

Schereneffekte auf der Sekundarstufe I? Zum Einfluss des Schultyps auf den Leistungszuwachs in Deutsch und Mathematik

Domenico Angelone, Universität Bern

Die vorliegende Längsschnittstudie (N = 950) untersucht mithilfe von Propensity Score Matching, ob vergleichbare Schülerinnen und Schüler im Verlauf der Sekundarstufe I in Abhängigkeit des besuchten Schultyps unterschiedliche Leistungszuwächse in Deutsch und Mathematik erzielen. Die Ergebnisse zeigen, dass sich die Leistungsdisparitäten im Verlauf der Sekundarstufe I als Folge schultypspezifischer Leistungszuwächse vergrössern. Neben den bekannten Herkunftseffekten auf die Leistungsentwicklung im frühen Schulverlauf kommen demnach auf der Sekundarstufe I zusätzliche Effekte der Schulstruktur hinzu.

Einleitung

Die institutionelle Ausgestaltung eines Schulsystems ist für die schulische Laufbahn, für Selektionsprozesse und damit auch für die Frage der Chancengleichheit von zentraler Bedeutung. Ein Hauptmerkmal des Schweizerischen Schulsystems auf der Sekundarstufe I ist die Aufteilung der Schülerinnen und Schüler in verschiedene Schultypen und Anspruchsniveaus. Die Aufteilung erfolgt in der Regel nach sechs Jahren Primarschule und basiert in den meisten Kantonen auf vorgängigen Leistungsbeurteilungen in Form von Schulnoten und einer Übergangsempfehlung der Lehrperson (EDK/IDES, 2015). Da schulische Leistungen stark mit der sozialen Herkunft verknüpft sind, geht die institutionelle Leistungsdifferenzierung immer auch mit sozialen Disparitäten in der Bildungsbeteiligung einher. Während die Gymnasien vorwiegend von leistungsstarken Schülerinnen und Schülern mit privilegierter sozialer Herkunft besucht werden, konzentrieren sich in den Schultypen mit Grundansprüchen leistungsschwache Schülerinnen und Schüler mit benachteiligter sozialer Herkunft (Angelone, Ramseier, & Moser, 2010; Baumert, Stanat, & Watermann, 2006; Dumont, Neumann, Maaz, & Trautwein, 2013).

Grundgedanke der Leistungsdifferenzierung auf der Sekundarstufe I ist, dass der Unterricht in leistungshomogenen Lerngruppen besser auf individuelle Lernvoraussetzungen abgestimmt werden kann und dadurch eine optimale Förderung aller Schülerinnen und Schüler möglich ist (Baumert et al., 2006). Vorliegende Befunde deuten jedoch darauf hin, dass die unterschiedlichen Schultypen der Sekundarstufe I differenzielle Lernumgebungen darstellen, die den Schülerinnen und Schülern unterschiedliche Chancen der Kompetenzentwicklung bieten. Als besonders lernförderliche Umgebungen erweisen sich die Schultypen mit höheren Anforderungen wie das Gymnasium, als weniger lernförderliche Umgebungen die Schultypen mit geringen Anforderungen wie die Realschulen (Neumann, Becker, & Maaz, 2014). Die Auswirkungen der gewollten Leistungsdifferenzierung scheinen somit im Widerspruch zu den gesetzten Zielen – einer bestmöglichen Förderung aller Schülerinnen und Schüler – zu stehen (Baumert et al., 2006). Vielmehr scheint die Leistungsdifferenzierung soziale Bildungsdisparitäten zu verstärken. Für die gegliederte Sekundarschule der Schweiz liegen dazu kaum längsschnittliche Befunde vor.

Die vorliegende Studie schliesst diese Lücke und untersucht am Beispiel der gegliederten Sekundarstufe I des Kantons Zürich, ob Schülerinnen und Schüler mit vergleichbaren Startvoraussetzungen im Verlauf der Sekundarstufe I in Abhängigkeit des besuchten Schultyps unterschiedliche Lernzuwächse in Deutsch und Mathematik erzielen. Vergrössern sich die Leistungsdisparitäten zwischen den unterschiedlich anspruchsvollen Bildungsgängen der Sekundarstufe I, bleiben sie bestehen oder verringern sie sich?

Schultypspezifische Lern- und Entwicklungsmilieus

Schultypspezifische Leistungsentwicklungen werden in der Literatur primär auf unterschiedliche *Lern- und Entwicklungsmilieus* zurückgeführt. Damit ist gemeint, dass Schülerinnen und Schüler unabhängig von individuellen Eingangs- und Lernvoraussetzungen «je nach besuchter Schulform differenzielle Entwicklungschancen erhalten, die schulmilieubedingt sind und sowohl durch den Verteilungsprozess als auch durch die institutionellen Arbeits- und Lernbedingungen und die schulformspezifischen pädagogisch-didaktischen Traditionen erzeugt werden» (Baumert et al., 2006, S. 99).

Mit Blick auf die lernmilieuprägenden Wirkungsmechanismen kann grundsätzlich zwischen *Komposition-* und *Institutionseffekten* unterschieden werden: Mit *Kompositionseffekten* sind Einflüsse des schulischen Lernkontextes auf die Leistungsentwicklung von Schülerinnen und Schülern gemeint, die auf Unterschiede in der leistungsbezogenen, sozialen oder lernbiographischen Zusammensetzung der Schülerschaft zurückführbar sind. Kompositionseffekte wirken sich in der Regel nicht direkt auf die schulischen Leistungen aus. Vermittelt

über Anpassungen des Unterrichts an das Leistungs- und Fähigkeitsniveau einer Klasse, sozialen Vergleichsprozessen innerhalb der Schülerschaft oder normativen Wertvorstellungen (z.B. Leistungsnorm) innerhalb der Peergroup und Elternschaft können sie diese jedoch beeinflussen (Baumert et al., 2006). Als lernförderliche Umgebungen gelten insbesondere ein hohes mittleres Leistungsniveau einer Klasse sowie eine privilegierte soziale Klassenzusammensetzung, wobei die beiden Merkmale stark miteinander konfundiert sind (Dumont et al., 2013). *Institutionseffekte* hingegen beschreiben Einflüsse des schulischen Lernkontextes auf die Leistungsentwicklung von Schülerinnen und Schülern, die unabhängig von Einflüssen der Klassenzusammensetzung auf schultypspezifische Lehrpläne, Fachkompetenzen der Lehrkräfte und Unterrichtskulturen zurückführbar sind (Baumert et al., 2006). So kann ein anspruchsvollerer Lehrplan an den Gymnasien zu grösseren Lernzuwächsen führen. Aber auch zwischen den Schultypen ohne formale Differenzierung der Lehrpläne, in denen leistungsgruppierete Schulklassen nach demselben Lehrplan, aber auf unterschiedlichen Anspruchsniveaus unterrichtet werden, können curriculare Effekte eine Rolle spielen (Bos & Scharenberg, 2010). Zudem erfolgt der Unterricht im Gymnasium durch Fachlehrpersonen, während in den Schultypen mit geringeren Anforderungen (z.B. Real- und Sekundarschulen) Klassenlehrpersonen meist fächerübergreifend unterrichten (Neumann et al., 2007). Gymnasiallehrkräfte sind im Vergleich zu den übrigen Lehrkräften der Sekundarstufe nicht nur fachlich, sondern auch fachdidaktisch besser ausgebildet (Kunter et al., 2011). Entsprechend sind auch qualitative Unterschiede in der Unterrichtsgestaltung feststellbar. Während im gymnasialen Unterricht Verstehensprozesse verstärkt angeregt und gefördert werden, fokussiert der Unterricht in den Schultypen mit geringen Leistungsanforderungen eher auf das Einüben von Routinen mit wenig Raum für kognitive Selbständigkeit (Kunter et al., 2011).

Kompositions- und Institutionseffekte lassen somit erwarten, dass in den anspruchsvolleren Schultypen der Sekundarstufe I vorteilhaftere Lern- und Entwicklungsmilieus vorherrschen und sich dadurch die Leistungsdisparitäten zwischen den Schultypen im Verlauf der Sekundarstufe I verstärken. Die bisherigen Forschungsbefunde zu schultypspezifischen Fördereffekten ergeben jedoch ein gemischtes Bild, insbesondere was die in dieser Untersuchung betrachteten Fachbereiche betrifft (Becker, 2009; Neumann et al., 2014). Im Bereich Mathematik deuten die bisherigen Befunde mehrheitlich auf eine Verstärkung der schultypspezifischen Leistungsdisparitäten hin, wobei vor allem für die Gymnasien höhere Lernzuwächse im Vergleich zu den weniger anspruchsvollen Schultypen feststellbar sind (Baumert et al., 2006; Bosen, Gröhlich, & Bos, 2009; Guill & Gröhlich, 2013; Köller & Baumert, 2001; van Ophuysen & Wendt, 2010). Es existieren jedoch auch Untersuchungen, die keine spezifische Förderwirkung des Gymnasiums finden (Baumert, Becker, Neumann, & Nikolova, 2009). Im Bereich Deutsch fallen die bisherigen Ergebnisse deutlich heterogener aus. Hinsichtlich der Entwicklung des Leseverständnisses beispiels-

weise finden sich sowohl Belege für schultypspezifische Leistungsentwicklungen zugunsten der anspruchsvolleren Schultypen (Gröhlich, Guill, Scharenberg, & Bos, 2010; Pfof & Artelt, 2013) wie auch für weitgehend parallele Lernzuwächse zwischen den verschiedenen Schultypen (Becker, McElvany, Lüdtke, & Trautwein, 2014; Köller, Schütte, Zimmermann, Retelsdorf, & Leucht, 2013; Retelsdorf, Becker, Köller, & Möller, 2012).

Für die Sekundarstufe I der Schweiz liegt bisher nur eine Längsschnittuntersuchung zu schultypspezifischen Leistungsentwicklungen in den deutschsprachigen Teilen der Kantone Wallis und Fribourg vor (Neumann et al., 2007). Untersucht wurde die Leistungsentwicklung im Fach Französisch als Fremdsprache im Verlauf der 8. Jahrgangsstufe. Auch nach Kontrolle individueller Eingangsvoraussetzungen zeigten sich deutliche Unterschiede in der Leistungsentwicklung zugunsten der Schülerinnen und Schüler der anspruchsvolleren Schultypen (Progymnasium und Sekundarschule). Inwiefern sich dieses Ergebnis auch für die Kernfächer Deutsch und Mathematik und für Leistungsentwicklung während der gesamten Sekundarschulzeit verallgemeinern lässt, ist aber bis jetzt ungeklärt.

Die vorliegende Studie untersucht am Beispiel des Zürcher Schulsystems, ob Schülerinnen und Schüler mit vergleichbaren Ausgangsleistungen und sozialer Herkunft im Verlauf der Sekundarstufe I je nach besuchtem Schultyp unterschiedliche Leistungszuwächse in Deutsch und Mathematik erzielen. Im Kanton Zürich wechseln die Schülerinnen und Schüler im Anschluss an die Primarschule in der Regel entweder ins Gymnasium oder in einen der beiden kognitiv unterschiedlich anspruchsvollen Schultypen der Sekundarschule – den Schultyp mit erweiterten Anforderungen (Abteilung A) und den Schultyp mit Grundanforderungen (Abteilung B). Um möglichst unkonfundierte Effekte des besuchten Schultyps auf die Leistungsentwicklung schätzen zu können, erfolgen die Analysen mithilfe von Propensity Score Matching (Baumert et al., 2009; Stuart, 2010). Schultypspezifische Leistungsentwicklungen werden zum einen für den Besuch des Gymnasiums im Vergleich zum Schultyp mit erweiterten Anforderungen und zum anderen für den Besuch des Schultyps mit erweiterten Anforderungen im Vergleich zum Schultyp mit Grundanforderungen überprüft. Ausgehend von den theoretischen Ausführungen und bisherigen Befunden werden sowohl für Deutsch wie auch für Mathematik jeweils höhere Lernzuwächse in den anspruchsvolleren Schultypen erwartet. Als Folge davon vergrößern sich die bildungsgangspezifischen Leistungsdisparitäten im Verlauf der Sekundarstufe I.

Daten und methodisches Vorgehen

Die empirischen Analysen basieren auf Daten der *Zürcher Längsschnittstudie*, eine für den Kanton Zürich repräsentative Längsschnittuntersuchung zur

Leistungsentwicklung in Deutsch und Mathematik während der obligatorischen Schulzeit mit bisher vier Erhebungswellen (Angelone, Keller, & Moser, 2013; Moser, Stamm, & Hollenweger, 2005). Die erste Erhebung fand im Jahr 2003 unmittelbar nach Schuleintritt statt. Danach folgten im Abstand von drei Jahren weitere drei Erhebungen am Ende der dritten Primarklasse, am Ende der sechsten Primarklasse – kurz vor Übergang in die Sekundarstufe I – und am Ende der Sekundarstufe I. Für die vorliegenden Analysen wurden Leistungsdaten der Erhebungen am Ende der Primarstufe und am Ende der Sekundarstufe I und Kontextdaten der Erhebung am Ende der dritten Primarklasse verwendet. Um Unterschiede im Leistungszuwachs möglichst präzise auf den Schultyp zurückführen zu können, wurden jene Schülerinnen und Schüler ($N = 220$) aus der Analyse ausgeschlossen, die im Verlauf der Sekundarstufe I den Schultyp gewechselt haben oder einen irregulären Schulverlauf (Repetition oder Akzeleration) aufwiesen. Dies führte in den einzelnen Schultypen jeweils zu keiner signifikanten Veränderung der Schülerzusammensetzung. Für die Analysen standen Daten zu insgesamt 950 Schülerinnen und Schülern zur Verfügung (vgl. Tabelle 1).

Eingangsselektivität in die Sekundarstufe I und Propensity Score Matching

Zur Schätzung kausaler Effekte des besuchten Schultyps auf die Leistungsentwicklung muss die Eingangsselektivität in die Sekundarstufe I kontrolliert werden (Baumert et al., 2009). Ein gängiges statistisches Verfahren zur Schätzung von Treatmenteffekten auf Basis von Beobachtungsdaten stellt das Propensity Score Matching (PSM) dar (Stuart, 2010). Grundgedanke des PSM ist es, unter Berücksichtigung relevanter Kontrollvariablen für alle Beobachtungen der Treatment- und Kontrollgruppe individuelle Wahrscheinlichkeiten (Propensity Score) für die Zugehörigkeit zur Treatmentgruppe zu bestimmen und anschliessend nur Beobachtungen der Treatment- und Kontrollgruppe mit möglichst ähnlichen Wahrscheinlichkeiten miteinander zu vergleichen.

Die Propensity Score Matching-Analysen zur Schätzung schultypspezifischer Leistungsentwicklungen erfolgen jeweils in drei Schritten (Baumert et al., 2009; Stuart, 2010): (1) In einem ersten Schritt wird für jede Beobachtung der Treatment- und Kontrollgruppe der Propensity Score (PS) bestimmt, das heisst die individuelle Wahrscheinlichkeit, der Treatmentgruppe anzugehören. In der vorliegenden Untersuchung bilden jeweils die Schülerinnen und Schüler des anspruchsvolleren Schultyps (z.B. Gymnasium) die Treatmentgruppe und jene des weniger anspruchsvolleren Schultyps (z.B. Schultyp mit erweiterten Anforderungen) die Kontrollgruppe. Der PS wurde mittels logistischer Regression unter Berücksichtigung einer Reihe selektionsrelevanter Schülermerkmale bestimmt. Dabei wurden die Testleistungen in Deutsch und Mathematik am Ende der Primarschule (T1), die Deutsch- und Mathematiknoten aus dem übergangsrelevanten Primarschulzeugnis, die soziale Herkunft, die Erstsprache und das

Geschlecht der Schülerinnen und Schüler in die Analysen einbezogen. Bei den metrischen Variablen – Testleistungen, Schulnoten und soziale Herkunft – wurden zur Berücksichtigung nichtlinearer Effekte auf den Übergang neben den linearen auch die quadratischen Terme einbezogen. In Anlehnung an Rubin und Thomas (1996) wurde der PS aufgrund besserer metrischer Eigenschaft in Logits transformiert. (2) In einem zweiten Schritt erfolgt das Matching der Treatment- und Kontrollgruppe anhand der geschätzten PS. Dazu wurde ein 10:1 Nearest Neighbour Matching mit Zurücklegen und einem Caliper von 0.1 Standardabweichungen des PS (Logit) verwendet. Für das Matching wurden nur Beobachtungen der Treatmentgruppe innerhalb der Area of Common Support (ACS) berücksichtigt (Morgan & Winship, 2007). Die ACS entspricht im Kontext von PSM dem Überlappungsbereich der PS-Verteilung in der Treatment- und Kontrollgruppe, d.h. jenem Bereich, in dem überhaupt vergleichbare Beobachtungen zwischen den beiden Gruppen vorliegen. Nur für die ACS sind Aussagen zu Treatmenteffekten ohne weitere Zusatzannahmen möglich (Baumert et al., 2009; Morgan & Winship, 2007). Zur Überprüfung der Matchingqualität wurden sowohl für den PS wie auch für die berücksichtigten Kontrollvariablen sog. standardisierte Bias vor und nach dem Matching berechnet. Der standardisierte Bias berechnet sich durch die Differenz der Mittelwerte in der Kontroll- und Treatmentgruppe im Verhältnis zur Standardabweichung in der Treatmentgruppe und sollte nach dem Matching möglichst minimiert werden, wobei Schwellenwerte von 0.25 Standardabweichungen als akzeptabel gelten (Rubin, 2001; Stuart, 2010). (3) In einem dritten Schritt erfolgt die Schätzung des kausalen Treatmenteffekts auf der Basis der gematchten Stichprobe. PSM erlaubt die Schätzung des sog. Average Treatment Effect on the Treated (ATT), d.h. des durchschnittlichen Effekts des Treatments innerhalb der Treatmentgruppe. Der ATT kann entweder durch einen (gewichteten) Mittelwertvergleich des Outcome zwischen der Kontroll- und Treatmentgruppe oder mittels Regression des Outcome auf die Treatmentvariable berechnet werden (Stuart, 2010). Das Matching wurde mit dem Stata-Modul *PSMATCH2* (Leuven & Sianesi, 2003) durchgeführt.

Instrumente und Variablen

Leistungswerte: Mit der Zürcher Längsschnittstudie wurden die schulischen Leistungen in Deutsch und Mathematik mithilfe lehrplanbasierter Leistungstests zu vier Messzeitpunkten erfasst (vgl. Angelone et al., 2013). Die Deutschtests umfassten Aufgaben aus den Lehrplanbereichen «Texte überarbeiten» (Sprachbetrachtung, Grammatik und Rechtschreibung) und «Texte lesen und verstehen». Mit den Mathematiktests wurden hauptsächlich elementare mathematische Fertigkeiten aus den Bereichen «Operationen», «Algebra», «Gleichungen / Ungleichungen», «Größen / Sachrechnen», «Proportionalität und Funktionen» und «Geometrie» geprüft. Am Ende der Sekundarstufe I wurde ein Multi-Matrix-Design mit drei verschiedenen anspruchsvollen Tests eingesetzt, die durch

Link-Items miteinander verbunden sind. Die Schülerinnen und Schüler lösten somit entsprechend des besuchten Schultyps unterschiedlich anspruchsvolle Testaufgaben. Mit Ausnahme der Leistungstests beim Schuleintritt wurde ein Anker-Item-Testdesign (Kolen & Brennan, 2014) eingesetzt, das die Skalierung der Deutsch- bzw. Mathematiktests am Ende der dritten und sechsten Primarklasse sowie am Ende der Sekundarstufe I auf einer einheitlichen Messskala ermöglichte. Die Testergebnisse zu den verschiedenen Erhebungszeitpunkten wurden jeweils pro Fachbereich nach dem Rasch-Modell skaliert. Die Testleistungen am Ende der dritten Primarklasse wurden auf einen Mittelwert von 650 Punkten und eine Standardabweichung von 100 Punkten standardisiert. Als Schätzer für die Schülerfähigkeiten dienen in der vorliegenden Untersuchung *Weighted Likelihood Estimates* (WLEs; vgl. Warm, 1989). Die WLE-Reliabilitäten der eingesetzten Tests liegen zwischen 0.82 und 0.95. Deckeneffekte in den Leistungstests am Ende der 9. Klasse sind nicht vorhanden. Die mittleren Lösungswahrscheinlichkeiten in Deutsch bzw. Mathematik betragen im Gymnasium 67 bzw. 44 Prozent, im Schultyp mit erweiterten Anforderungen 64 bzw. 37 Prozent und im Schultyp mit Grundanforderungen 64 bzw. 35 Prozent.

Tabelle 1: Deskriptive Beschreibung der Analytestichprobe (N = 950)

	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Missing</i> ¹
Schultyp auf Sekundarstufe I ¹					
- Gymnasium	19				0.0
- Erweiterte Anforderungen	48				0.0
- Grundanforderungen	33				0.0
Knaben ¹	46				0.0
Nichtdeutsche Erstsprache ¹	31				0.0
Soziale Herkunft (Index)	0.0	1.0	-2.9	3.2	0.0
Übergangsrelevante Primarschulnoten ²					
- Deutsch	4.7	0.6	3.0	6.0	2.1
- Mathematik	4.6	0.7	2.5	6.0	2.1
Deutschleistungen					
- Ende Primarschule (T1)	791	106	422	1143	0.0
- Ende Sekundarstufe I (T2)	888	107	366	1188	0.0
- Effektstärke Zuwachs pro Schuljahr ³	<i>d</i> = 0.31				
Mathematikleistungen					
- Ende Primarschule (T1)	889	103	511	1235	0.0
- Ende Sekundarstufe I (T2)	937	109	482	1221	0.0
- Effektstärke Zuwachs pro Schuljahr ³	<i>d</i> = 0.16				

Anmerkungen: ¹Prozentwerte. ²Halbnotenskala 6 bis 1: 6 (sehr gut); 5 (gut); 4 (genügend); 3 (ungenügend); < 3 (schwach). ³Effektstärke *d* bezieht sich auf die Standardabweichung der Ausgangsleistungen (T1).

Treatment: Das Treatment entspricht dem besuchten Schultyp auf der Sekundarstufe I. Dabei wurde zwischen dem Gymnasium und zwei kognitiv verschieden anspruchsvollen Schultypen der Sekundarschule unterschieden, dem Schultyp mit erweiterten Anforderungen (Abteilung A) und dem Schultyp mit Grundanforderungen (Abteilung B). Die Angaben zum Schultyp der Schülerinnen und Schüler entstammen der Bildungsstatistik des Kantons Zürichs (BISTA).

Kontrollvariablen: Zur Kontrolle der Eingangsselektivität in die Sekundarstufe I wurden neben den Testleistungen in Deutsch und Mathematik am Ende der Primarschule (T1), die übergangsrelevanten Deutsch- und Mathematiknoten, die soziale Herkunft, die gelernte Erstsprache und das Geschlecht der Schülerinnen und Schüler berücksichtigt. Die Noten stammen aus dem ersten Schulzeugnis der sechsten Primarklasse. Zur Bestimmung der sozialen Herkunft der Schülerinnen und Schüler wurde mittels Hauptkomponentenanalyse ein Index der ökonomischen und kulturellen Ressourcen des Elternhauses gebildet. Die ökonomischen Ressourcen wurden anhand der Wohnungsgrösse (Anzahl Zimmer pro Person) und die kulturellen Ressourcen anhand der Anzahl Bücher im Haushalt sowie dem höchsten Bildungsabschluss der Eltern erfasst. Der Index weist einen Mittelwert von 0 und eine Standardabweichung von 1 auf. Bei der gelernten *Erstsprache* der Schülerinnen und Schüler wurde zwischen der deutschen (entspricht der Unterrichtssprache) und nichtdeutschen Erstsprache unterschieden. Die Angaben zur sozialen Herkunft und zur Erstsprache wurden dem Elternfragebogen aus der zweiten Erhebungswelle am Ende der dritten Primarklasse entnommen. Der Anteil fehlender Werte in den Analysevariablen beträgt höchstens 2.1 Prozent (vgl. Tabelle 1).

Ergebnisse zu schultypspezifischen Förder- effekten in Deutsch und Mathematik

Gymnasium vs. Schultyp mit erweiterten Anforderungen
Zur Untersuchung unterschiedlicher Förderwirkungen zwischen dem Gymnasium und dem Schultyp mit erweiterten Anforderungen stehen eine Treatmentgruppe von 178 Gymnasiastinnen und Gymnasiasten und eine Kontrollgruppe von 452 Schülerinnen und Schülern des Schultyps mit erweiterten Anforderungen zur Verfügung. Tabelle 2 zeigt die Ergebnisse der logistischen Regression zur Schätzung des PS, d. h. der individuellen Wahrscheinlichkeit, auf der Sekundarstufe I das Gymnasium zu besuchen.

Tabelle 2: Ergebnisse der logistischen Regression zur Vorhersage des Übergangs ins Gymnasium (Treatmentgruppe) vs. Schultyp mit erweiterten Anforderungen (Kontrollgruppe)

	<i>b</i> -Koeff.	(SE)
Deutschleistung T1 (z-stand.)	0.924*	(0.365)
Deutschleistung T1 (quadriert)	0.314	(0.219)
Mathematikleistung T1 (z-stand.)	0.676	(0.363)
Mathematikleistung T1 (quadriert)	0.115	(0.189)
Deutschnote (z-stand.)	2.208*	(0.904)
Deutschnote (quadriert)	-0.198	(0.479)
Mathematiknote (z-stand.)	0.700	(0.393)
Mathematiknote (quadriert)	1.089***	(0.315)
Soziale Herkunft (z-stand.)	1.995***	(0.377)
Soziale Herkunft (quadriert)	-0.205	(0.218)
Nichtdeutsche Erstsprache (Ref.: Deutsche Erstsprache)	1.121*	(0.546)
Knaben (Ref.: Mädchen)	-0.231	(0.361)
Konstante	-6.358***	(0.659)
Pseudo R^2	0.691	
Anteil korrekt klassifizierter Fälle	0.924	
N	630	

Anmerkungen: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Die Ergebnisse zeigen, dass beim Übergang ins Gymnasium eine starke Selektion nach Leistung und sozialer Herkunft stattfindet. Die Testleistungen¹, die Schulnoten und die soziale Herkunft wirken sich signifikant positiv auf die Übergangschancen ins Gymnasium aus. Unabhängig davon wirkt sich eine nichtdeutsche Erstsprache positiv auf den Übergang aus – ein Ergebnis, das auch in anderen Untersuchungen feststellbar ist und auf vergleichsweise höhere Bildungsmotivationen von Migrantenfamilien zurückgeführt wird (vgl. z.B. Baumert et al., 2009; Kristen & Dollmann, 2009). Einzig das Geschlecht der Schulkinder weist keinen signifikanten Einfluss auf die Übergangschancen ins Gymnasium aus. Der Erklärungsbeitrag der berücksichtigten Hintergrundmerkmale ist mit einem Pseudo- R^2 von 69.1 Prozent und einen Anteil von 92.4 Prozent korrekt klassifizierten Fällen substantiell. Abbildung 1a zeigt die Verteilung des PS (Logit) im Gymnasium und im Schultyp mit erweiterten Anforderungen und die Grösse der ACS, d.h. jenen Bereich, in dem überhaupt vergleichbare Schülerinnen und Schüler zwischen dem Gymnasium und dem Schultyp mit erweiterten Anforderungen vorliegen.

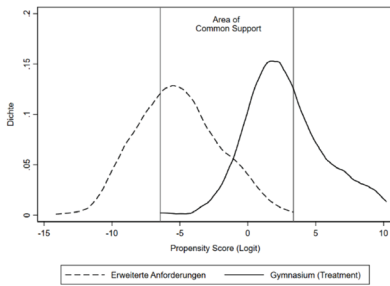


Abb. 1 a)

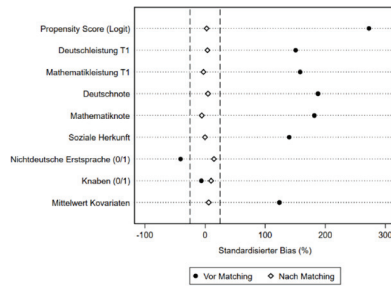


Abb. 1 b)

Abbildung 1 a): Verteilung des Propensity Score (Logit) im Gymnasium und im Schultyp mit erweiterten Anforderungen

Abbildung 1 b): Standardisierte Bias zwischen dem Gymnasium und dem Schultyp mit erweiterten Anforderungen vor und nach dem Matching (positive Werte zugunsten des Gymnasiums)

Wie aus der Abbildung hervorgeht, unterscheiden sich die PS-Verteilungen deutlich zwischen den beiden Schultypen. Im Gymnasium beträgt der Mittelwert des PS 2.5 Logits (SD = 2.7), gegenüber lediglich -4.8 Logits (SD = 2.9) im Schultyp mit erweiterten Anforderungen. Rund 63 Prozent der Gymnasiastinnen und Gymnasiasten ($N = 113$) befinden sich innerhalb der ACS und können somit für das Matching berücksichtigt werden. Durch das verwendete Matchingverfahren konnte jeder Beobachtung der Treatmentgruppe innerhalb der ACS mindestens eine hinreichend ähnliche Beobachtung der Kontrollgruppe zugewiesen werden. Abbildung 1b zeigt, dass das Matching zu einer ausreichenden Balancierung der Vergleichsgruppen geführt hat. Vor dem Matching lag der standardisierte Bias des PS (Logit) bei 271 Prozent, gegenüber lediglich 2 Prozent nach dem Matching (vgl. Tabelle 6 im Anhang). Bei den Hintergrundmerkmalen hat sich der Mittelwert der standardisierten Bias von 120 Prozent auf 10 Prozent reduziert und die standardisierten Bias aller Kovariaten liegen unter dem Schwellenwert von 25 Prozent (Rubin, 2001; Stuart, 2010).

Der abschliessende Analyseschritt ist der Leistungsvergleich zwischen der gematchten Treatment- und Kontrollgruppe. Tabelle 3 zeigt die Mittelwerte der Deutsch- und Mathematikleistungen am Ende der Sekundarstufe I.

Im Einklang mit den Erwartungen sprechen die Ergebnisse in Deutsch für einen substanziellen Entwicklungsvorteil der Gymnasiastinnen und Gymnasiasten. Der Lernvorteil im Gymnasium beträgt 0.34 Standardabweichungen der Ausgangsleistungen (T1) in der Gesamtstichprobe. Gemäss Cohens (1988) Kriterien zur Beurteilung von Effektstärken würde es sich hierbei um einen kleinen bis mittleren Unterschied handeln. Vergleicht man den Lernvorteil im

Gymnasium jedoch mit dem durchschnittlichen Lernzuwachs pro Schuljahr auf der Sekundarstufe I von $d = 0.31$ (vgl. Tabelle 1) – vergleichbare Zuwachsraten finden sich auch in anderen Untersuchungen (Becker et al., 2014) –, ist dieser durchaus als bedeutsam zu bezeichnen. Aufgrund der Ergebnisse in Mathematik hingegen ist keine spezifische Förderwirkung des Gymnasiums feststellbar. Am Ende der Sekundarstufe I unterscheiden sich die durchschnittlichen Mathematikleistungen nicht statistisch signifikant zwischen den beiden Vergleichsgruppen. Folglich öffnet sich die Leistungsschere zwischen dem Gymnasium und dem Schultyp mit erweiterten Anforderungen in Deutsch, nicht aber in Mathematik.

Tabelle 3: Mittelwertsunterschiede in den Deutsch- und Mathematikleistungen am Ende der Sekundarstufe I (T2) für vergleichbare Schülerinnen und Schüler des Gymnasiums und des Schultyps mit erweiterten Anforderungen (N = 113 Paare)

	Gymnasium (Treatmentgruppe)	Erweiterte Anforderungen (Kontrollgruppe)	Differenz (ATT)	(SE) ¹	Effekt- stärke d^2
Deutsch	995	959	36***	(9.0)	0.34
Mathematik	1034	1029	5	(10.3)	0.05

Anmerkungen: Gewichtet für multiples Matching. ¹Standardfehler durch Bootstrapping mit 100 Replikationen ermittelt. ²Effektstärke d bezieht sich auf die Standardabweichung der Ausgangsleistungen (T1) in der Gesamtstichprobe. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Schultyp mit erweiterten Anforderungen vs. Schultyp mit Grundanforderungen

Die Untersuchung unterschiedlicher Förderwirkungen zwischen dem Schultyp mit erweiterten Anforderungen und jenem mit Grundanforderungen basiert auf einer Treatmentgruppe von 452 Schülerinnen und Schülern des Schultyps mit erweiterten Anforderungen und einer Kontrollgruppe von 297 Schülerinnen und Schülern des Schultyps mit Grundanforderungen. Die Ergebnisse der logistischen Regression zur Schätzung des PS sind in Tabelle 4 dargestellt.

Auch beim Übergang in den Schultyp mit erweiterten Anforderungen ist eine starke positive Selektion nach Leistung (Testleistungen und Schulnoten) und sozialer Herkunft feststellbar. Zudem weisen die Knaben im Vergleich zu den Mädchen eine geringe Chance auf, im Anschluss an die Primarschule in einen Schultyp mit erweiterten Anforderungen zu wechseln.

Wiederum unterscheidet sich die Verteilung des PS (Logit) in den Vergleichsgruppen (vgl. Abbildung 2a). Im Schultyp mit erweiterten Anforderungen beträgt der mittlere PS 3.8 Logits (SD = 2.5), gegenüber lediglich -3.9 Logits (SD = 3.7 Logits) im Schultyp mit Grundanforderungen. Rund 46 Prozent ($N = 208$) der Schülerinnen und Schüler des Schultyps mit erweiterten Anforderungen fallen in die ACS und können somit für das Matching berücksichtigt werden. Wie Abbildung 2b zeigt, hat das Matching zu einer zufriedenstellenden Reduktion des Selektionsbias geführt.

Tabelle 4: Ergebnisse der logistischen Regression zur Vorhersage des Übergangs in den Schultyp mit erweiterten Anforderungen (Treatmentgruppe) vs. Schultyp mit Grundanforderungen (Kontrollgruppe)

	b-Koeff.	(SE)
Deutschleistung T1 (z-stand.)	1.216**	(0.398)
Deutschleistung T1 (quadriert)	0.116	(0.270)
Mathematikleistung T1 (z-stand.)	1.155***	(0.320)
Mathematikleistung T1 (quadriert)	0.262	(0.146)
Deutschnote (z-stand.)	2.059***	(0.376)
Deutschnote (quadriert)	-0.901*	(0.358)
Mathematiknote (z-stand.)	1.073***	(0.310)
Mathematiknote (quadriert)	-0.184	(0.237)
Soziale Herkunft (z-stand.)	0.472*	(0.201)
Soziale Herkunft (quadriert)	-0.017	(0.136)
Nichtdeutsche Erstsprache (Ref.: Deutsche Erstsprache)	0.304	(0.352)
Knaben (Ref.: Mädchen)	-1.182***	(0.313)
Konstante	3.603***	(0.363)
Pseudo R^2	0.680	
Anteil korrekt klassifizierter Fälle	0.907	
N	749	

Anmerkungen: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

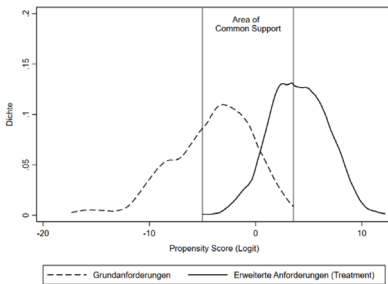


Abb. 2 a)

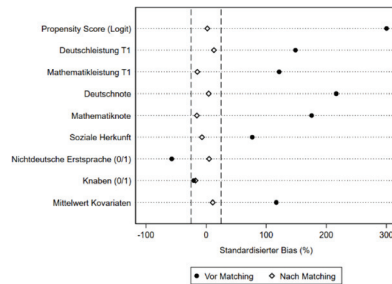


Abb. 2 b)

Abbildung 2 a): Verteilung des Propensity Score (Logit) im Schultyp mit erweiterten Anforderungen und im Schultyp mit Grundanforderungen

Abbildung 2 b): Standardisierte Bias zwischen dem Schultyp mit erweiterten Anforderungen und dem Schultyp mit Grundanforderungen vor und nach dem Matching (positive Werte zugunsten des Schultyps mit erweiterten Anforderungen)

Vor dem Matching lag der standardisierte Bias des PS (Logit) bei 306 Prozent, nach dem Matching bei 3 Prozent (vgl. Tabelle 7 im Anhang). Bei den Hintergrundmerkmalen hat sich der Mittelwert der standardisierten Bias von 115 Prozent auf 8 Prozent reduziert und die standardisierten Bias aller Kovariaten liegen unter dem Schwellenwert von 25 Prozent (Rubin, 2001; Stuart, 2010).

Tabelle 5: Mittelwertsunterschiede in den Deutsch- und Mathematikleistungen am Ende der Sekundarstufe I (T2) für vergleichbare Schülerinnen und Schüler des Schultyps mit erweiterten Anforderungen und Grundanforderungen (N = 208 Paare)

	Erweiterte Anforderungen (Treatmentgruppe)	Grundanforderungen (Kontrollgruppe)	Differenz (ATT)	(SE) ¹	Effektstärke d^2
Deutsch	873	851	22**	(7.2)	0.21
Mathematik	926	898	27***	(8.0)	0.26

Anmerkungen: Gewichtet für multiples Matching. ¹Standardfehler durch Bootstrapping mit 100 Replikationen ermittelt. ²Effektstärke d bezieht sich auf die Standardabweichung der Ausgangsleistungen (T1) in der Gesamtstichprobe. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Tabelle 5 schliesslich zeigt die Mittelwerte der Deutsch- und Mathematikleistungen am Ende der Sekundarstufe I. Im Einklang mit den Erwartungen zeigen die Ergebnisse, dass die schulischen Leistungen im Schultyp mit erweiterten Anforderungen in beiden Fachbereichen statistisch signifikant höher ausfallen als im Schultyp mit Grundanforderungen. In Deutsch beträgt der Lernvorteil im Schultyp mit erweiterten Anforderungen 0.21 Standardabweichungen der Ausgangsleistungen (T1) in der Gesamtstichprobe oder rund zwei Drittel des durchschnittlichen Lernzuwachses pro Schuljahr auf der Sekundarstufe I ($d = 0.31$, vgl. Tabelle 1). In Mathematik beträgt der Lernvorteil 0.26 Standardabweichungen. Dies entspricht zwar dem 1.6-fachen des durchschnittlichen Lernzuwachses pro Schuljahr auf der Sekundarstufe I. Allerdings gilt es zu berücksichtigen, dass dieser mit $d = 0.16$ im Vergleich zum Lernzuwachs in Deutsch – wie auch im Vergleich zum Lernzuwachs in Mathematik aus anderen Untersuchungen ($d = 0.3$ bis 0.5 , vgl. Becker, Lüdtke, Trautwein, & Baumert, 2006; Guill & Gröhlich, 2013; Köller & Baumert, 2001; van Ophuysen & Wendt, 2010) – deutlich geringer ausfällt.

Zusammenfassung und Konklusion

Mit der vorliegenden Untersuchung wurde mithilfe von Propensity Score Matching untersucht, ob und inwiefern vergleichbare Schülerinnen und Schüler im Verlauf der Sekundarstufe I in Abhängigkeit des besuchten Schultyps unterschiedliche Lernzuwächse in Deutsch und Mathematik erzielen. Die Ergebnisse der vorliegenden Analysen bestätigen die Befunde bisheriger Untersuchungen und zeigen auch für das gegliederte Schulsystem des Kantons Zürich Schereneffekte der Leistungsentwicklung zwischen den verschiedenen anspruchsvollen Bildungsgängen der Sekundarstufe I.

Bei gleichen Startvoraussetzungen lernen Schülerinnen und Schüler im Schultyp mit erweiterten Anforderungen im Verlauf der Sekundarschule mehr dazu als im Schultyp mit Grundanforderungen. In Deutsch beträgt der Lernvorteil im Schultyp mit erweiterten Anforderungen 0.21 Standardabweichungen der Ausgangsleistungen in der Gesamtstichprobe und entspricht rund zwei Dritteln des durchschnittlichen Lernzuwachses pro Schuljahr auf Sekundarstufe I ($d = 0.31$). In Mathematik beträgt der Lernvorteil 0.26 Standardabweichungen und entspricht sogar dem 1.6-fachen des durchschnittlichen jährlichen Lernzuwachses auf der Sekundarstufe I ($d = 0.16$). Allerdings gilt es dabei zu berücksichtigen, dass dieser in Mathematik im Vergleich zu Deutsch nur halb so gross ausfällt. Des Weiteren lernen Schülerinnen und Schüler im Gymnasium in Deutsch mehr dazu als vergleichbare Schülerinnen und Schüler im Schultyp mit erweiterten Anforderungen. Der Lernvorteil im Gymnasium beträgt 0.34 Standardabweichungen und entspricht ungefähr dem durchschnittlichen Lernzuwachs pro Schuljahr auf der Sekundarstufe I ($d = 0.31$).

Diese Ergebnisse deuten somit stark darauf hin, dass die Leistungsdifferenzierung in gegliederten Bildungssystemen einen Einfluss auf die Entstehung unterschiedlicher Lernkontexte ausübt, die den Schülerinnen und Schülern unabhängig von ihren individuellen Lernmöglichkeiten mehr oder weniger günstige Entwicklungschancen bieten. Dafür dürften sowohl institutionelle Unterschiede wie schultypspezifische Lehrpläne, Fachkompetenzen der Lehrkräfte und Unterrichtskulturen wie auch Unterschiede in der leistungsbezogenen und sozialen Komposition der Schülerschaft verantwortlich sein (Baumert et al., 2006; Neumann et al., 2007).

Einzig die Lernentwicklung in Mathematik unterscheidet sich aufgrund der vorliegenden Ergebnisse nicht zwischen dem Gymnasium und dem Schultyp mit erweiterten Anforderungen. Dieser Befund steht zwar im Widerspruch mit den Ergebnissen der meisten bisherigen Untersuchungen. Allerdings gilt es dabei zu berücksichtigen, dass mit der Erhebung vorwiegend elementare mathematische Fähigkeiten geprüft wurden. Diese gehören zwar zu den Basisqualifikationen, bilden jedoch nicht den Kern des gymnasialen Mathematikunterrichts. Um ein umfassenderes Bild spezifischer Fördereffekte des Gymnasiums zu erlangen, müssten auch anspruchsvollere mathematische Fähigkeiten getestet werden.

Die Ergebnisse zeigen jedoch auch, dass mathematische Basisqualifikationen im Gymnasium nicht besser gefördert werden als im Schultyp mit erweiterten Anforderungen (vgl. auch Baumert et al., 2009).

Einschränkend gilt es für die vorliegende Untersuchung zu erwähnen, dass sich die berichteten Befunde nicht ohne weitere Zusatzannahmen auf die gesamte Schülerschaft der Sekundarstufe I verallgemeinern lassen. Die Ergebnisse der Matching-Analysen beziehen sich auf eine Teilgruppe vergleichbarer Schülerinnen und Schüler, für die aufgrund ihrer schulischen Leistungen am Ende der Primarschule kein eindeutig vorgezeichneter Bildungsweg feststand. Zudem gilt es im Zusammenhang mit dem gewählten Matching-Verfahren auch zu berücksichtigen, dass ein Teil der festgestellten schultypspezifischen Förder-effekte auf unbeobachtete Eingangsselektivität in die Sekundarstufe I zurückführbar sein könnte.

Anmerkungen

- ¹ Linearer und quadratischer Term der Mathematikleistung sind gemeinsam signifikant ($p = 0.002$).

Anhang

Tabelle 6: Verteilung der Hintergrundkovariaten im Gymnasium und im Schultyp mit erweiterten Anforderungen vor und nach dem Matching

Vor Matching	Gymnasium (N = 178)		Erweiterte Anforderungen (N = 452)		Std. Bias %	Test auf Mittelwerts- differenz	
	M	SE	M	SE		t	p
Propensity Score (Logit)	2.54	0.20	-4.82	0.14	271	29.25	0.000
Deutschleistung T1 (z-stand.)	1.19	0.06	0.13	0.03	142	16.68	0.000
Mathematikleistung T1 (z-stand.)	1.15	0.06	0.06	0.04	143	15.77	0.000
Deutschnote (z-stand.)	1.14	0.04	0.23	0.03	188	17.75	0.000
Mathematiknote (z-stand.)	1.08	0.04	0.19	0.03	182	15.61	0.000
Soziale Herkunft (z-stand.)	0.97	0.05	0.00	0.04	140	13.48	0.000
Nichtdeutsche Erstsprache (0/1)	0.12	0.02	0.25	0.02	-41	-3.68	0.000
Knaben (0/1)	0.40	0.04	0.43	0.02	-6	-0.69	0.488
Mittelwert Kovariaten ³					120		
Nach Matching	Gymnasium (N=113)		Erweiterte Anforderungen (N = 113) ¹		Std. Bias %	Test auf Mittelwerts- differenz	
	M	SE ²	M	SE ²		t	p
Propensity Score (Logit)	1.01	0.16	0.97	0.16	2	0.16	0.876
Deutschleistung T1 (z-stand.)	0.92	0.06	0.82	0.07	15	1.25	0.210
Mathematikleistung T1 (z-stand.)	0.92	0.07	0.93	0.07	-1	-0.10	0.921
Deutschnote (z-stand.)	1.02	0.05	1.06	0.05	-8	-0.50	0.619
Mathematiknote (z-stand.)	0.94	0.04	1.00	0.05	-14	-1.03	0.302
Soziale Herkunft (z-stand.)	0.85	0.06	0.87	0.06	-3	-0.27	0.786
Nichtdeutsche Erstsprache (0/1)	0.12	0.03	0.06	0.03	18	1.47	0.140
Knaben (0/1)	0.39	0.04	0.44	0.04	-10	-0.84	0.400
Mittelwert Kovariaten ³					10		

Anmerkungen: ¹Gewichtet für multiples Matching. ²Standardfehler durch Bootstrapping mit 100 Replikationen ermittelt. ³Mittelwert (|Std. Bias %).

Tabelle 7: Verteilung der Hintergrundkovariaten im Schultyp mit erweiterten Anforderungen und im Schultyp mit Grundanforderungen vor und nach dem Matching

Vor Matching	Erweiterte Anforderungen (N = 452)		Grundanforderungen (N = 297)		Std. Bias	Test auf Mittelwertsdifferenz	
	M	SE	M	SE		%	t
Propensity Score (Logit)	3.80	0.12	-3.93	0.21	306	34.01	0.000
Deutschleistung T1 (z-stand.)	0.13	0.03	-0.87	0.04	141	19.19	0.000
Mathematikleistung T1 (z-stand.)	0.06	0.04	-0.81	0.04	111	15.67	0.000
Deutschnote (z-stand.)	0.23	0.03	-1.10	0.04	217	27.78	0.000
Mathematiknote (z-stand.)	0.23	0.03	-1.04	0.04	182	22.92	0.000
Soziale Herkunft (z-stand.)	0.00	0.04	-0.66	0.05	77	10.17	0.000
Nichtdeutsche Erstsprache (0/1)	0.25	0.02	0.50	0.03	-57	-7.21	0.000
Knaben (0/1)	0.43	0.02	0.53	0.03	-21	-2.77	0.006
Mittelwert Kovariaten ³					115		
Nach Matching	Erweiterte Anforderungen (N = 208)		Grundanforderungen (N = 208) ¹		Std. Bias	Test auf Mittelwertsdifferenz	
	M	SE ²	M	SE ²		%	t
Propensity Score (Logit)	1.60	0.11	1.56	0.15	3	0.21	0.834
Deutschleistung T1 (z-stand.)	-0.29	0.04	-0.31	0.05	2	0.20	0.843
Mathematikleistung T1 (z-stand.)	-0.32	0.05	-0.33	0.07	2	0.15	0.877
Deutschnote (z-stand.)	-0.17	0.04	-0.19	0.04	3	0.27	0.786
Mathematiknote (z-stand.)	-0.16	0.04	-0.12	0.05	-6	-0.63	0.526
Soziale Herkunft (z-stand.)	-0.22	0.06	-0.08	0.07	-16	-1.52	0.127
Nichtdeutsche Erstsprache (0/1)	0.33	0.03	0.25	0.05	16	1.52	0.128
Knaben (0/1)	0.52	0.03	0.57	0.04	-9	-1.03	0.304
Mittelwert Kovariaten ³					8		

Anmerkungen: ¹Gewichtet für multiples Matching. ²Standardfehler durch Bootstrapping mit 100 Replikationen ermittelt. ³Mittelwert (|Std. Bias %).

Literaturverzeichnis

- Angelone, D., Keller, F., & Moser, U. (2013). *Entwicklung schulischer Leistungen während der obligatorischen Schulzeit. Bericht zur vierten Zürcher Lernstandserhebung zuhanden der Bildungsdirektion des Kantons Zürich*. Zürich: Bildungsdirektion Kanton Zürich.
- Angelone, D., Ramseier, E., & Moser, U. (2010). Schulstruktur und Selektivität. In D. Angelone, E. Ramseier, C. Brühwiler, V. Morger, U. Moser, & E. Steiner (Hrsg.), *PISA 2006 in der Schweiz. Die Kompetenzen der Schülerinnen und Schüler im kantonalen Vergleich* (S. 72-99). Oberentfelden: Sauerländer Verlage AG.
- Baumert, J., Becker, M., Neumann, M., & Nikolova, R. (2009). Frühübergang in ein grundständiges Gymnasium – Übergang in ein privilegiertes Entwicklungsmilieu? *Zeitschrift für Erziehungswissenschaft*, 12(2), 189–215.
- Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In J. Baumert, P. Stanat, & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (S. 95–188). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, M. (2009). *Kognitive Leistungsentwicklung in differenziellen Lernumwelten. Effekte des gegliederten Sekundarschulsystems in Deutschland*. Berlin: Max-Planck-Inst. für Bildungsforschung.
- Becker, M., Lüdtke, O., Trautwein, U., & Baumert, J. (2006). Leistungszuwachs in Mathematik. Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? *Zeitschrift für Pädagogische Psychologie*, 20(4), 233–242.
- Becker, M., McElvany, N., Lüdtke, O., & Trautwein, U. (2014). Lesekompetenzen und schulische Lernumwelten. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 46(1), 35–50.
- Bonsen, M., Gröhlich, C., & Bos, W. (2009). Differentielle Lern- und Entwicklungsmilieus in der Hamburger Beobachtungsstufe? In W. Bos, M. Bonsen & C. Gröhlich (Hrsg.), *KESS 7 - Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Jahrgangsstufe 7* (S. 113–122). Münster: Waxmann.
- Bos, W., & Scharenberg, K. (2010). Lernentwicklung in leistungshomogenen und -heterogenen Schulklassen. In W. Bos, E. Klieme, & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 173–194). Münster: Waxmann.
- Cohen, J. (Ed.). (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dumont, H., Neumann, M., Maaz, K., & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen. Internationale und nationale Befunde. *Psychologie in Erziehung und Unterricht*, 60(3), 163–183.
- EDK/IDES. (2015). *Kantonsumfrage 2014-2015*. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) / Informations- und Dokumentationszentrum (IDES).
- Gröhlich, C., Guill, K., Scharenberg, K., & Bos, W. (2010). Differenzielle Lern- und Entwicklungsmilieus beim Erwerb der Lesekompetenz in den Jahrgangsstufen 7 und 8. In W. Bos & C. Gröhlich (Hrsg.), *KESS 8. Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 8* (S. 100–117). Münster: Waxmann.
- Guill, K., & Gröhlich, C. (2013). Individuelle Lernentwicklung im gegliederten Schulsystem der Bundesrepublik Deutschland. Fragen an die Sekundarstufe I. In K. Schwippert, M. Bosen & N. Berkemeyer (Hrsg.), *Schul- und Bildungsforschung. Diskussionen, Befunde und Perspektiven* (S. 51–69). Münster: Waxmann.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking. Methods and practices*. New York, NY: Springer.

- 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(2), 99–110.
- Köller, O., Schütte, K., Zimmermann, F., Retelsdorf, J., & Leucht, M. (2013). Starke Klasse, hohe Leistungen? Die Rolle der Leistungsstärke der Klasse für die individuellen Mathematik- und Leseleistungen in der Sekundarstufe I. *Psychologie in Erziehung und Unterricht*, 0(3), 184–197.
- Kristen, C., & Dollmann, J. (2009). Sekundäre Effekte der ethnischen Herkunft: Kinder aus türkischen Familien am ersten Bildungsübergang. In J. Baumert, K. Maaz, & U. Trautwein (Hrsg.), *Bildungsentscheidungen. Zeitschrift für Erziehungswissenschaften. Sonderheft 12* (S. 205–229). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Leuven, E., & Sianesi, B. (2003) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing (Version 4.0.11) [Computer software].
- Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge, UK: Cambridge University Press.
- Moser, U., Stamm, M., & Hollenweger, J. (Hrsg.). (2005). *Für die Schule bereit? Lesen, Wortschatz, Mathematik und soziale Kompetenzen beim Schuleintritt*. Oberentfelden: Sauerländer.
- Neumann, M., Becker, M., & Maaz, K. (2014). Soziale Ungleichheiten in der Kompetenzentwicklung in der Grundschule und der Sekundarstufe I. In K. Maaz, M. Neumann, & J. Baumert (Hrsg.), *Herkunft und Bildungserfolg von der frühen Kindheit bis ins Erwachsenenalter. Forschungsstand und Interventionsmöglichkeiten aus interdisziplinärer Perspektive* (Zeitschrift für Erziehungswissenschaft, Sonderheft 24, S. 167–203). Wiesbaden: Springer VS.
- Neumann, M., Schnyder, I., Trautwein, U., Niggli, A., Lüdtke, O., & Cathomas, R. (2007). Schulformen als differenzielle Lernmilieus: Institutionelle und kompositionelle Effekte auf die Leistungsentwicklung im Fach Französisch. *Zeitschrift für Erziehungswissenschaft*, 10(3), 499–420.
- Pfost, M., & Artelt, C. (2013). Reading literacy development in secondary school and the effect of differential institutional learning environments. In M. Pfost, C. Artelt, & S. Weinert (Hrsg.), *The development of reading literacy. Empirical findings from the Bamberg BiKS longitudinal studies* (S. 229–277). Bamberg: University of Bamberg Press.
- Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology*, 82(4), 647–671.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2, 169–188.
- Rubin, D. B., & Thomas, N. (1996). Matching using estimated Propensity Scores. Relating theory to practice. *Biometrics*, 52(1), 249.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21.
- Van Ophuysen, S., & Wendt, H. (2010). Zur Veränderung der Mathematikleistung von Klasse 4 bis 6. Welchen Einfluss haben Kompositions- und Unterrichtsmerkmale? *Zeitschrift für Erziehungswissenschaft, Sonderheft 12*, 302–327.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450.

Schlagworte: Leistungsentwicklung, Schultyp, Längsschnittstudie, Propensity Score Matching

Des effets de ciseaux dans le degré secondaire I? L'influence des filières scolaires sur l'amélioration des performances en allemand et en mathématiques

Résumé

La présente étude longitudinale (N = 950) utilise Propensity Score Matching pour déterminer si des élèves comparables améliorent leurs performances en allemand et en mathématiques de manière différente selon le type de filière fréquentée au cours du degré secondaire I. Les résultats montrent que les disparités de performances se creusent au cours du degré secondaire I, en raison de l'augmentation des performances spécifique au type de filière scolaire suivie. Outre les effets connus de l'origine socio-économique des élèves sur le développement des performances au cours des premières années de scolarité, d'autres effets liés à la structure de l'école interviennent donc au degré secondaire I.

Mots-clés: Développement des performances, filières scolaires, étude longitudinale, Propensity Score Matching

Divario di prestazioni nel livello secondario I? L'influenza della tipologia di scuola sul miglioramento delle prestazioni in tedesco e matematica

Riassunto

Questo studio longitudinale (N = 950) utilizza Propensity Score Matching per stabilire se alunni comparabili tra loro migliorino in modo diverso le loro prestazioni in tedesco e matematica nel livello secondario I, in base al tipo di scuola frequentata. I risultati mostrano come l'incremento delle disparità delle prestazioni nel corso della scuola secondaria I dipenda dall'aumento delle prestazioni specifico per il tipo di percorso scolastico seguito. Oltre al già riconosciuto impatto delle origini sociali sulle prestazioni scolastiche nei primi anni di scuola, si rilevano dunque ulteriori effetti dovuti alla struttura scolastica nel livello secondario I.

Parole chiave: Sviluppo delle prestazioni, tipologie di scuola, studio longitudinale, Propensity Score Matching

Fan spread effects at lower secondary school? The effect of school type on achievement gains in German and mathematics

Summary

Applying Propensity Score Matching this longitudinal study ($N = 950$) examines whether comparable students achieve different achievement gains depending on the school type attended. The results show increasing achievement disparities as a result of school type specific achievement gains. Therefore additional effects of the school structure at lower secondary level exist, beside the known effects of origin on achievement gains in the early school years.

Keywords: Achievement development, school type, longitudinal study, Propensity Score Matching

Domenico Angelone (lic, phil.), ist Doktorand an der Universität Bern (Institut für Erziehungswissenschaften, Abteilung Bildungssoziologie) und bei der Aufgabendatenbank der EDK (ADB) angestellt, die im Rahmen der Überprüfung der nationalen Bildungsziele in der Schweiz für die Koordination und Entwicklung standardisierter Leistungstests verantwortlich ist. Seine Forschungsgebiete sind soziale Bildungsungleichheiten, Bildungsmonitoring und Large-Scale-Assessments.

Universität Bern, Abteilung Bildungssoziologie, CH-3012 Bern
E-Mail: domenico.angelone@students.unibe.ch