

Leistungsvergleich der Neuen Mittelschule mit der AHS-Unterstufe und der Hauptschule

Heinz Leitgöb, Johann Bacher und Christoph Weber

1 Problemstellung

In der Diskussion über die Wirksamkeit der Neuen Mittelschule (NMS) wird regelmäßig die Frage aufgeworfen, ob diese sich durch den Einsatz der bereitgestellten finanziellen Mittel und den vorgegebenen Qualitätskriterien (siehe Helm, Kemethofer & Pieslinger, 2015) neben anderen angestrebten Wirkungszielen (z.B. Verbesserung der bildungsbezogenen Chancengerechtigkeit; siehe dazu Bruneforth, Weber & Bacher, 2012) auch in der Effektivität der Kompetenzvermittlung deutlich von den Hauptschule (HS) absetzt sowie zur Unterstufe der allgemein höherbildenden Schulen (AHS) vergleichbare Ergebnisse hervorbringt. Dieser Frage folgend wurde von Schreiner und Breit (2013, S. 92ff) auf Basis der Daten aus der letzten Bildungsstandarderhebung (BIST-E8 aus dem Jahr 2013) eine Gegenüberstellung der schulischen Performanz der SchülerInnen aus den Schultypen NMS, HS und AHS-Unterstufe (jeweils 8. Schulstufe) im getesteten Kompetenzbereich Englisch vorgenommen. Im Kern konnten keine relevanten Verteilungsunterschiede für die durch Mittelwertbildung auf Schulebene aggregierten Testleistungen zwischen der NMS und der HS identifiziert werden, während ein deutliches Leistungsgefälle zur AHS-Unterstufe zu beobachten ist (Schreiner & Breit, 2013, S. 93).

Dieser einfache Vergleich der Leistungsverteilungen zwischen NMS, AHS und AHS-Unterstufe ist mit dem Problem behaftet, dass die systematische Variation der sozialen Zusammensetzung der SchülerInnenschaft der jeweiligen Schulformen keine Berücksichtigung findet. So konnten in den vergangenen Jahren zahlreiche bildungssoziologische Untersuchungen den empirischen Nachweis erbringen, dass in den hierarchisch gegliederten und strukturell verhältnismäßig ähnlichen aufgebauten Bildungssystemen von Österreich und Deutschland eine starke Abhängigkeit des formalen Bildungserfolgs sowie der in den internationalen Bildungserhebungen gemessenen schulischen Kompetenzen von der sozialen und nationalen bzw. ethnischen Herkunft vorliegt (siehe etwa Bacher, 2003; 2005; 2009; 2010; Bacher, Leitgöb & Weber, 2012; Baumert, Maaz & Trautwein, 2009; Baumert, Stanat, Watermann, 2006a; Becker, 2009; Becker & Lauterbach, 2010; Becker & Reimer, 2010; Becker & Solga, 2012; Breit, 2009; Bruneforth et al., 2012; Diefenbach, 2007; Herzog-Punzenberger & Unterwurzacher, 2009; Leitgöb, Bacher, Bruneforth & Weber, 2014; Leitgöb & Lachmayr, 2012; Maaz, Baumert & Trautwein, 2009; Maaz, Neumann & Baumert, 2014; Müller-Benedict, 2007; Schimpl-Neimanns, 2000; Schreiner & Breit, 2006; Stubbe, 2009). Hierfür zeichnen sich soziale Selektionseffekte von Kindern in die verschiedenen Schultypen sowie soziale Kompositionseffekte in den Schulen verantwortlich (siehe Abschnitt 2).

Aufgrund der genannten Effekte sollte bei einem Vergleich der schulischen Leistungen der NMS mit der AHS-Unterstufe und der HS auf alle Fälle die soziale Zusammensetzung der SchülerInnenschaft in den Schulen der jeweiligen Schulformen berücksichtigt werden. Dies wird im vorliegenden Beitrag geleistet. So wird der Frage nachgegangen, welche Unterschiede in den Leistungsniveaus zwischen den NMS und der AHS-Unterstufe einerseits sowie der HS andererseits bei annähernd gleicher sozialer Komposition bestehen.

Ferner sollen – basierend auf der Annahme einer großen Heterogenität der Kompetenzniveaus innerhalb der NMS (es wird vermutet, dass es unter Berücksichtigung der sozialen Zusammensetzung der SchülerInnenpopulation manchen NMS gelingt, annähernd das Leistungsniveau von AHS-Unterstufenschulen zu erreichen, während andere NMS mit enormen Leistungsabständen konfrontiert sind) – jene Merkmale identifiziert werden, die zwischen NMS mit einer geringen und einer hohen Leistungsdifferenz zur AHS-Unterstufe differenzieren.

Zusammenfassend wird den folgenden Fragestellungen nachgegangen:

(I) Welche Leistungsunterschiede bestehen zwischen NMS und AHS-Unterstufe bei annähernd gleicher sozialer Zusammensetzung der Schule?

(II) Welche Leistungsunterschiede bestehen zwischen NMS und HS bei annähernd gleicher sozialer Zusammensetzung der Schule?

(III) Durch welche Merkmale unterschieden sich NMS mit geringem Leistungsabstand zur AHS-Unterstufe von jenen NMS mit großem Leistungsabstand zur AHS-Unterstufe?

2 Theoretische Überlegungen und Forschungsstand

Für systematische leistungsbezogene Divergenzen zwischen der NMS, der AHS-Unterstufe und der HS am Ende der Sekundarstufe I (8. Schulstufe) lassen sich neben den interessierenden institutionellen und unterrichtsbezogenen Unterschieden noch zumindest zwei Alternativerklärungen finden. Neben sozialen Selektionsprozessen, die bereits am Beginn der Sekundarstufe I (5. Schulstufe) zu variierenden Ausgangsniveaus in den kognitiven und schulischen Fähigkeiten zwischen den Schultypen führen, können die Kompetenzunterschiede (zumindest partiell) auch auf eine differentielle soziale Zusammensetzung der SchülerInnenschaft in der NMS, der AHS-Unterstufe und der HS zurückzuführen sein.

Als theoretische Grundlage für soziale Selektionseffekte von SchülerInnen in bestimmte Schulformen kann auf das von Boudon (1974) vorgeschlagene und etwa von Erikson und Jonsson (1996), Breen und Goldthorpe (1997), Esser (1999), Becker (2000) sowie Maaz, Baumert und Trautwein (2009) weiterentwickelte, Rational-Choice-basierte Konzept der Differenzierung des Einflusses der sozialen Schichtzugehörigkeit auf die elterlichen Bildungsentscheidungen in primäre und sekundäre soziale Herkunftseffekte zurückgegriffen werden.

Primäre Herkunftseffekte wirken in direkter Weise auf die Kompetenzentwicklung der Kinder und manifestieren sich in deren Schulleistungen. Bestehende sozialschichtbedingte Leistungsunterschiede lassen sich in der Hauptsache durch eine divergente Ausstattung der Herkunftsfamilien mit ökonomischem, kulturellem und sozialem Kapital (Bourdieu, 1983) begründen (Maaz & Nagy, 2009). Die Ungleichverteilung der genannten Ressourcen führt zu einer differentiellen Kompetenzentwicklung der Kinder, da alle drei Kapitalsorten für die Förderung der schulischen und kognitiven Entwicklung investiert bzw. zum Einsatz gebracht werden können. Weiterhin tragen schichtspezifisch geprägte familiäre Sozialisationsprozesse (z.B. die elterliche Vermittlung einer bildungsaffinen Einstellung und eine daraus resultierende Aktivierung der Kinder zu schulischem Lernen) zu einer ungleichen Ausgangsverteilung der Bildungschancen bei (Maaz & Nagy, 2009).

In der Folge wirken die primären Herkunftseffekte über die schulischen Leistungen der Kinder positiv auf die von den Eltern perzipierte Erfolgswahrscheinlichkeit des Abschlusses höherer Bildungsgänge. Darüber hinaus berechtigt ein guter Schulerfolg über die Erfüllung der formalen Anforderungen des Bildungssystems zur Partizipation an höheren Bildungsgängen.

Sekundäre Effekte der sozialen Herkunft determinieren unabhängig von den schulischen Leistungen bzw. dem erreichten Kompetenzniveau der Kinder die elterlichen Entscheidungskalküle im Zuge des Bildungswahlprozesses. „Bildungsentscheidungen werden im Kontext der eigenen sozialen Stellung getroffen und sind im Zusammenhang der familiären Bildungsvorstellungen und Lebensplanung zu verstehen“ (Maaz et al., 2009, S. 15). Demnach liegen, abhängig von der sozialen Schichtzugehörigkeit der Eltern, unterschiedliche Bewertungsschemata der Kosten und des Nutzens der möglichen Bildungsoptionen zugrunde.

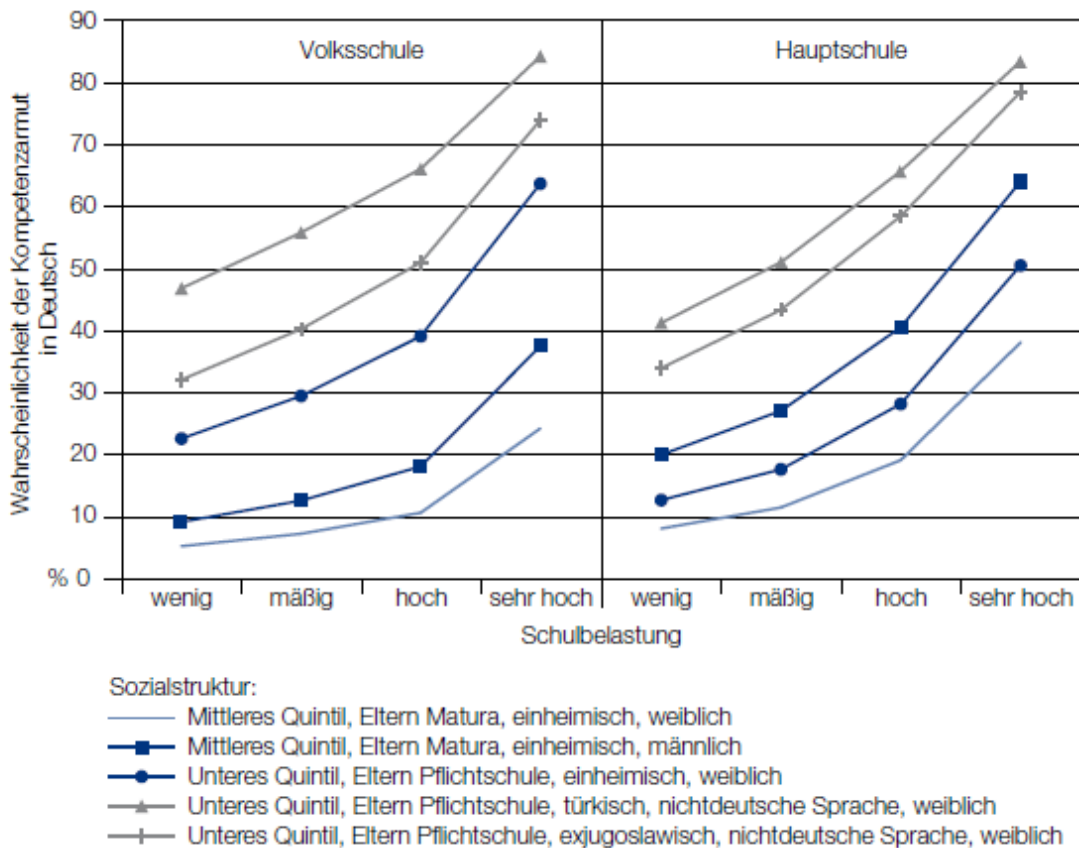
Für Österreich liegen für den ersten Bildungsübergang (Primarstufe in die Sekundarstufe I) stabile empirische Befunde über erhebliche primäre und sekundäre Herkunftseffekte vor (siehe etwa Bruneforth et al., 2012; Leitgöb et al., 2014). Theoriekonform fallen die sekundären Herkunftseffekte stärker aus als die primären, da die Fähigkeiten und Begabungen der Kinder nach vier Jahren Volksschule (VS) noch nicht treffsicher beurteilt werden können (Erikson & Jonsson, 1996). Diese empirischen Evidenzen beziehen sich allerdings ausschließlich auf die Entscheidung zwischen den beiden bisherigen Alternativen der AHS-Unterstufe und der HS. Für die NMS liegen bislang keine Befunde vor. Allerdings ist aus theoretischer Sicht zu erwarten, dass sich gerade in den ersten Jahren nach der Implementierung der NMS die Eltern von leistungsstarken Kindern aus höheren sozialen Schichten verstärkt gegen diese Schulform entscheiden, da aufgrund der fehlenden Erfahrungen mit der NMS und der damit verbundenen Unsicherheit der Nettonutzen der NMS nicht in zufriedenstellendem Maße abgeschätzt werden kann. Dies kann zu Befürchtungen führen, dass in der NMS die bisherigen guten schulischen Leistungen der Kinder nicht weiter aufrechterhalten werden können, das Prestige der NMS deutlich hinter jenem der AHS-Unterstufe liegt und so ein Umstieg auf einen maturaführenden Bildungsgang in der Sekundarstufe II erschwert werden könnte, die allgemein angestrebten Bildungsziele nicht bzw. nur mit einem höherem Aufwand erreicht werden könnten und somit die Gefahr von intergenerationalen Statusverlusten droht. Um diesen potentiellen Risiken von Anfang an entgegenzuwirken liegt es nahe, dass sich diese Eltern auch weiterhin für den bewährten Schultyp der AHS-Unterstufe entscheiden und der bisherige soziale Selektionsprozess in die AHS-Unterstufe auch zukünftig Bestand hat. Demgegenüber ist zu vermuten, dass die Entscheidung zwischen NMS und HS vornehmlich entlang anderer Kriterien getroffen wird. In diesem Zusammenhang sollten Faktoren wie etwa die räumliche Nähe zur nächstgelegenen NMS bzw. HS eine stärkere Rolle spielen.

Zusammenfassend bleibt festzuhalten, dass – insbesondere in den ersten Jahren – soziale Selektionseffekte in die NMS zu erwarten sind. Es ist davon auszugehen, dass vermehrt jene Kinder eine NMS besuchen werden, die die formalen Aufnahmekriterien in eine AHS-Unterstufe nicht erfüllt haben und/oder aus sozial schlechter gestellten Herkunftsfamilien kommen.

Der Einfluss der sozialen Zusammensetzung der MitschülerInnen auf den individuellen Lernerfolg von SchülerInnen wird in der Bildungsforschung als Kompositionseffekt bezeichnet. Im Zuge einer Auseinandersetzung mit dieser Thematik konnten jüngst Brune-

forth et al. (2012) sowohl für die VS als auch für die HS und die AHS-Unterstufe in Österreich starke Kompositionseffekte auf die individuellen Deutsch- und Mathematikleistungen ermitteln. Weiterhin ist eine erhebliche Wirkung der sozialen Komposition der SchülerInnenpopulation auf individuelle Kompetenzarmut – die Wahrscheinlichkeit, dass ein/e SchülerIn ein definiertes Kompetenzminimum nicht erreicht – zu beobachten. Die entsprechenden Befunde werden für die VS und die HS in Abbildung 1 wiedergegeben.

Abbildung 1: Wahrscheinlichkeit der Kompetenzarmut für SchülerInnen unterschiedlicher sozialer Herkunft in Schulen unterschiedlicher Belastungskategorien in der 4. Schulstufe (VS) und der 8. Schulstufe (HS)



Entnommen aus Bruneforth et al. (2012, S. 213)

Aus Abbildung 1 wird ersichtlich, dass die Wahrscheinlichkeit für Kompetenzarmut in allen definierten sozialen Gruppen ansteigt, wenn die Schule eine ungünstigere soziale Zusammensetzung der SchülerInnen¹ aufweist. Während etwa für einen einheimischen Hauptschüler der 8. Schulstufe, dessen Eltern als höchstes formales Bildungszertifikat einen Maturaabschluss besitzen und eine mittlere berufliche Position innehaben, die geschätzte Wahrscheinlichkeit für Kompetenzarmut im Falle einer günstigen sozialen Zusammensetzung (Kategorie „wenig“) bei 20 Prozent liegt, steigt diese auf über 60 Prozent bei einer sehr ungünstigen sozialen Zusammensetzung der SchülerInnenschaft (Kategorie „sehr hoch“). In Schulen mit einer sehr ungünstigen sozialen Zusammensetzung kann die geschätzte Wahrscheinlichkeit für Kompetenzarmut für manche SchülerInnen sogar auf über 80 Prozent ansteigen.

¹ Die soziale Zusammensetzung der SchülerInnenschaft wurde über einen von Bruneforth et al. (2012) im Rahmen des zweiten nationalen Bildungsberichts vorgeschlagenen Index erfasst, der sich aus vier Sozialindikatoren zusammensetzt: (i) geringe schulische Bildung der Eltern, (ii) geringer beruflicher Status der Eltern, (iii) Migrationshintergrund, (iv) andere zu Hause gesprochene Sprache. Für nähere Details zu diesem Sozialindex (Bruneforth et al. (2012, S. 198) sprechen von einem Index der sozialen Benachteiligung bzw. Belastung einer Schule) sei auf Abschnitt 3.2 verwiesen.

Die Ergebnisse indizieren, dass für den Schul- bzw. Lernerfolg eines/r SchülerIn somit nicht nur die eigene soziale Herkunft von Relevanz ist, sondern auch die soziale Zusammensetzung der Schule bzw. der Klasse. Die Wirkung der sozialen Schichtzugehörigkeit der Herkunftsfamilie bildet sich in den Niveauunterschieden der Verlaufskurven ab (primärer Herkunftseffekt). So sind etwa SchülerInnen mit Eltern aus bildungsfernen Schichten, unabhängig von der sozialen Herkunft ihrer MitschülerInnen, mit einer höheren Wahrscheinlichkeit von Kompetenzarmut betroffen, als SchülerInnen mit formal hoch gebildeten Eltern. Innerhalb der sozialen Ungleichheitsdimensionen kommen der formalen Bildung und dem beruflichen Status der Eltern die stärkste Wirkung zu (Bacher et al., 2012; Leitgöb et al., 2014). Der deutlich schwächer ausgeprägte Effekt des Migrationshintergrunds wird – zumindest partiell – über die vertikalen Schichtmerkmale (insbesondere über den beruflichen Status der Eltern; siehe Bacher et al., 2012) vermittelt. Eine eigenständige Wirkung des Migrationshintergrunds bleibt insbesondere dann bestehen, wenn die Migrationsgruppen feiner differenziert (z.B. nach Gruppen von Herkunftsländern) oder die Alltagssprache bzw. die zu Hause gesprochene Sprache als Migrationsindikator verwendet werden (Bruneforth et al., 2012; Herzog-Punzenberger & Schnell, 2012). Dies lässt sich auch durch Abbildung 1 empirisch belegen. So besitzen türkischstämmige Mädchen bei gleicher formaler Bildung der Eltern ein durchgehend höheres Risiko für Kompetenzarmut als einheimische Mädchen.

Eine umfassende Untersuchung von Kompositionseffekten im deutschen Bildungssystem (Sekundarstufe) wurde von Baumert, Stanat und Watermann (2006b) vorgenommen. Als soziale Kompositionsmerkmale auf Schulebene wurde auf den mittleren sozio-ökonomischen Status der Eltern sowie die Anteile der Eltern mit Hochschulreife bzw. ohne Berufsausbildung (= Indikatoren für den elterlichen Sozial- und Bildungsstatus), den Anteil der nicht Vollzeit erwerbstätigen Väter (= Indikator für belastende Familienverhältnisse), den Anteil der Eltern, die in der Familie nicht deutsch sprechen (= Indikator für den Migrationsstatus/die kulturelle Vielfalt), den Anteil der WiederholerInnen (= Indikator für lernbiographische Belastungen der SchülerInnen) und den durchschnittlichen KFT-Wert der SchülerInnen (= Indikator für das Leistungs- und kognitive Fähigkeitsniveau) rekuriert. Weiterhin wurden die Testleistungen in der Domäne Lesen (= Indikator für die schulische Leistung) und die Selbstwirksamkeitseinschätzung (= Indikator für die Persönlichkeitsentwicklung) als abhängige Variablen verwendet. Sämtliche Analysen wurden mit den im Rahmen von PISA 2000 erhobenen Daten (inklusive der nationalen Erweiterung PISA-E) und einem neu aufbereiteten Längsschnittdatensatz der BIJU-Studie (Bildungsverläufe und psychosoziale Entwicklung im Jugendalter)² durchgeführt.

Die Ergebnisse der geschätzten Mehrebenenmodelle offenbaren, dass die untersuchten Kompositionsmerkmale nicht voneinander isoliert, sondern in Kombination wirken. Ferner ist eine Variation dieser Wirkung über die Schulformen zu beobachten. Während für die gymnasiale Unterstufe keine nennenswerten Kompositionseffekte nachgewiesen werden können, liegen in der HS äußerst starke Effekte der sozialen Zusammensetzung der SchülerInnenschaft vor. Der stärkste Einfluss kann in diesem Zusammenhang für das elterliche Bildungsniveau, gefolgt vom Anteil an WiederholerInnen, dem mittleren kognitiven Fähigkeitsniveau und den belastenden Familienverhältnissen identifiziert werden. Im Gegensatz dazu kommt dem Migrationshintergrund, erfasst über die zu Hause gesprochene Sprache, als Kompositionsmerkmal keine Wirkung zu (S. 133ff).

² Siehe <http://www.uni-potsdam.de/quantmethoden/forschung/biju.html>.

In weiterführenden Interventionsanalysen konnte gezeigt werden, dass die Kompositionsmerkmale in Hauptschulen vermittelt über die Reaktionen der Lehrkräfte auf die soziale Zusammensetzung der SchülerInnen wirken. Auf Schulebene liegen zudem Effekte der Gewaltbereitschaft der SchülerInnen auf das Kompetenzniveau und die Persönlichkeitsentwicklung vor, allerdings besteht kein statistisch nachweisbarer Zusammenhang zwischen dem Anteil an gewaltbereiten SchülerInnen und der sozialen Zusammensetzung der SchülerInnenschaft (S. 146ff).

Die Befunde von Baumert et al. (2006b) weisen eine hohe Übereinstimmung mit den Ergebnissen einer Untersuchung von Rumberger und Palardy (2005) für die USA basierend auf Daten des National Educational Longitudinal Survey aus dem Jahr 1988 auf. Als abhängige Variablen dienten in diesem Fall die Kompetenzmessungen in den Domänen Mathematik, Naturwissenschaften, Lesen und Sozialwissenschaften sowie der Durchschnitt der vier Kompetenzbereiche für die 8., 10. und 12. Schulstufe. Die Ergebnisse der durchgeführten Analysen zeigen, dass der Effekt der sozialen Komposition auf Schulebene mitunter sogar stärker ausgeprägt ist als jener der sozialen Herkunft auf Individualebene. Analog zu Baumert et al. (2006b) kann die Wirkung der sozialen Zusammensetzung durch schulische Prozessmerkmale erklärt werden. Als bedeutsam erwiesen sich in diesem Zusammenhang der Umfang der Hausübungen, der Besuch von fortgeschrittenen (anspruchsvollen) Kursen, die Erwartungshaltung der Lehrkräfte und das Sicherheitsempfinden der SchülerInnen.

In einer rezenten Studie von Rjosk, Richter, Hochweber, Lüdtke, Klieme und Stanat (2014) können die Befunde bezüglich der mediiierenden Wirkung von schulischen Prozessmerkmalen lediglich partiell bestätigt werden. Anhand von Daten der vom Deutschen Institut für International Pädagogische Forschung (DIPF) durchgeführten DESI-Studie³ wurde die Fragestellung geprüft, ob der Einfluss der sozialen Zusammensetzung auf Klassenebene auf die Lesekompetenzen in Deutsch durch die Unterrichtsqualität erklärt werden kann. In diesem Zusammenhang wurde der mediiierende Einfluss der Betonung der Wichtigkeit der korrekten Verwendung der Sprache, eines unterstützenden schülerorientierten Unterrichts und eines strukturierten Unterrichts, in dem die Unterrichtsziele hervorgehoben und die aktive Mitwirkung der SchülerInnen gewährleistet wird, untersucht. Von den genannten Unterrichtsmerkmalen konnte lediglich für die Betonung der Wichtigkeit der korrekten Sprachverwendung eine intervenierende Wirkung nachgewiesen werden. Bestätigt werden die Befunde von Baumert et al. (2006b), dass auf Klassen- oder Schulebene nur die über vertikale soziale Ungleichheitsdimensionen definierte Zusammensetzung der Schülerschaft einen Einfluss auf das schulische Kompetenzniveau ausübt. Ein Kompositionseffekt des Migrationshintergrunds kann hingegen nicht beobachtet werden kann.

Ein stabiler Einfluss der sozialen Zusammensetzung der SchülerInnenschaft auf die schulische Performanz der SchülerInnen konnte auch von van Ewijk und Slegers (2010a, 2010b) im Rahmen von Metaanalysen festgestellt werden. Für den Migrationshintergrund ergeben sich ambivalente Resultate.

Die erläuterten empirischen Befunde bestätigen die Ergebnisse des sogenannten Coleman-Reports (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld & York, 1966) aus den 1960er Jahren hinsichtlich der Bedeutung der sozialen Zusammensetzung der

³ Deutsch-Englisch-Schülerleistungen-International; zu den Details siehe <http://www.dipf.de/de/forschung/projekte/deutsch-englisch-schuelerleistungen-international>.

SchülerInnenschaft. Während sich diese klassische und für die empirische Bildungsforschung richtungsweisende Arbeit verstärkt auf die Einflüsse des Freizeitverhaltens und der Peer-Group auf die schulischen Leistungen konzentriert, konnte in den diskutierten neueren Untersuchungen ergänzend die Bedeutsamkeit von einigen schulischen Prozessmerkmalen als Mediatoren zwischen den sozialen Kompositionsmerkmalen und den leistungsbezogenen Merkmalen herausgearbeitet. Allerdings sei darauf hingewiesen, dass der Freizeitbereich sowie die Peer Group in diesem Zusammenhang nach wie vor von Relevanz sind (siehe etwa Leitgöb, Paseka, Bacher & Altrichter, 2012).

Die diskutierten Einflüsse der sozialen Selektion von SchülerInnen in bestimmte Schulen bzw. Schulformen, der sozialen Herkunft auf der Individualebene und der sozialen Zusammensetzung auf Schul- und Klassenebene offenbaren, dass die soziale Komposition der SchülerInnen beim schulischen bzw. schultypenbezogenen Leistungsvergleich unbedingt berücksichtigt werden muss, um der Gefahr von Fehlschlüssen zu begegnen. Werden diese Effekte ignoriert, kann beispielsweise die gute Performanz einer Schule fälschlicherweise auf schulinterne Faktoren (z.B. auf qualifizierte und engagierte Lehrkräfte, die Unterrichtsqualität, strukturelle Rahmenbedingungen) zurückgeführt werden, obwohl diese ausschließlich durch eine günstige soziale Zusammensetzung der SchülerInnenpopulation bedingt ist. Daher sind in der empirischen Bildungsforschung mittlerweile sogenannte „faire“ Vergleiche üblich, bei denen versucht wird, die Ausgangssituation einer Schule zu berücksichtigen (siehe zusammenfassend Fiege, Reuther & Nachtigall, 2011). Im vorliegenden Beitrag wird dieser Tatsache Rechnung getragen, indem die Methode des Statistical Matching zum Einsatz gebracht wird (für eine kompakte allgemeine Einführung siehe Bacher, 2002; Gangl, 2010; zu den Details siehe Abschnitt 3).

3. Methodisches Vorgehen

3.1. Datenbasis

Zur Beantwortung der Fragestellungen werden die Bildungsstandardmessungen in Mathematik (BIST-M8; Schreiner & Breit, 2012) und Englisch (BIST-E8; Schreiner & Breit, 2013) herangezogen sowie die für die Evaluierung erhobenen Daten der NMS-Modellschulen (Abschlussverfahren von G1 und G2). Zur Abbildung der Schulleistungen wird aus Gründen der Datenverfügbarkeit auf die Weighted Likelihood Estimates (WLE; siehe etwa Warm, 1989) der allgemeinen Mathematikkompetenzen und der Lesekompetenzen in Englisch rekurriert. Details zu den Leistungsmessungen können dem publizierten Endbericht zur Evaluation der NMS (Eder, Altrichter, Hofmann & Weber, 2015) entnommen werden.

3.2. Statistical Matching

Um im Zuge des „fairen“ Vergleichs systematische Unterschiede in der schulischen Performanz zwischen den SchülerInnen der NMS, der AHS-Unterstufe und der HS ursächlich den Lernvermittlungsprozessen und den darauf wirkenden internen Strukturen und Prozessen in den einzelnen Schultypen zuschreiben zu können, müssen exogene Faktoren kontrolliert werden. Ist zu diesem Zweck eine randomisierte Zuordnung der SchülerInnen zu den Schulformen in Form eines experimentellen Designs aus ethischen, formalen

oder forschungspragmatischen Gründen nicht möglich, müssen jene (in Abschnitt 2 dargelegten) Prozesse, die für eine Selektion der SchülerInnen in die NMS, AHS-Unterstufe oder HS verantwortlich sind, explizit beachtet werden (zu den Grundlagen des Sample Selection Bias siehe etwa Heckman, 1979). Zudem gilt es, die daraus resultierenden Effekte der schultypenbezogenen Komposition der SchülerInnenschaft auf die Leistungen zu kontrollieren. Zur Berücksichtigung dieser Effekte spielen neben deren statistischer Kontrolle in multiplen Regressionsmodellen (auch als „Conditioning“ bezeichnet; für eine kritische Reflexion dieses Ansatzes siehe Morgan & Winship, 2007) nichtparametrische Matchingverfahren eine zentrale Rolle (siehe etwa Guo & Fraser, 2010; Morgan, 2013; Morgan & Winship, 2007; Schafer & Kang, 2008; Stuart, 2010).

Dem Matchingansatz liegt die Idee zugrundeliegende, über ein Set von Kovariaten \mathbf{x} eine der Treatmentgruppe (hier: die NMS-Schulen) statistisch möglichst ähnliche Vergleichsgruppe (hier: aus den verfügbaren Schulen der AHS-Unterstufe oder der HS) zu generieren, um – analog zur Forschungslogik in experimentellen Designs – den Einfluss potentieller exogener Faktoren auf die Schätzung des interessierenden kausalen Effekts zu eliminieren. Das Set von Kovariaten \mathbf{x} ist so zu wählen, dass es „eine hinreichende Teilmenge aller Merkmale enthält, die theoretisch wie empirisch das Auftreten von T^4 bestimmen, so dass die Verteilung der Treatmentbedingung T als konditional unabhängig von der Verteilung der abhängigen Variable Y betrachtet werden kann“ (Gangl, 2010, S. 933). Rosenbaum und Rubin (1983) formalisieren unter Verwendung der potentiellen Outcomes⁵ Y_0 und Y_1 die Bedingung wie folgt:

$$(Y_0, Y_1) \perp T | \mathbf{x} \text{ mit } 0 < P(T = 1 | \mathbf{x}) < 1 \quad (1)$$

Die bedingte Wahrscheinlichkeit der Zugehörigkeit zur Treatmentgruppe ($P(T = 1 | \mathbf{x})$) wird auch als Propensity Score⁶ bezeichnet. Ist die Bedingung aus (1) erfüllt⁷, gilt nach Rosenbaum und Rubin (1983; siehe dazu auch Heckman, 2005):

$$(Y_0, Y_1) \perp T | P(T = 1 | \mathbf{x}) \quad (2)$$

⁴ Die Variable T repräsentiert den Treatmentindikator ($T = 0$: Zugehörigkeit zur Vergleichsgruppe; $T = 1$: Zugehörigkeit zur Treatmentgruppe).

⁵ Gemäß der kontrafaktischen Nomenklatur (für eine Einführung siehe etwa Morgan & Winship, 2007) entspricht die abhängige Variable Y für ein Individuum/Objekt i dem Wert Y_{0i} , wenn $T_i = 0$ gegeben ist (i gehört der Vergleichsgruppe an) sowie dem Wert Y_{1i} , wenn $T_i = 1$ vorliegt (i gehört der Treatmentgruppe an). Der Umstand, dass lediglich einer der beiden Werte direkt beobachtet werden kann (i kann zwangsläufig entweder *nur* der Treatment- oder *nur* der Vergleichsgruppe zugeordnet werden), wird von Holland (1986) als „Fundamental Problem of Causal Inference“ deklariert und führte zur Bezeichnung von Y_{0i} und Y_{1i} als „Potential Outcomes“.

⁶ Der Propensity Score repräsentiert einen Balancing Score. Ein Balancing Score $b(\mathbf{x})$ ist allgemein eine Funktion des beobachtbaren Sets von Kovariaten \mathbf{x} , für die $\mathbf{x} \perp T | b(\mathbf{x})$ gilt (Rosenbaum & Rubin, 1983). Ist die Bedingung erfüllt, sind die Treatment- und die Vergleichsgruppe nach \mathbf{x} balanciert (zu den Details siehe Guo & Fraser, 2010, S. 55ff).

⁷ Ist dies nicht der Fall, kann der Treatmenteffekt nicht (vollständig) von den Selektionseffekten isoliert werden. Eine Verletzung der konditionalen Unabhängigkeitsbedingung kann auftreten, wenn nicht alle relevanten Kovariaten beobachtet bzw. berücksichtigt werden können (Stichwort „Selection on Observables“; Heckman & Robb, 1986), eine Kontaminierung der Kovariaten durch Messfehler vorliegt oder in \mathbf{x} eine extreme Heterogenität zwischen den Gruppen gegeben ist.

Diese Annahme der konditionalen Unabhängigkeit der Outcomevariable von der Zuweisung zum Treatment bei gegebenen Propensity Scores (Strongly Ignorable Treatment Assignment Assumption) stellt das Grundprinzip des Propensity Score Matchings (PSM; Rosenbaum & Rubin, 1983; 1985; Rubin, 1974) dar. Das Verfahren ist weit verbreitet und wird auch im vorliegenden Beitrag zur Anwendung gebracht.

Der Vorzug des Matchingverfahrens gegenüber dem parametrischen Ansatz des „Conditioning“ ist neben der für die Anwendung im Kontext der Bildungswahl überzeugenden Argumentation (explizite Modellierung der Zugehörigkeit zur Treatmentgruppe, im Gegensatz zum parametrischen Ansatz erschwerte Missspezifikationen durch nicht haltbare funktionale oder verteilungsbezogene Annahmen, Reduktion der Dimensionalität von einer Reihe von Kovariaten auf einen Propensity Score, es können explizite Balancierungschecks durchgeführt werden) von Becker, Lüdtke, Trautwein, Köller & Baumert (2012) dem Umstand geschuldet, dass es zur Klärung von Forschungsfrage (III) der Identifikation von NMS bedarf, die eine relative Leistungsnähe bzw. -ferne zur AHS-Unterstufe bei vergleichbarer sozialer Zusammensetzung der SchülerInnenschaft aufweisen.

Prinzipiell lässt sich das PSM in zwei zentrale Schritte unterteilen: (i) die Schätzung von Propensity Scores und (ii) die eigentliche Matchingprozedur.

(i) Die Schätzung der Propensity Scores. In diesem Schritt werden das Set von Kovariaten sowie die funktionale Form des Modells zur Schätzung der Propensity Scores ausgewählt. Im vorliegenden Fall werden jene vier Indikatoren, die den von Bruneforth et al. (2012) vorgeschlagenen Sozialindex für Schulen (von Bruneforth et al. als „Index der sozialen Benachteiligung“ bezeichnet) konstituieren und der Charakterisierung der sozialen Zusammensetzung der SchülerInnenschaft entlang vertikaler und horizontaler herkunftsbezogener Ungleichheitsdimensionen (siehe etwa Geißler, 2014) dienen, als Set von Kovariaten⁸ herangezogen:

x_1 = geringe schulische Bildung der Eltern (1 = maximal Pflichtschulabschluss, 0=sonst)

x_2 = geringer beruflicher Status der Eltern (1 = HISEI bis 33 Punkte, 0=sonst)

x_3 = Migrationshintergrund (1 = ja, 0 = nein)

x_4 = andere zu Hause gesprochene Sprache (1 = ja, 0 = nein)

Die Aggregation der $j = 4$ Individualindikatoren auf Schulebene k erfolgt über die Bildung von Mittelwerten

⁸ Operationalisierung der Kovariaten

x_1 : Der Indikator besitzt den Wert 1, wenn beide Elternteile als höchsten formalen Bildungsabschluss maximal einen Abschluss des Polytechnischen Lehrgangs (ISCED 3C) aufweisen.

x_2 : HISEI steht für „Highest International Socio-Economic Index“ und basiert auf dem ISEI (International Socio-Economic Index), einem von Ganzeboom, de Graaf und Treiman (1992) entwickelten Index zur Messung des sozioökonomischen Status von Personen mit einem Wertebereich von 16 (landwirtschaftliche Hilfskräfte, Reinigungskräfte) bis 90 (Richter). In Anlehnung an die Konvention im Large Scale Student Assessment PISA entspricht HISEI dem ISEI jenes Elternteils, der einen ISEI-Wert aufweist (OECD, 2012, S. 281). Der verwendete Cutpoint von 33 Punkten entspricht dem unteren Quintil in der Kalibrierungsstichprobe (BIST-BL4; Bruneforth et al., 2012, S. 198).

x_3 : Im Zuge der Operationalisierung des Migrationshintergrunds erfolgt ebenfalls eine Orientierung an PISA. Migrationshintergrund liegt vor, wenn beide Elternteile im Ausland geboren wurden (OECD, 2012, S. 281). Auf die Differenzierung in eine erste und zweite Migrationsgeneration wird an dieser Stelle verzichtet.

x_4 : Es wurde abgefragt, ob Deutsch die Alltagssprache der SchülerInnen ist.

$$\bar{x}_{jk} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{jik} \quad (3)$$

mit x_{jik} Wert des/der SchülerIn i der Schule k im Sozialindikator j .

Da alle Indikatoren 0/1-codierte dichotome Variable repräsentieren, lassen sich die Schulmittelwerte als Anteile interpretieren. So entspricht \bar{x}_{1k} beispielsweise dem Anteil an SchülerInnen in der achten Schulstufe der Schule k , deren Eltern maximal einen Pflichtschulabschluss besitzen.

Da sich die soziale Komposition der SchülerInnenschaft zwischen den Schulen der NMS und der AHS-Unterstufe als äußerst heterogen gestaltet, konnte im Zuge des PSM keine Balancierung der Gruppen in den vier Indikatoren erzielt werden.⁹ Aus diesem Grund wurde zur Erzeugung einer Vergleichsgruppe aus den verfügbaren AHS-Unterstufen nicht auf die vier Einzelindikatoren, sondern auf den Sozialindex als einziger Kovariate rekurriert. Der Sozialindex für eine Schule k wurde wie folgt berechnet: In einem ersten Schritt wurde für jeden der $i = 1, \dots, n_k$ SchülerInnen in Schule k ein Individualsozialindex s_{ik} ermittelt:

$$s_{ik} = 1 + \frac{1}{4} \sum_{j=1}^4 x_{jik} \quad (4)$$

Die Aggregationsvorschrift für den schulbezogenen Sozialindex S_k lautet:

$$S_k = \frac{100}{n_k} \sum_{i=1}^{n_k} s_{ik} \quad (5)$$

Durch die Multiplikation mit 100 weist der Sozialindex einen theoretischen Wertebereich zwischen 100 und 200 auf. Höhere Werte des Index implizieren eine stärkere Akkumulation von sozial benachteiligten SchülerInnen in einer Schule.

Zur Schätzung der Propensity Scores wird, getrennt für AHS-Unterstufe und HS, das folgende Logit-Modell spezifiziert:

$$P_k(NMS = 1 | \mathbf{x}_j) = \frac{\exp(\bar{x}_{jk}\boldsymbol{\beta})}{1 + \exp(\bar{x}_{jk}\boldsymbol{\beta})} \quad (6)$$

Wie bereits erläutert, wird für den Vergleich der NMS mit der AHS-Unterstufe der lineare Prädiktor aus (6) durch $\beta_0 + \beta_1 S_k$ ersetzt. An dieser Stelle sei explizit darauf hingewiesen, dass eine Verletzung von (1) bzw. (2) durch in den Daten nicht verfügbare und somit für das Matching fehlende Kovariaten nicht ausgeschlossen werden kann. So kann etwa der Einfluss der schulischen Performanz der SchülerInnen am Ende der Primarstufe auf die Entscheidung für einen bestimmten Schultyp über den primären Schichteffekt lediglich partiell abgebildet werden. Die gewählte Vorgehensweise wird allerdings dem Anspruch gerecht, einen „fairen“ Vergleich der Schulformen durch die Berücksichtigung der jeweils spezifischen sozialen Zusammensetzung der SchülerInnenschaft zu gewährleisten.

(ii) *Die Matchingprozedur.* Im zweiten Schritt des PSM wird versucht, Elemente in der Vergleichsgruppe zu identifizieren, die jenen der Treatmentgruppe in den geschätzten Propensity Scores möglichst ähnlich sind. Zu diesem Zweck wird im vorliegenden Beitrag auf

⁹ Für eine ausführliche Diskussion dieses Befundes sei auf Abschnitt 4 verwiesen.

das „Nearest Neighbor Matching“ (NNM; Rubin, 1973) zurückgegriffen. Die Paarbildung zwischen den Elementen der Treatment- und der Vergleichsgruppe erfolgt beim NNM auf Basis des Prinzips der maximalen Ähnlichkeit der Propensity Scores (zu den Details siehe etwa Guo & Fraser, 2010). Sei P_{κ} der Propensity Score einer Schule κ aus der Treatmentgruppe der NMS ($\kappa \in NMS$) und P_{κ^*} der Propensity Score einer Schule κ^* aus der Vergleichsgruppe der AHS-Unterstufen bzw. der HS ($\kappa^* \in AHS$ oder $\kappa^* \in HS$), dann gilt als Matchingkriterium, dass die absolute Differenz der Propensity Scores zwischen κ und κ^* ein Minimum unter allen in der Vergleichsgruppe verfügbaren AHS-Unterstufen bzw. HS darstellen muss:

$$d(P_{\kappa}, P_{\kappa^*}) = \min_{\kappa^*} \|P_{\kappa} - P_{\kappa^*}\| \quad (7)$$

Um ein Matching von einander hinsichtlich der Sozialstruktur sehr unähnlichen Schulen zu verhindern, wird ein Schwellenwert ε (auch als „Caliper“ bezeichnet) für $d(P_{\kappa}, P_{\kappa^*})$ definiert:

$$\|P_{\kappa} - P_{\kappa^*}\| < \varepsilon \quad (8)$$

Wird ε überschritten erfolgt keine Paarbildung. Rubin und Rosenbaum (1985) schlagen zur Festlegung der Größe des Calipers $\varepsilon = 0,25\sigma_p$ vor, wobei σ_p der Standardabweichung der Propensity Scores im Gesamtsample entspricht. Nach dieser Definition würde der Wertebereich des Calipers in den vorliegenden Daten zwischen 0,01 (BIST-M8; Vergleichsgruppe: HS) und 0,08 (BIST-M8; Vergleichsgruppe: AHS) variieren. Um eine optimale Ausschöpfung der Schulen bei angemessener Homogenität in den gemachten Paaren zu gewährleisten, wurde der Caliper für alle durchgeführten NNM auf 0,05 gesetzt, d.h. für eine Paarbildung muss in der jeweiligen Vergleichsgruppe (AHS bzw. HS) zumindest eine Schule enthalten sein, deren Propensity Score sich von jenem der Schule κ um maximal fünf Prozentpunkte unterscheidet.¹⁰

Hinsichtlich der Anzahl der zulässigen Matchingpartner pro NMS wurde auf ein 1:1 Matching rekurriert. Für jede Schule κ aus NMS wurde somit exakt eine Schule aus AHS bzw. HS gemäß den Kriterien (7) und (8) ausgewählt. Erfüllten mehrere Schulen aus den Vergleichsgruppen die Kriterien, entscheidet sich der NNM-Algorithmus für den „nächsten Nachbarn“ im Datensatz. Dies erfordert eine Zufallsanordnung der Schulen. Um den Umstand zu berücksichtigen, dass die erzeugte Vergleichsgruppe keine wohl definierte „reale“ Gruppe repräsentiert, sondern auf Schätzern (Propensity Scores) und Zufallsprozessen (Zufallsanordnung der Elemente im Datensatz) basiert, wird ein multiples PSM mit $m = 10$ Durchläufen realisiert.¹¹ Diese Vorgehensweise erzeugt – analog zur Logik der multiplen Imputation (siehe etwa Rubin, 1987; Little & Rubin, 2002) – Varianz in den interessierenden Parametern zwischen den m Durchläufen. Diese wird in die Schätzung der Standardfehler der Parameter aufgenommen und reflektiert die durch das Matching emergierenden Unsicherheiten. In Anlehnung an Rubin (1987; siehe auch Enders, 2010) werden die geschätzten Parameter über die $t = 1, \dots, m$ Durchläufe gemittelt ($\bar{\theta} =$

¹⁰ Wird zum Beispiel für eine NMS auf Basis des Logit-Modells aus (6) eine Wahrscheinlichkeit der Zugehörigkeit zur Gruppe der NMS von 40 Prozent prognostiziert, muss die geschätzte Wahrscheinlichkeit einer AHS-Unterstufe im Intervall (35, 45) Prozent liegen, um als statistischer Zwilling in Frage zu kommen.

¹¹ Zur exakten Reproduzierbarkeit der Ergebnisse der PSM wurde für jeden der zehn Durchläufe ein Seed vergeben. Um zu gewährleisten, dass die gewählten Seeds keiner Symptomatik folgen, wurden zehn Zufallszahlen über den frei verfügbaren Random Number Generator auf <http://www.random.org> (min: 1; max: 10.000.000) erzeugt.

$1/m \sum_{t=1}^m \hat{\theta}_t$). Die Gesamtvarianz (V_T) der Parameter wird durch Summenbildung der Within- (V_W) und der Between-Varianz (V_B) errechnet (SE_t^2 bezeichnet den quadrierten Standardfehler des Parameters in Durchlauf t):

$$V_T = V_W + V_B + \frac{V_B}{m} \quad \text{mit} \quad V_W = \frac{1}{m} \sum_{t=1}^m SE_t^2 \quad \text{und} \quad V_B = \frac{1}{m-1} \sum_{t=1}^m (\hat{\theta}_t - \bar{\theta})^2 \quad (9)$$

Da in den BIST-M8 Daten weniger AHS-Unterstufen als NMS enthalten sind, fiel die Entscheidung auf das Prinzip des Ziehens mit Zurücklegen. Schulen aus den beiden Vergleichsgruppen können somit mehreren NMS als Matchingpartner zur Verfügung stehen. Für die Umsetzung des PSM wurde das Stata-Modul `psmatch2` (Guo & Fraser, 2010; Nichols, 2007) eingesetzt.

3.4. Datenanalyse

Zur Beantwortung der Forschungsfragen (I) und (II) wird als zentrale statistische Kenngröße die Mittelwertdifferenz in den auf Schullevel aggregierten Testleistungen der allgemeinen Mathematikkompetenz (WLE_{Mk}) und der Lesekompetenz in Englisch (WLE_{Ek}) aus den beiden Bildungsstandardtestungen BIST-M8 und BIST-E8 zwischen den NMS und den AHS-Unterstufen bzw. den HS herangezogen. In diesem Zusammenhang wird vornehmlich auf den fairen Vergleich, d.h. auf die Berechnung der Mittelwertdifferenzen zwischen den NMS und den beiden gemachten Vergleichsgruppen abgestellt. In der kontrafaktischen Literatur (siehe etwa Morgan & Winship, 2007) werden diese Differenzen auch als Average Treatment Effects on the Treated (ATT) bezeichnet (zu den Details siehe etwa Schafer & Kang, 2008, S. 282). Die Effektschätzung (inklusive Standardfehler) erfolgt ebenfalls über das Stata-Modul `psmatch2`. Für den graphischen Abgleich der Verteilungen der mittleren Testleistungen der Schulen in den beiden Kompetenzbereichen zwischen der NMS und der AHS-Unterstufe bzw. der HS vor und nach dem Matching wird auf Kerndichteplots zurückgegriffen. Diese wurden mittels dem `kdensity`-Befehl in Stata (Kernel Funktion nach Epanechnikov (1969) mit fixer Bandbreite) erzeugt.

Für die Beantwortung von Forschungsfrage (III), wodurch sich NMS mit einem geringen Leistungsabstand zur AHS-Unterstufe von jenen mit einem großen Leistungsabstand unterscheiden, gilt es zunächst die beiden Extremgruppen zu identifizieren. Zu diesem Zweck wird zunächst für jedes gematchte Schulpaar mit $\kappa \in NMS$ und $\kappa^* \in AHS$ die absolute Differenz der in den Schulen gemittelten WLE -Scores im getesteten Kompetenzbereich c mit $c = \{E, M\}$ berechnet:

$$\Delta_{WLE_{c\kappa}}^+ = |WLE_{c\kappa} - WLE_{c\kappa^*}| \quad (10)$$

In einem nächsten Schritt werden die NMS aufsteigend nach den ermittelten absoluten Leistungsdifferenzen in eine Rangreihe gebracht und das erste ($Q_{0,2}$) sowie das vierte Quintil ($Q_{0,8}$) bestimmt. Jene 20 Prozent aller NMS, die unterhalb des ersten Quintils liegen ($\Delta_{WLE_{c\kappa}}^+ < Q_{0,2}$) und somit den im Betrag geringsten Leistungsabstand zur gematchten AHS-Unterstufe aufweisen, werden der Gruppe der „leistungsnahen“ NMS¹² zugeord-

¹² In der Folge wird das Kunstwort „leistungsnah“ als abgekürzte Bezeichnung für einen geringen relativen Leistungsabstand zur AHS-Unterstufe verwendet. Analog dazu steht „leistungsfern“ für einen großen relativen Leistungsabstand zur AHS-Unterstufe.

net. Demgegenüber werden jene 20 Prozent aller NMS, deren Leistungsdifferenz im Betrag das vierte Quintil überschreitet ($\Delta_{WLE_{ck}}^+ > Q_{0,8}$) der Gruppe der „leistungsfernen“ NMS zugewiesen.

Im Anschluss an die Identifikation jener NMS, die in mindestens 70 Prozent der zehn PSM-Durchläufe (d.h. in zumindest sieben Durchläufen) konsistent einer der beiden leistungsbezogenen Extremgruppen angehören¹³, werden deren relevante Daten aus den Datensätzen der Schlusserhebungen von G1 (für die NMS auf Basis von BIST-M8) und G2 (für die NMS auf Basis von BIST-E8) gefiltert. Während in den G1-Daten aufgrund der realisierten Totalerhebung alle NMS zur Verfügung stehen, basiert die G2-Erhebung auf einer Stichprobe, so dass lediglich für einen Teil der NMS aus den beiden Extremgruppen Informationen zur Verfügung stehen (für Details zu den Stichprobenplänen der G1- und G2-Erhebungen siehe Eder & Svecnik, 2015). Letztlich konnte ein Datensatz mit insgesamt 60 NMS erzeugt werden, der sich gemäß Tabelle 1 zusammensetzt.

Tabelle 1: Realisierter Datensatz für den Extremgruppenvergleich der NMS mit geringer relativer Leistungsdifferenz zur AHS (Δ_{WLE}^+ gering) und jenen mit hoher relativer Leistungsdifferenz zur AHS (Δ_{WLE}^+ groß)

	G1 (BIST-M8)	G2 (BIST-E8)	gesamt
Δ_{WLE}^+ gering	14	17	31
Δ_{WLE}^+ groß	14	15	29
gesamt	28	32	60

Im Zuge der Suche nach Erklärungsfaktoren für die Zugehörigkeit zur Gruppe der leistungsnahen bzw. leistungsfernen NMS¹⁴ wird zunächst bivariat mittels t-Tests für unabhängige Stichproben geprüft, ob sich die Schulen aus den beiden Gruppen systematisch hinsichtlich der sozialen Zusammensetzung, der mittleren kognitiven Fähigkeitsniveaus der SchülerInnen, der Konzeptmerkmale der NMS (siehe Helm, Kemethofer, Altrichter & Weber, 2015), der Anomie im Unterricht, der abweichenden Verhaltensweisen der SchülerInnen (siehe Eder, 2015; Eder & Springer, 2015) sowie der räumlichen Distanz zur nächstgelegenen AHS-Unterstufe voneinander unterscheiden. Zur Absicherung der Befunde gegen Verstöße der Normalverteilungsannahme werden ebenfalls nichtparametrische Tests (Wilcoxon-Mann-Whitney-Test) gerechnet. Die folgende Tabelle 2 stellt einen Überblick über die Analysevariablen zur Verfügung.

¹³ Für eine Abschätzung der Strenge dieses Inklusionskriteriums sei auf die Diskussion in Campbell und Kenny (1999, S. 37ff.) verwiesen.

¹⁴ An dieser Stelle sei nochmals explizit darauf hingewiesen, dass es sich um relative und nicht um absolute Leistungsnähe zur AHS-Unterstufe handelt.

Tabelle 2: Untersuchte Merkmale zur Identifikation von (Miss-)Erfolgsfaktoren für NMS in Bezug auf die Leistungsnähe zur AHS-Unterstufe

Merkmalsgruppe	Einzelmerkmale^(a)
soziale Zusammensetzung der Schule	Sozialindex (<i>S</i>)
kognitives Fähigkeitsniveau der Schule (Oberwimmer & Eder, 2015)	kognitiver Fähigkeitstest (KFT) in der 5. und 8. Schulstufe (verbal & nonverbal)
Konzeptmerkmale der NMS (Helm et al., 2015)	<i>Schulleitungen</i> : individueller Förderplan, Anteil an Oberstufenlehrkräften, Nachmittagsbetreuung <i>Klassenvorstände</i> : verschränktes Teamteaching, SchülerInnenorientierung, Möglichkeiten zum selbstregulierten Lernen
Unterrichtsmerkmale (Eder, 2015b)	Anomie im Unterricht (Fehlen von Struktur und Ordnung)
schul- und lernbezogenes Verhalten in der Schule (Eder & Oberwimmer, 2015)	abweichende Verhaltensweisen in der Schule, Gewalthandeln in der Schule, Opfererlebnisse in der Schule
räumliche Distanz zur nächstgelegenen AHS-Unterstufe	Distanz in km
Anteil an NMS-Modellklassen in der Schule	Anteil der NMS-Modellklassen an allen 4. Klassen der Schule

^(a) Zu Informationen über die Operationalisierung, Skalierung und (falls nötig) Aggregation der NMS-Konzeptmerkmale, Unterrichtsmerkmale sowie der Merkmale zum schul- und lernbezogenen Verhalten in der Schule sei auf die jeweiligen Beiträge in diesem Band verwiesen.

Gemäß den Befunden von Baumert et al. (2006b; siehe Abschnitt 2) ist zu erwarten, dass sich die Gruppe der NMS mit einem geringen relativen Leistungsabstand zur AHS-Unterstufe von der Gruppe der leistungsfernen NMS unterscheidet durch (i) eine ausgewogenere soziale Zusammensetzung SchülerInnenschaft (eine geringere Akkumulation von SchülerInnen aus sozial stark benachteiligten Herkunftsfamilien), (ii) ein höheres mittleres kognitives Fähigkeitsniveau (verbal & nonverbal) der SchülerInnen, (iii) ein geringeres Ausmaß an Anomie im Unterricht sowie (iv) weniger stark ausgeprägte abweichende Verhaltensweisen und ein geringeres Ausmaß an Gewalthandlungen durch die SchülerInnen. Darüber hinaus sind Unterschiede in der Implementierung der NMS-Konzeptmerkmale zu erwarten. So sollten sich individuelle Förderpläne für SchülerInnen, ein verschränktes Teamteaching, eine starke SchülerInnenorientierung in Form einer breiten aktiven Einbindung in den Unterricht, weitreichende Möglichkeiten für SchülerInnen zum selbstregulierten Lernen durch große Entscheidungsspielräume bezüglich der Unterrichtsgestaltung, das Angebot einer Nachmittagsbetreuung und ein gehobener Anteil an Oberstufenlehrkräften (Lehrkräfte aus maturaführenden Schulen in der Sekundarstufe II; z.B. AHS-Oberstufe, BHS), um der intendierten Durchmischung des Lehrkörpers zu entsprechen, zumindest in indirekter Weise positiv auf die Kompetenzvermittlungs- und Wissensaneignungsprozesse in den NMS auswirken und folglich zu geringen Leistungsunterschieden zur AHS-Unterstufe führen.

Die stärksten Effekte sind für die Anomie im Unterricht sowie die gewaltbezogenen Merkmale und das abweichende Verhalten der SchülerInnen in der Schule zu erwarten, da

diese Faktoren gemäß den Befunden von Baumert et al. (2006b) und in Übereinstimmung mit Befunden aus der empirischen Unterrichtsforschung (z.B. Helmke, 2007; 2012; Helmke & Schrader, 2008) eine direkte Wirkung auf die schulischen Leistungen entfalten sollten, während die soziale Komposition, das mittlere Fähigkeitsniveau der SchülerInnen und der Umsetzungsgrad der NMS-Konzeptmerkmale lediglich vermittelt über die Verhaltensweisen der SchülerInnen wirken sollten. Folglich ist auch nur eine schwache Wirkung durch die von Eder (2015b) gebildeten Implementierungscluster zu erwarten, da sie ebenfalls lediglich einen indirekten Einfluss ausüben.

Um die empirische Bedeutsamkeit der einzelnen Faktoren im multivariaten Fall evaluieren zu können, werden im Anschluss an die bivariaten Befunde ökologische Logit-Regressionsmodelle spezifiziert. Die abhängige Gruppenvariable wird so kodiert, dass die Wahrscheinlichkeit der Zugehörigkeit zu den NMS mit einem geringen relativen Leistungsabstand zur AHS-Unterstufe geschätzt wird (0 = leistungsferne NMS; 1 = leistungsnahe NMS). Da die Maximum Likelihood Schätzung (MLE) lediglich asymptotisch unverzerrt ist (Cordeiro & McCullagh, 1991; Firth, 1993; McCullagh & Nelder, 1989) und bei kleinen Stichproben einen erheblichen Finite Sample Bias in den geschätzten Parametern aufweisen kann, werden – aufgrund der gegebenen kleinen Fallzahl von 60 Schulen (siehe Tabelle 1) – die Befunde noch durch die Penalized Maximum Likelihood Schätzung (PMLE) abgesichert. Die PMLE stellt eine Erweiterung der ML-Funktion um einen Penalisierungsterm dar, der als Jeffreys Invariant Prior (Jeffreys, 1946) bekannt ist und die Logit-Koeffizienten „shrinks“, um deren im Betrag systematische Überschätzung im Zuge der MLE zu kompensieren (Firth, 1993). Der Shrinkage-Effekt ist umso ausgeprägter, je niedriger die Fallzahl und/oder je niedriger die Anzahl an Fällen mit einem interessierenden Ereignis sind (Rare Events Problem; siehe dazu ausführlich King & Zeng, 2001). Leitgöb (2013) konnte im Rahmen von Monte Carlo Simulationen nachweisen, dass die PMLE selbst bei sehr niedrigen Fallzahlen und äußerst seltenen Ereignissen unverzerrte Schätzer der Logit-Koeffizienten produziert. Die Logit-Modelle basierend auf der PMLE werden mit dem Stata-Modul `firthlogit` geschätzt. Die Ergebnisse werden aus platzökonomischen Gründen im Anhang (Tabelle A2) verortet.

Zur Identifikation potentieller Scheineffekte wird eine sequentielle bzw. hierarchische Modellierungsstrategie (siehe etwa Cohen, Cohen, West & Aiken, 2003, S. 158ff) gewählt, der die Spezifikation von vier genesteten Modellen zugrunde liegt. In einem ersten Schritt (Modell 1) wird lediglich die soziale Zusammensetzung der SchülerInnenschaft über die Modellierung des Sozialindex (*S*) berücksichtigt. In Modell 2 erfolgt eine Erweiterung um die kognitiven Fähigkeiten zu Schuleintritt (5. Schulstufe). Die Integration ausgewählter Konzeptmerkmale der NMS führt zu Modell 3. Neben diesen Faktoren, denen lediglich eine indirekte Wirkung zugeschrieben wird, komplettieren schließlich die Anomie im Unterricht sowie ausgewählte Merkmale hinsichtlich des abweichenden bzw. gewalttätigen Verhaltens in der Schule das Modell (Modell 4). Als Kontrollvariable wird in allen vier Modellen der Generationsindikator (0 = G1 bzw. identifiziert mit den BIST-M8 Daten; 1 = G2 bzw. identifiziert mit den BIST-E8 Daten; entspricht der Spaltenvariable in Tabelle 1) mitberücksichtigt. Die Selektion von NMS-Konzeptmerkmalen sowie von Verhaltensmerkmalen der SchülerInnen erfolgt auf Basis der bivariaten Befunde (jene mit den stärksten bivariaten Effekten werden für die multivariaten Analysen ausgewählt) und ist erforderlich, um die Anzahl der unabhängigen Variablen bei gegebener niedriger Fallzahl und niedriger Anzahl an Fällen in der interessierenden Gruppe der leistungsnahe NMS gering zu halten. Während Peduzzi, Concato, Kemper, Holford & Feinstein (1996) auf Basis von Monte Carlo Simulationen eine Unterschreitung von zehn „Ereignissen“ (im vorliegenden

Fall NMS in der leistungsnahen Gruppe) pro unabhängiger Variable (EPV) für bedenklich halten, erachten Vittinghoff und McCulloch (2006) den Wert als zu konservativ und identifizieren die kritische Grenze bei fünf EVP. Das Gesamtmodell (Modell 4) liegt mit sieben Kovariaten bei 31 NMS mit geringem Leistungsabstand (siehe Tabelle 1) und einem daraus resultierenden Verhältnis von 4,43 nur marginal unter dieser „Daumenregel“ von fünf EPV.

Um die Vergleichbarkeit der Logit-Koeffizienten über die genesteten Modelle mit divergierenden Residualvarianzen zu gewährleisten (zur Erläuterung dieser Problematik siehe Allison, 1999; Best & Wolf, 2012; Karlson, Holm & Breen, 2012; Cramer, 2003; Mood, 2010), wurde das von Karlson et al. (2012) vorgeschlagene Verfahren der Reskalierung der Logit-Koeffizienten zur Anwendung gebracht. Die Metrik der Koeffizienten des Gesamtmodells (Modell 4) dient als Kalibrierungsbasis für die Koeffizienten der Modelle 1 bis 3. Für die Schätzung der reskalierten Logit-Koeffizienten wird auf das *Stata*-Modul *khb* von Kohler, Karlson und Holm (2011) rekurriert. Die Ergebnisse werden in Tabelle A3 im Anhang abgebildet.

4. Ergebnisse

4.1. Leistungsunterschiede zur AHS-Unterstufe

Wie aufgrund der Befunde von Bruneforth et al. (2012) zu erwarten war, unterscheiden sich die NMS hinsichtlich der sozialen Zusammensetzung der SchülerInnenchaft in erheblichem Maße von den AHS-Unterstufen. Dies wird deutlich durch den in Tabelle 3 präsentierten Vergleich der Verteilungen der vier Einzelindikatoren sowie des kategorisierten Sozialindex zwischen NMS und AHS-Unterstufe auf Basis der auf Schulebene aggregierten BIST-M8 und BIST-E8 Daten zum Ausdruck gebracht. In allen den Sozialindex konstituierenden Einzelindikatoren hebt sich die NMS deutlich von der AHS-Unterstufe in der Akkumulation sozial benachteiligter SchülerInnen ab. Folglich unterscheiden sich auch die Verteilungen des Sozialindex nach den beiden Schultypen in hohem Maße voneinander. Während die überwiegende Mehrheit der AHS-Unterstufen (BIST-M8: 92,5%; BIST-E8: 84,8%) in die Kategorie mit einem gering ausgeprägten Sozialindex (bis max. 115 Punkte), fällt und somit durch eine sozial privilegierte SchülerInnenchaft charakterisiert sind, trifft dies für weniger als ein Viertel aller NMS (BIST-M8: 22,4%; BIST-E8: 22,0%) zu. Demgegenüber befinden sich in der Kategorie der sehr hoch belasteten Schulen (mehr als 135 Punkte) mehr als ein Fünftel aller NMS (BIST-M8: 26,9%; BIST-E8: 20,7%) und so gut wie keine AHS-Unterstufen (BIST-M8: 0,0%; BIST-E8: 1,9%). Sämtliche Mittelwertdifferenzen in den Einzelindikatoren bzw. Verteilungsunterschiede sind statistisch hoch signifikant ($p < 0,001$).

Aus den Befunden kann geschlossen werden, dass eine stark sozialschichtdeterminierte Selektion der SchülerInnen in die beiden Schulformen gegeben ist (für die zugrunde liegenden Mechanismen siehe Abschnitt 2) und die intendierte soziale Durchmischung der SchülerInnen in der NMS nach dem Prinzip einer gemeinsamen Schule aller 10- bis 14-Jährigen bislang nicht gelungen ist. Vielmehr scheinen die NMS der Generationen 1 und 2 durchschnittlich eine noch marginal ungünstigere soziale Zusammensetzung aufzuweisen als die HS, wie es in Abschnitt 4.2 zu zeigen gilt.

Tabelle 3: Soziale Zusammensetzung der SchülerInnenschaft von NMS und AHS-Unterstufe im Vergleich

BIST-M8				
Merkmal	NMS (n=67)	AHS (n=189)	Differenz (in PP) ^(a)	p-Wert
geringe schulische Bildung der Eltern (\bar{x}_1)	18,4%	4,1%	14,3	<0,001
geringer beruflicher Status der Eltern (\bar{x}_2)	43,4%	14,0%	29,4	<0,001
Migrationshintergrund (\bar{x}_3)	22,2%	9,0%	13,2	<0,001
andere zu Hause gesprochene Sprache (\bar{x}_4)	22,3%	7,3%	15,0	<0,001
Sozialindex (S)				
gering (100-115)	22,4%	92,5%	-70,1	
mittel (>115-125)	35,8%	5,9%	29,9	
hoch (>125-135)	14,9%	1,6%	13,3	
sehr hoch (>135)	26,9%	0,0%	26,9	<0,001
gesamt	100,0%	100,0%		
Mittelwert	126,6	108,6	18	<0,001
BIST-E8				
Merkmal	NMS (n=227)	AHS (n=211)	Differenz (in PP) ^(a)	p-Wert
geringe schulische Bildung der Eltern (\bar{x}_1)	16,3%	4,3%	12,0	<0,001
geringer beruflicher Status der Eltern (\bar{x}_2)	41,8%	15,7%	26,1	<0,001
Migrationshintergrund (\bar{x}_3)	22,3%	11,5%	10,8	<0,001
andere zu Hause gesprochene Sprache (\bar{x}_4)	22,0%	9,8%	12,2	<0,001
Sozialindex (S)				
gering (100-115)	22,0%	84,8%	-62,8	
mittel (>115-125)	39,7%	10,9%	28,8	
hoch (>125-135)	17,6%	2,4%	15,2	
sehr hoch (>135)	20,7%	1,9%	18,8	<0,001
gesamt	100,0%	100,0%		
Mittelwert	125,6	110,3	16	<0,001

^(a) PP ... Prozentpunkte; Ausnahme: Mittelwert des Sozialindex

Aufgrund der ausgeprägten Unterschiede in der sozialen Zusammensetzung der SchülerInnenschaft zwischen den NMS und den AHS-Unterstufen war es bei Verwendung der

vier Einzelindikatoren unter Festlegung des Calipers auf $\varepsilon = 0,05$ im Zuge der Matching-prozedur nicht möglich, eine angemessene Anzahl an statistischen Zwillingen aus den AHS-Unterstufen für die zur Verfügung stehenden NMS zu identifizieren.¹⁵ Eine drastische Erhöhung des Schwellenwertes auf 0,2 (BIST-M8) bzw. 0,1 (BIST-E8; siehe Tabelle 4) resultierte zwar in der Zuweisung eines statistischen Zwillinges zu jeder NMS, allerdings wurde eine äußerst geringe Anzahl von unterschiedlichen AHS-Unterstufen immer wieder gezogen und den NMS zugeordnet. In den BIST-M8 Daten wurde auf Basis des PSM eine AHS-Unterstufe insgesamt 59 der bestehenden 67 NMS (88,1%) zugewiesen, in den BIST-E8 Daten wurde die am häufigsten als statistischer Zwilling identifizierte AHS-Unterstufe immerhin 119 der 227 NMS (52,4%) zugeteilt. Die daraus resultierenden Konsequenzen liegen auf der Hand:

(i) Durch die Anwendung des PSM konnte keine erfolgreiche Balancierung der beiden Gruppen von Schulformen in den vier sozialstrukturellen Variablen erzielt werden. Der Betrag des standardisierten Bias – definiert als die prozentuelle Differenz der Stichprobenmittelwerte der NMS und der AHS-Unterstufen, relativiert an der Wurzel der mittleren Stichprobenvarianz aus den beiden Gruppen (zu den Details siehe Rosenbaum & Rubin, 1985; Berechnung mit dem Stata-Modul `pstest` nach der Ausführung von `pmatch2`) – ist insbesondere bei den migrationsbedingten Indikatoren als bemerkenswert zu bezeichnen (Migrationshintergrund (\bar{x}_3): BIST-M8: 76,0%; BIST-E8: 134,3%; nichtdeutsche Alltagssprache (\bar{x}_4): BIST-M8: 92,8%; BIST-E8: 85,7%). Interessant ist zudem die Tatsache, dass durch das PSM in der BIST-E8 Erhebung mit 227 NMS eine Balancierung der beiden schultypenbezogenen Gruppen hinsichtlich des elterlichen Bildungsindikators (\bar{x}_1) realisiert werden konnte, während in den BIST-M8 Daten mit lediglich 67 NMS ein beträchtlicher absoluter standardisierter Bias von 164,6 Prozent beobachtet werden kann.

(ii) Weiterhin stellt die Referenzgruppe der gematchten AHS-Unterstufen durch die exzessive Ziehung einer bzw. einer äußerst geringen Anzahl an Schulen eine in jeder Hinsicht außerordentlich homogene Gruppe mit einer artifiziell geringen Streuung des mittleren Kompetenzniveaus dar (BIST-M8: $s = 14,50$; BIST-E8: $s = 22,79$).¹⁶ Es kann nicht ausgeschlossen werden, dass diese AHS-Unterstufen zwar in einer oder zwei der sozialen Indikatorvariablen den NMS verhältnismäßig ähnlich sind, die SchülerInnen dieser Schulen aber dennoch sozial äußerst privilegiert sind und entsprechend hohe Mathematik- bzw. Lesekompetenzen in Englisch besitzen (z.B. AHS-Unterstufen mit einem internationalen Zweig und einem hohen Anteil von SchülerInnen mit Eltern aus dem diplomatischen Dienst oder internationalen Management).

Die Ausführungen unterstreichen, dass die soziale Zusammensetzung der AHS-Unterstufen nicht mit jener der NMS vergleichbar ist und das PSM auf Basis der vier Einzelindikatoren nicht zum intendierten Ergebnis einer sozialen Balancierung der beiden Gruppen von Schulen führt. Folglich muss die strenge Forderung einer Balancierung in allen vier Sozialindikatoren verworfen werden. Alternativ wird nun auf den Sozialindex zur Schätzung der Propensity Scores abgestellt. Aufgrund seiner additiven Natur (siehe Abschnitt

¹⁵ Es wurden mehrere Durchläufe mit unterschiedlichen Zufallsanordnungen der Schulen realisiert, die alle nahezu identische Ergebnisse hervorbrachten. Aus diesem Grund wird die Ergebnisdarstellung auf den ersten Durchlauf beschränkt.

¹⁶ Im Vergleich dazu weist die Gesamtpopulation aller AHS-Unterstufen deutlich höhere Streuungen auf (BIST-M8: $s = 29,65$; BIST-E8: $s = 26,31$).

3.3) liegt dem Sozialindex die implizite Annahme zugrunde, dass die vier den Index konstituierenden Sozialindikatoren weitgehend unabhängig voneinander (siehe Schnell, Hill & Esser, 2005, S. 172) bzw. mit gleichem Gewicht auf die Zieldimension der sozialen Benachteiligung einer Schule durch die Komposition der SchülerInnen wirken. Dies impliziert, dass zwei Schulen den gleichen Wert auf dem Sozialindex aufweisen können, die Belastung in den Einzelindikatoren allerdings erheblich zwischen den Schulen variieren kann. So ist es etwa möglich, dass eine NMS mit einem hohen Anteil an SchülerInnen mit Migrationshintergrund und nichtdeutscher Alltagssprache mit einer AHS-Unterstufe mit einem gehobenen Anteil an SchülerInnen aus bildungsfernen Herkunftsfamilien gemacht wird. Dies bedeutet, dass lediglich eine Balancierung der beiden Gruppen der NMS und der AHS-Unterstufen nach dem Sozialindex und nicht nach den vier Einzelindikatoren erwartet werden darf. Aus den bereits genannten Gründen erscheint ein exakteres Matching allerdings nicht realisierbar.

Die Auseinandersetzung mit dieser Komparativitätsproblematik zeigt nochmals in aller Deutlichkeit die massiven Schwächen „naiver“ (im Sinne von einfacher bzw. unbedingter) und – daraus folgend – die Notwendigkeit „fairer“ Leistungsvergleiche zwischen verschiedenen Schultypen auf, um systematische Unterschiede im mittleren schulischen Kompetenzniveau ursächlich den Lernvermittlungsprozessen und den darauf wirkenden internen Strukturen und Prozessen in den einzelnen Schulformen zuschreiben zu können. Theoretisch könnten beobachtbare Leistungsunterschiede nämlich ausschließlich durch die in Abschnitt 2 beschriebenen sozialen Selektions- und Kompositionseffekte zustande kommen. Einschränkend gilt es zu erwähnen, dass allerdings auch nicht ausgeschlossen werden kann, dass externe (nicht für das Matching verfügbare bzw. unberücksichtigte) Faktoren die Ursache für die Leistungsunterschiede zwischen den Schultypen darstellen. Mit Vorgriff auf die Ergebnisse aus Abschnitt 4.3 sind hier die kognitiven Fähigkeiten zum Eintritt in die Sekundarstufe I (5. Schulstufe) zu nennen. Da diese jedoch erheblich mit der sozialen Herkunft der SchülerInnen korrelieren (für die Korrelation auf der Aggregatenebene der Schulen siehe Tabelle A4) können systematische Unterschiede in den mittleren kognitiven Fähigkeitsniveaus zwischen den Schulformen partiell ausbalanciert werden.

Tabelle 4: Ergebnisse des PSM mit den vier Einzelindikatoren

	BIST-M8	BIST-E8
Empfängerdatei (NMS)	67	227
Spenderdatei (AHS-Unterstufe)	189	211
Anzahl der gefunden Paare	67	227
Anzahl der mehrfach verwendeten Schulen der Spenderdatei	2	14
Auftretenshäufigkeit der am häufigsten verwendeten Schule der Spenderdatei	59	119
minimale Distanz innerhalb der Paare	<0,001	<0,001
maximale Distanz innerhalb der Paare	0,19	0,07
festgelegter Caliper	0,2	0,1
Balancierung ¹⁷ nach		
geringe schulische Bildung der Eltern	nein	ja
geringer beruflicher Status der Eltern	nein	ja
Migrationshintergrund	nein	nein
andere zu Hause gesprochene Sprache	nein	nein
Sozialindex	nein	nein

Wird für die Generierung der Propensity Scores ausschließlich der Sozialindex verwendet, kann im Zuge des PSM unter Einhaltung des Schwellenwertes von $\varepsilon < 0,05$ für jede der 67 (BIST-M8) bzw. 277 (BIST-E8) NMS eine AHS-Unterstufe als statistischer Zwilling identifiziert werden (siehe Tabelle 5).¹⁸ Zudem konnte der Einfluss einzelner AHS-Unterstufen in der gematchten Vergleichsgruppe deutlich reduziert werden. So liegt die Anzahl der am häufigsten gezogenen AHS-Unterstufe nun bei 21 (BIST-M8) bzw. 17 (BIST-E8). Dies führte im Vergleich zur Matchingstrategie mit den vier Einzelindikatoren auch zu einem Varianzzuwachs im mittleren Kompetenzniveau der selektierten AHS-Unterstufen (BIST-M8: $s = 30,70$; BIST-E8: $s = 29,28$) und so zu einer Angleichung an die Streuungen in der Gesamtpopulation aller AHS-Unterstufen (siehe dazu Fußnote 16). Zudem konnte eine Balancierung nach dem Sozialindex, wie vermutet allerdings nicht nach den vier Einzelindikatoren (Ausnahme: nichtdeutsche Alltagssprache in den BIST-M8 Daten), erzielt werden.

Die hohe Erklärungskraft der beiden Logit-Modelle zur Schätzung der Propensity Scores (BIST-M8: $\text{Pseudo-}R^2 = 0,469$; BIST-E8: $\text{Pseudo-}R^2 = 0,342$; die Ergebnisse können Tabelle A1 aus dem Anhang entnommen werden) indiziert eine starke Abhängigkeit der ersten Bildungsentscheidung am Übergang von der Primarstufe in die Sekundarstufe I von

¹⁷ Die Balancierung der Samples in den Merkmalen wurde mittels t-Tests für unabhängige Stichproben sowie Produkt-Moment-Korrelationen geprüft.

¹⁸ Die Befunde in Tabelle 5 sowie in Tabelle 8 (für die HS) beziehen sich jeweils auf den ersten PSM-Durchlauf. Analog zu den Ausführungen in Fußnote 15 erwiesen sich auch hier die Ergebnisse über die insgesamt jeweils zehn Durchläufe als beachtenswert stabil (es liegt eine verhältnismäßig geringe Between-Varianz (V_B ; siehe (9)) vor), so dass die Darstellung der ersten Durchläufe ausreicht, um einen sehr guten Gesamtüberblick zu gewähren.

der sozialen Herkunft der SchülerInnen.¹⁹ Wie bereits kurz erläutert ist es in der Implementierungsphase der NMS als Schulversuch offensichtlich (noch) nicht in ausreichendem Maße gelungen, diesen Schultyp für Herkunftsfamilien mit einem privilegierten sozialen Status als attraktive Alternative zur AHS-Unterstufe präsentieren.

Tabelle 5: Ergebnisse des PSM mit dem Sozialindex

	BIST-M8	BIST-E8
Empfängerdatei (NMS)	67	227
Spenderdatei (AHS-Unterstufe)	189	211
Anzahl der gefunden Paare	67	227
Anzahl der mehrfach verwendeten Schulen der Spenderdatei	12	36
Auftretenshäufigkeit der am häufigsten verwendeten Schule der Spenderdatei	21	17
minimale Distanz innerhalb der Paare	0	0
maximale Distanz innerhalb der Paare	0,03	0,03
festgelegter Caliper	0,05	0,05
Balancierung nach		
geringe schulische Bildung der Eltern	nein	nein
geringer beruflicher Status der Eltern	nein	nein
Migrationshintergrund	nein	nein
andere zu Hause gesprochene Sprache	ja	nein
Sozialindex	ja	ja

Lenkt man den Blick auf die Ergebnisse zur Klärung der ersten Forschungsfrage – dem Leistungsvergleich zwischen der NMS und der AHS-Unterstufe –, so kann ein starkes Gefälle im mittleren Kompetenzniveau von der AHS-Unterstufe zur NMS beobachtet werden (siehe Tabelle 6). Während im „naiven“ Vergleich der beiden Schultypen ohne Berücksichtigung der sozialen Zusammensetzung der SchülerInnenschaft die AHS-Unterstufen jeweils deutlich über 100 Punkte vor den NMS liegen (BIST-M8: 110,4; BIST-E8: 119,6; es werden die Beträge der Differenzen ausgewiesen), reduziert sich der Betrag der Differenz nach dem PSM (entspricht dem ATT; siehe Abschnitt 3.4) als Effektmaß des „fairen“ Vergleichs um 35,4 Punkte (BIST-M8) bzw. 15 Punkte (BIST-E8) auf 75,0 Punkte bzw. 104,6 Punkte. Somit können von der Gesamtdifferenz der mittleren Fähigkeitsniveaus zwischen den beiden Schulformen 32,1 Prozent (BIST-M8) bzw. lediglich 12,5 Prozent (BIST-E8) auf systematische Divergenzen in der sozialen Komposition der SchülerInnenpopulation zwischen NMS und AHS-Unterstufe zurückgeführt werden. Da sämtliche Effekte hoch signifikant sind ($p < 0,001$), lässt sich Forschungsfrage (I) eindeutig beantworten: Auch unter weitgehender Kontrolle der sozialer Zusammensetzung bestehen beträchtliche Unterschiede in den Mathematik- und Lesekompetenzen in Englisch zwischen SchülerInnen aus

¹⁹ Diese Befunde werden durch Analysen auf der Basis von Individualdaten abgesichert (siehe etwa Bacher, 2009; Bruneforth et al., 2012; Kleine, Paulus & Blossfeld, 2009; Maaz & Nagy, 2009). Die Gefahr eines ökologischen Fehlschlusses ist somit nicht gegeben. Auf nähere Ausführungen wird an dieser Stelle jedoch verzichtet, da die Befunde lediglich ein Nebenprodukt der Matchingprozedur darstellen und nicht im Zentrum des Erkenntnisinteresses stehen.

der NMS und aus der AHS-Unterstufe. Um die Bedeutsamkeit dieser Unterschiede aufzuzeigen, sei zunächst drauf hingewiesen, dass die Differenz der schultypenbezogenen Mittelwerte von 75,0 Punkten bei den Mathematikkompetenzen beinahe exakt dem definierten Punkteintervall der Kompetenzstufe 1 bei BIST-M8 (440 bis 517 Punkte; Intervallbreite: 77 Punkte; siehe Schreiner & Breit, 2012, S. 21) entspricht. Ferner beträgt die Differenz von 104,6 Punkten in den englischen Lesekompetenzen beinahe zwei Drittel (62,3%) des GERS-Referenzniveaus A2²⁰ im Kompetenzbereich Lesen (406 bis 574 Punkte; Intervallbreite: 168 Punkte; siehe Schreiner und Breit, 2013, S. 27). Das verbleibende Leistungsgefälle muss somit als beträchtlich bezeichnet werden.

Tabelle 6: Testleistungen der NMS und AHS-Unterstufe im Vergleich

	NMS	AHS	Effekt	p
BIST-M8				
vor Matching („naiv“)	496,7	607,1	-110,4	<0,001
nach Matching („fair“)	496,7	571,7	-75,0	<0,001
BIST-E8				
vor Matching („naiv“)	484,2	603,8	-119,6	<0,001
nach Matching („fair“)	484,2	588,8	-104,6	<0,001

Um einen Einblick in die Verteilungen der mittleren Mathematik- und Lesekompetenzen in Englisch beider Schultypen vor und nach dem PSM zu erhalten, werden Kerndichteplots zur Verfügung gestellt (siehe Abbildungen 2a und 2b). Wie für Mathematik aus Abbildung 2a hervorgeht, besteht für die Verteilungen der NMS und der AHS-Unterstufen lediglich ein kleiner gemeinsam geteilter Bereich, der von etwa 500 Punkten (unteres Ende der Verteilung der AHS-Unterstufen) bis zu knapp 590 Punkten (oberes Ende der Verteilung der NMS) reicht. Während der Schnittpunkt der beiden Verteilungen vor dem Matching allerdings bei deutlich über 550 Punkten liegt, ist dieser nach dem Matching bei etwa bereits 540 Punkten zu verorten. Dies impliziert, dass nach dem Matching ein merklich höherer Anteil an AHS-Unterstufen im oberen mathematischen Kompetenzbereich der NMS zu finden ist, da einerseits die Masse der Verteilung der AHS-Unterstufen moderat nach links in einen niedrigeren Kompetenzbereich verschoben wurde und die Verteilung andererseits an Heterogenität eingebüßt hat. Die Verteilung nimmt nach dem Matching zudem eine bimodale Form an. Dies ist auf eine AHS-Unterstufe zurückzuführen, die insgesamt 21 Mal als statistischer Zwilling identifiziert wurde und für den zweiten (rechten) Peak verantwortlich ist. Weiterhin bleibt festzuhalten, dass der empirisch beobachtbare Leistungsbereich, in dem sich die NMS bewegen (von unter 400 Punkten bis zu etwa

²⁰ GERS steht für Gemeinsamer Europäischer Referenzrahmen für Sprachen. Zu den Details siehe Horak, Moser, Nezbeda und Schober (2010). Das GERS-Referenzniveau A2 für den Kompetenzbereich Lesen ist wie folgt definiert:

„SchülerInnen auf dem Niveau A2 können kurze, einfache persönliche Briefe, Karten oder E-Mails verstehen. Sie können vertrauten Alltagstexten (z.B. Schildern und Broschüren – auch aus dem Internet) die wichtigsten Informationen entnehmen und können einfache, alltägliche Anleitungen und Vorschriften verstehen. SchülerInnen können die Kompetenzen *globales Verstehen*, *selektives Verstehen* und *detailliertes Verstehen* nachweisen. Die SchülerInnen können außerdem die logische Organisation eines schriftlichen Textes und lexikalische Zusammenhänge einigermaßen verstehen.“ (Schreiner & Breit, 2013, S. 27).

580 Punkten), insbesondere nach dem Matching deutlich größer ist als jener der AHS-Unterstufen. Dies kann durch einen Vergleich der Streuungen belegt werden (NMS: $s = 46,19$; AHS-Unterstufe vor dem Matching: $s = 29,65$; AHS-Unterstufe nach dem Matching: $s = 30,70$).

Die Verteilungen der Lesekompetenzen in Englisch der beiden Schulformen (siehe Abbildung 2b) weisen vor dem Matching eine hohe Ähnlichkeit zu jenen der Mathematikkompetenzen auf. Wiederum decken sich die Kompetenzniveaus der absolut besten NMS mit jenen der AHS-Unterstufen im untersten Bereich der Verteilung (Schnittpunkt der Verteilungen bei knapp unter 550 Punkten). Im Vergleich zu den Mathematikkompetenzen fällt auf, dass die Verteilung der NMS trotz eines ähnlich ausgeprägten Ranges (von unter 400 Punkten bis zu knapp unter 580 Punkten) eine wesentlich geringere Streuung aufweist ($s = 34,46$; $|\Delta s| = 11,73$; $|\Delta s|$ steht hier für die absolute Differenz der Standardabweichungen der Mathematik- und Lesekompetenzen in Englisch in der NMS). Worauf diese stärkere Homogenität der mittleren Kompetenzniveaus in Englisch innerhalb des Schultyps der NMS letztlich zurückzuführen ist, kann auf Basis der vorliegenden Daten nicht geklärt werden. Weiterhin sei darauf hingewiesen, dass dieser erhebliche Streuungsunterschied für die AHS-Unterstufe nicht beobachtet werden kann (AHS-Unterstufe vor dem Matching: $s = 26,31$; $|\Delta s| = 3,34$; $|\Delta s|$ steht hier für die absolute Differenz der Standardabweichungen der Mathematik- und Lesekompetenzen in Englisch in der AHS-Unterstufe), d.h., dass die Schulen der AHS-Unterstufe durch eine höhere Homogenität in den Leistungsniveaus über die Kompetenzbereiche hinweg gekennzeichnet sind.

Nach dem Matching verschiebt sich der Schnittpunkt der beiden Verteilungen analog zu den Mathematikkompetenzen moderat nach links. Der Verteilung der AHS-Unterstufen verbleibt in diesem Fall, abgesehen von einem durch eine mehrmals gematchte Schule hervorgerufenen lokalen Maximum bei etwas mehr als 650 Punkten, unimodal.

Abbildung 2a: Gegenüberstellung der Leistungsverteilungen der NMS und der AHS-Unterstufe vor und nach dem Matching für BIST-M8 (PSM-Durchlauf 1)

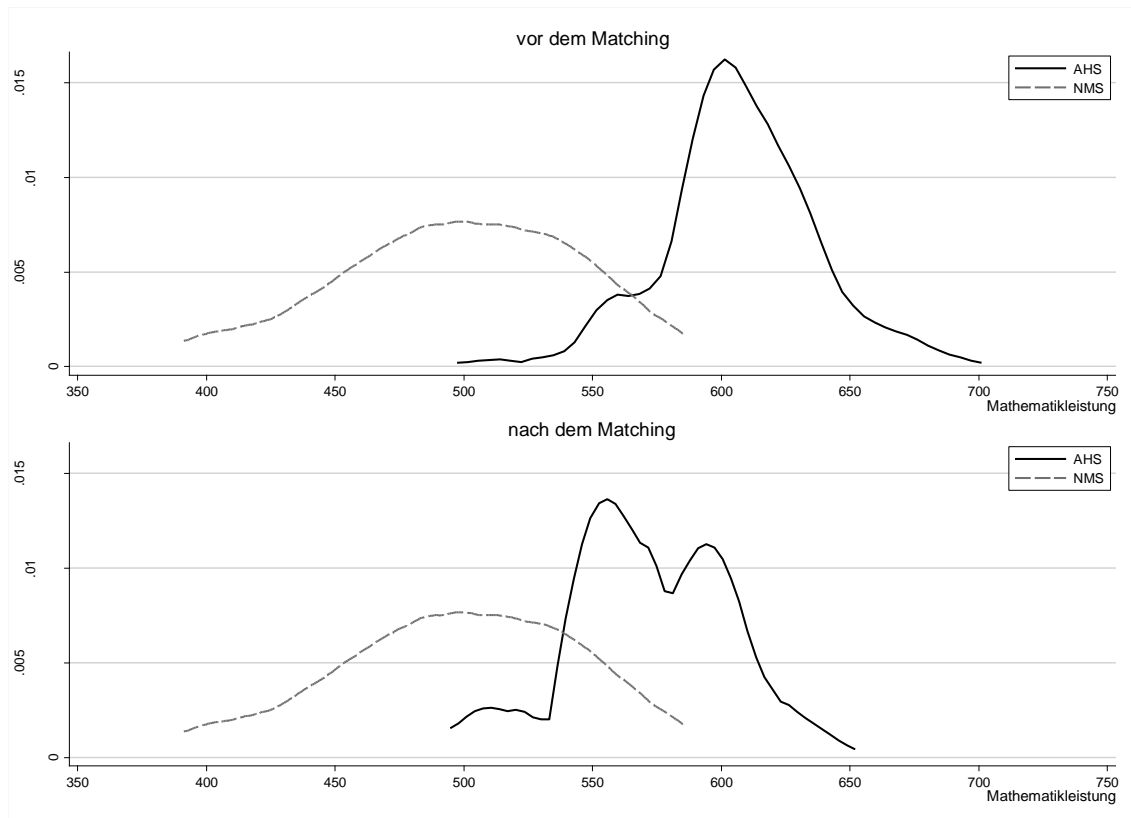
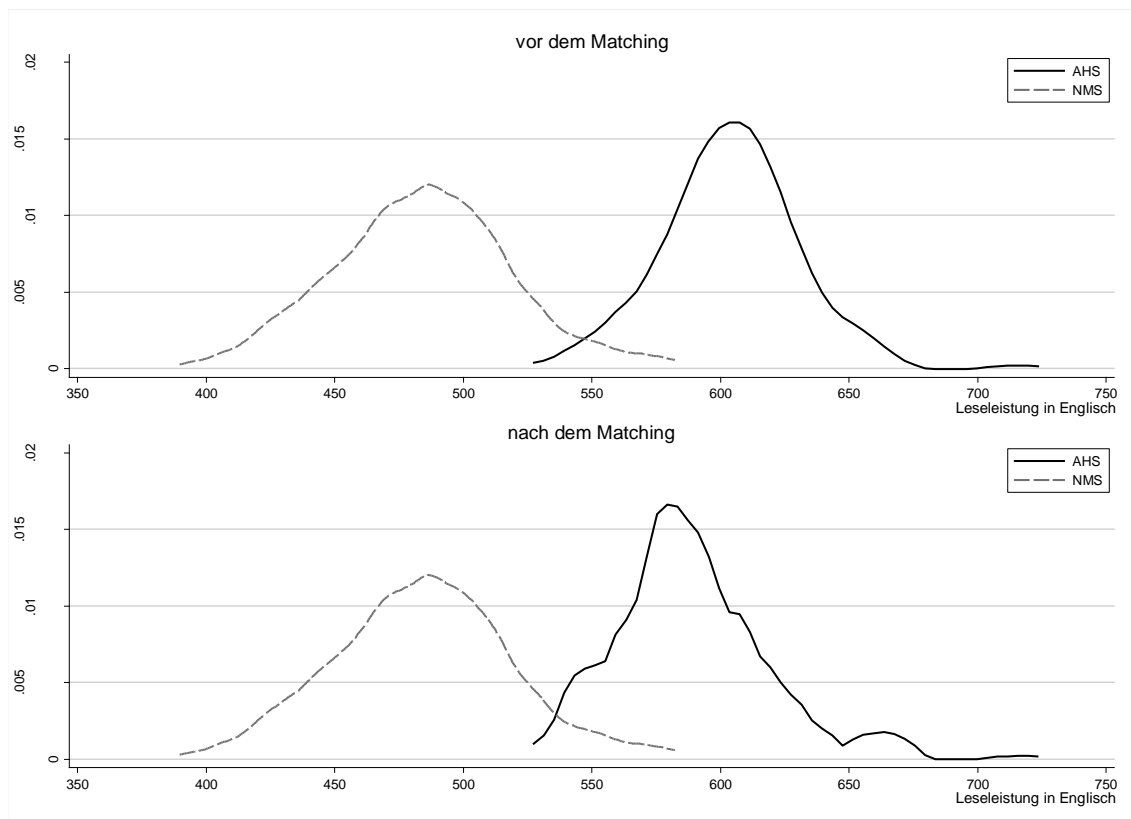


Abbildung 2b: Gegenüberstellung der Leistungsverteilungen der NMS und der AHS-Unterstufe vor und nach dem Matching für BIST-E8 (PSM-Durchlauf 1)



4.2. Leistungsunterschiede zur HS

In diesem Abschnitt sollen, analog zur Vorgehensweise bei der AHS-Unterstufe, die Leistungsniveaus der NMS und der HS einander gegenübergestellt werden. Vorab gilt es jedoch, einen Überblick über die soziale Komposition der SchülerInnenpopulation der beiden Schultypen bereitzustellen. Dies wird in Tabelle 7 geleistet. Wie klar zu erkennen ist, besitzen die NMS der Generationen 1 und 2 auch im Vergleich zu HS durchschnittlich eine sozial in stärkerem Maße benachteiligte Zusammensetzung ihrer SchülerInnen. Während sich die Differenzen in den typischen vertikalen Ungleichheitsdimensionen der formalen Bildung der Eltern (BIST-M8: 3,9 PP; BIST-E8: 1,8 PP) und deren beruflichem Status (BIST-M8: nicht signifikant; BIST-E8: -2,6 PP) in Grenzen halten und nicht konsistent in eine Richtung gehen, fallen ausgeprägte Unterschiede in den migrationsbezogenen Merkmalen in den Blick. Während der mittlere Anteil an SchülerInnen mit Migrationshintergrund in den HS bei 13,2 Prozent (BIST-M8) bzw. 12,9 Prozent (BIST-E8) liegt, können für die NMS jeweils fast doppelt so hohe Anteile beobachtet werden (BIST-M8: 22,2%; BIST-E8: 22,3%). Nahezu identische Befunde liegen für die mittleren Anteile an SchülerInnen mit nichtdeutscher Alltagssprache vor. Diese Differenzen sind vermutlich dem Umstand geschuldet, dass sich im Zuge der Implementierung der NMS im Jahr 2008 zunächst verstärkt Hauptschulen aus dem urbanen Bereich mit einem hohen Anteil an migrantischen SchülerInnen dafür entschieden haben, an dem Schulversuch zu partizipieren (siehe dazu Weber, Bacher, Altrichter & Leitgöb, 2015). Aus der holistischen Perspektive des Sozialindikators ergibt sich abschließend das folgende Bild: Der Anteil an NMS, die zumindest einen Wert von über 125 aufweisen und folglich den Kategorien „hoch belastet“ oder „sehr hoch belastet“ zuzuordnen sind, ist mit 41,8 Prozent beinahe doppelt so hoch wie jener HS mit 22,6 Prozent.

Zusammenfassend muss aus den Befunden abgeleitet werden, dass – wie bereits im Zuge der Erläuterungen zum Vergleich mit der AHS-Unterstufe angedeutet – die NMS in den ersten Jahren des Schulversuchs keineswegs die intendierte „gemeinsame Schule aller 10- bis 14-Jähren“ repräsentiert. Vielmehr stellen die SchülerInnen in der NMS eine hochselektive Population dar, die sich durch jeweils sehr hohe Anteile an migrantischen SchülerInnen sowie SchülerInnen aus bildungsfernen Herkunftsfamilien sogar noch deutlich von der SchülerInnenschaft der HS absetzt. Aus diesem Grund muss auch mit Blick auf die in der bildungspolitischen Diskussion zu beobachtende Praxis eindringlich davor gewarnt werden, die NMS mit einem Testlauf für die Gesamtschule gleichzusetzen.

Tabelle 7: Soziale Zusammensetzung der SchülerInnenschaft von NMS und HS im Vergleich

BIST-M8				
Merkmal	NMS (n=67)	HS (n=957)	Differenz (in PP) ^(a)	p-Wert
geringe schulische Bildung der Eltern (\bar{x}_1)	18,4%	14,5%	3,9	<0,001
geringer beruflicher Status der Eltern (\bar{x}_2)	43,4%	44,3%	-0,9	0,587
Migrationshintergrund (\bar{x}_3)	22,2%	13,2%	9,0	<0,001
andere zu Hause gesprochene Sprache (\bar{x}_4)	22,3%	12,9%	9,4	<0,001
Sozialindex (<i>S</i>)				
gering (100-115)	22,4%	28,1%	-5,7	
mittel (>115-125)	35,8%	49,3%	-13,5	
hoch (>125-135)	14,9%	11,6%	3,3	
sehr hoch (>135)	26,9%	11,0%	15,9	<0,001
gesamt	100,0%	100,0%		
Mittelwert	126,6	121,2	6	<0,001
BIST-E8				
Merkmal	NMS (n=227)	HS (n=818)	Differenz (in PP) ^(a)	p-Wert
geringe schulische Bildung der Eltern (\bar{x}_1)	16,3%	14,5%	1,8	0,006
geringer beruflicher Status der Eltern (\bar{x}_2)	41,8%	44,4%	-2,6	0,011
Migrationshintergrund (\bar{x}_3)	22,3%	12,9%	9,4	<0,001
andere zu Hause gesprochene Sprache (\bar{x}_4)	22,0%	12,4%	9,6	<0,001
Sozialindex (<i>S</i>)				
gering (100-115)	22,0%	31,4%	-9,4	
mittel (>115-125)	39,7%	46,1%	-6,4	
hoch (>125-135)	17,6%	13,0%	4,6	
sehr hoch (>135)	20,7%	9,5%	11,2	<0,001
gesamt	100,0%	100,0%		
Mittelwert	125,6	121,1	5	<0,001

^(a) PP ... Prozentpunkte; Ausnahme: Mittelwert des Sozialindex

Im Gegensatz zur AHS-Unterstufe können für die Schätzung der Propensity Scores (zu den Ergebnissen siehe Tabelle A1 im Anhang) aufgrund der Masse an zur Verfügung stehenden HS (BIST-M8: 957; BIST-E8: 818) sowie der deutlich geringer ausgeprägten Unterschiede in der sozialen Komposition die vier den Sozialindex konstituierenden Einzelindikatoren herangezogen werden. Wie aus Tabelle 8 entnommen werden kann, lassen sich aus der Vergleichsgruppe der HS zu allen 67 (BIST-M8) bzw. 227 (BIST-E8) NMS statistische Zwillinge unter Einhaltung des fixierten Calipers von $\varepsilon = 0,05$ finden. Weiterhin

konnte durch das PSM eine Balancierung der Gruppen der NMS und HS in allen vier Sozialindikatoren und folglich auch im Sozialindex erzielt werden.

Tabelle 8: Ergebnisse des PSM mit den vier Einzelindikatoren

	BIST-M8	BIST-E8
Empfängerdatei (NMS)	67	227
Spenderdatei (HS)	957	818
Anzahl der gefunden Paare	67	227
Anzahl der mehrfach verwendeten Schulen der Spenderdatei	4	48
Auftretenshäufigkeit der am häufigsten verwendeten Schule der Spenderdatei	3	6
minimale Distanz innerhalb der Paare	0	0
maximale Distanz innerhalb der Paare	0,03	0,04
festgelegter Caliper	0,05	0,05
Balancierung nach		
geringe schulische Bildung der Eltern	ja	Ja
geringer beruflicher Status der Eltern	ja	Ja
Migrationshintergrund	ja	Ja
andere zu Hause gesprochene Sprache	ja	Ja
Sozialindex	ja	Ja

Im „naiven“ Vergleich der mittleren Kompetenzniveaus in Mathematik und im Lesen englischer Texte zeigen sich vor dem Matching moderate allerdings statistisch hoch signifikante ($p < 0,001$) Unterschiede zugunsten der HS (BIST-M8: 18,3 Punkte; BIST-E8: 9,2 Punkte; es werden die Beträge der Differenzen ausgewiesen). Während sich durch das Matching der absolute Schultypeneffekt in Mathematik um 6,1 Punkte (33,3%) auf 12,2 Punkte reduziert und seine Signifikanz einbüßt ($p = 0,074$), vergrößert sich der Betrag der Differenz der mittleren Lesekompetenzen in Englisch um 3,0 Punkte auf ebenfalls 12,2 Punkte. In beiden Kompetenzbereichen ist somit im Zuge des „fairen“ Vergleichs nach Berücksichtigung der sozialen Komposition der SchülerInnenenschaft der gleiche Leistungsabstand zwischen der NMS und der HS (entspricht dem ATT; siehe Abschnitt 3.4) zu beobachten. Bezüglich des Signifikanzniveaus des Mathematikeffekts muss angemerkt, dass der entsprechende Standardfehler erheblich von der geringen Fallzahl ($67 * 2 = 134$ Schulen; $SE = 7,91$) nach oben gedrückt wird und der Effekt bei einer zu den BIST-E8 Daten vergleichbaren Fallzahl ($227 * 2 = 454$ Schulen; $SE = 4,41$) den Schwellenwert von $p < 0,05$ deutlich unterschreiten würde. Somit bleibt festzuhalten, dass durch das PSM und der damit verbundenen statistischen Kontrolle der beobachteten Unterschieden in der sozialstrukturellen Zusammensetzung der SchülerInnenpopulation zwischen der NMS und der HS keine vollständige Kompensation der Leistungsdifferenzen erzielt werden konnte.

Die vorliegenden Befunde erlauben die folgende Beantwortung der Forschungsfrage (II): Bei annähernd gleicher sozialer Zusammensetzung verzeichnet die NMS im Vergleich zur HS in der Tendenz leichte Defizite in den Mathematik- und Lesekompetenzen in Englisch.

Von einer deutlichen leistungsbezogenen Absetzung der NMS der Generationen 1 und 2 von der HS kann somit keine Rede sein.

Tabelle 9: Mittlere Testleistungen der NMS und HS im Vergleich

	NMS	HS	Effekt	P
BIST-M8				
vor Matching („naiv“)	496,7	515,0	-18,3	<0,001
nach Matching („fair“)	496,7	508,9	-12,2	0,074
BIST-E8				
vor Matching („naiv“)	484,2	493,4	-9,2	<0,001
nach Matching („fair“)	484,2	496,4	-12,2	0,003

Nimmt man einen graphischen Vergleich der Verteilungen der mittleren Kompetenzniveaus der NMS und HS auf der Basis von Kerndichteplots (Abbildungen 3a und 3b) vor, so zeigt sich, sowohl vor als auch nach dem Matching jeweils ein hoher Deckungsgrad dieser Verteilungen. Während beide Schulformen in beiden Kompetenzbereichen Schulen am unteren Ende des Leistungsspektrums mit deutlich unter 400 Punkten verzeichnen (in Mathematik ist der Anteil dieser Schulen in der NMS allerdings merklich höher), kann am oberen Ende im Gegensatz zu den NMS noch eine geringe Anzahl an HS deutlich jenseits der 600 Punkte-Marke beobachtet werden. Durch das PSM ändern sich die Verteilungen der HS mäßig. Während sich bei den Mathematikkompetenzen ansatzweise eine bimodale Verteilung entwickelt, zieht bei den Lesekompetenzen in Englisch ein kleines lokales Maximum am oberen Ende der als moderat linkssteil-rechtsschief zu bezeichnenden Verteilung (bei etwas über 650 Punkten) den Mittelwert minimal nach oben. Dieses Phänomen ist – allerdings schwächer ausgeprägt – bei den Mathematikleistungen ebenfalls zu beobachten. Hinsichtlich der Streuungen der gemessenen schulbezogenen Kompetenzen kann abschließend angemerkt werden, dass in Englisch (Lesen), insbesondere nach dem Matching, eine wesentlich höhere Leistungsheterogenität im Schultyp der HS (vor dem Matching: $s = 36,86$; nach dem Matching: $s = 48,75$) zu verzeichnen ist als in der NMS ($s = 34,46$). Für die Mathematikkompetenzen ist der gegenläufige Fall zu beobachten (NMS: $s = 46,19$; HS vor dem Matching: $s = 40,12$; HS nach dem Matching: $42,61$).

Abbildung 3a: Gegenüberstellung der Leistungsverteilungen der NMS und der HS vor und nach dem Matching für BIST-M8 (PSM-Durchlauf 1)

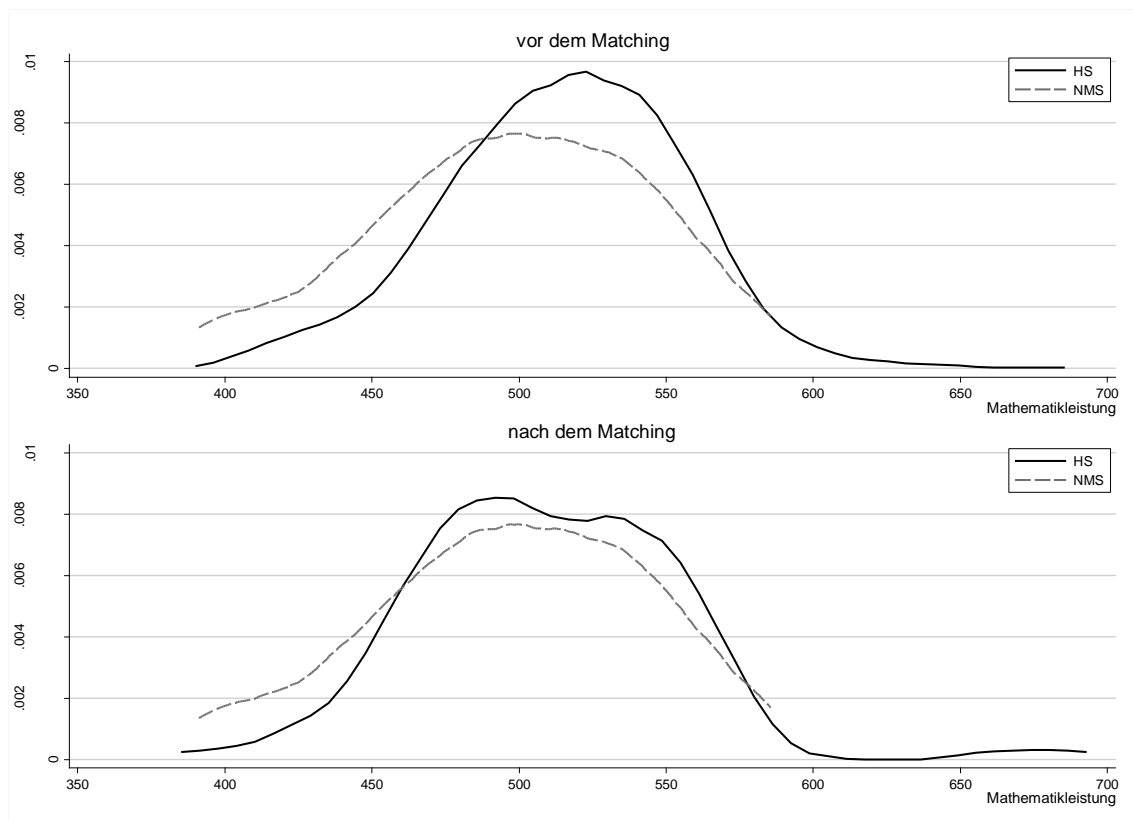
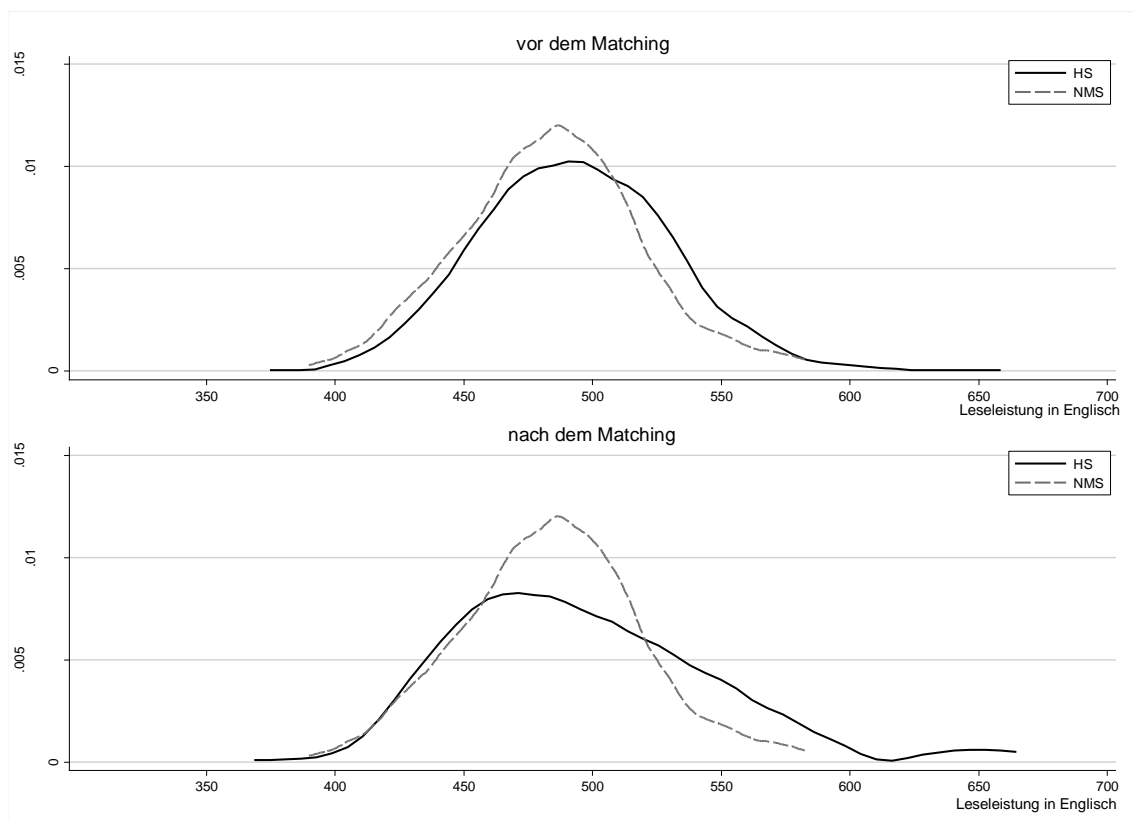


Abbildung 3b: Gegenüberstellung der Leistungsverteilungen der NMS und der HS vor und nach dem Matching für BIST-E8 (PSM-Durchlauf 1)



4.3. Unterschiede zwischen NMS mit geringem und großem Leistungsabstand zur AHS-Unterstufe

Neben dem Kompetenzvergleich zwischen der NMS und der AHS-Unterstufe sowie der HS liegt der zweite Fokus des Erkenntnisinteresses in diesem Beitrag gemäß Forschungsfrage (III) auf der Identifikation von Faktoren, die innerhalb der Schulform der NMS Einfluss auf das Leistungsniveau der Schulen (operationalisiert über den relativen Leistungsabstand zur AHS-Unterstufe; siehe Abschnitt 3.4) – positiv wie negativ – nehmen. Der vorliegende Abschnitt widmet sich der Beantwortung dieser Fragestellung.

In einem ersten Schritt soll auf Basis der Daten der Enderhebungen zu G1 und G2 bivariat untersucht werden, ob bzw. in welchem Maße sich die Gruppen der leistungsnahen bzw. leistungsfernen NMS in den Kompetenzbereichen Mathematik (8. Schulstufe), Deutsch und Englisch, den verbalen sowie nonverbalen kognitiven Kompetenzen zum Zeitpunkt des Eintritt in die Schule in der 5. Schulstufe, im Sozialindex, in den zentralen NMS-Konzeptmerkmalen, in der Anomie im Unterricht, den abweichenden Verhaltensweisen der SchülerInnen und in der räumlichen Nähe zur nächstgelegenen AHS-Unterstufe voneinander unterscheiden.

Erwartungsgemäß divergieren die leistungsnahen und die leistungsfernen NMS deutlich in den mittleren Kompetenzniveaus in der 8. Schulstufe. So beträgt die Differenz der Mittelwerte in allen drei Kompetenzbereichen über 50 Punkte. Auffallend ist, dass die beiden Gruppen in Deutsch die niedrigsten mittleren Leistungen aufweisen und der Abstand am geringsten ausgeprägt ist. In Hinblick auf die mittleren verbalen und nonverbalen kognitiven Fähigkeiten der SchülerInnen kann bereits zu Schuleintritt in der 5. Schulstufe ein signifikantes Gefälle von 3,27 bzw. 2,27 Punkten zugunsten der leistungsnahen NMS beobachtet werden. Der Befund legt die Vermutung nahe, dass auch die schulischen Leistungen der SchülerInnen aus leistungsnahen NMS bereits beim Einstieg in die Sekundarstufe I im Mittel höher sind als jene von SchülerInnen aus leistungsfernen NMS und erstgenannte Schulen somit einen „Startvorteil“ besitzen. Diese Defizite können von leistungsfernen NMS offensichtlich nicht kompensiert werden. Vielmehr vergrößert sich die Differenz der Mittelwerte in den verbalen und nonverbalen kognitiven Fähigkeiten zwischen den beiden Gruppen von NMS in der 8. Schulstufe auf 4,14 ($\Delta = 0,87$; $t = 1,388$; $df = 57$; $p = 0,171$) bzw. 4,87 ($\Delta = 2,60$; $t = 4,104$; $df = 57$; $p < 0,001$) Punkte.²¹ Weiterhin offenbaren sich deutliche Unterschiede zwischen den leistungsnahen und den leistungsfernen NMS hinsichtlich der sozialen Komposition der SchülerInnenschaft. Diese manifestieren sich in einer ausgeprägten Differenz der mittleren Sozialindizes (absolut: 12,25 Punkte). Die leistungsfernen NMS arbeiten folglich in noch wesentlich stärkerem Maße mit SchülerInnen aus sozial benachteiligten Herkunftsfamilien. Im Vergleich der mittleren Sozialindizes der leistungsnahen NMS mit den AHS-Unterstufen und den HS (siehe dazu die Tabellen 3 und 7) sind diese mit 122, 59 Punkten im Bereich der HS (jeweils etwa 121 Punkte) zu verorten. Etwas überraschend können bivariat keine signifikanten Effekte der NMS-Konzeptmerkmale nachgewiesen werden. Leistungsnahen und leistungsfernen NMS unterscheiden sich hinsichtlich der Erstellung individueller Förderpläne, dem Anteil an Oberstufenlehrkräften, der Bereitstellung von Nachmittagsbetreuung, dem Angebot des verschränkten Teamteachings sowie dem Ausmaß an SchülerInnenorientierung im Unterricht, der Gewährung von selbstreguliertem Lernen und dem Anteil an vorhandenen

²¹ Es ist somit keine Regression zur Mitte zu beobachten (siehe etwa Campbell und Kenny, 1999).

NMS-Modellklassen nicht systematisch voneinander. Demgegenüber sind signifikante Differenzen in Bezug auf das Ausmaß an Anomie im Unterricht zu erkennen. In leistungsnahen NMS scheinen Situationen der Verunsicherung und Desorganisation im Unterricht durchschnittlich seltener aufzutreten als in der Gruppe der leistungsfernen NMS. Dies gilt ebenso für abweichende bzw. gewalttätige Verhaltensweisen der SchülerInnen. In der Tendenz scheint abschließend noch die räumliche Entfernung zur nächstgelegenen AHS-Unterstufe zwischen den beiden Gruppen von NMS zu divergieren. Mit durchschnittlich 12,74 km liegen die leistungsnahen NMS geographisch deutlich weiter von der nächstgelegenen AHS-Unterstufe entfernt als die leistungsfernen NMS (7,46 km). Die Inkonsistenz zwischen dem parametrischen und dem nicht-parametrischen Signifikanztest ist bei dieser Variable vermutlich auf die Schiefe der Verteilung und einige wenige Extremwerte (drei Fälle mit über 40 km) zurückzuführen. Eine Interpretation der Variable als Proxy für den Urbanisierungsgrad des Schulstandortes indiziert, dass leistungsnahen NMS verstärkt in ländlich strukturierten Regionen zu finden sind. In Kombination mit dem Befund der heterogeneren sozialen Zusammensetzung der SchülerInnen in leistungsnahen NMS liegt die Vermutung nahe, dass aufgrund des fehlenden Angebots bzw. des langwierigen Anfahrtsweges zur nächstgelegenen AHS-Unterstufe verstärkt auch Kinder aus sozial privilegierten Herkunftsfamilien eine NMS besuchen und so der tendenzielle Effekt der relativen Leistungsnähe dieser Schulen zur AHS-Unterstufe (zumindest partiell) erklärt werden kann. Vermutlich hätten diese Kinder allerdings alternativ auch eine HS besucht, so dass der Effekt nicht eindeutig der NMS als Schulform zugeschrieben werden kann.

Tabelle 10: Unterschiede in der sozialen Zusammensetzung, den kognitiven Fähigkeiten, in den NMS-Implementierungsmerkmalen sowie in unterrichts- und verhaltensbezogenen Merkmalen zwischen NMS mit geringer relativer Leistungsdifferenz zur AHS (Δ_{WLE}^+ gering) und jenen mit hoher relativer Leistungsdifferenz zur AHS (Δ_{WLE}^+ groß)

Merkmal	Δ_{WLE}^+ gering (n=31)	Δ_{WLE}^+ groß (n=29)	$\Delta MW^{(a)}$	df	t-Wert	p-Wert	p-Wert ^(b)
Mathematikleistung (WLE)	498,25	433,51	64,74	58	5,78	<0,001	<0,001
Deutschleistung (WLE)	481,46	426,56	54,90	58	5,53	<0,001	<0,001
Englischleistung (WLE)	504,10	438,65	65,45	58	7,37	<0,001	<0,001
verbale Fähigk. 5. Schulst. (KFT)	43,63	40,36	3,27	58	3,99	<0,001	0,001
nonverb. Fähigk. 5. Schulst. (KFT)	48,58	46,21	2,37	58	3,46	<0,001	0,002
verbale Fähigk. 8. Schulst. (KFT)	46,20	42,06	4,14	58	5,34	<0,001	<0,001
nonverb. Fähigk. 8. Schulst. (KFT)	47,24	42,37	4,87	58	6,56	<0,001	<0,001
Sozialindex (S)	122,59	134,84	-12,25	58	-2,89	0,006	0,010
individueller Förderplan	0,67	0,86	-0,19	53	-1,67	0,102	0,100
Anteil Oberstufenlehrkräfte	0,18	0,16	0,02	50	0,37	0,716	0,420
Nachmittagsbetreuung	0,62	0,66	-0,04	56	-0,27	0,789	0,787
verschränktes Teamteaching	3,21	3,07	0,14	55	0,83	0,410	0,460
SchülerInnenorientierung	2,94	3,02	-0,08	56	-0,65	0,521	0,391
selbstreguliertes Lernen	4,11	4,31	-0,20	54	-1,15	0,255	0,251
Anteil an NMS-Modellklassen	0,20	0,24	0,04	58	0,36	0,720	0,372
Anomie im Unterricht	2,67	2,88	-0,21	58	-2,99	0,004	0,001
gewalttätiges Verhalten	2,59	2,95	-0,36	58	-2,53	0,014	0,003
Opfererfahrungen	2,28	2,68	-0,40	58	-3,38	0,001	0,003
abweichendes Verhalten	1,84	2,24	-0,40	58	-3,39	0,001	<0,001
räumliche Distanz zur nächstgelegenen AHS-Unterstufe	12,74	7,46	5,28	58	-1,51	0,136	0,016

(a) MW ... Mittelwert

(b) Wilcoxon-Mann-Whitney-Test (exakt)

Signifikante Differenzen ($p < 0,05$) werden kursiv markiert.

In einem finalen Schritt soll im Zuge der Modellbildung geprüft werden, ob die diskutierten Merkmale auch multivariat unter gegenseitiger statistischer Kontrolle empirischen Bestand haben. Da sich die PMLE-basierten Schätzer sowie deren Standardfehler (siehe Tabelle A2) nicht substantiell von jenen der MLE unterscheiden, dient Tabelle 11 als Grundlage zur Interpretation der Ergebnisse. Dem Generationsindikator (G2-Indikator) wird – auch vor dem Hintergrund, dass er in allen Modellen statistisch nicht signifikant ist – als Kontrollvariable im Zuge der inhaltlichen Interpretation keine Bedeutung beigemessen.

In Modell 1 wird die Zugehörigkeit zur Gruppe der leistungsnahen NMS lediglich auf den Sozialindex der Schulen (und den G2-Indikator) regressiert. Wie aufgrund der bivariaten Befunde zu erwarten war, liegt ein signifikanter negativer Effekt ($\beta(\text{KHB}) = -0,058$; $p = 0,016$)²² vor. Dieser verschwindet, wenn in Modell 2 auch die verbalen kognitiven Eingangsfähigkeiten (5. Schulstufe) kontrolliert werden ($\beta(\text{KHB}) = -0,032$; $p = 0,347$).

²² Um die Entwicklung der interessierenden Effekte über die Modelle hinweg nachverfolgen zu können, werden in Klammern die in Tabelle A3 aufbereiteten KHB-Koeffizienten berichtet. Die KHB-Koeffizienten repräsentieren die auf die Metrik des (vollen) Modells 4 reskalierten Logit-Koeffizienten der Modelle 1 bis 3. Zu den Details siehe Abschnitt 3.4.

Dies impliziert, dass sich leistungsnahe und leistungsferne NMS mit einem vergleichbaren kognitiven Fähigkeitsniveau nicht signifikant hinsichtlich der sozialen Zusammensetzung ihrer SchülerInnen (zumindest auf Basis des additiven Sozialindex) unterscheiden. Die verbalen kognitiven Fähigkeiten selbst wirken in erheblichem Maße auf die Gruppenzugehörigkeit ($\beta(\text{KHB}) = 0,550$; $p = 0,005$). So geht mit der Zunahme des KFT-Scores (Range: 60 Punkte) um eine Einheit ein durchschnittlicher Anstieg der Wahrscheinlichkeit der Gruppenzugehörigkeit zu den leistungsnahe NMS um 7,9 Prozentpunkte einher. Dieser Effekt bleibt verhältnismäßig konstant, wenn im Zuge von Modell 3 zudem die Erstellung individueller Förderpläne für die SchülerInnen als ausgewähltes Konzeptmerkmal der NMS Berücksichtigung findet ($\beta(\text{KHB}) = 0,503$; $p = 0,012$). Analog zu den bivariaten Ergebnissen ist der Effekt der individuellen Förderpläne auch multivariat nicht signifikant ($\beta(\text{KHB}) = -0,825$; $p = 0,346$).²³ Das Modell 4 repräsentiert eine Erweiterung von Modell 3 um das Ausmaß abweichender Verhaltensweisen der SchülerInnen in der Schule, Anomie im Unterricht und die räumliche Distanz zur nächstgelegenen AHS-Unterstufe. In Hinblick auf die explizite Beantwortung von Forschungsfrage (III) muss somit festgehalten werden, dass in diesem finalen Modell ein signifikanter Effekt lediglich für das verbale kognitive Fähigkeitsniveau der Schulen identifiziert werden kann. Im Vergleich zu Modell 3 hat sich dieser allerdings etwas abgeschwächt ($\beta(\text{KHB}) = -0,436$; $p = 0,033$). Die Effekte der anderen Variablen (inklusive des NMS-Konzeptmerkmals „individuelle Förderpläne“) vermögen die geforderte Signifikanzschwelle von $p < 0,05$ nicht zu unterschreiten. Eine tendenzielle Wirkung kann dennoch für die Anomie im Unterricht konstatiert werden, der in die erwartete Richtung verläuft. So reduziert ein steigendes Ausmaß an Anomie die Wahrscheinlichkeit der Zugehörigkeit zur Gruppe der leistungsnahe NMS.

Unter Beachtung des Korrelationsgefüges zwischen den unabhängigen Variablen (siehe Tabelle A4) kann für den Einfluss der kognitiven Ausgangsfähigkeiten der folgende Interpretationsversuch vorgenommen werden: NMS mit einem höheren kognitiven Fähigkeitsniveau weisen eine günstigere soziale Komposition der SchülerInnenschaft auf, die zum Teil auf die Entfernung zur nächstgelegenen AHS-Unterstufe und somit auf den Schulstandort im eher ländlich strukturierten Raum zurückzuführen ist. Diese Schulen gehören, gerade wegen der ausgeprägteren kognitiven Fähigkeiten der SchülerInnen, wiederum mit einer höheren Wahrscheinlichkeit der Gruppe der leistungsnahe NMS an. Multivariat führt dies zum Verlust der signifikanten Erklärungskraft dieser anderen Variablen, so dass nur mehr den kognitiven Fähigkeiten – also dem allgemeinen IQ – eine signifikante Wirkung zukommt. Dies ist vermutlich darauf zurückzuführen, dass die kognitiven Fähigkeiten der SchülerInnen direkt auf das Kompetenzniveau der Schulen wirken und die anderen Merkmale lediglich vermittelt. Davon unberührt kommt der Anomie im Unterricht zumindest in der Tendenz eine eigenständige negative Wirkung zu, die bereits bivariat in deutlich stärkerer Ausprägung zu beobachten ist. Wie Tabelle A4 zu entnehmen ist, besteht bivariat weiterhin ein signifikanter gegensinniger Zusammenhang zwischen dem mittleren kognitiven Niveau der Schulen sowie dem Ausmaß an Anomie im Unterricht. Dieser Befund kann auch durch empirische Evidenzen auf der Individualebene abgesichert werden, dass das Ausmaß devianter und delinquenter Verhaltensweisen bei Kindern und Jugendlichen mit dem kognitiven Fähigkeitsniveau variiert (siehe etwa McGloin & Pratt, 2003; McGloin, Pratt & Maahs, 2004).

²³ Anstelle der individuellen Förderpläne wurden auch die anderen NMS-Konzeptmerkmale iterativ in Modell 3 berücksichtigt. Es konnten ebenfalls keine signifikanten Effekte identifiziert werden.

Im Zuge einer abschließenden Evaluation der Modellpassung kann konstatiert werden, dass, gemessen an den Kriterien des Pseudo-R², dem Anteil korrekter Prognosen sowie der AUC, die Erklärungskraft über die Modelle beständig zugenommen hat. Ferner kann dem vollen Modell 4 (trotz lediglich eines signifikanten Effekts) mit einem Anteil an korrekten Prognosen von 77,8 Prozent und einer AUC von 0,825 eine zufriedenstellende Erklärungsleistung attestiert werden.

Tabelle 11: Ergebnisse der geschätzten Logit-Modelle zur Erklärung der Zugehörigkeit zur Gruppe der NMS mit geringer relativer Leistungsdifferenz zur AHS-Unterstufe (Extremgruppenvergleich)

Merkmal	Modell 1				Modell 2			
	β	p	AME ^(d)	p	β	p	AME	p
Konstante	6,450	0,010	-	-	-18,263	0,067	-	-
G2-Indikator	0,033	0,953	0,007	0,953	-0,467	0,471	-0,089	0,463
Sozialindex (S)	<i>-0,050</i>	<i>0,010</i>	<i>-0,011</i>	<i>0,001</i>	0,009	0,772	0,002	0,771
verb. Fähigk. 5. Schulst. (KFT)	-	-	-	-	<i>0,415</i>	<i>0,014</i>	<i>0,079</i>	<i>0,002</i>
LL ^(a)	-37,34				-33,150			
Pseudo-R ² ^(b)	0,102				0,189			
Anteil korrekter Prognosen	0,633				0,612			
AUC ^(c)	0,691				0,755			

Merkmal	Modell 3				Modell 4			
	β	p	AME	p	β	p	AME	p
Konstante	-19,418	0,088	-	-	-10,968	0,398	-	-
G2-Indikator	-0,407	0,546	-0,073	0,540	-0,741	0,326	-0,117	0,310
Sozialindex (S)	0,010	0,769	0,002	0,768	0,017	0,640	0,003	0,638
verb. Fähigk. 5. Schulst. (KFT)	<i>0,449</i>	<i>0,017</i>	<i>0,081</i>	<i>0,002</i>	<i>0,436</i>	<i>0,033</i>	<i>0,069</i>	<i>0,012</i>
individueller Förderplan	-0,917	0,265	-0,165	0,245	-1,028	0,241	-0,163	0,218
abweichendes Verhalten	-	-	-	-	-0,013	0,991	-0,002	0,991
Anomie im Unterricht	-	-	-	-	-3,031	0,106	-0,480	0,075
räumliche Distanz zur nächstgelegenen AHS-Unterstufe	-	-	-	-	-0,008	0,782	-0,001	0,781
LL	-28,741				-26,077			
Pseudo-R ²	0,231				0,303			
Anteil korrekter Prognosen	0,759				0,778			
AUC	0,797				0,824			

^(a) LL ... Log-Likelihood

^(b) nach McFadden (1973)

^(c) AUC ... Bereich unter der ROC-Kurve²⁴

^(d) AME ... durchschnittlicher marginaler Effekt

Lesehilfe (für Modell 1): Im Mittel reduziert sich die Wahrscheinlichkeit der Zugehörigkeit zur Gruppe der leistungsnahen NMS um 1,1 Prozentpunkte (-0,011), wenn der Sozialindex (theoretischer Wertebereich zwischen 100 und 200; empirischer Wertebereich zwischen 105,16 und 178,69) um eine Einheit steigt.

Signifikante Effekte ($p < 0,05$) werden kursiv markiert.

²⁴ Für eine Beschreibung sowie Details zur Ermittlung der ROC-Kurve und der AUC mittels Stata siehe Pepe, Longton und Janes (2009).

5. Zusammenfassung und Schlussfolgerungen

Das Ziel des vorliegenden Beitrags galt der Evaluation der im Rahmen der NMS vermittelten schulischen Kompetenzen in Mathematik und in Englisch (Lesen). Zu diesem Zweck wurden „faire“ Leistungsvergleiche mit den beiden anderen zentralen Schulformen der Sekundarstufe I, der AHS-Unterstufe und der HS, auf Schulebene durchgeführt. Zur Berücksichtigung der Auswirkungen der auf soziale Selektionseffekte zurückgehenden systematischen Unterschiede in der sozialen Zusammensetzung der SchülerInnenchaft auf das Kompetenzniveau der Schultypen wurde das Verfahren des Propensity Score Matchings (PSM) zur Anwendung gebracht. Im Anschluss wurden die Differenzen der mittleren Kompetenzniveaus zwischen den NMS und den gematchten AHS-Unterstufen bzw. HS ermittelt und auf Signifikanz geprüft. Eine Gegenüberstellung der Verteilungen der Kompetenzmerkmale zwischen den Schultypen wurde graphisch über die Darstellung von Kerndichteplots realisiert. Darüber hinaus wurde der Frage nachgegangen, hinsichtlich welcher Merkmale sich NMS mit einem geringen und einem großen relativen Leistungsabstand zur AHS-Unterstufe unterscheiden. Neben der sozialen Komposition wurde auf kognitive Fähigkeitsmerkmale der SchülerInnen, Konzeptmerkmale der NMS, unterrichtsbezogene Merkmale, verhaltensbezogene Merkmale der SchülerInnen (insbesondere die Häufigkeit abweichender und gewalttätiger Handlungen) und die räumliche Distanz zur nächstgelegenen AHS-Unterstufe abgestellt. Nach bivariaten Analysen basierend auf t-Tests und nichtparametrischen Wilcoxon-Mann-Whitney-Tests wurde der empirische Bestand der Effekte unter multivariaten Bedingungen über die Schätzung von Logit-Modellen geprüft. Die zentralen Ergebnisse der durchgeführten Analysen sollen kurz zusammengefasst werden.

- (i) Die NMS der Generationen 1 und 2 und die AHS-Unterstufe sind hinsichtlich der sozialen Zusammensetzung der SchülerInnen de facto nicht vergleichbar. Beide Schulformen unterscheiden sich in den sozialen Einzelindikatoren „geringe Bildung der Eltern“, „geringe berufliche Position der Eltern“, „Migrationshintergrund“ und „andere zu Hause gesprochene Sprache“ dermaßen, dass im Zuge des PSM keine Balancierung möglich war. Dies zeigt auf der andern Seite allerdings auch sehr deutlich die Problematik „naiver“ Leistungsvergleiche zwischen der NMS und der AHS-Unterstufe auf, da ein relevanter Teil der Unterschiede in den Kompetenzniveaus auf soziale Selektions- und Kompositionseffekte zurückzuführen ist (siehe Punkt (iii)).
- (ii) Dieser Vorbehalt kann für den alternativ vorgenommen „fairen“ Vergleich der beiden Schultypen basierend auf dem additiv aus den vier Einzelindikatoren generierten Sozialindex nicht vollständig entkräften. Zwar gelingen die Identifikation statistischer Zwillinge aus der Vergleichsgruppe der AHS-Unterstufen zu allen NMS und die Balancierung der beiden Gruppen hinsichtlich des Sozialindex, allerdings unterscheiden sich die gemachten Schulpaare hinsichtlich der dem Index zugrundeliegenden Einzelindikatoren zum Teil erheblich voneinander.
- (iii) Durch die Berücksichtigung der sozialen Komposition der SchülerInnenchaft über das PSM reduziert sich der Betrag der mittleren Leistungsdifferenz zwischen der NMS und der AHS-Unterstufe in Mathematik von 110,4 Punkten („naiver“ Vergleich) um 35,4 Punkte auf 75,0 Punkte. In Englisch (Lesen) wird eine deutlich schwächere Reduktion erzielt (von 119,6 Punkten um 15,0 Punkte auf 104,6 Punkte). Durch Unterschiede in der sozialen Zusammensetzung der SchülerInnenpopulation zwischen der NMS und der AHS-Unterstufe kann ein knappes Drittel (32,1%) des beträchtlichen beobachteten Niveauunterschieds in den Mathematikkompetenzen erklärt werden, in den Lesekompetenzen in Englisch lediglich ein

Achtel (12,5%). Wie auch der jeweils geringe Überschneidungsgrad der Flächen der Leistungsverteilungen zeigt, können – vor allem in den Lesekompetenzen in Englisch – nur die leistungsstärksten NMS mit den leistungsschwächsten AHS-Unterstufen mithalten.

- (iv) Durch eine hohe Anzahl an verfügbaren Schulen in der Vergleichsgruppe der HS und eine ähnliche soziale Komposition der SchülerInnenchaft zwischen NMS und HS – bedeutsame Unterschiede traten ausschließlich in den migrationsbezogenen Merkmalen auf – konnte für die Realisierung eines „fairen“ Leistungsvergleichs zwischen diesen beiden Schulformen die Prozedur des PSM auf Basis der vier sozialen Einzelindikatoren durchgeführt werden. Auch die Unterschiede in den schulischen Kompetenzniveaus sind zwischen den beiden Schultypen allenfalls als moderat zu bezeichnen. Im Zuge des „naiven“ Vergleichs vor Kontrolle der sozialen Zusammensetzung der SchülerInnen erzielt die NMS in Mathematik einen Mittelwert von 496,7 Punkten, die HS von 515,0 Punkten. Nach dem PSM reduziert sich der Mittelwert der HS um 6,1 Punkte auf 508,9 Punkte. Es verbleibt eine Leistungsdifferenz von 12,2 Punkten zugunsten der HS, die aufgrund der geringen Fallzahl jedoch die Signifikanzschwelle von $p < 0,05$ verfehlt. In den Lesekompetenzen in Englisch ergibt sich vor dem PSM für die NMS ein Mittelwert von 484,2 Punkten und für die HS von 493,4 Punkten. Nach der Kontrolle der sozialen Zusammensetzung im Rahmen des „fairen“ Vergleichs erhöht sich der Mittelwert der HS marginal auf 496,4 Punkte. Beide Differenzen zwischen den Schulformen (vor und nach dem PSM) sind statistisch signifikant. Selbst bei vergleichbarer Komposition der SchülerInnenchaft verzeichnet die NMS im Vergleich zur HS folglich zumindest tendenzielle Defizite in den evaluierten Kompetenzbereichen.
- (v) NMS mit einem geringen relativen Leistungsabstand zur AHS-Unterstufe unterscheiden sich von leistungsfernen NMS bivariat hinsichtlich der sozialen Zusammensetzung, den kognitiven Fähigkeiten und dem Ausmaß an dissozialer Verhaltensweisen (Gewalt, abweichendes Verhalten) ihrer SchülerInnen sowie der Auftretenshäufigkeit von desorganisierten und anomischen Zuständen während des Unterrichts. Im Mittel besitzen leistungsnahe NMS somit einen günstigeren Sozialindex, verfügen bereits von Beginn an (5. Schulstufe) über SchülerInnen mit höheren kognitiven Fähigkeiten, diese treten zudem seltener durch dissoziale Verhaltensweisen in Erscheinung und die Organisation des Unterrichts gestaltet sich unproblematischer. Hinsichtlich der untersuchten Implementierungsmerkmale lassen sich bivariat keine signifikanten Unterschiede festmachen, d.h. es scheint keine Rolle zu spielen, wie hoch der Anteil an Oberstufenlehrkräften ist, ob eine Nachmittagsbetreuung eingerichtet wurde, wie intensiv verschränktes Teamteaching praktiziert wird, ob individuelle Förderpläne erstellt werden oder welche konkreten Lernformen umgesetzt werden.
- (vi) Multivariat erweist sich mit dem kognitiven Ausgangsniveau der SchülerInnen lediglich ein Merkmal als bedeutsam, um einen Beitrag zur Aufklärung der Gruppenzugehörigkeit zu den leistungsnahe sowie leistungsfernen NMS zu leisten. Unter der Annahme, dass die kognitiven Fähigkeiten nur teilweise durch schul- sowie unterrichtsbezogene Maßnahmen kompensiert werden können, lässt sich festhalten, dass deren Förderung in der NMS derzeit nicht in ausreichendem Maße gelingt. Umgekehrt verweist dies allerdings auch auf Grenzen der Fördermöglichkeiten in der NMS.
- (vii) Abschließend muss nochmals explizit und in aller Deutlichkeit darauf hingewiesen werden, dass die NMS der Generationen 1 und 2 keineswegs die „gemeinsame

Schule aller 10 bis 14-Jährigen“ repräsentiert. Vielmehr stellt die SchülerInnen-schaft in der NMS eine sozial selektive Population aller in Österreich lebenden Kinder bzw. Jugendlichen dieser Altersgruppe dar, die insbesondere durch eine deutliche Überrepräsentanz von Kindern bzw. Jugendlichen mit Migrationshintergrund und nichtdeutscher Alltagssprache charakterisiert werden kann.

Aufgrund der Ergebnisse der auch in anderen Beiträgen ermittelten Bedeutung der kognitiven Fähigkeiten und damit der Intelligenz empfehlen wir eine aktive Auseinandersetzung mit diesem Thema sowohl auf Ebene der Politik als auch in der Bildungspraxis. Dieses Thema ist in Österreich bislang weitgehend tabuisiert. Durch die Aufrechterhaltung dieser Tabuisierung besteht die Gefahr einer Naturalisierung des Phänomens, bei der die Intelligenz eines Kindes als genetisch gegeben und somit nicht veränderbar betrachtet wird. Dies entspricht nicht dem internationalen Forschungsstand (siehe z.B. zusammenfassend Hanscombe, Trzaskowski, Haworth, Davis, Dale & Plomin, 2012; Nisbett, Aronson, Blair, Dickens, Flynn, Halpern & Turkheimer, 2012; Rohde & Thompson, 2007).

Aus praxisorientierter Perspektive sollte das Ziel einer aktiven Auseinandersetzung mit dem Thema der Verbindungslinien zwischen den kognitiven Fähigkeiten (der allgemeinen Intelligenz) von SchülerInnen und deren Schulleistungen sein, dass ein entsprechendes Grundlagenwissen aufgebaut und aufbereitet wird, das als Fundament für die Entwicklung konkreter (im schulischen Alltag anwendbarer) pädagogischer und didaktischer Konzepte zur gezielten Förderung von Kindern mit geringeren kognitiven Ausgangsfähigkeiten herangezogen werden kann. Um herkunftsbedingten Unterschieden in den kognitiven Fähigkeitsniveaus bereits in ihrer Genese zu begegnen, sollte eine verstärkte systematische Förderung von Kindern aus benachteiligten Herkunftsfamilien bereits möglichst früh ansetzen. Empirische Evidenz für die Wirkung dieser spezifischen frühkindlichen Interventionen kann etwa von Belsky, Lowe Vandell, Burchinal, Clarke-Stewart, McCartney, Tresch Owen und dem NICHD Early Child Care Research Network (2007) vorgelegt werden. Weiterhin müssen diese Bemühungen ihre nahtlose Fortsetzung im Bereich der Primar- und Sekundarstufe des Schulsystems finden.

Literatur

- Allison, P. D. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods and Research*, 28 (2), 186–208.
- Bacher, J. (2002). Statistisches Matching: Anwendungsmöglichkeiten, Verfahren und ihre praktische Umsetzung in SPSS. *ZA-Informationen*, 51, 38–66.
- Bacher, J. (2003). Soziale Ungleichheit und Bildungspartizipation im weiterführenden Schulsystem Österreichs. *Österreichische Zeitschrift für Soziologie*, 28 (3), 3–32.
- Bacher, J. (2005). Bildungsungleichheit und Bildungsbenachteiligung im weiterführenden Schulsystem Österreichs. Eine Sekundäranalyse der PISA 2000-Erhebung. *SWS-Rundschau*, 45 (1), 37–62.
- Bacher, J. (2009). Soziale Ungleichheit, Schullaufbahn und Testleistungen. In B. Suchán, C. Wallner-Paschon & C. Schreiner (Hrsg.), *PIRLS 2006: Die Lesekompetenz am Ende der Volksschule – Österreichischer Expertenbericht* (S. 79–102). Graz: Leykam.
- Bacher, J. (2010). Bildungschancen von Kindern mit Migrationshintergrund. Ist-Situation, Ursachen und Maßnahmen. *WISO*, 33 (1), 30–50.

- Bacher, J., Leitgöb, H. & Weber, C. (2012). Bildungsungleichheiten in Österreich. Vertiefende Analyse der PISA2009-Daten. In F. Eder (Hrsg.), *PISA 2009. Nationale Zusatzanalysen für Österreich* (S. 432–456). Münster: Waxmann.
- Baumert, J.; Brunner, M.; Lüdtke, O. & Trautwein, U. (2007). Was messen internationale Schulleistungsstudien? – Resultate kumulativer Wissenserwerbsprozesse. Eine Antwort auf Heiner Rindermann. *Psychologische Rundschau*, 58 (2), 118–145.
- Baumert, J., Maaz, K. & Trautwein, U. (Hrsg.). (2009). *Bildungsentscheidungen*. Sonderheft 12 der Zeitschrift für Erziehungswissenschaft. Wiesbaden: VS Verlag.
- Baumert, J., Stanat, P. & Watermann, R. (Hrsg.) (2006a). *Herkunftsbedingte Disparitäten im Bildungswesen. Vertiefende Analysen im Rahmen von PISA 2000*. Wiesbaden: VS Verlag.
- Baumert, J., Stanat, P. & Watermann, R. (2006b). Schulstruktur und die Entstehung differentieller Lern- und Entwicklungsmilieus. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen. Vertiefende Analysen im Rahmen von PISA 2000* (S. 95–188). Wiesbaden: VS Verlag.
- Becker, R. (2000). Klassenlage und Bildungsentscheidungen. Eine empirische Prüfung der Wert-Erwartungs-Theorie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 52 (3), 450–474.
- Becker, R. (Hrsg.) (2009). *Lehrbuch der Bildungssoziologie*. Wiesbaden: VS Verlag.
- Becker, R. & Lauterbach, W. (Hrsg.) (2010). *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit*. Wiesbaden: VS Verlag.
- Becker, B. & Reimer, D. (Hrsg.) (2010). *Vom Kindergarten bis zur Hochschule. Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie*. Wiesbaden: VS Verlag.
- Becker, R. & Solga, H. (Hrsg.) (2012). *Soziologische Bildungsforschung*. Sonderheft 52 der Kölner Zeitschrift für Soziologie und Sozialpsychologie. Wiesbaden: VS Verlag.
- Becker, M., Lüdtke, O., Trautwein, U., Köller, O. & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: do academic-track schools make students smarter? *Journal of Educational Psychology*, 104 (3), 682–699.
- Belsky, J., Lowe Vandell, D., Burchinal, M., Clarke-Stewart, A. K., McCartney, K., Tresch Owen, M. & NICHD Early Child Care Research Network (2007). Are there long-term effects of early child care? *Child Development*, 78 (2), 681–701.
- Best, H. & Wolf, C. (2012). Modellvergleich und Ergebnisinterpretation in Logit- und Probit-Regressionen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 64 (2), 377–395.
- Boudon, R. (1974). *Education, opportunity, and social inequality. Changing prospects in western societies*. New York: Wiley & Sons.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten, Soziale Welt, Sonderband 2* (S. 183–198). Göttingen: Schwartz.
- Breen, R. & Goldthorpe, J. H. (1997). Explaining educational differentials: towards a formal rational action theory. *Rationality and Society*, 9, 275–305.
- Breit, S. (2009). Kompetenzen von SchülerInnen und Schülern mit Migrationshintergrund. In C. Schreiner & U. Schwantner (Hrsg.), *PISA 2006. Österreichischer Expertenbericht zum Naturwissenschaftsschwerpunkt* (S. 146–158). Graz: Leykam.
- Bruneforth, M., Weber, C. & Bacher, J. (2012). Chancengleichheit und garantiertes Bildungsminimum in Österreich. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 189–227). Graz: Leykam.

- Campbell, D. T. & Kenny, D. A. (1999). *A primer on regression artefacts*. New York: The Guilford Press.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences*. Mahwah, Lawrence Erlbaum Associates.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. (1966). *Equality of educational opportunity*. Washington DC: U.S. Government Printing Office.
- Cordeiro, G. M. & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B (Methodological)*, 53 (3), 629–643.
- Cramer, J. S. (2003). *Logit models. From economics and other fields*. Cambridge: Cambridge University Press.
- Deary, I. J.; Spinath, F. M. & Bates T. C. (2006). Genetics of intelligence. *European Journal of Human Genetics*, 14, 690–700.
- Deary, I. J.; Strand, S.; Smith, P. & Fernandes C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21.
- Diefenbach, H. (2007). *Kinder und Jugendliche aus Migrantenfamilien im deutschen Bildungssystem. Erklärungen und empirische Befunde*. Wiesbaden: VS Verlag.
- Eder, F. (2015a). Schulform-Unterschiede im Schul- und Klassenklima. In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 179–202). Salzburg & Linz.
- Eder, F. (2015b). Auswirkungen der Implementierungsqualität. In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 203–224). Salzburg & Linz.
- Eder, F., Altrichter, H., Hofmann, F. & Weber, C. (Hrsg.) (2015). *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten*. Salzburg & Linz.
- Eder, F. & Oberwimmer, K. (2015). Veränderungen in den Unterrichtsprozessen und im schulbezogenen Verhalten der Schülerinnen und Schüler. In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 113–164). Salzburg & Linz.
- Eder, F. & Springer, K. (2015). Positive und negative Schulerfahrungen im Vergleich – eine Analyse kritischer Schulerfahrungen auf der Sekundarstufe I. In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 225–240). Salzburg & Linz.
- Eder, F. & Svecnik, E. (2015). Konzept und methodische Anlage der NMS-Evaluierung. In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 77–96). Salzburg & Linz.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.
- Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14 (1), 153–158.
- Erikson, R. & Jonsson, J. (1996). Explaining class inequalities in education: the Swedish test case. In R. Erikson & J. Jonsson (Hrsg.), *Can education be equalized? The Swedish test case in comparative perspective* (S. 1–63). Stockholm: Westview Press.
- Esser, H. (1999). *Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt: Campus Verlag.
- Fiege, C., Reuther, F. & Nachtigall, C. (2011). Faire Vergleiche? Berücksichtigung von Kontextbedingungen des Lernens beim Vergleich von Testergebnissen aus deutschen Vergleichsarbeiten. *Zeitschrift für Bildungsforschung*, 1 (1), 133–150.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80 (1), 27–38.

- Gangl, M. (2010). Nichtparametrische Schätzung kausaler Effekte mittels Matchingverfahren. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 931–961). Wiesbaden: VS Verlag.
- Ganzeboom, H., de Graaf, P. M. & Treiman, D. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21 (1), 1–56.
- Geißler, R. (2014). *Die Sozialstruktur Deutschlands*. Wiesbaden: VS Verlag.
- Guo, S. & Fraser, M. W. (2010). *Propensity score analysis. Statistical methods and applications*. Thousand Oaks: Sage.
- Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S. & Plomin, R. (2007). Socioeconomic status (SES) and children's intelligence (IQ): in a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS ONE*, 7 (2), 1–16.
- Hart S. A., Soden, B., Johnson, W., Schatschneider, C. & Taylor, J. (2013). Expanding the environment: gene × school-level SES interaction on reading comprehension. *Journal Child Psychology and Psychiatry*, 54 (10), 1047–1055.
- Heckman, J. J. (1979). Sample selection bias as specification error. *Econometrica*, 47 (1), 153–161.
- Heckman, J. J. (2005). The scientific model of causality. *Sociological Methodology*, 35, 1–97.
- Heckman, J. J. & Robb, R. (1986). Alternative methods for solving the problem of selection bias in evaluating the impact of treatment on outcomes. In H. Wainer (Hrsg.), *Drawing inferences from self-selected samples* (S. 63–107). New York: Springer.
- Helm, C., Kemethofer, D., Altrichter, H. & Weber, C. (2015). Effekte der NMS-Konzeptmerkmale auf die fachlichen Schülerleistungen. In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 285–302). Salzburg & Linz.
- Kemethofer, D., Pieslinger C. & Helm, C. (2015). Beschreibung der Konzeptmerkmale der Neuen Mittelschule. In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 39–56). Salzburg & Linz.
- Helmke, A. (2007). *Was wissen wir über guten Unterricht?* Zugriff am 03.08.2014 unter http://www.bildung.koeln.de/imperia/md/content/selbst_schule/downloads/andreas_helmke_.pdf.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Helmke, A. & Schrader, F.-W. (2008). Merkmale der Unterrichtsqualität: Potential Reichweite und Grenzen. *SEMINAR – Lehrerbildung und Schule*, 3, 14–47.
- Herzog-Punzenberger, B. & Unterwurzacher, A. (2009). Migration – Interkulturalität – Mehrsprachigkeit. Erste Befunde für das österreichische Bildungswesen. In W. Specht (Hrsg.), *Nationaler Bildungsbericht Österreich 2009. Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 161–182). Graz: Leykam.
- Herzog-Punzenberger, B. & Schnell, P. (2012). Die Situation mehrsprachiger Schüler/innen im österreichischen Schulsystem – Problemlagen, Rahmenbedingungen und internationaler Vergleich. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 229–267). Graz: Leykam.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81 (396), 945–960.

- Horak, A., Moser, W., Nezbeda, M. & Schober, M. (2010). *Der Gemeinsame europäische Referenzrahmen für Sprachen in der Unterrichtspraxis*. ÖSZ Praxisreihe, Heft 12. Graz: ÖSZ.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, Series A (Mathematical and Physical Sciences)*, 186 (1007), 453–461.
- Karlson, K. B., Holm, A. & Breen, R. (2012). Comparing regression coefficients between same-sample nested models using logit and probit: a new method. *Sociological Methodology*, 42, 286–313.
- King, G. & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9 (2), 137–163.
- Kleine, L., Paulus, W. & Blossfeld, H.-P. (2009). Die Formation elterlicher Bildungsentscheidungen beim Übergang von der Grundschule in die Sekundarstufe I. In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen, Zeitschrift für Erziehungswissenschaft, Sonderband 12/2009* (S. 103–125). Wiesbaden: VS Verlag.
- Kohler, U., Karlson, K. B. & Holm, A. (2011). Comparing coefficients of nested nonlinear probability models. *The Stata Journal*, 11 (3), 420–438.
- Leitgöb, H. (2013). The problem of modeling rare events in ML-based logistic regression. Assessing potential remedies via MC simulations. *Vortrag gehalten auf der 5th Conference of the European Survey Research Association (ESRA) in Ljubljana*. Zugriff am 03.08.2014 unter http://www.europeansurveyresearch.org/conf/uploads/494/678/167/PresentationLeitg_b.pdf.
- Leitgöb, H., Bacher, J., Bruneforth, M. & Weber, C. (2014). Primäre und sekundäre Ungleichheitseffekte in maturaführenden Schulen in Österreich. *Erziehung & Unterricht, 1-2/2014*: 48–57.
- Leitgöb, H. & Lachmayr, N. (2012). Bildungspartizipation von Kindern und Jugendlichen mit Migrationshintergrund in Österreich. Eine empirische Prüfung auf primäre und sekundäre Effekte der ethnischen Herkunft. *Kontraste, 8/2012*: 42–54.
- Leitgöb, H., Paseka, A., Bacher, J. & Altrichter, H. (2012). Geschlechtsspezifische Leistungsdifferenzen in Österreich. Vertiefende Analysen der PISA2009-Daten. In F. Eder (Hrsg.), *PISA 2009. Nationale Zusatzanalysen für Österreich* (S. 308–336). Münster: Waxmann.
- Little, R. J. A. & Rubin, D. B. *Statistical analysis with missing data*. New York: Wiley & Sons.
- Maaz, K., Baumert, J. & Trautwein, U. (2009). Genese sozialer Ungleichheit im institutionellen Kontext der Schule. Wo entsteht und vergrößert sich soziale Ungleichheit? In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen, Zeitschrift für Erziehungswissenschaft, Sonderband 12/2009* (S. 11–46). Wiesbaden: VS Verlag.
- Maaz, K. & Nagy, G. (2009). Der Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems: Definition, Spezifikation und Qualifizierung primärer und sekundärer Herkunftseffekte. In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen, Zeitschrift für Erziehungswissenschaft, Sonderband 12/2009* (S. 153–182). Wiesbaden: VS Verlag.
- Maaz, K. & Neumann, M. & Baumert, J. (Hrsg.) (2014). *Herkunft und Bildungserfolg von der frühen Kindheit bis ins Erwachsenenalter. Forschungsstand und Interventionsmöglichkeiten aus interdisziplinärer Perspektive*. Sonderheft 24 der Zeitschrift für Erziehungswissenschaft. Wiesbaden: VS Verlag.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models*. London: Chapman & Hall.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Hrsg.) *Frontiers in Econometrics* (S. 105–142). New York: Academic Press.

- McGloin, J. M. & Pratt, T. C. (2003). Cognitive ability and delinquent behavior among inner-city youth: a life-course analysis of main, mediating, and interaction effects. *International Journal of Offender Therapy and Comparative Criminology*, 47 (3), 253–271.
- McGloin, J. M., Pratt, T. C. & Maahs, J. (2004). Rethinking the IQ-delinquency relationship: a longitudinal analysis of multiple theoretical models. *Justice Quarterly*, 21 (3), 603–635.
- Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26 (1), 67–82.
- Morgan, S. L. (Hrsg.). (2013). *Handbook of causal analysis for social research*. Dordrecht: Springer.
- Morgan, S. L. & Winship, C. (2007). *Counterfactuals and causal inference. Methods and principles for social research*. New York: Cambridge University Press.
- Müller-Benedict, V. (2007). Wodurch kann die soziale Ungleichheit des Schulerfolgs am stärksten verringert werden? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 59 (4), 615–639.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J. & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51 (2), 77–101.
- Nichols, A. (2007). Causal inference with observational data. *The Stata Journal*, 7 (4), 507–541.
- Nisbett, R. E. (2010). *Intelligence and how to get it: why schools and cultures count*. New York: W. W Norton & Company, Inc.
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F. & Turkheimer, E. (2012). Intelligence. New findings and theoretical developments. *American Psychologist*, 67 (2), 130–159.
- Oberwimmer, K. & Eder, F. (2015). Förderung von Schülerinnen und Schülern mit unterschiedlicher kognitiver Leistungsfähigkeit. In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 349–368). Salzburg & Linz.
- OECD (2012). *PISA 2009 technical report*. OECD Publishing. Zugriff am 03.08.2014 unter <http://www.oecd.org/pisa/pisaproducts/50036771.pdf>.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49 (12), 1373–1379.
- Pepe, M. S., Longton, G. & Janes, H. (2009). Estimation and comparison of receiver operating characteristic curves. *The Stata Journal*, 9 (1), 1–16.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? *Psychologische Rundschau*, 57 (1), 69–86.
- Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E. & Stanat, P. (2014). Socioeconomic and language minority classroom composition and individual reading achievement: the mediating role of institutional quality. *Learning and Instruction*, 32, 63–72.
- Rohe, T. E. & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35, 83–92.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41–55.
- Rosenbaum, P. R. & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39 (1), 33–38.

- Rumberger, R. W. & Palardy, G. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record*, 107 (9), 1999–2045.
- Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, 29 (1), 159–183.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66 (5), 688–701.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley & Sons.
- Schafer, J. L. & Kang, J. (2008). Average causal effects from nonrandomized studies: a practical guide and simulated example. *Psychological Methods*, 13 (4), 279–313.
- Schimpl-Neimanns, B. (2000). Soziale Herkunft und Bildungsbeteiligung. Empirische Analysen zur herkunftsspezifischen Bildungsungleichheiten zwischen 1950 und 1989. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 52 (4), 636–669.
- Schreiner, C. & Breit, S. (2006). Kompetenzen von Schüler/innen mit Migrationshintergrund. In G. Haider & C. Schreiner (Hrsg.), *Die PISA-Studie. Österreichs Schulsystem im internationalen Wettbewerb* (S. 179–192). Wien: Böhlau.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- Schreiner, C. & Breit, S. (Hrsg.) (2012). *Standardüberprüfung 2012 Mathematik, 8. Schulstufe. Bundesergebnisbericht*. Zugriff am 03.08.2014 unter https://www.bifie.at/system/files/dl/01_BiSt-UE_M8_2013_Bundesergebnisbericht_gesamt.pdf.
- Schreiner, C. & Breit, S. (Hrsg.) (2013). *Standardüberprüfung 2013 Englisch, 8. Schulstufe. Bundesergebnisbericht*. Zugriff am 03.08.2014 unter https://www.bifie.at/system/files/dl/01_BiSt-UE_E8_2013_Bundesergebnisbericht_gesamt.pdf.
- Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science*, 25 (1), 1–21.
- Stubbe, T. C. (2009). *Bildungsentscheidungen und sekundäre Herkunftseffekte. Soziale Disparitäten bei Hamburger Schülerinnen und Schülern der Sekundarstufe I*. Münster: Waxmann.
- Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, 14 (6), 623–628.
- van Ewijk, R. & Sleegers, P. (2010a). Peer ethnicity and achievement: a meta-analysis into the compositional effect. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 21 (3), 237–265.
- van Ewijk, R. & Sleegers, P. (2010b). The effect of peer socioeconomic status on student achievement: a meta-analysis. *Educational Research Review*, 5, 134–150.
- Vittinghoff, E. & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and cox regression. *American Journal of Epidemiology*, 165 (6), 710–718.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54 (3), 427–450.
- Weber, C., Bacher, J., Altrichter, H. & Leitgöb, H. (2015). Sind Leistungsverbesserungen an den NMS-Modellschulen zu beobachten? In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten* (S. 241–264). Salzburg & Linz.

Anhang

Tabelle A1: Basisergebnisse der Logit-Modelle zur Schätzung der Propensity Scores

Merkmal	AHS				HS			
	BIST-M8 (n=256)		BIST-E8 (n=438)		BIST-M8 (n=1.024)		BIST-E8 (n=1.045)	
	β	p	β	p	β	p	β	p
Konstante	-31,45	<0,001	-22,09	<0,001	-2,181	0,001	-0,482	0,057
geringe schulische Bildung der Eltern (\bar{x}_1)					3,089	0,094	-0,078	0,943
geringer beruflicher Status der Eltern (\bar{x}_2)					-3,416	0,004	-3,230	<0,001
Migrationshintergrund (\bar{x}_3) andere zu Hause gespro- chene Sprache (\bar{x}_4)					-1,700	0,527	1,421	0,267
Sozialindex (S)	26,769	<0,001	19,192	<0,001				
LL ^(a)	-78,10		-199,72		-233,52		-510,30	
Pseudo-R ² (b)	0,469		0,342		0,056		0,067	

^(a) LL ... Log-Likelihood

^(b) nach McFadden (1973)

Signifikante Effekte ($p < 0,05$) werden kursiv markiert

Tabelle A2: Ergebnisse der Logit-Modelle zur Erklärung der Zugehörigkeit zur Gruppe der NMS mit geringer relativer Leistungsdifferenz zur AHS-Unterstufe auf Basis der PML-Schätzung (Extremgruppenvergleich)

Merkmal	Modell 1		Modell 2	
	β (PML)	p	β (PML)	p
Konstante	5,909	0,013	-16,362	0,082
G-Indikator	0,016	0,976	-0,434	0,483
Sozialindex (S)	-0,046	0,013	0,008	0,770
verb. Fähigk. 5. Schulst. (KFT)	-	-	0,371	0,019
PLL ^(a)	-31,499		-25,717	

Merkmal	Modell 3		Modell 4	
	β (PML)	p	β (PML)	p
Konstante	-17,039	0,108	-9,187	0,431
G2-Indikator	-0,381	0,548	-0,640	0,348
Sozialindex (S)	0,010	0,751	0,016	0,635
verb. Fähigk. 5. Schulst. (KFT)	0,390	0,022	0,357	0,050
individueller Förderplan	-0,783	0,304	-0,850	0,278
abweichendes Verhalten	-	-	-0,041	0,967
Anomie im Unterricht	-	-	-2,520	0,134
räumliche Distanz zur nächst- gelegenen AHS-Unterstufe	-	-	-0,006	0,808
PLL	-21,431		-15,599	

^(a) PLL ... penalisierte Log-Likelihood

Signifikante Effekte ($p < 0,05$) werden kursiv markiert.

Tabelle A3: Nach der KHB-Methode reskalierte Koeffizienten der Logit-Modelle zur Erklärung der Zugehörigkeit zur Gruppe der NMS mit geringer relativer Leistungsdifferenz zur AHS-Unterstufe (Extremgruppenvergleich)

Merkmal	Modell 1		Modell 2	
	β (KHB)	p	β (KHB)	p
G2-Indikator	0,193	0,785	-0,475	0,518
Sozialindex (S)	-0,058	0,016	0,032	0,347
verb. Fähigk. 5. Schulst. (KFT)	-	-	0,550	0,005

Merkmal	Modell 3		Modell 4 ^(a)	
	β (KHB)	p	β (KHB)	p
G2-Indikator	-0,528	0,472	-0,741	0,326
Sozialindex (S)	0,022	0,538	0,017	0,640
verb. Fähigk. 5. Schulst. (KFT)	0,503	0,012	0,436	0,033
individueller Förderplan	-0,825	0,346	-1,028	0,241
abweichendes Verhalten	-	-	-0,013	0,991
Anomie im Unterricht	-	-	-3,031	0,106
räumliche Distanz zur nächstgelegenen AHS-Unterstufe	-	-	-0,008	0,782

^(a) Die Metrik der Koeffizienten der Modelle 1 bis 3 wurde an das Gesamtmodell (Modell 4) angepasst. Aus diesem Grund entsprechen die Koeffizienten aus Modell 4 auch jenen aus Tabelle 12.

Signifikante Effekte ($p < 0,05$) werden kursiv markiert.

Tabelle A4: Korrelationsmatrix der Modellvariablen aus dem Gesamtmodell (Modell 4)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	1							
(2)	0,031 (0,813)	1						
(3)	-0,361 (0,005)	-0,100 (0,445)	1					
(4)	0,467 (<i><0,001</i>)	0,292 (0,025)	-0,785 (<i><0,001</i>)	1				
(5)	-0,224 (0,100)	-0,100 (0,476)	-0,057 (0,679)	-0,125 (0,368)	1			
(6)	-0,407 (0,001)	-0,237 (0,069)	0,331 (0,010)	-0,423 (0,001)	0,038 (0,784)	1		
(7)	-0,365 (0,004)	-0,072 (0,587)	0,274 (0,034)	-0,258 (0,048)	-0,063 (0,650)	0,653 (<i><0,001</i>)	1	
(8)	0,195 (0,136)	0,071 (0,593)	-0,436 (0,001)	0,273 (0,036)	0,120 (0,383)	-0,317 (0,014)	-0,377 (0,003)	1

(1) ... dichotome abhängige Variable (Zugehörigkeit zur Gruppe der NMS mit geringem relativen Leistungsabstand zur AHS-Unterstufe); (2) ... G2-Indikator, (3) ... Sozialindex der Schule (S); (4) ... verbale kognitive Fähigkeiten 5. Schulstufe (KFT); (5) ... individueller Förderplan für SchülerInnen; (6) ... abweichendes Verhalten der SchülerInnen in der Schule; (7) ... Anomie im Unterricht; (8) ... räumliche Distanz zur nächstgelegenen AHS-Unterstufe

Die p-Werte werden unter den Koeffizienten in Klammern ausgewiesen. Signifikante Effekte ($p < 0,05$) werden kursiv markiert.