

Frühübergang in ein grundständiges Gymnasium – Übergang in ein privilegiertes Entwicklungsmilieu? Ein Vergleich von Regressionsanalyse und Propensity Score Matching

Jürgen Baumert · Michael Becker · Marko Neumann · Roumiana Nikolova

Zusammenfassung: Der Beitrag geht der Frage nach, ob Schülerinnen und Schüler, die eine sechsjährige Grundschule bereits nach der 4. Klasse verlassen und auf ein grundständiges Gymnasium wechseln, höhere Lernzuwächse im Leseverständnis und in Mathematik erreichen. Auf der Grundlage von Daten der ELEMENT-Studie wurde die Leistungsentwicklung von Schülerinnen und Schülern aus Berliner grundständigen Gymnasien ($N=1758$) und Grundschulen ($N=3167$) während der 5. und 6. Jahrgangsstufe mithilfe von Regressions- und *Propensity Score Matching*-Analysen (PSM) modelliert. Nach Kontrolle von Eingangsunterschieden zwischen den Schulformen zeigten sich für das Leseverständnis keine statistisch signifikanten Unterschiede. Für die Mathematikleistung ließen sich kleine positive Effekte für die Frühübergänger lediglich mithilfe der Regressionsanalyse nachweisen. Das Ergebnis konnte im PSM nicht repliziert werden. Die Befunde sprechen gegen die Annahme, dass mit dem frühen Übergang auf ein grundständiges Gymnasium eine generelle Förderung der Lesefähigkeit und des mathematischen Verständnisses besonders leistungsfähiger Schülerinnen und Schüler erreicht wird. Methodische und inhaltliche Implikationen der Befunde und Grenzen ihrer Generalisierbarkeit werden diskutiert.

Schlüsselwörter: Übergang Gymnasium · Gegliedertes Schulsystem · Leistungsentwicklung · Schereneffekt · Leseverständnis · Mathematikleistung

Early transition into the academic track of secondary schooling – Transfer into a privileged learning environment? – A comparison of regression analysis and propensity score matching

Abstract: This article examines whether students' learning gains in reading and mathematics depend on the time of their transition from elementary school to the academic track of secondary schooling. Drawing on data from the ELEMENT study, the learning gains of fifth and sixth graders in elementary schools ($N=3167$) and academic-track schools with various curricular profiles ($N=1758$) in Berlin were modelled by regression and propensity score matching (PSM) analyses.

© VS-Verlag 2009

Die Reanalyse der ELEMENT-Daten geht auf eine Anregung der Senatsverwaltung für Bildung, Wissenschaft und Forschung in Berlin zurück. Wir danken Herrn Professor Dr. Dr. Rainer H. Lehmann, Humboldt-Universität zu Berlin, für die Zustimmung zu der Reanalyse sowie ihm und der Senatsverwaltung für die Bereitstellung des Datensatzes.

Prof. Dr. J. Baumert (✉) · M. Becker · M. Neumann · R. Nikolova (HU Berlin)
Max-Planck-Institut für Bildungsforschung, Forschungsbereich Erziehungswissenschaft und Bildungssysteme, Lentzeallee 94, 14195 Berlin, Deutschland
E-Mail: sekbaumert@mpib-berlin.mpg.de

When baseline differences between the school types were controlled, analyses for reading did not reveal statistically significant differences at all. Findings for mathematics were inconsistent depending on the method of analysis. Regression analysis showed statistically significant results of small effect size in favour of the academic-track schools. However, this finding could not be replicated by means of PSM. The study does not provide evidence for a generally positive effect of early transition to the academic track of secondary schooling with respect to mathematics and reading performance. Theoretical and methodological implications and limitations of the study are discussed.

Keywords: Academic school track · Learning gains · Mathematics achievement · Matthew effect · Reading comprehension · School tracking · Transition

1 Einleitung

Eines der umstrittensten Strukturmerkmale des deutschen Schulwesens ist die nur vierjährige Dauer der Grundschule und die sich daran anschließende frühe Verteilung der Schülerinnen und Schüler auf getrennte Sekundarschulformen. Diese Frühauslese wird als eine der maßgeblichen Ursachen für die hohen sozialen Disparitäten der Bildungsbeziehung und des Kompetenzerwerbs im Sekundarschulwesen angesehen (z. B. Becker & Schubert 2006). Für viele Kritiker des gegliederten Schulwesens gilt die sechsjährige Grundschule, die in den Ländern Berlin und Brandenburg anzutreffen ist und jetzt in Hamburg eingeführt wird, als Zwischenschritt zu einem gerechteren Schulwesen, das während der Pflichtschulzeit nur eine Schule für alle kennen sollte (Rösner 2005). Umso kritischer werden Ausnahmeregelungen in Berlin beurteilt, die einer Minderheit besonders leistungsstarker Schülerinnen und Schüler den frühen Übergang nach der 4. Klasse der Grundschule in ein sogenanntes grundständiges Gymnasium erlauben, das den Unterricht in der 5. Jahrgangsstufe beginnt. Berlin kennt den frühen Übergang auf ein grundständiges Gymnasium seit der Neustrukturierung des Schulwesens in den 1950er-Jahren. Er war in erster Linie ein Zugeständnis an humanistisch-bürgerliche Bildungstraditionen. Von der Möglichkeit des Frühübergangs machten in den letzten Jahren etwa 7 bis 8 Prozent der Schülerinnen und Schüler eines Altersjahrgangs mit steigender Tendenz Gebrauch.

Im Jahr 2003 vergab der damalige Berliner Senator für Schulwesen einen Auftrag an die Humboldt-Universität, anhand einer empirischen Untersuchung zu klären, welcher Personenkreis von der Möglichkeit des Frühübergangs Gebrauch macht und gleichzeitig die Kompetenzentwicklung von Schülerinnen und Schülern der 5. und 6. Jahrgangsstufe an Grundschulen und grundständigen Gymnasien zu vergleichen (Stryck & Lehmann 2003). In den Schuljahren 2002/2003 bis 2004/2005 führte Rainer H. Lehmann mit seiner Arbeitsgruppe an der Humboldt-Universität die Längsschnittstudie „Erhebungen zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin“ (ELEMENT) durch (Lehmann & Nikolova 2005a, 2005b). Im Frühjahr 2008 legten Lehmann und Lenkeit (2008) den primär deskriptiven Abschlussbericht der Studie vor. Darin berichten sie, dass Schülerinnen und Schüler an grundständigen Gymnasien während der 5. und 6. Jahrgangsstufe in den zentralen Domänen Lesen und Mathematik eine deutlich günstigere Kompetenzentwicklung zu verzeichnen hätten als Schülerinnen und

Schüler gleicher Ausgangsleistung an Grundschulen. Diese Fragestellung soll im Folgenden noch einmal aus stärker analytischer Perspektive aufgenommen werden.

2 Bildungsprogramme und Leistungsdifferenzierung

Alle modernen Dienstleistungsgesellschaften kennen die institutionelle Ausdifferenzierung von Bildungsprogrammen, die zu individuellen Bildungsprofilen, oftmals auch unterschiedlichen Abschlüssen und zu differenten Karrierepfaden führen. Fast immer ist diese Programmdifferenzierung mit einer impliziten oder expliziten Leistungsdifferenzierung und sozialen Entmischung verbunden. In Deutschland beginnt diese Differenzierung nach der 4. oder 6. Jahrgangsstufe sehr früh. Eine oft übersehene Besonderheit des deutschen Sekundarschulsystems ist die Tatsache, dass sich die gegliederte Sekundarstufe in den letzten Dekaden von einer Programm- zu einer – zumindest der Intention nach – reinen Leistungsdifferenzierung entwickelt hat. Noch in den 1960er-Jahren konnte man die Schulformgliederung bildungstheoretisch begründen, indem man eine Korrespondenz zwischen Begabungstypen und den curricularen Programmen der weiterführenden Schulen postulierte. Spätestens in den 1970er-Jahren setzte sich die Idee einer universellen, wissenschaftsorientierten Grundbildung für alle durch und entzog dem gegliederten System die bildungstheoretische Grundlage. Unter der bis heute gültigen Zielvorstellung der Durchlässigkeit des Systems kam es zu einer curricularen Angleichung der Schulformen durch Harmonisierung der Stundentafeln und Vereinheitlichung der Lehrpläne. Inhaltliche Verschiedenheit wurde durch Niveauabstufung ersetzt. Damit war der Schritt zur äußeren Leistungsdifferenzierung getan, die nicht mehr bildungstheoretisch, sondern nur noch didaktisch-methodisch begründet werden kann. In der Lehrerbildung und den didaktischen Traditionen der Schulformen haben sich noch Rudimente der älteren Programmdifferenzierung erhalten (Baumert, Cortina & Leschinsky 2008; Herrlitz et al. 2005). Dies zeigt, dass es bislang zu keinem spannungsfreien Übergang gekommen und mit beiden Traditionen – der Programm- und der Leistungsdifferenzierung – zu rechnen ist.

Das Berliner Schulsystem bietet mit den grundständigen Gymnasien eine historisch ältere Programmdifferenzierung an. Die Bildungsprogramme antworten entweder auf besonderes musikalisches oder sportliches Talent oder offerieren altsprachliche oder bilinguale Bildungsgänge, die – so die Annahme – besondere kognitive und motivationale Anforderungen stellten und zusätzliche Lern- und Übungszeit benötigten. Die Begründung für die institutionelle Trennung liegt dann auf der Hand. Da das Interesse an diesen Programmen auf eine Minderheit der Schülerinnen und Schüler beschränkt sei, könnten sie nur an ausgewählten Standorten angeboten werden. Dafür seien Grundschulen weniger geeignet, da in der Regel die fachliche Expertise im Lehrkörper fehle. Altsprachliche Angebote gibt es in Berlin an neun und bilinguale Züge an fünf grundständigen Gymnasien. Hinzu kommen zwei Standorte mit musikalischem und ein Standort mit sportlichem Schwerpunkt.

Einen Grenzfall stellen die sogenannten Schnellläuferzüge dar, die in den letzten Jahren an 13 Gymnasien grundständig eingerichtet wurden. Sie sollen besonders begabte Schülerinnen und Schüler in acht und demnächst in sieben Jahren zum Abitur führen, indem die Mittelstufe um ein Jahr gekürzt und das Curriculum entsprechend komprimiert wird. Man kann darüber streiten, ob es sich bei diesem Angebot tatsächlich um eine Pro-

grammdifferenzierung handelt, die einen besonderen Bildungsweg begründet. Das verdichtete Curriculum mag als Indiz für eine weitere Programmdifferenzierung dienen.

Alle grundständigen Bildungsangebote haben eines gemeinsam: Sie setzen die Überwindung hoher Leistungshürden in der Grundschule voraus. Zur Regelung des Frühübergangs in grundständige Gymnasien gibt es in Berlin institutionelle Leitlinien, die den einzelnen Schulen die Richtung vorgeben, aber auch Interpretations- und Ausgestaltungsspielraum lassen. Nach unseren Recherchen an den grundständigen Gymnasien lässt sich der Zuweisungsprozess folgendermaßen beschreiben:¹ Grundlage für den Übergang in ein grundständiges Gymnasium ist generell eine Gymnasialempfehlung der abgebenden Grundschule und ein Notendurchschnitt in den Fächern Deutsch, Mathematik und Sachkunde, der nur in Ausnahmefällen unter der Note 2,0 (gut) liegen darf. Für den Übergang in ein grundständiges Gymnasium mit sportlichem oder musikalischem Profil kommen zusätzliche Leistungsanforderungen im jeweiligen Bereich und längere Sport- bzw. praktische Musikerfahrungen hinzu. Die aufnehmenden Gymnasien führen ferner in unterschiedlicher Form Aufnahmegespräche, teilweise auch informelle Leistungsprüfungen durch. Bei einem Überhang an Bewerbern mit gleicher, ausreichender Qualifikation entscheidet dann das Los.

Die Frage nach der besonderen Förderungswirkung dieser grundständigen Angebote, die sich mit Ausnahme der Schnellläuferklassen ja gerade durch ein besonderes inhaltliches Profil auszeichnen, mit einer Untersuchung zur Leistungsentwicklung in den Kernbereichen der Lesekompetenz und Mathematik beantworten zu wollen, ist nicht unmittelbar naheliegend; denn bei einer Evaluation von curricularen Programmen wird man zuallererst prüfen, ob die jeweils spezifischen Bildungsziele erreicht werden – also ob ordentlich Latein gelernt wird, frühe Sicherheit im Englischen oder Französischen erreicht wird oder musikalische und sportliche Talente tatsächlich gefördert werden. Bei einem verkürzten Gymnasialangebot kann man die Frage nach der Zielerreichung überhaupt erst stellen, wenn die Vergleichsgruppe die Mittelstufe des Gymnasiums durchlaufen hat und nicht, solange der Schulformwechsel der Kontrollgruppe noch bevorsteht. Vergegenwärtigt man sich ferner, dass in den Kernbereichen Deutsch und Mathematik in Berlin die Standardstudentenafel der 5. und 6. Jahrgangsstufe der Grundschule auch für die grundständigen Gymnasien gilt, wird deutlich, dass mit dem Untersuchungsauftrag nicht die Überprüfung des spezifischen Anliegens der grundständigen Gymnasien gemeint sein kann. Geprüft werden sollte vielmehr, ob an den grundständigen Gymnasien differenzielle Lernumwelten entstehen, die den Schülerinnen und Schülern Vorteile auch beim Erwerb von Kernkompetenzen verschaffen, die vergleichbar leistungsfähigen Schülerinnen und Schülern an Grundschulen vorenthalten werden (Stryck & Lehmann 2003).

3 Schulformen als differenzielle Lern- und Entwicklungsmilieus: Die Forschungslage

Schulformen als differenzielle Lern- und Entwicklungsmilieus zu betrachten, haben erstmals Baumert und Köller (1998) vorgeschlagen und damit die Aufmerksamkeit von der Einzelschule wieder auf schulformbezogene Schulumwelten gelenkt, die im Wechselspiel zwischen leistungsmäßiger, sozialer und kultureller Selektion einerseits und bil-

dungstheoretischen und didaktischen Traditionen sowie unterschiedlichen Formen der Lehrerbildung andererseits entstehen. Auf der Grundlage der Längsschnittstudie „Bildungsverläufe und psychosoziale Entwicklung im Jugend- und jungen Erwachsenenalter“ (BIJU) konnten sie für verschiedene Fächer differenzielle Leistungsentwicklungen an Sekundarschulformen nachweisen. Auch bei Kontrolle der unterschiedlichen Eingangsselektivität der Schulformen traten typische Schereneffekte auf (Baumert et al. 1996; Baumert, Köller & Schnabel 2000; Köller & Baumert 2001, 2002).

Eine Reihe weiterer Untersuchungen brachte zusätzliche Evidenzen, die den Befund differenzieller Lernverläufe an Sekundarschulformen stützen. Kunter (2005) berichtet auf Grundlage der deutschen Längsschnittkomponente der TIMS-Studie für Mathematik einen deutlichen Schereneffekt in der 8. Klasse zugunsten der selektiveren Schulformen, der sich auch bei Kontrolle der Ausgangsbedingungen findet. Eine von Becker et al. (2006) vorgelegte Reanalyse desselben Datensatzes kommt unter Nutzung latenter Wachstumsmodelle zum gleichen Ergebnis. Pekrun et al. (2006) haben in Bayern eine Längsschnittuntersuchung (PALMA) durchgeführt, die es erlaubt, die Leistungsentwicklung von Schülerinnen und Schülern im Fach Mathematik über die gesamte Sekundarschulzeit zu verfolgen. Die einschlägigen Analysen zeigen ebenfalls den bekannten Schereneffekt (Pekrun et al. 2006).

In anderen Studien ist die Befundlage jedoch *nicht* eindeutig. Die beiden großen, auf Vollerhebungen eines Jahrgangs beruhenden Hamburger Längsschnittstudien – „Aspekte der Lernausgangslage und Lernentwicklung“ (LAU) und „Kompetenzen und Einstellungen von Schülerinnen und Schülern“ (KESS) – zeigen variierende Effekte in Abhängigkeit von Schulstufe, Unterrichtsfach und der Auswertungsmethode. Das Ergebnismuster ist in beiden Studien inkonsistent (Bonsen, Bos & Gröhlich 2007; Bos et al. 2006; Lehmann, Gänsfuß & Peek 1999; Lehmann et al. 2001). Keinen Schereneffekt fanden Schneider, Knopf und Stefanek (2002) sowie Schneider und Stefanek (2004) bei ihrer Analyse des Münchener LOGIK-Längsschnitts.

Neue Befunde wurden jüngst aus der Studie „Deutsch Englisch Schülerleistungen International“ (DESI) vorgelegt. Die Ergebnisse fallen für einzelne Kompetenzaspekte im Fach Deutsch unterschiedlich aus (Gailberger & Willenberg 2008; Neumann & Lehmann 2008). Dagegen deuten sich für das Fach Englisch differenzielle Leistungsentwicklungen in Abhängigkeit von der Schulform während der 9. Jahrgangsstufe an (Klieme 2006).

Fasst man die Ergebnisse zusammen, so wird man sagen müssen, dass die Befundlage über alle Fächer hinweg für die Schulformen der Sekundarstufe I keineswegs eindeutig ist. Am ehesten zeigen sich konsistente Befunde in den überregionalen Studien für die Fächer Mathematik und Englisch. Für die Entwicklung der Lesekompetenz ergibt sich kein geschlossenes Bild.

Zur Frage, inwieweit die Befunde einer institutionellen Leistungsdifferenzierung auch auf eine frühe Programmdifferenzierung in der kritischen Phase zum Ende der Grundschulzeit übertragbar sind, liegen bislang kaum aussagekräftige Befunde vor. Einen ersten Schritt haben Lehmann & Nikolova (2005b) und Lehmann & Lenkeit (2008) mit der Vorlage ihrer ELEMENT-Berichte geleistet. In einer deskriptiven Darstellung (Lehmann & Lenkeit 2008, S. 22–26 und S. 32–35) vergleichen sie den Zusammenhang zwischen Lese- bzw. Mathematikleistungen am Ende der Klassenstufe 4 und den entsprechenden Leistungen am Ende der Klassenstufe 6 für Schülerinnen und Schüler an grundständigen

Gymnasien sowie an Grundschulen. Von erkennbaren Unterschieden in den Kurvenverläufen schließen sie auf differenzielle Entwicklungsprozesse an Grundschulen und grundständigen Gymnasien. Eine Regressionsanalyse mit einem kleinen Satz an Kovariaten stützt den deskriptiven Befund. Die Autoren schreiben: „Aus der Perspektive des individuellen Lerner bzw. seiner Eltern ist sowohl für die Entwicklung des Leseverständnisses als auch für die Mathematik der Befund eindeutig: Unabhängig von der Lernausgangslage werden an den grundständigen Gymnasien jeweils deutlich höhere Lernfortschritte erzielt: Bei gleicher Lernausgangslage am Anfang der Klassenstufe 5 stellen sich die am Ende der Klassenstufe 6 erreichten Lernstände günstiger dar, wenn der Unterricht in einem grundständigen Gymnasium besucht wurde, als es anscheinend im Falle des Verbleibs auf der Grundschule der Fall wäre.“ (Lehmann & Lenkeit 2008, S. 38)

4 Fragestellungen

Im Folgenden sollen drei Fragestellungen untersucht werden. Zunächst soll die Schülerschaft an grundständigen Gymnasien im Vergleich zur Alterskohorte an Grundschulen hinsichtlich der Noten, Leistungen, motivationalen Orientierungen sowie sozialen und kulturellen Merkmale der Herkunftsfamilien beschrieben werden. Daran schließt sich die theoretische und empirische Rekonstruktion des Anwahl- bzw. Zuweisungsprozesses zu einem grundständigen Gymnasialprogramm an. Es soll geprüft werden, ob der Wechsel in ein grundständiges Gymnasium in substantieller Weise vorhersagbar ist. Die dritte Gruppe von Analysen beschäftigt sich mit der zentralen Fragestellung des Aufsatzes, ob in den Domänen Lesen und Mathematik differenzielle Leistungsentwicklungen an Grundschulen und grundständigen Gymnasien nachweisbar sind, die nicht auf unterschiedliche individuelle Lernraten zurückführbar sind (Becker 2008).

5 Methodisches Vorgehen

5.1 Untersuchungsanlage

Die Untersuchung „Erhebungen zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin“ (ELEMENT) wurde als Längsschnittstudie mit drei Messzeitpunkten angelegt. Mit der Studie sollte die Leistungs- und Motivationsentwicklung von Schülerinnen und Schülern im Übergangsbereich zwischen Primar- und Sekundarstufe an Berliner Grundschulen und die Übergangsprozesse an grundständige Gymnasien untersucht werden. Die Leistungsentwicklung wurde in den Fächern Deutsch und Mathematik vom Ende der 4. bis Ende der 6. Jahrgangsstufe erfasst. Die Leistungstests wurden durch Eltern- und Schülerbefragungen ergänzt. Erhebungen fanden für die Grundschüler im Juni und für die grundständigen Gymnasiasten im September 2003 und dann für alle Gruppen im Mai 2004 und Mai 2005 statt.

5.2 Stichprobe

Die Stichprobe von Viertklässlern an Grundschulen wurde im Jahre 2003 auf der Grundlage vollständiger Klassenlisten für das Land Berlin gezogen. Zu jeder gezogenen Klasse wurde bei mehrzügigen Grundschulen eine zweite Klasse derselben Schule hinzugezogen. Insgesamt wurden 140 Klassen in 71 Grundschulen in die Stichprobe aufgenommen.

An der Untersuchung nahmen – wie Tabelle 1 zeigt – zum ersten Messzeitpunkt 3293 Grundschülerinnen und Grundschüler teil. Von diesen wechselten 155 nach der Sommerpause an ein grundständiges Gymnasium. Von den verbleibenden 3138 Schülerinnen und Schülern traten 2761 in die 5. Jahrgangsstufe derselben Grundschule über. 377 der getesteten Viertklässler mussten die Klasse wiederholen oder wechselten ihre Klasse oder Schule. Sie wurden nicht in die Längsschnittstudie einbezogen. Die Basisstichprobe für Grundschulen umfasst damit 2761 Personen.

Tab. 1: Realisierte Stichprobe nach Erhebungszeitpunkt und Schulform

Messzeitpunkt (MZP)	Grundschule				Grundständiges Gymnasium			
	Teilnehmer an der Untersuchung (Bruttosample)	Teilnehmer am Längsschnitt (EL 4/5/6)	Verluste im Längsschnitt (EL 4/5/6)	Zugänge	Teilnehmer an der Untersuchung (Bruttosample)	Teilnehmer am Längsschnitt (EL 4/5/6)	Verluste im Längsschnitt (EL 4/5/6)	Zugänge
Juni/September 2003 ¹	3138 ²	2761 ³	(377)	–	1724	1681 ⁴	(43)	–
Mai 2004	3080			319	1701			20
Mai 2005	2963	2595	166	87	1700	1624	57	57
Insgesamt	3544	3167 ⁵	166	406	1794	1758 ⁵	57	77
Davon:								
1 MZP	502	125	–	–	101	65	–	–
2 MZPe	447	447	–	–	69	69	–	–
3 MZPe	2595	2595	–	–	1624	1624	–	–

1 Die erste Erhebung fand an Grundschulen im Juni 2003 (Ende Klassenstufe 4) und an grundständigen Gymnasien im September 2003 (Anfang Klassenstufe 5) statt.

2 An der ersten Grundschuluntersuchung nahmen 155 Personen teil, die zum Beginn der 5. Klasse an ein grundständiges Gymnasium wechselten. Sie wurden doppelt getestet und sind ausschließlich als Teil der Längsschnittstichprobe für das grundständige Gymnasium ausgewiesen.

3 Basisstichprobe für die Grundschulen nach Ausschluss von 377 Fällen, die im Juni 2003 untersucht wurden, aber nicht mit derselben Klasse in die 5. Jahrgangsstufe wechselten.

4 Basisstichprobe für die grundständigen Gymnasien nach Ausschluss von 43 Fällen, die im September 2003 untersucht wurden, im Laufe des Schuljahres aber ihre Klasse verlassen haben.

5 In der vorliegenden Untersuchung wurden nur die Abgänge nach der ersten Erhebungswelle (Grundschule $N=377$ bzw. grundständiges Gymnasium $N=43$) ausgeschlossen. Fehlende Angaben für die Abgänge nach der zweiten Erhebungswelle sowie die Zugänge zum zweiten und dritten Messzeitpunkt wurden imputiert.

Für die grundständigen Gymnasien wurde eine Vollerhebung der 5. Jahrgangsstufe im Schuljahr 2003/2004 angestrebt. Es wurden 59 fünfte Klassen an 31 öffentlichen grundständigen Einrichtungen zu Beginn der Klassenstufe 5 im September 2003 getestet. An der Untersuchung beteiligten sich 1724 Schülerinnen und Schüler. Darunter waren 155 Schülerinnen und Schüler, die bereits an der Untersuchung der Grundschulklassen im Mai 2003 teilgenommen hatten. Um Retest-Effekte zu vermeiden, haben wir bei den doppelt getesteten Schülern auf die Leistungsergebnisse der ersten Testung (Ende Klasse 4 in den Grundschulen) zurückgegriffen.

Die Ausschöpfungsquoten lagen für beide Teilstichproben auf der Schul- und Klassenzebene bei 100 Prozent. Die Teilnahme an den Leistungstests war für alle Schülerinnen und Schüler verpflichtend. In der ersten Erhebungswelle nahmen an den Grundschulen 95,5 und an den grundständigen Gymnasien 98,1 Prozent der Schülerinnen und Schüler der gezogenen Klassen an den Leistungsuntersuchungen teil. Die Ausschöpfungsquoten für den Schülerfragebogen lagen an den Grundschulen bei 87,9 und an grundständigen Gymnasien bei 93,6 Prozent. Den Elternfragebogen füllten an Grundschulen 80,3 und an grundständigen Gymnasien 91,3 Prozent der angeschriebenen Eltern aus.

Von der Basisstichprobe mit insgesamt 4443 Untersuchungsteilnehmern schieden 223 Schülerinnen und Schüler innerhalb der beiden Schuljahre aus der Untersuchung aus. Dies entspricht einer Verminderungsrate von 5,0 Prozent. Ein Teil dieser Verluste war vollständig zufallsbedingt, etwa wenn einzelne Klassen aufgelöst wurden oder die Daten einer ganzen Schule auf dem Postweg verloren gingen (vgl. Lehmann & Lenkeit 2008). Ein kleiner Teil der Ausfälle ging auf Umzüge der Eltern zurück. Ihnen stehen Zugänge in vergleichbarem Umfang gegenüber. Diese Mobilitätsprozesse führen vermutlich zu keiner ernsthaften Verzerrung der Stichprobe. Ein nicht unbedeutender Teil der Ausfälle ist jedoch auf Klassenwiederholungen, Schrägversetzungen von grundständigen Gymnasiasten an Grundschulen oder Übergänge aus der Grundschule an Förderschulen zurückzuführen und damit leistungsbedingt. Ebenso sind die Zugänge an Grundschulen zu großen Teilen Wiederholer oder schrägversetzte Gymnasiasten. Werden diese Stichprobenbewegungen nicht adäquat berücksichtigt, führen sie zu verzerrten Punktschätzungen. In der vorliegenden Untersuchung wurden nur die Abgänge nach der ersten Erhebungswelle (Grundschule $N=377$ bzw. grundständiges Gymnasium $N=43$) ausgeschlossen. Die fehlenden Angaben für die Abgänge nach der zweiten Erhebungswelle sowie für die Zugänge zum zweiten und dritten Messzeitpunkt wurden imputiert (vgl. Abschnitt 5.4 *Fehlende Daten*). Die Analysen basieren damit auf den Daten von 3167 Grundschülerinnen und Grundschulern sowie 1758 Schülerinnen und Schülern aus grundständigen Gymnasien.

5.3 Instrumente

Leistungstests: In der ELEMENT-Studie wurden zwei Leistungsmaße – Lesekompetenz und mathematisches Verständnis – längsschnittlich erfasst. Die Testaufgaben wurden aus der PIRLS/IGLU-Studie 2001 (Bos et al. 2003), der LAU-Untersuchung (Lehmann & Peek 1997; Lehmann, Gänsfuß & Peek 1999), der internationalen TIMS-Grundschulstudie (Bos et al. 2003) und aus der brandenburgischen Untersuchung QUASUM (Lehmann et al. 2000) übernommen. Es wurde ein Ankeritem-Design realisiert, das die Skalierung

der Lese- und Mathematiktests jeweils auf einer Metrik erlaubt. Eine genauere Beschreibung der Testkonstruktion findet sich bei Lehmann & Nikolova (2005a, 2005b) und Lehmann & Lenkeit (2008).

Die Tests wurden Rasch-skaliert. Die Skalierung wurde mit dem Programm ConQuest durchgeführt; für alle Analysen wurde das Partial-Credit-Modell benutzt (Wu, Adams & Wilson 1998). Die Personenparameter wurden als gewichtete Likelihood-Schätzer (Warm 1989) bestimmt. Die Logits wurden anschließend linear auf einen Mittelwert von 100 und eine Standardabweichung von 15 für den ersten Messzeitpunkt transformiert. Die Reliabilität der Lesetests lag zu den einzelnen Erhebungszeitpunkten zwischen $r_{KR20} = .85$ und $.86$, die des Mathematiktests zwischen $r_{KR20} = .84$ und $r_{KR20} = .93$.

Kognitive Grundfähigkeiten: Die kognitiven Grundfähigkeiten wurden zum ersten Messzeitpunkt einmalig mit zwei Untertests aus dem kognitiven Fähigkeitstest (KFT) erhoben (Heller & Perleth 2000). Die Subtests erfassen verbales und figurales Schlussfolgern und gelten als Markertest für fluide Intelligenz. Die Reliabilität der gemeinsam skalierten Subtests liegt bei $r_{KR20} = .94$.

Noten: Die Halbjahresnoten in der 4. Klasse für die Fächer Deutsch, Mathematik, Sachkunde, Musik und Sport wurden den Schulakten entnommen.

Motivation: Als Indikator für schulische Motivation wird das Leseinteresse – erfasst zum ersten Messzeitpunkt – herangezogen. Die Interessenskala besteht aus fünf Items mit einer internen Konsistenz von Cronbachs Alpha = $.78$.

Indikatoren des sozialen und kulturellen Milieus: Merkmale der sozialen, kulturellen und ethnischen Herkunft wurden über den Schüler- und Elternfragebogen erhoben. Der soziale Status der Familien wurde mit dem *International Social-Economic Index* (ISEI) erfasst (Ganzeboom & Treiman 2003). Als Indikator für den Status der Familie wird jeweils der höchste soziale Status des Elternpaares verwendet (HISEI). Das Bildungsniveau der Familie wird durch sieben hierarchisch geordnete Qualifikationsstufen indiziert, in die Elternangaben zu Schulabschlüssen und Berufsqualifikationen eingehen (vgl. Baumert & Schümer 2001). Der Migrationsstatus der Familie wird durch das Geburtsland der Eltern definiert. Wenn mindestens ein Elternteil im Ausland geboren wurde, stammt das Kind aus einer Migrationsfamilie. Als Indikatoren des kulturellen Kapitals werden Schülerangaben zum häuslichen Buchbestand und den im Haushalt verfügbaren Bildungsgütern herangezogen.

Besuchte Schulform in der 7. Jahrgangsstufe: Die Angaben zu den Übergangsentscheidungen nach der 6. Klasse wurden zum dritten Messzeitpunkt am Ende der 6. Klasse, zu dem die Übergangsentscheidung bereits gefallen war, durch Elternbefragung erfasst.

5.4 Fehlende Daten

Zum derzeitigen Stand der Forschung gilt die multiple Imputation (MI) als das angemessenste Verfahren des Umgangs mit fehlenden Werten (Graham, Cumsille & Elek-Fisk 2003; Lüdtke et al. 2007). In unserem Fall wurden für alle in die Analysen einbezogenen Variablen mit dem Programm NORM 2.03 (Schafer 1999) fünf vollständige Datensätze erzeugt, in denen die fehlenden Werte durch sogenannte *plausible values* ersetzt sind. Die Ergebnisse der Analysen für die fünf Datensätze wurden gemäß den Regeln von

Rubin (1987) zusammengefasst. Die entsprechenden Formeln sind in der Analyseoption *Type=Imputation* von *Mplus 5.0* (Muthén & Muthén 1998–2008) implementiert.

5.5 Datenanalyse

Analysestrategie: Im ersten Schritt wird die Klientel der grundständigen Gymnasialangebote im Vergleich zur Schülerschaft der anderen Sekundarschulformen multivariat beschrieben. Daran schließt sich die Modellierung des Frühübergangs mithilfe logistischer Regressionsanalysen an. Der dritte Schritt führt zum analytischen Zentrum des Aufsatzes. Hauptanliegen der Untersuchung ist es, zu einer möglichst unverzerrten und konsistenten Schätzung des sogenannten *Treatment Effects* (TE) der grundständigen Gymnasien im Vergleich zur Förderungswirkung der Lernumwelt an Grundschulen zu gelangen. Zu diesem Zweck sollen zwei methodische Vorgehensweisen, Kovarianz- bzw. Regressionsanalyse und *Propensity Score Matching* (PSM), miteinander verglichen werden.

Multivariate Beschreibung der Klientel: In einer multivariaten Analyse wird zunächst geprüft, ob sich die Klientel der unterschiedlichen grundständigen Gymnasien untereinander und im Vergleich zu anderen Sekundarschulformen systematisch unterscheidet. Als Kennziffern werden berücksichtigt: Noten, Schulleistungen, kognitive Grundfähigkeiten, schulbezogene Motivation und Merkmale der sozialen und kulturellen Herkunft.

Rekonstruktion des Frühübergangs auf ein grundständiges Gymnasium: Will man in Beobachtungsstudien, die keine randomisierte Zuweisung zu Experimental- und Kontrollgruppen erlauben, Schlüsse auf die Wirksamkeit pädagogischer Maßnahmen oder Programme ziehen, besteht die zentrale Herausforderung in der theoretisch und empirisch zutreffenden Rekonstruktion des jeweiligen Zuweisungsprozesses. Nur wenn der Zuweisungsprozess im Hinblick auf den gewünschten Effekt der Behandlung zu vernachlässigen ist, sind erwartungstreue und konsistente Schätzungen der Wirkung von Interventionen möglich (Morgan & Winship 2007).

Vor dem Hintergrund des bereits beschriebenen institutionalisierten Zuweisungsverfahrens schlagen wir ein theoretisches Modell des Frühübergangs vor, in das fünf Komponenten – Leistung, Motivation, soziale, kulturelle und ethnische Herkunft, Geschlecht und institutionelle Angebotsstruktur – eingehen, die – soweit dies möglich und nötig ist – multipel indikatorisiert werden. Für die Leistungskomponente stehen die Halbjahresnoten in der 4. Klasse in den Fächern Deutsch, Mathematik, Sachkunde, Musik und Sport und die Testleistungen für Lesefähigkeit, Mathematik und kognitive Grundfähigkeiten. Als motivationales Merkmal wird das Leseinteresse herangezogen. Die soziale, kulturelle und ethnische Herkunft wird durch fünf Variablen abgebildet: den höchsten Sozialstatus in der Familie (HISEI), den höchsten Bildungs- bzw. Berufsabschluss der Eltern, den Buchbestand und die im Haushalt verfügbaren Bildungsgüter sowie den Migrationsstatus. Als Indikator für die Angebotsstruktur dient die Anzahl grundständiger Gymnasien im Stadtbezirk der jeweils abgebenden Grundschule. Dabei wurden die alten, kleinräumigeren Bezirksgrenzen Berlins gewählt, die bis Ende 2003 galten. Die Modellschätzung zur Vorhersage des Übergangsverhaltens am Ende der 4. Jahrgangsstufe wird mit binären logistischen Regressionsanalysen vorgenommen.

Regressionsanalytische Schätzung differenzieller Fördereffekte: Zunächst wollen wir daran erinnern, dass mit der Kovarianz- oder Regressionsanalyse der durchschnittliche

Programmeffekt in der *gesamten* Population (*average treatment effect* – ATE) geschätzt wird, und nicht nur die durchschnittliche Behandlungswirkung auf diejenigen, die typischerweise in das jeweilige Programm gelangen (*average treatment effect for the treated* – ATT) (Heckman 2000; Morgan & Winship 2007). Um den durchschnittlichen Programmeffekt in der gesamten Population erwartungstreu und konsistent schätzen zu können, müssen unter anderem die folgenden Voraussetzungen erfüllt sein (vgl. Morgan & Winship 2007, S. 31 ff.): (1) Unter Kontrolle der Kovariaten müssen Merkmale des Zuweisungsprozesses (*assignment mechanism*) und Fehlerterm (*error term*) – in den alle anderen, nicht im *treatment* spezifizierten Ursachen eingehen – unkorreliert sein. Dies bedeutet, dass auch keine unbeobachtete Heterogenität zwischen Programm- und Kontrollgruppe vorhanden sein darf (*ignorability of assignment*). Da dies grundsätzlich nicht nachweisbar ist, hängt die Plausibilität der Erfüllung dieser Voraussetzung ausschließlich von der Qualität der theoretischen und empirischen Rekonstruktion des Zuweisungsprozesses ab. Eine Fehlspezifikation des Modells kann zu Verzerrungen der Schätzungen der Programmwirkung führen. (2) Die Programmwirkung ist für alle Ausprägungen der Kovariaten gleich. Es gibt also keine Interaktion zwischen Variablen, die zur Adjustierung verwendet werden, und Maßnahmen des Programms. (3) Die Wirkung des Programms auf einen Teilnehmer ist sowohl von der Art der Zuweisung als auch von den übrigen dem Programm zugewiesenen Personen unabhängig (*stable unit treatment value assumption* – SUTVA).

Die Erfüllung der ersten Voraussetzung ist in *allen* Beobachtungsstudien zentral für die Belastbarkeit kausaler Schlüsse. Wir haben ein theoretisches Modell des Frühübergangs vorgeschlagen, das anhand der verfügbaren Daten empirisch geprüft werden soll. Dabei kann allerdings nicht ausgeschlossen werden, dass zwischen der Motivstruktur der Eltern, ein grundständiges gymnasiales Bildungsprogramm auszuwählen, und einem leistungsfördernden Familienmilieu ein Zusammenhang besteht, der nicht durch die verfügbaren Variablen zur sozialen und kulturellen Herkunft abgebildet wird. Ebenso fehlt die Erfassung der praktischen sportlichen und musikalischen Betätigung der Grundschüler. Sind diese Merkmale im Hinblick auf das Leseverständnis und die mathematische Kompetenz leistungsrelevant, ist unser Modell unterspezifiziert und man müsste auch nach Adjustierung mit weiterhin bestehender unbeobachteter Heterogenität zwischen den Gruppen rechnen.

Die zweite Annahme der universellen Wirkung des Programms über alle Ausprägungen der Kovariaten hinweg ist für hochselektive Bildungsprogramme ausgesprochen unplausibel: Eine Schülerin oder einen Schüler mit deutlichen Leistungsschwächen in den Kernfächern einem fünf- bis sechsständigen Latein-, Französisch, Musik- oder Sportkurs oder in den Schnellläuferzügen einem erhöhten Durchnahmetempo zusätzlich zu den übrigen Belastungen auszusetzen, wird die Defizite in den Basiskompetenzen Lesen und Mathematik kaum beseitigen. Viel eher wäre mit Überforderung und generellem Versagen zu rechnen.

Ebenso wenig plausibel ist die dritte Annahme. Man kann nicht ausschließen, dass die Motivation, in ein grundständiges Gymnasium zu wechseln, selbst eine Voraussetzung für die Wirkung des jeweiligen Programms ist. Ebenso ist anzunehmen, dass Qualität und Wirkung des Bildungsprogramms von der Selektivität des Zugangs abhängen. Bei einer

Öffnung der Programme wird eine spezifische Wirkung des selektiv erzeugten Leistungsmilieus immer weniger wahrscheinlich.

Diese Überlegungen zeigen, dass die übliche regressionsanalytische Schätzung der differenziellen Förderwirkung unterschiedlicher Schulformen möglicherweise nicht zu verzerrungsfreien und konsistenten Ergebnissen führt. Sie verdeutlichen auch, dass die Frage nach dem durchschnittlichen Programmeffekt für die gesamte Population (ATE) in unserem Fall nicht sonderlich sinnvoll ist. Von Interesse ist vielmehr die durchschnittliche Wirkung der Programme für die typischerweise in ein grundständiges Gymnasium wechselnde Schülergruppe (ATT). Reduziert man den Anspruch in dieser Weise, lockert man die zweite und dritte Annahme. Dann bieten sich andere Vorgehensweisen, wie zum Beispiel *Matching*-Verfahren, an, die voraussetzungsärmer als die Regressions- oder Kovarianzanalyse sind. Dies spricht für einen expliziten Vergleich zwischen regressionsanalytischen und alternativen Auswertungsverfahren.

Im Folgenden wird zunächst eine regressionsanalytische Auswertungsstrategie verfolgt, die als ein HL-Modell in *Mplus* spezifiziert wurde, um der Mehrebenenstruktur Rechnung zu tragen. Das theoretisch spezifizierte und empirisch am besten angepasste Modell des Übergangsverhaltens wird vollständig in den Block der Kovariate aufgenommen (auf Level 1). Die Effekte des Besuchs eines grundständigen Gymnasiums sowie die Anzahl grundständiger Gymnasien im Wohnbezirk werden auf Level 2 (Ebene der Schulklassen) in das Modell aufgenommen und im Hinblick auf das Leseverständnis und die Mathematikkompetenz am Ende der 6. Jahrgangsstufe geschätzt.

Schätzung differenzieller Fördereffekte mit Propensity Score Matching: Eine Alternative zur ANCOVA stellt in vielen Fällen das sogenannte *Propensity Scores Matching* (PSM) dar, das von Ökonomen routinemäßig für Kausalanalysen genutzt wird (Heckman 2000). Es hat in jüngerer Zeit verstärkt auch Eingang in die Psychologie, jedoch noch kaum in die Erziehungswissenschaft gefunden (Schneider et al. 2007).

Im PSM werden auf Grundlage der durch das Zuweisungsmodell spezifizierten Merkmale für jede Person der gesamten Stichprobe Übergangswahrscheinlichkeiten in die Treatmentgruppe (in unserem Fall die Gruppe der Gymnasiasten) geschätzt – die sogenannten *Propensity Scores*. Mit deren Hilfe wird für jede Person der Treatmentgruppe ein vergleichbarer Fall in der Kontrollgruppe gesucht (Morgan & Winship 2007). Wie beim regressionsanalytischen Auswertungsverfahren hängt auch hier alles von der korrekten Rekonstruktion des Zuweisungsprozesses und damit der Validität der *Propensity Scores* ab. Werden allerdings die Übergangswahrscheinlichkeiten erwartungstreu und konsistent geschätzt und gelingt ein anschließendes *Matching* für den überwiegenden Teil der Programmbesucher, können belastbare Schlüsse auf die Wirksamkeit der grundständigen Gymnasialangebote für die typische Schülerschaft gezogen werden. Lassen sich keine differenziellen Fördereffekte nachweisen, gibt es darüber hinaus keinen guten Grund, solche für die übrige Grundschülerschaft zu erwarten.

Propensity Score Matching erfolgt üblicherweise in drei Schritten. Es werden zunächst die bedingten Zugangswahrscheinlichkeiten des durch das Zuweisungsmodells definierten Kovariatenvektors geschätzt und anschließend in Logits transformiert, da Logits bessere metrische Eigenschaften aufweisen als einfache bedingte Wahrscheinlichkeiten. Ein grafischer Vergleich der Verteilungen der *Propensity Scores* in der Programm- und Kontrollgruppe kann dann einen Eindruck von der Überlappung der beiden Verteilungen

geben. Dieser Vergleich erlaubt auch eine Schätzung der Größe der *Area of Common Support*, also des Bereichs, in dem – in unserem Fall – überhaupt vergleichbare Grundschüler und Gymnasiasten zu finden sind. Nur für die *Area of Common Support* sind vergleichende Wirkungsaussagen ohne weitere Zusatzannahmen möglich (King & Zeng 2006).

Im zweiten Schritt erfolgt das *Matching* der Fälle, wobei das doppelte Ziel verfolgt wird, eine möglichst balancierte Treatment- und Kontrollgruppe und gleichzeitig eine möglichst große Effizienz in der Ausschöpfung der Vergleichsgruppen zu erreichen. In der einschlägigen Literatur wird empfohlen, mehrere Matchingverfahren zu vergleichen, um zu einem optimierten Ergebnis zu gelangen (Morgan & Winship 2007). Der zweite Schritt schließt mit einer Erfolgskontrolle des Matchingprozesses durch einen Vergleich der Unterschiede auf den Zuweisungsvariablen ab.

Bei erfolgreichem *Matching* werden erst im dritten Schritt Kontroll- und Treatmentgruppe hinsichtlich der Kriteriumsvariablen verglichen. Kontrolle des *Selection Bias* und die Prüfung der theoretisch interessierenden Effekte sind analog zu experimentellen Designs zwei *unabhängige* Schritte der Datenanalyse (Rubin 2001).

Im Folgenden sollen drei Varianten des sogenannten *Caliper-Matching* vorgestellt werden. *Caliper-Matching* ist eine Form des *Propensity Score Matching*, bei dem von vornherein nur diejenigen Fälle berücksichtigt werden, die sich in ihren *Propensity Scores* hinreichend ähnlich sind und innerhalb einer festgelegten Toleranzgrenze liegen. Eine solche Vorgabe erhöht die Genauigkeit des *Matching*. Wir haben die *Caliper*-Breite auf eine Abweichung von maximal $\pm 0,10$ SD vom jeweiligen *Propensity Score* einer Treatmentperson festgelegt. Dies gilt als relativ strenge Grenze (Rosenbaum 2002). Alle drei *Caliper-Matching*-Varianten wurden als Prozeduren mit Zurücklegen spezifiziert, sodass derselbe Fall aus der Grundschülergruppe mehreren Personen der Treatmentgruppe zugeordnet werden kann. Im Vergleich zu Matchingprozeduren ohne Zurücklegen werden dadurch die jeweiligen Vergleichsgruppen besser ausgeschöpft (Rosenbaum 2002).

Verglichen werden: eine Eins zu Eins-Zuordnung, bei der jedem Treatmentfall lediglich ein Kontrollgruppenfall zugeordnet wird (*1:1 nearest-neighbor matching with caliper*) eine Mehrfachzuordnung, bei der jedem Treatmentfall bis zu zehn Fälle der Kontrollgruppe gewichtet zugewiesen werden können, wenn deren *Propensity Scores* innerhalb des vorab definierten *Caliper* liegen (*multiple nearest-neighbor matching with caliper*) und ein Verfahren, bei dem für die im *Caliper*-Bereich liegenden Fälle der Grundschüler zusätzlich Mahalanobis-Distanzen für zentrale Variablen berechnet werden, anhand derer zusätzlich über die Aufnahme in die gematchte Kontrollgruppe – bis zu zehn Grundschülern pro Treatmentperson – entschieden wird (*10:1 multiple matching with caliper and Mahalanobis metric*). Als zentrale Hintergrundvariablen, für die auf Basis der Mahalanobis-Distanzen gematcht wurde, wurden die Leseverständnis- und Mathematikleistung zum Ende der 4. bzw. Anfang der 5. Klasse, der sozioökonomische und soziokulturelle familiäre Hintergrund einbezogen (HISEI, Bücherbesitz und Bildungsgüter). Als zusätzliche Restriktion wurde das Geschlecht der Personen exakt gematcht. Diese zusätzlichen Maßnahmen werden in der Literatur diskutiert, um einen Nachteil des multiplen *Matching*, nämlich die häufigere Zuordnung auch weniger passender Fälle, auszugleichen (Caliendo & Kopeinig 2008; Rubin 1980).

Die Schätzung der *Propensity Scores* und das *Matching* wurden mit dem Programm-paket MatchIt 2.3-1 (Ho et al. 2008) unter Nutzung der fünf imputierten Datensätze durchgeführt.

6 Ergebnisse

6.1 Verteilung der Schülerschaft auf unterschiedliche Schulformen

Der Frühübergang an ein grundständiges Gymnasium erfolgt in Berlin mit dem Wechsel in die 5. Klasse. Am Ende der 6. Jahrgangsstufe folgt die Verteilung auf die übrigen Schulformen der Sekundarstufe I. Im Rahmen der ELEMENT-Studie wurden am Ende der 6. Klasse auch die Übergangsentscheidungen der Eltern für die besuchte Schulform ab der 7. Jahrgangsstufe erfasst. Damit eröffnet sich die Möglichkeit, die Schülerschaft der unterschiedlichen Sekundarschulformen nicht nur anhand ihrer stabilen sozialen und ethnischen Herkunftsmerkmale, sondern auch anhand ihrer Leistungs- und Motivationsprofile zwei Jahre vor dem Wechsel zu beschreiben und mit den Frühabgängern, die an grundständige Gymnasien gehen, zu vergleichen.

In Abb. 1 lassen sich nach einer multivariaten Varianzanalyse vier Merkmalsprofile unterscheiden.² Auffällig sind an dieser Profilbildung zwei Dinge: die institutionelle Trennung einer jeweils kleinen Elite- und Problemgruppe von unter zehn Prozent der Alterskohorte an den Rändern der Leistungs- und Sozialschichtverteilung in grundständigen Gymnasien und Hauptschulen und die klare, in allen Merkmalen signifikante Trennung

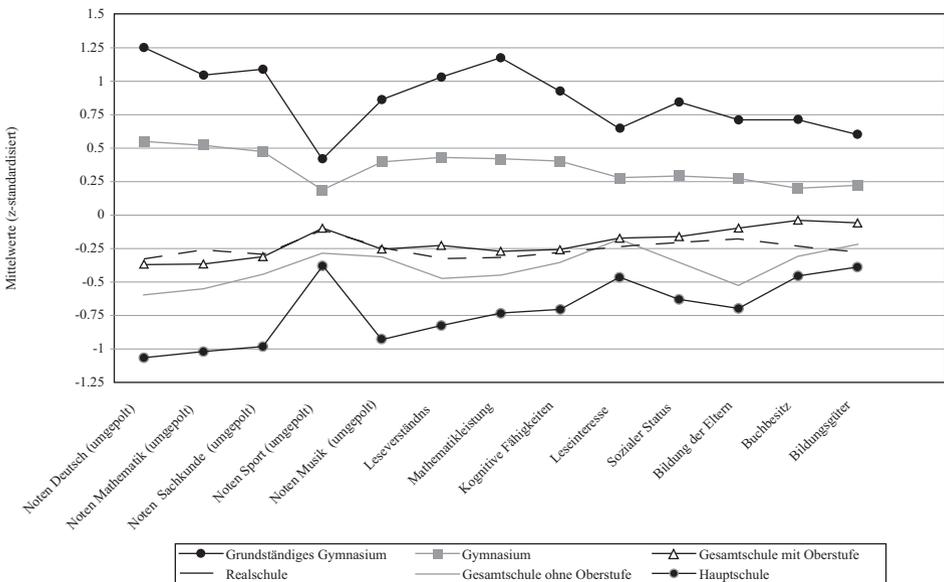


Abb. 1: Deskriptive Kennwerte zum ersten Messzeitpunkt der Schülerschaft unterschiedlicher Schulformen (Schulwahl am Ende der 6. Jahrgangsstufe; in der Grundschulstichprobe z-standardisierte Werte)

der regulären Gymnasiasten von der Schülerschaft sowohl der grundständigen Gymnasien als auch der übrigen Schulformen. In den Leistungsmerkmalen – also den Noten und Testergebnissen – betragen die Differenzen zwischen Frühübergängern und Schülerinnen und Schülern, die nach der 6. Klasse in ein Gymnasium wechseln, am Ende der 4. Jahrgangsstufe 0,50 bis 0,75 Standardabweichungen; in den sozialen Kennwerten liegen sie bei rund 0,50 Standardabweichungen. Insgesamt zeigt die Abb. 1 das Bild eines extrem stratifizierten Schulwesens, bei dem die Verteilungsergebnisse bereits am Ende der 4. Jahrgangsstufe weitgehend vorgezeichnet sind.

6.2 Vorhersage des Frühübergangs in ein grundständiges Gymnasialprogramm

Um die Auswirkungen der selektiven Einwahl in ein grundständiges Gymnasialprogramm von den Wirkungen des Programms trennen zu können, ist eine theoretisch und empirisch zutreffende Spezifikation des Zuweisungsprozesses notwendig, die expliziter Bestandteil der Effektschätzung wird.

Die Modellprüfung des von uns spezifizierten Übergangmodells erfolgte mit binären logistischen Regressionsanalysen. Abhängige Variable ist der dummy-codierte Besuch des grundständigen Gymnasiums (0=Grundschule, 1=grundständiges Gymnasium). Es wurden vier hierarchisch geschachtelte Modelle spezifiziert, deren Anpassungsgüte verglichen werden kann. Die Ergebnisse der Regressionsanalysen sind in Tab. 2 wiedergegeben. Für den Vergleich der Modellanpassung werden der Determinationskoeffizient (R^2)³, Akaikes Informationskoeffizient (AIC) und der Bayessche Informationsindex (BIC), der zusätzlich die Stichprobengröße berücksichtigt, verwendet.

Für die erste Modellschätzung (Tab. 2, Modell 1) wurden ausschließlich Leistungs- und Motivationsvariablen berücksichtigt. Ein Blick auf Tabelle 2 zeigt, dass Noten, Testleistungen und motivationale Orientierungen jeweils spezifische Vorhersagebeiträge leisten, auch wenn sie gemeinsam in die Regressionsgleichung eingehen. Unter den Noten ist die Deutschensur und unter den kognitiven Maßen die Mathematikleistung von besonderer Bedeutung. Der Erklärungsbeitrag der Leistungs- und Motivationskomponenten ist mit einem R^2 (latent) von 0.68 substantiell.

Im zweiten Analyseschritt wurde dieses Modell um Merkmale der sozialen Herkunft und Indikatoren für Bildungsnähe ergänzt (Tab. 2, Modell 2). Die Ergebnisse dieser Modellrechnung zeigen, welche Bedeutung dem Familienhintergrund über die Leistungsmerkmale des Kindes hinaus für die Übergangentscheidung an ein grundständiges Gymnasium zukommt. Mit der Berücksichtigung der Familienmerkmale verbessert sich die Modellanpassung in allen drei Kriterien (R^2 , AIC und BIC). Deutlich wird aber auch, dass die Übergangentscheidung primär aufgrund von Leistungs- und Motivationsmerkmalen getroffen wird. Die Ergebnisse bestätigen ferner den bereits in anderen Untersuchungen berichteten Befund, dass bei Kontrolle von Schülerleistung und Sozialschicht der Familie Zuwanderereltern zu anspruchsvolleren Bildungsgängen tendieren (Ditton, Krüskens & Schauenberg 2005). Das scheint in bemerkenswerter Weise auch für das grundständige Gymnasium zu gelten.

Die dritte Modellrechnung (Tab. 2, Modell 3) berücksichtigt zusätzlich das Geschlecht des Kindes. Überraschenderweise zeigt Modell 3 zumindest für die grundständigen Gym-

Tab. 2: (Fortsetzung)

Haupt- oder Real- schule ohne Berufsabschluss (0/1)	-0,58	0,36	0,56	-0,57	0,36	0,56	-0,44	0,36	0,65
Hauptschule und Lehre (0/1)	-0,59	0,27	0,55	-0,58	0,27	0,56	-0,62	0,30	0,54
Realschule und Lehre (0/1)	-0,60	0,17	0,55	-0,58	0,17	0,56	-0,57	0,18	0,57
Realschule und Fachschule (0/1)	-0,29	0,17	0,75	-0,28	0,17	0,75	-0,32	0,17	0,73
(Fach-)Abitur ohne Hochschule (0/1)	-0,26	0,14	0,77	-0,26	0,14	0,77	-0,32	0,16	0,73
Buchbestand (z-Werte)	0,13	0,06	1,14	0,13	0,06	1,14	0,13	0,07	1,14
Bildungsgüter (z-Werte)	0,26	0,07	1,30	0,26	0,07	1,30	0,27	0,07	1,31
Migrationsstatus (0/1)	0,47	0,18	1,60	0,47	0,18	1,61	0,55	0,18	1,73
Geschlecht (0 = männlich; 1 = weiblich)				-0,21	0,11	0,81	-0,27	0,11	0,76
Anzahl grundstän- diger Gymnasien im Bezirk							0,59	0,17	1,81
R ² (latent)	0,68			0,70			0,73		
AIC	3681,18			3532,71			3322,00		
BIC	3746,19			3669,25			3465,05		

Anmerkung: $p < .05$ als fett gedruckte Werte wiedergegeben.

nasien das noch traditionelle Muster, dass Jungen bei gleicher Leistung und gleicher familiärer Herkunft eine höhere Übergangschance als Mädchen aufweisen.

Wird schließlich in Analysemodell 4 (Tab. 2, Modell 4) die Angebotsstruktur im Wohnbezirk – es sind hier die alten Bezirksgrenzen Berlins zugrunde gelegt – berücksichtigt, sieht man, dass auch die Opportunität eine Rolle spielt. Die Erreichbarkeit des grundständigen Gymnasiums scheint die Anwahl eines grundständigen Gymnasiums mitzubestimmen. Das vierte, vollständige Regressionsmodell weist in allen Merkmalen die beste Modellanpassung auf.

6.3 Regressionsanalytische Überprüfung der differenziellen Förderwirkung grundständiger Gymnasialangebote

Das im vorherigen Abschnitt vorgestellte und an die Daten optimal angepasste Modell zur Vorhersage des Übergangs an grundständige Gymnasien wurde im nächsten Schritt als Kovariatenvektor für die regressionsanalytische Schätzung potenzieller Förderwirkungen der grundständigen Gymnasien auf die Lese- und Mathematikleistung am Ende der 6. Jahrgangsstufe verwendet. Der Besuch eines grundständigen Gymnasiums wurde dummy-codiert. Referenz- und Kontrollgruppe sind die Schülerinnen und Schüler, die bis zum Ende der 6. Jahrgangsstufe an einer Grundschule verbleiben.

Tabelle 3 weist die Ergebnisse der Regressionsanalyse für die Lese- und Mathematikleistungen am Ende der 6. Jahrgangsstufe aus. Ein erster Blick auf die Ergebnisse für das Leseverständnis zeigt, dass die drei Kernelemente – Leistung, Motivation und familiärer Hintergrund – leistungsrelevante Kovariaten darstellen. Die wichtigsten Kontrollvariablen im Leistungsbereich sind Lesefähigkeit, Mathematikleistung und kognitive Grundfähigkeiten am Ende der 4. Jahrgangsstufe sowie die Halbjahresnoten in Deutsch, Sachkunde und Musik aus der 4. Klasse. Als motivationale Variable ist das am Ende der 4. Jahrgangsstufe erhobene Leseinteresse für das Leseverständnis am Ende der 6. Jahrgangsstufe auch bei Kontrolle aller anderen Kovariaten prädiktiv. Die wichtigsten sozialen und kulturellen Hintergrundmerkmale sind der Bildungsabschluss der Eltern und die im Haushalt verfügbaren kulturellen Ressourcen als Indikator für das kulturelle Milieu der Familie. Ein wichtiges Ergebnis zeigt sich ferner für Schülerinnen und Schüler mit Migrationshintergrund. Bei Kontrolle von Leistungs-, Motivations- und sozialen Herkunftsmerkmalen scheint sich die Leistungsschere zwischen Schülerinnen und Schülern mit und ohne Migrationsgeschichte in den Schuljahren 5 und 6 nicht weiter zu öffnen. Der Befund zur differenziellen Förderung der grundständigen Gymnasien mag überraschen: Im Vergleich zur Grundschule ist kein differenzieller Fördereffekt des zweijährigen Besuchs eines grundständigen Gymnasiums auf die Lesekompetenz nachweisbar.

Die Ergebnisse der analogen Regressionsanalyse für Mathematik lassen sich folgendermaßen zusammenfassen: Im Leistungsbereich haben jetzt die am Ende der 4. Jahrgangsstufe erfasste Mathematikleistung, das Leseverständnis und die kognitiven Grundfähigkeiten sowie die Mathematik-, Sachkunde- und Musiknoten die stärkste Vorhersagekraft für die am Ende der 6. Jahrgangsstufe erreichte Mathematikleistung. Das Leseinteresse spielt für die Mathematik als Motivationsvariable keine Rolle. Das Vorhersagemuster der Merkmale der sozialen, kulturellen und ethnischen Herkunft ist im Großen und Ganzen mit den für das Leseverständnis berichteten Befunden vergleichbar.

Tab. 3: Ergebnisse der hierarchisch linearen Regression der Lese- und Mathematikleistung am Ende der 6. Jahrgangsstufe auf Kovariate des ersten Messzeitpunktes (Ebene 1, z-Werte in der Grundschulstichprobe standardisiert) und Schulform (Ebene 2: Grundständiges Gymnasium versus Grundschule als Referenz)

Prädiktoren/Zuweisungsmerkmale	Leseverständnis				Mathematikleistung					
	B	S.E.	Beta ¹	t	p	B	S.E.	Beta ¹	t	p
<i>Intercept/Mittel der Abweichungen</i>	110,00	0,49	8,16	226,07	0,000	117,03	0,63	6,81	181,28	0,000
Schülerebene										
Note Deutsch (umgepolt, z-Wert)	1,02	0,22	0,08	4,67	0,000	0,07	0,24	0,00	0,29	0,772
Note Mathematik (umgepolt, z-Wert)	0,37	0,21	0,03	1,76	0,079	3,74	0,30	0,23	12,54	0,000
Note Sachkunde (umgepolt, z-Wert)	1,13	0,24	0,09	4,61	0,000	0,49	0,27	0,03	1,83	0,067
Note Musik (umgepolt, z-Wert)	0,53	0,21	0,04	2,54	0,011	0,54	0,23	0,03	2,34	0,019
Note Sport (umgepolt, z-Wert)	-0,25	0,14	-0,02	-1,73	0,084	-0,02	0,17	-0,00	-0,14	0,893
Lesefähigkeit (z-Wert)	5,04	0,21	0,38	23,90	0,000	1,62	0,22	0,10	7,36	0,000
Mathematikleistung (z-Wert)	1,52	0,20	0,11	7,47	0,000	5,67	0,25	0,35	22,50	0,000
Kognitive Grundfähigkeiten (z-Wert)	1,02	0,17	0,08	6,08	0,000	1,99	0,20	0,12	10,22	0,000
Leseliteresse (z-Wert)	0,88	0,19	0,07	4,72	0,000	0,02	0,19	0,00	0,10	0,926
Sozialer Status (HISEI) (z-Wert)	0,27	0,23	0,02	1,16	0,248	0,38	0,22	0,02	1,68	0,093

Tab. 3: (Fortsetzung)

Bildung der Eltern (Ref.: Hochschulabschluss)										
Ohne Abschluss (0/1)	-1,30	0,76	-0,10	-1,72	0,085	-1,81	0,97	-0,11	-1,86	0,063
Haupt- oder Realschule ohne Berufsabschluss (0/1)	-0,44	0,96	-0,03	-0,46	0,645	-0,17	0,78	-0,01	-0,22	0,824
Hauptschule und Lehre (0/1)	-0,52	0,61	-0,04	-0,85	0,397	-1,27	0,74	-0,08	-1,74	0,087
Realschule und Lehre (0/1)	-1,19	0,44	-0,09	-2,70	0,007	-1,85	0,62	-0,12	-2,98	0,003
Realschule und Fachschule (0/1)	-1,14	0,48	-0,09	-2,38	0,017	-0,94	0,49	-0,06	-1,94	0,053
(Fach-)Abitur ohne Hoch- schule (0/1)	-0,56	0,42	-0,04	-1,33	0,185	-0,78	0,54	-0,05	-1,46	0,145
Buchbestand (z-Werte)	0,46	0,15	0,03	3,00	0,003	0,41	0,20	0,03	2,03	0,043
Bildungsgüter (z-Werte)	0,46	0,20	0,04	2,35	0,019	-0,03	0,24	-0,00	-0,12	0,903
Migrationsstatus (0/1)	-0,41	0,33	-0,03	-1,24	0,214	0,45	0,34	0,03	1,33	0,184
Geschlecht (0=männlich; 1=weiblich)	0,14	0,32	0,01	0,43	0,666	-1,97	0,33	-0,12	-6,01	0,000
R² innerhalb		0,64		51,97	0,000		0,69		65,96	0,000
Klassenebene										
Anzahl grundständiger Gym- nasien im Bezirk	0,03	0,15	0,00	0,21	0,837	-0,01	0,26	-0,00	-0,05	0,962
Besuch eines grundständigen Gymnasiums (0/1)	0,02	0,51	0,00	0,03	0,976	2,43	0,73	0,16	3,35	0,001
R² zwischen		0,001		0,19	0,848		0,11		1,90	0,058

ⁱ Nur abhängige Variable standardisiert.

Im Unterschied zur Lesekompetenz wird das Geschlecht jedoch auch bei Kontrolle aller anderen Kovariaten signifikant: Mädchen sind in der 5. und 6. Jahrgangsstufe in Mathematik weniger erfolgreich.

Die Ergebnisse zur Förderwirkung grundständiger Gymnasien fallen für die Mathematikleistungen positiv aus. Frühübergänger an grundständige Gymnasien erreichen am Ende der 6. Jahrgangsstufe im Mittel statistisch signifikant bessere Mathematikleistungen. Diese Effektstärke ist jedoch mit $d=0.16$ SD über zwei Schuljahre gesehen von eher geringer Größe.

Mit den beiden Regressionsanalysen wurde versucht, die durchschnittliche Förderwirkung grundständiger Gymnasialangebote für die Lese- und Mathematikleistung in der gesamten Altersgruppe (ATE) zu schätzen. Das Ergebnis ist für die Lesekompetenz negativ, für die Mathematikleistungen leicht positiv. Hinsichtlich der Belastbarkeit dieser Ergebnisse gilt der Vorbehalt, dass man mit Verletzungen der Voraussetzungen der Regressionsanalyse zu rechnen hat.

6.4 Schätzung differenzieller Förderwirkungen von grundständigen Gymnasialangeboten mit Verfahren des Propensity Score Matching

Abbildung 2a bildet die Verteilungen der *Propensity Scores* in der Grundschule und im grundständigen Gymnasium ab. Für die Schätzung der *Propensity Scores* wurde wiederum der Prädiktorsatz des am besten angepassten Übergangsmodells verwendet.

Der grafische Vergleich der Verteilungen gibt einen ersten Eindruck von ihrer Überlappungsbreite und der Größe der *Area of Common Support* – also des Bereichs, in dem überhaupt vergleichbare Grundschüler und Gymnasiasten zu finden sind. Die *Propensity Scores* sind in beiden Schulformen annähernd normal verteilt, wobei sich die Verteilungen in weiten Teilen überlappen. Dies bedeutet, dass praktisch die gesamte Treatmentgruppe, das heißt die Gruppe der Gymnasiasten, in die *Area of Common Support* fällt. Dies ist eine sehr gute Voraussetzung für ein erfolgreiches *Matching* und die Beantwortung unserer Fragestellung.

Ergebnisse der Matchingverfahren: Wir hatten im Methodenabschnitt 5.5 die Auswahl von drei alternativen Matchingprozeduren begründet und diese einzeln vorgestellt. Alle drei Matchingverfahren führen zu akzeptablen Ergebnissen. Die Ausschöpfungsquote der Treatmentgruppe ist sehr hoch. Für 1732 von 1758 Fällen wird ein Matchingpartner in der Grundschulstichprobe gefunden. Die wenigen Ausfälle liegen im Höchstleistungsbereich. Danach können Wirkungsaussagen für praktisch die gesamte Gruppe der grundständigen Gymnasiasten getroffen werden. Bei einer 1:1-Zuordnung umfasst die gematchte Kontrollgruppe 628 Grundschüler; lässt man Mehrfachzuordnungen zu, steigt die Probandenzahl in der Kontrollgruppe auf 1697. Vergleicht man die Matchingergebnisse anhand der einzelnen Prädiktoren des Übergangsmodells, lassen sich nach dem *Matching* bei *keinem* der Verfahren signifikante Mittelwertunterschiede zwischen Treatment- und Kontrollgruppe nachweisen. Eine Inspektion der standardisierten Mittelwertdifferenzen zeigt jedoch, dass weiterhin kleine Unterschiede bestehen bleiben. Die Unterschiede sind bei dem Matchingverfahren, das *Propensity Score*, Mahalanobis-Distanz- und exaktes *Matching* für Geschlecht verbindet, am geringsten. Diese Ergebnisse sind grafisch in Abb. 2b wiedergegeben.⁴ Die gemittelten standardisierten Mittelwertunterschiede betragen Null,

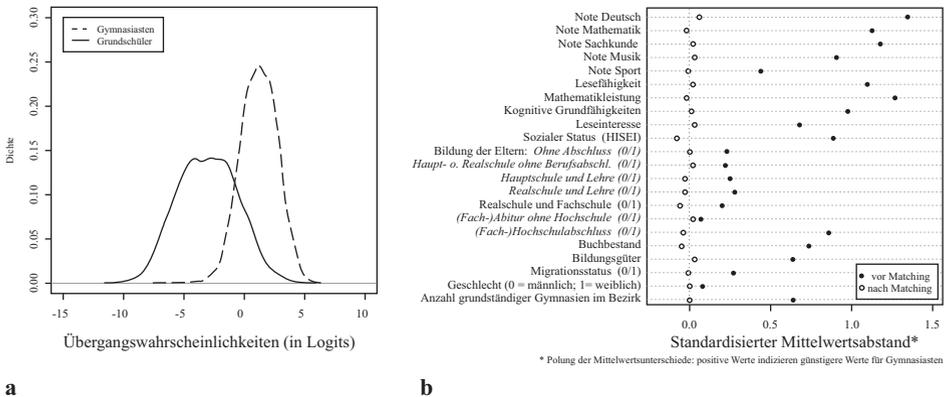


Abb. 2: a) Verteilung der Propensity Scores (Logits) in Grundschulen und Gymnasien b) Verteilung der Hintergrundvariablen (standardisierter Mittelwertsabstand) zwischen Grundschulern und Gymnasiasten vor und nach dem Matching (positive Werte indizieren günstigere Werte für die gymnasiale Schülerschaft)

wobei die Spannweite der Differenzen zwischen $-.08$ und $.06$ SD liegt. Die Unterschiede deuten unsystematisch in beide Richtungen. Sie können als vernachlässigbar erachtet werden (Rosenbaum 2002).

Prüfung der differenziellen Förderwirkung grundständiger Gymnasialangebote: Der abschließende Schritt ist der Leistungsvergleich zwischen der Kontroll- und Treatmentgruppe. Tabelle 4 weist die Ergebnisse für das Leseverständnis und die Mathematikkompetenz am Ende der 6. Jahrgangsstufe aus. Zusätzlich werden auch die Ergebnisse nach einjährigem Besuch des grundständigen Gymnasiums am Ende der 5. Jahrgangsstufe berichtet.

Die Ergebnisse der Mittelwertvergleiche sind eindeutig: In keinem Leistungsbereich sind Förderwirkungen des grundständigen Gymnasiums nachweisbar. Dies gilt sowohl für den ein- wie den zweijährigen Besuch dieser Einrichtungen. Für das Leseverständ-

Tab. 4: Mittelwertunterschiede in der Mathematikleistung und im Leseverständnis zwischen Grundschulern und Gymnasiasten nach *Matching* (ohne Korrektur für multiples Testen)

Leistungsbereiche	GS (N=1.697)		GY (N=1.732)		t	Std. M Diff**
	M	SE	M	SE		
Lesefähigkeit (Ende 5)	119,03	0,55	119,29	0,42	0,37	0,03
Lesefähigkeit (Ende 6)	123,16	0,73	123,25	0,48	0,10	0,01
Mathematikleistung (Ende 5)	123,43	0,87	124,26	0,58	0,79	0,06
Mathematikleistung (Ende 6)	133,56	1,07	133,89	0,67	0,26	0,02

Anmerkungen: GS: Grundschule; GY: Gymnasium.

* statistisch signifikant auf $\alpha = 5\%$ (ohne Korrektur für multiples Testen).

** Std. M Diff.: standardisierte Mittelwertsdifferenz.

nis ist dieser Befund konsistent mit den Ergebnissen der Regressionsanalyse. Für den Bereich der Mathematikkompetenz kann jedoch der regressionsanalytische Befund einer differenziellen Förderwirkung nicht repliziert werden.

7 Zusammenfassung und Diskussion

Sieben bis acht Prozent der Schülerinnen und Schüler eines Jahrgangs verlassen in Berlin die regulär sechsjährige Grundschule nach der 4. Klasse, um auf ein grundständiges Gymnasium zu wechseln. Aus dem Zusammenspiel von Schule und Elternhaus beim Frühübergang resultiert eine kognitiv, leistungsmäßig, motivational und hinsichtlich häuslicher Ressourcen hoch ausgelesene Schülerschaft an grundständigen Gymnasien, die für die Lehrkräfte eine intellektuelle Herausforderung, aber auch ein Erfolgsversprechen für den Unterricht darstellt.

Will man in Beobachtungsstudien für eine derart ausgelesene Schülerschaft belastbare Aussagen über die differenzielle Wirkung von Bildungsprogrammen treffen, ist die valide Rekonstruktion des Wahl- und Zuweisungsprozesses für eine erwartungstreue und konsistente Schätzung von Programmeffekten von entscheidender Bedeutung. In unseren Analysen konnte ein theoretisch begründetes Fünf-Komponenten-Modell zur Vorhersage des Frühübergangs erfolgreich angepasst werden.

Für die Überprüfung der Förderungswirkung der grundständigen Gymnasien wurden zwei unterschiedliche methodische Verfahren angewandt – Regressionsanalyse und Varianten des *Propensity Score Matching*. Im Vergleich zum *Propensity Score Matching* ist die Regressionsanalyse das voraussetzungsreichere Verfahren. In unserem Fall der Untersuchung des Frühübergangs dürften implizite Annahmen der Regressionsanalyse, insbesondere die der homogenen Wirkung der Programme über alle Ausprägungen der Variablen des Übergangsmodells hinweg und die *Stable Unit Treatment Value Assumption* (SUTVA), nicht erfüllt sein. Dagegen erwies sich der ELEMENT-Datensatz für Verfahren des *Propensity Score Matching* als besonders geeignet, da sich die Verteilungen der *Propensity Scores* weit überlappen, sodass der Großteil der Treatmentgruppe (Schülerinnen und Schüler an Gymnasien) in den Bereich des *Common Support* fällt, für den belastbare vergleichende Aussagen möglich sind. Unter den Varianten des *Propensity Score Matching* erwies sich ein relativ restriktives *Caliper-Matching* unter Verwendung von Mahalanobisdistanz- und exaktem *Matching* als Verfahren der Wahl. Für die gesamte Gruppe der grundständigen Gymnasiasten konnte eine sehr gut balancierte Kontrollgruppe gefunden werden.

Beide Analysestrategien – Regressionsanalyse und *Propensity Score Matching* – erbrachten für die Lesekompetenz konsistente Ergebnisse. In dieser weniger unterrichts-abhängigen Domäne scheint die Entwicklung von grundständigen Gymnasiasten und vergleichbaren Schülern der Grundschule völlig parallel zu verlaufen. In der unterrichts-abhängigen Domäne Mathematik zeigte sich in der Regressionsanalyse ein kleiner positiver Effekt, der jedoch im *Propensity Score Matching* nicht repliziert werden konnte. Da die gematchten Treatment- und Kontrollgruppen optimal balanciert sind, liegt der Schluss nahe, dass die Korrektur der Regressionsanalyse zu liberal ausfiel.

Bewertet man die Befunde insgesamt, so sind sie zunächst ein Kompliment für die Grundschule. Die Entwicklungskurven von Spitzenschülern verlaufen in der Grund-

schule und in der Unterstufe des grundständigen Gymnasiums parallel, und zwar nicht nur im Lesen, sondern – verlässt man sich auf das robustere Verfahren des PSM – auch in der unterrichtsabhängigen Domäne Mathematik. Für die grundständigen Gymnasien sind die Befunde ein Grund zur Nachdenklichkeit. Generell ist fraglich, ob die Gymnasien die Förderung der Lesekompetenz als akademische Aufgabe aller Fächer bislang überhaupt entdeckt haben. Die Befunde sprechen eher für einen Entwicklungsprozess der Lesekompetenz, der von den Vorleistungen der Schüler und des Elternhauses lebt und von dem die grundständigen Gymnasien profitieren, ohne ihn selbst aktiv zu fördern. Anders ist die Sachlage in Mathematik. Eine spezifische Förderung in diesem Fach ist kaum zu erwarten, wenn der curriculare Programmschwerpunkt auf Musik, Sport oder einer alten oder modernen Fremdsprache liegt. Kompositionseffekte, die ja über die Qualität des Unterrichts vermittelt werden müssen, stellen sich nicht von selbst ein. Hier wird auch deutlich, dass die Befunde zur differenziellen Förderungswirkung unterschiedlicher Sekundarschulformen, die primär für eine Niveaudifferenzierung nach Leistung stehen, nicht ohne Weiteres auf die Programmdifferenzierung der grundständigen Gymnasien – jedenfalls was die Unterstufe betrifft – übertragbar sind.

Wägt man aus einer etwas distanzierteren Perspektive Stärken und Schwächen der vorliegenden Studie ab, so wird man folgende Bilanz ziehen können: Stärken der Untersuchung sind die längsschnittliche Anlage, die es ermöglichte, alle Kovariaten vor Beginn des Treatments zu erheben und so eine bei Querschnittsstudien immer mögliche Überkontrolle zu vermeiden, sowie die Spezifikation und Anpassung eines theoretisch begründeten Zuweisungsmodells, das den Frühübergang in substanzieller Weise vorherzusagen erlaubte. Ebenso gehört das gelungene *Matching* mit der breiten *Area of Common Support* und der optimal balancierten Kontrollgruppe zu den Stärken des vorliegenden Beitrags. Aber die Grenzen der Studie liegen auch auf der Hand: Die für die Untersuchung ausgewählten Domänen des Leseverständnisses und der Mathematikleistung gehören zwar zu den Basiskompetenzen, treffen aber nicht den Kern der Bildungsprogramme der grundständigen Gymnasien. Die vorliegenden Befunde sagen also nichts über die Erfüllung des spezifischen Bildungsauftrages dieser Schulen aus. Ferner war es im Rahmen des vorliegenden Beitrags nur möglich, den Haupteffekt des Frühübergangs auf ein grundständiges Gymnasium zu überprüfen. Die differenzierte Analyse der unterschiedlichen Programmvarianten des grundständigen Gymnasiums steht noch aus.

Anmerkungen

- 1 Darüber hinaus wurden in den letzten Jahren einige Änderungen im Berliner Schulsystem vorgenommen (z. B. verpflichtende Leistungstests für die Aufnahme in Schnellläuferklassen), auf die im Folgenden nicht weiter eingegangen wird, insofern sie in den Zeitraum nach der hier vorgestellten Untersuchung fallen.
- 2 Die multivariaten Varianzanalysen wurden lediglich mit dem ersten *Plausible Value* und ohne Berücksichtigung der hierarchischen Struktur der Daten spezifiziert, da unseres Wissens nach keine Prozeduren vorliegen, die diese beiden Probleme im Rahmen varianzanalytischer Routinen integrieren können. Die Mehrfachtestung wurde hingegen durch eine Bonferroni-Korrektur des Signifikanzniveaus berücksichtigt. Ergebnistabellen zu Mittelwerten sind auf Anfrage bei den Autoren erhältlich.

- 3 Im Unterschied zu *Likelihood* basierten R^2 -Statistiken logistischer Modelle (z. B. Nagelkerkes Pseudo- R^2) wird die R^2 -Statistik in *Mplus* auf Grundlage aufgeklärter Varianz und Residualvarianz einer der dichotomen Variablen zugrunde gelegten kontinuierlichen Verteilung bestimmt (Muthén & Muthén 1998-2008).
- 4 Die Ergebnistabellen der beiden anderen Matchingverfahren sind bei den Autoren auf Anfrage zu erhalten.

Literatur

- Baumert, J. & Köller, O. (1998). Nationale und internationale Schulleistungsstudien: Was können sie leisten, wo sind ihre Grenzen? *Pädagogik*, 50 (6), 12–18.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. 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. 323–407). Opladen: Leske + Budrich.
- Baumert, J., Cortina, K. S. & Leschinsky, A. (2008). Grundlegende Entwicklungen und Strukturprobleme im allgemeinbildenden Schulwesen. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland: Strukturen und Entwicklungen im Überblick* (S. 53–130). Reinbek bei Hamburg: Rowohlt.
- Baumert, J., Köller, O. & Schnabel, K. U. (2000). Schulformen als differentielle Entwicklungsmilieus – eine ungehörige Fragestellung? Erwiderung auf die Expertise „Zur Messung sozialer Motivation in der BIJU-Studie“ von Georg Lind. In Bildungs- und Förderwerk der Gewerkschaft Erziehung und Wissenschaft im DGB (Hrsg.), *Messung sozialer Motivation: Eine Kontroverse* (S. 28–68). Frankfurt am Main: GEW.
- Baumert, J., Roeder, P. M., Gruehn, S., Heyn, S., Köller, O., Rimmele, R., Schnabel, K. U. & Seipp, B. (1996). Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). In K.-P. Treumann, G. Neubauer, R. Möller & J. Abel (Hrsg.), *Methoden und Anwendungen empirischer pädagogischer Forschung* (S. 170–180). Münster: Waxmann.
- Becker, M. (2008). *Kognitive Leistungsentwicklung in differenziellen Lernumwelten: Effekte des gegliederten Sekundarschulsystems in Deutschland*. Bisher unveröffentlichte Diss., Berlin: Freie Universität.
- Becker, M., Lüdtke, O., Trautwein, U. & Baumert, J. (2006). Leistungszuwachs in Mathematik: Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem. *Zeitschrift für Pädagogische Psychologie*, 20, 233–242.
- Becker, R. & Schubert, F. (2006). Soziale Ungleichheit von Lesekompetenzen: Eine Matching-Analyse im Längsschnitt mit Querschnittsdaten von PIRLS 2001 und PISA 2000. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 58, 253–284.
- Bonsen, M., Bos, W. & Gröhlich, C. (2007). *Differentielle Lern- und Entwicklungsmilieus in der Beobachtungsstufe?* Vortrag auf der 70. Tagung der Arbeitsgruppe empirische-pädagogische Forschung (AepF). Lüneburg, 10.–12. September 2007.
- Bos, W., Bonsen, M., Gröhlich, C., Jelden, D. & Rau, A. (2006). Erster Bericht zu den Ergebnissen der Studie „Kompetenzen und Einstellungen von Schülerinnen und Schülern – Jahrgangsstufe 7“ (KESS 7). Forschungsbericht für die Behörde für Bildung und Sport der Stadt Hamburg. Zugriff am 08.02.2008 unter <http://www.hamburger-bildungsserver.de/schulentwicklung/qualitaet/kess/KESS7.pdf>.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R. & Walther, G. (Hrsg.) (2003). *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.

- Caliendo, M. & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22, 31–72.
- Ditton, H., Krüsken, J. & Schauenberg, M. (2005). Bildungsungleichheit – der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft*, 8, 285–304.
- 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. G. & Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in cross-national comparison: A European working book for demographic and socio-economic variables* (S. 159–193). New York: Kluwer Academic Press.
- Graham, J. W., Cumsille, P. E. & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Bd. 2, S. 87–114). New York: John Wiley & Sons.
- Heckman, J. J. (2000). Causal parameters and policy analysis in economics: A twentieth century perspective. *The Quarterly Journal of Economics*, 115, 45–97.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4.–12. Klassen, Revision (KFT 4 12+R)*. Göttingen: Hogrefe.
- Herrlitz, H.-G., Hopf, W., Titze, H. & Cloer, E. (2005). *Deutsche Schulgeschichte von 1800 bis zur Gegenwart: Eine Einführung*. Weinheim: Juventa.
- Ho, D. E., Imai, K., King, G. & Stuart, E. A. (2008). *The MatchIt Package* (Version 2.3–1) [Computer Software]. Zugriff am 30.09.2008 unter: <http://gking.harvard.edu/matchit/>.
- King, G. & Zeng, L. (2006). The dangers of extreme counterfactuals. *Political Analysis*, 14, 131–159.
- Klieme, E. (2006). *Zusammenfassung zentraler Ergebnisse der DESI-Studie*. Zugriff am 09.04.2008 unter: www.dipf.de/desi/DESI_Ausgewahlte_Ergebnisse.pdf.
- Köller, O. & Baumert, J. (2001). Leistungsgruppierungen in der Sekundarstufe I: Ihre Konsequenzen für die Mathematikleistung und das mathematische Selbstkonzept der Begabung. *Zeitschrift für Pädagogische Psychologie*, 15, 99–110.
- Köller, O. & Baumert, J. (2002). Entwicklung schulischer Leistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie: Ein Lehrbuch* (S. 756–786). Weinheim: Psychologie Verlags Union.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.
- Lehmann, R. H. & Lenkeit, J. (2008). *Element. Erhebung zum Lese- und Mathematikverständnis. Entwicklung in den Jahrgangsstufen 4 bis 6 in Berlin: Abschlussbericht über die Untersuchungen 2003, 2004 und 2005 an Berliner Grundschulen und grundständigen Gymnasien*. Berlin: Humboldt Universität zu Berlin.
- Lehmann, R. H. & Nikolova, R. (2005a). *Element. Erhebung zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin: Bericht über die Untersuchung 2003 an Berliner Grundschulen und grundständigen Gymnasien*. Berlin: Humboldt Universität zu Berlin.
- Lehmann, R. H. & Nikolova, R. (2005b). *Lese- und Mathematikverständnis von Grundschülerinnen und Grundschülern am Ende der Klassenstufe 5: Befunde aus dem zweiten Erhebungszeitpunkt der ELEMENT-Untersuchung Schuljahr 2003/2004*. Berlin: Senatsverwaltung für Bildung, Jugend und Sport.
- Lehmann, R. H. & Peek, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lehmann, R. H., Gänsfuß, R. & Peek, R. (1999). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen – Klassenstufe 7*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.

- Lehmann, R. H., Peek, R., Gänsfuß, R. & Husfeldt, V. (2001). *Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 9 – Ergebnisse einer längsschnittlichen Untersuchung in Hamburg*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lehmann, R. H., Peek, R., Gänsfuß, R., Lutkat, S., Mücke, S. & Barth, I. (2000). *QuaSUM. Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik: Ergebnisse einer repräsentativen Untersuchung im Land Brandenburg*. Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau*, 58, 103–117.
- Morgan, S. L. & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles in social research*. Cambridge, UK: Cambridge University Press.
- Muthén, B. O. & Muthén, L. K. (1998–2008). *Mplus (Version 5.1)* [Computer software]. Los Angeles, CA: Muthén & Muthén.
- Neumann, A. & Lehmann, R. H. (2008). Schreiben Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (S. 83–103). Weinheim: Beltz.
- Pekrun, R., vom Hofe, R., Blum, W., Götz, T., Wartha, S., Frenzel, A. & Jullien, S. (2006). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA): Entwicklungsverläufe, Schülervoraussetzungen und Kontextbedingungen von Mathematikleistungen in der Sekundarstufe I. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 21–53). Münster: Waxmann.
- Rösner, E. (2005). *Rahmenkonzept einer allgemeinen Sekundarschule: Schulstrukturelle Konsequenzen aus demografischen Entwicklungen und gesellschaftlichen Erfordernissen in Nordrhein-Westfalen. Modell einer reformierten Sekundarstufenschule in Anlehnung an die „Gemeinschaftsschule Schleswig-Holstein“*. Dortmund: VBE-Verlag NRW.
- Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer.
- Rubin, D. B. (1980). Bias reduction using Mahalanobis-metric matching. *Biometrics*, 36, 293–298.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2, 169–188.
- Schafer, J. L. (1999). *NORM for Windows 95/98/NT: Multiple imputation of incomplete data under a normal model (Version 3.02)* [Software]. University Park, PA: Penn State Department of Statistics.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H. & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Washington, DC: American Educational Research Council.
- Schneider, W. & Stefanek, J. (2004). Entwicklungsveränderungen allgemeiner kognitiver Fähigkeiten und schulbezogener Fertigkeiten im Kindes- und Jugendalter. Evidenz für einen Schereneffekt? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36, 147–159.
- Schneider, W., Knopf, M. & Stefanek, J. (2002). The development of verbal memory in childhood and adolescence: Findings from the Munich Longitudinal Study. *Journal of Educational Psychology*, 94, 751–761.
- Stryck, T. & Lehmann, R. (2003). Erhebungen zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin (Element). Projektbeschreibung. Zugriff am 26.03.2009 unter: http://www.berlin.de/imperia/md/content/sen-bildung/schulqualitaet/schulleistungsuntersuchungen/element_projektbeschreibung.pdf.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1998). *ACER ConQuest: Generalized item response modeling software manual*. Melbourne: Australian Council for Educational Research.