to 10th grade: differential trajectories depending on school type, gender and socio-demographic background?

Abstract This paper examines the development of proficiency in reading from grades 9 to 10. In addition to the proficiency development in the overall population, relationships of proficiency gains with institutional (school type), familial (immigration background and socioeconomic family status) and individual characteristics (gender) were examined. In accordance with current results according to which the development of reading proficiency slows down in later stages of schooling, we found no meaningful proficiency increase in the overall sample. Furthermore, the explanatory variables showed no reliable relationships with proficiency growth. However, the analyses provided strong evidence that students' persistence of working thoroughly on the test indicated by position effects changed across measurement occasions. Furthermore, we found changes in position effects to be related to the explanatory variables considered in this paper. Stronger decreases in persistence were detected in male students, students with less favorable socio-economic background, and in students from nonacademic tracks. Non-consideration of test persistence led to biased estimates of effects of the explanatory variables school type and gender, which upon closer inspection were not related to real proficiency growth.

Keywords Reading comprehension · Competence development · Lower secondary education · Position effect · Test motivation

1 Einleitung

Lesen als Fähigkeit geschriebene Texte zu verstehen und zu nutzen gilt als eine Schlüsselkompetenz moderner Gesellschaften (Rychen und Salganik 2003). Dabei ist Lesekompetenz nicht nur fächerübergreifend wichtig für eine erfolgreiche Schullaufbahn (Härtig et al. 2015), sondern darüber hinaus wesentlich für die Teilhabe an gesellschaftlich-kultureller Praxis. So gilt Lesen nach Barker und Escarpit (1973) als effektives Mittel für die Aneignung, Organisation und Anwendung von Wissen. Die Bedeutung dieser Kulturtechnik als unverzichtbare Kompetenz im Be-

rufsalltag zeigt sich auch darin, dass Lesen zu einer der am häufigsten genannten Arbeitsanforderungen über verschiedenste Berufsfelder gehört (Klaukien et al. 2013). Dabei ist es wichtig zu erwähnen, dass – obwohl die grundlegende Vermittlung der Lesekompetenz im Grundschulalter erfolgt – die Förderung der Lesekompetenz auch zu den zentralen Aufgaben von Schule in der Sekundarstufe gehört (vgl. Artelt et al. 2001). Die Relevanz von Leseförderung auch jenseits der Grundschule wird durch die in Deutschland wiederholt gefundenen hohen Anteile sehr leistungsschwacher Schülerinnen und Schüler in den PISA-Studien (zusammenfassend Hohn et al. 2013) sowie vergleichbare Anteile im Erwachsenenalter (Zabal et al. 2013) deutlich.

Es gibt inzwischen einige Studien, die sich mit Lesekompetenz bzw. spezifischeren Lesefähigkeiten in der Sekundarstufe I beschäftigen. Längsschnitte zur Modellierung von Leistungszuwächsen bleiben in dieser Altersklasse aber eine Ausnahme. Zudem fokussierten die vorhandenen Studien eher den Zeitraum zu Beginn der Sekundarstufe I – Entwicklungen nach der sechsten Klassenstufe wurden unseres Wissens bislang kaum untersucht (für Ausnahmen siehe z. B. Cameron et al. 2015; Gailberger und Willenberg 2008). So ist weitgehend ungeklärt, ob die Lesekompetenzentwicklung gegen Ende der Sekundarstufe I stagniert oder ob noch substantielle Lernzuwächse zu verzeichnen sind. Zudem stellt sich die Frage, ob sich auch in dieser Altersgruppe Leistungsunterschiede zwischen Schülerinnen und Schülern in Abhängigkeit von Schulform, Geschlecht, Zuwanderungshintergrund sowie soziokulturellem Hintergrund noch weiter vergrößern, ob sie stabil sind, oder ob sie sogar geringer werden. Bisher existieren auch zur Relevanz dieser Prädiktoren in der hier relevanten Altersgruppe nur wenige Arbeiten (siehe für Ausnahmen z. B. Cameron et al. 2015; Gailberger und Willenberg 2008; Nikolova 2011; Robinson und Lubienski 2011). Im vorliegenden Beitrag wurde die Entwicklung der Lesekompetenz im letzten Abschnitt der Sekundarstufe I (Klassenstufe 9 bis 10) an einer großen Stichprobe im Rahmen des PISA-Längsschnitts 2012/2013 untersucht, der eine Messwiederholung nach einem Schuljahr umfasst.

Im Fokus des Beitrags stehen sowohl inhaltliche Fragestellungen zur Kompetenzentwicklung (s. oben), als auch methodische Fragestellungen, die sich Einflüssen von längsschnittlichen Änderungen der Persistenz der Testbearbeitung widmen (z. B. deMars 2007), die mittels (Item-)Positionseffekten (Leary und Dorans 1985) operationalisiert wurden. Konkret gehen wir der Frage nach inwieweit die untersuchten Determinanten der Kompetenzentwicklung mit der Veränderung der Persistenz der Bearbeitung der Leistungstests assoziiert sind und somit zu „Scheineffekten“ hinsichtlich deren Vorhersageleistung für die Kompetenzentwicklung führen können (vgl. Nagy et al. 2017).

In den nachfolgenden Abschnitten fassen wir den aktuellen Stand der Forschung zur Kompetenzentwicklung im Bereich Lesen in der Sekundarstufe I zusammen. In diesem Zusammenhang diskutieren wir auch eine über die Zeit veränderliche Persistenz der Testbearbeitung als eine mögliche Störvariable, die sich auf die Ergebnisse von Längsschnittanalysen zur Kompetenzentwicklung auswirken kann. Anschließend präzisieren wir die Fragestellungen des Beitrags, präsentieren die Datengrundlage der Auswertungen und stellen ein längsschnittliches Analysemodell vor, dass die Erfassung der Effekte der untersuchten Kovariaten auf die Veränderung der Persistenz der Testbearbeitung erlaubt (vgl. Nagy et al. 2017; Nagy et al.

2016). Das Modell basiert auf der Item Response Theorie (IRT) und basiert auf der Annahme, dass sich eine geringe Persistenz der Testbearbeitung in einer über den Testverlauf graduellen Abnahme der Lösungswahrscheinlichkeit der Testitems manifestiert, die mit sogenannten Positionseffekten kompatibel ist (Leary und Dorans 1985).

2 Entwicklung von Lesekompetenzen in der Sekundarstufe I

Die grundlegende Entwicklung von Lesefähigkeiten findet im Vor- und Grundschulalter statt. Dabei zeigt sich etwa in dem Modell von Lundberg (2002), dass diese Entwicklung eng mit dem allgemeinen Spracherwerb verbunden ist. Dass diese Entwicklung aber nicht mit Eintritt in die Sekundarstufe I abgeschlossen ist, wird deutlich, wenn man die Ergebnisse für Deutschland aus den PISA-Studien mit denen im Grundschulalter aus PIRLS (Progress in International Reading Literacy Study) vergleicht. Wenngleich dies mit einigen methodischen Herausforderungen verbunden ist, haben Artelt et al. (2008) einen vorsichtigen Vergleich der Ergebnisse beider Studien vorgenommen. Die vergleichsweise gute Ausgangsposition deutscher Schülerinnen und Schüler im Lesen, die in PIRLS gefunden wurde, konnte bei PISA nicht bestätigt werden. Entsprechend folgern Artelt et al., dass die Förderung der Lesekompetenz in der Sekundarstufe nicht optimal funktioniert.

Wie sich die Lesekompetenz im Verlauf der Sekundarstufe I – vor allem jenseits der sechsten Klassenstufe – entwickelt ist allerdings nach wie vor eher selten untersucht worden. In den wenigen längsschnittlichen Studien, die Erkenntnisse hierzu liefern, ist die Befundlage zudem recht heterogen. So zeigten sich etwa in den LAU- (Aspekte der Lernausgangslage und der Lernentwicklung; Lehmann et al. 2004) oder KESS-Studien (Kompetenzen und Einstellungen von Schülerinnen und Schülern; Nikolova 2011) in deskriptiven Analysen jeweils moderate bis hohe Zuwächse im Lesen, wobei die Zuwächse von Klasse 9 nach 11 größer waren als die Zuwächse von Klasse 7 nach 8 bzw. Klasse 8 nach 10. In anderen Arbeiten zeigte sich dagegen, dass Leistungszuwächse im Lesen im Verlauf der Sekundarstufe in höheren Klassenstufen eher geringer wurden (Bloom et al. 2008; Cameron et al. 2015; Retelsdorf et al. 2011) oder dass die Leseleistung zunächst anstieg, sich am Ende der Sekundarstufe aber sogar rückläufig entwickelte (van Gelderen et al. 2007). Auch in der Messwiederholung im Rahmen der DESI-Studie (Deutsch Englisch Schülerleistungen International) zeigten sich nur sehr kleine Leistungszuwächse vom Anfang bis zum Ende der 9. Klassenstufe (Gailberger und Willenberg 2008).

Zusammenfassend zeigte sich in bisherigen Studien also ein eher heterogenes Bild der Entwicklung der Lesekompetenz in der Sekundarstufe. In der Gesamtschau überwiegen aber Studien, die für das Ende der Sekundarstufe nur noch geringe Zuwächse oder eine Stagnation der Entwicklung der Lesekompetenz erwarten lassen. Neben dieser durchschnittlichen Entwicklung stellt sich die Frage nach differenziellen Entwicklungsverläufen in Abhängigkeit von institutionellen, familiären und individuellen Merkmalen. In diesem Beitrag haben wir uns auf die Merkmale Schulform, Geschlecht, Zuwanderungshintergrund und soziokultureller Hintergrund konzentriert, da diese häufig im Fokus von querschnittlich angelegten Large-Scale-

Studien stehen (z. B. OECD 2016), jedoch bis heute vergleichsweise selten im Kontext von Längsschnittstudien zur Entwicklung der Leseleistung am Ende der Sekundarstufe I betrachtet wurden. Von diesen Merkmalen ist bekannt, dass sie in enger Beziehung mit dem Stand der Leseleistungen stehen und zentral für eine Beschreibung leistungsschwacher Schülerinnen und Schüler in PISA sind (OECD 2016). Somit stellt sich die Frage nach deren Zusammenhang mit der Leistungsentwicklung in der letzten Phase der Sekundarstufe I. Dem Forschungsstand zur Bedeutung dieser Merkmale für die Lesekompetenz widmen wir uns im folgenden Abschnitt.

3 Differenzielle Entwicklungen in Abhängigkeit von Schulform, Geschlecht und Merkmalen des soziodemografischen Hintergrunds

Effekte der Schulform auf die Kompetenzentwicklung von Schülerinnen und Schülern wurden bereits häufig untersucht. Dabei zeigte sich für Mathematik das recht einheitliche Bild größer werdender Unterschiede zwischen Schülerinnen und Schülern an Gymnasien und anderen Schulformen zugunsten der Gymnasien (im Überblick s. Becker 2009). Im Bereich der Lesekompetenz ist die Befundlage zu solchen Schereneffekten deutlich heterogener. Hier zeigten sich unterschiedliche Ergebnisse zwischen Studien, aber auch innerhalb einzelner Studien in Abhängigkeit vom jeweils konkret betrachteten Aspekt der Lesekompetenz. So zeigte sich etwa bei Retelsdorf et al. (2012) im Verlauf von der fünften zur achten Klasse ein Schereneffekt für Dekodiergeschwindigkeit, allerdings nicht für das Textverstehen. Pfost et al. (2010) fanden andererseits einen Schereneffekt von der fünften bis zur sechsten Klasse im Leseverständnis, aber nicht im Wortschatz, während sich vor dem Übergang auf die weiterführenden Schulen in beiden Bereichen eine positivere Leistungsentwicklung der späteren Gymnasiastinnen und Gymnasiasten zeigte. In den bereits erwähnten Studien LAU und KESS wurden insgesamt inkonsistente Ergebnisse für Schereneffekte in der Lesekompetenz berichtet (Lehmann et al. 2004; Nikolova 2011). Während in LAU zum Teil sogar größere Leistungszuwächse an Nicht-Gymnasien gefunden wurden, zeigten sich in KESS eher parallele Entwicklungsverläufe für Gymnasien, Real- und Gesamtschulen – lediglich die Hauptschulen fielen leicht zurück. Zusammenfassend bleibt die Befundlage zu Schereneffekten im Lesen uneindeutig.

Bezogen auf Geschlechterunterschiede im Lesen schlussfolgerte Hyde (2014) in ihrem aktuellen Überblicksbeitrag, dass der Unterschied zwischen Jungen und Mädchen im Lesen eher klein ist. Im Gegensatz dazu findet sich in einer aktuellen Auswertung von PISA-Daten bei Reilly (2012) ein deutlicher Vorteil von Mädchen gegenüber Jungen über alle Länder hinweg ($d = 0,44$). Neben diesen widersprüchlichen Aussagen über Niveauunterschiede im Lesen finden sich in den existierenden längsschnittlichen Analysen zu Geschlechterunterschieden in der Sekundarstufe neben Differenzen im Ausgangswert allenfalls Hinweise auf kleine Effekte im Verlauf zugunsten der Mädchen (z. B. Gailberger und Willenberg 2008; Retelsdorf und Köller 2014; Retelsdorf et al. 2011). Diese Ergebnisse passen gut zu der Studie von Cameron et al. (2015), die einen größer werdenden Unterschied zugunsten der Mädchen etwa bis zur dritten oder vierten Klasse feststellten – im weiteren Verlauf

bis zur achten Klasse blieb dieser weitgehend stabil. Robinson und Lubiencki (2011) fanden im Mittel eher stabile Geschlechterunterschiede im Lesen vom Kindergarten bis zur achten Klasse. Bei Betrachtung der leistungsschwachen Schülerinnen und Schüler zeigte sich im Verlauf allerdings ein kleiner bis mittlerer Schereneffekt zugunsten der Mädchen. Schließlich zeigten sich in einer Studie von De Fraine et al. (2007), die ebenfalls nichtlineare Entwicklungsverläufe sprachlicher Fähigkeiten in Abhängigkeit des Geschlechts betrachteten, stetige Zuwächse für die Mädchen von der siebten bis zur 12. Klasse, während für die Jungen etwa bis zur achten Klasse zunächst ein Leistungsrückgang zu verzeichnen war, bevor sie in Klasse 12 fast das Niveau der Mädchen erreichten.

Niedrigere Werte in der Lesekompetenz wurden auch für Schülerinnen und Schüler mit Zuwanderungshintergrund im Vergleich zu denen ohne Zuwanderungshintergrund wiederholt festgestellt (z. B. Lesaux und Kieffer 2010; Marx und Stanat 2012). Allerdings zeigten sich auch hier in der Arbeit von Cameron et al. (2015) zunächst größer werdende Unterschiede bis zur vierten oder fünften Klasse, bevor diese bis zur achten Klasse etwa stabil blieben. Auch hierzu passen weitere längsschnittliche Studien, die Zusammenhänge des Zuwanderungshintergrunds mit dem Ausgangsniveau, nicht aber mit der weiteren Entwicklung fanden (z. B. Kigel et al. 2015; Retelsdorf und Köller 2014; Retelsdorf et al. 2011). Baumert et al. (2012) fanden sogar einen Kompensationseffekt für Schülerinnen und Schüler mit Zuwanderungshintergrund, der zur Folge hatte, dass die Unterschiede im Verlauf von der vierten zur sechsten Klasse geringer wurden.

Schließlich zeigte sich bei verschiedenen Aspekten des soziokulturellen Hintergrunds ein ähnliches Bild wie beim Geschlecht und dem Zuwanderungshintergrund. In der Grundschule lassen sich zunehmende Unterschiede zugunsten von Kindern aus besser gestellten Familien in der Lesekompetenz beobachten, die je nach Untersuchung bereits zu Beginn der Grundschulzeit (Aikens und Barbarin 2008) oder später ab der fünften Klasse (Cameron et al. 2015) stabil blieben. Entsprechend zeigten sich in längsschnittlichen Studien in der Sekundarstufe vor allem Zusammenhänge des soziokulturellen Hintergrunds mit dem Ausgangsniveau und keine oder nur sehr kleine Effekte auf die längsschnittliche Kompetenzentwicklung (z. B. Kigel et al. 2015; Retelsdorf und Köller 2014; Retelsdorf et al. 2011).

Zusammenfassend ist die Befundlage bei allen hier berücksichtigten Prädiktoren der Entwicklung der Lesekompetenz uneinheitlich. Es scheint aber in der Tendenz zu gelten, dass Unterschiede zwischen den jeweils relevanten Gruppen vor allem zu Beginn der Grundschulzeit größer werden und dass diese Unterschiede dann im weiteren Verlauf der Schulzeit auf einem stabilen Niveau bleiben.

4 Veränderung der Persistenz der Testbearbeitung im Längsschnitt als potentielle Störquelle

Zusätzlich zu den Ergebnissen früherer Untersuchungen zu differenziellen Entwicklungen der Lesekompetenz stellt sich die Frage, ob etwaige Unterschiede auf Aspekte der Testbearbeitung (z. B. Persistenz der Testbearbeitung) zurückzuführen sind. So erklären z. B. De Fraine et al. (2007) den oben berichteten zeitweiligen Rückgang

sprachlicher Fähigkeiten bei Jungen mit einer möglichen niedrigen Testmotivation der Jungen in der achten Klasse. Diese Argumentation erscheint im Lichte der in diesem Band berichteten Befunde nicht unplausibel. Nagy et al. 2017 berichteten Testkontexteffekte, die in Übereinstimmung mit Effekten der Positionen der Items in den im PISA-Längsschnitt 2012/2013 eingesetzten Tests ausfielen. Die dort berichteten Ergebnisse dokumentieren einen Anstieg der Positionseffekte vom ersten zum zweiten Messzeitpunkt, wobei dieses Muster für den Bereich Lesen besonders stark ausgeprägt war.

(Item-)Positionseffekte führen in der Regel dazu, dass die Lösungswahrscheinlichkeit eines Items mit zunehmender Nähe zum Ende des Tests sinkt (z. B. Meyers et al. 2009). Diese Effekte werden als Indikatoren einer suboptimalen Persistenz der Testbearbeitung diskutiert und können somit als eine Facette der Motivation der Testbearbeitung verstanden werden (Debeer et al. 2014). Dementsprechend ist die Stärke von Positionseffekten prinzipiell variabel und kann sich somit zwischen Gruppen und/oder Messzeitpunkten unterscheiden (Debeer et al. 2014; DeMars 2007; Hartig und Buchholz 2012).

Wie Nagy et al. (2017) gezeigt haben, können Positionseffekte und deren längsschnittliche Veränderung mit individuellen und kontextuellen Merkmalen interagieren. In solchen Situationen kann die Nichtberücksichtigung von Positionseffekten zu falschen Schlussfolgerungen über individuelle und kontextuelle Prädiktoren der Kompetenzveränderung führen. Eine Abnahme der Persistenz der Testbearbeitung über die betrachteten Messzeitpunkte impliziert eine Abnahme der mittleren Lösungswahrscheinlichkeit der bearbeiteten Testitems. Insofern die Abnahme der Lösungswahrscheinlichkeiten, die sich infolge von Positionseffekten (d. h. der geringer werdenden Persistenz der Testbearbeitung) ergibt, nicht kontrolliert wird, wird diese irrtümlich einer Abnahme des Kompetenzniveaus zugeschrieben (vgl., Nagy et al. 2016). Ändert sich die Persistenz der Testbearbeitung im Längsschnitt in Abhängigkeit einer Kovariate (z. B. stärkere längsschnittliche Abnahme bei männlichen Schülern), könnte dieser Zusammenhang irrtümlich einem Effekt der Kovariate auf die Kompetenzentwicklung zugeschrieben werden, sofern die entsprechenden Zusammenhänge der Kovariate mit Positionseffekten unberücksichtigt bleiben. Nagy et al. (2017) haben diesen Effekt am Beispiel von Schulformunterschieden illustriert. Im Prinzip können sich jedoch derartige Verzerrungen für jede der in diesem Beitrag betrachteten Erklärungsvariablen ergeben. Aus diesem Grund haben wir im vorliegenden Beitrag die Effekte der betrachteten Kovariaten auf die Persistenz der Testbearbeitung (Operationalisiert durch Effekte der Position von Itemblöcken bzw. Itemclustern) explizit berücksichtigt.

5 Die vorliegende Studie

Längsschnittliche Untersuchungen zur Entwicklung der Lesekompetenz am Ende der Sekundarstufe sind rar. Auch in der ersten Messwiederholung im Rahmen von PISA 2003 (PISA-I-Plus; Prenzel et al. 2006) wurde die Lesekompetenz nicht berücksichtigt. Vor diesem und dem eingangs referierten Hintergrund sollte in der vorliegenden Studie mehreren Forschungsfragen nachgegangen werden. Erstens wollten wir

prüfen, ob gegen Ende der Pflichtschulzeit überhaupt noch Zuwächse in der Lesekompetenz zu verzeichnen sind. Gerade vor dem Hintergrund der Arbeiten, die eine sich über die Schullaufbahn verlangsamende Kompetenzentwicklung zeigen (Bloom et al. 2008; Cameron et al. 2015), sind dabei nicht unbedingt weitere Zuwächse zu erwarten.

Zweitens sollten differenzielle Leistungsentwicklungen in Abhängigkeit von Schulform, Geschlecht, Zuwanderungshintergrund und soziokulturellem Hintergrund untersucht werden. Für alle vier Merkmale lassen sich aus der Literatur keine eindeutigen Erwartungen ableiten – erwartbar wären auf Grundlage früherer Studien am ehesten sich leicht öffnende Leistungsscheren oder parallele Entwicklungsverläufe. Allerdings sind auch kompensatorische Entwicklungen denkbar, bei denen eine vormals benachteiligte Gruppe zu den übrigen Schülerinnen und Schülern aufschließt (vgl. etwa Baumert et al. 2012).

Schließlich sollte drittens untersucht werden, inwieweit die hier untersuchten Prädiktoren der Leistungsentwicklung mit Positionseffekten und deren längsschnittlicher Veränderung interagieren. Diese Analysen sind auf zweierlei Weise bedeutsam. Zunächst liefern sie Hinweise für Assoziationen der Testbearbeitungspersistenz (operationalisiert durch Positionseffekte) und deren Veränderung mit den untersuchten Prädiktorvariablen. Insofern derartige Assoziationen vorliegen, werden Effekte der Hintergrundvariablen auf die Leistungsentwicklung unter Kontrolle der Bearbeitungspersistenz zudem realistischer geschätzt.

Die Berücksichtigung von Positionseffekten als Indikatoren der Bearbeitungspersistenz erscheint gerade im Kontext der Entwicklung der Leseleistung in der letzten Phase der Sekundarstufe I wichtig. Wie zuvor ausgeführt, lassen sich auf Ebene der Gesamtstichprobe eher kleine Leistungszuwächse erwarten. Geht man ferner davon aus, dass (reale) Kompetenzverluste unplausibel sind, folgen daraus geringe Effekte der Prädiktorvariablen auf die Leistungsentwicklung, da sich die mittleren Zuwächse einzelner Teilgruppen (z. B. Geschlechtergruppen) zu einem relativ geringen Gesamtwachstums addieren müssen. In solchen Datenkonstellationen können bereits relativ schwache Verzerrungen der Effektschätzungen zu falschen Rückschlüssen über die Leistungsentwicklung führen (vgl. Nagy und Neumann 2010). Hinzu kommt, dass die Ergebnisse von Nagy et al. (2017) darauf hindeuten, dass der Bereich Lesen im PISA-Längsschnitt 2012/2013 besonders stark von Positionseffekten betroffen ist, sodass sich diese besonders stark auf die Ergebnisse von Zuwachsschätzungen auswirken können.

6 Methode

6.1 Stichprobe

Grundlage der vorliegenden Auswertung ist eine Teilstichprobe des von Heine et al. (2017) beschriebenen Samples. Konkret haben wir uns auf diejenigen Schülerinnen und Schüler ($n = 4954$) konzentriert, die zumindest an einem der beiden Messzeitpunkte des PISA-Längsschnitts 2012/2013 Items zur Erfassung der Lesekompetenz bearbeitet haben. Das analysierte Sample kann als eine Zufallsauswahl der Aus-

gangsstichprobe ($N = 6584$) verstanden werden, da sie eine Funktion der bearbeiteten Testbooklets ist, auf die die Schülerinnen und Schüler zufällig verteilt wurden. Die Eingrenzung der Stichprobe ist eine Folge der Entscheidung die Leistungsentwicklung im Bereich Lesen mittels IRT-Modellen auszuwerten, die zumindest einen Teilmenge valider Itemantworten voraussetzen. Aufgrund der designbedingten „Zufallsauswahl“ des Analysesamples, ist nicht mit systematischen Verzerrungen der Befunde gegenüber der Zielpopulation zu rechnen.

Die vorliegende Datenkonstellation bietet im Zusammenhang mit IRT basierten Analyseverfahren die Möglichkeit auch Angaben von Schülerinnen und Schüler zu verwenden, die den Lesetest zu nur einem Messzeitpunkt bearbeitet haben. Fehlende Werte zum ersten Messzeitpunkt können als „missing completely at random“ (Rubin 2004) verstanden werden, da sie sich ausschließlich in Folge der zufallsbedingten Zuordnung von Schülerinnen und Schülern zu Testbooklets ergeben. Fehlende Werte zum ersten Messzeitpunkt wirken sich somit nicht systematisch auf die Abschätzung der Verteilung der Lesekompetenz in Klassenstufe 9 aus. Fehlende Werte zum zweiten Messzeitpunkt ergaben sich demgegenüber als eine Folge der zufälligen Zuordnung zu Testbooklets und des selektiven Stichprobenausfalls, wobei letzterer maßgeblich vom Kompetenzniveau in Klassenstufe 9 gesteuert wurde (vgl. Heine et al. 2017). Die zum zweiten Messzeitpunkt vorliegenden fehlenden Werte können sich somit auf die Abschätzung der Kompetenzverteilung des zweiten Messzeitpunkts und somit auf die Kompetenzveränderung auswirken, sofern die zentralen Determinanten des Ausfallsprozesses unberücksichtigt bleiben (Rubin 2004). Aus diesem Grund haben wir uns für eine mehrdimensionale IRT-Modellierung entschieden, die beide Messzeitpunkte berücksichtigt. Diese Analysestrategie wirkt Verzerrungen der Analyseergebnisse entgegen, die sich bei einer Ausblendung der Determinanten des selektiven Stichprobenausfalls (d. h. den Ausgangsniveaus der Kompetenzen) einstellen würden.

6.2 Instrumente

6.2.1 Lesekompetenztests

Das zur Erfassung der Leseleistung verwendete Testdesign ist im Detail im Beitrag von Nagy et al. (2017; Tab. 1) dargestellt. Die Leseleistung wurde in der ersten Erhebungswelle mittels dreier Itemcluster (L1, L2 und L3) erhoben, die 14 (L1 und L3) bzw. 15 Einzelitems (L2) umfassten. In der zweiten Welle wurden nur zwei der drei ursprünglichen Itemcluster eingesetzt (L2 und L3). Das zur ersten Welle verwendete Bookletdesign war hinsichtlich der Position der Itemcluster ausbalanciert. Die drei Itemcluster verteilten sich auf insgesamt neun von 13 Booklets, wovon sechs Booklets ein Itemcluster und drei Booklets zwei Itemcluster enthielten.

Die Balance der Itemclusterpositionen wurde zur zweiten Welle nur approximativ erfüllt (für Details vgl. Nagy et al. 2017). Insgesamt sieben von 18 Booklets umfassten Items zur Lesekompetenz und nur ein Booklet umfasste ein Itemcluster (L3) an der ersten Position. Diese Datenkonstellation hatte zur Folge, dass die hinsichtlich der ersten Itemclusterposition definierten Zuwächse mit einer vergleichsweise gerin-

gen Präzision erfasst wurden. Zum zweiten Messzeitpunkt umfassten sechs Booklets jeweils ein Itemcluster und ein Booklet zwei Itemcluster zum Bereich Lesen.

6.2.2 Hintergrundvariablen

Die Schulform wurde als dichotome Variable erfasst mit den Ausprägungen Gymnasium (46 %) und Nicht-Gymnasium (54 %). Dabei setzte sich die Gruppe der Nicht-Gymnasiasten zusammen aus Schülerinnen und Schülern an Realschulen, Schulen mit mehreren Bildungsgängen und integrierten Gesamtschulen. Sonderschulen, Hauptschulen oder anderen Schulen mit voraussichtlichem Hauptschulabschluss wurden aus den Analysen ausgeschlossen, da sie nicht Bestandteil der in der im PISA-Längsschnitt 2012/2013 avisierten Zielpopulation sind (vgl. Heine et al. 2017).

Der Zuwanderungshintergrund wurde für unsere Analysen ebenfalls dichotom erfasst. In Übereinstimmung mit den internationalen PISA-Berichten (z. B., OECD 2014) wurde Schülerinnen und Schülern dann ein Zuwanderungshintergrund (12 %) zugeschrieben, wenn beide Elternteile, oder sie selbst und mindestens ein Elternteil, im Ausland geboren wurden. Schülerinnen und Schüler mit nur einem im Ausland geborenen Elternteil wurden nicht zur Gruppe der Schülerschaft mit Zuwanderungshintergrund gezählt. Wir haben uns bewusst für eine dichotome Kategorisierung entschieden, da die verwendeten Messmodelle vergleichsweise hoch parametrisiert sind und jede weitere Kategorie zu einer Vielzahl zusätzlicher Parameter geführt hätte, die auf Grundlage des relativ geringen Datenumfangs (vgl. Tab. 1 in Nagy et al. 2017) nicht akkurat hätten geschätzt werden können.

Zur Erfassung des soziokulturellen Hintergrunds wurde der im Rahmen der PISA-Studie erhobene „International Socio-Economic Index of Occupational Status“ (ISEI; Ganzeboom et al. 1992) verwendet. Der ISEI ist ein in der internationalen Schulleistungsforschung etabliertes Maß des mit der Ausübung eines Berufs assoziierten sozioökonomischen Status. Der ISEI wurde getrennt aus den Angaben der ausgeübten Berufe der Mütter und Väter berechnet und die höchste Ausprägung wurde zur Erfassung des sozioökonomischen familiären Hintergrunds herangezogen (HISEI). Da der HISEI-Index einen weiten Wertebereich aufweist, wurde er zur besseren Interpretierbarkeit der Analysen in der Gesamtstichprobe z -standardisiert ($M = 0$; $SD = 1$).

Insgesamt lagen auf den Indikatoren des Zuwanderungshintergrunds und des sozioökonomischen familiären Hintergrunds eine Reihe von fehlenden Werten vor (vgl. Heine et al. 2017). Um den Ausschluss von Personen mit fehlenden Angaben auf einer dieser Variablen zu vermeiden, wurden die Variablen im Vorfeld der Analysen in der verwendeten Teilstichprobe imputiert. Konkret haben wir die Methode der multiplen Imputation (Rubin 2004) verwendet, wobei wir eine Imputationsmethode herangezogen haben, die die kategoriale Natur einzelner Variablen und darüber hinaus eine Reihe weiterer Hintergrundvariablen berücksichtigte (Schulform, Geschlecht, Ausgangsleistungen in den PISA-Tests). Wir haben 25 Ersetzungen der fehlenden Werte vorgenommen und die Ergebnisse der 25 Imputationen wurden entsprechend der Regeln von Rubin zu einem Gesamtergebnis integriert.

7 Statistische Analysen

Die Auswertung der Daten erfolgte mittels IRT-Modellen. Wir haben uns gegen die Verwendung von Plausible Values entschieden, da die Analysen für den Bereich Lesen Hinweise auf starke längsschnittliche Varianzzunahmen erbrachten, die möglicherweise Positionseffekten geschuldet sein könnten (vgl. Nagy et al. 2017). Das in diesem Beitrag verwendete Modell berücksichtigt mögliche Interaktionen zwischen den untersuchten Prädiktoren der Leistungsentwicklung und Bookleteffekten. Das Verfahren ermöglicht es somit festzustellen, inwieweit etwaige Effekte auf die Leistungsentwicklung letztlich auf längsschnittliche Veränderungen von Positionseffekten zurückzuführen sind.

Im vorliegenden Modell wurde der Logit der Wahrscheinlichkeit einer korrekten Antwort auf einem (dichotomen) Item einer Person i auf Item j des Itemclusters c in Booklet b_t zum Messzeitpunkt t , y_{ijcb_t} , dargestellt als

$$\text{Log} \left[\frac{P(y_{ijcb_t} = 1)}{P(y_{ijcb_t} = 0)} \right] = \theta_{it} + v_{ict} - \beta_j, \quad (1)$$

wobei β_j die Schwierigkeit des Items j (fixiert auf die Ergebnisse der Itemkalibrierung von Nagy et al. 2017) und θ_{it} die Fähigkeitsausprägung der Person i zum Zeitpunkt t indiziert. v_{ict} ist eine Platzhaltervariable, deren Ausprägung vollständig von externen Variablen (Booklet-Indikatoren und Interaktionen mit den Prädiktoren der Leistungsentwicklung) determiniert wird. Der im PISA-Längsschnitt 2012/2013 eingesetzte Lesetest setzte sich zum ersten Messzeitpunkt ($t = 1$) aus drei und zum zweiten Messzeitpunkt ($t = 2$) aus zwei Itemclustern zusammen, sodass zum ersten Messzeitpunkt drei (v_{11} , v_{21} und v_{31}) und zur zweiten Welle zwei Platzhaltervariablen (v_{22} und v_{32}) in das Modell aufgenommen wurden.

Positionseffekte wurden über Bookleteffekte approximiert (Nagy et al. 2017). Die Effekte wurden mittels dichotomer Bookletindikatoren ermittelt, wobei zu jedem Messzeitpunkt ein Booklet als Referenz diente. Die Fähigkeitsvariablen wurden mittels folgender Regressionsgleichung modelliert

$$\theta_{it} = \alpha_t + \delta_t x_i + \sum_{b_t=1}^{B_t-1} (\gamma_{0b_t} d_{ib_t} + \gamma_{1b_t} d_{ib_t} x_i) + \zeta_{it}, \quad (2)$$

während die Ausprägungen der Platzhaltervariablen wie folgt modelliert wurden

$$v_{ict} = \sum_{b_t=1}^{B_t-1} (\gamma_{0cb_t} d_{ib_t} + \gamma_{1cb_t} d_{ib_t} x_i). \quad (3)$$

In Gl. 2 steht α_t für den zeitspezifischen Regressionsintercept und δ_t indiziert das zeitspezifische Regressionsgewicht der Prädiktorvariablen x . ζ_{it} ist ein Abweichungsterm mit einem Mittelwert von 0 (Residuum). In den Gln. 2 und 3 stehen die Variablen d_{ib_t} für die Bookletindikatoren der Person i zum Messzeitpunkt t , wobei zu jedem Messzeitpunkt $B_t - 1$ Indikatoren gebildet wurden (mit b_t als Bookletvariable zum Messzeitpunkt t). Dementsprechend stehen die γ_{0b_t} - und γ_{0cb_t} -Parameter

für zeitspezifische Bookleteffekte auf die Fähigkeitsausprägungen und Platzhaltervariablen, während die Parameter $\gamma_{1b,t}$ und $\gamma_{1cb,t}$ die Interaktion zwischen Booklets und x erfassen.

Die in Gl. 2 ausgewiesenen Booklet- und Interaktionseffekte wurden frei geschätzt. Demgegenüber wurde der Großteil der in Gl. 3 ausgewiesenen Effekte auf die Platzhaltervariablen auf 0 restringiert. Diese Restriktionen ergaben sich aus Gründen der Modellidentifikation und stellten sicher, dass die ermittelten Bookleteffekte Auskunft über Positionseffekte liefern. Die konkrete Parametrisierung des Modells ist in Tab. 1 anhand der Zerlegung der querschnittlichen Effekte einer Kovariaten x auf die positions- und clusterspezifische Testleistung dargestellt. Die dort definierten Parameter ω_{pct} stehen für den Effekt einer Kovariaten x auf die Leistungen in Cluster c an Position p ($p = 1, 2, 3, 4$) zum Messzeitpunkt t .

Aus der dortigen Darstellung geht hervor, dass zum ersten Messzeitpunkt der Effekt der Kovariate x auf das Itemcluster L3, das in Booklet 3 an der ersten Position vorgelegt wurde, als Referenz zur Bestimmung des Effekts diente (d. h. $\omega_{1,3,1} = \delta_1$).¹ Die Effekte der Kovariate auf die Itemcluster L2 und L1 an der ersten Position wurden in Referenz dazu bestimmt, wobei die entsprechenden γ -Parameter den Unterschied des Effekts der Kovariate auf die Itemcluster relativ zum Itemcluster L3 darstellen (d. h. Interaktionseffekt der Kovariate mit Itemcluster L2 und L3). Die Effekte der Kovariate auf Itemcluster, die an späteren Positionen vorgelegt wurden (d. h. $p > 1$) wurden ebenso relativ zum Referenzeffekt δ_1 ausgedrückt. Beispielsweise indiziert der Effekt $\gamma_{1,3,2,1}$ den Unterschied des Effekts der Kovariate x auf das Itemcluster L3, das in Booklet 2 zum ersten Messzeitpunkt dargeboten wurde, relativ zum Effekt δ_1 .

Um die Schätzung des Modells zu beschleunigen wurden die Effekte der Kovariaten x in Booklets, die zwei Itemcluster zum Bereich Lesen umfassten, auf die zuletzt dargebotenen Cluster mittels eines bookletspezifischen Abweichungsterms dargestellt. Der Parameter $\gamma_{1c(b_t=h)t}^*$ ist definiert als die Differenz der Interaktionsterme des zuletzt in der Sequenz dargebotenen Clusters zum davor bearbeiteten Cluster in Booklet $b_t = h$. Beispielsweise entspricht der Interaktionseffekt $\gamma_{1,2,2,1}$ (der Effekt von x auf Cluster $c = 2$ in Booklet $b_t = 2$ zum Zeitpunkt $t = 1$; vgl. Gl. 2) in der vorgenommenen Parametrisierung (Tab. 1) der Summe $\gamma_{1,2,2,1} = \gamma_{1,3,2,1} + \gamma_{1,2,2,1}^*$.

Die so ermittelten ω -Parameter können über Itemcluster gemittelt werden, wobei $\omega_{p,t}$ dann für den über Itemcluster gemittelten Effekt von x zum Messzeitpunkt t an Position p steht. Der Vergleich der $\omega_{p,t}$ -Parameter ist insofern informativ, als dass er Hinweise für die Interaktion von Positionseffekten mit der Kovariate x liefert.

¹ Die Wahl des Referenzclusters und des Referenzbooklets ist aus statistischer Perspektive arbiträr. Wir haben uns zum ersten Messzeitpunkt für das Cluster L3 und das Booklet B03 aus pragmatischen Gründen entschieden, da das Booklet B03 nur ein Cluster zum Bereich Lesen umfasst (L3) und somit die Parametrisierung des Modells vereinfacht. Wir haben uns aus den gleichen Gründen zum zweiten Messzeitpunkt für das Cluster L2 und das Booklet B16 als Referenz entschieden.

Tab. 1 Darstellung der Zerlegung der Effekte einer Kovariate x auf die positions- und clusterspezifischen Testleistungen (ω_{pct})

	Position 1	Position 2	Position 3	Position 4
<i>Messzeitpunkt 1</i>				
Booklet 02 N3-L3- M7-L2	-	$\omega_{2,3,1} = \delta_1 + \gamma_{1,3,2,1}$	-	$\omega_{4,2,1} = \delta_1 + \gamma_{1,3,2,1} + \gamma_{1,2,2,1}^*$
Booklet 03 L3-M6- N1-M3	$\omega_{1,3,1} = \delta_1$	-	-	-
Booklet 04 M6-M7- L1-M4	-	-	$\omega_{3,1,1} = \delta_1 + \gamma_{1,1,4,1}$	-
Booklet 06 M1-M2- L2-M6	-	-	$\omega_{3,2,1} = \delta_1 + \gamma_{1,2,6,1}$	-
Booklet 08 N2-L2- M4-N1	-	$\omega_{2,2,1} = \delta_1 + \gamma_{1,2,8,1}$	-	-
Booklet 09 L2-M3- M5-L1	$\omega_{1,2,1} = \delta_1 + \gamma_{1,2,9,1}$	-	-	$\omega_{4,1,1} = \delta_1 + \gamma_{1,2,9,1} + \gamma_{1,1,9,1}^*$
Booklet 11 M4-M5- L3-M2	-	-	$\omega_{3,3,1} = \delta_1 + \gamma_{1,3,11,1}$	-
Booklet 12 N1-L1- M2-N3	-	$\omega_{2,1,1} = \delta_1 + \gamma_{1,1,12,1}$	-	-
Booklet 13 L1-M1- N2-L3	$\omega_{1,1,1} = \delta_1 + \gamma_{1,1,13,1}$	-	-	$\omega_{4,3,1} = \delta_1 + \gamma_{1,1,13,1} + \gamma_{1,3,13,1}^*$
<i>Messzeitpunkt 2</i>				
Booklet 02 M4-M5- L3-M2	-	-	$\omega_{3,3,2} = \delta_2 + \gamma_{1,3,2,2}$	-
Booklet 03 N2-L3-L2- N3	-	$\omega_{2,3,2} = \delta_2 + \gamma_{1,3,3,2}$	$\omega_{3,2,2} = \delta_2 + \gamma_{1,3,3,2} + \gamma_{1,2,3,2}^*$	-
Booklet 06 M7-L2-BI- BI	-	$\omega_{2,2,2} = \delta_2 + \gamma_{1,2,6,2}$	-	-
Booklet 11 BI-BI-L3- N2	-	-	$\omega_{3,3,2} = \delta_2 + \gamma_{1,3,11,2}$	-
Booklet 12 BI-BI-M4- L3	-	-	-	$\omega_{4,3,2} = \delta_2 + \gamma_{1,3,12,2}$
Booklet 16 L2-M7-BI- BI	$\omega_{1,2,2} = \delta_2$	-	-	-

Tab. 1 Darstellung der Zerlegung der Effekte einer Kovariate x auf die positions- und clusterspezifischen Testleistungen (ω_{pct}) (Fortsetzung)

	Position 1	Position 2	Position 3	Position 4
Booklet 17	–	–	–	$\omega_{4,2,2} = \delta_2 +$
MR-M7-				$\gamma_{1,2,17,2}$
N3-L2				

M Mathematik, *N* Naturwissenschaften, *L* Leseverständnis, *BI* Itemcluster der Tests zur Überprüfung der Bildungsstandards

Über die Positionen invariante $\omega_{p,t}$ -Parameter sind ein Indiz dafür, dass die Kovariate nicht mit den Positionseffekten interagiert, sodass der mittlere Parameter $\omega_{..t}$ (d. h. Mittelwert der Effekte über Cluster und Positionen zur Welle t) eine adäquate Darstellung des querschnittlichen Effekts von x liefert.

Die so definierten ω -Parameter können verwendet werden, um den Effekt von x auf die Leistungsveränderung abzuschätzen. So lassen sich die Leistungsveränderungen in Abhängigkeit von x getrennt nach Positionen bestimmen: $\Delta_{\omega_p} = \omega_{p,2} - \omega_{p,1}$.² Der über Positionen gemittelte Effekt auf die Leistungsänderung lässt sich zudem als $\Delta_{\omega_{..}} = \omega_{..2} - \omega_{..1}$ berechnen. Eine Situation, in der sich die positionsspezifischen Δ_{ω_p} -Werte deutlich voneinander unterscheiden, indiziert, dass die in Abhängigkeit von x ermittelten Leistungszuwächse auch zeitspezifischen Interaktionen zwischen x und den Positionseffekten geschuldet sind. Demgegenüber indiziert ein Ergebnismuster, das über Positionen konstante Effektschätzungen aufzeigt, dass die Änderung der Positionseffekte über die Zeit nicht von x abhängt und $\Delta_{\omega_{..}}$ eine valide Abschätzung des Effekts von x auf die Leistungsveränderung ermöglicht. Diese Situation ist aus inferenzstatistischer Sicht von Vorteil, da die $\Delta_{\omega_{..}}$ -Parameter mit einer größeren Genauigkeit geschätzt werden (d. h. geringere Standardfehler aufweisen), während die positionsspezifischen Δ_{ω_p} -Parameter im vorliegenden Design vergleichsweise ungenau geschätzt werden.

Das hier skizzierte Modell lässt sich im Prinzip um eine Vielzahl von Erklärungsvariablen und deren Interaktionen mit Bookleteffekten erweitern. Aus Gründen der Sparsamkeit haben wir uns jedoch auf Modelle mit maximal zwei Kovariaten beschränkt. Alle Modelle wurden mittels der Statistiksoftware Mplus 7.4 (Muthén und Muthén 1998–2012) geschätzt.

8 Ergebnisse

Die Auswertung der Leistungsentwicklung erfolgte in drei Schritten. Erstens wurde die Entwicklung in der Gesamtgruppe ausgewertet, wobei die Höhe der Positionseffekte und deren längsschnittliche Zu- oder Abnahme berücksichtigt wurden. Dieses Modell lieferte eine Abschätzung der mittleren Zu- oder Abnahme der Leistungs-

² Diese Interpretation der Δ_{ω_p} -Parameter ergibt sich unmittelbar aus der Beziehung zwischen den zeitspezifischen Regressionseffekten einer Kovariate auf ein wiederholt gemessenes Outcome. Die Differenz der zeitspezifischen Regressionsgewichte der Kovariate entspricht dem Regressionsgewicht der Differenz auf die Kovariate.

stände, der Leistungsvariabilität zu beiden Messzeitpunkten sowie der Variabilität der Leistungsveränderung.

Zweitens wurde das Ausgangsmodell um die Schulform als Erklärungsvariable ergänzt. Hierbei wurden auch die Interaktionen zwischen Schulform- und Bookleteffekten zugelassen. Zwar berichten bereits Nagy et al. (2017) die entsprechenden Schulformeffekte, allerdings liefert der hier verwendete Modellierungsansatz zusätzlich eine Abschätzung der Leistungsvariabilität und der Variabilität der Leistungsänderung nach Kontrolle der Schulform und deren Interaktionen mit den Bookleteffekten. Die entsprechenden Ergebnisse wurden gemeinsam mit den im Ausgangsmodell ermittelten Varianztermen verwendet, um den Erklärungsbeitrag der Schulform und deren Interaktionen mit den Positionseffekten abzuschätzen. Parallel hierzu wurde das Ausgangsmodell auch jeweils separat für die anderen betrachteten Erklärungsvariablen (d. h. Geschlecht, Zuwanderungshintergrund und sozioökonomischer Hintergrund) erweitert.

Drittens wurden die um die einzelnen Erklärungsvariablen erweiterten Modelle im letzten Schritt um die Schulformvariable und deren Interaktionen mit den Bookleteffekten ergänzt. Wir haben dieses Modell spezifiziert, da der Zuwanderungshintergrund und der sozioökonomische Hintergrund stark mit der besuchten Schulform kovariieren und sich somit die Frage nach deren eigenständigem Erklärungsbeitrag stellt. Wir haben jedoch auf eine gemeinsame Modellierung aller Erklärungsvariablen verzichtet, da die vorliegenden Daten keine belastbare Grundlage für ein derart hoch parametrisiertes Modell bieten.

In allen Modellen lag ein robuster Effekt einer Kovariate auf Lernzuwächse vor, wenn ein von 0 abweichender Effekt auf die Leistungsentwicklung an der ersten Testheftposition vorlag (d. h. Δ_{ω_1} -Parameter). Insofern sich der Parameter Δ_{ω_1} statistisch signifikant von 0 unterscheidet, bietet er einen robusten Hinweis für den um Positionseffekte bereinigten Effekt einer Kovariate auf die Leistungsentwicklung. Da dieser Test aufgrund der eher fragilen Datenkonstellation jedoch eine geringe Teststärke aufweist, kann der Effekt einer Kovariate auf die Leistungsentwicklung bei einer Situation, in der die positionsspezifischen Δ_{ω_p} -Parameter weitgehend invariant über Testpositionen ausfallen (d. h., wenn keine Hinweise für das Vorliegen von Bookleteffekten mit der Kovariate vorliegen), auch anhand des über Positionen und Cluster gemittelten $\Delta_{\omega..}$ -Parameters inferenzstatistisch evaluiert werden.

9 Leistungsentwicklung in der Gesamtstichprobe und Effekte der besuchten Schulform

In Tab. 2 sind die mittleren Testleistungen zu beiden Zeitpunkten getrennt nach Testheftpositionen abgetragen. Die über Positionen hinweg abnehmenden Mittelwerte dokumentieren die Positionseffekte. Mit Ausnahme der dritten Position unterschieden sich die Leistungen nicht zwischen den Messzeitpunkten ($\widehat{\Delta}_{\omega_p}$ -Parameter in Tab. 2). Dieses Befundmuster indiziert, dass mit Ausnahme der dritten Position die Höhe der Positionseffekte unverändert blieb und dass die mittlere Leistung im Gesamtsample unverändert ausfiel ($\widehat{\Delta}_{\omega..}$ -Parameter). Allerdings lag im Gesamtsample

Tab. 2 Testleistungen und Leistungszuwächse in der Gesamtstichprobe (mittlere Testleistungen) und Effekte der Gymnasialzugehörigkeit auf Leistungsstände und Leistungszuwächse getrennt nach Testheftpositionen

	Mittlere Testleistungen				Effekt der Gymnasialzugehörigkeit									
	$\hat{\omega}_{p-1}$	$\hat{\omega}_{p-2}$	$\hat{\omega}_{p-1}$	$\hat{\omega}_{p-2}$	$\hat{\omega}_{p-1}$	$\hat{\omega}_{p-2}$	$\hat{\omega}_{p-1}$	$\hat{\omega}_{p-2}$	$\hat{\omega}_{p-1}$	$\hat{\omega}_{p-2}$	$\hat{\omega}_{p-1}$	$\hat{\omega}_{p-2}$	$\hat{\omega}_{p-1}$	$\hat{\omega}_{p-2}$
Pos. 1	0,44	0,46	0,07)**	0,02	1,02	1,03	0,03)	0,03)	1,02	1,03	0,03)	0,03)	0,01	0,01
Pos. 2	0,14	0,15	0,06)**	0,01	1,17	1,55	0,03)	0,03)	1,17	1,55	0,03)	0,03)	0,37	0,37
Pos. 3	0,17	-0,02	0,03)**	-0,19	1,17	1,47	0,03)	0,06)**	1,17	1,47	0,03)	0,06)**	0,30	0,30
Pos. 4	-0,27	-0,22	0,04)**	0,05	1,31	1,67	0,03)	0,07)	1,31	1,67	0,03)	0,07)	0,36	0,36
Var	0,98	1,60	0,03)**	0,60	0,64	1,03	-	0,07)**	0,64	1,03	-	0,07)**	0,55	0,55

* $p \leq 0,05$; ** $p \leq 0,01$

ein deutlicher Anstieg der Leistungsvariabilität vor, der sich in einer vergleichsweise hohen Variabilität der individuellen Leistungsänderungen niederschlug.

Tab. 2 umfasst auch die Ergebnisse der Effekte der Gymnasialzugehörigkeit auf die Testleistungen. Die querschnittlichen Befunde indizierten eine Interaktion der Schulform mit den Positionseffekten. Die Leistungsvorteile der Gymnasiastinnen und Gymnasiasten nahmen über die Positionen hinweg zu, wobei das Muster zum zweiten Messzeitpunkt besonders prononciert war ($\hat{\omega}_{p,1}$ - und $\hat{\omega}_{p,2}$ -Parameter). Dementsprechend unterschieden sich Schulformunterschiede in den Zuwachsschätzungen zwischen Positionen ($\hat{\Delta}\omega_p$ -Parameter). Die mittleren Zuwächse unterschieden sich an der ersten Position nicht zwischen den Schulformen, während für die Testheftpositionen zwei bis vier größere Leistungszuwächse in der Gymnasialgruppe ermittelt wurden. Dieses Muster führte dazu, dass die über Positionen

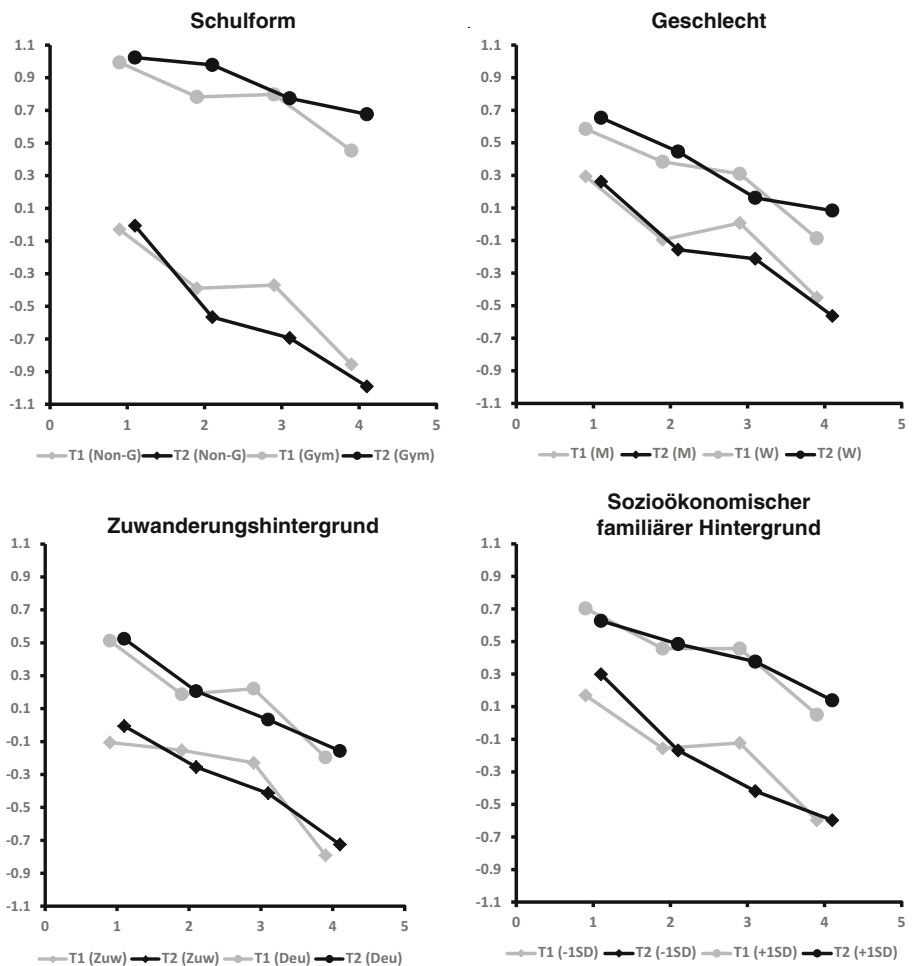


Abb. 1 Mittlere Testleistungen in Welle 1 und 2 nach Testheftpositionen für unterschiedliche Ausprägungen der untersuchten Kovariaten

gemittelten Leistungszuwächse statistisch signifikant zugunsten der Gymnasiasten ausfielen ($\widehat{\Delta}_{\omega..}$ -Parameter). Aufgrund der Unterschiedlichkeit der Effekte auf den Leistungszuwachs in Abhängigkeit der Testheftpositionen ließ sich der $\widehat{\Delta}_{\omega..}$ -Parameter nicht sinnvoll interpretieren. Im vorliegenden Fall ist davon auszugehen, dass die Gruppenunterschiede in den Leistungsänderungen eher Schulformunterschieden in der Veränderung von Positionseffekten als tatsächlichen Unterschieden in Lerngewinnen geschuldet waren. So unterschieden sich Leistungszuwächse an der ersten Testheftposition nicht zwischen Schulformen ($\widehat{\Delta}_{\omega p_1} = 0,01$; $p = 0,965$). Dieses Ergebnismuster wurde bereits von Nagy et al. (2017) berichtet und ist in kompakter Form in Abb. 1 wiedergegeben.

Nach Kontrolle der Schulform und deren Interaktion mit den Positionseffekten verringerten sich die Variabilität der querschnittlichen Leistungsstände und der Leistungsänderungen. Die Variabilität der Leseleistungen zur ersten und zweiten Welle nahm gegenüber dem ersten Modell um 34,4 % bzw. 35,4 % ab, während sich die Variabilität der Änderungsraten um 7,6 % verringerte. Dabei gilt zu berücksichtigen, dass die Abnahme der Streuung der Zuwachsschätzung nahezu vollständig auf die Kontrolle der Interaktionseffekte zwischen Bookleteffekten und der Schulform zurückgeführt werden konnte, da sich die Schulformen nicht in den mittleren Leistungsänderungen voneinander unterschieden.

10 Geschlechterunterschiede in der Leistungsentwicklung

Die Ergebnisse zu den quer- und längsschnittlichen Geschlechterunterschieden finden sich in Tab. 3. Zu beiden Messzeitpunkten fanden sich Leistungsvorteile zugunsten der Mädchen, wobei die Unterschiede zwischen den Positionen variierten. Wie aus Tab. 3 ersichtlich ist, lagen die größten Geschlechterunterschiede an den Positionen zwei und vier vor, wobei dieses Muster zum zweiten Messzeitpunkt stärker ausgeprägt war. Dieses Muster führte dazu, dass sich der Zusammenhang zwischen Geschlecht und Leistungsänderung zwischen den Positionen unterschied. Die Vorzeichen der $\widehat{\Delta}_{\omega p}$ -Parameter indizierten zwar durchweg höhere Lerngewinne für Mädchen, der Geschlechterunterschied im Zuwachs wurde jedoch nur für die vierte Testheftposition statistisch signifikant. Der über Positionen gemittelte Effekt indizierte jedoch einen statistisch signifikant größeren Leistungsgewinn für Mädchen. Demnach nahmen die Geschlechterunterschiede um $\widehat{\Delta}_{\omega..} = 0,14$ Einheiten auf der Logit-Metrik zu, was einer Vergrößerung der Geschlechterdifferenz gegenüber der über Positionen gemittelten Geschlechterdifferenz zum ersten Messzeitpunkt von $\widehat{\omega}_{..1} = 0,36$ um rund 39 % entspricht.

Eine genauere Betrachtung der geschlechterspezifischen Leistungsverläufe (Abb. 1) deutete jedoch an, dass die Geschlechterunterschiede aller Wahrscheinlichkeit nach mit Geschlechterunterschieden in der Änderung von Positionseffekten konfundiert waren. So nahmen in der Gruppe der Jungen die Positionseffekte von der ersten zur zweiten Welle deutlich zu, während sich die Positionseffekte in der Gruppe der Mädchen kaum zwischen den Wellen unterschieden. Berücksichtigt man zudem die Tatsache, dass die Leistungsänderungen zwischen den Geschlechtergruppen an der ersten Testheftposition geringer ausfielen ($\widehat{\Delta}_{\omega p_1} = 0,10$; $p = 0,502$),

Tab. 3 Effekte des Geschlechts, des Zuwanderungshintergrunds und des sozioökonomischen Hintergrunds auf Leistungsstände und Leistungszuwächse getrennt nach Testheftpositionen

	Ohne Kontrolle des Bildungsgangs				Nach Kontrolle des Bildungsgangs			
	$\hat{\omega}_{p-1}$ (SE)	$\hat{\omega}_{p-2}$ (SE)	$\hat{\Delta}_{\omega_p}$ (SE)	$\hat{\Delta}_{\omega_{..}}$ (SE)	$\hat{\omega}_{p-1}$ (SE)	$\hat{\omega}_{p-2}$ (SE)	$\hat{\Delta}_{\omega_p}$ (SE)	$\hat{\Delta}_{\omega_{..}}$ (SE)
<i>Geschlecht (I = weiblich)</i>								
Pos. 1	0,29 (0,06)***	0,39 (0,14)***	0,10 (0,12)***	0,14 (0,15)	0,30 (0,07)**	0,32 (0,13)***	0,02 (0,14)	0,12 (0,06)*
Pos. 2	0,48 (0,07)***	0,60 (0,12)***	0,12 (0,11)***	0,14 (0,14)	0,42 (0,07)**	0,57 (0,11)***	0,16 (0,12)	0,12 (0,06)*
Pos. 3	0,30 (0,06)***	0,37 (0,11)***	0,07 (0,13)***	0,14 (0,12)	0,30 (0,07)**	0,35 (0,09)***	0,05 (0,11)	0,12 (0,06)*
Pos. 4	0,37 (0,07)***	0,65 (0,13)***	0,28 (0,14)***	0,14 (0,14)**	0,37 (0,07)**	0,61 (0,12)***	0,24 (0,13)*	0,12 (0,06)*
Var	0,95 (0,03)***	1,52 (0,08)***	0,59 (0,07)***	–	0,61 (0,02)***	0,97 (0,06)***	0,54 (0,06)	–
<i>Zuwanderungshintergrund (I = Zuwanderungshintergrund)</i>								
Pos. 1	-0,62 (0,10)***	-0,53 (0,27)**	0,09 (0,18)***	-0,00 (0,28)	-0,45 (0,09)***	-0,31 (0,25)	0,14 (0,26)	0,01 (0,11)
Pos. 2	-0,34 (0,11)***	-0,46 (0,18)***	-0,12 (0,19)***	-0,00 (0,20)	-0,20 (0,09)**	-0,28 (0,16)*	-0,09 (0,18)	0,01 (0,11)
Pos. 3	-0,45 (0,11)***	-0,45 (0,19)***	0,00 (0,21)***	-0,00 (0,20)	-0,31 (0,10)***	-0,28 (0,17)*	0,03 (0,18)	0,01 (0,11)
Pos. 4	-0,59 (0,11)***	-0,57 (0,21)***	0,03 (0,22)***	-0,00 (0,23)	-0,34 (0,10)***	-0,40 (0,20)**	-0,06 (0,22)	0,01 (0,11)
Var	0,95 (0,03)***	1,57 (0,08)***	0,59 (0,07)***	–	0,63 (0,02)***	1,02 (0,06)***	0,55 (0,06)***	–
<i>Sozioökonomischer Hintergrund</i>								
Pos. 1	0,27 (0,03)***	0,17 (0,07)**	-0,10 (0,06)***	0,02 (0,07)	0,09 (0,03)***	-0,04 (0,07)	-0,13 (0,08)*	-0,05 (0,04)
Pos. 2	0,31 (0,04)***	0,33 (0,06)***	0,02 (0,05)***	0,02 (0,07)	0,13 (0,03)***	0,04 (0,07)	-0,09 (0,07)	-0,05 (0,04)
Pos. 3	0,29 (0,04)***	0,40 (0,05)***	0,11 (0,06)**	0,02 (0,06)*	0,10 (0,03)***	0,13 (0,06)**	0,03 (0,06)	-0,05 (0,04)
Pos. 4	0,32 (0,04)***	0,37 (0,07)***	0,05 (0,08)***	0,02 (0,08)	0,10 (0,04)***	0,09 (0,07)	-0,01 (0,08)	-0,05 (0,04)
Var	0,89 (0,03)***	1,47 (0,08)***	0,58 (0,07)***	–	0,63 (0,02)***	1,01 (0,06)***	0,54 (0,06)***	–

* $p \leq 0,10$; ** $p \leq 0,05$; *** $p \leq 0,01$

kann geschlossen werden, dass die Daten keine belastbaren Hinweise für geschlechterspezifische Lernzuwächse lieferten. Die Ergebnisse des um die Effekte der Schulform ausgebauten Modells (Tab. 3) unterstützten die Vermutung ausbleibender Geschlechterunterschiede in Lernzuwächsen, da sich die Geschlechterunterschiede in den mittleren Leistungsänderungen an der ersten Position nun noch stärker an 0 annäherten ($\widehat{\Delta}_{\omega_{p1}} = 0,02$; $p = 0,872$).

Das Geschlecht erklärte 3,5 % bzw. 5,0 % der Leistungsvariabilität zu den beiden Messzeitpunkten und 1,6 % der Variabilität in den Leistungsänderungen, wobei das letztgenannte Ergebnis nahezu ausschließlich der Interaktion der Erklärungsvariable mit den Positionseffekten geschuldet war. Die entsprechende Interaktion führte zudem dazu, dass das Geschlecht und die Schulform gemeinsam 9,7 % der Variabilität der Unterschiede in den Leistungsänderungen erklärten, wobei der Erklärungsbeitrag fast vollständig auf die Positionseffekte zurückging.

11 Zuwanderungshintergrund und Leistungsentwicklung

Wie die in Tab. 3 dargestellten Befunde zeigen, war der Zuwanderungshintergrund zu beiden Messzeitpunkten deutlich mit der Testleistung assoziiert. Die Leistungsunterschiede nahmen zu beiden Messzeitpunkten an den Positionen zwei und drei gegenüber der ersten Position ab, und befanden sich an der vierten Position auf dem Niveau der Ausgangsposition. Wie aus Abb. 1 hervorgeht, war dieser Befund auf die stärker ausgeprägten Positionseffekte bei Schülerinnen und Schülern ohne Zuwanderungshintergrund in der ersten Erhebung zurückzuführen.

Insgesamt lagen keine belastbaren Anzeichen dafür vor, dass sich die Leistungszuwächse zwischen Schülerinnen und Schülern mit und ohne Zuwanderungshintergrund unterschieden. Die Schätzungen der positionsspezifischen Effekte auf die Leistungsänderung fielen durchweg nahe 0 aus ($\widehat{\Delta}_{\omega_p}$ -Parameter), was zu einer Gesamtschätzung von 0 führte. Der Zuwanderungshintergrund und dessen Interaktion mit den Positionseffekten erklärte insgesamt rund 2,8 % und 1,8 % der Leistungsvariabilität zu den beiden Messzeitpunkten, sowie 1,6 % der Variabilität in den Leistungszuwächsen. Die vergleichsweise geringe Varianzerklärung ist hauptsächlich darauf zurückzuführen, dass der Anteil von Schülerinnen und Schülern mit Zuwanderungshintergrund aufgrund der von uns gewählten dichotomen Kategorisierung nur relativ gering war.

Nach Kontrolle der besuchten Schulform verringerten sich die querschnittlichen Effekte des Zuwanderungshintergrunds (Tab. 3). Aufgrund des größeren Anteils fehlender Itemantworten zum zweiten Messzeitpunkt wurde ein Effekt nicht signifikant und zwei Effektschätzungen wurden nur auf dem Niveau von $p \leq 0,10$ statistisch signifikant. Die Datenkonstellation hatte zur Folge, dass die Standardfehler der Effekte des Zuwanderungshintergrunds auf die Leistungsänderungen sehr groß ausfielen, sodass keine robusten Schlussfolgerungen über dessen Effekt auf den Leistungszuwachs gezogen werden konnten. Gemessen an der erklärten Varianz schienen die Effekte des Zuwanderungshintergrunds auf die Leistungsänderung bei Kontrolle der besuchten Schulform vernachlässigbar zu sein, da diese Variable zu keinem nen-

nenswertem Anstieg der erklärten Varianz gegenüber der alleinigen Betrachtung der Schulform führte (8,4 % vs. 7,6 %).

12 Sozioökonomischer Hintergrund und Leistungsentwicklung

Der sozioökonomische familiäre Hintergrund war erwartungsgemäß zu beiden Messzeitpunkten mit den Leseleistungen assoziiert (Tab. 3). Während die Assoziationen zum ersten Messzeitpunkt kaum von den Positionen abhingen, zeigte sich zum zweiten Messzeitpunkt ein Anstieg des Zusammenhangs über Positionen. Dieses Ergebnismuster führte dazu, dass der Effekt des sozioökonomischen Hintergrunds auf die Leistungszuwächse zwischen Positionen variierte. Der Effekt war nur zur dritten Position positiv und auf dem Niveau von $p \leq 0,10$ signifikant ($\widehat{\Delta}_{\omega_3} = 0,11$; $p = 0,082$). Diese inkonsistente Befundlage schien in erster Linie der Interaktion des sozioökonomischen Hintergrunds mit den Positionseffekten zum zweiten Messzeitpunkt geschuldet zu sein. Dieser Befund ist in Abb. 1 grafisch dargestellt, wobei stärkere Positionseffekte für Schülerinnen und Schüler mit geringerem sozioökonomischem Hintergrund erkenntlich sind. Zusammengefasst lieferten die Daten jedoch keine belastbaren Hinweise für einen Zusammenhang zwischen dem sozioökonomischen familiären Hintergrund und dem Kompetenzzuwachs. Tatsächlich schien der sozioökonomische Hintergrund eher im Querschnitt mit der Leseleistung (8,9 % und 8,2 % erklärte Varianz zu den Wellen eines und zwei), als mit der Leistungsänderung assoziiert zu sein (3,1 % erklärte Varianz).

Der Zusammenhang zwischen dem sozioökonomischen Hintergrund und den Testleistungen nahm nach Kontrolle der Schulform deutlich ab (Tab. 3). Die vom Betrag her geringen, aber statistisch signifikanten Effekte variierten zum ersten Messzeitpunkt kaum zwischen Positionen ($\widehat{\omega}_{p,1}$ -Parameter). Im Gegensatz dazu schienen die Zusammenhänge des sozioökonomischen Status mit der zur zweiten Welle gemessenen Leistung weitgehend Positionseffekten geschuldet zu sein. Die Effekte nahmen der Tendenz nach über Positionen zu ($\widehat{\omega}_{p,2}$ -Parameter), wobei sich nur ein einzelner Effekt als statistisch signifikant erwies. Nach Kontrolle der Schulform fiel der Effekt des sozioökonomischen Hintergrunds auf den Leistungszuwachs an der ersten Position negativ aus und war auf dem Niveau von $p \leq 0,10$ ($p = 0,095$) signifikant. Dieser Effekt deutete einen disparitätsmindernden Leistungsverlauf an, der jedoch aufgrund der unpräzisen Effektschätzung mit großer Vorsicht interpretiert werden muss. Diese Interpretation wird auch durch den Prozentsatz der erklärten Varianz nahegelegt, der zum ersten und zweiten Messzeitpunkt 35,5 % und 36,5 % betrug, was den inkrementell zur Schulform geringen Erklärungsbeitrag des sozioökonomischen Hintergrunds verdeutlicht. Ein ähnliches Ergebnis ergab sich auch für die Leistungsänderung, wobei hier durch die Hinzunahme des Prädiktors die Erklärungsleistung der Schulform (7,6 %) auf immerhin 10,5 % gesteigert werden konnte.

13 Diskussion

Gegenstand des vorliegenden Beitrags war die Untersuchung ausgewählter Prädiktoren der Entwicklung der Lesekompetenz im letzten Abschnitt der Sekundarstufe I. Unser Augenmerk lag auf zentralen institutionellen (Schulform), familiären (Zuwanderungshintergrund und sozioökonomischer Hintergrund) und individuellen Merkmalen (Geschlecht), die im Zentrum vieler früherer Querschnittuntersuchungen standen, aber vergleichsweise selten im Kontext der Entwicklung der Leseleistung am Ende der Sekundarstufe I untersucht wurden. Zudem haben wir Effekte der Veränderung der Testbearbeitungspersistenz (Debeer et al. 2014) berücksichtigt, die mit tatsächlichen Lernzuwächsen konfundiert sein können. Zu diesem Zweck haben wir versucht die Effekte der Kovariaten auf die Leistungsentwicklung von deren Zusammenhängen mit der Veränderung der Positionseffekten zu separieren (vgl. Nagy et al. 2017). Wir haben uns für dieses Vorgehen entschieden, da der im PISA-Längsschnitt 2012/2013 verwendete Lesetest besonders stark von Positionseffekten betroffen war, die zudem deutlich zum zweiten Messzeitpunkt zunahm (vgl. Nagy et al. 2017). Vor dem Hintergrund dieser Ergebnisse erschienen Zusammenhänge zwischen Änderungen in der Testbearbeitungspersistenz und anderen Kovariaten plausibel.

Insgesamt haben wir keine starken Zusammenhänge zwischen den untersuchten Kovariaten und der Lesekompetenzentwicklung erwartet. Wie an früherer Stelle erläutert, impliziert eine Situation, in der sich die Entwicklung der Leseleistung im Laufe der Sekundarstufe I deutlich verlangsamt oder eventuell sogar ganz abflacht (Baumert et al. 2012; Bloom et al. 2008) und größere (reale) Kompetenzverluste ausgeschlossen werden, maximal schwache Effekte von Kovariaten auf die Leistungsentwicklung. Tatsächlich indizieren unsere Ergebnisse, dass die mittleren Leistungszuwächse in der Gesamtgruppe nahe 0 ausfielen und dass keine der betrachteten Variablen robuste Zusammenhänge mit der Leseentwicklung aufwies. Allenfalls deutete sich an, dass Schülerinnen und Schüler mit einem geringeren sozioökonomischen familiären Hintergrund in der Tendenz höhere Lerngewinne aufwiesen und somit ihre Leistungsrückstände teilweise kompensierten (vgl., Baumert et al. 2012). Der Befund konnte jedoch nur für die erste Testheftposition nachgewiesen werden und war nur auf dem Niveau von $p \leq 0,10$ signifikant. Es bedarf somit einer erneuten Prüfung, bevor belastbare Schlüsse über disparitätsmindernde Effekte der Beschulung am Ende der Sekundarstufe I gezogen werden können.

Insgesamt indizierten unsere Analysen, dass drei der vier untersuchten Kovariaten mit der Veränderung der Testbearbeitungspersistenz assoziiert waren. Dabei gilt zu berücksichtigen, dass das von uns verwendete IRT-Modell neben Positionseffekten auch andere Formen von Testkontexteffekten abbildet (Brennan 1992; Nagy et al. 2016). Gleichwohl lieferten unsere Ergebnisse deutliche Hinweise dafür, dass die ermittelten Effekte auf die Entwicklung der Lesekompetenz hauptsächlich Positionseffekten geschuldet waren (vgl. Nagy et al. 2017). Die stärksten Zusammenhänge mit der längsschnittlichen Veränderung der Bearbeitungspersistenz ergaben sich mit der Schulform (vgl. Nagy et al. 2017). Die Ausblendung von Schulformunterschieden im Anstieg der Positionseffekte führte dazu, dass ein Scheineffekt in Form eines Schulformunterschieds in (realen) Lerngewinnen auftrat (Nagy et al. 2017). Ein vergleichbarer Befund ergab sich hinsichtlich des Geschlechts. Wie wir

in diesem Beitrag zeigen konnten, führte die Ausblendung von Positionseffekten zu einem Ergebnismuster, das einen höheren Lerngewinn bei Mädchen gegenüber Jungen nahelegte. Die vorgenommene differenzierte Analyse zeigte aber, dass an Positionen, die nicht von Positionseffekten betroffen sind, dieser Unterschied nicht repliziert werden konnte. Schließlich lieferten unsere Analysen auch Hinweise dafür, dass Kovariaten-Booklet-Interaktionen im Prinzip reale, aber gegenläufige Effekte auf die Leistungsentwicklung verdecken können. So trat der negative Effekt des sozioökonomischen Hintergrunds auf die Leistungsentwicklung erst nach Kontrolle von Positionseffekten in Erscheinung. Wie bereits aufgeführt, bedarf dieser Befund aufgrund der schwachen Datenlage einer erneuten Prüfung bevor definitivere Rückschlüsse gezogen werden können.

Unsere Ergebnisse lieferten zudem Hinweise dafür, dass die Effekte der Positionen der Itemcluster nicht notwendigerweise linear verlaufen (vgl. Abb. 1) und das sich die Form der Verläufe in Abhängigkeit der betrachteten Kovariaten unterscheiden kann. So ähnelte der Leistungsverlauf über Itemclusterpositionen bei weniger leistungsstarken Gruppen (nichtgymnasiale Schulformen, männliche Schüler und Schülerinnen und Schüler mit einem weniger leistungsförderlichen sozialen Hintergrund) einer Treppenfunktion. Dieses Muster könnte eine Auffrischung von motivationalen Ressourcen indizieren, die sich in Folge der Testpause einstellte, die in der Mitte der Tests gewährt wurde (ca. 15 min). Die Treppenfunktion war in den leistungsstärkeren Gruppen (Gymnasien, weibliche Schülerinnen und Schüler mit leistungsförderlichem familiären Hintergrund) nicht anzutreffen und die Testleistungen zeigten in der Tendenz einer geringere Abnahme über Itemclusterpositionen. Dieser Befund weist darauf hin, dass die Motivation für eine gewissenhafte Testbearbeitung in diesen Schülergruppen höher ausgeprägt ist, und diese Schülerinnen und Schüler resistenter gegenüber Impulsen für eine Reduktion der Anstrengung sind. Da wir mit den vorgenommenen Auswertungen Neuland betreten haben, bedarf es einer Replikation der Befunde in unabhängigen Stichproben, bevor eine belastbare Interpretation der Effektmuster vorgenommen werden kann.

Unabhängig von der Form der Muster der Effekte der Kovariaten über Itemclusterpositionen, liefern unsere Befunde jedoch deutliche Hinweise dafür, dass die hinsichtlich der ersten Itemclusterpositionen definierten Kompetenzausprägungen am wenigsten von einer suboptimalen Persistenz der Testbearbeitung betroffen waren und somit die bestmögliche Annäherung an die tatsächlichen Kompetenzausprägungen und Kompetenzveränderungen liefern. Die in diesem Beitrag berichteten Effekte auf den Kompetenzzuwachs, fielen jedoch aufgrund der fragilen Datenkonstellation, die sich infolge der schwachen Verankerung der Kompetenzmetrik zum zweiten Messzeitpunkt einstellte, relativ unpräzise aus (d. h. vergleichsweise große Standardfehler). Nichtsdestotrotz schätzen wir das Risiko für systematisch verzerrte Ergebnisse als sehr gering ein, da die von Nagy et al. (2017) berichteten Befunde zeigten, dass die Schwierigkeiten der Items des zur Verankerung verwendeten Clusters einen hohen Grad an Invarianz über Itemclusterpositionen und Schulformen aufwiesen. Aus diesen Gründen vermuten wir, dass die hier berichteten Ergebnisse sich in ähnlicher Weise auch bei günstiger gestalteten Testdesigns (s. unten) einstellen würden.

14 Theoretische und Praktische Implikationen

In der Gesamtschau indizieren die vorliegenden Befunde, dass alle untersuchten Variablen nicht, oder nur schwach (z. B. sozioökonomischer Status), mit den tatsächlichen Lerngewinnen von Schülerinnen und Schülern am Ende der Sekundarstufe I assoziiert sind. Dieses Ergebnismuster ist vor den Hintergrund früherer Untersuchungen plausibel (s. oben). Aus einer methodischen Perspektive geben die Ergebnisse aber deutliche Hinweise für das Risiko von Scheineffekten auf die Entwicklung der Leseleistung in späteren Phasen der Beschulung. Wie Bloom et al. (2008) darlegten, zeichnet sich die Leistungsentwicklung in dieser Phase durch geringe Lernraten aus, sodass auch vergleichsweise schwache Effekte auf die Leistungsentwicklung theoretisch und praktisch relevant sind. Unsere Befunde zeigen, dass zumindest die für den Bereich der Leseleistung ermittelten Ergebnisse hinsichtlich Positionseffekten, die als Indikatoren der Testbearbeitungspersistenz gelten (Debeer et al. 2014), verzerrt sein können. Diese Situation führen insofern zu einem Dilemma, da die zu erwartenden realen Effekte auf die Kompetenzentwicklung gering sind und mit (ebenfalls numerisch geringen) Artefakten des Testbearbeitungsverhaltens konfundiert sein können.

Ein möglicher Ausweg aus diesem Dilemma ist die explizite Analyse von Störeffekten. Ein solches Vorgehen setzt geeignete Testdesigns und Modellierungsansätze voraus. In diesem Beitrag haben wir das in den PISA-Erhebungen eingesetzte rotierte Bookletdesign genutzt, um Positionseffekte zu identifizieren (vgl. Nagy et al. 2017). Da Lesen in der PISA-Längsschnitterhebung 2012/2013 nicht die Hauptdomäne darstellte, wurde dieser Kompetenzbereich mit einer vergleichsweise geringen Itemanzahl erhoben. Eine Konsequenz dieser Datensituation ist eine geringe Teststärke des verwendeten statistischen Ansatzes. Dem Verlust der Teststärke ließe sich zum einen durch die Optimierung des Erhebungsdesigns (Weirich et al. 2014) und der Verwendung alternativer Indikatoren der Testbearbeitungsmotivation, wie beispielsweise Bearbeitungszeiten von Testitems (z. B. Goldhammer et al. 2014), begegnen. Letztere sind aber in der Haupterhebung des PISA-Längsschnitts 2012/2013 aufgrund der papierbasierten Erhebung nicht verfügbar.

15 Schlussfolgerungen und Ausblick

Die Frage nach der Verringerung oder Vergrößerung von schulformspezifischen Unterschieden in der Leseleistung, von herkunftsbedingten Disparitäten sowie von Geschlechterdisparitäten im Verlauf der Beschulung ist ein zentrales Thema der Schulleistungsforschung. Die hier vorgestellte Auswertung des PISA-Längsschnitts 2012/2013 liefert keine Hinweise auf einen Anstieg der entsprechenden Disparitäten. Vielmehr zeichnete sich ein Bild ab, wonach die zum Teil sehr hohen Leistungsunterschiede vom 9. bis zum 10. Schuljahr weitgehend konstant bleiben.

Bei der Interpretation der Befunde gilt jedoch zu berücksichtigen, dass wir die Leistungsentwicklung an Hauptschulen außer Acht lassen mussten (vgl. Heine et al. 2017). Somit ist nicht geklärt, inwieweit diese Schlussfolgerungen bei einer zusätzlichen Berücksichtigung dieser Schülergruppe Bestand haben. Zur Klärung der Frage

bedarf es Längsschnittstudien, die auch einen früheren Zeitraum der Beschulung (z. B. ab Klassenstufe 8) berücksichtigen und ein Erhebungsdesign verwenden, das die Kontrolle der unerwünschten Störeffekte der Testbearbeitungspersistenz ermöglicht.

Ebenso haben wir in diesem Beitrag auf eine Auswertung von Kompetenzstufen (Watermann und Klieme 2002) verzichtet, die im Bereich von Large-Scale-Studien üblich ist. Obwohl dieser Ansatz durch seine Transparenz und einfache Kommunizierbarkeit besticht, setzt er hohe Anforderungen an die Belastbarkeit der Leistungsdaten. Da der Bereich Lesen im PISA-Längsschnitt 2012/2013 mit nur wenigen Items erhoben wurde, die zudem stark von Positionseffekten und deren Interaktion mit Hintergrundvariablen betroffen waren, haben wir von entsprechenden Auswertungen abgesehen. Diese sollten zukünftigen Studien, die auf einen breiteren Kranz von Items zurückgreifen können, vorbehalten werden.

Literatur

- Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: the contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*, 235–251.
- Artelt, C., Drechsel, B., Bos, W., & Stubbe, T. C. (2008). Lesekompetenz in PISA und PIRLS/IGLU – ein Vergleich. *Zeitschrift für Erziehungswissenschaft, 10*(Sonderheft), 35–52.
- Artelt, C., Stanat, P., Schneider, W., & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–137). Opladen: Leske + Budrich.
- Barker, R., & Escarpit, R. (Hrsg.). (1973). *The book hunger*. Paris: UNESCO.
- Baumert, J., Nagy, G., & Lehmann, R. H. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools? *Child Development, 83*, 1347–1367.
- Becker, M. (2009). *Kognitive Leistungsentwicklung in differenziellen Lernumwelten: Effekte des gegliederten Sekundarschulsystems in Deutschland*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*, 289–328.
- Brennan, R. L. (1992). The Context of Context Effects. *Applied Measurement in Education, 5*, 225–264.
- Cameron, C. E., Grimm, K. J., Steele, J. S., Castro-Schilo, L., & Grissmer, D. W. (2015). Nonlinear Gompertz curve models of achievement gaps in mathematics and reading. *Journal of Educational Psychology, 107*, 789–804.
- Debeer, D., Buchholz, J., Hartig, J., & Janssen, R. (2014). Student, school, and country differences in sustained test-taking effort in the 2009 PISA reading assessment. *Journal of Educational and Behavioral Statistics, 39*, 502–523.
- De Fraine, B., Van Damme, J., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology, 32*, 132–150.
- DeMars, C. E. (2007). Changes in rapid-guessing behavior over a series of assessments. *Educational Assessment, 12*, 23–45.
- Gailberger, S., & Willenberg, H. (2008). Leseverstehen Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 60–71). Weinheim: Beltz.
- Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social science research, 21*, 1–56.
- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*, 608–626.

- Hartig, J., & Buchholz, J. (2012). A multilevel item response model for item position effects and individual persistence. *Psychological Test and Assessment Modeling*, 54, 418–431.
- Härtig, H., Heitmann, P., & Retelsdorf, J. (2015). Analyse der Aufgaben zur Evaluation der Bildungsstandards in Physik – Differenzierung von schriftsprachlichen Fähigkeiten und Fachlichkeit. *Zeitschrift für Erziehungswissenschaft*, 18, 763–779.
- Heine, J. H., Nagy, G., Meinck, S., Zühlke, O., Mang, J. (2017). Empirische Grundlage, Stichprobenausfall und Adjustierung im PISA-Längsschnitt 2012–2013. *Zeitschrift für Erziehungswissenschaft*. doi:[10.1007/s11618-017-0756-0](https://doi.org/10.1007/s11618-017-0756-0)
- Hohn, K., Schiepe-Tiska, A., Sälzer, C., & Artelt, C. (2013). Lesekompetenz in PISA 2012: Veränderungen und Perspektiven. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 217–244). Münster: Waxmann.
- Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65, 373–398.
- Kigel, R. M., McElvany, N., & Becker, M. (2015). Effects of immigrant background on text comprehension, vocabulary, and reading motivation: a longitudinal study. *Learning and Instruction*, 35, 73–84.
- Klaukien, A., Ackermann, D., Helmschrott, S., Rammstedt, B., Solga, H., & Wößmann, L. (2013). Grundlegende Kompetenzen auf dem Arbeitsmarkt. In B. Rammstedt (Hrsg.), *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich. Ergebnisse von PIAAC 2012* (S. 127–166). Münster: Waxmann.
- Leary, L. F., & Dorans, N. J. (1985). Implications for altering the context in which test items appear: a historical perspective on an immediate concern. *Review of Educational Research*, 55, 387–423.
- Lehmann, R. H., Hunger, S., Ivanov, S., & Gänsfuß, R. (2004). *Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 11. Ergebnisse einer längsschnittlichen Untersuchung in Hamburg*. Hamburg: Behörde für Bildung und Sport.
- Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal*, 47, 596–632.
- Lundberg, I. (2002). The child's route into reading and what can go wrong. *Dyslexia*, 8, 1–13.
- Marx, A., & Stanat, P. (2012). Reading comprehension of immigrant students in Germany: research evidence on determinants and target points for intervention. *Reading and Writing*, 25, 1929–1945.
- Meyers, J. L., Miller, G. E., & Way, W. D. (2009). Item position and item difficulty change in an IRT-based common item equating design. *Applied Measurement in Education*, 22, 38–60.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7. Aufl.). Los Angeles CA: Muthén & Muthén.
- Nagy, G., & Neumann, M. (2010). Psychometrische Aspekte des Tests zu den voruniversitären Mathematikleistungen in TOSCA-2002 und TOSCA-2006: Unterrichtsvalidität, Rasch-Homogenität und Messäquivalenz. In U. Trautwein, M. Neumann, G. Nagy, O. Lüdtke & K. Maaz (Hrsg.), *Schulleistungen von Abiturienten: Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand* (S. 281–306). Wiesbaden: VS.
- Nagy, G., Lüdtke, O., & Köller, O. (2016). Modeling test context effects in longitudinal achievement data: Examining position effects in the longitudinal German PISA 2012 assessment. *Psychological Tests and Assessment Modeling*, 58, 641–670.
- Nagy, G., Lüdtke, O., Köller, O., Heine, J. H. (2017). IRT-Skalierung der Tests im PISA-Längsschnitt 2012/2013: Auswirkungen von Testkontexteffekten auf die Zuwachsschätzung. *Zeitschrift für Erziehungswissenschaft*. doi:[10.1007/s11618-017-0749-z](https://doi.org/10.1007/s11618-017-0749-z)
- Nikolova, R. (2011). Lesekompetenz und Einstellungen zum Deutschunterricht. In U. Vieluf, S. Ivanov & R. Nikolova (Hrsg.), *KESS 10/11. Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen am Ende der Sekundarstufe I und zu Beginn der gymnasialen Oberstufe* (S. 31–73). Hamburg: Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung.
- OECD (2014). *PISA 2012 results: What students know and can do student performance in mathematics, reading and science*. Paris: OECD.
- OECD (2016). *PISA: Low-performing students: Why they fall behind and how to help them succeed*. Paris: OECD Publishing.
- Pfost, M., Karing, C., Lorenz, C., & Artelt, C. (2010). Schereneffekte im ein- und mehrgliedrigem Schulsystem: Differenzielle Entwicklung sprachlicher Kompetenzen am Übergang von der Grund- in die weiterführende Schule? *Zeitschrift für Pädagogische Psychologie*, 24, 259–272.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R. H., Leutner, D., Neubrand, M. et al. (Hrsg.). (2006). *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS ONE*, 7(7), e39904.

- Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: a longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology*, 82, 647–671.
- Retelsdorf, J., & Köller, O. (2014). Reciprocal effects between reading comprehension and spelling. *Learning and Individual Differences*, 30, 77–83.
- Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, 21, 550–559.
- Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, 48, 268–302.
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. Bd. 81. New York: John Wiley & Sons.
- Rychen, D. S., & Salganik, L. H. (Hrsg.). (2003). *Key competencies for a successful life and a well functioning society*. Göttingen: Hogrefe & Huber.
- Van Gelderen, A., Schoonen, R., Stoel, R. D., De Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, 99, 477–491.
- Watermann, R., & Klieme, E. (2002). Reporting results of large-scale assessment in psychologically and educationally meaningful terms: Construct validation and proficiency scaling in TIMSS. *European Journal of Psychological Assessment*, 18, 190–203.
- Weirich, S., Hecht, M., & Böhme, K. (2014). Modeling item position effects using generalized linear mixed models. *Applied Psychological Measurement*, 38, 535–548.
- Zabal, A., Martin, S., Klaukien, A., Rammstedt, B., Baumert, J., & Klieme, E. (2013). Grundlegende Kompetenzen der erwachsenen Bevölkerung in Deutschland im internationalen Vergleich. In B. Rammstedt (Hrsg.), *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich. Ergebnisse von PIAAC 2012* (S. 31–76). Münster: Waxmann.