

„Bildung und Bildungsungleichheit an Halb- und Ganztagschulen“

Rolf Strietholt · Veronika Manitius · Nils Berkemeyer · Wilfried Bos

Online publiziert: 27. Mai 2015
© Springer Fachmedien Wiesbaden 2015

Zusammenfassung Der intensive Ausbau ganztägiger Bildung und Betreuung ist eines der kostspieligsten und meist diskutierten Schulreformprogramme der vergangenen Jahre. Ein erklärtes Ziel ist eine Verbesserung von Schülerkompetenzen und der Abbau von Bildungsungleichheit. Der empirische Beleg hierfür steht aber noch aus. Hieran anknüpfend widmet sich die vorliegende Arbeit der Frage, ob die Organisationsform von Schulen einen Effekt auf Schülerkompetenzen hat. Hierzu reanalysieren wir Daten aus IGLU, TIMSS und PISA und vergleichen das Leistungsniveau und herkunftsbedingte Disparitäten an Halb- und Ganztagschulen. Für einen fairen Vergleich der Schulen beider Organisationsformen matchen wir zunächst Schulen, die sich nicht in Bezug auf die Schülerkomposition, Region und Schulform unterscheiden. Die Ergebnisse unserer Mehrebenenanalysen lassen sich so zusammenfassen, dass wir an Ganztagschulen weder ein höheres Leistungsniveau, noch eine Reduktion von Bildungsungleichheiten beobachten.

Schlüsselwörter Ganztag · Schülerleistung · PISA · TIMSS · IGLU

Dr. R. Strietholt (✉) · Prof. Dr. W. Bos
Institut für Schulentwicklungsforschung, TU Dortmund,
Vogelpothsweg 78,
44227 Dortmund, Deutschland
E-Mail: strietholt@ifs.tu-dortmund.de

Prof. Dr. W. Bos
E-Mail: officebos@ifs.tu-dortmund.de

Dr. V. Manitius · Prof. Dr. N. Berkemeyer
Fakultät für Sozial- und Verhaltenswissenschaft, Friedrich-Schiller-Universität Jena,
Fürstengraben 11,
07443 Jena, Deutschland
E-Mail: veronika.manitius@uni-jena.de

Prof. Dr. N. Berkemeyer
E-Mail: nils.berkemeyer@uni-jena.de

Educational outcomes and educational inequalities at half- and all-day schools

Abstract The massive expansion of all-day schools that can be observed recently is one of the most costly and most widely discussed school/education reform programs over the last years. One of the declared aims is to improve the students competencies and reduce educational inequalities. However, there is little empirical research that provides evidence in this respect. Against this background, the present study investigates whether the organizational form of schools have an effect on student achievement. Therefore, we use/reanalyse data from PIRLS, TIMSS, and PISA and compare levels of performance and origin-related disparities at half- and all-day schools. To have a fair comparison of both types of schools we match half- and all-day schools based on their social composition, region, and school type. The results of our multilevel analyses indicate that there are no effects at all-day schooling on levels of achievement and educational inequalities.

Keywords All-day school · PISA · PIRLS · Student performance · TIMSS

1 Hintergrund

Im Zuge der bildungspolitischen Reaktionen auf das schlechte Abschneiden der Schülerinnen und Schüler in Deutschland in der ersten PISA-Erhebung 2000 wurde auch der Auf- und Ausbau des Ganztags als Unterstützungsprogramm forciert (KMK 2002; Kuhlmann und Tillmann 2009). Welche Erwartungen mit dem Ganztags verknüpft sind, welche theoretischen Annahmen ihm zugeschrieben werden und inwiefern sich diese empirisch bestätigen, ist Gegenstand des Beitrags. Dabei wird der Fokus auf die Effekte von ganztägiger Bildung und Betreuung auf die Schülerleistung und die Kompensation von Benachteiligung gelegt. Nachfolgend werden wir zunächst programmatische Zuschreibungen an Ganztagschulen und theoretische Annahmen hinter den an sie gerichteten Erwartungen referieren, um danach den empirischen Forschungsstand zu den Wirkungen von Ganztagschule auf die Leistungen von Schülerinnen und Schülern, und im Hinblick auf die Kompensation von Benachteiligung darzustellen.

1.1 Programmatische Zuschreibungen und theoretische Überlegungen

In dem bildungspolitischen Diskurs um die Konsequenzen aus den Befunden rezenter Schulleistungsstudien wurde mit dem Start des „Investivprogramms Zukunft Bildung und Betreuung“ (IZBB) auch der Ausbau von Ganztagschulen und die Entwicklung von ganztagsschulischen Förderkonzepten forciert. Im Rahmen dieses Programms stellte die Bundesregierung vier Milliarden Euro für den Auf- und Ausbau der Ganztagsinfrastruktur (2003 bis 2009) zur Verfügung, ergänzend investieren die Bundesländer finanzielle Eigenleistungen von ca. 400 Millionen Euro in Konzeptentwicklungen und Zusammenführungen des Modells Ganztagschule mit anderen Reformprojekten (vgl. BMBF 2009). Diese hohen Fördersummen implizieren bereits

die substanziellen sozial- und bildungspolitischen Hoffnungen, die mit Ganztagschulen verknüpft sind. Die Begründungslinien für ihre Einführung heben vor allem auf sich ändernde Bildungsaufgaben im Zuge komplexer werdender gesellschaftlicher Anforderungen ab. Dies bezieht sich auf ein Verständnis von Ganztagschulen, die einen Beitrag zur Sicherstellung der soziokulturellen Infrastruktur leisten sollen (z. B. der Vereinbarkeit von Familie und Beruf entgegenkommen); die zur Stärkung der sozialerzieherischen Funktion von Schule (z. B. die Schaffung von Angeboten für die Praxis von Partizipation und Demokratie lernen) und zur Verbesserung von Schulqualität beitragen soll (z. B. die Entwicklung einer Lernkultur, die die Differenzierung von Lernarrangements im Unterricht stärkt) und die auf die Wahrung der Bildungschancen aller Kinder und Jugendlichen abhebt (z. B. die Intensivierung der Förderung vor allem auch der Lernschwächeren) (vgl. Radisch und Klieme 2004; Holtappels 2006, 2009). Anhand der bildungspolitischen Begründungslinien, die sich im Wesentlichen aus sozial-, arbeitsmarkt- und bildungsspezifischen Argumenten zusammensetzen (vgl. Appel 2004) wird ersichtlich, dass die Leistung von Ganztagschule eng mit gesellschaftlichen Bedürfnissen verknüpft wird, für die Ganztagschulen nicht die einzige, aber eine adäquate Problemlösung darstellen soll. Allerdings stellt Ganztagschule kein reines „Vollzug-Instrument“ (Ipfling 2005, S. 300) für die Bearbeitung gesellschaftlicher Entwicklungen dar. Bildungstheoretisch wird vielmehr argumentiert, dass Ganztagschule in einem Wechselverhältnis zu den gesamtgesellschaftlichen Entwicklungen steht, indem sie einerseits Angebote zur Bearbeitung gesellschaftlicher Problemlagen schafft, andererseits aber auch vor dem Hintergrund ihrer Zielsetzung als Schule, nämlich die Mündigkeit des Subjekts zu befördern (vgl. Klafki 1991), sich kritisch daran orientieren muss, dass ihre Angebote pädagogische Ansprüche erfüllen, Einzelschule also auch selbst Initiator ihrer eigenen pädagogischen Gestaltung ist.

Sowohl die bildungstheoretischen, als auch schultheoretischen Auseinandersetzungen mit der Ganztagschule beschreiben diese in erster Linie funktional als über die generell beschriebenen Funktionen von Schule (Halbtagschule) hinausgehend. Dabei soll Ganztagschule gleichwohl „Schule“ bleiben (vgl. Rekus 2005); also die auf Teilhabe an Gesellschaft bezogenen Funktionen von Schule erfüllen (Integration, Allokation, Qualifikation und Enkulturation, vgl. Fend 2006). Durch neu gewonnene Zeitkapazitäten im Vergleich zur Halbtagsbesuchung besteht eine spezifische theoretische Annahme nun darin, dass durch eine effiziente Umorganisation von Schule, die diesen neuen Gestaltungsrahmen nutzt, konkreten gesellschaftlichen Interessenlagen begegnet werden kann (vgl. Rekus 2005). Gemeint ist hierbei u. a. die Möglichkeit, durch Ganztagschule verstärkt lebensweltliche Bezüge zu berücksichtigen: Indem sie beispielsweise die Übernahme der Hausaufgabenbetreuung professionell übernimmt und damit die Eltern als „Nicht-Profis“ entlastet, weist sie einen familienergänzenden Charakter auf und bedient gleichzeitig ein arbeitsmarktspezifisches Interesse, nämlich die verbesserte Vereinbarkeit von Familie und Beruf. Eine kompensatorische Wirkung wird der Ganztagschule zudem im Hinblick auf Familien mit nicht hinreichenden Erziehungskompetenzen zugeschrieben, wenn sie Erziehungshilfen anbietet (vgl. Scholz und Reh 2009).

Neuere Arbeiten versuchen sich in der gerechtigkeits-theoretisch fundierten Beschreibung von Ganztagschule, wonach der Ganztagschule aus einer system-

orientierten Perspektive heraus so genannte „Integrationskraft“ zugesprochen wird, indem sie zusätzliche zeitliche Möglichkeitsräume für die soziale Integration von Kindern bietet. Hier wird die Annahme plausibilisiert, dass im Zuge des Ganztags nicht allein schulisch geprüfte Fähigkeiten weiterentwickelt werden, sondern auch Fähigkeiten kultureller und politischer Teilhabe und so persönliche Kontakte und Freundschaften im Raum ganztägiger Bildung eine neue und umfanglichere Bedeutung für das Erfahren von (gerechtigkeitstheoretisch bedeutsamen) Anerkennungsprozessen erhalten können (vgl. Berkemeyer et al. 2012; Berkemeyer und Manitus 2013).

In schul-, bildungs- und gerechtigkeitstheoretischen, sowie pädagogischen Argumentationslinien wird das Verhältnis von Ganztagsschulen zur Gesellschaft diskutiert. Dabei wird das Beziehungsgeflecht zwischen Ganztagsschule und konkreten gesellschaftlichen Teilbereichen (wie z. B. Familie, Arbeitsmarkt, Freizeit) modelliert, und die Funktionen und Leistungen von Ganztagsschule erörtert. Gleichzeitig lässt sich aber aus allen genannten theoretischen Richtungen argumentieren, dass Ganztagsschule letztlich auch Schule ist und somit auch die klassischen Funktionen von Schule erfüllen muss: Mit Blick etwa auf die Qualifikationsfunktion, womit auch die Leistungen von Schülerinnen und Schülern einbezogen sind, ist es hier vor allem die Bildungspolitik, die diesbezüglich Hoffnungen und Erwartungen artikuliert, dass explizit Ganztagsschule einen Beitrag zur Verbesserung der Kompetenzen von Kindern und Jugendlichen erbringt (Ahn 2005; Schavan 2005). Es wird argumentiert, dass Ganztagsschulen potenziell mehr Lernzeit zur Verfügung steht und damit auch größere Gestaltungsmöglichkeiten, um Leistungen der Schülerinnen und Schüler zu verbessern. Von dem längeren Verbleib im schulischen Kontext sollen zudem auch Schülerinnen und Schüler mit Migrationshintergrund profitieren, da sie mehr Zeit in einem deutschsprachigen Umfeld verbringen (vgl. Merckens und Schröder-Lenzen 2010).

Mit Blick auf die konkreten Wirkungen von Ganztagsschule finden sich in den entsprechenden theoretisch-konzeptionellen Arbeiten vorwiegend Wirkungsmodelle von Ganztagsschule, in denen in Kontext-, Input-, Prozess-, und Outputdimensionen zur Ganztagsschule ihre Ergebnisqualität als eine Outputvariable verortet wird (vgl. Fischer et al. 2011). Diese Modellierung entspricht üblichen Effektivitätsvorstellungen zur Schule (vgl. Stufflebeam 1972; Creemers et al. 2000; Creemers und Kyriakides 2008). Die Erweiterung des Bildungsmodells zur Ganztagsschule hinsichtlich des Outputs erfolgt nun, indem neben den Wirkungen auf Schülerebene (Leistung) auch zusätzlich noch erzieherische Wirkungen hinsichtlich etwa der Entwicklung von sozialen Kompetenzen betont werden (vgl. Fischer et al. 2011). In seinen Arbeiten zu einem umfassenderen Wirkungsmodell von Ganztagsschule kommt Radisch (2009) zu dem Schluss, dass ganztagschulische Wirkung sich nicht allein über das zeitliche „Mehr“ gegenüber Halbtagschule ausmachen lässt, sondern vielmehr in Wirkungsmodellen von Ganztagsschule stärker Umfeldfaktoren, die Angebotsstrukturen und -gestaltung und ihre Rolle auf ausdifferenzierte Wirkungen (fachleistungsbezogen, fachübergreifend, fachunabhängig) einzubeziehen sind. Folglich werden hier auch explizit andere Wirkungsbereiche, als ausschließlich die Fachleistung betont.

Gleichwohl ist sowohl in den bildungspolitischen Erwartungen, als auch den bisherigen theoretisch-konzeptionellen Ansätzen zum schulischen Ganztag die

Annahme zu entnehmen, „dass Ganztagschulen einen konkreten und keinesfalls nur programmatischen Kontext bieten, in dem sich eine neue individualisierende sowie Bildung und Erziehung zusammenführende Praxis etablieren kann“, (Maykus et al. 2011, S. 128) wobei die Auswirkung von Ganztagschule auf Schülerleistung ein expliziter Wirkungsbereich sein soll. Was nun tatsächlich Leistungen und Wirkungen von Ganztagschulen sind, ist inzwischen Gegenstand vielzähliger Forschung.

1.2 Forschung zu den Effekten von Ganztagschule

Im Rahmen der Forschung zu ganztägiger Bildung und Betreuung wurde mittlerweile eine Reihe von Forschungsarbeiten vorgelegt. Der empirische Nachweis für den Beitrag von deutschen Ganztagschulen zur Verbesserung der Leistungen von Schülerinnen und Schüler auf Grundlage repräsentativer Daten steht bislang jedoch weitgehend aus. Kleinere Studien finden jedoch Hinweise auf einen entsprechenden Effekt der Ganztagschule.

Insgesamt hat sich mittlerweile eine über mehrere Institute getragene breite Forschungslandschaft entwickelt, die differenzierte Erkenntnisse zum schulischen Ganztag bereithält (z. B. Fischer et al. 2011; Steiner 2011; Stecher et al. 2011; Rauschenbach et al. 2012; Züchner und Fischer 2014; Willems und Holtappels 2014).

Mit Fokus auf die Wirkungen von Ganztagschule im Hinblick auf eine Verbesserung der Schülerleistungen und ganztagsschulischer Kompensationsleistungen – bezüglich der Benachteiligung von Kindern und Jugendlichen – stellt sich die Forschungslage jedoch nicht eindeutig dar. In querschnittlichen Untersuchungen haben bisherige Vergleiche der Fachleistungen von Schülerinnen und Schülern an Halb- und Ganztagschulen keine Wirkungen der Ganztagschule nachweisen können. Radisch et al. finden in ihren Analysen der IGLU-Daten von 2001 für den Primarbereich keinen signifikanten Unterschied in den Kompetenzen von Halb- und Ganztagschülern (vgl. Radisch et al. 2006). Mittels Daten aus IGLU 2006 konnten Holtappels et al. (2010) zwar ein besseres Abschneiden in der Leseleistung bei Schülern ohne intensive Teilnahme an Ganztagsangeboten ermitteln, dieser Befund wurde jedoch insbesondere über Selektionseffekte erklärt (Ganztags*teilnahme* als remediales Förderinstrument innerhalb von Schulen). Sichtbar wurde in dieser Untersuchung, dass der Ganztag offenbar verstärkt von Kindern aus benachteiligten Elternhäusern genutzt wird. Neuere Untersuchungen aus den letzten StEG-Daten zeigen jedoch insgesamt ein differentes Bild hinsichtlich der sozialen Selektivität in der Ganztags*teilnahme*: Demnach bestehen vor allem im Grundschulbereich herkunftsspezifische Differenzen; insbesondere wird der Ganztag frequentiert von Kindern, die nicht aus „Kernfamilien“ stammen, wo also kein Elternteil Hausfrau oder Hausmann ist (vgl. Steiner 2011).

Um längsschnittlich generierte Befunde zum Ganztag zu betrachten, erscheint der Rückgriff auf die größte Studie zum Ganztag, StEG, naheliegend. Diese sah aber bislang keine Erhebung von Leistungsdaten vor und weist auch kein Kontrollgruppendesign auf. Allerdings wurde in StEG die Entwicklung des Notenerwerbs analysiert. Die Autoren stellten fest, dass eine häufigere Teilnahme (mehrere Tage die Woche) am Ganztag eine förderliche Entwicklung auf die Schulnoten hat (vgl. Kuhn und Fischer 2011). Auch findet sich ein schwächerer Zusammenhang zwischen der sozia-

len Herkunft und der Mathematiknote an gebundenen Ganztagschulen (Züchner und Fischer 2014). Problematisch bleibt jedoch die Messung der schulischen Performanz über Schulnoten, was eher als gering valide einzuschätzen ist und keine Aussage darüber erlaubt, ob hinter einer positiven Notenentwicklung auch ein Kompetenzzuwachs steht.

Für die Betrachtung des Kompetenzerwerbs an Ganztagschulen in längsschnittlichen Designs liegen bislang nur wenige Befunde aus kleineren Studien vor. Für den Grundschulbereich lassen sich im Vergleich der Leseleistung von einer nicht repräsentativen Stichprobe Berliner Halbtags- und offenen Ganztagschulen zwar höhere Kompetenzwerte der Schülerinnen und Schüler mit Ganztagssteilnahme finden, allerdings handelt es sich nach Einschätzung der Autoren hierbei um eine positiv selektierte Gruppe, die bereits zum ersten Messzeitpunkt (Ende des ersten Schuljahres) eine höhere Lesefähigkeit aufweisen (vgl. Bellin und Tamke 2010). Für Schweizer Ganztagsgrundschulen untersuchte Schüpbach (2014) ebenfalls längsschnittlich angelegt die Leistungsentwicklung von ganz- und halbtägig beschulten Kindern. Während zum ersten Messzeitpunkt (Ende des ersten Schuljahres) keine Unterschiede in der Leseleistung zu finden waren, zeigte sich ein höherer Zuwachs in der Entwicklung der Leistung der Kinder im Ganztags am Ende des dritten Schuljahres. Es wurde außerdem der Frage nachgegangen, ob der Ganztags kompensatorische Effekte hat. Dies zeigte sich jedoch nicht; es fanden sich keine Effekte des Ganztags auf die Kopplung des Zusammenhangs von sozialer Herkunft und Schulleistung (Schüpbach 2014). Insgesamt handelt es sich in dieser Studie um eine Stichprobe mit überproportional hohem Anteil an Kindern aus Elternhäusern mit hohem sozioökonomischen Hintergrund.

Insbesondere fokussiert auf den Erwerb sprachlicher Kompetenzen von Kindern mit Migrationshintergrund, fragen Reinders et al. (2011) danach, ob Ganztagschule hier eine bessere Förderung leistet. Die Autoren finden keinen signifikanten Unterschied in der Leseleistung von Halbtags- und Ganztagschülern. Ebenso zeigten sich keine Auswirkungen von konkret implementierten Sprachförderangeboten auf die Leseleistung. Jedoch scheint die Zusammensetzung der Schülerschaft ab einem Migrationsanteil von über 70% einen negativen Effekt auf die Veränderungen der Leseleistung der Schülerinnen und Schüler zu haben. Auch hier kommen die verantwortlichen Forscher zu Einschränkungen hinsichtlich der Aussagekraft ihrer Befunde, da sich Selektionseffekte durch die Freiwilligkeit in der Studienteilnahme ergaben.

Wenn Variablen wie der Migrationshintergrund und auch die soziale Lage statistisch kontrolliert würden, ließe sich dann immer noch ein Unterschied zwischen Ganztags- und Halbtagsschulen feststellen? Untersuchungen der IGLU Studie 2006 zeigen, dass bei Kontrolle der Schulkomposition in Bezug auf den sozioökonomischen Hintergrund, den Anteil an Migranten sowie den durchschnittlichen kognitiven Fähigkeiten (Kognitiver Fähigkeitstest, KFT) keine Unterschiede mehr zwischen Ganztags- und Halbtagsschülerinnen und -schülern erkennbar sind (vgl. Holtappels et al. 2010). Die Aufnahme des KFTs als Kontrollvariable auf Schulebene kann allerdings auch kritisch betrachtet werden, da Schule nicht nur Fähigkeiten in einzelnen Domänen bzw. Fächern beeinflusst, sondern auch kognitive Fähigkeiten im Allgemeinen.

2 Fragestellungen der Studie

- Vor dem Hintergrund der hohen Erwartungen, die an Ganztagschulen gerichtet werden, ist es erstaunlich, dass zu den Effekten ganztägiger Bildung und Betreuung auf Schülerleistungen kaum Forschungsbefunde vorliegen. Die Leitfragen der vorliegenden Arbeit schließen an dieses Forschungsdesiderat an und fragen nach den Effekten der Organisationsform auf das Leistungsniveau, sowie auf herkunftsbedingte Disparitäten innerhalb der Schule. Folgende Teilfragen werden separat für die Grundschule und für weiterführende Schulen in den Domänen Lesen, Mathematik und Naturwissenschaften untersucht: Hat die Organisationsform Ganztags vs. Halbtag einen Effekt auf das Leistungsniveau von Schulen?
- Hat die Organisationsform einen Effekt auf die Kopplung zwischen sozioökonomischem Hintergrund und Schülerleistung in Schulen, sowie innerschulische Leistungsunterschiede zwischen Kindern mit und ohne Migrationshintergrund?

Um unser Untersuchungsdesign zu beschreiben, scheint es zunächst sinnvoll zwischen Analysen auf Individual- oder Organisationsebene zu unterscheiden. Diese Unterscheidung ist sowohl inhaltlich wie auch methodisch bedeutsam. Aus inhaltlicher Perspektive handelt es sich um eine bedeutsame Unterscheidung, da im Rahmen der ganztägigen Bildung und Betreuung nicht alle Angebote verpflichtend sind. Eine der Grundideen von Ganztagschule ist es, individuelle Fördermöglichkeiten zu schaffen und Wahlmöglichkeiten zu bieten. In Analysen auf Individualebene kann nun der Frage nachgegangen werden, welche spezifischen Effekte die von Schülerinnen und Schülern *genutzten* Ganztagsangebote haben. Dabei können auch bestimmte Qualitätsmerkmale betrachtet werden, die in der Schulrealität kaum implementiert sind. In diesem Kontext sei darauf verwiesen, dass die zusätzlichen Zeitbudgets an Ganztagschulen ganz unterschiedliche genutzt werden können. Demnach sind Studien zu den Effekten *spezifischer* Qualitätsmerkmale keinesfalls uninteressant, vielmehr können sie wichtige Erkenntnisse zur Weiterentwicklung von Ganztagschulen liefern. Gleichzeitig können Studien zu einzelnen Qualitätsmerkmalen keine Antwort auf die Frage geben, welchen *Gesamteffekt* der Ausbau von Ganztagschulen in der Realität hat. In der vorliegenden Studie interessieren wir uns (im Kontext des massiven Ausbaus von Ganztagschulen) genau für diese Frage. Letztlich ist das IZBB ein Reformprogramm, das auf Schulebene ansetzt. Daher schätzen wir den Gesamteffekt auf Schulebene.

Die Wahl der Analyseebene ist auch aus methodischer Sicht bedeutsam, denn die Identifikation der Effekte des Ganztags auf Organisationsebene hat den Vorteil, dass Selektionsprozesse auf Individualebene qua Design umgangen werden. In der Regel werden innerhalb von Schulen insbesondere leistungsschwache Schülerinnen und Schüler gezielt gefördert, es liegen aber selten differenzierte Informationen dazu vor, wie einzelne Schülerinnen und Schüler individuell gefördert werden. Wenn man diesen Selektionsprozess nicht berücksichtigen kann, wird der Effekt von Ganztags unterschätzt. Analysen auf Aggregatebene umgehen solche Selektionsmechanismen innerhalb der Schulen. Allerdings können auf Schulebene andere Selektionsprobleme auftreten und die Analysen verzerren, wenn etwa insbesondere in sozialen Brennpunkten (als gezielte Interventionsmaßnahme) Halbtagsschulen in Ganztagschulen

umgewandelt werden (vgl. Kuhlmann und Timmann 2009). Stuart (2007) diskutiert in diesem Kontext die Möglichkeit mithilfe des Propensity Score Matchings (PSM) vergleichbarer Schulen auszuwählen, die sich in Bezug auf andere Kovariaten nicht voneinander unterscheiden. Aufgrund früherer Forschungsbefunde und theoretischer Erwägungen berücksichtigen wir als Kovariaten die Schülerkomposition in den Schulen, das Bundesland und in der Sekundarstufe zudem die Schulform (vgl. Abschn. 3.5. Analyseplan).

3 Anlage und methodisches Vorgehen

3.1 Daten

Für unsere Analysen nutzen wir repräsentative Daten aus der deutschen Beteiligung an drei großflächig angelegten Schulleistungstudien, die 2006 und 2007 in Deutschland durchgeführt wurden: In TIMSS 2007 und IGLU 2006 wurden Kinder der vierten Jahrgangsstufe getestet und in PISA-E 2006¹ die 15-jährigen Jugendlichen (Bos et al. 2007, 2008; Prenzel et al. 2008). Neben den Leistungstests wurden den Schülerinnen und Schülern, sowie deren Eltern zusätzliche Hintergrundfragebögen vorgelegt. Die Stichprobe von TIMSS 2007 umfasst 5200 Schülerinnen und Schüler aus 247 Schulen, an IGLU 2006 nahmen 7899 Schülerinnen und Schülern aus 405 Schulen teil und die PISA 2006-E Stichprobe basiert auf 33.809 15-jährigen aus insgesamt 1460 Schulen.

3.2 Variablen auf Individualebene

Schülerleistung. Die Schülerinnen und Schüler wurden in IGLU im Lesen, in TIMSS in Mathematik und Naturwissenschaften und in PISA in allen drei Domänen getestet. Für die Schülerleistung wurden jeweils fünf Plausible Values geschätzt.

Sozioökonomischer Status und Migrationshintergrund. Der sozioökonomische Status der Schülerinnen und Schüler wurde mithilfe des HISEI (Highest International Socio-Economic Index of Occupational Status) im Elternhaus erfasst (Ganzeboom et al. 1992), in dem Berufe in Bezug auf das notwendige Ausbildungsniveau und das typische Einkommen klassifiziert werden. Von den Rohwerten wurde 40 abgezogen, sodass ein transformierter Wert von 0 einem HISEI von 40 (=z. B. Verkäufer) entspricht; die Transformation macht die Analysen inhaltlich besser interpretierbar. Als Indikator für den Migrationshintergrund verwenden wir das Geburtsland der Eltern. Die Referenzkategorie sind Kinder, von denen beide Eltern in Deutschland geboren wurden. Wir unterscheiden zudem zwischen Kindern, von denen ein Elternteil, oder beide Elternteile im Ausland geboren wurden.

3.3 Variablen auf Schulebene

Organisationsform. In der vorliegenden Studie verwenden wir die KMK-Definition von Ganztagschulen und definieren eine Schule als Ganztagschule, wenn die

Schülerinnen und Schüler verpflichtend an mindestens drei Tagen pro Woche am Ganztagsbetrieb teilnehmen. Diese Information wurde in allen drei Studien bei den Schulleitungen erfragt. An TIMSS nehmen 52 Ganztagschulen teil (insgesamt 247 Schulen), bei IGLU 2006 sind es 80 Ganztagschulen (insgesamt 405 Schulen) und bei PISA 2006-E 294 Ganztagschulen (insgesamt 1460 Schulen). Gleichzeitig sei darauf verwiesen, dass sich hinter dieser Minimaldefinition eine gewisse Heterogenität verbirgt, denn es gibt beispielsweise viele offene, aber nur wenige vollgebundenen Ganztagsgrundschulen (vgl. Tab. 1). In den von uns berücksichtigten Stichproben nehmen auch nicht alle Schülerinnen und Schüler an speziellen Förderangeboten teil. In der Grundschule nehmen jeweils etwa 30% der Schülerinnen und Schüler an Hausaufgabenhilfen teil; die Teilnahmequote an Fördergruppen ist ähnlich hoch. Letztlich sei noch darauf verwiesen, dass ein gewisser Anteil der Schulen erst vor wenigen Jahren den Ganztagsbetrieb aufgenommen hat. Wir zählen diese Schulen trotzdem zu den Ganztagschulen, da die Schülerinnen und Schüler zumindest in der vierten Klasse eine Ganztagschule besucht haben.²

Schulkomposition. Um die Schülerkomposition an den einzelnen Schulen möglichst umfassend zu beschreiben, greifen wir auf eine ganze Reihe von Merkmalen zurück, die sich im Rahmen der Ungleichheits- und Schuleffektivitätsforschung als erklärungsstarke Prädiktoren für Leistungsunterschiede gezeigt haben. Hierzu gehören soziale, kulturelle und ökonomische Kapitalien der Kinder; das Geschlecht, Migrationshintergrund und Sprache, sowie der Besuch frühkindlicher Bildungseinrichtungen (vgl. Bourdieu 1983; Wagemaker 1996; Shonkoff und Phillips 2000; Stanat und Kunter 2001; Baumert et al. 2006; Esser 2006; Konsortium Bildungsberichterstattung 2006; Cunha und Heckman 2007). Diese Indikatoren wurden bei den Kindern oder deren Eltern erfragt.

Tab. 1 Ausgewählte Merkmale des Ganztagsbetriebs

	PIRLS 06	TIMSS 07	PISA-E 06
<i>Form des Ganztagsbetriebs^a</i>			
voll gebundene	13,4%	7,8%	32,3%
teilweise gebunden	18,3%	28,9%	26,7%
offene	68,3%	63,3%	41,0%
<i>Im Ganztagsbetrieb seit mindestens^a</i>			
1 Jahr	29,8%	29,9%	–
2 Jahren	15,5%	23,0%	–
3 Jahren	13,1%	8,0%	–
>3 Jahren	41,7%	39,1%	–
<i>Teilnahme an ausgewählten Förderangeboten^b</i>			
Hausaufgabenhilfe	29,4%	31,8%	–
Fördergruppe	29,9%	31,3%	–

eigene Berechnungen mit ungewichteten Daten.

^aSchulleiterangaben in allen Studien.

^bSchülerangaben in PIRLS und Elternangaben in TIMSS, angegeben sind jeweils Anteile der Schülerinnen und Schüler, die an mindestens einem Tag pro Woche die entsprechenden Förderangeboten wahrnehmen.

Schulform und Bundesland. Während der Stichprobenziehung wurde in allen Studien das Bundesland vermerkt. Zusätzlich wurde in der Sekundarstufenstudie PISA zusätzlich der Schultyp vermerkt (Hauptschule, Realschule, Gymnasium, Integrierte Gesamtschule, Schulen mit mehreren Bildungsgängen, Berufliche Schule).

Vorleistung. Des Weiteren liegen in IGLU und PISA Informationen über das Leistungsniveau der Schülerschaft vor der Einschulung bzw. vor dem Wechsel in die Sekundarstufe I vor. In IGLU wurden den Eltern fünf Fragen vorgelegt, in denen danach gefragt wurde, ob die Kinder bereits vor der Einschulung einzelne Buchstaben, Wörter oder sogar Sätze lesen und schreiben konnten. In PISA wurde erfasst, ob die Kinder nach der Grundschule eine Übergangsempfehlung für das Gymnasium erhalten haben. In TIMSS liegen keine vergleichbaren Informationen vor. Wenngleich die Elterneinschätzung und die Übergangsempfehlung messfehlerbehaftet sind, so bieten sie zumindest einen Proxy für die Vorleistung. Beispielweise finden Maaz und Nagy (2009) Korrelationen in Höhe von $r > 0,50$ zwischen Gymnasialempfehlung und Schülerleistung (in Mathematik und Naturwissenschaften) am Ende der Grundschule. Für die retrospektiv eingeschätzte vorschulische kindliche Fähigkeit liegen keine Informationen zu Zusammenhängen mit standardisierten Tests vor. Allerdings berichten Gustafsson et al. (2013), dass vorschulische kindliche Fähigkeiten nach Kontrolle von elterlicher Bildung und Büchern im Elternhaus in Deutschland einen geringen, aber signifikanten Effekt auf die Leseleistungen am Ende der Grundschule haben (standardisierter Pfadkoeffizient: .16).

Um die Schulkomposition zu beschreiben, wurden die Angaben der Kinder und Eltern pro Schule aggregiert. Bei kategorialen Variablen wurden zunächst Dummyvariablen gebildet und in einem zweiten Schritt der Mittelwert für die Schulen berechnet (=Anteile); für kontinuierliche Variablen wurde direkt der Mittelwert pro Schule berechnet. Tab. 2 bietet eine Übersicht, welche Variablen in den drei Studien verfügbar sind.³

3.4 Fehlende Daten

Einige Schulleiterinnen und -leiter haben keine Angaben zur Organisationsform der Schule gemacht bzw. den Schulfragebogen überhaupt nicht ausgefüllt. Diese Schulen wurden aus den Analysen ausgeschlossen. Fehlende Angaben einzelner Schülerinnen und Schüler oder ihrer Eltern wurden mit dem R-Paket MICE imputiert (Van Buuren und Groothuis-Oudshoorn 2009). Neben den oben beschriebenen Variablen zur Schulkomposition wurden auch die Leistungsscores der Schülerinnen und Schüler in das Imputationsmodell aufgenommen. Anhand der vollständigen Datensätze wurden die Daten der Individualebene wie zuvor beschrieben aggregiert, um die Schulkomposition zu beschreiben.

3.5 Analysestrategie

Um den Effekt der Organisationsform auf Schülerleistung zu schätzen, müssen Kompetenzen an Halb- und Ganztagschulen beobachtet werden. Dies kann am besten im Rahmen von Mehrebenenmodellen beschrieben werden. In Modell (1) hängt die

Tab. 2 Übersicht über die verwendeten Variablen

Studie	PIRLS 06	TIMSS 07	PISA-E 06
Leistungsscore: Lesen	√	–	√
Leistungsscore: Mathematik	–	√	√
Leistungsscore: Naturwissenschaften	–	√	√
Organisationsform	√	√	√
Geschlecht	√	√	√
Sozioökonomischer Status (HISEI)	√	√	√
Elterliche Bildung	√	√	√
Geburtsland Kind	√	√	√
Geburtsland Eltern	√	√	√
Familiensprache nicht Deutsch	√	√	√
Bücher im Elternhaus	√	√	√
Bundesland	√	√	√
Familieneinkommen	√	√	–
Dauer Besuch Kindergarten	√	√	–
Besuch kultureller Veranstaltungen	√	√	–
Funktionieren des familiären Netzwerks	√	√	–
Vorschulische schriftsprachliche Fähigkeiten	√	–	–
Bundesland	√	√	√
Schulform	–	–	√
Vorschulische schriftsprachliche Fähigkeiten	√	–	–
Übergangsempfehlung: Gymnasium	–	–	√

Die aufgeführten Merkmale wurden in den drei Studien mithilfe unterschiedlicher Fragen erhoben.

Leistung y von Schüler i aus Schule s von seinem Migrationsstatus und sozioökonomischen Hintergrund ab. Mit dem Parameter β_{0s} schätzen wir die Leistung einer Schülerin oder eines Schülers mit einem HISEI von 40 (s. 3.2.2. zur Transformation) und ohne Migrationshintergrund, mit β_{1s} schätzen wir, um wie viel Punkte die Leistung steigt, wenn der HISEI um einen Punkt steigt, β_{2s} bzw. β_{3s} beschreiben den Leistungsunterschied zwischen Kindern ohne Migrationshintergrund und Kindern von denen ein Elternteil bzw. beide Eltern im Ausland geboren wurden.

$$y_{is} = \beta_{0s} + \beta_{1s} HISEI_{is} + \beta_{2s} Mig1_{is} + \beta_{3s} Mig2_{is} + e_{is} \quad (1)$$

Um den Effekt der Organisationsform zu untersuchen, modellieren wir β_{0s} , β_{1s} , β_{2s} und β_{3s} als abhängige Variablen auf Schulebene und regressieren sie auf die Organisationsform. Die Variable GT beschreibt die Organisationsform und ist 0, wenn die Schule eine Halbtagsschule ist und 1, wenn es sich um eine Ganztagschule handelt:

$$\beta_{0s} = \gamma_{00} + \gamma_{01} GT_s + u_0 \quad (2)$$

$$\beta_{1s} = \gamma_{10} + \gamma_{11} GT_s + u_1 \quad (3)$$

$$\beta_{2s} = \gamma_{20} + \gamma_{21} GT_s + u_2 \quad (4)$$

$$\beta_{3s} = \gamma_{30} + \gamma_{31}GT_s + u_3 \quad (5)$$

Die vier β -Koeffizienten in Formel (1) können durch die Terme ersetzt, die in den Formeln (2)–(5) auf der linken Seite stehen:

$$\begin{aligned} y_{is} = & \gamma_{00} + \gamma_{01}GT_s + \gamma_{10}HISEI_{is} + \gamma_{11}HISEI_{is} * GT_s \\ & + \gamma_{20}Mig1_{is} + \gamma_{21}Mig1_{is} * GT_s + \gamma_{30}Mig2_{is} + \gamma_{31}Mig2_{is} * GT_s \\ & + u_0 + u_1HISEI_{is} + u_2Mig_{-1}_{is} + u_3Mig_{-2}_{is} + e_{is} \end{aligned} \quad (6)$$

Zur Beantwortung unserer ersten Forschungsfrage schätzen wir mit γ_{01} den Unterschied im Leistungsniveau zwischen Halb- und Ganztagschulen. Zur Beantwortung der Fragen, ob an Ganztagschulen die Kopplung zwischen Schulleistung und sozialer Herkunft bzw. Migrationshintergrund geringer ist, als an Halbtagschulen, betrachten wir mit γ_{11} , γ_{21} und γ_{31} drei Interaktionseffekte zwischen Organisationsform und Schülerherkunft.

Die kritische Voraussetzung, damit mithilfe von Modell (6) tatsächlich Effekte der Organisationsform identifiziert werden können, ist, dass sich Halb- und Ganztagschulen nicht in Bezug auf anderen Prädiktoren der Schülerleistung unterscheiden dürfen. Mit anderen Worten muss die Organisationsform mit den Fehlertermen u_0 , u_1 , u_2 und u_3 unkorreliert sein. Es liegt nahe zu fragen, ob diese Annahme haltbar ist, denn im Rahmen der Bildungsforschung sind Selektionseffekte wohlbekannt. Typischerweise finden sich vergleichsweise viele Kinder aus privilegierten Elternhäusern in Einrichtungen der frühkindliche Bildung, besuchen Reformschulen oder wechseln häufiger auf das Gymnasium (z. B. Watermann et al. 2005; Baumert et al. 2009; Schlotter und Wößmann 2010). Auch im Rahmen der Forschung zur Teilnahme am Ganzttag deuten erste Forschungserkenntnisse darauf hin, dass der Schülerhintergrund mit der Teilnahme am Ganzttag korrespondiert (vgl. Holtappels et al. 2010). Des Weiteren verläuft die Umwandlung in Ganztagschulen unterschiedlichen schnell in den einzelnen Bundesländern (Kuhlmann und Tillmann 2009).

Um trotz der beschriebenen potenziellen Konfundierungen von Organisationsform, Region und Schülerkomposition mit Modell (6) die Effekte der Organisationsform zu identifizieren, matchen wir zunächst Halb- und Ganztagschulen. Hierzu nutzen wir den Propensity Score Matching (PSM)-Ansatz (Rosenbaum und Rubin 1983; Becker 2011). Die Grundidee hierbei ist, zwei Gruppen zu bilden, die sich ausschließlich anhand des untersuchten Treatments (hier: Organisationsform) unterscheiden, nicht aber in Bezug auf andere Prädiktoren der Schülerleistung. Hierzu müssen alle potenziellen Prädiktoren benannt, erfasst und im Rahmen der Konditionierung mithilfe des PSM verwendet werden.

Das Matching verläuft in zwei Schritten. In einem ersten Schritt wird ein logistisches Regressionsmodell geschätzt, in dem wir das Treatment (Organisationsform) auf die Variablen zur Schulkomposition, die Region und den Schultyp regressieren. Die Regressionsparameter geben darüber Auskunft, inwiefern sich Halb- und Ganztagschulen in Bezug auf die Schulkomposition, das Bundesland und die Schulform unterscheiden. Anhand der Parameterschätzungen kann für alle Schulen ein sogenannter Propensity Score berechnet werden, hierfür wird für jede Schule ein Sum-

menscore aus den Variablen zur Schulkomposition, Bundesland und Schulformen gebildet, wobei die Variablen mit den Regressionsparametern gewichtet werden. Im zweiten Schritt bietet die Verteilung der Propensity Scores den Ausgangspunkt, wenn zwei Gruppen von Ganz- und Halbtagsschulen gematcht werden, die sich möglichst ähnlich sind. In der Literatur werden eine Reihe von Algorithmen diskutiert, wie vergleichbare Gruppen gebildet werden. Wir verwenden in der vorliegenden Studie den Optimal Matching Algorithmus (Hansen und Olson Klopfer 2006).⁴

Zunächst führen wir das Propensity Score Matching durch, wobei in IGLU, TIMSS und PISA auf die in den jeweiligen Studien verfügbaren Informationen zurückgegriffen wurde (vgl. Tab. 2). Danach schätzen wir separat für Primar- und Sekundarstufe sowie im Lesen, in Mathematik und in Naturwissenschaften fünf Mehrebenenmodelle mit jeweils einem anderen Plausible Value. Danach werden die Schätzungen nach den Regeln von Rubin (1987) für die unterschiedlichen Domänen und Studien zusammengefasst. Das Matching erfolgt mit dem R-Paket MatchIt (Ho et al. 2011) und die Mehrebenenmodelle werden mit Mplus geschätzt (Muthén und Muthén 1998–2010).

4 Ergebnisse

4.1 Ergebnisse der Konditionierung

Ein Vergleich von Halb- und Ganztagschulen ist nur dann zulässig, wenn sich die Schulen beider Organisationsformen nicht in Bezug auf andere Prädiktoren der Schülerleistung unterscheiden. In den Abb. 1, 2 und 3 finden sich die Ergebnisse des Propensity Score Matchings für die drei von uns betrachteten Stichproben. Während in den Zeilen die verwendeten Variablen zur Schulkomposition angeführt werden, zeigen die Punkte und Kreuze Unterschiede zwischen den Halb- und Ganztagschulen vor bzw. nach dem Matching an. Bei stetigen Variablen sind standardisierte Mittelwertunterschiede (standardisiert an der Standardabweichung der Treatmentgruppe = Ganztagschulen) und bei diskreten Variablen Anteile angegeben. Je geringer die Unterschiede zwischen den beiden Gruppen sind, desto näher sind die Punkte bzw. Kreuze an der Vertikalen über der Null. Die gestrichelten Linien links und rechts sind die Grenzen, ab wo der Unterschied größer als 10% oder 0,1 Standardabweichung (SD) ist. Im Folgenden bezeichnen wir Unterschiede außerhalb dieses Korridors als substanziell und betrachten geringere Unterschiede als vernachlässigbar. In der Forschungsliteratur variieren Empfehlungen für vernachlässigbare Unterschiede zwischen konservativen Empfehlungen von 3–5% und liberale von bis zu 0,25 einer Standardabweichung (vgl. Becker 2011). Da wir in den vorliegenden Analysen auf Schulebene matchen, ist die Stichprobengröße naturgemäß kleiner als beim Matching auf Individualebene und es ist schwieriger in kleineren Stichproben zwei ausbalancierte Gruppen zu finden. Vor diesem Hintergrund interpretieren wird die von uns vorgeschlagenen Grenzen (10%/ 0,1 SD) als relativ konservativ, obwohl wir über den in der Literatur genannten Untergrenzen bleiben.⁵

Ein Vergleich beider Organisationsformen vor dem Matching (in den Abb. Kreise) deckt auf, dass die Schülerkomposition an Halb- und Ganztagschulen insgesamt

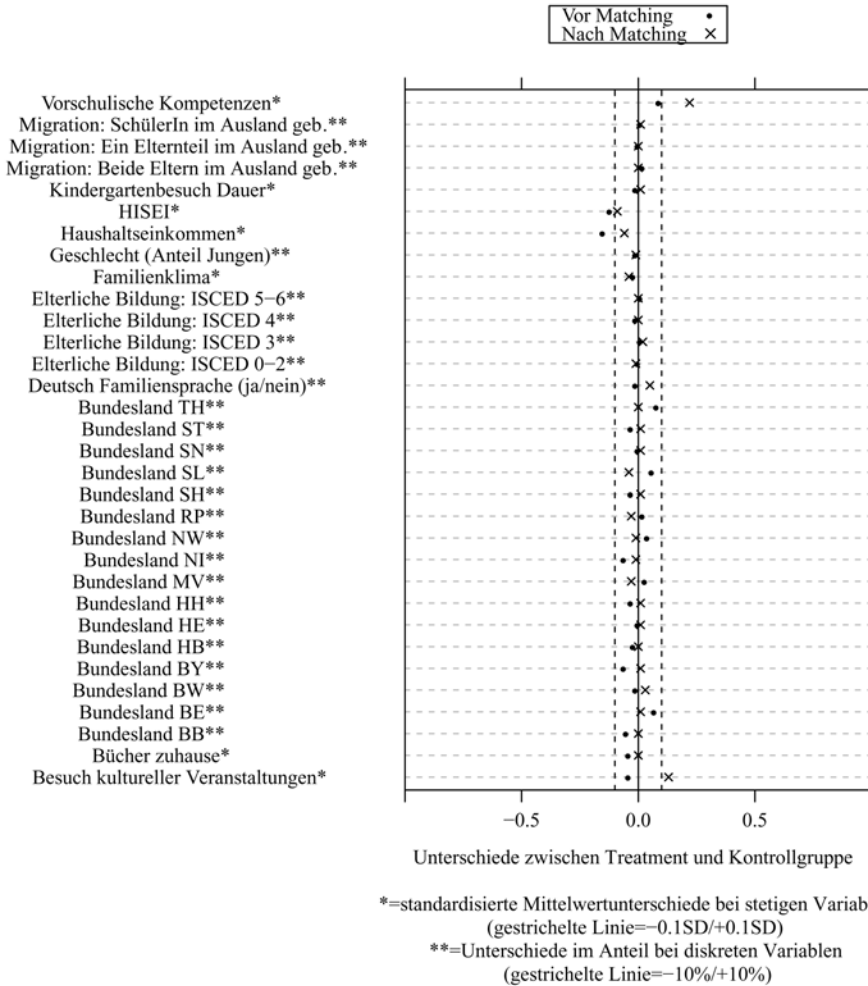


Abb. 1 Ergebnisse des Propensity Score Matchings – IGLU 2006

relativ ausgeglichen ist. Dieser Befund ist für die weitere Interpretation unserer Analysen von Bedeutung, denn wenn mithilfe von Beobachtungsdaten Programmeffekte geschätzt werden, so wird zwischen dem Average Treatment Effects (ATE) und dem Average Treatment Effect on the Treated (ATT) unterschieden. Diese Unterscheidung folgt der Überlegung, dass Programme oder Maßnahmen differenzielle Effekte für bestimmte Schülergruppen haben können. In Experimenten werden ATEs für die betrachtete Population geschätzt, wenn aber eine selbstselektierte Teilnehmergruppe an einer Maßnahme teilnimmt und qua Matching eine Vergleichsgruppe gebildet wird, so wird der ATT geschätzt. Da zwischen Halb- und Ganztagschulen keine größeren Unterschiede beobachtet werden, fallen ATE und ATT – zumindest annähernd – zusammen und wir halten es für gerechtfertigt, die Ergebnisse unserer Analysen auf die gesamte Population deutscher Schulen zu verallgemeinern.

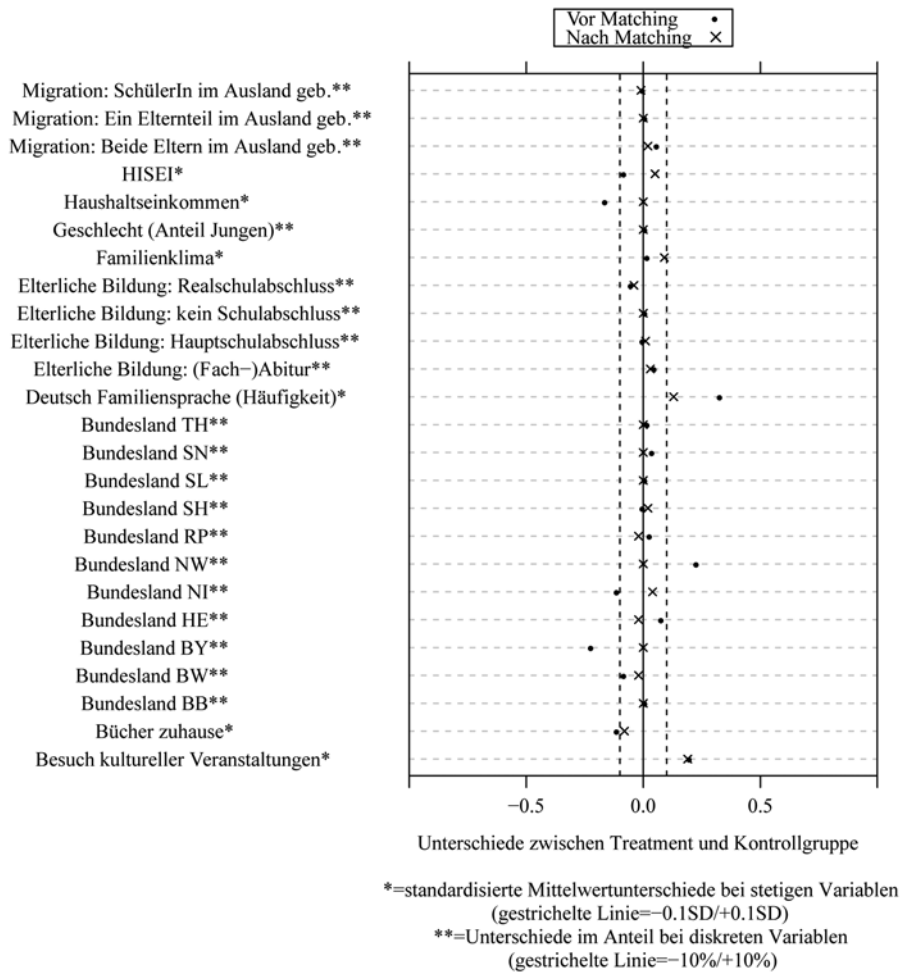


Abb. 2 Ergebnisse des Propensity Score Matchings – TIMSS 2007

Halb- und Ganztagschulen unterscheiden sich in einer Reihe von Schulmerkmalen voneinander. In IGLU lassen sich in Bezug auf den HISEI und der Haushaltseinkommen substanzielle Unterschiede (10% < bzw. 0,1 SD <) beobachten und in TIMSS beim Haushaltseinkommen, der Familiensprache, den zuhause verfügbaren Büchern, sowie bei den Besuchen kultureller Veranstaltungen. In der PISA-Stichprobe werden bei den Variablen zur Schulkomposition keine Unterschiede von größer als 10% bzw. 0,1 SD beobachtet, allerdings ist der Anteil von Ganztagschulen bei den Gesamtschulen deutlich erhöht und bei den Realschulen deutlich niedriger. In Bezug auf die Bundesländer sind in PIRLS und PISA alle Länder relativ ausgeglichen, aber in der TIMSS-Stichprobe werden größere Unterschiede beobachtet.

Nach dem Matching (in den Abb. Kreuze) sind auch in der TIMSS-Stichprobe in etwa gleich viele Bundesländer in den beiden Gruppen von Halb- und Ganztags-

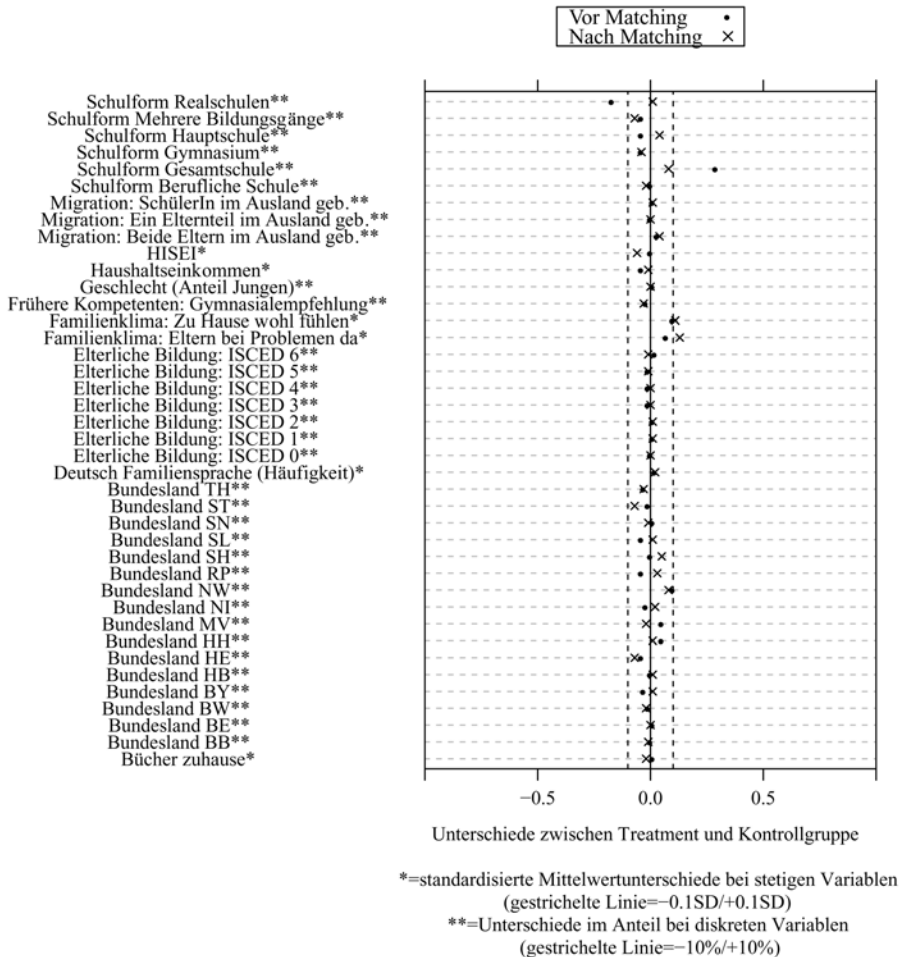


Abb. 3 Ergebnisse des Propensity Score Matchings – PISA-E 2006

schulen. In PISA werden keine substanziellen Schulformunterschiede mehr beobachtet. Die Verteilung der Variablen zur Schulkomposition ist nach dem Matching insgesamt gut ausbalanciert, sodass bei Vergleichen zwischen Halb- und Ganztagschulen Verzerrungen durch die Schulform, die Bundesländer sowie die beobachteten Variablen zur Schulkomposition weitgehend ausgeschlossen werden können. Bei einer genaueren Betrachtung gibt es in IGLU, TIMSS und PISA jeweils noch zwei Variablen, bei denen sich Halb- und Ganztagschulen nach dem Matching noch um mehr als 0,1 SD unterscheiden. Aus diesem Grund schätzen wir zusätzlich zu dem in Abschn. 3.5. beschriebenen Modell auch noch einmal zusätzlich Modelle, bei denen wir die jeweiligen unbalancierten Schulmerkmale als Kontrollvariablen aufnehmen. Das heißt, wir regressieren die Koeffizienten β_{0s} , β_{1s} , β_{2s} und β_{3s} (vgl. 2–5) nicht nur auf die Organisationsform, sondern bspw. in IGLU zusätzlich auf die vorschulische

Kompetenzen und den Besuch kultureller Veranstaltungen. Die gematchten Datensätze sind die Grundlage für die Mehrebenenanalysen.

4.2 Ergebnisse der Mehrebenenanalysen

Die Ergebnisse der Schätzungen der Random Slope-Mehrebenenmodelle sind in Tab. 3 zusammengefasst. In der Primar- und der Sekundarstufe wurden separat für die Kompetenzdomänen Lesen, Mathematik und Naturwissenschaften (NaWi) sechs Modelle geschätzt, dabei steht der Intercept für die mittlere Leistung eines Kindes ohne Migrationshintergrund, mit einem HISEI von 40, das eine Halbtagschule besucht. Die Regressionsgewichte für die beiden Migrationsvariablen (Mig1 = ein Elternteil im Ausland geboren/Mig2 = beide Eltern im Ausland geboren) und den HISEI sind Indikatoren für die Bildungsungleichheit.

Um zu untersuchen, ob die Organisationsform einen Effekt auf die Schülerleistung hat, betrachten wir für das Leistungsniveau den Parameter GT und für die Bildungs-

Tab. 3 Ergebnisse der Mehrebenenanalysen

	PIRLS 06		TIMSS 07		PISA-E 06	
N _{Schulen}	160	104	104	588	588	588
N _{Schüler}	3101	2185	2185	14685	14685	14685
Domäne	Lesen	Mathe	NaWi	Lesen	Mathe	NaWi
Model	(1)	(2)	(3)	(4)	(5)	(6)
<i>Feste Effekte</i>						
Intercept	543,89** (3,39)	521,82** (3,83)	529,80** (4,32)	496,74** (4,11)	501,62** (3,59)	518,46* (3,65)
GT	-9,04 (5,75)	-4,67 (6,76)	-1,76 (6,67)	-5,91 (5,77)	-7,85 (5,33)	-9,25 (5,29)
Mig1	-18,31** (5,04)	-14,27 (10,14)	-30,97* (12,40)	-13,08* (5,50)	-20,84** (5,50)	-18,76** (4,97)
Mig2	-28,69** (5,41)	-27,08** (6,09)	-51,96** (8,95)	-46,29** (4,14)	-44,32** (3,46)	-63,22** (3,59)
HISEI	1,09** (0,11)	1,23** (0,15)	1,37** (0,16)	0,44 ** (0,07)	0,65** (0,06)	0,62** (0,06)
GT*Mig1	-3,20 (7,46)	-3,49 (13,64)	-3,78 (17,11)	1,76 (7,91)	-0,25 (7,03)	4,00 (6,66)
GT*Mig2	-0,23 (7,62)	-7,81 (10,02)	-15,07 (12,50)	9,80 (5,79)	8,25 (4,71)	13,00** (5,00)
GT*HISEI	0,23 (0,13)	-0,04 (0,21)	-1,76 (6,67)	0,03 (0,09)	-0,03 (0,09)	0,00 (0,10)
<i>Zufallseffekte</i>						
σ^2_e	2774,54** (127,80)	3151,70** (147,15)	4087,51** (159,04)	5035,55** (85,00)	4314,57** (74,77)	4673,58** (67,64)
σ^2_{u0}	936,58** (272,55)	576,73 (305,43)	502,40* (199,39)	4218,70** (238,2)	3645,06** (215,96)	3712,78** (209,70)
σ^2_{u1}	0,00 (0,03)	0,62 (388,12)	79,48 (576,37)	1958,43* (833,5)	1568,84* (646,03)	735,37* (288,21)
σ^2_{u2}	1,17 (102,66)	173,34 (451,54)	1101,17 (733,56)	1318,50** (323,87)	559,46* (224,28)	443,11 (316,62)
σ^2_{u3}	0,01 (0,10)	0,03 (0,13)	0,03 (0,17)	0,10 (0,08)	0,16* (0,08)	0,14* (0,07)

Es sind nicht standardisierte Parameter angegeben und clusterrobuste Standardfehler in Klammern.

Signifikanzniveaus sind durch * = 0,05 und ** = 0,01 gekennzeichnet.

ungleichheit die Interaktion zwischen der Organisationsform und die Herkunftsvariablen (GT*Mig1, GT*Mig2, GT*HISEI).

Als erstes Ergebnis unserer Analysen betrachten wir die sechs Schätzungen des Parameters GT, welche zwischen -1,76 (Modell 2) und -9,25 (Modell 6) liegen. Somit beobachten wir in allen von uns betrachteten Stichproben an Ganztagschulen ein geringeres Leistungsniveau als an Halbtagsschulen. Allerdings unterscheidet sich keine der Parameterschätzungen signifikant von null, damit bieten unsere Analysen keine empirische Evidenz dafür, dass die Organisationsform der Schule einen Effekt auf das Leistungsniveau hat.

In Bezug auf herkunftsbedingte Disparitäten in der Leistung von Schülerinnen und Schülern an Halb- und Ganztagschulen betrachten wir in den sechs Modellschätzungen jeweils drei Interaktionseffekte zwischen Organisationsform und Schülerhintergrund. In den Modellen (1–5) unterscheidet sich keiner der Interaktionseffekte signifikant von null. Auch in Modell (6) unterscheiden sich zwei der drei Parameter nicht signifikant von null. Die einzige Ausnahme bildet der Parameter GT*Mig2. Wenn Schülerinnen und Schüler eine Halbtagsschule besuchen und beide Eltern im Ausland geboren sind, so schneiden sie um 63,22 Punkte schlechter ab als Kinder, deren Eltern beide in Deutschland geboren wurden. An Ganztagschulen ist dieser Abstand um 13,00 Punkte geringer und dieser Unterschied ist statistisch signifikant. Würde man nur diesen einen Parameter betrachten, so könnte man vermuten, dass Migranten an Ganztagschulen besser gefördert werden, als an Halbtagsschulen. Vor dem Hintergrund, dass 17 von 18 Parametern nicht signifikant werden, schlagen wir vor, diesen Befund mit großer Zurückhaltung zu interpretieren. In der Zusammenschau bieten die Analysen unseres Erachtens höchstens schwache Evidenz dafür, dass die Organisationsform in einem Teilwirkungsbereich Einfluss auf die Bildungsungleichheit hat.

4.3 Ergänzende Analysen

Die Tatsache, dass nach dem Propensity Score Matching in allen drei Datensätzen jeweils zwei Variablen in den Halb- und Ganztagschulen noch nicht ausbalanciert waren, legt weitere Analysen nahe, um möglichst faire Vergleiche zwischen Halb- und Ganztagschulen zu ermöglichen. In einer Reihe weiterer Mehrebenenanalysen haben wir daher die jeweiligen beiden unbalancierten Variablen als Kontrollvariablen in die Modelle aufgenommen, bei denen die standardisierten Mittelwertunterschiede nach dem Matching den Schwellenwert von 0.10 SD bzw. 10% erreichen.

Man kann auch noch strengere Kriterien an die Vergleichbarkeit von Halb- und Ganztagschulen anlegen, hierbei ist allerdings zu beachten, dass durch die zusätzliche Aufnahme jeder weiteren Kontrollvariable vier zusätzliche Modellparameter geschätzt werden müssen (einen für den Intercept und drei für die Slopes). Aufgrund der auf Schulebene vergleichsweise geringen Stichprobengröße, kann dies dazu führen, dass etwaige Effekte nicht mehr inferenzstatistisch abgesichert werden können. Folglich muss zwischen möglichst strengen Kriterien und sparsamen Modellen abgewogen werden. Vor diesem Hintergrund wurden in weiteren Analysen nicht alle, sondern nur die Variablen als Kontrollvariablen aufgenommen, bei denen die standardisierten Mittelwertunterschiede, nach dem Matching, den von Caliendo und

Tab. 4 Ergebnisse der Mehrebenenanalysen mit Kontrollvariablen (standardisierte Mittelwertunterschiede > 0,10 SD bzw. > 10%)

	PIRLS 06		TIMSS 07		PISA-E 06	
N _{Schulen}	160	104	104	588	588	588
N _{Schüler}	3101	2185	2185	14685	14685	14685
Domäne	Lesen	Mathe	NaWi	Lesen	Mathe	NaWi
Model	(1)	(2)	(3)	(4)	(5)	(6)
<i>Feste Effekte</i>						
Intercept	543,69** (3,21)	522,85** (3,26)	530,98** (3,90)	495,02** (3,98)	500,07** (3,47)	516,94** (3,51)
GT	-8,28 (5,45)	-4,27 (6,01)	-1,24 (6,25)	-3,08 (5,56)	-5,38 (5,14)	-6,54 (5,10)
Mig1	-19,06** (5,49)	-14,79 (10,03)	-32,06* (15,06)	-12,8* (5,47)	-20,31** (5,46)	-18,06** (5,01)
Mig2	-33,32** (5,88)	-27,91** (6,30)	-52,87** (8,25)	-45,99** (4,16)	-44,07** (3,48)	-62,82** (3,68)
HISEI	1,04** (0,10)	1,20** (0,15)	1,33** (0,16)	0,44** (0,07)	0,64** (0,06)	0,62** (0,06)
GT*Mig1	-4,31 (7,26)	-3,63 (12,19)	-4,37 (18,02)	1,49 (7,88)	-0,51 (6,96)	3,98 (6,66)
GT*Mig2	-1,98 (8,11)	-11,13 (10,30)	-18,39 (11,84)	9,67 (5,81)	8,10 (4,61)	13,08** (5,07)
GT*HISEI	0,26 (0,14)	-0,06 (0,21)	-0,21 (0,23)	0,02 (0,09)	-0,03 (0,09)	0,00 (0,10)
Kontrollvariablen	$\sqrt{(2)^1}$	$\sqrt{(2)^2}$	$\sqrt{(2)^2}$	$\sqrt{(2)^3}$	$\sqrt{(2)^3}$	$\sqrt{(2)^3}$
<i>Zufallseffekte</i>						
σ_e^2	2716,48** (99,08)	3137,09** (141,95)	4035,03** (186,49)	5033,93** (84,95)	4324,95** (75,88)	4666,17** (65,53)
σ_{u0}^2	804,69** (239,76)	339,36* (167,61)	321,25* (133,39)	3859,66** (227,52)	3347,72** (201,61)	3344,77** (191,42)
σ_{u1}^2	0,04 (158,17)	1,00 (495,81)	364,06 (814,01)	1946,6* (813,51)	1498,63* (631,04)	690,45* (288,36)
σ_{u2}^2	432,42 (256,32)	114,14 (332,21)	442,1 (460,74)	1329,62** (329,32)	424,28 (327,81)	538,11** (187,86)
σ_{u3}^2	0,01 (0,11)	0,02 (0,12)	0,03 (0,14)	0,09 (0,08)	0,16* (0,08)	0,13* (0,06)

Es sind nicht standardisierte Parameter angegeben und clusterrobuste Standardfehler in Klammern.

Signifikanzniveaus sind durch * = 0,05 und ** = 0,01 gekennzeichnet.

Kontrollvariablen auf Schulebene sind: ¹=Besuch kultureller Veranstaltungen, Vorschulische Kompetenzen; ²=Migration: Besuch kultureller Veranstaltungen, Familiensprache Deutsch (Häufigkeit); ³=Familienklima: Zu Hause wohl fühlen, Familienklima: Eltern bei Problemen da.

Kopeinig (2008) vorgeschlagenen (strengen) Schwellenwert von 0.03 SD bzw. 3% überschreiten. Dies trifft bei IGLU auf sechs, bei TIMSS auf sieben und bei PISA auf acht Kovariaten zu.

Die Ergebnisse dieser Analysen sind in Tab. 4 und 5 dargestellt.⁶ Wiederum bestätigt sich der Befund der vorangegangenen Analysen: Im Leistungsniveau an Halb- und Ganztagschulen können keine signifikanten Unterschiede beobachtet werden. Und auch in Bezug auf Bildungsungleichheit bietet sich ein unverändertes Bild. Es können wiederum (mit einer einzigen Ausnahme in Tab. 4 und zwei in Tab. 5) keine signifikanten Unterschiede in der Kopplung zwischen Herkunft und Leistung beobachtet werden.

Tab. 5 Ergebnisse der Mehrebenenanalysen mit Kontrollvariablen (standardisierte Mittelwertunterschiede > 0.03 SD bzw. > 3 %)

	PIRLS 06	TIMSS 07		PISA-E 06		
N_{Schulen}	160	104	104	588	588	588
$N_{\text{Schüler}}$	3101	2185	2185	14685	14685	14685
Domäne	Lesen	Mathe	NaWi	Lesen	Mathe	NaWi
Model	(1)	(2)	(3)	(4)	(5)	(6)
<i>Feste Effekte</i>						
Intercept	545,72** (2,89)	476,66** (23,29)	479,15** (24,42)	545,77** (4,23)	548,89** (4,03)	566,32** (3,87)
GT	-8,70 (4,64)	-0,45 (4,96)	2,02 (5,29)	-0,67 (3,38)	-1,67 (3,00)	-2,58 (2,93)
Mig1	-16,10** (5,44)	-2,24 (54,23)	8,1 (76,94)	3,38 (11,77)	-4,07 (11,87)	-3,62 (20,40)
Mig2	-34,76** (5,56)	-23,71 (33,24)	-39,98 (45,22)	-35,08** (7,88)	-42,16** (7,26)	-51,79** (4,68)
HISEI	0,98** (0,11)	2,20** (0,85)	2,49* (1,04)	0,17 (0,12)	0,45** (0,13)	0,49** (0,13)
GT*Mig1	-0,07 (7,26)	-4,74 (12,28)	-3,49 (16,13)	2,90 (7,65)	-0,10 (6,92)	-6,54 (21,88)
GT*Mig2	-0,69 (7,49)	-11,33 (9,20)	-17,74 (10,52)	12,25* (5,65)	8,19 (4,45)	13,27** (3,00)
GT*HISEI	0,29 (0,14)	-0,12 (0,20)	-0,25 (0,23)	-0,02 (0,09)	-0,04 (0,09)	-0,02 (0,10)
Kontroll-variablen	$\sqrt{(6)^1}$	$\sqrt{(7)^2}$	$\sqrt{(7)^2}$	$\sqrt{(8)^3}$	$\sqrt{(8)^3}$	$\sqrt{(8)^3}$
<i>Zufallseffekte</i>						
σ^2_e	2715,68** (99,58)	3120,58** (138,02)	4070,10** (156,40)	5029,11** (84,21)	4323,86** (75,23)	4681,04** (69,16)
σ^2_{u0}	407,85** (132,01)	172,03* (78,12)	176,34* (77,75)	1023,31** (80,96)	793,84** (69,01)	709,34** (59,75)
σ^2_{u1}	0,49 (214,62)	0,70 (642,23)	5,15 (693,56)	1590,82* (726,67)	1196,73* (587,28)	399,66 (237,88)
σ^2_{u2}	251,56 (216,49)	54,95 (209,36)	205,23 (296,95)	1103,26** (281,14)	317,12 (247,68)	206,99 (322,88)
σ^2_{u3}	0,01 (0,10)	0,02 (0,11)	0,01 (0,16)	0,07 (0,07)	0,13 (0,07)	0,13* (0,06)

Es sind nicht standardisierte Parameter angegeben und clusterrobuste Standardfehler in Klammern. Signifikanzniveaus sind durch * = 0,05 und ** = 0,01 gekennzeichnet. Kontrollvariablen auf Schulebene sind: ¹ = Besuch kultureller Veranstaltungen, Familienklima, Haushaltseinkommen, HISEI, Migration: Deutsch Familiensprache (ja/nein), Vorschulische Kompetenzen; ² = Bücher zuhause, Elterliche Bildung: Realschulabschluss, Elterliche Bildung: (Fach-)Abitur, Familienklima, HISEI, Migration: Familiensprache Deutsch (Häufigkeit), Besuch kultureller Veranstaltungen; ³ = HISEI, Familienklima: Zu Hause wohl fühlen, Familienklima: Eltern bei Problemen da, Schulform Hauptschule, Schulform Mehrere Bildungsgänge, Schulform Gesamtschule, Schulform Berufliche Schule.

5 Zusammenfassung und Diskussion

In unserer Studie betrachteten wir die Primar- und die Sekundarstufe, sowie Kompetenzen in den Domänen Lesen, Mathematik und Naturwissenschaften. Neben dem Leistungsniveau analysierten wir Bildungsungleichheit an Halb- und Ganztagschulen. Unsere Analysen beruhen auf repräsentativen Stichproben aus drei großflächig angelegten Schulleistungsstudien. Zusammenfassend bieten die Ergebnisse so gut wie keine empirische Evidenz dafür, dass die Organisationsform von Schule einen

Effekt auf die Schülerleistung hat. Trotz des erweiterten Zeitbudgets ist das Leistungsniveau an Ganztagschulen nicht höher und die Bildungsungleichheit nicht geringer als an Halbtagschulen.

Diese Ergebnisse müssen vor dem Hintergrund betrachtet werden, dass mit dem Ausbau von Ganztagschulen vielfältige Ziele verfolgt werden. Neben ebenso wichtigen Aspekten, wie der Förderung der Vereinbarkeit von Familie und Beruf oder der sozialen Entwicklung, sind Schülerkompetenzen nur ein Kriterium, wenn nach den Effekten von ganztägiger Bildung und Betreuung gefragt wird. Des Weiteren sind Rückschlüsse auf die Effekte bestimmter Maßnahmen und Programme mit Beobachtungsstudien eine methodologische Herausforderung. Die kritische Annahme der vorliegenden Analysen ist, dass es mithilfe des Propensity Score Matchings gelungen ist, zwei Gruppen von Halb- und Ganztagschulen zu bilden, die sich nicht in Bezug auf andere Prädiktoren von Schülerleistung unterscheiden (conditional independence assumption). Wie plausibel ist diese Annahme? In den von uns verwendeten Schulleistungsstudien können wir auf ein ausführliches Repertoire an Informationen zu unterschiedlichen Merkmalen zur Schulkomposition zurückgreifen. Interessanterweise werden bereits vor dem Matching kaum systematische Unterschiede zwischen Halb- und Ganztagschulen beobachtet. Das Matching verringert die beobachteten Differenzen weiter. Des Weiteren liegen in IGLU und PISA Informationen zur Vorleistung vor. Wenngleich die Elterneinschätzung (IGLU) und die Übergangsempfehlung (PISA) sicherlich messfehlerbehaftet sind, bieten diese Informationen durchaus die Möglichkeit ein rudimentäres Value Added Modell zu schätzen und somit auch den Einfluss unbeobachteter konfundierter Variablen zu kontrollieren.

Trotz dieser Einschränkungen drängt sich die Frage auf, inwiefern der Ausbau von Ganztagschulen als Reformmaßnahme wirkt. Dabei stellt sich weniger die Frage, ob ganztägige Bildung und Betreuung wirken könnte, sondern ob sie in der gegenwärtigen Ausgestaltung wirkt. In der vorliegenden Untersuchung haben wir keine Modellprogramme (bzw. Modellschulen) untersucht, sondern gegenwärtige Veränderungen und Entwicklungen in ganz Deutschland betrachtet. Hierbei gilt es zu beachten, dass als die von uns verwendeten Daten erhoben wurde, ein Großteil der Ganztagschulen erst kürzlich von einer Halb- in eine Ganztagschulen umgewandelt wurden (vgl. Höhmann et al. 2007). Dabei muss kritisch angemerkt werden, dass Fördermittel vor allem in Räumlichkeiten (z. B. Mensen) investiert wurden. Wenngleich eine entsprechende Infrastruktur wohl als notwendige Bedingung für den Ganztagsbetrieb gelten kann, so gibt es gleichzeitig weitere Gelingens-Bedingungen. Beispielsweise scheint es zumindest plausibel, dass das zusätzliche pädagogische Personal eine fundierte akademische Ausbildung benötigt, damit Schülerinnen und Schüler tatsächlich besser lernen. Forschungsbefunde deuten darauf hin, dass dies oftmals nicht der Fall ist (Höhmann et al. 2007). Insofern halten wir es für naiv zu schlussfolgern, dass Ganztagschulen nicht wirken können. Vielmehr interpretieren wir unsere Ergebnisse dahingehend, dass die gegenwärtige organisatorische, pädagogische oder personelle Ausgestaltung ganztägiger Bildung und Betreuung in der aktuellen Form nicht zu den erhofften Effekten hinsichtlich einer Verbesserung von fachlichen Schülerkompetenzen führt.

Schlussendlich weisen wir auf einen nach wie vor existenten Forschungsbedarf hin. Wenn in der Begleitforschung zum milliardenschweren „Investitionsprogramm

Zukunft Bildung und Betreuung“ darauf verzichtet wird, Tests in den zentralen Kompetenzdomänen durchzuführen und das bisherige Studiendesign keine Vergleichsgruppe von Halbtagsschulen vorsieht, so sind die Motive hierfür kaum wissenschaftlicher Natur. Mit Sorge beobachten wir auch, dass mittlerweile darauf verzichtet wird, die Stichproben internationaler Schulleistungsstudien zu erweitern. Die sogenannten Ländervergleiche im Rahmen der Überprüfung der Erreichung der Bildungsstandards bieten hierfür kaum einen Ersatz, da im Rahmen dieser Untersuchungen nur einige wenige individuelle Schülerinformationen erhoben werden, nicht aber Schulmerkmale, wie etwa die Organisationsform. Zur Verdeutlichung: unsere Kritik soll nicht als Kritik am Ausbau von Ganztagschulen verstanden werden. Vor dem Hintergrund unserer Befunde muss jedoch über die Qualität und die Art und Weise des Ausbaus nachgedacht werden. Hierbei allein den Schulen die Verantwortung zu geben, mag bildungspolitisch vorstellbar sein, aus wissenschaftlicher Perspektive scheint es allerdings geboten, unter Berücksichtigung der Mehrebenenstruktur des Schulsystems über Ausbauege, sowie Qualitätssicherungs- und -entwicklungsmaßnahmen nachzudenken.

Anmerkungen

- 1 Die PISA-Daten wurden vom Forschungszentrum am IQB Berlin bereitgestellt (Prenzel, M. et al. (2010): Programme for International Student Assessment 2006 (PISA 2006). Version: 1. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_PISA_2006_v1).
- 2 In PISA wurde nicht erhoben, seit wann die Schulen als Ganztagschule organisiert sind und an welchen Ganztagsangeboten die Schülerinnen und Schüler teilnehmen.
- 3 Eine genaue Aufstellung aller Items ist bei den Autoren auf Anfrage zu erhalten.
- 4 Eine Grundvoraussetzung für das Matching ist, dass es ausreichend Vergleichsschulen gibt. In Bezug auf die Bundesländer ist das nicht immer der Fall. Während sich in der PISA-E-Stichprobe in allen Bundesländern mehr Halbtags- als Ganztagschulen in der Stichprobe befinden, enthält die (viel kleinere) TIMSS-Stichprobe in Bremen ($n=2$), Hamburg ($n=4$), Mecklenburg-Vorpommern ($n=2$) und Sachsen-Anhalt ($n=5$) ausschließlich Halbtagsschulen, in Berlin ($n=7$) hingegen nur Ganztagschulen. Diese Bundesländer wurden für die Analysen mit TIMSS-Daten komplett entfernt. Des Weiteren befanden sich im Saarland, NRW, Brandenburg und Sachsen mehr Halbtagsschulen als Ganztagschulen in der Stichprobe. In diesem Fall wurden zufällig genau so viele Ganztagschulen ausgewählt und gelöscht, bis die Anzahl der Ganztagschulen der Zahl der Halbtagsschulen entspricht. In der IGLU-Stichprobe trifft dasselbe Problem auf, allerdings nur in Berlin. Den 16 Berliner Ganztagsgrundschulen, stehen nur 8 Halbtagsschulen gegenüber, somit wurden zufällig 8 Ganztagschulen von den Analysen ausgeschlossen.
- 5 Bei den stetigen Kovariaten betrachten wir neben Mittelwertunterschieden zusätzlich noch die Streuung der Variablen in den beiden Gruppen. Hierbei sollte das Verhältnis der Varianzen in der Treatment- und der Kontrollgruppe (σ_{TG}/σ_{KG}) zwischen 0,5 und 2 liegen (vgl. Becker 2011). Das trifft auch auf fast alle Variablen zu, wobei der Anteil an Eltern mit Realschulabschluss in der TIMSS-Stichprobe die einzige Ausnahme bildet: Hier liegt die Varianz in der Gruppe der Ganztagschulen bei $\sigma_{TG}=0,028$ und in Halbtagsschulen bei $\sigma_{KG}=0,013$ ($\sigma_{TG}/\sigma_{KG}=2,24$). Da über alle drei Studien hinweg nur eine einzige Variable (knapp) außerhalb der vorgeschlagenen Grenzwerte liegt, schlagen wir vor diesen Unterschied nicht zu überinterpretieren.
- 6 Bei den Modellschätzungen mit Kontrollvariablen ist bei den TIMSS-Daten in Naturwissenschaft (Tab. 5, Modell 3) einer von fünf Durchläufen mit den unterschiedlichen Plausible Values nicht konvergiert. Die Ergebnisse beruhen auf Analysen mit den anderen vier Plausible Values.

Literatur

- Ahnen, D. (2005). Zur Gegenwart und Zukunft der Ganztagschulen. In T. Fitzner, T. Schlag, & M. W. Lallinger (Hrsg.), *Ganztagschule – Ganztagsbildung. Politik-Pädagogik-Kooperationen* (S. 125–137). Bad Boll: Evangelische Akademie.
- Appel, S. (2004). *Handbuch Ganztagschule. Konzeption, Einrichtung und Organisation*. Schwalbach: Wochenschau Verlag.
- Baumert, J., Stanat, P., & Watermann, R. (Hrsg.) (2006). *Herkunftsbedingte Disparitäten im Bildungswesen. Vertiefende Analysen im Rahmen von PISA 2000*. Wiesbaden: Springer VS.
- Baumert, J., Becker, M., Neumann, M., & Nikolova, R. (2009). Frühübergang in ein grundständiges Gymnasium – Übergang in ein privilegiertes Entwicklungsmilieu? Ein Vergleich von Regressionsanalyse und Propensity Score Matching. *Zeitschrift für Erziehungswissenschaft*, 12(2), 189–215.
- Becker, M. (2011). Matching-Verfahren und Gruppenvergleiche. In S. Maschke & L. Stecher (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online. Fachgebiet: Methoden der empirischen erziehungswissenschaftlichen Forschung, Quantitative Methoden*. Weinheim: Beltz.
- Bellin, N., & Tamke, F. (2010). Bessere Leistungen durch Teilnahme am offenen Ganztagsbetrieb? *Empirische Pädagogik*, 24(2), 93–112.
- Berkemeyer, N., Bos, W., & Manitiuss, V. (2012). Chancenspiegel. Zur Leistungsfähigkeit und Chancengerechtigkeit der deutschen Schulsysteme. In Bertelsmann Stiftung und Institut für Schulentwicklungsforschung (Hrsg.), *Chancenspiegel. Zur Leistungsfähigkeit und Chancengerechtigkeit der deutschen Schulsysteme* (S. 1–192). Gütersloh: Verlag Bertelsmann Stiftung.
- Berkemeyer, N., & Manitiuss, V. (2013). Gerechtigkeit als Kategorie der Analyse von Schulsystemen – das Beispiel Chancenspiegel. In K. Schwippert, M. Bonsen, & N. Berkemeyer (Hrsg.), *Schul- und Bildungsforschung: Diskussionen, Befunde und Perspektiven* (Festschrift für Wilfried Bos, S. 223–240). Münster: Waxmann.
- BMBF (2009) = Bundesministerium für Bildung und Forschung. (2009). *Gut angelegt. Das Investitionsprogramm Zukunft Bildung und Betreuung*. Berlin: Bundesministerium für Bildung und Forschung.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K., & Valtin, R. (Hrsg.) (2007). *IGLU 2006. Lesekompetenz von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C., & Walther, G. (Hrsg.) (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 183–198). Göttingen: Schwartz.
- Caliendo, M., & Kopeining, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Survey*, 22(1), 31–72.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness*. London: Routledge.
- Creemers, B. P. M., Scheerens, J., & Reynolds, D. (2000). Theory development in school effectiveness research. In C. Teddlie & D. Reynolds (Hrsg.), *The international handbook of school effectiveness research* (S. 283–298). London: Falmer Press.
- Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. *American Economic Review*, 97(2), 31–47.
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt a. M.: Campus Verlag.
- Fend, H. (2006). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen*. Wiesbaden: Springer VS.
- Fischer, N., Holtappels, H. G., Klieme, E., Rauschenbach, T., Stecher, L., & Züchner, I. (Hrsg.) (2011) *Ganztagschule: Entwicklung, Qualität, Wirkungen. Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagschulen (StEG)*. Weinheim: Juventa.
- Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56.
- Gustafsson, J.-E., Yang-Hansen, K., & Rosén, M. (2013). Effects of home background on student achievement in reading, mathematics, and science at the fourth grade. In M. O. Martin & A. L. Mullen (Hrsg.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade—implications for early learning* (S. 181–287). Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

- Hansen, B. B., & Klopfer Olson, S. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, 15(3), 609–627.
- Ho, D., Imai, K., King, G., & Stuart, E. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*. <http://gking.harvard.edu/matchit>. Zugegriffen: 26. Juni 2012.
- Höhmann, K., Grewe, M., & Strietholt, R. (2007). Ganztägiges Lernen in Deutschland – Ausgangslagen und Rahmendaten. Gründung und Ausstattung. In H.-G. Holtappels, E. Klieme, T. Rauschenbacher, & L. Stecher (Hrsg.), *Ganztagsschule in Deutschland. Ergebnisse der Ausgangserhebung der „Studie zur Entwicklung von Ganztagsschulen“ (StEG)* (S. 70–76). Weinheim: Juventa.
- Holtappels, H. G. (2006). Stichwort: Ganztagsschule. *Zeitschrift für Erziehungswissenschaft*, 9(1), 5–29.
- Holtappels, H. G. (2009). Qualitätsmodelle – Theorie und Konzeptionen. In I. Kamski, H. G. Holtappels, & T. Schnetzer (Hrsg.), *Qualität von Ganztagsschule. Konzepte und Orientierungen für die Praxis* (S. 11–25). Münster: Waxmann.
- Holtappels, H. G., Radisch, F., Rollett, W., & Kowoll, M. E. (2010). Bildungsangebot und Schülerkompetenzen in Ganztagsschulen. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, I. Tarelli, & R. Valtin (Hrsg.), *IGLU 2006– die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens* (S. 165–198). Münster: Waxmann.
- Ipfing, H. J. (2005). Voraussetzungen und Bedingungen für die Errichtung von Ganztagsschulen. In V. Ladenthin (Hrsg.), *Die Ganztagsschule* (S. 299–309). Weinheim: Juventa.
- Klafki, W. (1991). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung u. kritisch-konstruktive Didaktik* (2. erw. Aufl.). Weinheim: Beltz.
- Konsortium Bildungsberichterstattung (2006). *Bildung in Deutschland. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann.
- Kuhlmann, C., & Tillmann, K.-J. (2009). Mehr Ganztagsschulen als Konsequenz aus PISA? Bildungspolitische Diskurse und Entwicklungen in den Jahren 2000 bis 2003. In F.-U. Kolbe, S. Reh, B. Fritzsche, T.-S. Idel, & K. Rabenstein (Hrsg.), *Ganztagsschule als symbolische Konstruktion. Fallanalysen zu Legitimationsdiskursen in schultheoretischer Perspektive* (S. 159–177). Wiesbaden: Springer VS.
- Kuhn, H. P., & Fischer, N. (2011). Entwicklung der Schulnoten in der Ganztagsschule. Einflüsse der Ganztagesteilnahme und der Angebotsqualität. In N. Fischer, H. G. Holtappels, E. Klieme, T. Rauschenbach, L. Stecher, & I. Züchner (Hrsg.), *Ganztagsschule: Entwicklung, Qualität, Wirkungen. Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagsschulen (StEG)* (S. 207–226). Weinheim: Juventa.
- KMK (2002) = Kultusministerkonferenz. (2002). *PISA 2000– Zentrale Handlungsfelder. Zusammenfassende Darstellung der laufenden und geplanten Maßnahmen in den Ländern*. <http://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2002/massnahmen.pdf>. Zugegriffen: 12. Sep. 2012
- Maaz, K., & Nagy, G. (2009). Der Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems: Definition, Spezifikation und Quantifizierung primärer und sekundärer Herkunftseffekte. In J. Baumert, K. Maaz, & U. Trautwein (Hrsg.), *Bildungsentscheidungen (Zeitschrift für Erziehungswissenschaft: Sonderheft 12, S. 153–182)*. Wiesbaden: Springer VS.
- Maykus, S., Böttcher, W., Liesegang, T., & Altermann, A. (2011). Individuelle Förderung in der Ganztagschule. Theoretisch-empirische Reflexionen zu Anspruch und Konsequenzen eines (sozial-)pädagogischen Programms im professionellen Handeln. In L. Stecher, H.-H. Krüger, & T. Rauschenbach (Hrsg.), *Ganztagsschule – Neue Schule? Eine Forschungsbilanz (Zeitschrift für Erziehungswissenschaft: Sonderheft 15, S. 125–142)*. Wiesbaden: Springer VS.
- Merkens, H., & Schründer-Lenzen, A. (2010). *Lernförderung unter den Bedingungen des Ganztags in Grundschulbereich. Die Untersuchung Ganztagsorganisation im Grundschulbereich (GO)*. Münster: Waxmann.
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus. Statistical analysis with latent variables. User's guide*. Los Angeles: Muthén & Muthén.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (Hrsg.) (2008). *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich*. Münster: Waxmann.
- Radisch, F. (2009). *Qualität und Wirkung ganztägiger Schulorganisation*. Weinheim: Juventa.
- Radisch, F., & Klieme, E. (2004). Wirkungen ganztägiger Schulorganisation. Bilanz und Perspektiven der Forschung. *Die Deutsche Schule*, 96(2), 153–169.
- Radisch, F., Klieme, E., & Bos, W. (2006). Gestaltungsmerkmale und Effekte ganztägiger Angebote im Grundschulbereich. Eine Sekundäranalyse zu Daten der IGLU-Studie. *Zeitschrift für Erziehungswissenschaft*, 9(1), 30–50.

- Rauschenbach, T., Arnoldt, B., Steiner, C., & Stolz, H.-J. (2012). *Ganztagschule als Hoffnungsträger für die Zukunft? Ein Reformprojekt auf dem Prüfstand. Expertise des Deutschen Jugendinstituts (DJI) im Auftrag der Bertelsmann Stiftung*. Gütersloh: Verlag Bertelsmann Stiftung.
- Reinders, H., Gogolin, I., Gresser, A., Schnurr, S., Böhmer, J., & Bremm, N. (2011). Ganztagschulbesuch und Integration von Kindern mit Migrationshintergrund im Primarbereich: Erste Näherungen an empirische Befunde einer vergleichenden Untersuchung. *Ganztagschule – Neue Schule? Eine Forschungsbilanz* (Zeitschrift für Erziehungswissenschaft: Sonderheft 15, S. 163–184). Wiesbaden: Springer VS.
- Rekus, J. (2005). Theorie der Ganztagschule – praktische Orientierungen. In V. Ladenthin (Hrsg.), *Die Ganztagschule* (S. 299–309). Weinheim: Juventa.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41–55.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley + Sons.
- Schavan, A. (2005). Schulen im Aufbruch. In T. Fitzner, T. Schlag, & M.W. Lallinger (Hrsg.), *Ganztagschule – Ganztagsbildung. Politik-Pädagogik-Kooperationen*. Bad Boll: Evangelische Akademie.
- Schlotter, M., & Wöbmann, L. (2010). *Frühkindliche Bildung und spätere kognitive und nicht-kognitive Fähigkeiten. Deutsche und internationale Evidenz*. München: Ifo Institut für Economic Research.
- Scholz, J., & Reh, S. (2009). Verwahrloste Familien – Familiarisierte Schulen. Zum Verhältnis von Schule und Familie in den Diskursen der deutschen Schulgeschichte seit 1800. In F.-U. Kolbe, S. Reh, B. Fritzsche, T.-S. Idel, & K. Rabenstein (Hrsg.), *Ganztagschule als symbolische Konstruktion. Fallanalysen zu Legitimationsdiskursen in schultheoretischer Perspektive* (S. 159–177). Wiesbaden: Springer VS.
- Schüpbach, M. (2014). Extended education and social inequality in Switzerland: Compensatory effects? An analysis of the development of language achievement with regard to structural and process-related aspects of social background. *Journal of Education Research Online*, 6(3), 95–114.
- Shonkoff, J. P., & Phillips, D. (2000). *From neurons to neighborhoods. The science of early child development*. Washington D.C: National Academy Press.
- Stanat, P., & Kunter, M. (2001). Geschlechterunterschiede in Basiskompetenzen. 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. 251–269). Opladen: Leske + Budrich.
- Stecher, L., Krüger, H.-H., & Rauschenbach, T. (2011). *Ganztagschule, Neue Schule? Eine Forschungsbilanz* (Zeitschrift für Erziehungswissenschaft: Sonderheft 15). Wiesbaden: Springer VS.
- Steiner, C. (2011). Teilnahme am Ganztagsbetrieb. Zeitliche Entwicklung und mögliche Selektionseffekte. In N. Fischer, H. G. Holtappels, E. Klieme, T. Rauschenbach, L. Stecher, & I. Züchner, (Hrsg.), *Ganztagschule: Entwicklung, Qualität, Wirkungen. Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagschulen (StEG)* (S. 57–75). Weinheim: Juventa.
- Stuart, E. A. (2007). Estimating causal effects using school-level datasets. *Educational Researcher*, 36, 187–198
- Stufflebeam, D. L. (1972). Evaluation als Entscheidungshilfe. In C. Wulf (Hrsg.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (S. 113–145). München: R. Piper & Co. Verlag.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2009). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1–67.
- Wagemaker, H. (1996). *Are girls better readers? Gender differences in reading literacy in 32 countries*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Watermann, R., Thurn, S., Tillmann, K.-J., & Stanat, P. (Hrsg.) (2005). *Die Laborschule im Spiegel ihrer PISA-Ergebnisse*. Weinheim: Juventa.
- Willems, A., & Holtappels, H. G. (2014). Pädagogische Prozessqualität an Ganztagschulen: Ausgewählte Befunde des bundesweiten StEG-Bildungsmonitoring 2011/2 zu Zielen und Konzepten von Ganztagsgrund- und Sekundarstufenschulen. In K. Drossel, R. Strietholt, & W. Bos (Hrsg.), *Empirische Bildungsforschung und evidenzbasierte Reformen im Bildungswesen* (S. 327–348). Münster: Waxmann.
- Züchner, I., & Fischer, N. (2014). Kompensatorische Wirkungen von Ganztagschulen – Ist die Ganztagschule ein Instrument zur Entkopplung des Zusammenhangs von sozialer Herkunft und Bildungserfolg? *Zeitschrift für Erziehungswissenschaft*, 14(2), 349–367.