

Platzierungsentscheidungen beim Übergang in die gymnasiale Oberstufe – Folgen der Umstellung auf Zweigliedrigkeit des Sekundarschulsystems in Berlin

Jürgen Baumert  · Marko Neumann · Hanna Dumont · Michael Becker · Anna Bachsleitner · Kai Maaz · Olaf Köller

Online publiziert: 19. Juni 2019
© Der/die Autor(en) 2019

Zusammenfassung Ein zentraler Trend in der Entwicklung der Bildungsbeteiligung in Deutschland ist der kontinuierliche Anstieg der Quote der Hochschulzugangsberechtigten. Mit der Umstellung des Sekundarschulsystems auf Zweigliedrigkeit scheint diese Expansion neue Dynamik zu erhalten. Gleichzeitig wird das Problem der Standardsicherung und der Vergleichbarkeit von Berechtigungen sichtbar. In einem quasi-experimentellen Design untersucht der vorliegende Beitrag am Fall der flächendeckenden Einführung eines Zweisäulensystems in Berlin die Struktur von Platzierungsentscheidungen und die Einhaltung von Standards bei Öffnung

Prof. Dr. J. Baumert (✉)
Max-Planck-Institut für Bildungsforschung, Lentzeallee 94, 14195 Berlin, Deutschland
E-Mail: sekbaumert@mpib-berlin.mpg.de

Dr. M. Neumann · Dr. H. Dumont · Dr. M. Becker · A. Bachsleitner · Prof. Dr. K. Maaz
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation,
Rostocker Straße 6, 60323 Frankfurt am Main, Deutschland

Dr. M. Neumann
E-Mail: marko.neumann@dipf.de

Dr. H. Dumont
E-Mail: dumont@dipf.de

Dr. M. Becker
E-Mail: becker@dipf.de

A. Bachsleitner
E-Mail: bachsleitner@dipf.de

Prof. Dr. K. Maaz
E-Mail: maaz@dipf.de

Prof. Dr. O. Köller
IPN-Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik,
Olshausenstraße 62, 24118 Kiel, Deutschland
E-Mail: koeller@ipn.uni-kiel.de

des Zugangs zur gymnasialen Oberstufe. Mit der Strukturreform vollzog sich ein sprunghafter Anstieg der Quote der Oberstufenzugangsberechtigungen, der mit einer Verminderung von Leistungsanforderungen einherging. Die Platzierungsentscheidungen der Lehrkräfte beruhen auf der Integration unterschiedlicher Bewertungen. In die Entscheidungen gehen schulübergreifende und schulinterne Leistungsurteile sowie Einschätzungen der Lernmotivation und des Arbeits- und Schulverhaltens der Schüler/-innen ein. Mit der Liberalisierung des Oberstufenzugangs blieb die standortübergreifende Akkuratheit der Leistungsurteile trotz gleichzeitigen Absenkens der für den Übergang in die Oberstufe erforderlichen Leistungsvoraussetzungen stabil. Schulspezifische Maßstäbe gewannen zusätzliche Bedeutung. Die Relevanz von motivationalen Merkmalen änderte sich nicht. Die Überprüfung, inwieweit Platzierungsentscheidungen durch organisatorische und soziale Kontextbedingungen der Schulen beeinflusst werden, ergab ein differenziertes Bild.

Schlüsselwörter Leistungsstandards · Urteile von Lehrkräften · Zweigliedrigkeit · Gymnasiale Oberstufe · Hochschulreife

Teachers' placement decisions regarding the transition to upper secondary education: Consequences of the switch to a two-track secondary system in Berlin

Abstract A developing trend in participation in the educational system in Germany over time has been the steady increase in the proportion of students eligible to enter higher education. Since the switch to a two-track secondary system, this development seems to have taken on new momentum. At the same time, problems have emerged with respect to the maintenance of standards and comparability of qualifications. In a quasi-experimental design, this article uses the introduction of a two-track secondary system in Berlin to examine the nature of teachers' placement decisions, as well as maintenance of standards when opening up access to college-level programs. The structural reforms led to a jump in the proportion of students eligible for upper secondary education, associated with a drop in the required achievement level. Teachers' placement decisions integrate various judgments: assessments of student achievement, judgements of cognitive abilities, and evaluations of students' learning motivation and social behavior. Our results show that when access to the upper secondary track was opened up, the cross-school accuracy of teacher assessments of student achievement remained stable, despite the drop in the achievement level necessary to enter upper secondary education. School-specific measures became more relevant. There was no change in the relevance of motivational characteristics. Our analysis of the extent to which teachers' placement decisions were affected by the organizational and social context of the school revealed mixed findings.

Keywords Academic track · Achievement standards · Eligibility for higher education · Teacher judgments · Two-track system

1 Problemstellung

Für die Entwicklung des Sekundarschulsystems in Deutschland sind zwei langfristige Trends kennzeichnend. Dies ist einmal die Entkopplung von nominellem Bildungsgang und Schulabschluss, die maßgeblich zur Öffnung des gegliederten Schulsystems beigetragen hat (Baumert et al. 2003a; Schuchart 2006; Trautwein et al. 2008), und zum anderen die Einrichtung bzw. der Ausbau alternativer Zugangswege zur Hochschulreife außerhalb des Gymnasiums, die den Prozess der Entkopplung von Schulform und Abschluss unterstützen (Köller et al. 2004). Beide Entwicklungen haben den Anstieg der Quote der Hochschulzugangsberechtigten in Deutschland befördert (Autorengruppe Bildungsberichterstattung 2016). In der letzten Dekade scheint dieser Prozess infolge der in den meisten Bundesländern erfolgten oder sich vollziehenden Umstellung des Sekundarschulsystems auf Zweigliedrigkeit zusätzliche Dynamik zu erhalten, und zwar insbesondere dann, wenn neben dem Gymnasium ein zweiter, direkter Weg in eine gymnasiale Oberstufe und zur Hochschulreife angeboten wird (Holtappels 2017; Tillmann 2017).

In Deutschland erfüllen Schulabschlüsse eine doppelte Funktion: sie bescheinigen schulisch erworbene Qualifikationen und berechtigen gleichzeitig zum Besuch sich anschließender Bildungs- bzw. Ausbildungsgänge. Dies gilt sowohl für die allgemeine Hochschulreife als auch für den Mittleren Schulabschluss (inkl. des Erwerbs der Berechtigung zum Übergang in die gymnasiale Oberstufe) oder den Hauptschulabschluss bzw. die Berufsbildungsreife. Diese Verbindung von Abschlussqualifikation und Zugangsberechtigung geht von der stillschweigenden und in gewissem Maße kontrafaktischen Voraussetzung aus, dass die Vergabe von Abschlüssen nach vergleichbaren Kriterien erfolgt und die Vergabeinstitution die Gleichwertigkeit von Abschlüssen sowohl im Hinblick auf die Einhaltung von Standards als auch hinsichtlich der Äquivalenz der erteilten Noten garantiert (Baumert et al. 2003b, S. 322). Mit der Öffnung des Bildungssystems stellt sich demnach in meritokratischer Hinsicht verstärkt die Frage nach der Sicherung *gemeinsamer* Standards.

Die zum Schuljahr 2010/11 in Berlin durchgeführte flächendeckende Schulstrukturreform bietet eine Gelegenheit, diese Frage unter den Bedingungen der schulischen Infrastruktur einer Großstadt zu untersuchen. Die Berliner Schulstrukturreform schuf mit der Einführung der Zweigliedrigkeit ein Zweisäulensystem, in dem zwei parallele Sekundarschulformen – das Gymnasium und die Integrierte Sekundarschule (ISS) – mit unterschiedlichen curricularen Akzentsetzungen, unterschiedlicher Unterrichtsorganisation und in unterschiedlicher Zeittaktung zu allen Abschlüssen einschließlich der allgemeinen Hochschulreife führen. Mit dieser Reform verfolgte das Land Berlin unter anderem das Ziel, langfristig eine Erhöhung der Abiturientenquote zu erreichen (Abgeordnetenhaus Berlin 2009). Um den Übergang in die gymnasiale Oberstufe nach der 10. Jahrgangsstufe zu erleichtern und den Weg zur Hochschulreife auch in der zweiten Säule als einheitlichen Bildungsgang erkennbar zu machen, führen die ISS entweder eine eigene Oberstufe oder haben eine verpflichtende Kooperation mit einer gymnasialen Oberstufe im beruflichen oder allgemeinbildenden Schulsystem einzugehen.

Vor dem Übergang in die gymnasiale Oberstufe liegt jedoch am Ende der 10. Klasse eine formale Schwelle: Es muss eine *Übergangsberechtigung* erteilt werden, de-

ren Leistungsvoraussetzungen länderübergreifend abgestimmt sind (KMK 1993). Das deutsche Schulsystem kennt also nach dem Übergang von der Grundschule auf die weiterführenden Schulen am Ende der Sekundarstufe I eine weitere leistungsbewehrte Hürde, die mit der Einführung eines Zweisäulensystems zusätzliche Bedeutung erhält.

Bemerkenswerterweise ist dieser zweite Übergang zwar verschiedentlich im Hinblick auf soziale Disparitäten der Bildungsbeteiligung (Maaz et al. 2004, 2017; Hillmert und Jacob 2005, 2010; Trautwein et al. 2011; Combet 2013), aber kaum hinsichtlich der Struktur der Entscheidungsfindung und der Einhaltung von Leistungsanforderungen untersucht worden (Trautwein et al. 2008; Neumann et al. 2017a). Die jüngste Untersuchung von Neumann et al. (2017a), die im folgenden Literaturüberblick genauer dargestellt wird, analysiert auf der Grundlage einer quasi-experimentellen Begleitstudie zur Berliner Schulstrukturreform (Maaz et al. 2013; Neumann et al. 2017b) den mit der Systemumstellung verbundenen Anstieg der Quote der Übergangsberechtigungen und das erreichte Leistungsniveau der Übergangsberechtigten. An die deskriptiven Befunde dieser Untersuchung schließt der vorliegende Beitrag unter Nutzung desselben Datensatzes an. Im Folgenden sollen die mit der Schulstrukturreform verbundenen Veränderungen der Leistungsvoraussetzungen für den Zugang zur gymnasialen Oberstufe und potentielle Veränderungen der Struktur der Platzierungsentscheidungen bei Öffnung des Zugangs zur Hochschulreife multivariat untersucht werden. Im Zentrum der Untersuchung steht einerseits die Frage, welche Leistungen an nichtgymnasialen Schulen vor und nach der Schulstrukturreform sowie im Vergleich zum Gymnasium für den Erhalt der Berechtigung zum Übergang in die Oberstufe erbracht werden. Zum anderen wird in Anlehnung an vorhandene Untersuchungen zum Grundschulübergang danach gefragt, welche leistungsbezogenen, motivationalen und schulkontextuellen Merkmale sich als empirisch bedeutsam für die Übergangsberechtigung erweisen und inwieweit sich diesbezüglich Veränderungen in Folge der Schulstrukturreform ergeben haben.

2 Forschungslage

2.1 Öffnung von Bildungswegen zur Hochschulreife und die Sicherung von vergleichbaren Maßstäben bei Platzierungsentscheidungen

Im deutschen Berechtigungswesen ist mit einer Öffnung von Bildungswegen, die zur Hochschulreife führen, die Frage der Standardsicherung und der Äquivalenz der Notenvergabe verbunden. Für die allgemeine Hochschulreife liegt mittlerweile eine Reihe von Untersuchungen aus verschiedenen Bundesländern vor, die zeigen, dass beide Anforderungen nur eingeschränkt erfüllt werden. Köller et al. (1999) haben zum ersten Mal eklatante Leistungs- und Benotungsunterschiede an Oberstufen von Gymnasien und Gesamtschulen in Nordrhein-Westfalen nachgewiesen. Es folgten dann ähnliche Befunde aus Baden-Württemberg, Hamburg und Schleswig-Holstein (Köller et al. 2004; Trautwein et al. 2007b; Leucht et al. 2016). So lag etwa das mittlere mathematische Leistungsniveau der in der TOSCA-Studie (Köller

et al. 2004) untersuchten Abiturienten/-innen in Baden-Württemberg an beruflichen Gymnasien mit sozialpädagogischer Ausrichtung rund $d=1,3$ Standardabweichungen unter dem Leistungsniveau allgemeinbildender Gymnasien (Watermann et al. 2004). Auch an den anderen beruflichen Gymnasialzweigen (agrар-, ernährungs- und wirtschaftswissenschaftlich) fielen die Leistungen deutlich niedriger aus. Lediglich an den technischen Gymnasien wurde in Mathematik ein vergleichbares Leistungsniveau wie an den allgemeinbildenden Gymnasien erzielt. Dabei spiegelten sich die Leistungsunterschiede zwischen den Gymnasialzweigen kaum in entsprechenden Notenunterschieden wider (Watermann et al. 2004). In ähnlicher Weise konnten auch für Oberstufenschüler/-innen an Gesamtschulen in Hamburg Leistungsrückstände im Vergleich zu den allgemeinbildenden Gymnasien nachgewiesen werden (Trautwein et al. 2007b). Dieses Bild wiederholte sich auch in Schleswig-Holstein beim Vergleich der Abiturleistungen an allgemeinbildenden und beruflichen Gymnasien (Leucht et al. 2016). Die großen Leistungsunterschiede sind im Wesentlichen Folge der unterschiedlichen Eingangsselektivität beim Übergang in die Oberstufe. Sie verdeutlichen jedoch die Spannung zwischen der Öffnung von Bildungswegen einerseits und der Sicherung von gemeinsamen Leistungsstandards und einer auf meritokratischen Prinzipien beruhenden Verteilung von Zertifikaten andererseits.

Dies gilt nicht erst für den Erwerb des Abiturs, sondern in vergleichbarer Weise bereits für die Erteilung der Berechtigung zum Übergang in die gymnasiale Oberstufe. Trautwein et al. (2008) untersuchten an deutschsprachigen Schulen des Kantons Freiburg (Schweiz) den Übergang von Sekundarschulen einerseits und dem Progymnasium andererseits auf die gymnasiale Oberstufe. Anhand des Vergleichs von Testleistungen, die allerdings bereits am Ende der Grundschule und damit mehrere Schuljahre vor dem Übergang in die Oberstufe erhoben wurden, kommen sie zu der Vermutung, dass sich das Leistungsniveau von Übergängern aus beiden Schulformen systematisch unterscheidet und mit der Öffnung der Bildungswege unterschiedliche Leistungsstandards zur Anwendung kommen könnten. Einen ähnlichen Befund berichten Trautwein et al. (2011) in ihrer Untersuchung zur Öffnung des Sekundarschulsystems in Baden-Württemberg. Sie fanden, dass das mittlere Niveau der kognitiven Grundfähigkeiten von Schülern/-innen mit einer Übergangsberechtigung für die gymnasiale Oberstufe an Realschulen um ungefähr eine halbe Standardabweichung unter dem entsprechenden Wert übergangsberechtigter Gymnasiasten/-innen lag (Tab. 2). In jüngster Zeit haben Neumann et al. (2017a) auf der Basis des auch der vorliegenden Arbeit zugrunde liegenden Datensatzes die Vergabe der Übergangsberechtigungen für die gymnasiale Oberstufe an nichtgymnasialen Schulformen vor und nach der Schulstrukturreform in Berlin untersucht. Sie belegten einen abrupten Anstieg der Berechtigungsquoten um 17,6 Prozentpunkte nach der Strukturreform und ein Absinken der mittleren Leistungen der Übergangsberechtigten je nach Fach um $d=-0,13$ (Englisch), $d=-0,17$ (Mathematik) und $d=-0,27$ (Lesekompetenz Deutsch) Standardabweichungen. An Schulen ohne gymnasiale Oberstufe schien der Leistungsrückgang besonders ausgeprägt zu sein. Ergänzend berichten sie auf der Basis explorativer Analysen die Anteile der Übergangsberechtigten, die an ISS in allen untersuchten Kernfächern unter dem Leistungsniveau des 8. Perzentils an Gymnasien blieben. Dieses Leistungsniveau betrachteten sie als gymnasiale Mindesterwartung. Die Ergebnisse zeigten, dass die angesetzten Leistungserwartun-

gen von erheblichen Anteilen (je nach Modellierung zwischen 27 und 43 %) der übergangsberechtigten Schülerinnen und Schülern an den ISS nicht erreicht wurden. Zudem wurde deutlich, dass die Notenvergabe in den Erweiterungskursen der ISS liberaler als an Gymnasien erfolgte. Der Zusammenhang zwischen Testleistung und der Wahrscheinlichkeit, eine Übergangsberechtigung zu erhalten, und die Lokalisation der für die Erteilung der Übergangsberechtigung kritischen Leistungsschwelle wurden ebenso wie die Urteilskomponenten der Platzierungsentscheidung nicht untersucht.

2.2 Leistungsurteile und Platzierungsentscheidungen von Lehrkräften

Die Erteilung einer Berechtigung für den Übergang in die gymnasiale Oberstufe beruht – wie auch die Schulformempfehlung am Ende der Grundschule – *formal* auf kollegial abgestimmten Urteilen von Lehrerinnen und Lehrern. Die endgültige Entscheidung fällt in der Klassen- oder Jahrgangskonferenz. In beiden Fällen beeinflussen oder entscheiden die Platzierungsurteile Schulkarrieren. Am besten untersucht sind Platzierungsentscheidungen beim Übergang von der Grundschule in die Sekundarschule in gegliederten Schulsystemen. Überraschenderweise sind vergleichbare Untersuchungen zum Übergang in die Sekundarstufe II selten. Wir vermuten, dass Platzierungsurteile von Lehrkräften an beiden Übergangsschwellen eine ähnliche *Struktur* aufweisen, auch wenn sich die diagnostische Kompetenz von Lehrkräften unterschiedlicher Schulformen möglicherweise unterscheidet (Karing 2009). Deshalb berichten wir zunächst die Befundlage für Übergangsempfehlungen am Ende der Grundschulzeit.

Mittlerweile liegen verschiedene qualitative Studien vor, die über die Entscheidungskriterien, die Lehrkräfte beim Übergang von der Grundschule in weiterführende Schulen nach eigenen Angaben anwenden, Auskunft geben (Nölle et al. 2009; Pohlmann 2009; Böhmer et al. 2015). Diese Untersuchungen belegen, dass Lehrkräfte in erster Linie Leistungsinformationen (aktueller Leistungsstatus, Leistungsentwicklung, aber auch die Einschätzung der kognitiven Fähigkeiten), leistungsbezogene motivationale Merkmale (Lernmotivation, Interesse, Anstrengungsbereitschaft), das Arbeitsverhalten und selbstregulative Fähigkeiten (Aufmerksamkeit, Konzentrationsfähigkeit, Verlässlichkeit, emotionale Stabilität), das Sozialverhalten (Hilfsbereitschaft, Kooperationsfähigkeit, allgemeines Schulverhalten) sowie potenzielle Unterstützung in der Familie und den Elternwunsch berücksichtigen. Böhmer et al. (2015) konnten in einer experimentellen Studie zum Abrufverhalten von Informationen zeigen, dass leistungsbezogene Informationen einschließlich der Informationen zum Arbeitsverhalten zuerst und am häufigsten abgerufen werden, gefolgt von Informationen zum Sozialverhalten und zur elterlichen Unterstützung. Direkte Informationen zur sozialen und ethnischen Herkunft wurden an letzter Stelle herangezogen.

Welche dieser von Lehrkräften für relevant gehaltenen Urteilkriterien die Platzierungsentscheidung bestimmen, wurde in verschiedenen quantitativen Untersuchungen zum Übergang nach der Grundschule überprüft (vgl. zusammenfassend Dumont et al. 2014; Klapproth et al. 2013). In diesen Analysen erwiesen sich leistungsbezogene Informationen (Noten, Testleistungen, kognitive Grundfähigkeiten) als die wichtigsten Prädiktoren (Bos et al. 2004; Baeriswyl et al. 2006; Arnold et al. 2007;

Driessen et al. 2008; de Boer et al. 2010; Neugebauer 2011; Schneider 2011; Stubbe et al. 2012; Dumont et al. 2013a; Klapproth et al. 2013). Es ist bemerkenswert, dass Testleistungen und kognitive Grundfähigkeiten – also Informationen, die den Lehrkräften in der Regel nicht direkt zur Verfügung stehen – über die Schulnoten hinaus bedeutsame spezifische Erklärungsbeiträge für Platzierungsurteile leisten. Für die kognitiven Grundfähigkeiten gilt dies auch bei Kontrolle von Noten *und* Testleistungen (Bos et al. 2004; Arnold et al. 2007; Driessen et al. 2008; de Boer et al. 2010; Stubbe et al. 2012). Die Effektstärken sind allerdings klein. Diese Befunde weisen darauf hin, dass in die Urteilsfindung von Lehrkräften durchaus klassen- und schulübergreifende Leistungsgesichtspunkte und Einschätzungen der allgemeinen kognitiven Befähigung von Schülerinnen und Schülern eingehen, die *nicht* schon in den Noten zum Ausdruck kommen. Ergänzend hinzu treten das Arbeitsverhalten, motivationale Orientierungen und schulrelevante Persönlichkeitsmerkmale der Schüler/-innen sowie die potenzielle Unterstützung im Elternhaus und der Schulwunsch der Eltern. Soziales Verhalten, das in zwei Untersuchungen Berücksichtigung fand, erwies sich als nicht prädiktiv.

2.3 Akkuratheit von Leistungsurteilen und Platzierungsentscheidungen

Platzierungsentscheidungen beruhen primär, aber nicht ausschließlich auf leistungsbezogenen Urteilen. Die Genauigkeit vor allem von Leistungsurteilen gilt als Indikator für die diagnostische Kompetenz von Lehrkräften (Spinath 2005; Artelt und Rausch 2014). Schrader und Helmke (1987) haben diesbezüglich drei Urteils-komponenten unterschieden: Die Niveauebene, die sich auf die Tendenz zur systematischen Über- oder Unterschätzung von Schülermerkmalen bezieht, die Differenzierungskomponente, die die korrekte Einschätzung von Merkmalsstreuungen betrifft, und die Rangkomponente, mit der erfasst wird, inwieweit Lehrkräfte die Rangfolge von Schülerinnen und Schülern hinsichtlich der Ausprägung eines Merkmals korrekt einschätzen. Die Rangkomponente gilt in der Literatur als „zentraler Akkuratheitsindikator [...] und als Maß für die diagnostische Sensitivität im eigentlichen Sinne“ (Spinath 2005, S. 86). Wenn im Folgenden im Anschluss an die übliche Begriffsverwendung ohne weitere Spezifikation von Akkuratheit gesprochen wird, ist diese Urteilskomponente gemeint.

Nimmt man die Korrelation zwischen Lehrerurteil und objektiv gemessenen Testleistungen als Maß für Urteilsakkuratheit, kommt die Mehrzahl der einschlägigen Untersuchungen zu dem Ergebnis, dass Lehrkräfte im Mittel gute Diagnostiker von Schülerleistungen sind. Südkamp et al. (2012) berichten in ihrer Metaanalyse eine mittlere Korrelation von $r=0,63$. Die Befunde einer früheren Metaanalyse von Hoge und Coladarci (1989) liegen in ähnlicher Größenordnung. Wie Südkamp et al. (2012) weiterhin zeigen, fällt die Akkuratheit bei informierten Urteilen mit einer mittleren Korrelation von $r=0,76$ höher aus als bei Urteilen mit schwacher Informationsbasis ($r=0,61$). Diese Befunde sind für die vorliegende Untersuchung bedeutungsvoll. Denn Entscheidungen über Laufbahneempfehlungen oder die Vergabe von Zugangszertifikaten beruhen auf informierten und reflektierten Entscheidungen, die unter Einbeziehung eines breiten Kranzes von Leistungsinformationen getroffen werden.

Spezielle Untersuchungen zur Akkuratheit von Platzierungsentscheidungen im Sinne der Rangkomponente der Urteilgenauigkeit von Lehrkräften (Korrelation zwischen *vorgängig* gemessener Testleistung und Entscheidung) liegen mit Ausnahme der neuen Arbeit von Steinmayr et al. (2017), die für die kognitiven Grundfähigkeiten eine Korrelation von $r=0,43$ und für die Lese- bzw. Mathematikleistung Korrelationen von $r=0,51$ bzw. $r=0,48$ berichten, unseres Wissens nicht vor. Die Frage spielt jedoch bei allen Untersuchungen zu den Determinanten von Platzierungsurteilen, über die in Abschn. 2.2 berichtet wurde, eine Rolle. Hier gehören objektive Testleistungen zu den stärksten Prädiktoren von Platzierungsentscheidungen. Im Rahmen der Untersuchung der prognostischen Validität von Grundschulempfehlungen sind drei Arbeiten zu finden, die den Zusammenhang zwischen Empfehlung und *später* erreichten Testleistungen thematisieren.¹ Für Hamburg (KESS-Studie) berichten Ivanov und Nikolova (2010) Leistungsunterschiede in Orthografie und Mathematik zwischen Gymnasiasten/-innen mit und ohne entsprechender Empfehlung von $d=0,87$ in der 7. Jahrgangsstufe. Dies entspricht einer punktbiserialen Korrelation zwischen Übergangsempfehlung und Testleistung von ungefähr $r=0,40$. In Bayern und Hessen (BIKS-Längsschnitt) fanden Pfost et al. (2017) in einer vergleichbaren Untersuchung in Mathematik Leistungsunterschiede von $d=0,57$ und im Lesen von $d=0,37$. Die entsprechenden Korrelationen betragen $r=0,27$ bzw. $r=0,18$. Einen engeren Zusammenhang berichten Klapproth et al. (2013) für Luxemburg. Sie teilten Schüler/-innen aufgrund von Testwerten, die in der 9. Jahrgangsstufe in den Fächern Deutsch, Französisch und Mathematik erhoben wurden, am Schnittpunkt der Leistungsverteilungen der akademisch bzw. beruflich orientierten Schulformen in zwei Gruppen (dies entspricht etwa einem Mediansplit) und verglichen die Zuordnung mit den Übertrittsempfehlungen der Grundschullehrkräfte am Ende der 6. Jahrgangsstufe. Dabei kommen sie auf Fehlklassifikationen von 21 %. Nach dem von Rosenthal und Rubin (1982) vorgeschlagenen binomialen Effektstärkenmaß entspricht dies einer punktbiserialen Korrelation von $r=0,58$. Mit Ausnahme der Luxemburger Untersuchung gilt für diese Arbeiten jedoch die Einschränkung, dass falsch-negative Entscheidungen nicht berücksichtigt werden konnten und sich die Angaben damit nur auf eine Schulform beziehen.

2.4 Soziale Kontexte, Leistungsurteile und Platzierungsentscheidungen von Lehrkräften

Leistungsurteile von Lehrkräften sind auch von sozialen Kontexten abhängig. Dies gilt gleichermaßen für Notenvergabe und Platzierungsentscheidungen. Untersucht wurde vor allem, inwieweit die leistungsmäßige, soziale und ethnische Zusammensetzung von Schulklassen oder Schulen das Niveau der Leistungsurteile von Lehrkräften beeinflusst. Dagegen gibt es kaum Arbeiten, die die Auswirkungen von Kompositionsmerkmalen auf die Akkuratheit der Urteile von Lehrkräften im Sinne

¹ Als weitere Validitätskriterien für die Grundschulempfehlung wurden der Verbleib auf bzw. der Abschluss an der nach der Grundschule besuchten Schulform (Roeder 1997; Tiedemann und Billman-Machecha 2010) und die in der Sekundarstufe erreichten Noten (Baeriswyl et al. 2009; Neuenschwander und Malti 2009) herangezogen.

der Rangkomponente untersuchen – eine Frage, die für den vorliegenden Beitrag besonders relevant ist.

Übersichten über die Forschung zu Niveaueffekten sind bei Neumann et al. (2010), Dumont et al. (2013b) und Westphal et al. (2016) zu finden. Die Befundlage ist im Hinblick auf die leistungsmäßige Komposition von Referenzgruppen international und national relativ eindeutig: Es lässt sich regelmäßig ein Referenzgruppeneffekt nachweisen, bei dem die Anwendung eines gruppeninternen Maßstabs dazu führt, dass bei variierenden mittleren Leistungen von Gruppen die individuellen Urteile mit steigendem Leistungsniveau einer Gruppe im Vergleich zu gleichen Leistungen in schwächeren Gruppen strenger ausfallen. Der Referenzgruppeneffekt lässt sich sowohl bei der Notenvergabe als auch bei Bildungsgangsempfehlungen konsistent nachweisen (schon Marsh 1987, S. 288; Dompnier et al. 2006; Trautwein et al. 2007a; Tiedemann und Billmann-Mahecha 2007; Maaz et al. 2008; Caro et al. 2009; Gröhlich und Guill 2009; Wagner et al. 2009; Hochweber et al. 2014; Westphal et al. 2016; Meissel et al. 2017). Südkamp und Möller (2009) konnten diesen Effekt auch experimentell in ihrem simulierten Klassenzimmer zeigen.

Abweichende Befunde fanden Ready und Wright (2011) in der Kindergartenkohorte der *Early Childhood Longitudinal Study-K* (ECLS-K). Ihre Auswertungen ergaben, dass mit steigendem Leistungsniveau einer Lerngruppe die Sprachentwicklung der Kinder *überschätzt* und bei sinkendem Leistungsniveau dementsprechend *unterschätzt* wird. Sie schließen auf einen Generalisierungseffekt.

Im Hinblick auf die soziale und ethnische Komposition sind die Befunde gemischt. In einer Reihe von Untersuchungen trat im Hinblick auf die soziale Zusammensetzung der jeweiligen Referenzgruppe ein Generalisierungseffekt auf: Mit steigendem mittleren Sozialstatus der Bezugsgruppe fielen die Leistungsurteile der Lehrkräfte positiver aus. Dies gilt sowohl für die Vergabe von Einzelnoten als auch für Platzierungsentscheidungen (Maaz et al. 2008; Gröhlich und Guill 2009; Schulze et al. 2009; Ready und Wright 2011; Hochweber et al. 2014; Westphal et al. 2016). Keinen Effekt der auf Klassenebene aggregierten Bildungsorientierung der Eltern bzw. ihres kulturellen Besitzes fanden Tiedemann und Billmann-Mahecha (2007) sowie Wagner et al. (2009).

Im Hinblick auf die ethnische Komposition von Schulklassen liegen widersprüchliche Ergebnisse vor. Tiedemann und Billmann-Mahecha (2007), Schulze et al. (2009), Ready und Wright (2011) fanden keinen ethnischen Kompositionseffekt, während Caro et al. (2009) bei einer Reanalyse der ELEMENT-Daten einen positiven Effekt feststellen konnten. Martínez et al. (2009) berichten inkonsistente Befunde.

Von besonderem Interesse sind die wenigen Untersuchungen, die den Zusammenhang zwischen Komposition der Lerngruppe und der Akkuratheit von Leistungsurteilen analysieren. Martínez et al. (2009) konnten für Dritt- und Fünftklässler der ECLS-K-Untersuchung zeigen, dass mit steigendem Anteil von Schülerinnen und Schülern mit geringen Englischkenntnissen (LEP), die Akkuratheit der Leistungsbeurteilungen zurückging. Die einzige in Deutschland zu diesem Thema vorliegende Untersuchung von Hochweber et al. (2014) berichtet einen nichtlinearen positiven Zusammenhang zwischen der leistungsmäßigen Komposition einer Schulklasse und der Akkuratheit der Notenvergabe im Fach Mathematik. Das mittlere Bildungsni-

veau der Eltern einer Schulklasse veränderte die Akkuratheit der Noten dagegen nicht.

3 Entwicklung und Differenzierung der Fragestellungen

Im Anschluss an die deskriptiven Befunde von Neumann et al. (2017a) sollen im vorliegenden Beitrag die mit der Berliner Schulstrukturreform verbundenen systematischen Verschiebungen der kritischen Leistungsschwelle für den Übertritt in die gymnasiale Oberstufe und die Struktur der Platzierungsentscheidungen und ihre Veränderungen multivariat untersucht werden. Im Folgenden werden die Fragestellungen konkretisiert.

3.1 Komponenten der Entscheidung über die Übergangsberechtigung für die gymnasiale Oberstufe

Die Erteilung der Übergangsberechtigung für die gymnasiale Oberstufe beruht formal auf einer kollektiven Entscheidung der Klassen- oder Jahrgangskonferenz. Im Unterschied zur Übergangsempfehlung am Ende der Grundschulzeit, mit der die Passung zwischen individuellem Entwicklungsstand und den Anforderungen der weiterführenden Schulen optimiert werden soll, hat die Übergangsberechtigung für die gymnasiale Oberstufe in erster Linie qualitätssichernde Funktionen. Notwendige Voraussetzung für die Erteilung einer Übergangsberechtigung ist der Mittlere Schulabschluss (MSA). Qualifizierende Bedingungen sind am Gymnasium die Versetzung in die 11. Jahrgangsstufe und an Schulen mit mehreren Bildungsgängen bestimmte Kombinationen von Kursen mit erhöhtem Anspruchsniveau und an allen Schulformen die Erfüllung von Mindestvoraussetzungen in den Fachnoten. Die detaillierten und komplizierten Übertrittsregelungen vor und nach der Schulstrukturreform sind im Anhang ausgewiesen (vgl. ausführlich auch Neumann et al. 2017a).

Die Ermessensspielräume der Lehrkräfte sind am Ende der Sekundarstufe I im Vergleich zur Grundschulempfehlung äußerst begrenzt. Werden die formalen Voraussetzungen erfüllt, muss die Übergangsberechtigung erteilt werden. Ein Ermessen spielt zu diesem Zeitpunkt nur noch in Grenzfällen eine Rolle. Die eigentliche *pädagogische* Entscheidung über den Übertritt zur gymnasialen Oberstufe wird an Schulen mit mehreren Bildungsgängen und damit an allen ISS in Berlin bereits am Ende der 9. Jahrgangsstufe vorbereitet, wenn über die Zuweisung zu den für die Abschlüsse notwendigen Kursen bzw. Kursniveaus der 10. Jahrgangsstufe und damit auch über die ungefähr zu erwartenden Noten entschieden wird. Wenn man also nach den Selektionsmechanismen an der Schwelle zur gymnasialen Oberstufe an nichtgymnasialen Schulen fragen will, empfiehlt es sich, bereits die Situation beim Übergang in die 10. Jahrgangsstufe in den Blick zu nehmen. Bei der Vergabe von Übergangsberechtigungen in die gymnasiale Oberstufe handelt sich um einen Entscheidungsprozess, der sich über ein Schuljahr erstreckt und in der formalen Entscheidung seinen Abschluss findet. Wir erwarten, dass in diesem Entscheidungsprozess weitgehend dieselben *expliziten und impliziten* Urteilkriterien zur Anwendung kommen, die bei der Vergabe der Grundschulempfehlung identifiziert werden konn-

ten (vgl. Abschn. 2.2). Vor diesem Hintergrund untersuchen wir im vorliegenden Beitrag, inwiefern sich der Erhalt der Übergangsberechtigung für die gymnasiale Oberstufe durch folgende Faktoren vorhersagen lässt:

1. Schulübergreifende Leistungsstandards, die sich über den Zusammenhang zwischen objektiven Leistungs- und Fähigkeitstests und Berechtigungsvergabe beschreiben lassen,
2. schulinterne, auf den Jahrgang bezogene Leistungsgesichtspunkte, die in der Rangordnung der Noten zum Ausdruck kommen, sowie
3. motivationale Merkmale und das Arbeits- und Schulverhalten der Schüler/-innen.

3.2 Erteilung von Übergangsberechtigungen am Gymnasium und an nichtgymnasialen Schulen: Zusammenhang zwischen Leistung und Platzierungsentscheidung

Der Idee nach ist das Gymnasium ein einheitlicher, akademisch orientierter Bildungsgang, der vom Abschluss der Grundschule bis zum Abitur führt, auch wenn im Gymnasium alle anderen Schulabschlüsse erreicht werden können. Die Übergangsauslese am Ende der Grundschulzeit soll sicherstellen, dass dieses Ziel auch von der großen Mehrheit der Gymnasiasten/-innen erreicht werden kann. In diesem Rahmen ist die Erteilung der Übergangsberechtigung für die gymnasiale Oberstufe eine formale Hürde, deren Überwindung keine besonderen Probleme bereiten sollte. In einer ganz anderen Situation befinden sich Schülerinnen und Schüler an nichtgymnasialen Schulen, deren modales Schulziel nicht das Abitur, sondern der MSA ist. Hier ist die Erteilung der Übergangsberechtigung eine Selektionsschwelle, die nur eine Minderheit überwinden will und kann. Dementsprechend lagen die Basisraten der Übergangsberechtigungen in Berlin im untersuchten Jahrgang vor der Schulstrukturreform im Schuljahr 2011/12 im Gymnasium bei rund 96 % und an nichtgymnasialen Schulen bei 24 % (vgl. Neumann et al. 2017d).

Die Ergebnisse von Neumann et al. (2017a, 2017d) zeigen nach der Schulstrukturreform weitgehende Stabilität am Gymnasium und einen Anstieg der Quote der Übergangsberechtigten an nichtgymnasialen Schulen auf 41 %, der an Schulen mit eigener Oberstufe besonders ausgeprägt zu sein scheint. Gleichzeitig sinken die mittleren Leistungen der Übergangsberechtigten je nach Fach zwischen gut einer zehntel und knapp einer drittel Standardabweichung. Dies spricht für eine mit der Öffnung der Oberstufe verbundene Absenkung von Leistungsanforderungen, ohne dass der Nachweis explizit geführt würde. Im Anschluss an diese deskriptiven Befunde soll im vorliegenden Beitrag der Zusammenhang zwischen Fachleistungen und Platzierungsentscheidung systematisch analysiert werden.

Aufgrund des höheren Leistungsniveaus und der größeren Leistungshomogenität sind die Übergangsberechtigungen an Gymnasien extrem schief verteilt: In der Regel wird die kritische Hürde übersprungen. Daraus folgt notwendigerweise, dass die Fachleistungen in den Kernfächern an Gymnasien am Ende der 9. Jahrgangsstufe weit weniger entscheidungsprädiktiv sind als im nichtgymnasialen Bereich. Bei sich überlappenden Leistungsverteilungen zwischen den Schulformen heißt dies allerdings auch, dass leistungsschwächere Gymnasiasten/-innen, deren Gesamtleis-

tung im Leistungsbereich von Schülern/-innen an nichtgymnasialen Schulen liegt, eine höhere Wahrscheinlichkeit besitzen sollten, eine Übergangsberechtigung für die Oberstufe zu erhalten, als Schüler/-innen gleicher Leistungsstärke an nichtgymnasialen Schulen.

Die Befunde von Neumann et al. (2017a) sprechen für sinkende Leistungsanforderungen bei Öffnung des Zugangs zur gymnasialen Oberstufe. Um dies explizit zu überprüfen und das Ausmaß der Absenkung der Zugangsvoraussetzungen zu quantifizieren, sollen kritische Leistungsschwellen vor und nach der Strukturreform verglichen werden. Als kritische Leistungsschwelle für die Erteilung der Übergangsberechtigung zur Oberstufe betrachten wir das Gesamtleistungsniveau in den übergangsrelevanten Kernfächern, bei dem die Wahrscheinlichkeit, eine Übergangsberechtigung zu erhalten, 50 % überschreitet. Die Differenz zwischen den Leistungsniveaus ist ein Maß für die Verschiebung von Zugangsanforderungen vor und nach Schulstrukturreform und im Vergleich zum Gymnasium.

3.3 Auswirkungen der Schulstrukturreform auf die Struktur der Platzierungsentscheidungen im nichtgymnasialen Bereich

Die Umstellung auf Zweigliedrigkeit des Schulsystems erfolgte in Berlin durch die Umgründung von Schulen der nichtgymnasialen Schulformen in Verbindung mit Schulschließungen, Schulfusionen und gelegentlichen Neugründungen (vgl. Baumert et al. 2017). Dies bedeutet, dass in den neuen, formal äquivalenten ISS die mehrgliedrige Schulstruktur bedingt durch die unterschiedliche Umgründungsgeschichte deutlich erkennbar bleibt. Wir vermuten, dass die formale Äquivalenz aller ISS und der damit verbundene Anspruch, einen alternativen Weg zur Hochschulreife anzubieten, nicht nur zu einer Annäherung der Übergangsquoten, sondern bei weiterbestehenden Leistungsunterschieden zwischen den nichtgymnasialen Schulen auch zu einer Verringerung der Bedeutung standortübergreifender Leistungsmaßstäbe und damit auch zu einer verminderten Akkuratheit der in den Übergangsberechtigungen implizierten Leistungsurteilen führen. Dies sollte sich in einem Rückgang des Zusammenhanges von standardisierten Leistungsmaßen und der Vergabe der Übergangsberechtigung an den nichtgymnasialen Schulen äußern. Die Bedeutung schulinterner Bewertungsmaßstäbe sollte gleichzeitig zunehmen, was in einem Anstieg des Zusammenhangs von Noten und Berechtigungsvergabe resultieren sollte. Welche Rolle die Einschätzung allgemeiner kognitiver Fähigkeiten, die bei der Vergabe von Grundschulempfehlungen über Noten und Testleistungen hinaus entscheidungsrelevant ist (vgl. Abschn. 2.2), bei der Erteilung von Übergangsberechtigungen spielt, ist unklar. Soweit in die Erteilung der Übergangsberechtigung Urteile über die Lernmotivation und Schulanpassung der Schülerinnen und Schüler eingehen, dürfte sich daran durch die Schulstrukturreform nichts ändern.

Bei Kurszuweisungen und einer jahrgangsbezogenen Notenvergabe stellt die jeweilige Jahrgangsstufe einer Schule die primäre Referenzgruppe dar. Wir erwarten deshalb bei steigendem mittleren Leistungsniveau eines Jahrgangs einen Referenzgruppeneffekt, der dazu führt, dass die Zuweisung zu Kursen mit erhöhtem Anspruchsniveau selektiver gehandhabt und Noten entsprechend strenger vergeben werden. Daraus folgt bei Kontrolle der individuellen Leistungsvoraussetzungen ein

negativer Zusammenhang zwischen dem mittleren Leistungsniveau einer Schule und der Wahrscheinlichkeit, an ISS eine Berechtigung zum Übergang in die gymnasiale Oberstufe zu erhalten.

Im Hinblick auf die soziale Zusammensetzung der Schülerschaft einer Schule rechnen wir an ISS mit keinem schulinternen Bezugsgruppeneffekt. Ob mit steigender Sozialschicht der Schülerschaft und damit größerer Bildungsnähe der Eltern die Bereitschaft der Lehrkräfte zunimmt, den Weg in die gymnasiale Oberstufe zu öffnen, ist unklar. Sofern dies zutreffen sollte, müsste ein positiver Kontexteffekt des mittleren sozialen Milieus einer Schule auftreten, der auch dann nachweisbar ist, wenn Leistungsvoraussetzungen und Merkmale der soziokulturellen Herkunft auf individueller Ebene konstant gehalten werden.

Im Hinblick auf die ethnische Zusammensetzung der Schülerschaft erlaubt die Forschungslage keine schlüssigen Annahmen. Auch im Hinblick auf die Frage, ob soziale Kontexte die Akkuratheit der Leistungsurteile von Lehrkräften beeinflussen, legen die vorliegenden Befunde keine gerichtete Hypothese nahe (vgl. aber Martínez et al. 2009; Hochweber et al. 2014).

4 Methode

4.1 Datengrundlage

Die Datengrundlage des vorliegenden Beitrags bildet die BERLIN-Studie, mit der Implementation und Auswirkungen der Umstellung des Sekundarschulsystems von Fünf- auf Zweigliedrigkeit in Berlin untersucht werden (Maaz et al. 2013; Neumann et al. 2017b). Die BERLIN-Studie ist längsschnittlich in einem quasi-experimentellen Kontrollgruppendesign angelegt. Zum Ende des Schuljahres 2010/11 wurden an öffentlichen Schulen Berlins 15-jährige Schüler/-innen der Jahrgangsstufen 7–12 und nicht 15-jährige Schüler/-innen, die eine 9. Klasse besuchten, untersucht. Die Schüler/-innen gehörten zu dem Jahrgang, der als letzte Kohorte das fünfgliedrige Schulsystem in Berlin vollständig durchlaufen hatte. Diese Kohorte bildet die Kontrollgruppe der BERLIN-Studie. Zum Ende des Schuljahres 2013/14 wurde die Untersuchung an einer nach Alter und Klassenstufe parallelisierten Stichprobe wiederholt. Diese Schüler/-innen gehörten dem Jahrgang an, der als 2. Kohorte das reformierte Sekundarschulsystem absolviert hatte. Diese Schüler/-innen bilden als Reformkohorte die Experimentalgruppe der BERLIN-Studie. Alle Schüler/-innen der 9. Jahrgangsstufe werden in beiden Kohorten bis zum Ende der beruflichen Erstausbildung bzw. bis zum Abitur längsschnittlich verfolgt. Anlage, Instrumentierung und Durchführung der Studie wurden von Becker et al. (2017) ausführlich beschrieben. Im vorliegenden Beitrag greifen wir auf diese Beschreibung zurück, beschränken uns jedoch auf die Informationen, die zum Verständnis des vorliegenden Beitrags notwendig sind.

Die Stichprobenziehungen wurden jeweils im Rahmen der ersten Erhebungswelle beider Kohorten durchgeführt. Sie folgen dabei einem mehrstufigen, stratifizierten Stichprobenplan. In einem ersten Schritt wurden nach Schulform stratifizierte Zufallsstichproben von Schulen gezogen. Die Schulstichprobe der Kontrollkohorte um-

fasst 29 Gymnasien, 29 Hauptschulen, 35 Gesamtschulen und 23 Realschulen. Zur Stichprobe der Reformkohorte gehören 29 Gymnasien und 70 ISS. In einem zweiten Schritt wurden innerhalb der einzelnen Schulen Zufallsstichproben von Schüler/-innen gezogen, und zwar jeweils 25 15-Jährige aus den Jahrgängen 7–12 und 10 nicht 15-jährige Neuntklässler/-innen. Die realisierte Stichprobe umfasste in der Kontrollkohorte $N=2359$ Schüler/-innen der 9. Jahrgangsstufe und $N=2377$ 15-Jährige. In der Reformkohorte beläuft sich die realisierte Stichprobe auf $N=1925$ Neuntklässler/-innen und $N=2119$ 15-Jährige. Die Ausschöpfungsquoten lagen je nach Schulform und Kohorte zwischen 82,9 und 91,3 % (zur Prüfung der Selektivität und zu weiteren detaillierten Stichprobenangaben vgl. Becker et al. 2017). Für die Analysen dieses Beitrags werden gewichtete Stichproben der Neuntklässler/-innen benutzt. Die Gewichte bilden die Populationsverhältnisse korrekt ab.

4.2 Untersuchungsinstrumente

Personenmerkmale, familialer Hintergrund und Schulbiografie Angaben zum *Geschlecht* der an der Untersuchung teilnehmenden Schüler/-innen wurden aus den Schulakten übernommen. Der *Sozialstatus* der Familie der Untersuchungsteilnehmerinnen und -teilnehmer wird durch den *International Socio-Economic Index of Occupational Status* (ISEI) bestimmt. Dabei wird jeweils der höhere Wert der beiden Elternteile zugrunde gelegt (HISEI). Die dafür verwendeten Berufsangaben beruhen auf Auskünften im Schülerinnen- und Schülerfragebogen. Der *Migrationshintergrund* der Befragten wurde in drei Ausprägungen operationalisiert (0=beide Eltern in Deutschland geboren, 1=ein Elternteil in Deutschland geboren, 2=beide Eltern im Ausland geboren). Auch diese Angaben beruhen auf Auskünften im Fragebogen der Schülerinnen und Schüler. Als *schulbiografische Merkmale* werden die *Schulformzugehörigkeit* und die am Ende der 10. Jahrgangsstufe erteilte *Berechtigung zum Übergang in die gymnasiale Oberstufe* berücksichtigt. Die Berechtigung zum Übergang in die gymnasiale Oberstufe wurde dem Jahreszeugnis am Ende der 10. Jahrgangsstufe entnommen.

Leistungsmerkmale *Schulleistungen* wurden in der BERLIN-Studie in beiden Kohorten über die Testleistungen in den Domänen Deutsch (Leseverstehen), Mathematik, Naturwissenschaften und Englisch (Leseverstehen) erhoben. Die Testitems wurden im Rahmen des *Partial Credit Models* in beiden Kohorten konkurrenzfähig einparametrisch skaliert. Die Personen-Parameter wurden als *Plausible Values* (PVs) geschätzt und in der Kontrollkohorte auf einen Mittelwert von 100 und eine Standardabweichung von 30 standardisiert. Die Reliabilitäten der Tests liegen je nach Domäne und Kohorte zwischen $r_{EAP}=0,88$ und $r_{EAP}=0,93$. Um Kollinearität zu vermeiden und den Leistungsstatus sparsam zu modellieren, werden im vorliegenden Beitrag die Ergebnisse der Leistungstests faktorenanalytisch zu einem einzigen Gesamtwert zusammengefasst, der ebenfalls in der Kontrollkohorte auf einen Mittelwert von 100 und eine Standardabweichung von 30 standardisiert wurde. Die Bildung eines Index für das Gesamtleistungsniveau lässt sich rechtfertigen, da in einer Hauptkomponentenanalyse das Kaiser-Kriterium und der *Scree*-Test für eine einfaktorielle Lösung

sprechen. Mit der ersten unrotierten Hauptkomponente werden 88 % der Gesamtvarianz erklärt.

Kognitive Grundfähigkeiten wurden mit dem Kognitiven Fähigkeitstest (KFT) erhoben (Heller und Perleth 2000). In den folgenden Analysen wird ein Gesamtwert verwendet, in den die beiden Untertests, die mit 20 Items verbales und 25 Items figurales Schlussfolgern erfassen, mit gleichem Gewicht eingehen. Die Reliabilitäten der beiden Testteile des KFT liegen bei $r_{KR20}=0,75$ (Kontrollkohorte) bzw. $r_{KR20}=0,76$ (Reformkohorte) für den verbalen und bei $r_{KR20}=0,91$ (beide Kohorten) für den figuralen Subtest. Die Reliabilität des Gesamtwertes beträgt in beiden Kohorten $r_{KR20}=0,91$.

Als weiteres Leistungsmerkmal wird die am Ende der 9. Jahrgangsstufe erreichte *Durchschnittsnote* in den Fächern Deutsch, Englisch und Mathematik herangezogen. Die Noten wurden den Jahrgangszeugnissen entnommen. Um den schulspezifischen Leistungsstatus von Schülern/-innen zu erfassen, wird im vorliegenden Beitrag die mittlere Jahrgangsnote der besuchten Schule von der individuellen Durchschnittsnote abgezogen.

Motivationale Orientierung und Schulverhalten Die Schulmotivation und schulische Anpassung wurden in der BERLIN-Studie in mehreren Dimensionen erhoben. Für den vorliegenden Beitrag wird aus Gründen sparsamer Modellierung und um Kollinearität zu vermeiden, für die Motivation nur die Lernfreude, die mit 4 Items erfasst wurde, herangezogen (Beispielitem: „Lernen macht mir Spaß“) (Gresch et al. 2010). Die interne Konsistenz dieser Skala beträgt in der Kontroll- und Reformkohorte $\alpha=0,82$. Als Indikator für (unzureichende) schulische Anpassung werden im vorliegenden Beitrag Schulversäumnisse, die mit vier Fragen erfasst wurden, verwendet (Beispielitems „... die Schule geschwänzt“) (Trautwein et al. 2010; Gresch et al. 2010). Die interne Konsistenz dieser Skala liegt in der Kontrollkohorte bei $\alpha=0,85$ und in der Reformkohorte bei $\alpha=0,83$.

Institutionelle Merkmale und Merkmale der Zusammensetzung der Schülerschaft Als institutionelles Merkmal wird sowohl in der Kontroll- als auch in der Reformkohorte neben der *Schulform* die *Verfügbarkeit einer eigenen gymnasialen Oberstufe* berücksichtigt. Zur Kennzeichnung der Zusammensetzung der Schülerschaft werden Schulleistungen, soziale Herkunft und Migrationsstatus in aggregierter Form auf Schulebene herangezogen.

4.3 Fehlende Werte und Mehrebenenstruktur

Große Feldstudien sind in der Regel mit dem Problem eines partiellen Datenausfalls durch Nichtteilnahme oder Antwortverweigerung konfrontiert. Dies gilt auch für beide Kohorten der BERLIN-Studie (vgl. Becker et al. 2017). Fehlende Werte treten bei Angaben, die in den Schulakten enthalten sind (Geschlecht, Alter), kaum auf und sind praktisch zu vernachlässigen. Dies gilt auch für die Noten, die den Zeugnissen entnommen wurden. Für die in den Zeugnissen verzeichneten Angaben hinsichtlich der Berechtigung zum Übergang in eine gymnasiale Oberstufe liegt der Anteil fehlender Werte bei 25,1 % in der Kontrollkohorte und bei 9,2 % in der Reformko-

horte. Bei Leistungsdaten beträgt der Anteil fehlender Werte je nach Kohorte bis zu 25,2% und bei Fragebogenangaben zu Motivation und Schulanpassung auf Ebene der gebildeten Skalen bis zu 23,1 bzw. 23,9%. Der Datenausfall ist nicht zufällig. Die fehlenden Werte hängen systematisch mit Leistungs- und Herkunftsmerkmalen zusammen. Insofern wurden in der BERLIN-Studie fehlende Werte bei Antwortverweigerung bzw. Nichtteilnahme unter Nutzung eines breiten Hintergrundmodells multipel imputiert. Multiple Imputation setzt lediglich ein *bedingt* zufälliges Fehlen von Daten voraus (Lüdtke und Robitsch 2011). Das breite Hintergrundmodell und die Tatsache, dass wichtige Personen- und Leistungsinformationen, die den Schulakten und Zeugnissen entnommen wurden, praktisch vollständig vorliegen, sprechen für die Erfüllung dieser Voraussetzung. Das verwendete Imputationsverfahren – *Multivariate Imputation by Chained Equations (MICE)* – ist darüber hinaus relativ robust gegenüber Verletzungen dieser Annahme (Schafer und Graham 2002; Van Buuren und Groothuis-Oudshoorn 2011). Es wurden fünf Datensätze erzeugt. Analyseergebnisse werden im vorliegenden Beitrag nach dem Vorschlag von Rubin (1987) integriert. In dem hierarchischen Untersuchungsdesign der BERLIN-Studie liegt Abhängigkeit der Personen innerhalb von Schulen vor. Dies führt abhängig von der Intraklassenkorrelation (ICC) zu einer Unterschätzung der Standardfehler. Diesem Problem kann man durch eine explizite Modellierung der Mehrebenenstruktur oder durch die Schätzung robuster Standardfehler (*Huber-White-* oder *Sandwich-Schätzer*) begegnen. Im vorliegenden Beitrag verwenden wir beide Vorgehensweisen.

4.4 Analytisches Vorgehen

Im Ergebnisteil werden zunächst deskriptive Befunde berichtet, die einen Populationsvergleich der untersuchten Kohorten erlauben und Auskunft über die Verteilung der Berechtigungen zum Übergang in die gymnasiale Oberstufe differenziert nach Schulform, Kohortenzugehörigkeit und Organisationsform der besuchten Schule geben. Im anschließenden analytischen Teil beruhen die Auswertungen im Wesentlichen auf Regressionsanalysen im linearen Wahrscheinlichkeitsmodell (*linear probability model* [LPM]) mit der Oberstufenzugangsberechtigung als abhängiger Variable.

Bei dichotomen abhängigen Variablen ist die Nutzung logistischer Regressionen die übliche Analyseentscheidung. Unsere Fragestellungen verlangen den Vergleich von geschachtelten Modellen innerhalb einer Stichprobe/Kohorte und vor allem den Vergleich von Modellen über Stichproben/Kohorten hinweg. Gerade dies ist jedoch bei nichtlinearen Regressionen aufgrund von Voraussetzungen der Modellidentifikation nicht oder nicht direkt möglich (Winship und Mare 1984; Karlson et al. 2010; Mood 2010; Best und Wolf 2012; Kern und Stein 2015). Um Koeffizienten in nichtlinearen Modellen vergleichbar zu machen, gibt es verschiedene Vorschläge, die bei Kern und Stein (2015) diskutiert werden.

Solange sich – wie in unserem Fall – das Hauptinteresse der Analysen primär auf die Richtung von Effekten und ihre zufallskritische Absicherung sowie die Schätzung *durchschnittlicher* Effekte richtet und die Berücksichtigung des nichtlinearen Zusammenhangs von nachgeordneter Bedeutung ist, bieten sich zwei unterschiedli-

che Wege zur Lösung der beschriebenen Probleme an. Im Rahmen der logistischen Regressionsanalysen lassen sich durchschnittliche marginale Effekte berechnen (*average marginal effects* [AMEs]), die den Durchschnittseffekt eines Prädiktors als Mittelwert aller marginalen Effekte über alle Beobachtungen hinweg wiedergeben. AMEs approximieren im Wesentlichen den linearen Trend und sind über Modelle hinweg vergleichbar (Mood 2010; Best und Wolf 2012). Sie sind – intuitiv unmittelbar einleuchtend – als Veränderungen der Wahrscheinlichkeit zu interpretieren, dass die abhängige Variable die Ausprägung 1 annimmt.

Der zweite Weg ist die Schätzung von LPMs, für die die Einschränkungen der logistischen Regression nicht gelten. Da LPMs erwartungstreue und konsistente Schätzungen des Effekts einer Variablen auf $P(y=1)$ liefern, werden sie in der Ökonomik häufig verwendet (Wooldridge 2010). Korrekte Standardfehler lassen sich über Sandwichschätzer ermitteln. LPM-Koeffizienten stimmen mit AMEs praktisch überein, solange die Prädiktoren nicht grob von der Normalverteilung abweichen (Mood 2010). Bei Verletzung der Normalitätsannahme können Verschätzungen auftreten, die mit steigender Interkorrelation der Prädiktoren zunehmen (Best und Wolf 2012). LPMs haben gegenüber AMEs den Vorteil, dass nicht nur die Regressionskoeffizienten, sondern auch die *Intercepts* direkt über Modelle und Gruppen hinweg vergleichbar sind.

Aufgrund dieser Überlegungen und des Umstandes, dass die in unseren Analysen verwendeten Prädiktoren mit einer Ausnahme (Schulversäumnisse) annähernd normalverteilt sind, berichten wir im Ergebnisteil die Resultate von linearen Wahrscheinlichkeitsmodellen mit robusten Standardfehlern. Um die Belastbarkeit der Befunde abzusichern, haben wir als Sensitivitätsanalyse in logistischen Regressionen AMEs bei gleichzeitiger KHB-Korrektur (Karlson et al. 2010) geschätzt, die im Anhang ausgewiesen sind. Wenn für die Beantwortung von theoretischen Fragestellungen der nichtlineare Zusammenhang von Bedeutung ist, werden logistische Regressionen genutzt.

Wir spezifizieren vornehmlich Zweigruppenvergleiche entweder zwischen Gymnasien und nichtgymnasialen Schulen oder zwischen Kontroll- und Reformkohorte. Für die Schätzung von schulischen Kontexteffekten werden hierarchisch-lineare Regressionsmodelle, bei denen die individuelle und die schulische Ebene unterschieden werden, jeweils in Zweigruppenvergleichen spezifiziert. Die Kontextanalysen werden doppelt manifest durchgeführt, da Mplus bei Gruppenvergleichen (*mixture models*) keine doppelt latente Effektschätzung erlaubt (Marsh et al. 2009, 2012). Dabei werden sowohl *random-intercept*- als auch *random-intercept/random-slope*-Modelle angepasst. Interaktionen zwischen Schulform- bzw. Kohortenzugehörigkeit und den als Prädiktoren spezifizierten Variablen werden mit dem *Wald*-Test geprüft. Die linearen Wahrscheinlichkeitsmodelle wurden in Mplus (Version 7.4) spezifiziert. Die Berechnung der AMEs erfolgte im Programmpaket STATA (Version 13). Für die übrigen Analysen wurden SPSS (Version 24), R und Conquest 4 benutzt. Die Korrelationen auf individueller Ebene der in den Analysen verwendeten Variablen sind in Tab. A.1 im Anhang A ausgewiesen.

Tab. 1 Stichprobenbeschreibung für die Neuntklässler/-innen in der Kontroll- und Reformkohorte

| | Gesamt N= 4892 | | Gymnasium N= 2040 | | Nicht-Gymnasium N= 2852 | |
|--|----------------------|--------------------|----------------------|--------------------|----------------------------|--------------------|
| | Kontroll- kohorte | Reform- kohorte | Kontroll- kohorte | Reform- kohorte | Kontroll- kohorte | Reform- kohorte |
| <i>Personenmerkmale</i> | | | | | | |
| Alter (M; SD) | 15,7; 0,72 | 15,4*; 0,65 | 15,4; 0,59 | 15,2*; 0,58 | 15,9; 0,72 | 15,5*; 0,68 |
| Geschlecht (männlich) | 50,9 | 51,9 | 46,5 | 46,5 | 53,9 | 55,9 |
| <i>Höchster Schulabschluss der Eltern</i> | | | | | | |
| Maximal Haupt- schulabschluss | 15,9 | 16,4 | 5,3 | 10,9 | 23,2 | 20,4 |
| Realschulabschluss | 29,0 | 25,4 | 17,9 | 16,2 | 36,6 | 32,3 |
| Fachhochschulreife | 8,7 | 6,1* | 8,4 | 5,6 | 8,8 | 6,5* |
| Abitur | 40,8 | 47,1 | 64,3 | 63,6 | 24,5 | 34,9* |
| Andere | 5,7 | 5,0 | 4,1 | 3,7 | 6,9 | 6,0 |
| <i>Höchster Sozialstatus der Eltern</i> | | | | | | |
| HISEI (M; SD) | 52,9; 20,5 | 51,8; 20,2 | 62,8; 18,8 | 59,4; 19,9 | 45,9; 18,6 | 46,1; 18,4 |
| <i>Migrationshintergrund (MGH)</i> | | | | | | |
| Kein MGH | 51,6 | 50,3 | 61,6 | 55,1 | 44,6 | 46,7 |
| Einseitiger MGH | 14,8 | 18,6* | 14,5 | 17,8 | 15,0 | 19,2* |
| Beidseitiger MGH | 33,6 | 31,1 | 23,9 | 27,1 | 40,4 | 34,0 |
| <i>Übergangsvoraussetzung</i> | | | | | | |
| Oberstufenzugangs- berechtigung | 53,2 | 62,6 | 95,8 | 91,7 | 23,5 | 41,1* |
| Ohne Oberstufe | – | – | – | – | 18,4 | 33,7* |
| Mit Oberstufe | – | – | – | – | 32,1 | 52,6* |
| <i>Fachleistungen</i> | | | | | | |
| Leistungsindex (M; SD) | 100,0; 30,0 | 97,4; 30,0 | 123,3; 18,9 | 117,4*; 20,2 | 83,8; 25,3 | 82,5; 27,3 |
| <i>Motivationale Orientierung</i> | | | | | | |
| Lernfreude (M; SD) (Range: 1–4) | 2,36; 0,70 | 2,22*; 0,66 | 2,38; 0,69 | 2,23*; 0,63 | 2,35; 0,70 | 2,21*; 0,68 |
| Schulversäumnisse (M; SD) (Range: 1–4) | 1,71; 0,86 | 1,49*; 0,72 | 1,58; 0,73 | 1,32*; 0,46 | 1,80; 0,93 | 1,62*; 0,84 |

M Angaben in % bzw. Mittelwerten, SD Standardabweichungen

* $p < 0,05$

5 Ergebnisse

5.1 Deskriptive Ergebnisse

Wenn man zwei Schülerjahrgänge, die im Abstand von drei Jahren – vor und nach der Schulstrukturreform – untersucht wurden, vergleichen will, ist es vordringlich zu prüfen, ob und inwieweit sich wichtige Populationsparameter verändert haben. Tab. 1

stellt zentrale Merkmale aufgebrochen nach Schulform und Kohortenzugehörigkeit vergleichend zusammen.

Die Ergebnisse zeigen eine hohe Stabilität der basalen Populationsparameter. Die feststellbaren Veränderungen der motivationalen Orientierungen sind nicht konsistent. So sind mit der Schulstrukturreform sowohl die Lernfreude als auch die Schulversäumnisse leicht zurückgegangen. Deutliche Unterschiede werden im mittleren Alter der Neuntklässler/-innen und bei den erteilten Oberstufenzugangsberechtigungen sichtbar. In der Reformkohorte sind die Schüler/-innen, die die 9. Jahrgangsstufe besuchen, im Mittel um vier Monate jünger. Dies ist teilweise auf eine Umstellung der Einschulung von einer Stichtags- auf eine Geburtsjahresregelung im Jahr 2004 zurückzuführen und teilweise Folge einer Abschaffung der Klassenwiederholung an ISS.

Tab. 1 gibt auch – differenziert nach Schulform – einen Überblick über die Verteilung der Oberstufenzugangsberechtigungen in der Kontroll- und Reformkohorte. Die Daten weisen einen leichten Rückgang der Berechtigungsquote an Gymnasien und einen starken Anstieg der Zulassungsquoten an ISS aus. Wie Neumann et al. (2017a, 2017c) bereits gezeigt haben, stieg an ISS die Quote der Übergangsberechtigten von der Kontroll- zur Reformkohorte insgesamt um rund 18 Prozentpunkte von 23,5 auf 41,1 %, an ISS ohne Oberstufe von 18,4 auf 33,7 % und an ISS mit eigener Oberstufe – alle umgegründete IGS – von 32,1 auf 52,6 %.

5.2 Multivariate Ergebnisse

Die Untersuchung der in Abschn. 3 entwickelten Fragestellungen erfolgt in zwei Schritten. In Abschn. 5.2.1 werden der Zusammenhang zwischen Fachleistungen und der Erteilung einer Übergangsberechtigung an Gymnasien und nichtgymnasialen Schulen und dabei insbesondere die Lokalisation der kritischen Übergangsschwelle und ihre Veränderung durch die Schulstrukturreform untersucht. Diese Fragestellung wurde in Abschn. 3.2 theoretisch begründet. Darauf folgt in Abschn. 5.2.2 die Untersuchung der strukturellen Komponenten der Platzierungsentscheidungen im nichtgymnasialen Bereich und ihrer mit der Systemumstellung verbundenen Veränderungen. Diese Fragestellungen wurden in den Abschn. 3.1 und 3.3 entwickelt.

5.2.1 *Vergleich der Erteilung von Übergangsberechtigungen am Gymnasium und im nichtgymnasialen Bereich vor und nach der Schulstrukturreform: Zusammenhang zwischen Fachleistungen und Übergangsberechtigung*

Der Zusammenhang zwischen Fachleistungen und Erteilung der Übergangsberechtigung an Gymnasien und nichtgymnasialen Schulen wurde regressionsanalytisch untersucht. Tab. 2 fasst die Ergebnisse der linearen Regressionen (LPM) von der Oberstufenzugangsberechtigung (nein/ja) auf Kohortenzugehörigkeit bzw. Kohortenzugehörigkeit und Schulleistungen zusammen. Es wurden zwei Modelle mit und ohne Kontrolle von Schulleistungen spezifiziert. Für beide Modelle wurde ein Zweigruppenvergleich (Gymnasium vs. Nichtgymnasium) durchgeführt. Ergänzend wurde für beide Modelle eine separate Analyse nur mit Schüler/-innen an nichtgymnasialen Schulen vorgenommen.

Tab. 2 Ergebnisse der Regressionen von der Oberstufenzugangsberechtigung (nein = 0/ja = 1) auf Kohortenzugehörigkeit und Schulleistungen (lineare Wahrscheinlichkeitsmodelle (LPM), Zweigruppenvergleiche (Gymnasium vs. Nichtgymnasium), nicht standardisierte Regressionskoeffizienten (B))

| Prädiktoren | Modell 1 | | Modell 1a | Modell 2 | | Modell 2a |
|---|--------------------|---------------------------------|--------------------|--------------------|---------------------------------|--------------------|
| | Gymnasium | Nichtgymnasium | Nichtgymnasium | Gymnasium | Nichtgymnasium | Nichtgymnasium |
| | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> |
| <i>Kohortenzugehörigkeit</i> | | | | | | |
| Kontrollkohorte M3 (Referenz) | – | – | – | – | – | – |
| Reformkohorte M2 | –0,041 (0,040) | 0,175** (0,033) ^a | 0,156** (0,035) | –0,036 (0,042) | 0,187** (0,021) | 0,178** (0,026) |
| <i>Organisationsform</i> | | | | | | |
| Ohne Oberstufe (Referenz) | – | – | – | – | – | – |
| Mit Oberstufe | – | – | 0,137** (0,046) | – | – | 0,026 (0,031) |
| Interaktion Oberstufe × Kohorte | – | – | 0,052 (0,065) | – | – | 0,022 (0,043) |
| <i>Schulleistung</i> | | | | | | |
| Leistungsfaktor (z-standardisiert) ^d | – | – | – | 0,026 (0,022) | 0,265** (0,012) ^c | 0,226** (0,015) |
| <i>Intercept</i> | 0,958** (0,017) | 0,235** (0,022) ^b | 0,184** (0,025) | 0,937** (0,026) | 0,368** (0,017) | 0,221** (0,015) |
| R ² | 0,003 (0,006) | 0,050** (0,019) | 0,064** (0,018) | 0,007 (0,006) | 0,351** (0,024) | 0,285** (0,020) |

Robuste Standardfehler (SE) in Klammern

** $p < 0,01$

^aUnterschied zwischen Gymnasium und Nichtgymnasium im Wald-Test signifikant: $\text{Chi}^2 = 17,7$; $\text{df} = 1$; $p < 0,01$

^bUnterschied zwischen Gymnasium und Nichtgymnasium im Wald-Test signifikant: $\text{Chi}^2 = 709,8$; $\text{df} = 1$; $p < 0,001$

^cUnterschied zwischen Gymnasium und Nichtgymnasium im Wald-Test signifikant: $\text{Chi}^2 = 93,1$; $\text{df} = 1$; $p < 0,001$

^dIn den Modellen 1 und 2 in der Gesamtstichprobe, in den Modellen 1a und 2a in der nichtgymnasialen Teilstichprobe standardisiert

Im Modell 1 wird in der Differenz der *Intercepts* die vor der Schulreform an Gymnasien und nichtgymnasialen Schulen extrem unterschiedliche Wahrscheinlichkeit, eine Berechtigung für den Übergang in die gymnasiale Oberstufe zu erhalten, sichtbar. Die Übergangsquoten betragen am Gymnasium 95,8 % und an den nichtgymnasialen Schulen 23,5 % (vgl. auch Tab. 1). Die Differenz betrug also 72,3 Prozentpunkte. Dass dieser Unterschied statistisch signifikant wird, ist trivial. Modell 1 zeigt ferner, dass der Rückgang der Berechtigungsquoten am Gymnasium von der Kontroll- zur Reformkohorte um 4,1 Prozentpunkte von rund 96 auf 92 % angesichts der Schiefe der Verteilung nicht zufallskritisch abzusichern ist, während der Anstieg der Berechtigungsquoten an den ISS um 17,5 Prozentpunkte signifikant wird ($b = 0,175^{**}$). Modell 1a bestätigt darüber hinaus inferenzstatistisch die unabhängig von der Kohortenzugehörigkeit erheblich höheren Berechtigungschancen an nichtgymnasialen Schulen mit eigener Oberstufe (Haupteffekt der Organisations-

form ($b = 0,137^{**}$) bei nicht signifikanter Interaktion mit der Kohortenzugehörigkeit ($b = 0,052$). Der Öffnungsprozess hat also in ähnlicher Weise an ISS mit und ohne gymnasiale Oberstufe stattgefunden.

Im Modell 2 wird anhand des in Abschn. 4 beschriebenen Leistungsfaktors das Leistungsniveau in vier Kernfächern kontrolliert. Der Faktorwert ist ein Indikator für das Gesamtleistungsniveau in den übergangsrelevanten Fächern (Deutsch, Mathematik, Englisch und Naturwissenschaften). Das Gesamtleistungsniveau ist ein guter Prädiktor für die Erteilung der Übergangsberechtigung im nichtgymnasialen Bereich. Mit dem Anstieg des Leistungsniveaus um eine Standardabweichung in der Gesamtstichprobe der Neuntklässler/-innen erhöht sich die Wahrscheinlichkeit, eine Übergangsberechtigung zu erhalten, um $b = 0,265^{**}$ – also um 26,5 Prozentpunkte. Im Gymnasium hat die Gesamtleistung dagegen keinen signifikanten Vorhersagewert ($b = 0,026$): Fast alle Schüler/-innen schaffen es. Die Differenz ist statistisch signifikant. Modell 2a zeigt ergänzend, dass bei Kontrolle des Leistungsniveaus die erhöhten Berechtigungsquoten an nichtgymnasialen Schulen mit Oberstufe (Modell 1a) nicht mehr nachweisbar sind. An Gesamtschulen bzw. ISS mit eigener Oberstufe gibt es also keinen leistungsunabhängigen Übergangsbonus. Die LPM-Koeffizienten unterscheiden sich von in logistischen Regressionsanalysen geschätzten AMEs in ihrer praktischen Bedeutung nicht (vgl. Tab. A.2 im Anhang A).

Besondere Aufmerksamkeit verdient die unterschiedliche Vorhersagekraft des Leistungsniveaus in den Kernfächern für die Erteilung einer Übergangsberechtigung im Gymnasium und im nichtgymnasialen Bereich. Um dies anschaulich zu machen, soll der Zusammenhang zwischen Leistungsniveau und Übergangsberechtigung über die gesamte Leistungsverteilung hinweg betrachtet werden. Für diesen Zweck ist das lineare Wahrscheinlichkeitsmodell nicht geeignet (vgl. Abschn. 4.4). Um einen nichtlinearen Zusammenhang abbilden zu können, haben wir logistische Regressionen von der Übergangsberechtigung auf das Fachleistungsniveau getrennt nach Schulform und Kohorte durchgeführt und die leistungsabhängige Wahrscheinlichkeit des Erhalts einer Oberstufenzugangsberechtigung für das Gymnasium und den nichtgymnasialen Bereich nach Kohorte differenziert geschätzt und grafisch dargestellt (vgl. Abb. 1).

Als kritische Übergangsschwelle definieren wir den jeweiligen Wendepunkt der Ogiven, an dem mit dem erreichten Leistungsniveau eine 50-prozentige Wahrscheinlichkeit, eine Übergangsberechtigung zu erhalten, verbunden ist. Die Gesamtleistung wurde in der gewichteten Gesamtstichprobe der Neuntklässler/-innen in der Kontrollkohorte auf einen Mittelwert von $\bar{x} = 100$ und eine Standardabweichung von $SD = 30$ standardisiert.

Vergleicht man zunächst für den nichtgymnasialen Bereich die Lage der Wendepunkte in der Kontroll- und Reformkohorte, wird ersichtlich, dass die kritische Zugangsschwelle von einem Leistungswert von 109 Punkten in der Kontroll- auf einen Wert von 92 Punkten in der Reformkohorte gesunken ist. Dies bedeutet, dass mit der Schulstrukturereform der kritische Schwellenwert an den ISS mit $d = 0,57$ um mehr als eine halbe Standardabweichung abgesenkt wurde. Ein Blick auf die Grafen des Gymnasiums zeigt dagegen Stabilität. Gleichzeitig wird aber auch erkennbar, wie wenig das Gesamtleistungsniveau in den Kernfächern den Erhalt bzw. den Nichterhalt der Übergangsberechtigung am Gymnasium vorhersagt. Der Ver-

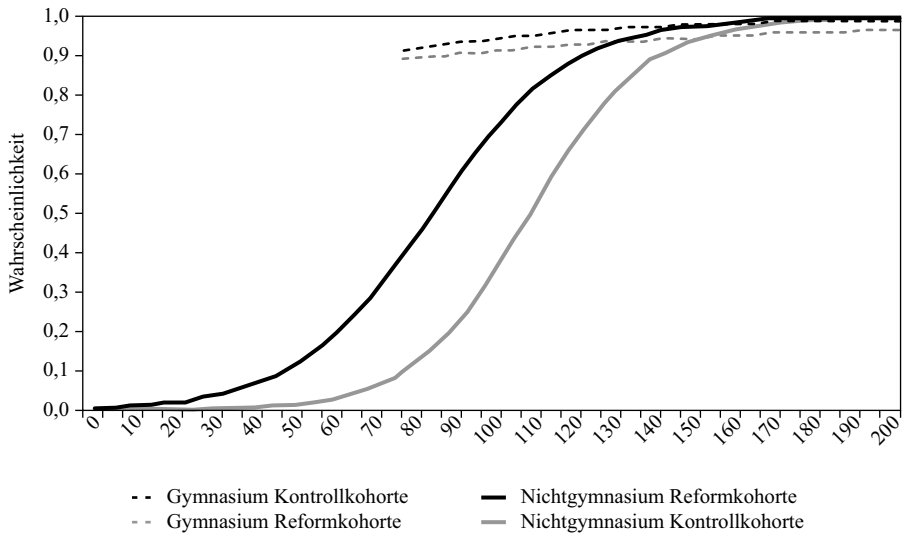


Abb. 1 Wahrscheinlichkeit der Erteilung einer Übergangsberechtigung für die gymnasiale Oberstufe in Abhängigkeit von der Schulleistung nach Schulform und Kohortenzugehörigkeit

lauf der Kurven ist extrem flach. Wenn Schüler/-innen am Gymnasium am Ende der 10. Jahrgangsstufe keine Übergangsberechtigung erhalten, ist dies eher eine Folge von Teilleistungsversagen als die Konsequenz eines insgesamt niedrigen Leistungsniveaus. Die Grafen illustrieren aber auch den in Tab. 2, Modell 2 nachgewiesenen Befund extrem unterschiedlicher Regressionssteigungen in beiden Schulbereichen: Wenn Schüler/-innen am Gymnasium ein Gesamtleistungsniveau aufweisen, das in den typischen Leistungsbereich nichtgymnasialer Schulen fällt, haben sie bessere Chancen auf eine Übergangsberechtigung als Schüler/-innen mit gleichen Leistungen im nichtgymnasialen Bereich. Der kritische Schwellenwert, mit mehr als 50-prozentiger Wahrscheinlichkeit eine Übergangsberechtigung zu erhalten, liegt im nichtgymnasialen Bereich bei 109 Leistungspunkten in der Kontroll- und bei 92 Leistungspunkten in der Reformkohorte. Die entsprechenden Wahrscheinlichkeiten für eine Übergangsberechtigung betragen bei diesen Leistungen im Gymnasium hingegen rund 95 bzw. 92 %. Gesamtleistungen von 109 Punkten oder weniger weisen vor und nach der Reform nur rund 28 % der Gymnasiasten/-innen auf. Bei einem Schwellenwert von 92 Punkten liegt der entsprechende Anteil von Gymnasiasten/-innen nur noch bei knapp 8 %. Diese Schüler/-innen werden im Vergleich zu Schülern/-innen an nichtgymnasialen Schulen nachsichtiger beurteilt. Ihr Anteil hat sich mit der Reform infolge der Absenkung der kritischen Übergangsschwelle im nichtgymnasialen Bereich deutlich verringert.

5.2.2 Erteilung der Berechtigung zum Übergang in die gymnasiale Oberstufe im nichtgymnasialen Bereich: Struktur der Platzierungsentscheidungen

Im Folgenden soll zunächst die Vergabe der Berechtigungen zum Übergang in die gymnasiale Oberstufe im nichtgymnasialen Bereich in Abhängigkeit von Leistung,

Tab. 3 Ergebnisse der Regression von der an nichtgymnasialen Schulen erworbenen Oberstufenzugangsberechtigung (0 = nein/1 = ja) auf Schulleistung, kognitive Fähigkeiten und motivationale Merkmale (lineare Wahrscheinlichkeitsmodelle (LPM), Zweigruppenvergleiche, Mittelwerte (M), nicht standardisierte Regressionskoeffizienten (B))

| Prädiktoren/ Intercept | Modell 0 | | Modell 1 | | Modell 2 | |
|---|----------------------|---------------------------------|----------------------|---------------------------------|----------------------|----------------------------------|
| | Kontroll- kohorte | Reform- kohorte | Kontroll- kohorte | Reform- kohorte | Kontroll- kohorte | Reform- kohorte |
| | M (SE) | M (SE) | B (SE) | B (SE) | B (SE) | B (SE) |
| <i>Schulleistungen/Intelligenz</i> | | | | | | |
| Leistungsfaktor (z-standardisiert) ^a | – | – | 0,216** (0,014) | 0,248** (0,015) ^c | 0,137** (0,014) | 0,140** (0,021) |
| Durchschnittsnote (Abweichung vom Schulmittel) | – | – | – | – | –0,144** (0,022) | –0,210** (0,023) ^d |
| Kognitive Fähigkeiten (z-standardisiert) ^a | – | – | – | – | 0,043** (0,013) | 0,045** (0,018) |
| <i>Motivation</i> | | | | | | |
| Lernmotivation (z-standardisiert) ^a | – | – | – | – | – | – |
| <i>Deviantes Verhalten</i> | | | | | | |
| Schulversäumnisse (z-standardisiert) ^a | – | – | – | – | – | – |
| <i>Mean/Intercept</i> | 0,235** (0,022) | 0,411** (0,024) ^b | 0,231** (0,014) | 0,418** (0,016) | 0,233** (0,014) | 0,420** (0,016) |
| R ² | – | – | 0,222** (0,022) | 0,304** (0,030) | 0,298** (0,027) | 0,414** (0,020) |

Lernmotivation und Schulverhalten der Schüler/-innen untersucht werden. Dabei steht die Akkuratheit der Leistungsurteile im Mittelpunkt. Daran anschließend wird die Abhängigkeit der Platzierungsentscheidungen von Kontextmerkmalen geprüft.

Leistung, kognitive Grundfähigkeiten und Motivation als Komponenten von Platzierungsentscheidungen im nichtgymnasialen Bereich Tab. 3 fasst die Ergebnisse der im Zweigruppenvergleich (Kontroll- vs. Reformkohorte) durchgeführten linearen Regressionen von der an nichtgymnasialen Schulen erworbenen Oberstufenzugangsberechtigung auf Schulleistungen, kognitive Grundfähigkeiten und motivationale Merkmale zusammen. Im leeren Modell 0 werden noch einmal die Mittelwerte der dichotom kodierten Oberstufenzugangsberechtigung in der Kontroll- und Reformkohorte für nichtgymnasiale Schulen gegenübergestellt. In Modell 1 wird als Prädiktor das Fachleistungsniveau in die Regressionsgleichung aufgenommen. Auch unter Kontrolle des Fachleistungsniveaus ändert sich der Unterschied zwischen den Kohorten, eine Übergangsberechtigung zu erhalten, praktisch nicht. Dies bedeutet, dass die stark erhöhten Berechtigungsquoten in der Reformkohorte *ausschließlich* auf gesunkene Leistungsanforderungen bei der Erteilung einer Zugangsberechtigung zurückzuführen sind (vgl. die Illustration in Abb. 1). Dennoch ist das mit objektiven Tests erfasste Gesamtleistungsniveau ein guter Prädiktor für den Erhalt einer Übergangsberechtigung. In der Kontrollkohorte steigt die Wahrscheinlichkeit,

Tab. 3 (Fortsetzung)

| Prädiktoren/ Intercept | Modell 3 | | Modell 4 | |
|---|----------------------|---------------------------------|----------------------|----------------------------------|
| | Kontroll- kohorte | Reform- kohorte | Kontroll- kohorte | Reform- kohorte |
| | <i>B</i> (SE) | <i>B</i> (SE) | <i>B</i> (SE) | <i>B</i> (SE) |
| <i>Schulleistungen/Intelligenz</i> | | | | |
| Leistungsfaktor (z-standardisiert) ^a | 0,159** (0,015) | 0,177** (0,020) | 0,128** (0,014) | 0,135** (0,021) |
| Durchschnittsnote (Abweichung vom Schulmittel) | – | – | –0,135** (0,022) | –0,205** (0,023) ^f |
| Kognitive Fähigkeiten (z-standardisiert) ^a | 0,059** (0,013) | 0,085** (0,019) ^e | 0,045** (0,012) | 0,046** (0,018) |
| <i>Motivation</i> | | | | |
| Lernmotivation (z-standardisiert) ^a | 0,030** (0,011) | 0,028* (0,013) | 0,013 (0,010) | 0,004 (0,011) |
| <i>Deviantes Verhalten</i> | | | | |
| Schulversäumnisse (z-standardisiert) ^a | –0,042** (0,010) | –0,044** (0,015) | –0,031** (0,009) | –0,023 (0,015) |
| <i>Mean/Intercept</i> | 0,232** (0,014) | 0,417** (0,016) | 0,235** (0,015) | 0,417** (0,016) |
| R ² | 0,247** (0,023) | 0,331** (0,028) | 0,304** (0,027) | 0,417** (0,020) |

Robuste Standardfehler (SE) in Klammern

** $p < 0,01$; * $p < 0,05$

^aIn der nichtgymnasialen Teilstichprobe kohortenübergreifend z-standardisiert

^bDifferenz zwischen Kontroll- und Reformkohorte signifikant (Wald-Test): $\chi^2 = 77,7$; $df = 1$; $p < 0,001$

^cDifferenz zwischen Kontroll- und Reformkohorte nicht signifikant (Wald-Test): $\chi^2 = 2,23$; $df = 1$; $p = 0,14$

^dDifferenz zwischen Kontroll- und Reformkohorte signifikant (Wald-Test): $\chi^2 = 9,77$; $df = 1$; $p = 0,002$

^eDifferenz zwischen Kontroll- und Reformkohorte nicht signifikant (Wald-Test): $\chi^2 = 1,24$; $df = 1$; $p = 0,26$

^fDifferenz zwischen Kontroll- und Reformkohorte signifikant (Wald-Test): $\chi^2 = 5,20$; $df = 1$; $p = 0,02$

eine Übergangsberechtigung zu erhalten, mit dem Anstieg des Leistungsniveaus um eine Standardabweichung (in der Leistungsverteilung der nichtgymnasialen Schüler/-innen) um $b = 0,216^{**}$. In der Reformkohorte ist der Effekt mit $b = 0,248^{**}$ tendenziell noch etwas größer. Der Unterschied lässt sich jedoch über alle Schulen hinweg nicht zufallskritisch absichern (vgl. aber die hierarchisch-linearen Analysen in Tab. 4). Die Analysen wurden mit den einzelnen Fachleistungen getrennt und gemeinsam wiederholt. Die in Tab. A.3 im Anhang A zusammengefassten Ergebnisse rechtfertigen noch einmal die Bildung eines Gesamtwertes. Der Zusammenhang zwischen Fachleistung und Übergangsberechtigung ($r = 0,47$ in der Kontroll- und $r = 0,55$ in der Reformkohorte) ist ein Indikator für die Akkuratheit der Urteile der Lehrkräfte, die mit der Reform und der Liberalisierung des Oberstufenzugangs entgegen unserer Ausgangsvermutung nicht geringer geworden ist. In der Tendenz deutet sich sogar eine leichte Zunahme der Urteilsakkuratheit an.

Tab. 4 Ergebnisse der hierarchisch-linearen Regression von der an nichtgymnasialen Schulen erworbenen Oberstufenzugangsberechtigung ($nein = 0/ja = 1$), auf Schulleistung, kognitive Fähigkeiten, Herkunftsmerkmale, Motivation, Schulverhalten und Merkmale der Schülerschaft der besuchten Schule (lineare Wahrscheinlichkeitsmodelle (LPM), Zweigruppenvergleiche (Kontroll- vs. Reformkohorte), nicht standardisierte Regressionskoeffizienten (B), robuste Standardfehler (SE) in Klammern)

| Prädiktoren/Intercept | Modell 1 | | Modell 2 | | Modell 3 | | Modell 4 | |
|---|----------------------------------|----------------------------------|----------------------------------|--------------------------------|----------------------------------|--------------------------------|----------------------------------|--------------------------------|
| | Kontrollkohorte <i>B (SE)</i> | Reformkohorte <i>B (SE)</i> | Kontrollkohorte <i>B (SE)</i> | Reformkohorte <i>B (SE)</i> | Kontrollkohorte <i>B (SE)</i> | Reformkohorte <i>B (SE)</i> | Kontrollkohorte <i>B (SE)</i> | Reformkohorte <i>B (SE)</i> |
| Individuelle Ebene | | | | | | | | |
| <i>Schulleistungen/Intelligenz</i> | | | | | | | | |
| Leistungsfaktor (z-standardisiert) ^a | 0,070** (0,017) | 0,097** (0,022) | 0,070** (0,017) | 0,097** (0,022) | 0,076** (0,017) | 0,097** (0,022) | - | - |
| Durchschnittsnote (Abweichung vom Schulmittel) | -0,138** (0,020) | -0,257** (0,021) ^c | -0,138** (0,020) | -0,257** (0,021) | -0,137** (0,020) | -0,257** (0,021) | -0,139** (0,020) | -0,253** (0,021) |
| Kognitive Fähigkeiten (z-standardisiert) ^a | 0,012 (0,009) | 0,007 (0,011) | 0,009 (0,009) | 0,008 (0,012) | 0,016 (0,009) | 0,007 (0,011) | 0,010 (0,009) | 0,006 (0,011) |
| <i>Herkunftsmerkmale</i> | | | | | | | | |
| Sozialstatus (HISEI) (z-standardisiert) ^a | - | - | 0,010 (0,009) | -0,001 (0,016) | - | - | - | - |
| Beidseitiger MGH | - | - | - | - | 0,051** (0,021) | 0,000 (0,036) | - | - |

Tab. 4 (Fortsetzung)
Prädiktoren/Intercept

| | Modell 1 | | Modell 2 | | Modell 3 | | Modell 4 | |
|--|---------------------------------------|-------------------------------------|---------------------------------------|-------------------------------------|---------------------------------------|-------------------------------------|---------------------------------------|-------------------------------------|
| | Kontroll- kohorte <i>B (SE)</i> | Reform- kohorte <i>B (SE)</i> | Kontroll- kohorte <i>B (SE)</i> | Reform- kohorte <i>B (SE)</i> | Kontroll- kohorte <i>B (SE)</i> | Reform- kohorte <i>B (SE)</i> | Kontroll- kohorte <i>B (SE)</i> | Reform- kohorte <i>B (SE)</i> |
| <i>Motivation</i> | | | | | | | | |
| Lernmotivation (z-standardisiert) ^a | 0,016 (0,009) | 0,010 (0,012) | 0,015 (0,009) | 0,010 (0,012) | 0,013 (0,009) | 0,010 (0,012) | 0,016 (0,009) | 0,009 (0,012) |
| <i>Deviantes Verhalten</i> | | | | | | | | |
| Schulversäumnisse (z-standardisiert) ^a | -0,111 (0,007) | -0,009 (0,014) | -0,011 (0,007) | -0,009 (0,013) | -0,011 (0,007) | -0,009 (0,014) | -0,011 (0,007) | -0,013 (0,014) |
| <i>R²</i> (innerhalb) | 0,197** (0,030) | 0,413** (0,025) | 0,199** (0,030) | 0,414** (0,025) | 0,199** (0,029) | 0,415** (0,025) | 0,128** (0,029) | 0,309** (0,035) |
| Schulebene | | | | | | | | |
| <i>Berechtigung auf:</i> | | | | | | | | |
| Gymnasiale Oberstufe (nein/ja) | -0,114 (0,095) | 0,016 (0,128) | -0,108 (0,097) | 0,043 (0,128) | -0,029 (0,110) | 0,023 (0,143) | -0,165 (0,090) | 0,027 (0,133) |
| Mittlere Schulleistung (z-standardisiert) ^b | 0,176** (0,064) | 0,075 (0,099) ^d | 0,181* (0,065) | 0,034 (0,095) | 0,129* (0,066) | 0,071 (0,095) | 0,217** (0,061) | 0,071 (0,103) |
| Mittlerer Sozialsta- tus der Eltern (HISEI, z-standardisiert) ^b | - | - | -0,014 (0,019) | 0,005 (0,038) | - | - | - | - |

Tab. 4 (Fortsetzung)

| Prädiktoren/Intercept | Modell 1 | | Modell 2 | | Modell 3 | | Modell 4 | |
|--|----------------------------------|--------------------------------|----------------------------------|--------------------------------|----------------------------------|--------------------------------|----------------------------------|--------------------------------|
| | Kontrollkohorte <i>B (SE)</i> | Reformkohorte <i>B (SE)</i> | Kontrollkohorte <i>B (SE)</i> | Reformkohorte <i>B (SE)</i> | Kontrollkohorte <i>B (SE)</i> | Reformkohorte <i>B (SE)</i> | Kontrollkohorte <i>B (SE)</i> | Reformkohorte <i>B (SE)</i> |
| Anteil SuS mit beidseitigem MGH | - | - | - | - | 0,049 (0,049) | 0,009 (0,101) | - | - |
| <i>Slope Leistungsfaktor auf:</i> | | | | | | | | |
| Mittlere Schulleistung (z-standardisiert) ^b | - | - | - | - | - | - | 0,070** (0,008) | 0,070** (0,008) |
| Intercept (Berechtigung) | 0,134* (0,062) | 0,400** (0,094) | 0,524* (0,099) | 0,482** (0,136) | 0,165** (0,063) | 0,402** (0,096) | 0,048 (0,057) | 0,371** (0,095) |
| Intercept (Slope Leistung) | - | - | - | - | - | - | 0,081** (0,017) | 0,089** (0,023) |
| R ² (zwischen: Berechtigung) | 0,530** (0,098) | 0,471** (0,132) | 0,524** (0,099) | 0,482** (0,136) | 0,553** (0,097) | 0,477** (0,131) | 0,592** (0,098) | 0,524** (0,124) |
| R ² (zwischen: Slope Leistung) | - | - | - | - | - | - | 0,531** (0,113) | 0,531** (0,113) |

** $p < 0,01$; * $p < 0,05$

^aIn der nichtgymnasialen Teilstichprobe kohortenübergreifend z-standardisiert

^bIn der nichtgymnasialen Teilstichprobe kohortenübergreifend auf Schulebene z-standardisiert

^cDifferenz zwischen Kontroll- und Reformkohorte signifikant (Wald-Test): $\text{Chi}^2 = 3,70$; $df = 1$; $p = 0,05$

^dDifferenz zwischen Kontroll- und Reformkohorte nicht signifikant (Wald-Test): $\text{Chi}^2 = 0,74$; $df = 1$; $p = 0,39$

Um einen angemessenen Eindruck von der komplexen Struktur der Urteile der Lehrkräfte zu erhalten, wurden die Modelle 2–4 spezifiziert, in denen vier weitere aus der Grundschulforschung bekannte Urteilskomponenten in die Regressionsgleichung aufgenommen wurden. Dem Modell 2 sind die *spezifischen* Effekte des Testleistungsniveaus, der kognitiven Grundfähigkeiten und der schulinternen Notenvergabe zu entnehmen. Alle drei Urteilskomponenten sind je für sich bedeutsam. Dabei bleibt die Akkuratheit der Platzierungsentscheidung entgegen unserer Vermutung über die Kohorten stabil, während der schulinterne Notenmaßstab mit der Schulstrukturereform erwartungsgemäß an Entscheidungsrelevanz gewinnt. Die Einschätzung der allgemeinen kognitiven Fähigkeiten spielt offensichtlich bei der Erteilung der Übergangsberechtigung – wie auch bei der Vergabe der Grundschulempfehlung – eine eigene Rolle, die sich mit der Schulstrukturereform nicht ändert.

In Modell 3 werden neben der Testleistung und den kognitiven Grundfähigkeiten zwei motivationale Variablen in die Regressionsgleichung eingeführt. Die motivationalen Merkmale sind erwartungsgemäß in beiden Kohorten für die Erteilung einer Übergangsberechtigung in spezifischer Weise prädiktiv. Die Effektstärke ist jeweils klein ($b = 0,03^{**}$ bzw. $b = 0,04^{**}$), aber in der Kumulation durchaus bemerkenswert.

Eine gewisse Modifizierung ergibt sich im Gesamtmodell 4, in dem zusätzlich die Benotung in den Kernfächern Deutsch, Mathematik und Englisch berücksichtigt wird. Von den fünf Prädiktoren – drei Leistungs- und zwei Motivationsmerkmale – leisten in beiden Kohorten nur die drei Leistungsindikatoren jeweils einen spezifischen, statistisch und praktisch signifikanten Erklärungsbeitrag. Beide Motivationsmerkmale sind dagegen – mit Ausnahme des devianten Verhaltens in der Kontrollkohorte – mit der Notenvergabe konfundiert.

Bei diesen Analysen ist herauszuheben, dass sich die im linearen Wahrscheinlichkeitsmodell gefundenen Koeffizienten in ihrer Größenordnung nicht substantiell von den in einer logistischen Regressionsanalyse geschätzten AMEs mit KHB-Korrektur unterscheiden, und zwar auch dann nicht, wenn ein Prädiktor von der Normalverteilung (wie im Fall der Schulversäumnisse) deutlich abweicht oder die Prädiktoren stark interkorreliert sind (wie im Fall Schulleistung und kognitive Grundfähigkeiten) (vgl. die Tab. A.4 im Anhang A).

Soziale Kontexte und die Erteilung von Übergangsberechtigungen im nichtgymnasialen Bereich Um abschließend zu überprüfen, ob die Urteile, mit denen Lehrkräfte an nichtgymnasialen Schulen über die Vergabe einer Übergangsberechtigung für die gymnasiale Oberstufe entscheiden, von der Schülerzusammensetzung des betroffenen Jahrgangs einer Schule – also von sozialen Kontexten – abhängen, wurden vier hierarchisch-lineare Regressionsanalysen durchgeführt, die Effekte innerhalb und zwischen Schulen trennen. Auf der Ebene innerhalb von Schulen werden alle Prädiktoren des Gesamtmodells in Tab. 3 berücksichtigt. In Modellen, in denen die ethnische oder soziale Zusammensetzung der Schulen als Kontextvariablen analysiert werden, wird auf individueller Ebene ebenfalls für ethnische und soziale Herkunft kontrolliert. Um die Kontexteffekte korrekt zu schätzen, werden Gesamtleistung sowie ethnische und soziale Herkunft auf individueller Ebene am Gesamtmittelwert zentriert (Raudenbush und Bryk 2002, S. 140). Die Ergebnisse der Analysen sind in Tab. 4 zusammengefasst.

In Modell 1 werden als Kontextmerkmale die Organisationsform der Schule (eigene Oberstufe nein/ja) und die in der nichtgymnasialen Stichprobe auf Schulebene schulübergreifend z -standardisierte mittlere Fachleistung herangezogen. Erwartet wurde ein Referenzgruppeneffekt, nach dem mit steigendem Leistungsniveau der Bezugsgruppe strenger geurteilt wird (vgl. Abschn. 3.3). Im Hinblick auf die Auswirkungen der Organisationsform wurden keine gerichteten Hypothesen formuliert. Es ist unklar, ob eine eigene Oberstufe die Erteilung von Übergangsberechtigungen bei Berücksichtigung der individuellen Schülermerkmale fördert oder unter dem Gesichtspunkt der Standardsicherung eher bremst. Die Ergebnisse zeigen, dass das Vorhandensein einer gymnasialen Oberstufe nach Kontrolle der Beurteilungskomponenten auf individueller Ebene keine Bedeutung für die Bereitschaft hat, Übergangsberechtigungen zu vergeben. Nichtgymnasiale Schulen mit und ohne Oberstufe legen bei der Erteilung der Berechtigungen vergleichbare Maßstäbe an. Im Hinblick auf den Bezugsgruppeneffekt widersprechen die Befunde den Erwartungen, obwohl die Leistungsvarianz zwischen den nichtgymnasialen Schulen erheblich ist ($ICC_{(M3)}=0,45$ und $ICC_{(M2)}=0,39$). Weder in der Kontroll- noch in der Reformkohorte lässt sich der erwartete Effekt nachweisen. Die Urteile von Lehrkräften werden mit steigendem mittlerem Leistungsniveau der Schülerschaft nicht strenger. Vielmehr findet sich in der *Kontrollkohorte* ein *positiver* Zusammenhang zwischen mittlerem Leistungsniveau und der Vergabe von Übergangsberechtigungen – ein Befund, den Ready und Wright (2011) für US-amerikanische Grundschulen für die Notenvergabe nachgewiesen hatten. Diese Effekte treten auch dann auf, wenn auf individueller Ebene Noten *nicht* kontrolliert werden. Weitere Kontextmerkmale – wie die soziale oder ethnische Zusammensetzung der Schülerschaft einer Schule – spielen, wie die Modelle 2 und 3 zeigen, für die Vergabe von Übergangsberechtigungen weder in der Kontroll- noch in der Reformkohorte eine Rolle. Zusätzlich wird in beiden Modellen durch die Trennung von inner- und zwischen schulischen Prozessen sichtbar, dass in den kognitiven Grundfähigkeiten ein schulübergreifender Maßstab zum Ausdruck kommt, der innerhalb von Schulen nach Kontrolle von Testleitung und Noten keine zusätzliche Erklärungskraft besitzt, während objektive Testleistungen über pädagogisch begründete Noten hinaus auch eine innerschulisch bedeutsame Urteils Komponente abbilden.

Im Modell 4 werden ein *random-intercept/random-slope*-Modell spezifiziert und die *cross-level*-Interaktion zwischen mittlerem Leistungsniveau eines Schuljahrgangs und der Steigung der Regression von der erteilten Übergangsberechtigung auf die individuelle Gesamtleistung geschätzt. Dieses Modell prüft, inwieweit die Akkuratheit von Platzierungsurteilen vom Leistungskontext einer Schule abhängt. Analog zu den Befunden von Martínez et al. (2009) und Hochweber et al. (2014) nimmt die Akkuratheit der Urteile von Lehrkräften bei der Übergangsberechtigung in die gymnasiale Oberstufe mit steigendem mittleren Leistungsniveau des betroffenen Jahrgangs einer Schule zu. Mit dem Anstieg der mittleren Leistung um eine Standardabweichung auf Schulebene verstärkt sich der Zusammenhang zwischen Gesamtleistung und Übergangsberechtigung auf individueller Ebene um $\Delta = 0,070^{**}$ – in der Kontrollkohorte von $b = 0,083$ auf $b = 0,153$ und in der Reformkohorte von $b = 0,089$ auf $b = 0,159$. Mit steigendem Leistungsniveau einer Schule

gelingt es offensichtlich *besser*, schulübergreifende Leistungsmaßstäbe zur Geltung zu bringen. Dies gilt unabhängig vom Vorhandensein einer eigenen Oberstufe.

6 Zusammenfassung und Diskussion

Ziel des vorliegenden Beitrags war es, die im Zuge der Umstellung auf das zweigliedrige Schulsystem in Berlin eingetretenen Veränderungen bei der Erteilung der Berechtigung für den Übergang in die gymnasiale Oberstufe, insbesondere die systematische Verschiebung der kritischen Leistungsschwelle und Veränderungen der Struktur der Platzierungsentscheidungen, multivariat zu untersuchen. Eine im vorliegenden Beitrag vorgenommene Überprüfung des Zusammenhangs zwischen Gesamtleistung in den Kernfächern, Kohortenzugehörigkeit und Wahrscheinlichkeit, eine Übergangsberechtigung zu erhalten, zeigte, dass der Anstieg der Berechtigungsquote an den nichtgymnasialen Schulen ausschließlich auf die Absenkung von Leistungsanforderungen zurückzuführen ist. Es traten starke Haupteffekte der Gesamtleistung und Kohortenzugehörigkeit auf. Eine Interaktion zwischen Gesamtleistung und Kohorte ließ sich nicht nachweisen, was für vergleichbare Zusammenhänge zwischen Gesamtleistung und Berechtigungsvergabe vor und nach der Schulstrukturreform spricht. Dabei wurde jedoch die kritische Leistungsschwelle, an der die Wahrscheinlichkeit für den Erhalt einer Übergangsberechtigung 50 % beträgt, um mehr als eine halbe Standardabweichung ($d=0,57$) herabgesetzt. Im Gymnasium blieben die Berechtigungsquoten stabil bei einer leichten Absenkung des mittleren Gesamtleistungsniveaus ($d=-0,22$). Dieses Ergebnis verdeutlicht, dass sich das zu Beginn des Beitrags skizzierte Problem, bei der Öffnung alternativer Wege zur Hochschulreife *gemeinsame* Standards zu sichern, mit der Strukturreform in Berlin verstärkt hat.

Nichtgymnasiale Schulen mit eigener Oberstufe erteilen erwartungsgemäß häufiger Übergangsberechtigungen für die gymnasiale Oberstufe als Schulen ohne Oberstufe. Daran hat sich mit der Schulstrukturreform, wie die nicht signifikante Interaktion zwischen Kohorte und Organisationsform zeigt, nichts geändert. Bei Kontrolle des Fachleistungsniveaus ist dieser Unterschied jedoch nicht mehr nachweisbar. Vermutlich sind die unterschiedlichen Berechtigungsquoten primär auf positive Eingangsselektivität möglicherweise aber auch teilweise auf gute Förderung zurückzuführen. Die vorliegende Studie erlaubt dazu keine finalen Aussagen. Festzuhalten ist, dass die ISS mit eigener Oberstufe weder Schrittmacher bei der Öffnung der gymnasialen Oberstufe noch Hüter stabiler Leistungsanforderungen während der Expansion sind.

Das Gymnasium wurde durch die Schulstrukturreform in der Mittelstufe vermutlich kaum tangiert. Das Gesamtleistungsniveau der Neuntklässler/-innen am Gymnasium sank zwar nach der Reform um 0,22 SD (relativ zur Standardabweichung der Gesamtkohorte; vgl. Tab. 1). Für die Annahme, dass dies eine direkte Folge der Strukturreform sein könnte, gibt es jedoch keine Hinweise (vgl. Becker et al. 2017). Der Übergang in die Oberstufe ist im Gymnasium nach wie vor die Regel. Über 90 % der Schülerinnen und Schüler erhalten die Berechtigung zum Übergang in die Qualifikationsphase der gymnasialen Oberstufe. Das Gesamtleistungsniveau von

Schüler/-innen ist bei dieser Verteilungsschiefe für die Vergabe der Übergangsberechtigung nur wenig prädiktiv, da die kritische Leistungsschwelle praktisch immer überschritten wird. Ein Scheitern beim Übergang in die Oberstufe ist demnach am Gymnasium eher auf Teilleistungsversagen als auf generelle Leistungsschwäche zurückzuführen. Dies bedeutet zugleich, dass Schüler/-innen, deren Gesamtleistung im typischen Leistungsbereich der ISS liegt, am Gymnasium einen Übergangsbonus erhalten. Dieser Bonus ist mit der Strukturreform und der Liberalisierung der Berechtigungsvergabe im nichtgymnasialen Bereich deutlich zurückgegangen.

Die Platzierungsentscheidung, die Lehrkräfte bei der Erteilung einer Zugangsberechtigung für die gymnasiale Oberstufe an ISS zu treffen haben, beruht auf der Integration unterschiedlicher Bewertungen, die im Laufe des 10. Schuljahrs vorgenommen werden. Die Struktur dieser Urteile ähnelt dem Urteilmuster, das bei der Erteilung der Grundschulempfehlung identifiziert wurde. In die Platzierungsentscheidung am Ende der Sekundarstufe I gehen Leistungsurteile, bei denen schulübergreifende und schulinterne Leistungsmaßstäbe Anwendung finden, sowie Einschätzungen der Leistungsmotivation und des Arbeits- und Schulverhaltens der Schüler/-innen ein. Die erklärungsstärksten Prädiktoren sind die am Ende der 9. Jahrgangsstufe erhobene Gesamtttestleistung in vier Fächern und die zum selben Zeitpunkt erfasste Durchschnittsnote in drei Kernfächern. Bemerkenswerterweise leisten darüber hinaus auch die kognitiven Grundfähigkeiten einen spezifischen Erklärungsbeitrag. Dies kann vorsichtig als Hinweis auf durch Tests und Noten nicht erfasste Leistungsvarianz interpretiert werden, könnte aber auch entsprechend den Grundschulbefunden ein Indikator für eine abschlussrelevante Einschätzung kognitiver Fähigkeiten durch Lehrkräfte sein. Die standortübergreifende Akkuratheit der leistungsbezogenen Urteilskomponente (Rangkorrelation zwischen Gesamtleistung und Platzierung) war vor der Systemumstellung mit $r=0,47$ an nichtgymnasialen Schulen moderat und erreichte nicht das durchschnittliche Akkuratheitsniveau von informierten Leistungsurteilen von Lehrkräften von $r=0,76$, das Südkamp et al. (2012) in ihrer Metaanalyse berichten. Die Korrelation liegt jedoch in gleicher Größenordnung, wie sie Steinmayr et al. (2017) bei Grundschulempfehlungen fanden. Bei der Interpretation dieses Befundes ist allerdings der zeitliche Abstand von einem Schuljahr zwischen Erfassung der Testleistung und der formalen Erteilung der Übergangsberechtigung zu berücksichtigen. Mit der Schulstrukturreform stieg die standortübergreifende Genauigkeit der Leistungsurteile an ISS wider Erwarten tendenziell auf $r=0,55$ an. Gleichzeitig gewannen die idiosynkratischen Maßstäbe von Einzelschulen erwartungsgemäß größere Bedeutung. Dies ist einmal ein Hinweis darauf, dass die Schulformen des gegliederten Systems mit ihren unterschiedlichen Traditionen und Ansprüchen je nach Umgründungsgeschichte der ISS auch bei Zweigliedrigkeit noch präsent sind, zum anderen aber auch ein Indikator für eine bei Ausweitung des Zugangs zur gymnasialen Oberstufe erhöhte Sensibilität für eine korrekte Leistungsabstufung innerhalb der eigenen Schule. Die Relevanz von schulübergreifend erfassten motivationalen Merkmalen und Merkmalen des allgemeinen Schulverhaltens änderte sich für die Erteilung von Übergangsberechtigungen mit der Schulstrukturreform nicht. Im Unterschied zu den entsprechenden Ergebnissen der Untersuchungen zu Platzierungsentscheidungen am Ende der Grundschule wurden

die motivationalen UrteilsKomponenten jedoch vollständig durch die Notenvergabe vermittelt.

Die Überprüfung, inwieweit Platzierungsentscheidungen von Lehrkräften an nichtgymnasialen Schulen durch organisatorische und soziale Kontextbedingungen beeinflusst werden, ergab ein differenziertes Bild. Im Unterschied zur Grundschulempfehlung konnte der erwartete Referenzgruppeneffekt, wonach bei steigendem mittlerem Leistungsniveau der Einzelschule die Erteilung der Übergangsberechtigung für die gymnasiale Oberstufe restriktiver gehandhabt wird, nicht nachgewiesen werden. Vielmehr waren Schulen mit höherem Leistungsniveau in der Kontrollkohorte eher bereit, eine Zugangsberechtigung für die gymnasiale Oberstufe zu erteilen. Hier wirkte möglicherweise ein Generalisierungseffekt, den Ready und Wright (2011) für US-amerikanische Grundschulen berichten. Die Vergabe eines Berechtigungsbonus bei steigendem mittleren Leistungsniveau des Schuljahrgangs setzt ein Bewusstsein für institutionelle Leistungsunterschiede voraus, das in stratifizierten Systemen auch zu erwarten ist. Insofern ist es durchaus naheliegend, dass dieser Generalisierungseffekt an Grundschulen in der USA und im gegliederten Schulsystem in Deutschland, aber hier nicht an Grundschulen auftritt (Westphal et al. 2016). Nach der Schulstrukturreform sind Einflüsse des Leistungskontextes nicht mehr nachweisbar – weder als Generalisierungs- noch als Referenzgruppeneffekt. Mit dem nominellen Äquivalenzanspruch aller ISS scheint die wahrgenommene institutionelle Stratifizierung des Systems zurückzugehen, sodass der Generalisierungseffekt nicht mehr auftritt. Gleichzeitig scheint aber die institutionelle Gliederung noch hinreichend präsent zu sein, um das Auftreten eines Referenzgruppeneffekts zu verhindern.

Von besonderem Interesse ist der Nachweis einer *cross-level*-Interaktion zwischen dem mittleren Leistungsniveau einer ISS und der Akkuratheit der Platzierungsentscheidung im Sinne des Zusammenhangs zwischen Gesamttestleistung und Urteilsfindung. Mit steigendem Leistungsniveau einer Schule spielen standortübergreifende Leistungsmaßstäbe bei der Erteilung von Übergangsberechtigungen offenbar eine größere Rolle. Dies trifft unabhängig von der Organisationsform der Schule (eigene Oberstufe nein/ja) zu. Das heißt jedoch nicht, dass Urteile mit steigendem Leistungsniveau der Schule strenger würden; es verändert sich nur die Struktur der Urteilsfindung, nicht die kritische Leistungsschwelle. Die ethnische und soziale Zusammensetzung der Schülerschaft ist vor und nach der Schulreform für die Akkuratheit der Erteilung von Übergangsberechtigungen ohne Bedeutung.

6.1 Grenzen der Untersuchung

Mit der BERLIN-Studie wurde in Deutschland zum ersten Mal eine quasi-experimentelle Evaluation einer flächendeckenden Schulstrukturreform realisiert. Dennoch ist auf die Grenzen der Untersuchungsanlage hinzuweisen. Untersucht wurde als Reformkohorte der zweite Schülerjahrgang, der die neu strukturierte Sekundarstufe I in Berlin vollständig durchlaufen hatte. In einem langfristigen Reformprozess ist dies ein relativ früher Zeitpunkt, der noch kein abschließendes Urteil erlaubt, dafür aber Korrekturen im Prozess ermöglicht. Es ist noch offen, welche Maßnahmen das Land Berlin ergreifen wird, um eine neue Balance zwischen Öffnung von Bildungswe-

gen und Standardsicherung zu finden. Insbesondere ist über die Weiterentwicklung der gymnasialen Oberstufen und deren Verzahnung mit den ISS noch nicht entschieden. Insofern sind die in diesem Beitrag berichteten Befunde Ergebnisse einer Zwischenbilanz (Neumann et al. 2017c).

Ebenso sind die Grenzen der Generalisierbarkeit der Befunde zu beachten. Berlin ist ein Stadtstaat mit einer für Großstädte und Ballungsgebiete charakteristischen Schulentwicklung. Dies gilt für die schulische Infrastruktur, die größere Wahlmöglichkeiten für Eltern eröffnet und Wettbewerb zwischen Schulen verstärkt (vgl. Neumann et al. 2013), ebenso wie für die Entwicklung des relativen Schulbesuchs in der Sekundarstufe. In Flächenstaaten und dort vor allem in strukturschwachen Gebieten sind andere Strukturbedingungen gegeben. Dies schränkt die Generalisierbarkeit der vorliegenden Ergebnisse ein.

Der vorliegende Beitrag untersucht die Struktur von institutionalisierten Platzierungsentscheidungen beim Übergang in die gymnasiale Oberstufe und dabei insbesondere die Rolle von Leistungsinformationen und die Veränderung kritischer Übergangsschwellen. Soziale und ethnische Disparitäten beim Erwerb von Übergangsberechtigungen und damit möglicherweise verbundene herkunftsbedingte Urteilsverzerrungen werden nicht behandelt (vgl. dazu Baumert et al. 2018).

Schließlich ist eine weitere Einschränkung der vorliegenden Untersuchung zu nennen. Leistungsdaten und Schulnoten – die wichtigsten Prädiktoren für die Erteilung der Übergangsberechtigung in die gymnasiale Oberstufe – wurden ebenso wie die motivationalen Merkmale am Ende der 9. Jahrgangsstufe erhoben. Dies ist einerseits ein Vorteil, da damit der Entscheidungsprozess, der sich über die 10. Jahrgangsstufe erstreckt, in Rechnung gestellt und nicht nur sein formaler Abschluss betrachtet wird. Darüber hinaus muss man aufgrund der formal festgelegten Notenanforderungen, die bei Erteilung der Übergangsberechtigung erfüllt sein müssen, davon ausgehen, dass Noten zumindest in Grenzbereichen am Ende der 10. Jahrgangsstufe auch strategisch im Hinblick auf die zu erteilende Berechtigung vergeben werden, so dass die Erteilung der Übergangsberechtigung und die Notenvergabe interagieren. Damit sind die Abschlussnoten kein unabhängiger Prädiktor der Übergangsberechtigung. Dies spricht für die Nutzung früher erhobener Daten. Andererseits kann dies aber auch zu einer Unterschätzung der Bedeutung sowohl der objektiven Leistung als auch der schulinternen Notenvergabe führen. Dies ist bei der Interpretation der Ergebnisse zu berücksichtigen.

Förderung Die Untersuchung wird durch Mittel der Senatsverwaltung für Bildung, Wissenschaft und Forschung (SenBWF) des Landes Berlin und der Jacobs-Foundation (Projekt-Nr. 2013-1083) gefördert.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Funding Open access funding provided by Max Planck Society.

Anhang A

Tab. A.1 Korrelationsmatrix auf individueller Ebene (Neuntklässler an nichtgymnasialen Schulen)

| | BERGO | LEIST | NOTEN | KFT | MOT | ABS |
|-------------------------------|-------|-------|--------|--------|--------|--------|
| Übergangsberechtigung (BERGO) | 1,00 | 0,494 | -0,474 | 0,413 | 0,113 | -0,247 |
| Gesamtleistung (LEIST) | - | 1,00 | -0,430 | 0,680 | 0,126 | -0,278 |
| Notendurchschnitt (NOTEN) | - | - | 1,00 | -0,382 | -0,206 | 0,235 |
| Kognitive Fähigkeiten (KFT) | - | - | - | 1,00 | 0,087 | -0,176 |
| Lernmotivation (MOT) | - | - | - | - | 1,00 | -0,123 |
| Schulversäumnisse (ABS) | - | - | - | - | - | 1,00 |

N = 2852

Tab. A.2 Ergebnisse der Regressionen von der Oberstufenzugangsberechtigung (nein = 0/ja = 1) auf Kohortenzugehörigkeit, Organisationsform der Schule und Schulleistungen (KHB, Average Marginal Effects (AMEs))

| Prädiktoren | Modell 1 | | Modell 1a | Modell 2 | | Modell 2a |
|------------------------------------|-------------------|--------------------|--------------------|-------------------|--------------------|--------------------|
| | Gymnasium | Nichtgymnasium | Nichtgymnasium | Gymnasium | Nichtgymnasium | Nichtgymnasium |
| | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> |
| <i>Kohortenzugehörigkeit</i> | | | | | | |
| Kontrollkohorte M3 (Referenz) | - | - | - | - | - | - |
| Reformkohorte M2 | -0,041 (0,040) | 0,170** (0,021) | 0,165** (0,029) | -0,036 (0,042) | 0,183** (0,022) | 0,191** (0,029) |
| <i>Organisationsform</i> | | | | | | |
| Ohne Oberstufe (Referenz) | - | - | - | - | - | - |
| Mit Oberstufe | - | - | 0,155** (0,031) | - | - | 0,036 (0,031) |
| Interaktion Oberstufe × Kohorte | - | - | 0,010 (0,041) | - | - | -0,020 (0,040) |
| <i>Schulleistung</i> | | | | | | |
| Leistungsfaktor (z-standardisiert) | - | - | - | 0,025 (0,021) | 0,267** (0,010) | 0,229** (0,009) |

Robuste Standardfehler (SE) in Klammern

** $p < 0,01$

Tab. A.3 Ergebnisse der Regression von der an nichtgymnasialen Schulen erworbenen Oberstufenzugangsberechtigung (0 = nein/1 = ja) auf Schulleistungen (lineare Wahrscheinlichkeitsmodelle (LPM), Zweigruppenvergleiche, nicht standardisierte Regressionskoeffizienten (B))

| Prädiktoren/Intercept | Modell 1–4 ^b | | Modell 5 | | Modell 6 | |
|--|-------------------------|--------------------|----------------------|--------------------|----------------------|--------------------|
| | Kontroll- kohorte | Reform- kohorte | Kontroll- kohorte | Reform- kohorte | Kontroll- kohorte | Reform- kohorte |
| | <i>M (SE)</i> | <i>M (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> | <i>B (SE)</i> |
| Leistungsfaktor (z-standardisiert) ^a | – | – | – | – | 0,216** (0,014) | 0,248** (0,015) |
| Lesen Deutsch (z-standardisiert) ^a | 0,192** (0,015) | 0,226** (0,015) | 0,039 (0,031) | 0,028 (0,024) | – | – |
| Lesen Englisch (z-standardisiert) ^a | 0,103** (0,014) | 0,230** (0,014) | 0,080** (0,022) | 0,088** (0,027) | – | – |
| Mathematik (z-standardisiert) ^a | 0,199** (0,013) | 0,234** (0,015) | 0,087** (0,028) | 0,105** (0,023) | – | – |
| Naturwissenschaften (z-standardisiert) ^a | 0,201** (0,015) | 0,238** (0,014) | 0,035 (0,038) | 0,057 (0,035) | – | – |
| <i>R²</i> | | | | | | |
| Lesen Deutsch (Modell 1) | 0,170** (0,023) | 0,243** (0,029) | 0,226** (0,022) | 0,309** (0,030) | 0,222** (0,022) | 0,304** (0,030) |
| Lesen Englisch (Modell 2) | 0,189** (0,019) | 0,257** (0,029) | – | – | – | – |
| Mathematik (Modell 3) | 0,187** (0,012) | 0,263** (0,018) | – | – | – | – |
| Naturwissenschaften (Modell 4) | 0,189** (0,020) | 0,263** (0,028) | – | – | – | – |

Robuste Standardfehler (SE) in Klammern

** $p < 0,01$ ^aIn der nichtgymnasialen Teilstichprobe kohortenübergreifend z-standardisiert^bJeweils alternativ ein einziger Leistungsprädiktor

Tab. A.4 Ergebnisse der logistischen Regressionen von der an nichtgymnasialen Schulen erworbenen Oberstufenzugangsberechtigung (nein/ja) auf Leistungs- und Motivationsmerkmale (KHB, Average Marginal Effects (AMEs))

| Prädiktoren/Intercept | Modell 1 | | Modell 2 | |
|--|----------------------------------|--------------------------------|----------------------------------|--------------------------------|
| | Kontrollkohorte <i>B</i> (SE) | Reformkohorte <i>B</i> (SE) | Kontrollkohorte <i>B</i> (SE) | Reformkohorte <i>B</i> (SE) |
| <i>Schulleistungen/Intelligenz</i> | | | | |
| Leistungsfaktor (<i>z</i> -standardisiert) ^a | 0,232* (0,011) | 0,244* (0,012) | 0,163* (0,015) | 0,145* (0,021) |
| Durchschnittsnote (Abweichung vom Schulmittel) | – | – | –0,167* (0,020) | –0,218* (0,022) |
| Kognitive Fähigkeiten (<i>z</i> -standardisiert) | – | – | 0,019*** (0,011) | 0,028 (0,017) |
| <i>Motivation</i> | | | | |
| Lernmotivation (<i>z</i> -standardisiert) | – | – | – | – |
| <i>Deviantes Verhalten</i> | | | | |
| Schulversäumnisse (<i>z</i> -standardisiert) | – | – | – | – |
| Prädiktoren/Intercept | Modell 3 | | Modell 4 | |
| | Kontrollkohorte <i>B</i> (SE) | Reformkohorte <i>B</i> (SE) | Kontrollkohorte <i>B</i> (SE) | Reformkohorte <i>B</i> (SE) |
| <i>Schulleistungen/Intelligenz</i> | | | | |
| Leistungsfaktor (<i>z</i> -standardisiert) ^a | 0,179* (0,014) | 0,180* (0,020) | 0,146* (0,016) | 0,138* (0,021) |
| Durchschnittsnote (Abweichung vom Schulmittel) | – | – | –0,154* (0,020) | –0,212* (0,023) |
| Kognitive Fähigkeiten (<i>z</i> -standardisiert) | 0,038* (0,011) | 0,071* (0,017) | 0,022** (0,011) | 0,029*** (0,017) |
| <i>Motivation</i> | | | | |
| Lernmotivation (<i>z</i> -standardisiert) | 0,030* (0,010) | 0,029** (0,012) | 0,012 (0,010) | 0,005 (0,011) |
| <i>Deviantes Verhalten</i> | | | | |
| Schulversäumnisse (<i>z</i> -standardisiert) | –0,073* (0,013) | –0,053* (0,015) | –0,060* (0,013) | –0,032** (0,016) |

Robuste Standardfehler (SE) in Klammern

* $p < 0,01$; ** $p < 0,05$; *** $p < 0,10$ ^aIn der nichtgymnasialen Teilstichprobe kohortenübergreifend *z*-standardisiert

Anhang B

Vorgaben für die Vergabe der Übergangsberechtigung in die gymnasiale Oberstufe an nichtgymnasialen Schulen vor und nach der Berliner Schulstrukturreform (Quelle: Neumann et al. 2017a, S. 263 ff.)

Vorgaben vor der Schulstrukturreform

Auszug Regelungen zum Erwerb der Übergangsberechtigung **an Gesamtschulen** (§ 57 Sek-I-VO vom 19.01.2005 in der Fassung vom 11.12.2007):

1. Schülerinnen und Schüler der Gesamtschule gehen in die gymnasiale Oberstufe über, wenn:
 1. sie den mittleren Schulabschluss bestanden haben,
 2. an mindestens drei Kursen des oberen Anspruchsniveaus, darunter mindestens zwei der drei Fächer Deutsch, Mathematik und erste Fremdsprache, teilgenommen haben und
 3. mit den Jahrgangsnoten am Ende der Jahrgangsstufe 10 die Leistungsanforderungen nach Absatz 2 erfüllt werden.

Wer an mindestens drei Kursen des oberen Anspruchsniveaus teilgenommen hat, geht in die Einführungsphase über; bei Teilnahme an mindestens fünf Kursen des oberen Anspruchsniveaus ist der unmittelbare Übergang in die Qualifikationsphase grundsätzlich verpflichtend.

2. Die Leistungsanforderungen nach Absatz 1 Satz 1 Nr. 3 werden erfüllt, wenn:
 1. im leistungsdifferenzierten Unterricht in Kursen des oberen Anspruchsniveaus mindestens neun Punkte, in den Kursen des unteren Anspruchsniveaus mindestens neun Punkte sowie in den nicht leistungsdifferenziert unterrichteten Fächern mindestens vier Punkte erreicht werden,
 2. insgesamt in den ohne Leistungsdifferenzierung unterrichteten Fächern mindestens 64 Punkte und in allen Fächern insgesamt 120 Punkte erreicht werden und
 3. die Jahrgangsnote in höchstens einem Fach, ausgenommen in den Fächern Deutsch, Mathematik und erste Fremdsprache, „ungenügend“ lautet.

Die Leistungen nach Satz 1 Nr. 1 dürfen in höchstens zwei Fächern unterschritten werden, wobei nur eine Unterschreitung in den Fächern Deutsch, Mathematik und erste Fremdsprache vorliegen darf.

Sofern Jahrgangsnoten in weniger als 14 Fächern vorliegen, verringern sich die Punktsommen gemäß Satz 1 Nr. 2 für jedes nicht beurteilte Fach des leistungsdifferenzierten Unterrichts um neun Punkte und des nicht leistungsdifferenzierten Unterrichts um acht Punkte. Werden die Gesellschaftswissenschaften nicht als Lernbereich unterrichtet, so erhöhen sich die Punktsommen gemäß Satz 1 Nr. 2 um jeweils acht Punkte.

Regelungen für **Haupt- und Realschulen:**

Die entsprechenden Regelungen für Haupt- und Realschulen fanden sich in der Verordnung über die gymnasiale Oberstufe (VO-GO vom 18.04.2007, zuletzt geändert am 11.08.2011). Danach konnten Schülerinnen und Schüler aus der *Hauptschule* bei Vorhandensein des mittleren Schulabschlusses aufgenommen werden, „wenn in allen Fächern Jahrgangleistungen mit einem Durchschnitt² von 3,0 oder besser erreicht werden, dabei darf die Summe der Fächer Deutsch, erste Fremdsprache und Mathematik (a) nicht größer als fünf oder (b) wenn eines dieser Fächer in Leistungsstufe A³ unterrichtet wurde, nicht größer als sechs, (c) wenn zwei dieser Fächer in Leistungsstufe A unterrichtet wurden, nicht größer als sieben oder (d) wenn drei dieser Fächer in Leistungsstufe A unterrichtet wurden, nicht größer als acht und kein weiteres Fach darf schlechter als ausreichend bewertet sein“ (§ 5 Abs. 2 Satz 1 VO-GO). An der *Realschule* wurde die Übergangsberechtigung bei Vorliegen des mittleren Schulabschlusses erteilt, „wenn in allen Fächern Jahrgangleistungen mit einem Durchschnitt von 3,0 oder besser erreicht wurden, dabei müssen in den Fächern Deutsch, erste Fremdsprache und Mathematik mindestens befriedigende Leistungen vorliegen und kein weiteres Fach darf schlechter als ausreichend bewertet sein“ (§ 5 Abs. 2 Satz 2 VO-GO).

Vorgaben nach der Schulstrukturreform

Auszug Regelungen zum Erwerb der Übergangsberechtigung an **Integrierten Sekundarschulen** (§ 48 Sek-I-VO vom 31.03.2010 in der Fassung vom 26.01.2011):

1. Schülerinnen und Schüler der Integrierten Sekundarschule gehen in die gymnasiale Oberstufe über, wenn sie
 1. den mittleren Schulabschluss erworben haben,
 2. in mindestens drei Fächern des leistungsdifferenzierten Unterrichts, darunter mindestens zwei der Fächer Deutsch, Mathematik und erste Fremdsprache, auf dem E-Niveau unterrichtet wurden und
 3. mit den Jahrgangsnoten am Ende der Jahrgangsstufe 10 die Leistungsanforderungen nach Absatz 2 erfüllen.

In die zweijährige Form der gymnasialen Oberstufe geht auf Antrag über, wer nach Leistungsfähigkeit und Leistungsbereitschaft eine erfolgreiche Mitarbeit in der zweijährigen Form erwarten lässt; über den Antrag entscheidet die Klassenkonferenz oder der Jahrgangsausschuss.

2. Die Leistungsanforderungen nach Absatz 1 Satz 1 Nummer 2 werden erfüllt, wenn bei Umrechnung der erreichten Punkte im leistungsdifferenzierten Unterricht in Noten des E-Niveaus

² Bei der Ermittlung des Durchschnitts bleiben dabei die Leistungen in Musik, Bildende Kunst und Sport unberücksichtigt. Dies gilt sowohl für die Haupt- als auch für die Realschulen.

³ In Leistungsstufe A wurde an den Hauptschulen auf die Standards des mittleren Schulabschlusses vorbereitet.

1. in mindestens drei Fächern des leistungsdifferenzierten Unterrichts, darunter mindestens zwei der Fächer Deutsch, Mathematik und erste Fremdsprache, mindestens befriedigende Leistungen erzielt werden und
2. in allen anderen Fächern im Durchschnitt der Noten mindestens 3,0 erreicht wird und kein Fach schlechter als ausreichend bewertet wurde.

Literatur

- Abgeordnetenhaus Berlin (2009). *Beschluss: Weiterentwicklung der Berliner Schulstruktur* (Abgeordnetenhaus Berlin, Drucksache 16/2479). Berlin: Abgeordnetenhaus Berlin.
- Arnold, K.-H., Bos, W., Richert, P., & Stubbe, T.C. (2007). Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 271–297). Münster: Waxmann.
- Artelt, C., & Rausch, T. (2014). Accuracy of teacher judgments. In S. Krolak-Schwerdt, S. Glock & M. Böhmer (Hrsg.), *Teachers' professional development: Assessment, training, and learning* (S. 27–43). Rotterdam: Sense Publishers.
- Autorengruppe Bildungsberichterstattung (2016). *Bildung in Deutschland 2016: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann.
- Baeriswyl, F., Trautwein, U., Wandeler, C., & Lüdtke, O. (2009). Wie gut prognostizieren subjektive Lehrerempfehlungen und schulische Testleistungen beim Übertritt die Mathematik- und Deutschleistung in der Sekundarstufe I? In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen* (S. 352–372). Wiesbaden: VS.
- Baeriswyl, F., Wandeler, C., Trautwein, U., & Oswald, K. (2006). Leistungstest, Offenheit von Bildungsgängen und obligatorische Beratung der Eltern: Reduziert das Deutschfreiburger Übergangsmodell die Effekte des sozialen Hintergrunds bei Übergangsentscheidungen? *Zeitschrift für Erziehungswissenschaft*, 9(3), 373–392. <https://doi.org/10.1007/s11618-006-0056-6>.
- Baumert, J., Cortina, K.S., & Leschinsky, A. (2003a). Grundlegende Entwicklungen und Strukturprobleme im allgemeinbildenden Schulwesen. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland: Strukturen und Entwicklungen im Überblick* (S. 52–147). Reinbek: Rowohlt.
- Baumert, J., Neumann, M., Böse, S., & Zunker, N. (2017). Implementation der Berliner Schulstrukturenform. In M. Neumann, M. Becker, J. Baumert, K. Maaz & O. Köller (Hrsg.), *Zweigigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin* (S. 81–126). Münster: Waxmann.
- Baumert, J., Trautwein, U., & Artelt, C. (2003b). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 261–331). Opladen: Leske + Budrich.
- Baumert, J., Dumont, H., Becker, M., Neumann, M., Bachsleitner, A., Köller, O., & Maaz, K. (2018). Soziokulturelle und geschlechtsspezifische Selektivität von Übergangsberechtigungen für die gymnasiale Oberstufe in mehr- und zweigliedrigen Schulsystemen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 70(4), 593–628.
- Becker, M., Neumann, M., Radmann, S., Jansen, M., Nagy, G., Borzиковsky, C., Kropf, M., Köller, O., Maaz, K., & Baumert, J. (2017). Durchführung, Datengrundlage, Erhebungsinstrumente und statistische Methoden. In M. Neumann, M. Becker, J. Baumert, K. Maaz & O. Köller (Hrsg.), *Zweigigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin* (S. 55–80). Münster: Waxmann.
- Best, H., & Wolf, C. (2012). Modellvergleich und Ergebnisinterpretation in Logit- und Probit-Regressionen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 64, 377–395. <https://doi.org/10.1007/s11577-012-0167-4>.
- de Boer, H., Bosker, R.J., & van der Werf, M.P.C. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, 102, 168–179. <https://doi.org/10.1037/a0017289>.

- Böhmer, I., Hörstermann, T., Gräsel, C., Krolak-Schwerdt, S., & Glock, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung: Welcher Urteilsregel folgen Lehrkräfte? *Journal for Educational Research Online*, 7(2), 59–81. <https://doi.org/10.1007/s11577-012-0167-4>.
- Bos, W., Voss, A., Lankes, E.-M., Schwippert, K., Thiel, O., & Valtin, R. (2004). Schullaufbahneempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *Einige Länder der Bundesrepublik Deutschland in nationalen und internationalen Vergleich* (S. 191–228). Münster: Waxmann.
- Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–68. <https://doi.org/10.18637/jss.v045.i03>.
- Caro, D. H., Lenkeit, J., Lehmann, R., & Schwippert, K. (2009). The role of academic achievement growth in school track recommendations. *Studies in Educational Evaluation*, 35, 183–192. <https://doi.org/10.1016/j.stueduc.2009.12.002>.
- Combat, B. (2013). Zum Einfluss von primären und sekundären Effekten der sozialen Herkunft beim zweiten schulischen Übergang in der Schweiz. Ein Vergleich unterschiedlicher Dekompositions- und Operationalisierungsmethoden. *Schweizerische Zeitschrift für Bildungswissenschaften*, 35(3), 447–471.
- Dompnier, B., Pansu, P., & Bressoux, P. (2006). An integrative model of scholastic judgments: Pupils' characteristics, class context, halo effect and internal attributions. *European Journal of Psychology of Education*, 21, 119–133. <https://doi.org/10.1007/BF03173572>.
- Driessen, G., Sleegers, P., & Smit, F. (2008). The transition from primary to secondary education: Meritocracy and ethnicity. *European Sociological Review*, 24, 527–542. <https://doi.org/10.1093/esr/jcn018>.
- Dumont, H., Maaz, K., Neumann, M., & Becker, M. (2014). Soziale Ungleichheiten beim Übergang von der Grundschule in die Sekundarstufe I: Theorie, Forschungsstand, Interventions- und Fördermöglichkeiten. In S. Hellekamps, J.-L. Le Cam & A. Conrad (Hrsg.), *Schulbücher und Lektüren in der vormodernen Unterrichtspraxis* (Zeitschrift für Erziehungswissenschaft: Sonderheft 17, S. 141–165). Wiesbaden: Springer VS.
- Dumont, H., Neumann, M., Becker, M., Maaz, K., & Baumert, J. (2013a). Der Übergangsprozess von der Grundschule in die Sekundarstufe I vor und nach der Schulstrukturreform in Berlin: Die Rolle primärer und sekundärer Herkunftseffekte. In K. Maaz, J. Baumert, M. Neumann, M. Becker & H. Dumont (Hrsg.), *Die Berliner Schulstrukturreform: Bewertung durch die beteiligten Akteure und Konsequenzen des neuen Übergangsverfahrens von der Grundschule in die weiterführenden Schulen* (S. 133–207). Münster: Waxmann.
- Dumont, H., Neumann, M., Maaz, K., & Trautwein, U. (2013b). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen: Internationale und nationale Befunde. *Psychologie in Erziehung und Unterricht*, 3, 163–183. <https://doi.org/10.2378/peu2013.art14d>.
- Gresch, C., Maaz, K., & Baumert, J. (2010). *Skalenhandbuch „Übergang“ 2006*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Gröhllich, C., & Guill, K. (2009). How stable are reference group effects of secondary school track recommendations? *Journal for Educational Research Online*, 1(1), 154–171.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4–12-R)*. Göttingen: Hogrefe.
- Hillmert, S., & Jacob, M. (2005). Institutionelle Strukturierung und inter-individuelle Variation: Zur Entwicklung herkunftsbezogener Ungleichheiten im Bildungsverlauf. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 57(3), 414–442.
- Hillmert, S., & Jacob, M. (2010). Selections and social selectivity on the academic track: A lifecourse analysis of educational attainment in Germany. *Research in Social Stratification and Mobility*, 28(1), 59–76.
- Hochweber, J., Hosenfeld, I., & Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *Journal of Educational Psychology*, 106, 289–300. <https://doi.org/10.1037/a0033829>.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313.
- Holtappels, H. G. (2017). Schulsystem im Umbruch. Entwicklungen in zentralen Feldern der Schulstruktur in Deutschland. In H. G. Holtappels (Hrsg.), *Schulsystem im Umbruch. Entwicklung und Qualität des Schulsystems: Neue empirische Befunde und Entwicklungstendenzen* (S. 11–52). Münster: Waxmann.
- Ivanov, S., & Nikolova, R. (2010). Prognostische Validität der Schullaufbahneempfehlung aufgrund von Leistungsdaten. In W. Bos & C. Gröhllich (Hrsg.), *Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 8* (S. 128–141). Münster: Waxmann.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsreich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 197–209.

- Karolson, K. B., Holm, A., & Breen, R. (2010). *Total, direct, and indirect effects in logit models* (Centre for strategic educational research (CSER), working paper series, Bd. No. 0005). Aarhus: Aarhus University.
- Kern, C., & Stein, P. (2015). Comparing coefficients of nonlinear multivariate regression models between equations. *Survey Research Methods*, 9(3), 159–167. <https://doi.org/10.18148/srm/2015.v9i3.6211>.
- Klapproth, F., Glock, S., Krolak-Schwerdt, S., Martin, R., & Böhmer, M. (2013). Prädiktoren der Sekundarschulempfehlung in Luxemburg: Ergebnisse einer Large-Scale-Untersuchung. *Zeitschrift für Erziehungswissenschaft*, 16(2), 355–379. <https://doi.org/10.1007/s11218-012-9197-z>.
- KMK (1993). Vereinbarung über Schularten und Bildungsgänge im Sekundarbereich I. Beschluss der Kultusministerkonferenz vom 03.12.1993 i.d.F. vom 25.09.2014. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1993/1993_12_03-VBSek-I.pdf. Zugegriffen: 25. Okt. 2016.
- Köller, O., Baumert, J., & Schnabel, K. U. (1999). Wege zur Hochschulreife: Offenheit des Systems und Sicherung vergleichbarer Standards. Analysen am Beispiel der Mathematikleistungen von Oberstufenschülern an integrierten Gesamtschulen und Gymnasien in Nordrhein-Westfalen. *Zeitschrift für Erziehungswissenschaft*, 2(3), 385–422.
- Köller, O., Watermann, R., Trautwein, U., & Lüdtke, O. (Hrsg.). (2004). *Wege zur Hochschulreife in Baden-Württemberg: TOSCA – Eine Untersuchung an allgemeinbildenden und beruflichen Gymnasien*. Opladen: Leske + Budrich.
- Leucht, M., Kampa, N., & Köller, O. (Hrsg.). (2016). *Fachleistungen beim Abitur: Vergleich allgemeinbildender und beruflicher Gymnasien in Schleswig-Holstein*. Münster: Waxmann.
- Lüdtke, O., & Robitzsch, A. (2011). Umgang mit fehlenden Daten in der empirischen Bildungsforschung. In S. Maschke & L. Stecher (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online: Fachgebiet: Methoden der empirischen erziehungswissenschaftlichen Forschung, Quantitative Forschungsmethoden*. Weinheim: Juventa. <https://doi.org/10.3262/EEO07100150>
- Maaz, K., Bachsleitner, A., Baumert, J., Dumont, H., Neumann, M., Radmann, S., & Becker, M. (2017). Soziokulturelle Disparitäten des Bildungserfolgs vor und nach der Schulstrukturreform in Berlin. In M. Neumann, M. Becker, J. Baumert, K. Maaz & O. Köller (Hrsg.), *Zweigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin* (S. 361–424). Münster: Waxmann.
- Maaz, K., Baumert, J., Neumann, M., Becker, M., & Dumont, H. (Hrsg.). (2013). *Die Berliner Schulstrukturreform: Bewertung durch die beteiligten Akteure und Konsequenzen des neuen Übergangsverfahrens von der Grundschule in die weiterführenden Schulen*. Münster: Waxmann.
- Maaz, K., Chang, P. H., & Köller, O. (2004). Führt institutionelle Vielfalt zur Öffnung im Bildungssystem? Sozialer Hintergrund und kognitive Grundfähigkeit der Schülerschaft an allgemein bildenden und beruflichen Gymnasien. In O. Köller, R. Watermann, U. Trautwein & O. Lüdtke (Hrsg.), *Wege zur Hochschulreife in Bade-Württemberg. TOSCA – Eine Untersuchung an allgemeinbildenden und beruflichen Gymnasien* (S. 153–203). Opladen: Leske + Budrich.
- Maaz, K., Neumann, M., Trautwein, U., Wendt, W., Lehmann, R., & Baumert, J. (2008). Der Übergang von der Grundschule in die weiterführende Schule: Die Rolle von Schüler- und Klassenmerkmalen beim Einschätzen der individuellen Lernkompetenz durch die Lehrkräfte. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30(3), 519–548.
- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79, 280–295.
- Marsh, H., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44, 764–802.
- Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, 47(2), 106–124.
- Martínez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*, 14, 78–102. <https://doi.org/10.1080/10627190903039429>.
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. <https://doi.org/10.1016/j.tate.2017.02.021>.
- 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. <https://doi.org/10.2307/40602478>.
- Neuenschwander, M. P., & Malti, T. (2009). Selektionsprozesse beim Übergang in die Sekundarstufe I und II. *Zeitschrift für Erziehungswissenschaft*, 12(1), 216–232.

- Neugebauer, M. (2011). Werden Jungen von Lehrerinnen bei den Übergangsempfehlungen für das Gymnasium benachteiligt? Eine Analyse auf Basis der IGLU-Daten. In A. Hadjar (Hrsg.), *Geschlechtsspezifische Bildungsungleichheiten* (S. 235–260). Wiesbaden: VS.
- Neumann, M., Baumert, J., Radmann, S., Becker, M., Maaz, K., & Ohl, S. (2017a). Leistungs- und Bewertungsstandards beim Erwerb der Übergangsberechtigung in die gymnasiale Oberstufe im neu strukturierten Berliner Sekundarschulsystem. In M. Neumann, M. Becker, J. Baumert, K. Maaz & O. Köller (Hrsg.), *Zweigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin* (S. 255–294). Münster: Waxmann.
- Neumann, M., Becker, M., Baumert, J., Maaz, K., & Köller, O. (Hrsg.). (2017b). *Zweigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin*. Münster: Waxmann.
- Neumann, M., Becker, M., Baumert, J., Maaz, K., Köller, O., & Jansen, M. (2017c). Das zweigliedrige Berliner Sekundarschulsystem auf dem Prüfstand. In M. Neumann, M. Becker, J. Baumert, K. Maaz & O. Köller (Hrsg.), *Zweigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin* (S. 469–502). Münster: Waxmann.
- Neumann, M., Kropf, M., Becker, M., Albrecht, R., Maaz, K., & Baumert, J. (2013). Die Wahl der weiterführenden Schulen im neu geordneten Berliner Übergangsverfahren. In K. Maaz, J. Baumert, M. Neumann, M. Becker & H. Dumont (Hrsg.), *Die Berliner Schulstrukturreform. Bewertung durch die beteiligten Akteure und Konsequenzen des neuen Übergangsverfahrens von der Grundschule in die weiterführenden Schulen* (S. 87–132). Münster: Waxmann.
- Neumann, M., Lühe, J., Becker, M., & Baumert, J. (2017d). Schulbiografische Verläufe und Abschlussspirationen vor und nach der Berliner Schulstrukturreform. In M. Neumann, M. Becker, J. Baumert, K. Maaz & O. Köller (Hrsg.), *Zweigliedrigkeit im deutschen Schulsystem: Potenziale und Herausforderungen in Berlin* (S. 127–154). Münster: Waxmann.
- Neumann, M., Milek, A., Maaz, K., & Gresch, C. (2010). Zum Einfluss der Klassenzusammensetzung auf den Übergang von der Grundschule in die weiterführenden Schulen. In K. Maaz, J. Baumert, C. Gresch & N. McElvany (Hrsg.), *Der Übergang von der Grundschule in die weiterführende Schule: Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (S. 229–252). Bonn: Bundesministerium für Bildung und Forschung.
- Nölle, I., Höstermann, T., Krolak-Schwerdt, S., & Gräsel, C. (2009). Relevante diagnostische Informationen bei der Übergangsempfehlung – die Perspektive der Lehrkräfte. *Unterrichtswissenschaft*, 37(4), 294–319.
- Pfost, M., Rausch, T., Schiefer, I.M., & Artelt, C. (2017). Zur Entwicklung von Gymnasiastinnen und Gymnasiasten ohne Gymnasialempfehlung. *Zeitschrift für Erziehungswissenschaft*, 21(3), 511–534.
- Pohlmann, S. (2009). *Der Übergang am Ende der Grundschulzeit – Zur Formation der Übergangsempfehlung aus der Sicht der Lehrkräfte*. Münster: Waxmann.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.
- Ready, D.D., & Wright, D.L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities. *American Educational Research Journal*, 48, 335–360. <https://doi.org/10.3102/0002831210374874>.
- Roeder, P.M. (1997). Entwicklung vor, während und nach der Grundschulzeit: Literaturüberblick über den Einfluss der Grundschulzeit auf die Entwicklung in der Sekundarschule. In F.E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 405–421). Weinheim: Beltz.
- Rosenthal, R., & Rubin, D.B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74(2), 166–169.
- Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177.
- Schneider, T. (2011). Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschulempfehlung. *Zeitschrift für Erziehungswissenschaft*, 14(3), 371–396.
- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27–52.
- Schuchart, C. (2006). Die Bedeutung der Entkopplung von Schulart und Schulabschluss für die Schullaufbahnplanung aus Elternsicht. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 26(4), 403–419.
- Schulze, A., Wolter, F., & Unger, R. (2009). Bildungschancen von Grundschulern: Die Bedeutung des Klassen- und Schulkontextes am Übergang auf die Sekundarstufe I. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 61, 411–435. <https://doi.org/10.1007/s11577-009-0072-7>.

- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85–95. <https://doi.org/10.1024/1010-0652.19.12.85>.
- Steinmayr, R., Michels, J., & Weidinger, A.F. (2017). Fa(ir)bulous. Faire Beurteilung des Leistungspotenzials von Schülerinnen und Schülern. Abschlussbericht des Forschungsprojekts. <https://www.stiftung-mercator.de/de/publikation/studie-fairbulous/>. Zugegriffen: 07. Juni 2019.
- Stubbe, T.C., Bos, W., & Euen, B. (2012). Der Übergang von der Primar- in die Sekundarstufe. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 209–226). Münster: Waxmann.
- Südkamp, A., & Möller, J. (2009). Referenzgruppeneffekte im Simulierten Klassenraum: Direkte und indirekte Einschätzungen von Schülerleistungen. *Zeitschrift für Pädagogische Psychologie*, 23, 161–174. <https://doi.org/10.1024/1010-0652.23.34.161>.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>.
- Tiedemann, J., & Billmann-Mahecha, E. (2007). Zum Einfluss von Migration und Schulklassenzugehörigkeit auf die Übergangsempfehlung für die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 10(1), 108–120. <https://doi.org/10.1007/s11618-007-0009-8>.
- Tiedemann, J., & Billmann-Mahecha, E. (2010). Wie erfolgreich sind Gymnasiasten ohne Gymnasialempfehlung? Die Kluft zwischen Schullaufbahneempfehlung und Schulformwahl der Eltern. *Zeitschrift für Erziehungswissenschaft*, 13(4), 649–660. <https://doi.org/10.1007/s11618-010-0146-3>.
- Tillmann, K.-J. (2017). Gleichwertigkeit im Zwei-Säulen-Modell? Das Beispiel Gemeinschaftsschule im Saarland. In H. G. Holtappels (Hrsg.), *Schulsystem im Umbruch. Entwicklung und Qualität des Schulsystems: Neue empirische Befunde und Entwicklungstendenzen* (S. 105–118). Münster: Waxmann.
- Trautwein, U. et al. (2010). *Transformation des Sekundarschulsystems und akademische Karrieren (TOSCA). Dokumentation der eingesetzten Items und Skalen*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Trautwein, U., Baeriswyl, F., Lüdtke, O., & Wandeler, C. (2008). Die Öffnung des Schulsystems: Fakt oder Fiktion? Empirische Befunde zum Zusammenhang von Grundschulübertritt und Übergang in die gymnasiale Oberstufe. *Zeitschrift für Erziehungswissenschaft*, 11(4), 648–665. <https://doi.org/10.1007/s11618-008-0043-1>.
- Trautwein, U., Köller, O., Lehmann, R., & Lüdtke, O. (2007a). Öffnung von Bildungswegen, erreichtes Leistungsniveau und Vergleichbarkeit von Abschlüssen. In U. Trautwein, O. Köller, R. Lehmann & O. Lüdtke (Hrsg.), *Schulleistungen von Abiturienten: Regionale, schulformbezogene und soziale Disparitäten* (S. 11–29). Münster: Waxmann.
- Trautwein, U., Köller, O., Lehmann, R., & Lüdtke, O. (2007b). *Schulleistungen von Abiturienten: Regionale, schulformbezogene und soziale Disparitäten*. Münster: Waxmann.
- Trautwein, U., Nagy, G., & Maaz, K. (2011). Soziale Disparitäten und die Öffnung des Sekundarschulsystems. *Zeitschrift für Erziehungswissenschaft*, 14(4), 445–463. <https://doi.org/10.1007/s11618-011-0220-5>.
- Wagner, W., Helmke, A., & Schrader, F.-W. (2009). Die Rekonstruktion der Übergangsempfehlung für die Sekundarstufe I und der Wahl des Bildungsgangs auf der Basis des Migrationsstatus, der sozialen Herkunft, der Schulleistung und schulklassenspezifischer Merkmale. In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen* (Zeitschrift für Erziehungswissenschaft: Sonderheft 12, S. 183–204). Wiesbaden: VS.
- Watermann, R., Nagy, G., & Köller, O. (2004). Mathematikleistungen in allgemein bildenden und beruflichen Gymnasien. In O. Köller, R. Watermann, U. Trautwein & O. Lüdtke (Hrsg.), *Wege zur Hochschulreife in Baden-Württemberg: TOSCA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien* (S. 205–283). Opladen: Leske + Budrich.
- Westphal, A., Becker, M., Vock, M., Maaz, K., Neumann, M., & McElvany, N. (2016). The link between teacher-assigned grades and classroom socioeconomic composition: The role of classroom behavior, motivation, and teacher characteristics. *Contemporary Educational Psychology*, 46, 218–227. <https://doi.org/10.1016/j.cedpsych.2016.06.004>.
- Winship, C., & Mare, R.D. (1984). Regression models with ordinal variables. *American Sociological Review*, 49(4), 512–525.
- Wooldridge, J.-M. (2010). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.