

HERKUNFT ZENSIERT?

Leistungsdiagnostik und soziale Ungleichheiten in der Schule

Eine Studie im Auftrag der Vodafone Stiftung Deutschland

Prof. Dr. Kai Maaz, Universität Potsdam

Prof. Dr. Franz Baeriswyl, Universität Freiburg (Schweiz)

Prof. Dr. Ulrich Trautwein, Universität Tübingen

HERKUNFT ZENSIERT?

Leistungsdiagnostik und soziale Ungleichheiten in der Schule

Eine Studie im Auftrag der Vodafone Stiftung Deutschland

Prof. Dr. Kai Maaz, Universität Potsdam

Prof. Dr. Franz Baeriswyl, Universität Freiburg (Schweiz)

Prof. Dr. Ulrich Trautwein, Universität Tübingen

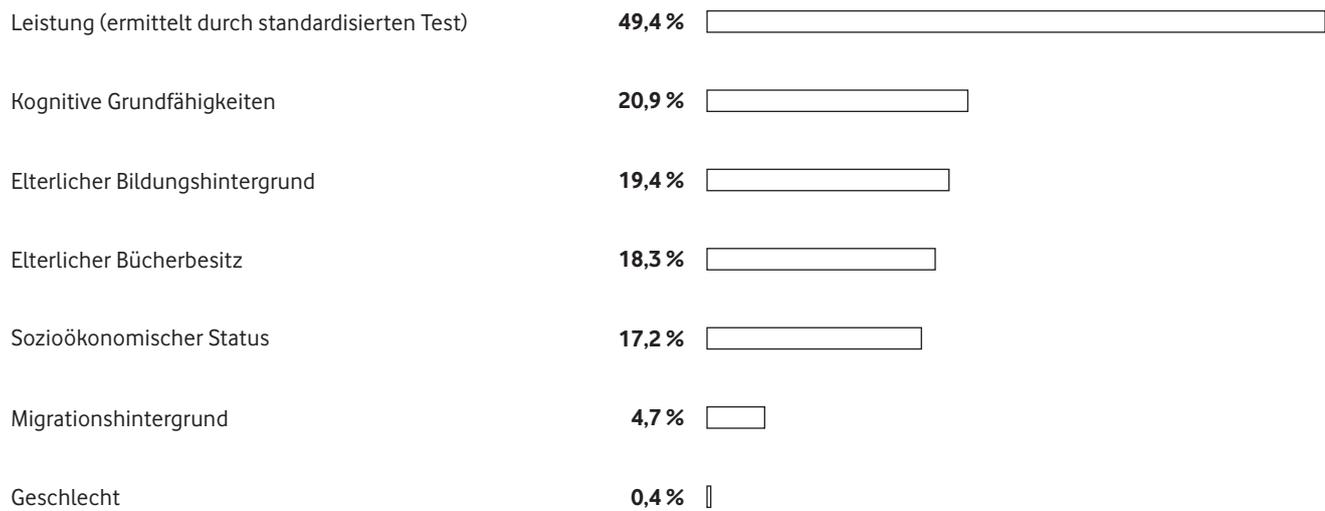
Inhalt

Zentrale Ergebnisse.....	4
1. Einleitung	8
2. Theoretischer und empirischer Hintergrund.....	10
2.1. Leistungsbewertungen.....	10
2.2. Sozialer Hintergrund	12
2.3. Primäre und sekundäre Herkunftseffekte	13
2.4. Psychometrische Merkmale von Leistungsbewertungen und das Auftreten systematischer Verzerrungen.....	14
2.5. Exkurs: Effekte gesellschaftlichen Wandels und wissenschaftlichen Fortschritts auf die Leistungsdiagnostik am Beispiel Schweiz.....	16
2.6. Empirische Befunde aus jüngeren Schulleistungsstudien.....	22
3. Soziale Herkunft und Leistungsbewertung: Empirische Studien zu zentralen Fragestellungen.....	26
3.1. Teilstudie 1: <i>Soziale Herkunft und Schulnoten: Gibt es substanzielle Zusammenhänge?.....</i>	26
3.2. Teilstudie 2: <i>Anstrengungsbereitschaft und Gewissenhaftigkeit: Effekte auf die Notengebung?</i>	34
3.3. Teilstudie 3: <i>Quantifizierung primärer und sekundärer Herkunftseffekte der Benotung auf die Empfehlung und den Übergang auf das Gymnasium.....</i>	40
3.4. Teilstudie 4: <i>Simulation der Neutralisation von primären und sekundären Effekten der Notengebung – Wie groß wäre der Anstieg der Gymnasialquote für Kinder aus Arbeiterfamilien?</i>	44
3.5. Teilstudie 5: <i>Konsequenzen der Simulation von primären und sekundären Effekten der Notengebung für den Übergang auf das Gymnasium</i>	50
3.6. Teilstudie 6: <i>Bewertungsunterschiede am Ende der gymnasialen Oberstufe</i>	54
3.7. Teilstudie 7: <i>Übertrittsempfehlungen aus Sicht der Lehrkräfte: Pro Klasse im Durchschnitt eine Gymnasialempfehlung zu viel</i>	62
3.8. Teilstudie 8: <i>Wird von Lehrkräften das familiäre Umfeld bei der Übertrittsentscheidung (zu) wichtig genommen?.....</i>	65
3.9. Teilstudie 9: <i>Persistierende Geschlechterdisparitäten: Verhaltensunterschiede und die Wirkung von Stereotypen.....</i>	68
3.10. Teilstudie 10: <i>Wirkungen primärer und sekundärer Herkunftseffekte sowie von Merkmalen der Lehrperson auf die Benotung schulischer Leistungen</i>	77
4. Resümee, Ausblick und Einschränkungen	86
Anhang 1: Beschreibung der TIMSS-Übergangsstudie.....	88
Anhang 2: Beschreibung der ELEMENT-Studie	94
Anhang 3: Beschreibung der TOSCA-Studie	98
Anhang 4: Das Bildungssystem in der Schweiz.....	102
Abbildungs- und Tabellenverzeichnis	104
Literaturverzeichnis	108

Zentrale Ergebnisse

Herkunft wird mit zensiert

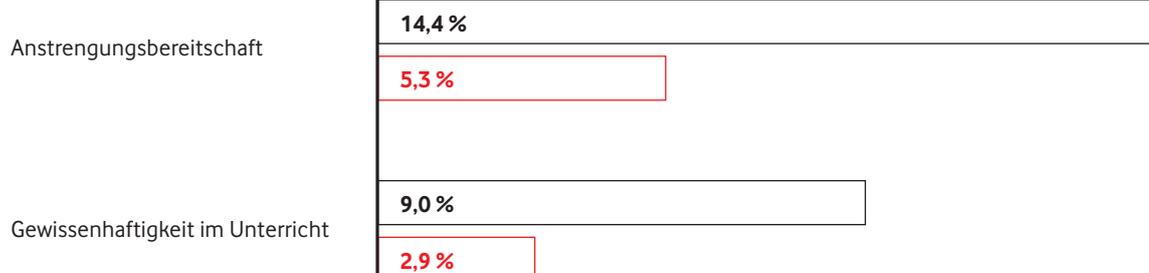
Welche Merkmale erklären Unterschiedlichkeit bei der Benotung
(Anteil der Varianz an der Durchschnittsnote)



In Noten fließt nicht nur die Leistung ein

Einfluss motivationaler Merkmale der Schüler
auf die Durchschnittsnote (Korrelation, Angabe in Prozent)

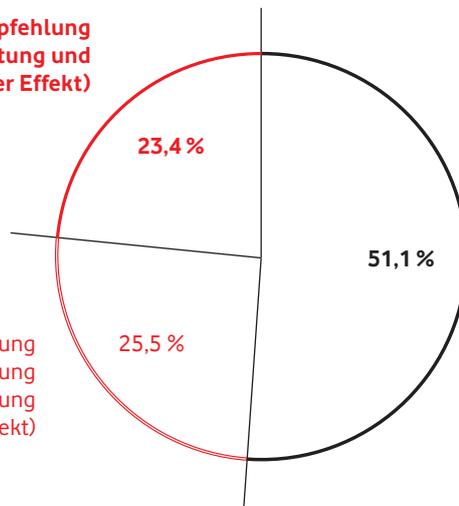
Einfluss motivationaler Merkmale der Schüler auf das Ergebnis
eines standardisierten Leistungstests (Korrelation, Angabe in Prozent)



Wie sich die soziale Herkunft auf die Vergabe der Schulempfehlung auswirkt (Prozentuale Zerlegung des Herkunftseffekts bei der Empfehlungsvergabe)

Unterschiedliche Schulempfehlung
der Lehrkräfte bei gleicher Leistung und
gleichen Noten (Inkrementeller Effekt)

Unterschiedliche Empfehlung
wegen ungleicher Benotung
der Lehrer bei gleicher Leistung
(Indirekter sekundärer Effekt)



Unterschiedliche Schulempfehlung
wegen unterschiedlicher Leistung
je nach sozialer Schichtzugehörigkeit
(Primärer Effekt)

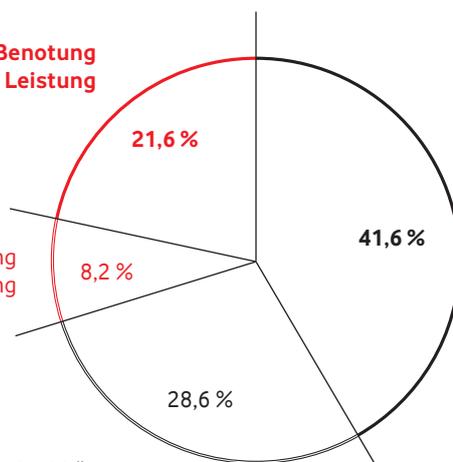
→ Wenn Benotungen und Schulempfehlungen sozial neutral wären, könnte der Einfluss der Sozialschicht auf die Schulempfehlung um die Hälfte reduziert werden!

Wie sich die soziale Herkunft auf den tatsächlichen Übergang auf die weiterführende Schule auswirkt (Prozentuale Zerlegung des Herkunftseffekts beim Übergang)

Unterschiedliche Benotung
der Lehrer bei gleicher Leistung

Unterschiedliche Schulempfehlung
der Lehrer bei gleicher Leistung

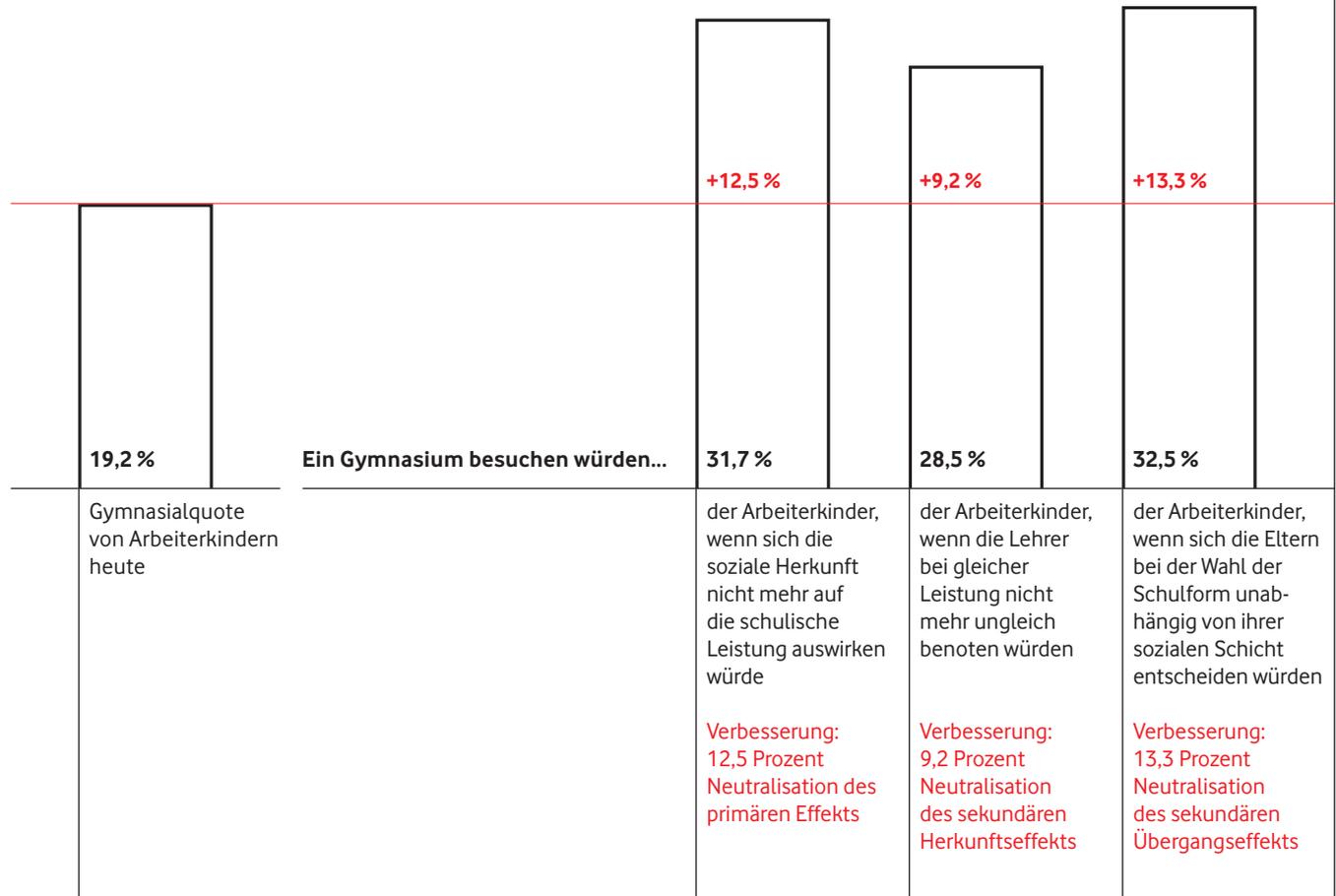
Schichtabhängige
Entscheidungen der Eltern



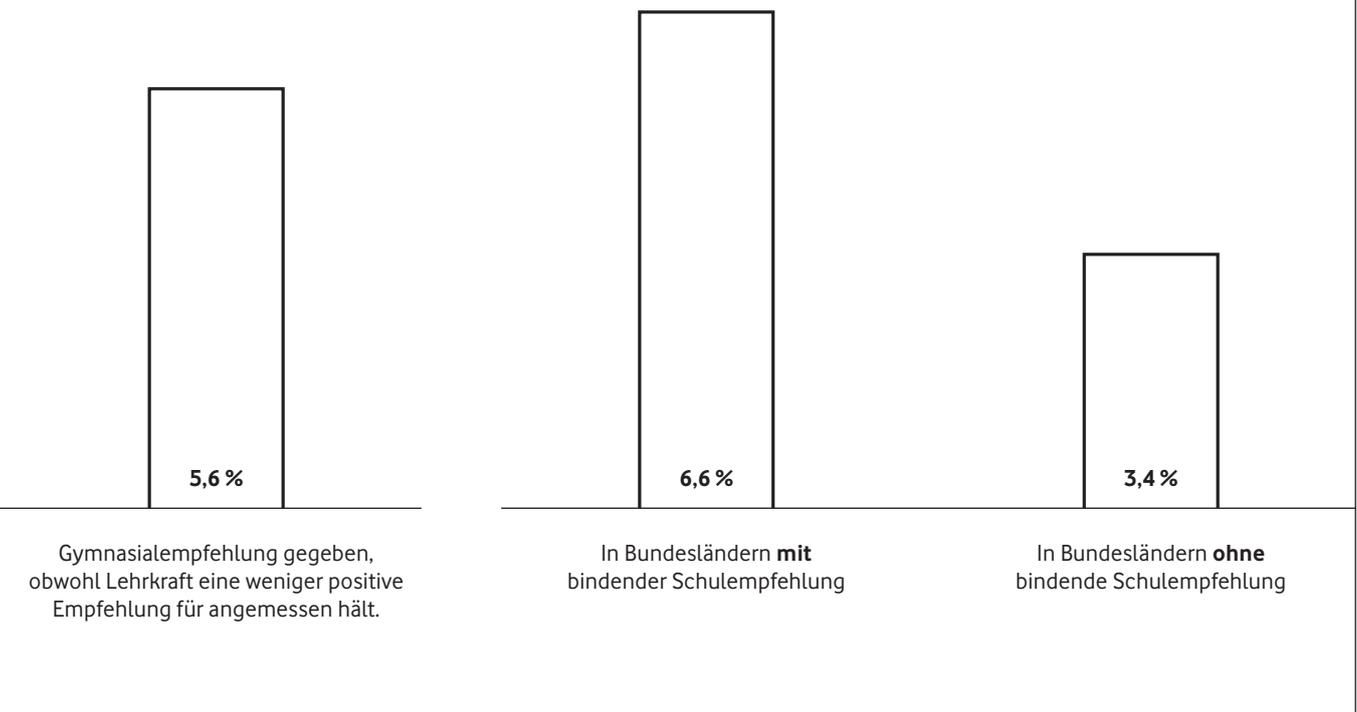
Unterschiedliche Leistung
(Primärer Effekt)

→ Die soziale Herkunft wirkt sich vor allem auf die Leistungen und damit auf den Übergang aus. Aber auch bei gleicher Leistung wirken sich die Benotungspraxis und die Schulempfehlungen der Lehrkräfte auf den Übergang aus (zusammen 29,8 Prozent). Eine große Rolle spielt auch das Entscheidungsverhalten der Eltern (anteiliger Effekt: 28,6 Prozent).

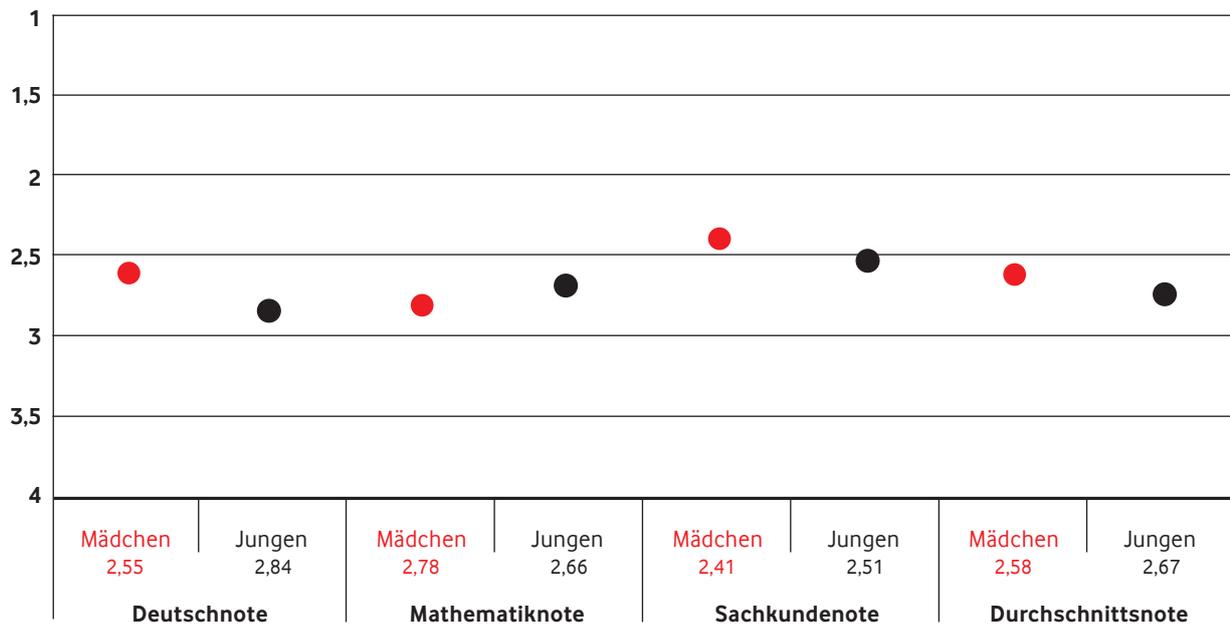
Wie würde sich die Quote der Arbeiterkinder auf dem Gymnasium verändern, wenn die Wirkung der sozialen Herkunft neutralisiert würde?



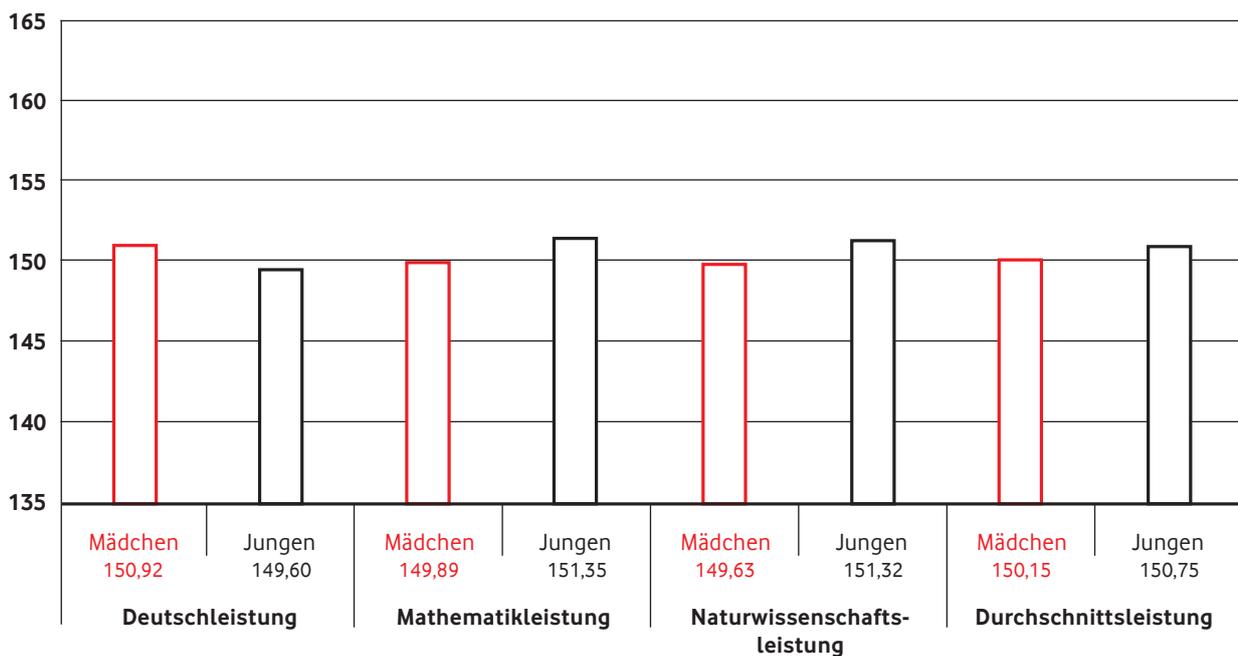
Wie oft geben Lehrer eine Gymnasialempfehlung, obwohl sie diese eigentlich nicht für angemessen halten?



Mädchen bekommen im Durchschnitt bessere Noten



Mittelwerte der Leistungstests differenziert nach Geschlecht



1. Einleitung

Kaum ein Thema bewegt Eltern und Kinder mehr als die Bewertung der Leistung der Schülerinnen und Schüler durch die Lehrkräfte. Es gibt eine Fülle von anekdotischen Belegen dafür, dass die Leistungsbewertung oftmals ungerecht ist, und die erlebte Ungerechtigkeit wirkt oft noch lange nach. Befragt man Studierende, so sind fast alle in der Lage, zumindest eine Begebenheit aus der Schule zu schildern, in denen eine „falsche“ Note erteilt wurde.

Besonders relevant sind Leistungsbewertungen, wenn sie unmittelbare Konsequenzen für die weitere Schullaufbahn der Schülerinnen und Schüler haben. Dies ist beispielsweise dann der Fall, wenn die Versetzung in die nächste Klassenstufe gefährdet ist oder wenn mit der Leistungsbewertung grundlegende Entscheidungen über den weiteren Bildungsweg verbunden sind, wie es beispielsweise bei Übertrittsempfehlungen der Fall ist. Das letztgenannte Thema ist entsprechend auch von großem Interesse für die empirische Forschung, und diese Forschung hat, gerade in dem vergangenen Jahrzehnt, eine Fülle von Ergebnissen hervorgebracht. Dabei konnte immer wieder gezeigt werden, dass die Leistungsdiagnostik nicht frei von Fehlern ist. Das Zusammenwirken von sogenannten systematischen und unsystematischen Fehleinflüssen sorgt dafür, dass die Qualität von Noten und Übertrittsempfehlungen aus psychometrischer Sicht insgesamt nicht befriedigend ausfällt.

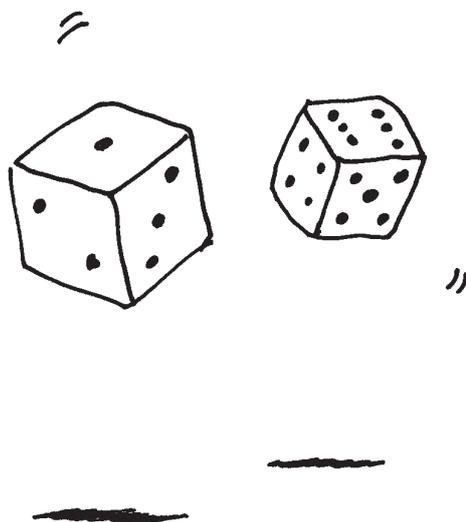
In der vorliegenden Expertise steht die Frage im Mittelpunkt, inwieweit bei der Leistungsdiagnostik auch Effekte des sozialen Hintergrunds der Schülerinnen und Schüler eine Rolle spielen, insbesondere in Bezug auf die kritische Phase beim Übertritt von der Grundschule in die Bildungsgänge der Sekundarstufe I. In anderen Worten: Wie stark hängt die Bewertung der Leistung einer Schülerin/eines Schülers – auch – davon ab, welchen familiären

Hintergrund sie/er besitzt? Werden Schülerinnen und Schüler mit eher ungünstigem sozialen Hintergrund bei vergleichbaren Leistungen ungünstiger bewertet als ihre Mitschülerinnen und Mitschüler mit günstigerem sozialen Hintergrund?

Die Expertise gliedert sich in zwei große Abschnitte. In einem ersten großen Abschnitt (Kapitel 2) wird ins Thema eingeführt, indem die wichtigsten Begriffe geklärt werden und die vorliegende Literatur aufbereitet wird. Da in Hinblick auf die generelle Rolle und psychometrische Qualität von Noten sowie in Bezug auf soziale Disparitäten bei Übertrittsentscheidungen bereits eine Fülle gut zugänglicher Original- und Übersichtsarbeiten vorliegen, sind die entsprechenden Passagen knapp gefasst. Im zweiten großen Abschnitt (Kapitel 3) werden dann neue bzw. ergänzte Analysen zu insgesamt zehn zentralen Fragen zu (sozialen) Ungleichheiten bei der Notenvergabe und den Übertrittsempfehlungen vorgestellt. Die Expertise schließt mit einem knappen Resümee.

Ein Teil der in dieser Expertise vorgestellten Befunde entstammen derzeit noch laufenden Forschungsarbeiten an den Universitäten Potsdam, Freiburg (Schweiz) und Tübingen, die maßgeblich von Christiane Fiege, Michaela Kropf, Marko Neumann und Christian Wandeler vorangetrieben werden und in wissenschaftliche Artikel münden sollen. Wir haben von dem intellektuellen Input unserer Mitarbeiterinnen und Mitarbeiter sehr profitiert und danken ihnen ausdrücklich dafür. Bei Katharina Konietzko möchten wir uns für die administrative Unterstützung bedanken. Jürgen Baumert danken wir herzlich für die vielen Gespräche über die Effekte des sozialen Hintergrunds, die einen Niederschlag in manchen der hier untersuchten Teilfragestellungen fanden. Wir danken ihm darüber hinaus für einen großzügigen Datenzugang. Ab

schließlich möchten wir der Vodafone-Stiftung herzlich dafür danken, dass sie uns eine größtmögliche Freiheit bei der Schwerpunktsetzung der Expertise gegeben hat und die Durchführung ergänzender Analysen explizit begrüßt hat, auch auf die immer bestehende Gefahr hin, dass diese Analysen nur wenig zum Erkenntnisfortschritt beitragen. Dieses Vertrauen hat dazu geführt, dass – für eine Expertise ist diese Schwerpunktsetzung durchaus unüblich – das Verhältnis von „neuen“ zu „alten“ Erkenntnissen zugunsten der „neuen“ Befunde ausfällt.



2. Theoretischer und empirischer Hintergrund

In diesem Kapitel wird der theoretische und empirische Hintergrund der Expertise erarbeitet. Dabei wird herausgearbeitet, was unter Leistungsbewertungen zu verstehen ist und welche Funktionen Noten haben. Anschließend werden wichtige Aspekte zur Beschreibung und Erfassung des sozialen Hintergrunds genannt. Da sich in den Analysen zum Einfluss der sozialen Herkunft auf den Bildungserwerb die Trennung zwischen primären und sekundären Herkunftseffekten als sehr nutzbringend erwiesen hat, werden die zentralen theoretischen Konzeptionen kurz beschrieben. Abschließend wird auf psychometrische Merkmale von Leistungsbewertungen eingegangen. In einem Exkurs soll beispielhaft für die Schweiz thematisiert werden, welche Konsequenzen die Befunde wissenschaftlicher Studien für die Qualität Leistungsdiagnostik haben.

2.1. Leistungsbewertungen

Unter dem Begriff „*Schulleistung*“ (eigentlich besser: „Schülerleistung“) eines Kindes können alle Leistungs- und Verhaltensaspekte subsumiert werden, denen unter den jeweiligen Rahmenbedingungen ein positiver Wert zugeschrieben wird. Zu den Schulleistungen gehört damit ein breites Spektrum von Verhaltensweisen und Leistungen wie beispielsweise das Rechnen, Lesen und Schreiben, aber auch motivationale Merkmale wie die Mitarbeit im Unterricht bzw. die sorgfältige Erledigung der Hausaufgaben sowie das soziale Verhalten.

Schulnoten (auch „Zensuren“ genannt) sind Variablen, mit denen diese Leistungs- und Verhaltensaspekte in Bewertungen/Zahlen überführt werden. In der Regel sind Lehrkräfte für diese Bewertung zuständig; sie erteilen die Noten nach einem mehr oder weniger expliziten und strukturierten Bewertungsprozess. In die Noten fließen –

je nach Funktion der jeweiligen Note, der Vorgabe durch die zuständigen Behörden sowie der persönlichen Präferenz der Lehrkräfte – unterschiedliche Verhaltens- und Leistungsaspekte mit unterschiedlicher Gewichtung ein.

Schulnoten haben mehrere Funktionen. Tent (2006) unterscheidet zwischen einer *pädagogischen Funktion* und einer *gesellschaftlichen Funktion*. Im Sinne der pädagogischen Funktion erfolgt (1) eine Sozialisation durch Noten, da die Schülerinnen und Schüler mit Normen und Leistungsvergleichen vertraut gemacht werden, liefern (2) Noten wichtige Rückmeldungen für die Betroffenen, die u.a. als Hinweise für Handlungsnotwendigkeiten wirken und (3) motivieren und disziplinieren Noten, indem sie Lernanreize bieten. Um ihrer pädagogischen Funktion gerecht zu werden, müssen Noten nicht zwingend ein objektives Abbild der Wirklichkeit darstellen. Anders sieht es in Hinblick auf die zweite Funktion der Noten, der *gesellschaftlichen Funktion*, aus. Noten besitzen in dieser Hinsicht Berechtigungsfunktionen, sie sollen die Verteilung attraktiver Güter (z.B. Studienplätze, Ausbildungsplätze) nach Leistungskriterien ermöglichen und damit meritokratische Verteilungsmechanismen erlauben (Klassifikations-, Allokations- und Selektionsfunktion). Noten haben darüber hinaus eine Kontrollfunktion, beispielsweise indem sie strukturelle oder pädagogische Veränderungen in den Schulen reflektieren.

Schulnoten fließen ganz wesentlich in die zum Ende der Grundschule erteilten *Übertrittsempfehlungen* (auch Schulformempfehlungen, Übertrittsgutachten o.ä. genannt) ein. In manchen Bundesländern bzw. entsprechenden Gebietseinheiten in der Schweiz und Österreich sind die Übertrittsempfehlungen eine direkte Funktion der Schulnoten in als besonders relevant betrachteten Schulfächern, in anderen werden sie ergänzt um weitere Beurteilungsaspekte (Maaz, Neumann, Trautwein, Wendt, Lehmann & Baumert, 2008; Baeriswyl et al., 2006).

Auch wenn sich durch die voranschreitende Entkopplung von Bildungsgang und Bildungsabschluss Korrekturmöglichkeiten auftun (vgl. Köller, Baumert & Schnabel, 1999; Baumert, Trautwein & Artelt, 2003; Köller, Watermann, Trautwein & Lüdtke, 2004; Trautwein, Baeriswyl, Lüdtke & Wandeler, 2008) und bestehende flexible Anschlussoptionen für Bildungslaufbahnen einen späteren Erwerb von angestrebten Abschlüssen ermöglichen, handelt es sich bei dem Übergang von der Grundschule in die weiterführende Schule noch immer um eine Weichenstellung, deren Bedeutung für die Bildungsbiografien von Schülerinnen und Schülern nicht unterschätzt werden sollte (Baumert, Maaz & Trautwein, 2009; Ditton & Krüsken, 2006; Schnabel, Alfeld, Eccles, Köller & Baumert, 2002).

Dies erklärt, warum der Übertrittsempfehlung eine sehr hohe Bedeutung zukommt. Eine Besonderheit ist der Verbindlichkeitsgrad der Empfehlung, der sich erheblich zwischen den Ländern unterscheiden kann. Ein Teil der Länder (Bremen, Hamburg, Hessen, Niedersachsen, Rheinland-Pfalz, Schleswig-Holstein) gibt nach Abschluss des vorgesehenen Beratungsprozesses den Elternwillen frei: Eltern können ihr Kind an der Schulform ihrer Wahl anmelden, auch wenn die Grundschule – abweichend vom Elternwunsch – eine weniger anspruchsvolle Schulform empfiehlt. Andere Länder (Baden-Württemberg, Bayern, Nordrhein-Westfalen, Saarland, Sachsen, Sachsen-Anhalt) sehen für den Fall, dass sich Eltern für eine anspruchsvollere Schulform als von der Grundschule empfohlen entscheiden, eine formelle Eignungsüberprüfung in Form von Probeunterricht oder Tests vor, die bei Misserfolg zur Abweisung des Schülers oder der Schülerin führt – wir bezeichnen die Übertrittsempfehlung in diesen Bundesländern deshalb (etwas vereinfachend) als „bindend“ (vgl. Füssel, Gresch, Baumert & Maaz, 2010). Der Verbindlichkeitsgrad der Übergangsempfehlung wird immer dann potenziell handlungsbedeutsam, wenn die Empfehlung hinter den Schullaufbahnwünschen der Eltern zurückbleibt oder Eltern diese Diskrepanz befürchten. Besteht unabhängig von der Empfehlung letztlich Schulformwahlfreiheit der Eltern, können sie ihren Bildungswunsch ohne zusätzliche Eignungsüberprüfung erfüllen. Die Bedeutung der Übertrittsempfehlung ist folglich besonders groß in solchen Bundesländern, in denen die Übertrittsempfehlung eine bindende Wirkung hat.

Leistungsbewertungen können auch durch *standardisierte Schulleistungstests* erfolgen. Hiermit sind in der Regel Tests gemeint, denen ein aufwändiger Konstruk-

tionsprozess vorangeht, der eine hohe psychometrische Qualität (siehe unten) gewährleistet, und die für eine Bewertung der Schulleistung nach einer für alle Schülerinnen und Schüler (auch über unterschiedliche Klassen hinweg) gleichen Bewertungsgrundlage sorgen. Besonders bekannt sind die Schulleistungstests aus Studien wie PISA oder TIMSS. Zu nennen sind darüber hinaus auch speziell für den Einsatz beim Übertritt in die Sekundarstufe konstruierte Leistungstests sowie kommerziell angebotene Tests, die u.a. zur Identifikation von Leistungsproblemen eingesetzt werden. Oftmals wird standardisierten Leistungstests zugesprochen, dass sie die Leistungsfähigkeit von Schülerinnen und Schülern besonders „objektiv“ erfassen und deshalb besonders „gerecht“ seien. Dies ist jedoch nur in dem Maße der Fall, wie sie eine hohe psychometrische Qualität aufweisen und tatsächlich diejenigen Leistungs- und Verhaltensmerkmale erfassen, die für die Notengebung bzw. Übertrittsentscheidung relevant sind; man spricht hier auch von der „Validität“ eines Tests. Diese Validität ist sicherlich nur in beschränktem Maße gegeben, da Schulleistungstests in aller Regel nur die kognitiven Leistungsaspekte erfassen.

Einen Spezialfall von standardisierten Schulleistungstests stellen Tests dar, die die kognitiven Grundfähigkeiten bzw. Teilaspekte der psychometrischen Intelligenz erfassen. Im engeren Sinne sind sie keine Schulleistungstests, da sie gerade auch darauf abzielen, solche Begabungaspekte zu messen, die relativ wenig von der Qualität und Quantität der Beschulung abhängig sind. Je stärker die Schulleistungen von Kindern in unterschiedlichen Klassen durch Unterschiede in der Qualität des Unterrichts beeinflusst sind, desto wichtiger kann der Einsatz von Tests zur Erfassung der kognitiven Grundfähigkeit als ein Regulativ sein, das dafür sorgt, vorhandenes Begabungspotenzial nicht zu übersehen. Da die Bearbeitung von Tests zur Erfassung kognitiver Grundfähigkeiten in aller Regel nur geringe sprachliche Kompetenzen erfordern, können sie darüber hinaus helfen, solche Schülerinnen und Schüler zu identifizieren, bei denen sprachliche Defizite höheren Leistungen im Wege stehen. Dies betrifft vor allem – aber nicht nur – Schülerinnen und Schüler mit Migrationshintergrund. Allerdings muss einschränkend bemerkt werden, dass es eine Reihe von Hinweisen darauf gibt, dass auch Tests zur Erfassung der kognitiven Grundfähigkeit von der Quantität und Qualität der Beschulung beeinflusst werden (vgl. Becker 2009; Ceci, 1991), so dass auch beim Einsatz dieser Tests kein „objektiver“ Wert für das Leistungspotenzial resultiert.

Klassenarbeiten (auch Klausuren genannt) sind schriftliche Leistungsprüfungen, die in der Regel von der jeweiligen Lehrkraft konstruiert und ausgewertet werden. Sie weisen eine hohe curriculare Passung auf, da sie sich auf den Unterrichtsstoff der zurückliegenden Wochen beziehen. Ihre psychometrische Qualität wird in der Regel nicht geprüft.

Zur Schulleistung von Schülerinnen und Schülern gehört neben den Leistungen, die die Schülerinnen und Schüler in Klassenarbeiten oder standardisierten Leistungstest zeigen, eine Reihe weiterer Leistungs- und Verhaltensaspekte, die ebenfalls in die Noten einfließen und oftmals als „*mündliche Note*“ bezeichnet werden. Hierzu gehören u.a. die Mitarbeit im Unterricht, die Erledigung der Hausaufgaben sowie das unterrichtsbezogene und soziale Verhalten in der Schule. Welche dieser Aspekte in welcher Weise notenwirksam werden, bleibt oftmals der jeweiligen Lehrkraft überlassen.

2.2. Sozialer Hintergrund

In den vergangenen Jahren hat sich in der Bildungs- und Sozialforschung eine zunehmend differenzierte Erfassung des sozialen Hintergrunds durchgesetzt. Möchte man die Effekte der sozialen Herkunft umfassend und adäquat erfassen, sollte dies nach Möglichkeit nicht anhand nur einer Variable, sondern über mehrere Dimensionen erfolgen, da nur dadurch die relative Bedeutung der einzelnen Dimensionen sowie der Gesamteffekt des sozialen Hintergrunds abgeschätzt werden können. Dies trifft sowohl für den Kompetenzerwerb (vgl. Baumert, Watermann & Schümer, 2003; Jungbauer-Gans, 2004, 2006) als auch die Notengebung und Übertrittsempfehlungen zu. In der folgenden Beschreibung wichtiger Indikatoren der sozialen Herkunft folgen wir eng der Darstellung in Maaz (2006), Maaz, Baumert und Cortina (2008) sowie Maaz, Trautwein und Dumont (2011).

Die soziale Herkunft von Schülerinnen und Schülern wird üblicherweise mithilfe der sozioökonomischen Stellung ihrer Familien bestimmt, das heißt mithilfe von Daten zur relativen Position ihrer Eltern in einer sozialen Hierarchie, deren Ordnungsprinzipien in der Verfügung über finanzielle Mittel, Macht oder Prestige bestehen. Da Informationen über Einkommensverhältnisse, Macht und soziale Anerkennung von Individuen nicht einfach zu erhalten sind, wird die sozioökonomische Stellung in aller Regel über die Berufstätigkeit erfasst, die Hinweise

auf jeden der drei Aspekte ihrer Stellung in der sozialen Hierarchie geben kann (Ganzeboom, De Graaf, Treiman & De Leeuw, 1992). Erst seit einigen Jahren werden – im Anschluss an die Arbeiten von Bourdieu (1983) und Coleman (1988, 1996) – zuweilen auch das „kulturelle“ und das „soziale“ Kapital der Familien mitberücksichtigt, wodurch ein etwas differenzierteres Bild des sozialen Hintergrunds entsteht. Die Begriffe kulturelles bzw. soziales Kapital werden für alle kulturellen und sozialen Ressourcen benutzt, die die Handlungsmöglichkeiten von Personen erweitern und folglich auch ihre sozioökonomische Stellung positiv beeinflussen können.

Coleman geht davon aus, dass Bildungseinrichtungen nur dann erfolgreich arbeiten können, wenn die Kinder und Jugendlichen soziales Kapital besitzen, das heißt, wenn sie in einem Netzwerk sozialer Beziehungen groß geworden sind, das die Übernahme sozial anerkannter Ziele, Werte und Einstellungen fördert und unterstützt. Normalerweise wird soziales Kapital in der Familie gebildet, in Verwandtschafts- und Nachbarschaftsgruppen, in religiösen oder ethnischen Gruppen, in Vereinen, Betrieben oder politischen Parteien. Durch die Struktur der sozialen Beziehungen in diesen Gemeinschaften wird ein Netz aus wechselseitigen Erwartungen und Verpflichtungen erzeugt, das Vertrauen bildet und Zusammenarbeit ermöglicht. Bourdieu (1983) definiert das soziale Kapital als die „Gesamtheit der aktuellen und potenziellen Ressourcen, die mit dem Besitz eines dauerhaften Netzes von mehr oder weniger institutionalisierten Beziehungen gegenseitigen Kennens oder Anerkennens verbunden sind“ (S. 190). Es handelt sich dabei also um Ressourcen, die zur Akkumulation von ökonomischem Kapital beitragen können. Da das soziale Kapital – zumindest bei den gängigen Operationalisierungen – nur einen schwachen Zusammenhang mit Schulleistungen aufweist, wird es in

den meisten Schulleistungsstudien nicht näher analysiert (Maaz, Trautwein, Gresch & Lüdtke, 2009).

Unter kulturellem Kapital subsumiert Bourdieu alle Kulturgüter und kulturellen Ressourcen, die – als symbolische Machtmittel – dazu beitragen, dass in einem sozialen System die Qualifikationen, Einstellungen und Wertorientierungen vermittelt werden, die das System zu seiner Bestandserhaltung braucht. Bei Kulturgütern und kulturellen Ressourcen handelt es sich keineswegs nur um Sachgüter wie Kunstwerke oder Literatur, sondern auch um institutionalisierte Formen potenzieller Macht wie zum Beispiel Bildungszertifikate oder Titel. Insbesondere gehören zu den kulturellen Ressourcen die Wahrnehmungs-, Denk- und Handlungsschemata, die eine Person verinnerlicht hat. Das kulturelle Kapital kann in drei Formen auftreten: (1) als verinnerlichtes bzw. inkorporiertes Kapital (was von Bourdieu auch als „Habitus“ bezeichnet wird), (2) als Besitztum von kulturellen Gütern wie z.B. Bücher und (3) in institutionalisiertem Zustand z.B. in Form von Bildungszertifikaten.

Um die Stellung von Personen in der sozialen Hierarchie einer Gesellschaft zu bestimmen, wurde eine Reihe von Maßen entwickelt (Berufsrangskalen und Kategoriensysteme). Berufsprestigemaße beruhen auf einer empirisch gewonnenen Einschätzung der gesellschaftlichen Anerkennung von Berufen. Heute gehören Prestigemaße zum Standardrepertoire sozialwissenschaftlicher Umfragen (Ganzeboom, de Graaf & Treiman, 1992). Der wohl bekannteste international vergleichbare Index für das Berufsprestige ist der Treiman-Index (Treiman, 1977), der auf einer Berufsklassifikation von 1968 (ISCO-68) beruht und zum Standard Index of Occupational Prestige Scale (SIOPS) weiterentwickelt wurde (Ganzeboom & Treiman, 1996).

Neben dem Prestigemaß haben Ganzeboom und Mitarbeiter (1992) ein weiteres international vergleichbares Messinstrument entwickelt, das den sozioökonomischen Status der beruflichen Tätigkeit misst. Der International Socio-Economic Index of Occupational Status (ISEI) hat den Vorzug, dass er die ökonomische Stellung vom Berufsprestige trennt. Dem liegt die Annahme zugrunde, dass Schul- und Berufsbildung über Berufe in Einkommen und in Chancen zur Teilhabe an Macht umgesetzt werden; da Berufe bestimmte Qualifikationen voraussetzen und zu bestimmten Einkommen führen, sind sie als Vermittler zwischen Bildungsabschlüssen und Einkommenslagen zu betrachten.

Während die Prestigemaße und die sozioökonomischen Indizes von relativ feinen quantifizierbaren Unterschieden

zwischen Berufen ausgehen, betonen andere Ansätze zur Erfassung der sozioökonomischen Stellung die qualitativen Differenzen zwischen sozialen Klassen. In der soziologischen Forschung hat sich das von Erikson, Goldthorpe und Portocarero (1979) entwickelte Kategoriensystem durchgesetzt (sog. EGP-Klassen) (Erikson & Goldthorpe, 1992). Für den internationalen Vergleich liegt mit der European Socio-economic Classification (ESeC) ein weiteres Kategoriensystem vor, das auf dem EGP-Klassenschema basiert (vgl. Hausen, Jungblut, Müller, Pollak & Wirth, 2006; Müller, Wirth, Bauer, Pollak & Weiss, 2006). Mayer und Aisenbrey (2007) haben darüber hinaus eine Klassifikation (German Employment Status Class Schema [GEC]) entwickelt, die explizit auf das deutsche Beschäftigungssystem für innerdeutsche Analysen ausgerichtet ist.

Die differenzierte Konzeption der sozialen Herkunft steht im Einklang mit der Perspektive, dass Statusmerkmale wie z.B. die sozioökonomische Stellung einer Familie alleine nicht ausreichen, um den Zusammenhang zwischen sozialer Herkunft und Schulleistungen zu verstehen. So konnten Baumert, Watermann und Schümer (2003) anhand von Daten der PISA-Studie zeigen, dass der Effekt der Sozialschichtzugehörigkeit auf die Schulleistungen zu einem großen Teil über familiäre Prozessmerkmale wie der kulturellen und kommunikativen Praxis in der Familie vermittelt wird. Sie kommen daher zu dem Schluss, dass erst die simultane Berücksichtigung von Struktur- und Prozessmerkmalen des familiären Hintergrunds eine angemessene Vorstellung der anzutreffenden sozialen Disparitäten liefert (Baumert et al., 2003b; Maaz & Watermann, 2007).

2.3. Primäre und sekundäre Herkunftseffekte

Einigkeit besteht in der Bildungs- und Sozialstrukturforschung dahingehend, dass die entscheidenden Stationen für die Entstehung von Bildungsungleichheiten die Gelenkstellen von individuellen Bildungsverläufen bzw. die entsprechenden Übergänge im Bildungssystem sind (Baumert & Schümer, 2001; Breen & Goldthorpe, 1997; Bellenberg & Klemm, 1998; Ditton, 1992; Schnabel, Alfeld, Eccles, Köller & Baumert, 2002). Besonders gut untersucht ist der Übergang von der Grundschule in die weiterführenden Schulen. Ein Großteil der vorhandenen Studien stützt sich dabei auf das Modell der primären und sekundären Disparitäten von Boudon (1974). Als primäre Herkunftseffekte werden Einflüsse der sozialen Herkunft

bezeichnet, die sich direkt auf die Kompetenzentwicklung der Heranwachsenden auswirken und in den unmittelbar übergangsrelevanten schulischen Leistungen (insbesondere der Schulnoten) der Kinder sichtbar werden. Als sekundäre Herkunftseffekte werden jene sozialen Disparitäten bezeichnet, die unabhängig von der Kompetenzentwicklung und dem erreichten Kompetenzniveau, aus unterschiedlichen Bildungsaspirationen und einem unterschiedlichen Entscheidungsverhalten Angehöriger verschiedener Sozialschichten resultieren (hierzu: Maaz, Baumert & Trautwein, 2009; Maaz & Nagy, 2009).

Die theoretische Konzeption primärer und sekundärer Effekte hat sich für die Analyse von Bildungsungleichheiten bewährt. Für die Berechnung der entsprechenden Effektgrößen ist der verwendete Leistungsindikator von besonderer Wichtigkeit. Werden Schulnoten als Leistungsindikator verwendet, können differenzielle Benotungseffekte zu Verzerrungen bei der Berechnung von primären und damit auch sekundären Herkunftseffekten führen. Dies trifft insbesondere dann zu, wenn sich Effekte der sozialen Herkunft auf die Benotung feststellen lassen. Sowohl der primäre als auch der sekundäre Effekt würde unterschätzt werden.

Für den Übergang in die Sekundarstufe I konnten in verschiedenen Studien primäre und sekundäre Effekte der sozialen Herkunft belegt werden (u. a. Arnold et al., 2007; Baumert & Schümer, 2001; Becker, 2000, 2003; Bos et al., 2004; Ditton, 2005, 2007; Ditton, Krüskens & Schauenberg, 2005; Ehmke et al., 2004; Ehmke, Hohensee, Siegle & Prenzel, 2006; Lehmann, Peek & Gänsfuß, 1997; Maaz & Nagy, 2009; Merckens & Wessel, 2002; Pietsch, 2007; Stocké, 2007). Die Ergebnisse zeigen zusammenfassend, dass Kinder aus sozial weniger begünstigten Familien im Vergleich zu Kindern aus sozial privilegierten Elternhäusern (1) über niedrigere schulische Kompetenzen und Noten verfügen, (2) bei gleichen Leistungen von den Lehrkräften schlechter bewertet werden, (3) auch unter Kontrolle der Schulleistungen und Noten geringere Chancen auf den Erhalt einer Gymnasialempfehlung haben und (4) Eltern ihr Kind schließlich bei Kontrolle von Leistungen seltener auf ein Gymnasium schicken.

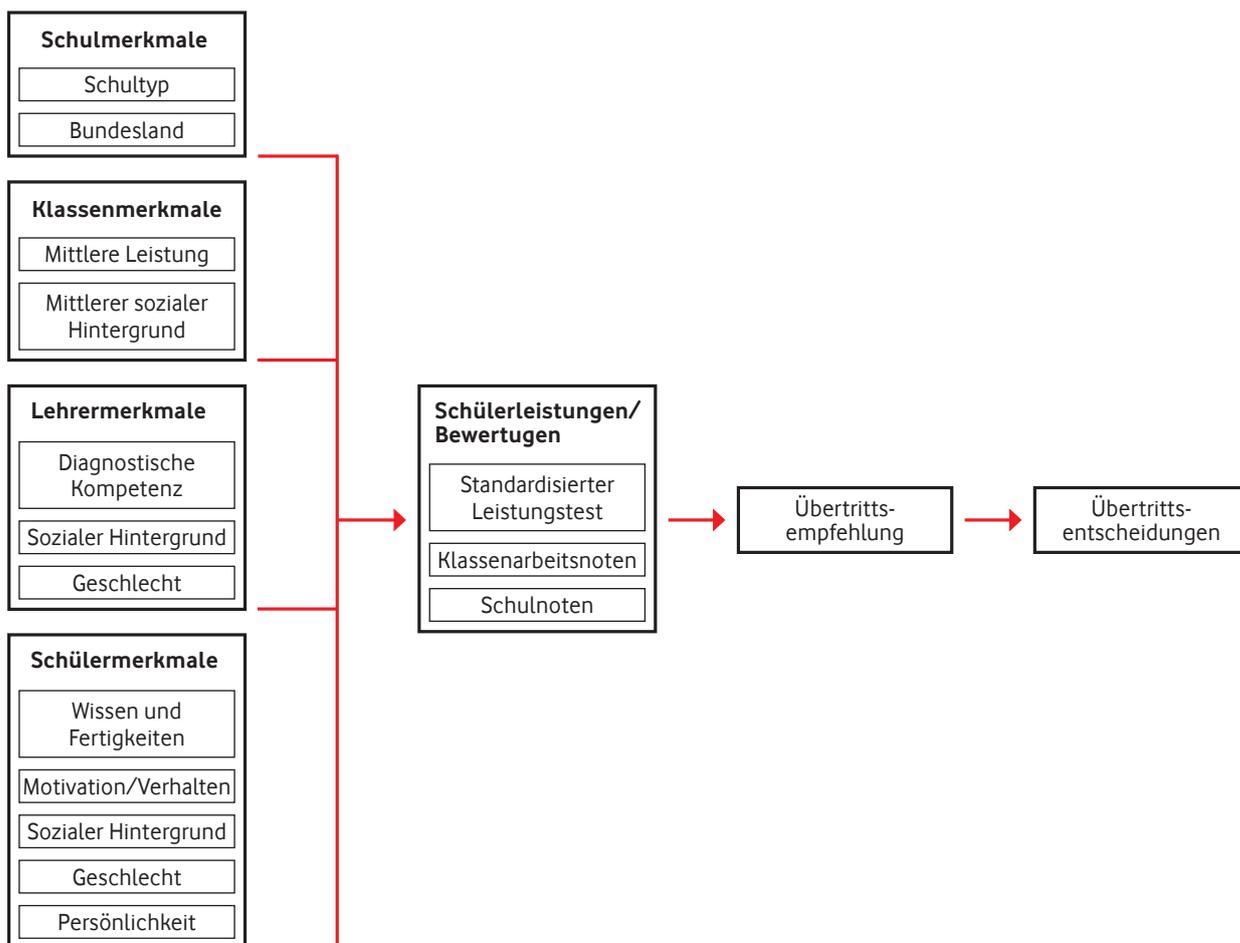
Neben Effekten der sozialen Herkunft auf den Übergang findet die Berücksichtigung des Migrationshintergrunds große Beachtung. Dieses Interesse resultiert nicht zuletzt aus den deskriptiven Betrachtungen, aus denen hervorgeht, dass Kinder mit Migrationshintergrund seltener auf dem Gymnasium vertreten sind als Kinder ohne Migrationshintergrund und auch geringere Chancen ha-

ben, das Abitur zu erwerben. Dieser deskriptive Befund muss etwas genauer betrachtet werden. In Analogie zu den primären und sekundären sozialen Herkunftseffekten lassen sich auch migrationsspezifische primäre Effekte – beispielsweise wenn Kinder aufgrund sprachlicher Barrieren schlechtere schulische Leistungen erzielen sowie migrationsspezifische sekundäre Effekte, wenn Kinder bei gleichen Leistungen aufgrund ihrer Migration eine andere Schulform besuchen als Kinder ohne Migration (vgl. Gresch & Becker, 2010). Die einschlägigen Analysen, die es ermöglichen, primäre und sekundäre Effekte bei Kindern mit Migrationshintergrund zu trennen, zeigen, dass der in bivariaten Analysen dokumentierte Migrationseffekt nicht mehr nachweisbar war, wenn für die soziale Herkunft kontrolliert wurde (Gresch & Becker, 2010; Kristen & Dollmann, 2009). Damit ist der Migrationseffekt im Grunde ein sozialer Effekt. Wenn zusätzlich die Leistungen der Schüler berücksichtigt wurden, kehrte sich der negative Migrationseffekt in einen positiven Effekt, zumindest für ausgewählte Zuwandergruppen (türkischstämmige Schülerinnen und Schüler; Gresch & Becker, 2010). Legt man das Leistungsprinzip zugrunde, kann auf der Grundlage dieser Befunde nicht von einer systematischen Benachteiligung von Kindern mit Migrationshintergrund durch das Bildungssystem gesprochen werden.

Die Betrachtung des Übergangsmechanismus ist dann von Bedeutung, wenn Kinder nicht nur leistungsbezogen auf bestimmte Bildungsgänge oder Schulen übergehen, sondern zusätzlich auch leistungsfremde Merkmale wie die soziale Herkunft diesen Übergang mitsteuern. Dass leistungsfremde Merkmale der Schülerinnen und Schüler Einfluss auf Bildungsübergänge nehmen, ist vielfach gezeigt worden, sie determinieren sie jedoch nicht.

2.4. Psychometrische Merkmale von Leistungsbewertungen und das Auftreten systematischer Verzerrungen

An den Noten und Übertrittsempfehlungen von Lehrkräften in den mehrgliedrigen Schulsystemen der deutschsprachigen Länder Deutschland, Österreich und Schweiz ist viel Kritik geübt worden. Bei aller Unterschiedlichkeit, die sich auch innerhalb der Staaten zwischen einzelnen Bundesländern/Kantonen finden, gilt, dass den diagnostischen Fähigkeiten der Lehrkräfte eine entscheidende Rolle zukommt. Aus psychologisch-diagnostischer Sicht (vgl. Ingenkamp, 1969; vgl. Schrader, 2006) ist schon früh

Abbildung 1: Ein vereinfachtes Arbeitsmodell zu möglichen Einflussfaktoren auf die Leistungsbewertung

kritisiert worden, dass die Objektivität, Reliabilität und Validität der eingesetzten Verfahren vielfach nicht den klassischen Gütekriterien genügen würden und dass es den meisten Lehrkräften an einer systematischen Ausbildung ihrer diagnostischen Fähigkeiten fehle. Dies dürfte u. a. dazu beitragen, dass der (unsystematische) Fehleranteil bei Noten und Übertrittsempfehlungen substantiell ausfällt (vgl. z. B. Baeriswyl, Wandeler & Christ, 2008).

Übersichtsarbeiten zur Objektivität, Reliabilität und Validität von Schulnoten (z. B. Ingenkamp, 1971; Jäger, 2004; Ziegenspeck, 1999) zeigen deutlich, dass diese die üblicherweise an die Individualdiagnostik angelegten Qualitätsmaßstäbe nicht erreichen. Allerdings sind Schulnoten in psychometrischer Hinsicht vielleicht auch nicht so defizitär, wie manchmal der Eindruck erweckt wird. Verschiedene Arbeiten haben gezeigt, dass unterschiedliche Lehrkräfte bei der Beurteilung der Leistung von Schülerinnen und Schülern einer bestimmten Schulklasse insgesamt zu ähnlichen Rangreihen kommen. Dies ist auch der Fall bei Bewertungen von Aufsätzen, die gemeinhin als besonders problematisch beschrieben werden (vgl. Birkel & Birkel, 2002). Gleichzeitig treten trotz einer relativ hohen Übereinstimmung in einer Häufigkeit substantielle Streuungen in den Bewertungen auf, die in Hinblick auf die angestrebte Gleichbehandlung aller Schülerinnen und Schüler nicht akzeptabel sind. Unter anderem wegen der den Lehrkräf-

ten nur teilweise bewussten fehlenden Vergleichbarkeit von Leistungen über unterschiedliche Klassen hinweg fallen zudem die Zusammenhänge zwischen Noten und standardisierten Leistungstest bei Einbezug vieler Klassen vergleichsweise mager aus (z. B. Baumert et al., 2003a).

Dabei gilt bei aller berechtigten Kritik an Leistungsbewertungen durch Lehrkräfte zu berücksichtigen, dass natürlich auch Ergebnisse aus standardisierten Schulleistungstests nicht beanspruchen können, ein „valides“ Abbild der Leistung zu vermitteln. Wenn der Zusammenhang zwischen standardisiertem Leistungstest und Schulnote nur gering ausfällt, so kann dies beispielsweise auch daran liegen, dass die Validität eines bestimmten Leistungstests in Bezug auf das intendierte Schulcurriculum begrenzt ist, da er vielmehr zur Messung „allgemeiner“ Kompetenzen entwickelt wurde. Zudem mögen in Schulnoten zu einem relativ großen Anteil auch Aspekte von Motivation und Verhalten einfließen, die typische standardisierte Leistungstests nicht erfassen. In dem Maße, wie die Berücksichtigung solcher Aspekte gewollt ist, verringert sich die Validität eines standardisierten Leistungstests. Dies zeigt, wie wichtig es ist, bei der Prüfung von unerwünschten Disparitäten sehr genau zu spezifizieren, welche Schülermerkmale, Leistungsbewertungen etc. sowie welche Prozesse betrachtet werden.

(→ **ABBILDUNG 1**)

In der Abbildung 1 sind Variablen bzw. Variablenbündel aufgeführt, die in Hinblick auf die Identifikation von unerwünschten Disparitäten von Belang sein können. Neben einer Reihe von Schülermerkmalen werden die Ebene der Lehrkräfte, die Klassenebene und die Schulebene differenziert, die allesamt einen Effekt auf Leistungsbeurteilungen aller Art haben können. Die Abbildung erhebt keinen Anspruch auf Vollständigkeit; im Gegenteil: Es ist davon auszugehen, dass weitere Faktoren relevant sind, beispielsweise bei den Schüler- oder Lehrermerkmalen, und dass es eine Vielzahl weiterer „Effekte“ gibt, beispielsweise direkte Effekte von Schülermerkmalen auf die Übertrittsempfehlung oder die Übertrittsentscheidung. Trotzdem hilft diese bereits stark reduzierte Abbildung der Wirklichkeit, zentrale Forschungsdesiderata in der Analyse von sozialen Disparitäten bei der Leistungsbeurteilung zu identifizieren. So differenziert die Abbildung zwischen einer Reihe von Schülermerkmalen, die bewertungsrelevant sein könnten; dabei sind neben Wissen und Fähigkeiten motivationale und verhaltensbezogene Aspekte, deren Relevanz für bestimmte Schulleistungen unbestritten ist, aber auch der soziale Hintergrund genannt, der durchaus mit Schulnoten assoziiert sein darf, aber unter meritokratischen Gesichtspunkten nach Möglichkeit – nach Kontrolle anderer Faktoren – keinen eigenständigen Beitrag zur Erklärung von Leistungsbewertungen haben sollte.

Die Abbildung 1 wird im Kapitel 3 dabei helfen, die jeweiligen Forschungsfragen genauer zu spezifizieren.

2.5. Exkurs: Effekte gesellschaftlichen Wandels und wissenschaftlichen Fortschritts auf die Leistungsdiagnostik am Beispiel Schweiz

Welche Konsequenzen hatten wissenschaftliche Erkenntnisse zur Qualität von Noten und Übertrittsentscheidungen sowie gesellschaftliche Veränderungen, die sich u.a. in veränderten Anforderungen an diagnostische Prozesse niederschlugen, für die Notengebung und Übertrittsentscheidungen? Im Folgenden soll dies beispielhaft für die Schweiz dargestellt werden.

2.5.1. EIN BLICK IN DIE NAHE VERGANGENHEIT

Lakonisch könnte festgestellt werden, dass die größte Veränderung bezüglich der schulischen Beurteilung während den ersten sieben Jahrzehnten des vergangenen Jahrhunderts in der Umkodierung der Notenwerte bestand: Seit den fünfziger Jahren bedeutete nicht mehr die Note

1 „sehr gut“, sondern die Note 6. Die Note 4 bedeutet auf der 6er Skala noch „genügend“ und die 3 bedeutet „ungenügend“. Erst im letzten Drittel des 20. Jahrhunderts, angeregt durch die Praxis in den Rudolf Steiner Schulen und durch die Forschungsergebnisse aus Deutschland (Fend, 1969; Ingenkamp, 1971; Schelsky, 1957; Tent, 1969), formierten sich private Bewegungen zur Abschaffung der Noten, so etwa die einflussreiche Vereinigung von Lehrpersonen und Eltern „Schule ohne Noten“.

Im Folgenden ist die Frage, welche Grundlagen zur schulischen Beurteilung an den Lehrerseminaren vermittelt wurden von Interesse. Dazu sollen überblicksartig drei wichtige Vertreter berücksichtigt werden. Emil Achermann verfasste ein Standardwerk zur Schulmethodik, das von Hans Aebli's Schriften abgelöst wurde. Dazu hat August Flammer Ende der sechziger und anfangs der siebziger Jahre als erster in der Schweiz eine empirische Forschung zur Notengebung begründet.

2.5.2. DIE BEITRÄGE VON EMIL ACHERMANN

Emil Achermann, der 1950 eine in der Schweiz weit beachtete „Methodik des Volksschulunterrichts“ veröffentlichte, widmete in seiner dritten Auflage (Achermann, 1966) dem Zeugnis drei von insgesamt 366 Seiten. Die Beurteilung der Schülerleistung oder das Prüfen wird weder im allgemeinen Teil der Methodik, noch in den fachspezifischen erwähnt. Daraus ist zu schließen, dass die Lehrperson einzig mit dem Quartalszeugnis dem Schüler seine Leistungseinschätzung offenbart. Er ordnet die Bedeutung des Zeugnisses auch als ein für das Kind bedeutsames amtliches Dokument ein:

Zeugnisse sind Spiegelbild für den Schüler, Brief der Schule an die Eltern, amtliche Urkunde, mitunter Schlüssel für das Leben. Es gibt Zwischenzeugnisse und Jahreszeugnisse; Zwischenzeugnisse sind mehr vertraulicher Art, Jahreszeugnisse sind amtliche Urkunden. (Achermann, 1966, S. 18)

Achermann bezeichnet die Ziffernote eher als ein Instrument, das mit Vorsicht aufzunehmen ist und geradezu gefährlich sein kann:

Wenn du Ziffern anwendest, denke immer an ihren Sinn! – Man unterscheidet Leistungs-, Fleiß- und Betragennoten. Was sagst du zu den zwei letzten? Denke vor der Beantwortung dieser Frage an die Schläuheit der Schüler, an die Unzulänglichkeit des Lehrerurteils; Eintragungen

ins Strafregister erlöschen nach einer bestimmten Zeit der Bewährung; Zeugnisse bleiben fürs Leben. (Achermann, 1966, S. 18)

Mit der Unzulänglichkeit des Lehrerurteils spricht er die Ergebnisse der späteren Untersuchungen dazu an. Das Zeugnis als amtliches Dokument kann stigmatisierender sein als ein Eintrag ins Strafregister. Er anerkennt die Qualifikations- und Allokationsnorm des Zeugnisses, wenn er von der „Steignorm“ spricht. Er warnte gleichzeitig davor:

Ich habe von Berufes wegen jährlich dreimal Zeugnisse auszustellen. Und am liebsten möchte ich jedem Kinde auf sein „Lebensbüchlein“ schreiben „Vorsicht bei Gebrauch!“ Warum? Weil die Zensuren von einem Menschen stammen, der selber voller Fehler und Schwächen ist und weil die die anderen Menschen nur so sehen kann, wie er sie sieht, darum. Uns Schulmeistern wird die Sache ja einfach gemacht. Statt vieler Worte gibt es da Zahlen. Zahlen sind eindeutig, fünf ist nicht sechs, und drei ist „ungenügend“, daran gibt es nichts zu deuteln. Sechs Zahlen sind es, sechs Tablarern bei der Apfelhürde gleich. Und nun zu bestimmen, auf welches Tablar jeder Schüler gehört. ... Und hat man sie nach bestem Wissen und Gewissen eingeordnet, dann kann es passieren, dass man es erst noch falsch gemacht hat. Wenn man später wieder Nachschau hält, dann sind vielleicht gerade bei den edlen Sorten, von denen man sich so viel versprach, etliche faul, indessen sich die minderen Sorten vielleicht ganz wacker halten. ... Man muss ganz fleißig Nachschau halten und dennoch trifft man's nicht in jedem Falle ganz. Das Sortieren ist eine schwierige Arbeit, und bedrückend ist es geradezu, wenn man fürchten muss, dass der Entscheid Schicksal spielen werde. (Achermann, 1966, S. 20 f.)

Wenn auch seine Ausdrucksweise, mit dem Vergleich des Sortierens von „edleren“ und „minderen“ Äpfeln zu anschaulich und wertend ausgefallen ist, macht er in seiner Methodenlehre auf die mangelnde prognostische Validität der Zeugnisnoten, ihre Ungenauigkeit in Form von scheinbar genauen Ziffernoten und auf schicksalshafte Selektionsentscheide aufmerksam. Als Pädagoge legt er mehr Wert auf das direkte Gespräch als auf Zeugnisnoten:

Es gibt Leute, denen sind Zeugnisse so etwas wie Göttersprüche, sie treiben einen wahren Kult um sie. ... Es gibt auch Menschen, die stehen über dem Alltag und somit

über dem Zeugnis. Oft sind es gerade die, die den Alltag am gründlichsten kennen, die schon Großes geleistet haben. Es hat etwas befreiendes, wenn ein Prinzipal die Zeugnisse auf die Seite schiebt und Aug in Aug mit dem Bewerber zu sprechen beginnt. ... Und ebenso ist ein Zeugnis kein endgültiges Urteil über einen Menschen, das ihm nun sein Leben lang wie ein Schatten auf den Fersen folgt. Wäre es das, dann sollte man die Zeugnisse mit sofortiger Wirkung verbieten. (Achermann, 1966, S. 20)

2.5.3. DIE KRITIK AN DER LEISTUNGSDIAGNOSTIK VON AEBLI

Mehr als zwanzig Jahre später wird in einer Reihe von Publikationen (Tent, 1969; Ingenkamp, 1971; Biermann, 1976; Ziegenspeck, 1978; Fend, 1974), aufgrund empirischer Untersuchungen der Schulnoten, ihre vollständige Unzulänglichkeit aufgezeigt. Mit der Entwicklung der klassischen Testtheorie (vgl. Amelang & Schmidt-Atzert, 2006) entstanden Gütekriterien, die bei der Erfassung menschlicher Eigenschaften wie Persönlichkeitsmerkmale und Leistungsfähigkeit beachtet und minimal erfüllt werden müssen. So wurde schon früh ersichtlich, dass selbst die Kriterien einer objektiven Leistungsauswertung nicht erreicht werden. Dazu sind Wahrnehmungsfehler und Einstellungen der Lehrperson verantwortlich. Dieselben Leistungen werden zudem von verschiedenen Lehrpersonen unterschiedlich benotet. So kann dieselbe Leistung, ausgedrückt in Punktzahlen ungenügend oder gut sein. Die Anforderungen der Reliabilität, der Zuverlässigkeit oder der Inhaltsvalidität, der inhaltlichen Gültigkeit konnten mit Schulprüfungen und der entsprechenden Benotung nicht mal annäherungsweise erfüllt werden. Während dieser Zeitspanne der siebziger Jahre entsteht in der Schweiz eine rege Tätigkeit der Lehrplanentwicklung. Lehrpläne werden nach den Anforderungen operationalisierter Lernziele/Lehrziele aufgebaut. Dadurch soll die Lehrperson auch eine verlässlichere Grundlage für die Leistungserfassung erhalten. Das Erreichen operationalisierter Lernziele kann mit größerer Genauigkeit beurteilt werden. Diese, vom Behaviorismus geprägte Bewegung, peilt einerseits wissenschaftlich entwickelte Leistungstests als Messinstrumente an und andererseits entwickelt sich die Forderung nach einer der Normalverteilung entsprechenden Beurteilung und Benotung von Schülerleistungen.

In den siebziger Jahren des letzten Jahrhunderts entwickelte Hans Aebli eine Allgemeine Didaktik (Aebli, 1973) auf der Grundlage der kognitiven Psychologie. Damit hob er sich von der behavioristisch ausgerichteten

Lehrplanentwicklung ab und fokussierte auf die Lernprozesse. Implizit wurde dadurch angenommen, dass diese, von Piagets Epistemologie und Entwicklungspsychologie geprägte Didaktik zu qualitativ besserem Lernen führt. Daher widmete sich Aebli vor allem der Didaktik zur Förderung der Lernprozesse. Im letzten Kapitel wandte er sich denn auch der Beurteilung und dem Prüfen von Schülerleistungen zu. Eine kognitionspsychologisch fundierte Didaktik propagierend, wertet er die pädagogische Bedeutung des Prüfens sehr differenziert.

Prüfungen gehören zu den umstrittensten pädagogischen Maßnahmen. Dass dem so ist, versteht man leicht, wenn man sich in unseren Schulen umsieht. Es ist wahr, dass Prüfungen, z.T. aus Unkenntnis, aber auch aus tieferreichenden Gründen, vielerorts in einer Weise angewendet werden, die nicht nur ihr Ziel verfehlt, sondern den übri-gen erzieherischen und didaktischen Absichten geradezu entgegenwirken. ... Ja, der Kampf gegen die Prüfungen wird z.T. so unsachlich geführt, dass man fragen muss, ob es nicht auch eine legitime Funktion der Schulprüfung gebe und ob sich daraus eine Praxis ableiten lasse, welche dem Lernen und der Entwicklung der jungen Menschen diene, statt ihnen zu schaden. Wir meinen, dass dies möglich sei. (Aebli, 197811, S. 325)

Als Antwort darauf schlägt er vor, Prüfungsaufgaben auf die Lernziele und den im Unterricht behandelten Stoff abzustützen. Diese Schulprüfungen hebt Aebli von den Schulleistungstests ab:

Nach dem Vorausgehenden kann kein Zweifel bestehen, dass im Handel erhältliche Tests die vom Lehrer konstruierte Prüfung keinesfalls ersetzen können, sind diese doch nie auf den vorangehenden Lernprozess bezogen. ... Sie (die Schulleistungstests; Anmerkung der Autoren) beziehen sich, wie gesagt, auf die grundlegenden Stoffe einer gegebenen Schulstufe, und sie haben vor allem die Funktion, dem Lehrer zu sagen, wie das Leistungsniveau seiner Schüler sich zum durchschnittlichen Leistungsniveau der Schüler dieser Stufe verhält. (Aebli, 197811, S. 334)

Für die Aufgabenkonstruktion fordert er in erster Linie, die Inhaltsvalidität zu beachten und vielseitige Aufgabenformen, von unterschiedlichen Anspruchsniveaus eng an die behandelten Lernziele ausgerichtet zu formulieren.

Beim Benoten und Bewerten entscheidet sich nun, wie Aebli die Forderung nach diesem inhaltvaliden Prüfen löst, und auf welche Bezugsnorm er sich abstützt. Dazu hält er fest:

Schulnoten darf man nicht gemäß einer im Voraus festgelegten Skala, die eine lineare Abhängigkeit von Fehlerzahl und Noten annimmt, festsetzen. Wir brauchen ein gedankliches Modell, das sichert, dass die Durchschnittsnoten in einer größeren Zahl von Prüflingen von Jahr zu Jahr gleich bleiben und welches auch den extremen Leistungen eine klare Bedeutung gibt. Das ist nur möglich, wenn man Noten auf eine wohlumschriebene Bezugsgruppe bezieht. ... Die Altersgruppe soll die einzige Bezugsgruppe sein. (Aebli, 197811, S. 363; 371)

Auch in der letzten Ausgabe, den „Grundlagen des Lehrens“ (Aebli, 1987) baut er das Benotungsmodell auf der gaußschen Normalverteilung auf. Die Punkteverteilung soll an einer repräsentativen Stichprobe von Klassen erhoben werden. Da erhalten 5 Prozent die beste Note 6, 10 Prozent eine 5.5, 20 Prozent eine 5 (gut), 30 Prozent eine 4.5, 20 Prozent eine 4 (genügend), 10 Prozent eine 3.5 (ungenügend) und 5 Prozent eine 3 (ungenügend). Die Leistungsverteilung der einzelnen Klasse wird mit dieser Stichprobenverteilung verglichen und daran gemessen. So können in einer leistungsstarken Klasse mehr als 5% eine sehr gute und allenfalls niemand eine ungenügende Note erhalten. Bei einer leistungsschwachen Klasse würden allenfalls mehr als 15% ungenügende Noten vergeben. So löste Aebli das Dilemma einer normalverteilungsorientierten Benotung.

Aebli's Anliegen zur Beurteilung schulischer Leistungen ist mannigfaltig: Er möchte dem Notenwert eine mehr oder weniger normalverteilungsorientierte Grundlage zuordnen, so dass eine Note einen verlässlichen Informationswert erhält. Danach wäre die häufigste Note eine 4.5 (30%). Er ringt damit um eine gerechte Benotung im Sinn einer vergleichbaren, nicht um eine sozial gerechte Beurteilung und Benotung. Indem er von einer lernzielnahen Leistungserfassung ausgeht, postuliert er implizit eine sozial gerechte Beurteilung. Auf die in dieser Zeit stark soziologisch geprägte Fragestellung nach der Schule als Reproduktionsinstanz (Fend, 1974) der vorherrschenden sozialen Schichtung und ihre Chancenverteilung geht Aebli nicht ein.

2.5.4. FLAMMER UND DIE DIFFERENZIERUNG DER FUNKTIONEN VON NOTEN

Ähnlich wie Aebli ist Flammer um die möglichst objektive Aussagekraft der Ziffernote als Rangreihe besorgt (Flammer, Perrig-Chiello & Rüeegg, 1983). Flammer kümmert sich um eine Funktionsentflechtung der Note:

Das ist unsere zentrale Kritik an der gegenwärtigen Schulnotenpraxis, nämlich dass die meisten von uns versuchen, der Schulnote eine solche Menge von verschiedenen Funktionen zuzuordnen, dass sie keine Funktion mehr richtig erfüllen kann. Sie muss durch diese Ansprüche hoffnungslos verzerrt werden; niemand mehr kann sie ernst nehmen, und doch nehmen sie (fast) alle sehr ernst, zu ernst, weil sie durch die vielen zugeordneten Funktionen so wichtig geworden ist. Es ergibt sich ein verhängnisvolles Paradoxon: Je mehr Funktionen der Schulnote zugeordnet werden, desto bedeutsamer wird sie – aber auch desto wertloser! (Flammer, Perrig-Chiello & Rüeegg, 1983, 15).

Unter diesen multiplen Gegenständen der Schulnote nennen sie:

- Notendurchschnitt während einer Schulperiode
- den Leistungsstand am Ende eines Schuljahres
- den Leistungsfortschritt
- die geschätzte Leistungsfähigkeit oder/und
- Fleiß, Anstrengung.

Tatsächlich stellte Flammer (1971) in einer Befragung von Lehrpersonen der Primarstufe all diese Zuschreibungen fest. Dazu kamen noch die Funktionen der Ermutigung und der Prognose, der im nächsten Jahr voraussichtlich erreichten Leistung. Flammer, Perrig-Chiello & Rüeegg (1983) stellen nicht nur an die Schulprüfungen, sondern auch an die Noten die Anforderung, dass sie den Gütekriterien der klassischen Testtheorie genügen. Demnach sollen Noten objektiv, reliabel und valide sein. Ihre Literaturanalyse zeigt denn auch, dass Schulnoten diesen Ansprüchen nie entsprechen konnten. So schlagen sie vor: „Eine Verbesserung der Reliabilität wäre demnach zu erreichen, indem ein größerer Konsens zwischen den Lehrern erzielt würde in Bezug auf das, was zu benoten ist und wie die Skala zu handhaben ist (Flammer, Perrig-Chiello & Rüeegg, 1983, S. 21). Zur Validität kommen sie zur Erkenntnis und Forderung, „... dass die Noten eine hohe kurzfristige Validität haben sollten, dass an sie aber

keine Forderung nach langfristiger Validität gestellt werden sollte.“ (S. 23) Dazu kommen sie zur Einsicht:

Natürlich ist die ungenügende Validität der Schulnote nicht nur eine logisch-technische Konsequenz aus der ungenügenden Definition, sondern auch eine Folge von „Fehlern“ der Wahrnehmung und Beurteilung der Schüler und ihres Verhaltens. Wahrnehmen und Beurteilen sind komplexe psychologische Vorgänge, die im Zusammenhang stehen mit der Persönlichkeitsstruktur des Beurteilers und seinen Erfahrungen, aber auch mit seiner sozialen Situation und seiner Beziehung zum Beurteilten. (Flammer, Perrig-Chiello & Rüeegg, 1983, 23).

Im Folgenden entwickeln sie Thesen im Sinne von neuen Forderungen an die Funktion der Noten: Als einzige Funktion sehen sie „die Dokumentation der summativen Evaluation zuhanden des Promotionsentscheides“ (S. 27). So sollen nach ihnen Noten „nur das Leistungsniveau, das Schüler am Ende der betreffenden Unterrichtsperiode erreicht hat“ (S. 29) dokumentieren und Schulzeugnisse seien nur am Ende des Schuljahres, im Rahmen des Promotionsentscheides (S.31) zu rechtfertigen. Nach dieser Forderung zur massiven Funktionseinschränkung folgt jene nach einem „gemeinsamen Code“ (S. 36) um das Bewertungsergebnis mitzuteilen. Rein technisch betrachtet ist der Code die Ziffernote, aber die Bezugsnorm, worauf sich die Ziffer bezieht, ist unklar. So befragte Flammer (1971) Gymnasial- und Primarlehrer nach der Note für einen durchschnittlichen Schüler in einer durchschnittlichen Klasse. Dazu erhielt Flammer Angaben zwischen dem Notendurchschnitt 3,9 und 5,0. Daraus folgerten sie „Offensichtlich ist eine präzise Definition der Notenskala nötig und die Bereitschaft aller Beteiligten, sich daran zu halten.“ (Flammer, Perrig-Chiello, Rüeegg & 1983, 39). Ähnlich wie Aebli eine zuverlässige Notenverteilung forderte, erhoben Flammer, Perrig-Chiello und Rüeegg (1983) in allen Kantonen der Schweiz die Noten der beiden Jahre vor dem Übertritt in die Sekundarschule. So konnten sie für jeden Kanton eine empirische Notenverteilung mit dem Median und den Quartilabständen erarbeiten, die den Lehrpersonen als Referenznorm für ihre Notengebung dienen sollte – sie schufen eine kantonale Referenznorm für die Notenverteilung. Aus dem Gesamt entwickelten sie eine „Schweizerische Referenznorm für Schulzeugnisnoten“:

(→ TABELLE 1)

Tabelle 1: Referenznorm für Schulzeugnisnoten

Notenhäufigkeiten der Fächer Rechnen und Sprache schriftlich in den beiden letzten Klassen vor dem Übertritt in die Sekundarstufe (alle Kantone)

Note	Anteil in %	Summierte %
6,0	3,29	3,29
5,5	15,45	18,74
5,0	25,49	44,23
4,5	23,94	68,17
4,0	17,69	85,86
3,5	8,95	94,81
3,0	3,97	98,78
2,5	0,92	99,70
2,0	0,27	99,97
1,5	0,03	100,00
1,0	0,00	100,00

Quelle: Flammer, Perrig-Chiello & Rüeegg (1983, S. 165)

Verglichen mit Aebli's theoretischer Notenverteilung, weist Flammers empirische Verteilung weniger 6er (sehr gut), dafür 5 Prozent mehr 5,5er und 5 Prozent mehr 5er (gut) auf. Der Median liegt sowohl bei Aebli wie bei Flammer in der Note 4.5, obwohl die empirische Verteilung 6 Prozent weniger 4,5er und 2 Prozent weniger 4er aufweist. Aebli's Annahme von insgesamt 15% ungenügende Noten kommt der empirischen Feststellung von 14 sehr nahe. Aebli setzt also eine leicht „strengere“ Notenverteilung an, indem er von einer angepassteren „Normalverteilung“ ausgeht als Flammer beobachten konnte.

Gemeinsames Merkmal von Aebli's und Flammers Postulaten ist, dass beide eine soziale Bezugsnorm der Notenverteilung zu Grunde legen. Noten haben einen Aussagewert über den Prozentrang eines Schülers, einer Schülerin innerhalb eines Faches. Flammer kümmert sich sehr stark um eine Funktionsentflechtung der Noten und fordert eine Trennung von Aspekten der Motivation und Förderung (formative Beurteilung) von einer strikten Lernstandsbestätigung am Ende des Schuljahres. So fordert Flammer auch, dass die Zeugnisnote am Ende des Schuljahres nicht auf einem genauen Durchschnitt der Prüfungsleistungen während des Jahres beruht, sondern das ungefähre Leistungsniveau des Schülers bzw. der Schülerin ausdrückt (Flammer, Perrig-Chiello & Rüeegg, 1983). Damit verlegt er das Augenmerk von der technischen Messgenauigkeit auf die Lehrperson als „Messinstrument“, die Schülerleistungen dank ihrer Professionalität zuverlässig einschätzen

kann. Die Frage der sozialen Gerechtigkeit wird auch von Flammer nicht explizit angegangen.

2.5.5. SIPRI – EIN PROJEKT, DAS VIELES IN BEWEGUNG SETZTE

1978 startete die EDK (Schweizerische Konferenz der kantonalen Erziehungsdirektoren) das Projekt SIPRI (Überprüfung der Situation der Primarschule; Heller, 1986) mit dem Teilprojekt „Funktionen und Formen der Schülerbeurteilung“. Das Projekt kann insofern als „Bewegung“ bezeichnet werden, weil das Rahmenthema „Schülerbeurteilung“ von Steuergruppen unterstützt mit freiwilligen Schulteams in „bottom up“ Prozessen entwickelt wurde. Hauptanliegen war die grundlegende Entflechtung der formativen (fördernden, beratenden) von der summativen (Lernstands-) Beurteilung. Vor allem galt der Grundsatz, die formative Beurteilung nicht zu benoten und Noten höchstens den als summativ geplanten Lernstandserhebungen zuzuordnen. Als zweites Hauptanliegen galt es, eine lernzielbezogene Schülerbeurteilung aufzubauen. Damit wurde die Diskussion um die Messgenauigkeit innerhalb der sozialen Bezugsnorm, wie sie Aebli und Flammer führten, obsolet. Im Interesse der Schulentwicklung stand die Sach- oder Kriteriumsnorm oder das Interesse am Erreichen von Lernzielen und an deren Inhalt. Damit wurde auch Rheinbergs Forschung zur Wirkung der sozialnormorientierten Beurteilung versus der individualnormorientierten durch die Inhalts- oder Lernzielorientierung erweitert (vgl. dazu Mischo & Rheinberg, 1995; Rheinberg, 2001). Zentral wurden Fragen, wie man das Erfüllen eines Lernzieles überprüfen kann und wann eine gezeigte Leistung als „erfüllt“ oder „genügend“ beurteilt werden kann. Wird diese Frage gestellt, erübrigt sich der Disput um eine normalverteilte Notengebung. Individueller Erfolg besteht hier in der Bewältigung von Aufgaben, die lernzielorientiert konstruiert worden sind. Schulerfolg ist ablesbar im Prozentsatz von Schülerinnen und Schülern einer Klasse, welche die verschiedenen Anforderungsstufen erfolgreich bewältigt haben. Noten müssen unter diesem Aspekt eher als unpassendes Kommunikationsmittel betrachtet werden, weil in dieser Ziffer diese inhaltliche Aussage nicht kodiert werden kann. Heute wissen wir, dass die Leistungsbeurteilung von PISA dieser Logik folgt.

Mit dem Fokussieren weg von der Sozialnorm und der Genauigkeit des Messens und Benotens im Schulalltag hin auf die Sach- oder Kriteriumsnorm wurde die Frage nach der Leistung, nach dem Können und dem „Was“ des Kön-

nens wichtig. Damit verbunden wurde die in den 70er Jahren soziologisch geführte Diskussion um die soziale Gerechtigkeit unter pädagogischen Aspekten. Obwohl auch das SIPRI-Projekt soziale Gerechtigkeit im Schulsystem implizierte, indem innerhalb der Schülerbeurteilung der Förderaspekt stark gemacht wurde, brachten erst entsprechende PISA-Studien diesen Aspekt wiederum hervor.

Der 1999 erschienene Trendbericht „Mehr fördern, weniger auslesen“ (Vögeli-Mantovani, 1999) der Schweizerischen Koordinationsstelle für Bildungsforschung berichtet eine Bestandsaufnahme zur schulischen Beurteilung in der Schweiz und zu den Übergangsverfahren von der Primar- in die Sekundarschule. Dieser Bericht ist keine Evaluationsstudie der SIPRI-Bewegung, darf aber als Bestandsaufnahme nach 15 Jahren Dissemination des SIPRI-Gedankenguts gewertet werden. So stellte Vögeli-Mantovani (1999) fest, dass zum Zeitpunkt der Bestandsaufnahme 1998/99 mit wenigen Ausnahmen die Ziffernoten für die ersten zwei bis drei Jahren der Primarschule durch institutionalisierte Elterngespräche und Angaben zur Erfüllung der Lernziele im Zeugnis ersetzt worden sind. Mittels direkter Kontakte zu den Eltern sollte die Qualität des Lernprozesses und die Kompetenzerweiterung über die curricular festgelegten Bereiche hinaus gewürdigt werden. Unter anderem ging es dabei auch um die Förderung der Chancengerechtigkeit:

Selbst wenn die formale Chancengleichheit gewährleistet wird, wirkt sich die Barriere des sozialen Milieus auf die Chancenausnutzung aus. Geringe Bildungsambitionen verhindern die Ausschöpfung des Fähigkeitspotentials und der Chancen im Wettbewerb mit anderen um Leistungen ... In der Schule erfährt die Leistungsorientierung eine Verengung, wenn das Handlungs- bzw. Leistungsergebnis mehr gewichtet wird als der Weg des Lernens und Leistens. Zudem werden kognitive Leistungen höher bewertet als andere (Vögeli-Mantovani, 1999, S. 72).

Von den direkten Beurteilungsgesprächen mit den Eltern versprach man sich u.a. auch einen Abbau sozialer Ungerechtigkeiten, indem Eltern aus bildungsfernerem Milieu die Bedeutung der Schule und die schulische Leistung ihres Kindes verständlicher gemacht werden können. Indem die Beurteilung auf die Erreichung der Lernziele und die individuellen Fortschritte ausgerichtet wurde (Vögeli-Mantovani, 1999, S. 73), können Eltern die gemachten Fortschritte und die Schwierigkeiten ihres Kindes an konkreten Lerninhalten dargestellt werden. Besonders die Darstellung der Fortschritte (Individualnorm-

Orientierung) sollte auch das Selbstwertgefühl der Eltern stärken und sie zur Unterstützung ihres Kindes bei der Schularbeit motivieren.

In fast allen Kantonen wurden die Ziffernoten in den ersten zwei bis drei Primarschuljahren abgeschafft und mit Wortetiketten zur Erfüllung der Lernziele ersetzt. Die Fokussierung der Beurteilung auf die Sach- oder Kriteriumsnorm ist hier unverkennbar. Mit der Überwindung der Sozialnorm versprach man sich eine Verminderung sozialer Ungerechtigkeiten.

Ebenfalls mit Ziel einer besseren Chancengerechtigkeit wurde versucht, die jährlichen Promotionsbedingungen vom Erreichen eines Notendurchschnitts abzukoppeln. So wechselte beispielsweise der deutschsprachige Kantonsteil Freiburg von der Reglementierung der Promotionsbestimmung zur Reglementierung der Nicht-Promotion: In der Regel sind alle Schülerinnen und Schüler für die nächste Klasse promoviert. Bei einer systematischen Über- oder Unterforderung des Kindes muss die Lehrperson einen begründeten Antrag auf pädagogische Fördermaßnahmen stellen. Eine Klassenrepetition wird dann allenfalls im Sinne einer pädagogischen Fördermaßnahme angeordnet. In den meisten Kantonen wurden die jährlichen Promotionsordnungen in ähnlichem Sinne verändert, um die Repetentenquote zu senken (Vögeli-Mantovani, 1999).

Als dritten Schwerpunkt erhob Vögeli-Mantovani (1999) das Übertrittsverfahren von der Primar- in die Sekundarschule in den Kantonen. Die SIPRI-Bewegung postulierte hierzu das Ersetzen von Aufnahmeprüfungen durch vergleichende Orientierungsarbeiten, welche die Primarlehrperson über den Leistungsstand ihrer Klasse im größeren Gruppenvergleich informieren. Der selektiven Aufnahmeprüfung wurde ein großes Potential sozialer Ungerechtigkeit zugeschrieben. Anstelle der Leistungsprüfung sollte das globale Lehrerurteil in Form einer Zuweisungsempfehlung treten. Zudem sollte der Elternwunsch für den Übertritt des Kindes gleichwertig zur Lehrerempfehlung berücksichtigt werden. Vögeli-Mantovani (1999) zeigt auf, dass sämtliche Kantone ihr Übertrittsverfahren in die Sekundarschule verändert haben und dass sich alle ein multikriteriales Verfahren geschaffen haben. Offizielle Entscheidungsgrundlagen dazu sind: der Leistungsstand des Kindes, bezogen auf die Lernziele, allenfalls Noten; Lehrerempfehlung; Eltern- und Schülerwunsch; Übertrittsgespräche mit den Eltern und vereinzelt eine Prüfung zur definitiven Entscheidung. In den meisten Kantonen entscheidet der „Instanzenweg“ bei Uneinigkeit.

Seit Erscheinen des Trendberichts von Vögeli-Mantovani (1999) wurde die Beurteilungspraxis in der Schweiz nicht mehr systematisch beschrieben. Vor allem die Praxis der Notengebung wurde wissenschaftlich nicht weiter bearbeitet. Ein informelles Netzwerk „Beurteilung“, das von der Koordinationsstelle für Bildungsforschung, namentlich von Vögeli-Mantovani verwaltet wird, bildet eine Plattform für Informationsaustausch zu kantonalen Entwicklungen in Bezug auf die Beurteilungs- und Benotungspraxis. Der Informationsaustausch dazu zeigt, dass sich folgendes Gedankengut zur Beurteilung von Schülerleistungen in sehr vielen Kantonen verankert hat:

- Beurteilungen auf der Primarschule werden sach-, bzw. kriteriumsnormorientiert vergeben. Unabhängig davon wie die Beurteilungen und Bewertungen mitgeteilt werden – vorwiegend mit lernzielorientierten Wortetiketten – oder auch mit Noten, soll die Genügend-Limite auf das Erreichen der Lernziele bezogen werden. Die Gruppen- oder Sozialnorm wird weniger in der Sekundarschule, vermehrt noch am Gymnasium als statistische Verteilungsform herangezogen.
- Aufnahmeprüfungen sind ausschließlich noch beim Übertritt von der Primarschule ins Langzeitgymnasium auszumachen.
- Beim Übertritt in die meistens dreigliedrige Sekundarschule wird keine oder vereinzelt eine Schulleistungsprüfung als „Vergleichsprüfung“ ohne Selektionskraft durchgeführt. Inwiefern dadurch jedoch mehr soziale Chancengerechtigkeit entstanden ist, ist umstritten. Baeriswyl et al. (2006) konnten zeigen, dass die globale Lehrerempfehlung und die Elternempfehlung stärker vom sozioökonomischen Status der Eltern geprägt ist als das Ergebnis des standardisierten Leistungstests.
- Beachtenswert ist die Tatsache, dass sämtliche kantonalen Bildungsgesetze für die Volksschule, die während der letzten dreißig Jahre erlassen wurden, im Zweckartikel ausschließlich die Förderung des Kindes vorgeben. Es findet sich kein Selektionsauftrag in einem Gesetz der obligatorischen Volksschule.

Zusammenfassend kann festgehalten werden, dass die unter dem Eindruck der Psychometrie entstandenen Forderungen nach Messgenauigkeit von Schulleistungen, die

sich an die einzelne Lehrperson richtete, verklungen sind. Die Wende zu einer pädagogischen Beurteilung, die sich in erster Linie am Erreichen der Lernziele ausrichtet, hat sich durchgesetzt. Die Teilnahme an den PISA-Studien und deren Ergebnisse bieten Grundlagen für Schweiz weite Steuerungsinstrumente. Zudem haben sich Institute gebildet, die kantonsübergreifend das leistungsvergleichende Testen von Schulklassen anbieten. Die Frage, inwieweit die unterschiedlichen Bewertungssysteme soziale Disparitäten steigern oder vermindern, hat bei der Weiterentwicklung der Bewertungspraxis in der Schweiz bis heute eine eher nachrangige Rolle gespielt.

2.6. Empirische Befunde aus jüngeren Schulleistungsstudien

2.6.1. BEFUNDE AUS SCHULLEISTUNGSFORSCHUNG

Leistung und Leistungsbewertung sind zentrale Themen der großen nationalen und internationalen Schulleistungsstudien der letzten Dekade. Besonders für den Primar und Sekundarschulbereich I liegen mittlerweile auch differenzierte Analysen von Schulnoten sowie zum Zusammenhang von sozialer Herkunft und schulischen Bewertungen vor (Arnold, Bos, Richert & Stubbe, 2007; Baumert et al., 2003a; Bos et. al, 2004; Hochweber, 2010; Klieme, 2003; Lehmann, Peek, Gänsfuß, Lukat, Mücke & Barth, 2000; Lehmann, Peek & Gänsfuß, 1997; Schöps, Walter, Zimmer & Prenzel, 2006; Stubbe & Bos, 2008).

Für die Primarstufe haben Stubbe und Bos (2008, vgl. auch Bos et. al, 2004) auf der Datengrundlage der IGLU 2001-Studie (vgl. Bos et al., 2003) im Rahmen einer Modellierung des Grundschulübergangs die Notenvergabe in den Fächern Deutsch und Mathematik untersucht. Neben den erzielten Testleistungen wurden dabei auch die kognitive Grundfähigkeiten, das Geschlecht und verschiedene Merkmale der familiären Herkunft mit einbezogen. Zudem gingen motivationale Schülermerkmale (die durch die Eltern eingeschätzte Anstrengungsbereitschaft und Prüfungsängstlichkeit) in die Analysen ein. Die stärksten Effekte auf die Notenvergabe hatten die Testleistungen der Schülerinnen und Schüler. Aber auch für die kognitiven Grundfähigkeiten, das Geschlecht sowie die motivationalen Schülermerkmale (hier insbesondere die Anstrengungsbereitschaft) konnten Effekte festgestellt werden. Interessant war weiterhin, dass auch nach Berücksichtigung dieser Merkmale Einflüsse der familiären Herkunft

auf die Notenvergabe nachweisbar waren, deren Ausmaß in Relation zu den Effekten der Testleistung jedoch vergleichsweise gering ausfiel. Nichtsdestotrotz liefert die IGLU-Studie damit Hinweise auf leistungsfremde Einflussfaktoren des familiären Hintergrundes bei der Leistungsbewertung. Diese Befunde konnten mit den Daten der IGLU-2006-Studie repliziert werden (Arnold, Bos, Richert & Stubbe, 2007)

Auch für den Sekundarschulbereich I konnten substantielle Korrelationen zwischen Fachnoten und den Ergebnissen aus den standardisierten Leistungstests festgestellt werden (Baumert et al., 2003a; Hochweber, 2010; Klieme, 2003; Schöps, Walter, Zimmer & Prenzel, 2006). Dieser Befund entspricht auch dem Alltagsverständnis, demzufolge Kinder mit guten Ergebnissen in den Leistungstests auch gute Schulnoten aufweisen sollten. Des Weiteren weisen die Studien darauf hin, dass für die Benotung auch bildungsgangspezifische Referenznormen zum Tragen kommen. Das bedeutet, dass die Notenvergabe in jedem Bildungsgang innerhalb eines eigenen Referenzrahmen erfolgt, der auf unterschiedlichen Leistungsniveaus beruht (vgl. Baumert et al., 2003a). Die Note 3 an der Hauptschule ist also nicht mit einer Note 3 an der Realschule oder am Gymnasium vergleichbar. Gleichwohl differenzieren die Noten innerhalb einer Schulform nach der Testleistung. Gleichzeitig konnte in verschiedenen Studien gezeigt werden, dass sich innerhalb einer Schulform die mit den Noten verbundenen Leistungsverteilungen zum Teil deutlich überlappen. So kann man mit der gleichen Leistung an einer Schule eine zwei, an einer anderen Schule eine fünf erreichen.

Eine Reihe von Studien konnte überdies klare Hinweise auf differierende Bewertungsmaßstäbe zwischen den Bundesländern erbringen. Entsprechende Befunde finden sich sowohl für die Grundschule (vgl. z.B. Bos et al., 2004), die Mittelstufe (vgl. z.B. Baumert et al., 2003a) und die gymnasiale Oberstufe (vgl. Neumann, Nagy, Trautwein & Lüdtke, 2009).

Vergleichsweise wenige Studien haben sich mit dem Einfluss leistungsfremder Indikatoren auf Schulnoten auseinandergesetzt. Die wenigen bislang vorliegenden Befunde deuten auf einen sozialen Herkunftseffekt dergestalt hin, dass Kinder aus sozial weniger begünstigten Familien strenger bewertet werden als Kinder aus sozial begünstigten Familien (Lehmann et al. 1997, 2000; Hochweber, 2010). Darüber hinaus konnte in verschiedenen Studien einen Geschlechtereffekt festgestellt werden, wonach

Mädchen bei gleichen Leistungen in standardisierten Tests vielfach bessere Noten als Jungen bekommen (Hochweber, 2010; Schöps, Walter, Zimmer & Prenzel, 2006). Problematisch an den entsprechenden Studien ist oftmals der Verzicht auf den Einbezug von Indikatoren der schulischen Mitarbeit, die einen Teil der entsprechenden Disparitäten erklären könnten (vgl. Hannover & Kessels, 2011).

2.6.2. REFERENZGRUPPENEFFEKTE SOWIE LEHRER- UND SCHULEFFEKTE

Neben den angeführten sozialen Disparitäten bei der Benotung und den Übertrittsempfehlungen wurde in der jüngeren Vergangenheit auch die Bedeutung der bereits angesprochenen Referenzgruppeneffekte bei der Leistungsbewertung untersucht. Zudem wurden zwei weitere, bislang aber kaum empirisch untersuchte Quellen für nicht-zufällige Unterschiede bei der Übertrittsempfehlung betrachtet: interindividuelle Unterschiede zwischen einzelnen Lehrpersonen im Sinne von Härte-/Mildeeffekten (vgl. Maier, 2007) sowie systematische Unterschiede zwischen einzelnen Schulen oder Schulklassen.

Forschung zu systematischen Referenzgruppeneffekten bei der Benotung können bis auf Ingenkamp (1969, 1971) zurückgeführt werden, der hervorhob, dass bei der Benotung meist klasseninterne Vergleichsrahmen zur Anwendung kommen. Tatsächlich lieferten in jüngerer Vergangenheit eine Reihe von Studien aus Deutschland und der Schweiz (z.B. Maaz, Neumann et al., 2008; Milek et al., 2009; Neumann, Milek, Maaz & Gresch, 2010, Tiedemann & Billmann-Mahecha, 2007; Trautwein & Baeriswyl, 2007) Hinweise darauf, dass Lehrkräfte die Schulnoten sowie die eng mit den Schulnoten verbundenen Übertrittsempfehlungen für die einzelne Schülerin/den einzelnen Schüler nicht unabhängig vom Leistungsniveau ihrer/seiner Klassenkameraden vergeben. In leistungsstärkeren Lerngruppen erhalten Schülerinnen und Schüler dabei für die gleiche Leistung eine schlechtere Note wie ein vergleichbarer Schülerinnen und Schüler in einer leistungsschwachen Klasse. So kann es beispielsweise vorkommen, dass Schülerinnen und Schüler, die in einer sehr leistungsstarken Grundschulklasse unterrichtet werden, u.U. eine Hauptschulempfehlung erhalten, obwohl ihre individuellen Schulleistungen eigentlich im Kernbereich der typischerweise Realschulempfohlenen liegen.

Darüber hinaus gibt es auch Hinweise auf Unterschiede zwischen einzelnen Lehrkräften im Sinne von stabilen Härte-/Mildeeffekten, die jedoch bislang vor allem in Form

anekdotischer Berichte vorliegen. Eine wichtige Ausnahme stellt die Studie von Maier (2007) dar. Maier ging der Frage nach, „ob über einen längeren Zeitraum hinweg innerhalb einer Grundschule systematische Lehrereffekte (Urteilstendenzen) bei der Vergabe von Hauptschul-, Realschul- und Gymnasialempfehlungen nachgewiesen werden können“ (S. 275). Zu diesem Zwecke prüfte Maier (2007), ob sich die über mehrere Jahre hinweg ermittelten, durchschnittlichen Übergangsquoten einzelner Grundschullehrkräfte innerhalb von Schulen voneinander unterscheiden. Maier (2007) bediente sich der Daten des Statistischen Landesamtes Baden-Württemberg für den Zeitraum von 1991 bis 2006, zu denen auch die Übergangsquoten für jede Schule gehören. Leider liegen diese Übergangsquoten nicht für einzelne Lehrkräfte vor. Deshalb musste Maier für seine Analysen eine letztlich nicht prüfbare Modellannahme treffen: In einzügigen Grundschulen sollte derselbe Lehrer tendenziell alle zwei Jahre eine Klasse in Klassenstufe 4 als Klassenlehrer unterrichten, nicht aber jedes Jahr. Wenn es nun tatsächlich Lehrerunterschiede in Übertrittsquoten gibt, die als stabile Härte-/Mildeeffekte interpretiert werden können, sollten die Unterschiede der Übertrittsquoten im Einjahresabstand größer ausfallen als im Zweijahresabstand. Bei größeren Schulen mit mehreren Parallelklassen sollte sich dieses Muster als weniger stark ausgeprägt erweisen, da sich differentielle Übertrittsquoten über die unterschiedlichen Lehrkräfte ausmitteln sollten. Tatsächlich bestätigten die Ergebnisse der Studie Maiers (2007) *cum grano salis* die Annahmen des Autors. Einschränkend ist allerdings auf Grenzen der Studie hinzuweisen. So wurden für zentrale Analysen keine Kennwerte der statistischen Analysen berichtet. Darüber hinaus ist die Kalkulation der verwendeten statistischen Größen nicht immer einfach nachvollziehbar und die Größe des beobachteten Effekts wurde nicht berichtet. Die vielleicht größte Beschränkung besteht aber darin, dass es nicht möglich ist, die zentralen Modellannahmen zur Stabilität und Wechsel von Klassenlehrern empirisch zu prüfen. Diese und weitere Grenzen der Studie wurden vom Autor ausführlich dokumentiert.

Baeriswyl, Wandeler und Trautwein (2011) konnten die Annahme, dass sich Lehrkräfte systematisch im Sinne von Härte-/Mildeeffekten unterscheiden, weiter erhärten. Diese Autoren nutzten dafür die ungewöhnlich gute empirische Datenlage, die im Rahmen einer über viele Jahre laufenden wissenschaftlichen Begleitung der Übertrittsentscheidungen im Kanton Freiburg (Schweiz)

entstand. Insgesamt lagen Daten aus 16 Schuljahren von 209 Lehrkräften aus 30 Primarschulen im Kanton Freiburg (Schweiz), die in 631 Klassen bei insgesamt 10654 Schülern eine Übertrittsempfehlung aussprachen. Baeriswyl et al. spezifizierten Mehrebenenmodelle zur Vorhersage von Übertrittsentscheidungen, bei denen für die individuelle Schülerleistung aus einem standardisierten Schulleistungstest kontrolliert wurde. In diesen Mehrebenenanalysen fanden sich, wie erwartet, Belege für Referenzgruppeneffekte (bei gleicher individueller Leistung fiel die Übertrittsempfehlung in leistungsstarken Klassen niedriger aus). Darüber hinaus zeigten sich jedoch auch Unterschiede in substantieller Größe zwischen einzelnen Lehrkräften. In anderen Worten: Bei vergleichbarer Schulleistung der Schülerinnen und Schüler gab es Lehrkräfte, die über mehrere Kohorten hinweg eher vielen oder eher wenigen Schülerinnen und Schüler eine Empfehlung für hohe oder niedrige Schulformen gaben. Auch zwischen den untersuchten Schulen gab es Unterschiede in dem Sinne, dass beispielsweise an manchen Schulen die „Schwelle“ für eine Empfehlung fürs Progymnasium niedriger lag als bei anderen Schulen.

Zusammenfassend lässt sich resümieren, dass es neben systematischen Effekten des sozialen Hintergrunds auch weitere systematische Effekte gibt, die im Widerspruch zu meritokratischen Grundsätzen bei der Übertrittsentscheidung stehen. Insbesondere die Referenzgruppeneffekte auf Klassenebene sind dabei von nicht zu vernachlässigender Größe. Auch wenn die vorgelegte Expertise Referenzgruppeneffekten im Folgenden keine größere Aufmerksamkeit mehr zukommen lässt, da ihr Schwerpunkt auf sozialen Disparitäten liegt, soll nochmals deren Bedeutung bei allen Überlegungen zur Verbesserung der Gerechtigkeit bei Übergangsprozessen hervorgehoben werden.

3. Soziale Herkunft und Leistungsbewertung: Empirische Studien zu zentralen Fragestellungen

In diesem Kapitel werden zehn aktuelle Fragen zum Zusammenhang zwischen Merkmalen der sozialen Herkunft und der Leistungsbewertung thematisiert und jeweils in einer Teilstudie untersucht. Die Teilstudien folgen einer einheitlichen Struktur. Nach Formulierung der zu untersuchenden Frage wird in knapper Form auf den Forschungshintergrund eingegangen. Daran schließt sich eine kurze Beschreibung der verwendeten Daten sowie der Analysestrategie an. Im Ergebnisteil werden die Hauptbefunde zusammengefasst und abschließend wird ein kurzes Resümee gezogen.

3.1. Teilstudie 1: Soziale Herkunft und Schulnoten: Gibt es substanzielle Zusammenhänge?

Frage 1: Schlägt sich in den Leistungsbewertungen der Lehrkräfte neben der Schulleistung eines Schülers oder einer Schülerin auch die soziale Herkunft der Kinder und der Migrationsstatus nieder?

3.1.1. HINTERGRUND

Ein zentrales Strukturmerkmal des deutschen Bildungssystems ist die frühe Leistungsdifferenzierung, die sich in den unterschiedlichen Schulformen bzw. Bildungsgängen des Sekundarschulsystems ausdrückt. Während in Berlin und in Brandenburg nach der sechsten Jahrgangsstufe differenziert wird, erfolgt die Differenzierung in den anderen Bundesländern bereits nach der vierten Klassenstufe, zuweilen abgemildert durch Orientierungsstufen (z.B. in Mecklenburg-Vorpommern). Ähnliche Formen der Differenzierung, wenn auch zeitlich etwas verzögert, finden sich vor allem in den anderen deutschsprachigen Staaten (vgl. Maaz, Neumann & Trautwein, 2009). Die Leistungsdifferenzierung geht einher mit dem Erwerb unter-

schiedlicher Abschlusszertifikate, die wiederum die nachfolgenden Ausbildungsoptionen beeinflussen und damit auch berufliche Karrieren bahnen. Der Wechsel von der Grundschule in die jeweilige weiterführende Schule ist ohne Zweifel eine der zentralen Statuspassagen im Leben junger Menschen (Arnold, Bos, Richert & Stubbe, 2007; Maaz, Baumert & Trautwein, 2009; Maaz, Gresch, McElvany, Jonkmann & Baumert, 2010).

In diesem Übergangsprozess kommt den Schulnoten eine besondere Rolle zu. Über die Schulnoten wird die Leistungsfähigkeit eines Kindes dokumentiert. Die elterliche Einschätzung über den Leistungsstand ihres Kindes wird somit stark von den Schulleistungen beeinflusst. In allen Ausformungen des Transitionsprozesses sind die von den Grundschulen erstellten Empfehlungen über den Besuch einer weiterführenden Schule von maßgeblicher Bedeutung. Diese Empfehlung kann entweder bindend sein oder einen empfehlenden Charakter haben. Auch wenn die Empfehlung ausschließlich empfehlenden Charakter hat, orientieren sich Eltern bei der Entscheidungsgenese an den Empfehlungen. Die Empfehlungen basieren in den meisten Bundesländern in erster Linie auf Schulnoten.

Das deutsche Bildungssystem ist stark föderal geprägt. Die einzelnen Bundesländer unterscheiden sich nicht nur in Hinblick auf die Zahl der angebotenen Bildungsgänge im Sekundarschulsystem, sie unterscheiden sich teilweise auch recht deutlich in den konkreten Übergangsregelungen (KMK, 2009; Füßel, Gresch, Baumert & Maaz, 2010; Kropf, Gresch & Maaz, 2010), zum Beispiel beim Entscheidungsspielraum der Eltern. Unterschiede gibt es ebenfalls in der Art und Weise, wie die Leistungsfähigkeit der Schülerinnen und Schüler festgestellt wird, etwa durch Einsatz standardisierter Leistungstests oder durch das Erstellen von ausführlichen Grundschulgutachten, die Aspekte wie das Arbeitsverhalten oder die Motivation berücksichtigen (Ditton & Krüskens, 2006). In allen Bundesländern

jedoch kommt den Schulnoten für die Bildungsgangempfehlung eine starke Bedeutung zu. So fanden Bos et al. (2004; Arnold et al., 2007) im Rahmen der Internationalen Grundschul-Lese-Untersuchung IGLU einen besonders engen Zusammenhang zwischen Schulnoten und Schullaufbahneempfehlungen, wobei zugleich die Leistungen in standardisierten Tests substanziell mit Noten und der Schullaufbahneempfehlung assoziiert waren.

Befunde zur Reliabilität und Validität von Schulnoten (z.B. Baumert et al., 2003a) weisen allerdings ebenfalls darauf hin, dass der Zusammenhang zwischen Schulnoten und Testleistungen keineswegs perfekt ist, sondern teilweise durch den Einfluss von individuellen Schülermerkmalen wie Geschlecht und soziale Herkunft überlagert wird.

In dieser Teilstudie gehen wir der Frage nach, ob sich ein Einfluss der sozialen Herkunft auf die Notengebung am Ende der Grundschulzeit nachweisen lässt, wenn für die unterschiedliche, mit der Sozialschicht der Schülerinnen und Schüler assoziierte Leistung in einem standardisierten Leistungstest kontrolliert wird. Es handelt sich um eine Replikation und Erweiterung vorliegender Forschungsbefunde mithilfe von repräsentativen Daten aus Berlin bzw. dem gesamten Bundesgebiet.

3.1.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Mit den Daten der in Berlin durchgeführten Untersuchung „Erhebungen zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin“ (ELEMENT) (Lehmann & Lenkeit, 2008) sollen zentrale Befunde zum Einfluss individueller Schülermerkmale auf die Notenvergabepraxis der Grundschullehrkräfte am Ende der sechsten Klassenstufe untersucht werden. In dieser längsschnittlich angelegten Untersuchung wurde die Entwicklung der Deutsch- und Mathematikleistungen der Schülerinnen und Schüler des Primarbereichs von der vierten bis zur sechsten Klasse untersucht (vgl. Lehmann & Lenkeit, 2008). In Jahrgangsstufe 6 wurden zudem die Englischleistungen querschnittlich erhoben. Die Leistungstests wurden durch Schüler-, Eltern- und Lehrerbefragungen ergänzt.

Für die Analyse der Zusammenhangsstruktur in Bildungssystemen in vierjähriger Grundschulzeit wird auf die am Berliner Max-Planck-Institut für Bildungsforschung durchgeführte Studie „Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (ÜBERGANG) (Maaz, Baumert, Gresch & McElvany, 2010;

Becker, Gresch, Baumert, Watermann, Schnitger, & Maaz, 2010) zurückgegriffen. Weitere Informationen zur Studienanlage finden sich im Anhang 1.

Folgende Operationalisierungen liegen den Analysen mit der TIMSS-Übergangsstudie zugrunde.

Sozioökonomische Stellung der Eltern. Zur Beschreibung der sozioökonomischen Stellung dient der auf den elterlichen Berufangaben basierende Internationale Sozioökonomische Index (ISEI), den Ganzeboom, DeGraaf, Treiman & De Leeuw (1992) entwickelt haben. Für die Analysen wurde der höchste sozioökonomische Index von Vater bzw. Mutter gewählt.

Mathematik- und Naturwissenschaftsleistung. Die Leistungen in diesen Domänen wurden über das internationale Design der TIMSS-2007 erfasst (vgl. Anhang 1). Beide Tests erwiesen sich als hoch reliabel (Cronbachs Alpha von $\alpha = 0,83$ für Mathematik und einem Cronbachs Alpha von $\alpha = 0,80$ für Naturwissenschaft) (vgl. Walther, Selter, Bensen & Bos, 2008; Wittwer, Saß & Prenzel, 2008).

Deutschleistung. Für die Erfassung der Deutschleistung wurden 446 Items der Normierung Deutsch Primarstufe 2007 herangezogen. Die Items verteilen sich relativ homogen auf vier Kompetenzbereiche (*Lesen, Hören, Sprachgebrauch* und *Rechtschreibung*), sodass die aus dem Modell ermittelte Kompetenzverteilung quasi ein *Composite* dieser vier Bereiche darstellt. Aus ConQuest wurde eine WLE-Reliabilität von 0,81 geschätzt.

Durchschnittsleistung. Darüber hinaus wurde ein Leistungsindikator aus den drei Subdomänen gebildet, dessen Gesamtscore erwartungsgemäß hinreichend reliabel ($\alpha = 0,80$) war.

Als Indikator der *kognitiven Leistungsfähigkeit* wurde der figurale Subtest N2 des Kognitiven Fähigkeits-Tests für 4. Klassen (KFT 4–12+R; Heller & Perleth, 2000) verwendet. Die Tests wurden in zwei parallelisierten Versionen A und B administriert. Die Reliabilität dieser Tests liegt für Version A bei Cronbachs $\alpha = 0,92$ und für Version B bei $\alpha = 0,93$ (vgl. Heller & Perleth, 2000).

Gymnasialübergang. Die Informationen zum Übergang in die Sekundarstufe I entstammten einem Elternfragebogen, der am Ende der 4. Klasse eingesetzt wurde. Bei fehlenden Angaben wurde zudem auf den Schülerfragebogen zurückgegriffen.

Noten. Die Noten der Schülerinnen und Schüler in den Fächern Mathematik, Deutsch und Sachkunde basierten auf Angaben der Schule. Aus diesen drei Noten wurde eine Durchschnittsnote gebildet. Um ein Kriterium für

Tabelle 2: Vorhersage der Durchschnittsnote in der Bildungsgangempfehlung durch individuelle Schülermerkmale des sozio-kulturellen familiären Hintergrund und zentraler Leistungsindikatoren (ELEMENT)

	Modell 1	Modell 2
	<i>b</i>	<i>b</i>
Individualebene		
Geschlecht (Ref. Jungen)		
Mädchen	-0,26	-0,30
Ausbildung der Eltern (Ref. Lehre)		
ohne Abschluss	-0,06	-0,09
Fachschulabschluss	0,04	0,05
FH-Abschluss	-0,10	-0,04
Uni-Abschluss	-0,18	-0,02
Migration (Ref. beide Eltern in Dt. geb.)		
ein Elternteil im Ausland geboren	0,04	-0,02
beide Elternteile im Ausland geboren	0,13	-0,03
ISEI ⁺	-0,18	-0,06
Leistungsindex Klasse 6 ⁺		-0,44
Kognitive Grundfähigkeit (KFT) ⁺		-0,08
<i>R</i> ² (Level 1)	0,11	0,46
<i>R</i> ² (Level 2)	0,43	0,60

Anmerkungen: ⁺ Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

„gute“ und „schlechte“ Noten zu erhalten, wurde der Durchschnittswert 2,33 als Cut-Off-Kriterium festgelegt. Der Wert korrespondiert mit dem Notendurchschnitt, der in Bayern für den Erhalt einer Gymnasialempfehlung erforderlich ist.

Bildungs- und Ausbildungsmerkmale wurden von den Eltern direkt erfasst.

Folgende Operationalisierungen liegen den Analysen mit der ELEMENT-Übergangsstudie zugrunde (vgl. Anhang 2).

In der Elementstudie wurden familiäre Hintergrundmerkmale wie in der TIMSS-Übergangsstudie direkt von den Eltern mit einem Fragebogen erfasst. Der sozioökonomische Status wurde auch mit dem Socio-Economic Index of Occupational Status (ISEI) (den Ganzeboom et al., 1992) erhoben. Die Leistungsindikatoren wurden wie folgt operationalisiert.

Schulischer Leistungsindex. Zur Bestimmung der Leseleistungen am Ende der sechsten Klasse wurden Tests zum Leseverstehen eingesetzt, die bereits in der Hamburger Studie LAU für Klassenstufe 6 und 7 (Aspekte der Lernausgangslage und der Lernentwicklung; Lehmann, Gänsfuß & Peek, 1999) sowie in der IGLU-Studie (Bos et

al., 2003) zum Einsatz kamen. Der Test bestand aus vier kurzen Prosatexten und einem Gebrauchstext, zu denen anschließend Fragen zum Verständnis gestellt wurden. Unter den insgesamt 37 Fragen befanden sich 6 Aufgaben mit offenem Antwortformat, die restlichen Fragen wurden im Multiple Choice-Format vorgelegt. Die Reliabilität des Tests war mit .84 (Kuder-Richardson Formula 20) zufrieden stellend. Die Mathematikleistung wurde mit Instrumenten aus der LAU-Untersuchung (Lehmann et al., 1999), der IGLU-Studie (Bos et al., 2003) sowie der Brandenburgischen Mathematikuntersuchung QuaSUM (vgl. Lehmann et al., 2001) gemessen. Die insgesamt 49 Aufgaben stammten zum größten Teil aus den Stoffgebieten Arithmetik und Geometrie. Die Reliabilität des Tests fiel mit .92 (Kuder-Richardson Formula 20) gut aus. Beide Tests wurden unter Nutzung des Partial-Credit-Modells Rasch-skaliert. Bei den Personenparametern handelt es sich jeweils um gewichtete Likelihood-Schätzer (Warm, 1989). Um einen validen Leistungsindikator für Berücksichtigung der leistungsbezogenen Komposition der Klasse zu bekommen und der Multikollinearität auf Klassenebene zu begegnen, wurde aus den Testleistungen der

Tabelle 3: Korrelationsmuster alle leistungsbezogener Schülermerkmale

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Durchschnittsleistung	1	0,86	0,83	0,84	0,50	0,71	0,64	0,65	0,58
(2) Mathematikleistung		1	0,57	0,60	0,48	0,64	0,54	0,62	0,51
(3) Deutschleistung			1	0,52	0,40	0,61	0,60	0,54	0,48
(4) Naturwissenschaftsleistung				1	0,39	0,54	0,48	0,48	0,47
(5) Kognitive Grundfähigkeit					1	0,46	0,40	0,43	0,38
(6) Durchschnittsnote						1	0,89	0,88	0,88
(7) Mathematiknote							1	0,67	0,69
(8) Deutschnote								1	0,64
(9) Sachkundenote									1

Anmerkung: Alle Korrelationen sind auf dem 5-Prozentniveau statistisch signifikant.

Schülerinnen und Schüler in Lesen und Mathematik ein Leistungsindex gebildet, indem die beiden Leistungswerte nach einer Standardisierung an der Gesamtstichprobe ($N = 3.008$) gemittelt wurden.

Kognitive Grundfähigkeiten. Die kognitiven Grundfähigkeiten der Grundschüler wurden am Ende der vierten Klasse (erste Erhebungswelle) mit zwei Untertests des Kognitiven Fähigkeitstests (KFT, Heller & Perleth, 2000) erhoben. Die Subtests erfassen verbales und figürliches Schlussfolgern und gelten als Markertest für fluide Intelligenz. Die Reliabilität des Gesamttests betrug .93 (Kuder-Richardson Formula 20).

Durchschnittsnote. Als Indikator für die Noten wurde der in den Bildungsgangempfehlungen errechnete Notendurchschnitt verwendet. Dieser berücksichtigt Noten ab dem Schuljahr 5. Noten aus dem ersten Halbjahr der Klassenstufe 6 gehen doppelt in die Berechnung ein. Darüber hinaus werden die Mathematik-, Deutsch- sowie die Note der ersten Fremdsprache doppelt gewichtet.

3.1.3. ERGEBNISSE

Mit den Daten der ELEMENT-Studie können Schulnoten in einem Bildungssystem mit sechsjähriger Grundschulzeit analysiert werden. Zur Vorhersage der Schulnoten der Schülerinnen und Schüler wurde der Notendurchschnitt, wie er in der Bildungsgangempfehlung Eingang findet, berücksichtigt. Die Note wurde für diese Analysen nicht rekodiert. Somit steht ein hoher Wert für eine schlechte Note und ein niedriger Wert für eine gute. Tabelle 2 fasst die zentralen Befunde zusammen. Es zeigte sich, dass die individuelle Leistung der standardisierten Tests mit den Schulnoten assoziiert waren. Kinder aus sozial begünstigten Familien erhielten darüber hinaus bessere Noten als Kinder aus sozial weniger begünstigten Familien. Bei Hinzunahme der individuellen Testleistungen in das Regressionsmodell verringert sich der Effekt des sozioökonomischen Status von $b = -0,18$ auf $b = -0,06$,

erreichte aber weiterhin das Signifikanzkriterium von $p < 0,05$. Diese partielle Mediation deutet darauf hin, dass der Einfluss der sozialen Herkunft auf die Notenvergabe zumindest zu Teilen auf unterschiedliche Leistungen, wie sie mit einem standardisierten Test erfasst werden, zurückzuführen sind. Der auch bei Kontrolle zentraler Leistungsmerkmale bestehende Herkunftseffekt bedeutet aber auch, dass Kinder aus sozial begünstigten Familien in der Grundschule bessere Schulnoten bekommen als Kinder aus sozial weniger begünstigten Familien und zwar auch dann, wenn die Kinder vergleichbare Ergebnisse in den standardisierten Leistungstests erzielt haben. Anders sieht es beim Migrationshintergrund der Kinder aus. Dieser hat keinen signifikanten Einfluss auf die Notengebung. Damit findet die These, wonach Kinder mit Migrationshintergrund bei der Notengebung benachteiligt werden, keine empirische Evidenz.

In einem zweiten Schritt sollen Noten in Bildungssystemen mit vierjähriger Grundschulzeit auf der Basis der TIMSS-Übergangsstudie betrachtet werden. Dies betrifft die große Mehrheit der Bundesländer. Aktuell haben nur Berlin und Brandenburg eine sechsjährige Grundschulzeit. Sieht man von Mecklenburg-Vorpommern ab, wo der fünften und sechsten Jahrgangsstufe eine besondere Bedeutung zukommt, werden in 13 Ländern im Laufe des vierten Schuljahres die Weichen für die weitere Bildungslaufbahn gestellt.

(→ TABELLE 2)

Tabelle 3 stellt die Zusammenhängestruktur der einzelnen leistungsbezogenen Schülermerkmale dar. Erwartungsgemäß fanden sich hohe bis sehr hohe Zusammenhänge zwischen den Leistungsindikatoren und den Schulnoten. Der engste Zusammenhang resultierte mit $r = 0,88$ zwischen der Durchschnittsnote und der Mathematiknote.

(→ TABELLE 3)

Im Folgenden werden die Anteile der durch individuelle Schülermerkmale aufgeklärten Varianz an der Durchschnittsnote berichtet. Aus Tabelle 4, welche die Ergebnisse für die Durchschnittsleistung darstellt, geht hervor, dass der kombinierte Leistungsindikator den größten Anteil der Varianz an der Durchschnittsnote erklärt (49,4%), wenn man die Prädiktorvariablen jeweils einzeln einführt. Gefolgt wird er von den kognitiven Grundfähigkeiten (20,9%) sowie dem elterlichen Bildungshintergrund (19,4%). Der elterliche Bücherbesitz klärt 18,3 Prozent der Varianz an der Durchschnittsnote auf. Der sozioökonomische Status allein bindet 17,2 Prozent der Varianz. Die geringsten Varianzanteile erklären der Migrationshintergrund und das Geschlecht.

(→ TABELLEN 4 UND 5)

Die Anteile aufgeklärter Varianz der zuvor beschriebenen Indikatoren unterscheiden sich zum Teil zwischen den drei Fachnoten. Abgesehen von dem kombinierten Leistungswert erklären die Fachleistungen die größten Varianzanteile in Mathematik und in Deutsch. Die kognitive Grundfähigkeit bindet in Mathematik mehr Varianz als in den beiden anderen Fachnoten. Das Geschlecht erklärt insgesamt nur wenig Varianz, wobei es deutliche Unterschiede zwischen den Fachnoten gibt.

Der aus den Leistungen in den Domänen Deutsch, Mathematik und Naturwissenschaft gebildete mittlere Leistungsindikator ist der stärkste Prädiktor für die Vorhersage der mittleren Grundschulnoten. Die mit einem standardisierten Leistungstest gemessenen Leistungen erklären insgesamt ca. 49 Prozent der Variabilität der Schulnoten (M1). Die zusätzliche Berücksichtigung der kognitiven Grundfähigkeiten, die einen zusätzlichen Effekt aufweisen, verbessert die Varianzaufklärung um lediglich 2,4 Prozent (M2).

In den Modellen 3 und 4 wurden Indikatoren der sozialen Herkunft mit in das Modell aufgenommen. Dabei wird der von Baumert und Kollegen getroffenen Unterscheidung zwischen familiären Struktur- und Prozessmerkmalen Rechnung getragen (vgl. Baumert, Watermann & Schümer, 2003). In der jüngeren Bildungsforschung hat sich diese differenzierte Konzeption sozialer Herkunftsmerkmale bewährt, da in verschiedenen Untersuchungen gezeigt werden konnte, dass Statusmerkmale wie z.B. die sozioökonomische Stellung einer Familie alleine nicht ausreichen, um den Zusammenhang zwischen sozialer

Herkunft und Bildungserfolg zu verstehen. So konnten Baumert, Watermann und Schümer (2003) anhand von Daten der PISA-Studie zeigen, dass der Effekt der Sozialschichtzugehörigkeit auf die Schulleistungen zu einem großen Teil über familiäre Prozessmerkmale wie der kulturellen und kommunikativen Praxis in der Familie vermittelt wird. Sie kommen daher zu dem Schluss, dass erst die simultane Berücksichtigung von Struktur- und Prozessmerkmalen des familiären Hintergrunds eine angemessene Vorstellung der anzutreffenden sozialen Disparitäten liefert (Baumert et al., 2003b; Maaz & Watermann, 2007; Jungbauer-Gans, 2004, 2006). Bisher gibt es keine Erkenntnis darüber, ob der Zusammenhang zwischen Merkmalen der sozialen Herkunft und der Notengebung in einem ähnlich differenzierten Zusammenhang steht.

Modell 3 berücksichtigt zunächst nur die familiären Strukturmerkmale wie den sozioökonomischen Status und die Bildungsqualifikation der Eltern. Es zeigten sich signifikante Effekte für den sozioökonomischen Status und das allgemeine Bildungsniveau der Eltern. Bei gleichen schulischen Leistungen werden Kinder aus sozial begünstigten und bildungsnahen Familien weniger streng benotet als Kinder aus sozial schwächeren Familien. Für das berufliche Qualifikationsniveau der Eltern konnte kein signifikanter Effekt nachgewiesen werden. Dies ist unter anderem auf die Konfundierung der verschiedenen Strukturmerkmale zurückzuführen. Mit den drei Indikatoren konnten insgesamt ca. 3,4 Prozent zusätzlich an Varianz aufgeklärt werden. In Modell 4 wurden zusätzlich die familiären Prozessmerkmale wie der elterliche Bücherbesitz und die kulturellen Aktivitäten der Eltern berücksichtigt. Es konnte ein schwacher aber signifikanter Effekt der Partizipation an der gesellschaftlichen Hochkultur festgestellt werden. Die Integration familiärer Prozessmerkmale führte auch zu keiner substanziellen Modellverbesserung. Die im Modell bereits enthaltenen Variablen blieben insgesamt stabil. Diese Befunde deuten darauf hin, dass anders als beim Kompetenzerwerb und in Entscheidungssituationen soziale Hintergrundmerkmale nicht differenziert auf die Schulnoten wirken.

Die Ergebnisse deuten des Weiteren auf einen Geschlechtereffekt hin, wonach Mädchen in der Leistungsbewertung besser abschneiden als Jungen – auch dann, wenn die individuelle Testleistung kontrolliert wird (vgl. Teilstudie 9).

(→ TABELLE 6)

Tabelle 4: Anteile aufgeklärter Varianz durch Schülermerkmale an der Durchschnittsnote auf Individualebene

	Durchschnittsnote
Kombinierter Leistungswert	49,4
Kognitive Grundfähigkeit (KFT)	20,9
Geschlecht	0,4
Sozioökonomischer Status (ISEI)	17,2
Elterliche Ausbildung, schulisch	19,4
Elterliche Ausbildung, beruflich	7,0
Bücherbesitz	18,3
Kulturelle Aktivitäten Eltern	11,1
Migration	4,7

Tabelle 5: Anteile aufgeklärter Varianz durch Schülermerkmale an den Fachnoten in Mathematik, Deutsch und Sachkunde auf Individualebene

	Mathematiknote	Deutschnote	Sachkundenote
Kombinierter Leistungswert	41,5	40,6	33,2
Mathematikleistung	38,8	29,1	25,8
Deutschleistung	28,3	35,6	23,3
Naturwissenschaftsleistung	22,5	22,5	21,8
Kognitive Grundfähigkeit (KFT)	18,6	15,9	14,3
Geschlecht	0,4	2,8	0,4
Sozioökonomischer Status (ISEI)	13,1	14,0	13,2
Elterliche Ausbildung, schulisch	13,5	16,7	15,2
Elterliche Ausbildung, beruflich	4,7	6,1	5,5
Bücherbesitz	12,4	15,2	15,4
Kulturelle Aktivitäten Eltern	6,6	9,6	10,0
Migration	2,4	4,0	4,8
Gewissenhaftigkeit im Unterricht	9,3	11,9	10,5
Anstrengungsbereitschaft	7,0	6,2	5,6

Tabelle 6: Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum sozio-kulturellen familiären Hintergrund und zentrale Leistungsindikatoren

	Bivariat	M1	M2	M3	M4
		<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Kombinierter Leistungswert ⁺	0,56	0,56	0,50	0,44	0,43
Kognitive Grundfähigkeit (KFT) ⁺	0,36		0,11	0,09	0,09
Geschlecht (Ref. Mädchen)					
Jungen	-0,10			-0,10	-0,10
Sozioökonomischer Status (ISEI) ⁺	0,33			0,08	0,06
Ausbildung, schulisch (Ref. Abitur)					
Hauptschulabschluss	-0,99			-0,33	-0,29
Mittlere Reife	-0,44			-0,11	-0,09
Fachhochschulreife	-0,32			-0,09	-0,08
Ausbildung, beruflich (Ref. Uni-Abschluss)					
nicht universitär Abschlüsse	0,54			0,00	-0,01
Bücherbesitz (Ref. 26-100)					
weniger als 25	-0,45				-0,08
101 bis 250	0,33				0,03
251 bis 500	0,53				0,03
Mehr als 500	0,70				0,04
Kulturelle Aktivitäten Eltern ⁺	0,26				0,04
<i>R</i> ²		49,40	51,00	55,00	55,40

Anmerkungen: ⁺ Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; SES = sozioökonomischer Status; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

Bezüglich des Migrationshintergrunds der Grundschülerinnen und -schüler konnte in der bivariaten Analyse ein Effekt zu Ungunsten der Kinder mit Migrationshintergrund ausgewiesen werden (vgl. Tabelle 7, M1). Allein aus der deskriptiven Betrachtung lassen sich aber keine Rückschlüsse auf migrationsspezifische Benachteiligungen bei der Benotungspraxis schließen, da Migrationsmerkmale sowohl mit der sozialen Herkunft als auch mit Leistungsmerkmale konfundiert sein können. Tabelle 7 stellt die Befunde der migrationsspezifischen Analysen zusammenfassend dar. Berücksichtigt man allein den Migrationsstatus, lässt sich ein negativer Migrationseffekt nachweisen. Um zu prüfen, ob dieser Effekt auf andere Schülermerkmale zurückzuführen ist, werden schrittweise die Leistungsindikatoren und die Merkmale der sozialen Herkunft in das Modell aufgenommen. In Modell 2 wird zusätzlich für die Leistung kontrolliert. Der Migra-

tionseffekt verringert sich von $b = -0,45$ auf $b = -0,07$. Diese deutliche Reduzierung des Effekts deutet darauf hin, dass Kinder mit Migrationshintergrund insgesamt weniger gute Ergebnisse in den Leistungstests erzielt haben und dies bei den Lehrkräften auch zu unterschiedlichen Benotungen geführt hat. Auch wenn sich der Migrationseffekt deutlich reduziert hat, ist er weiterhin statistisch signifikant. In Modell 3 werden schließlich die verschiedenen Indikatoren der sozialen Herkunft berücksichtigt. Hier zeigte sich, dass bei Kontrolle der Leistungsmerkmale und der sozialen Herkunft kein zusätzlicher Migrationseffekt besteht. Migrationseffekte, wie sie in deskriptiven Analysen sichtbar werden, sind den vorliegenden Befunden zufolge nicht auf migrationsspezifische Merkmale zurückzuführen, sondern auf leistungs- und soziale Herkunftseffekte.

(→ TABELLE 7)

Tabelle 7: Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum Migrationsstatus, zum sozio-kulturellen familiären Hintergrund und zentraler Leistungsindikatoren

	M1	M2	M3
	<i>b</i>	<i>b</i>	<i>b</i>
Kombinierter Leistungswert [†]		0,50	0,42
Kognitive Grundfähigkeit (KFT) [†]		0,11	0,09
Geschlecht (<i>Ref. Mädchen</i>)			
Jungen		-0,11	-0,10
Sozioökonomischer Status (ISEI) [†]			0,06
Ausbildung, schulisch (<i>Ref. Abitur</i>)			
Hauptschulabschluss			-0,29
Mittlere Reife			-0,09
Fachhochschulreife			-0,08
Ausbildung, beruflich (<i>Ref. Uni-abschluss</i>)			
nicht universitär Abschlüsse			-0,01
Bücherbesitz (<i>Ref. 26-100</i>)			
weniger als 25			-0,08
101 bis 250			0,03
251 bis 500			0,03
Mehr als 500			0,04
Kulturelle Aktivitäten Eltern [†]			0,04
Migration (<i>Ref. maximal ein Elternteil in Deutschland geboren</i>)			
beide Eltern im Ausland geboren	-0,45	-0,07	0,00
<i>R</i> ²	4,70	51,50	55,40

Anmerkungen: [†] Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; SES = sozioökonomischer Status; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

3.1.4. RESÜMEE

Die Befunde dieser Teilstudie mit aussagekräftigen Daten zum Ende der Grundschulzeit replizieren den Befund, dass sich ein direkter Effekt der sozialen Herkunft auf die Notengebung der Lehrkräfte in dem Sinne nachweisen lässt, dass auch nach Kontrolle der Leistung in standardisierten Leistungstests noch eine statistisch signifikante Assoziation zwischen Herkunft und Note gegeben ist. Dieser Zusammenhang fällt zwar insgesamt nicht sehr groß aus, ist aber statistisch bedeutsam. Keine differenziellen Befunde konnten zu familiären Struktur- und Prozessmerkmalen identifiziert werden. Für den Migrationshintergrund wurden keine direkten Effekte auf die Notenvergabe nachgewiesen, die auf eine mögliche Benachteiligung von Kindern mit Migrationshintergrund gerade an der Übergangsschwelle von der Primarstufe in die Sekundarstufe hindeuten würden.

Bei dem in dieser Teilstudie vorgestellten Zusammenhang zwischen Herkunft und Schulnoten handelt es sich um einen „Bruttoeffekt“, bei dem quasi unterstellt wird, dass die Ergebnisse in einem standardisierten Leistungstest ein valides Abbild der Gesamtleistung darstellen. Dies ist jedoch, wie oben dargestellt, nicht der Fall, da standardisierte Leistungstests immer nur einen Ausschnitt der benotungsrelevanten Fähigkeiten und Verhaltensweisen erfassen. Es wird deshalb in weiteren Teilstudien analysiert werden, welche Rolle beispielsweise die Mitarbeit im Unterricht und bei den Hausaufgaben für die Notenvergabe spielt und ob diese auch Herkunftseffekte bei der Notenvergabe erklären können.

3.2. Teilstudie 2: Anstrengungsbereitschaft und Gewissenhaftigkeit: Effekte auf die Notengebung?

Frage 2: Wie stark trägt das Schülerverhalten (Anstrengungsbereitschaft und Gewissenhaftigkeit im Unterricht) zur Notengebung bei? Lassen sich durch motivationale Schülermerkmale die dokumentierten Herkunftseffekte auf die Notengebung erklären?

3.2.1. HINTERGRUND

Der Zusammenhang zwischen Schulnoten und Ergebnissen aus Leistungstests ist mittlerweile gut dokumentiert (u.a. Arnold, Bos, Richert & Stubbe, 2010; Baumert, Trautwein & Artelt, 2003; Hochweber, 2010; Stubbe & Bos,

2008). In der vorherigen Teilstudie wurde herausgearbeitet, dass der Zusammenhang zwischen Schulnoten und Testleistungen alles andere als perfekt ist, sondern teilweise durch den Einfluss von individuellen Schülermerkmalen und soziale Herkunft überlagert sein kann. Ferner ist zu berücksichtigen, dass die Basis von Noten und die Basis von Leistungstests nicht deckungsgleich sind (vgl. Kapitel 2.4 und Abbildung 1). Noten dienen der Leistungsbewertung von Schülerinnen und Schülern. Eine Schulnote, wie sie beispielsweise in einem Zeugnis dokumentiert wird, setzt sich aus verschiedenen Einzelleistungen der Schülerinnen und Schüler zusammen. Hierzu zählen neben schriftlichen Arbeiten (z.B. Klassenarbeiten) auch mündliche Beiträge der Schülerinnen und Schüler im Unterricht und die Sorgfalt bei den Hausaufgaben. Greift man bei der Vorhersage von Schulnoten ausschließlich auf standardisierte Leistungstests zurück, kann nur eine Komponente – die schriftliche Leistung der Schülerin bzw. des Schülers – berücksichtigt werden. Der in vielen Studien nachgewiesene „Effekt“ der sozialen Herkunft auf die Notenvergabe könnte demnach eine Folge einer sogenannten „Unterspezifikation des Modells“ sein.

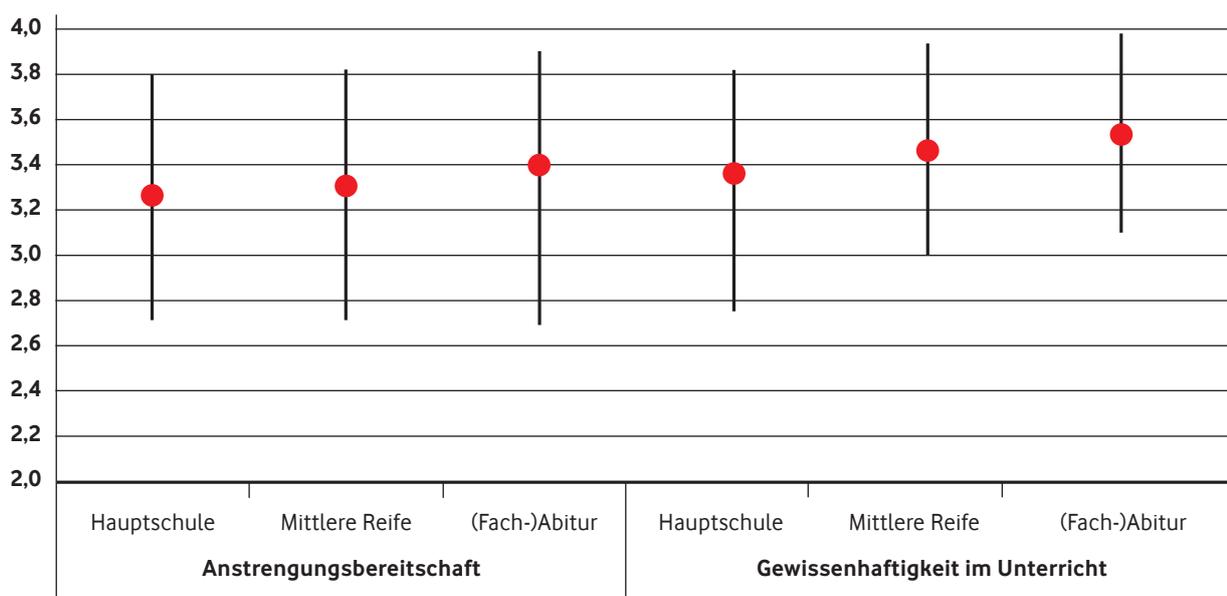
Darüber hinaus ist bereits aus früheren Studien bekannt, dass bei der Notenvergabe keine klassenübergreifenden Standards greifen (vgl. Ingenkamp, 1993). Lehrkräfte orientieren sich vorwiegend am mittleren Leistungsniveau der jeweiligen Klasse. Der klasseninterne Bezugsrahmen führt dazu, dass Bewertungsmaßstäbe über verschiedene Klassen hinweg unterschiedlich sind. Somit gibt es Klassen, in denen die mittlere Note identisch ausfällt, deren anhand von standardisierten Tests ermittelten mittleren Leistungen jedoch deutlich differieren.

Wir untersuchen daher in dieser Teilstudie, ob motivationale Schülermerkmale in einem substanziellen Zusammenhang mit der Notengebung stehen. Damit ist zunächst die Frage verbunden, wie groß dieser Zusammenhang ausfällt. In einem zweiten Schritt wird die Frage nach sozialschichtspezifischen Ausprägungen der motivationalen Merkmale gestellt. Inwieweit können sie den in der vorigen Teilstudie dokumentierten sozialen Herkunftseffekt erklären?

3.2.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Für die Analyse wird auf die aktuellste bundesweit repräsentative Studie zum Übergang in die Sekundarstufe I zurückgegriffen, die am Berliner Max-Planck-Institut für Bildungsforschung unter dem Namen „Der Übergang von

Abbildung 2: Mittelwerte und Standardabweichungen für die Schülermerkmale Gewissenhaftigkeit im Unterricht und Anstrengungsbereitschaft differenziert nach dem Bildungsgrad der Eltern



Anmerkung: Die roten fett gedruckten Punkte bilden den Mittelwert ab, die vertikalen Linien beschreiben den Bereich, in dem rund zwei Drittel der jeweiligen Gruppe liegen (eine Standardabweichung unterhalb und oberhalb des Mittelwerts).

der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (ÜBERGANG; Maaz, Baumert, Gresch & McElvany, 2010) durchgeführt wurde. Weitere Informationen zur Studienanlage finden sich im Anhang 1.

Gewissenhaftigkeit im Unterricht wurde von den Schülerinnen und Schülern als Selbsteinschätzung im Rahmen des Schülerfragebogens erfasst. Die Skala Gewissenhaftigkeit im Unterricht setzt sich aus 6 Items zusammen und weist eine gute interne Konsistenz auf ($\alpha = 0,82$). Beispielitem: „Im Unterricht bin ich in der Regel aufmerksam“. Anstrengungsbereitschaft im Unterricht wurde auch als Selbsteinschätzung erfasst. Die Skala bildet sich aus 4 Items, die insgesamt hinreichend intern konsistent sind ($\alpha = 0,75$). Beispielitem: „Ich kann sehr gut aufpassen, wenn ich etwas lernen will“.

3.2.3. ERGEBNISSE

Die im Kapitel 3.1 vorgestellten Analysen haben auch bei Kontrolle der Schülerleistung eine Assoziation zwischen sozialer Herkunft und Schulnoten nachgewiesen. Im Folgenden sollen spezifische Schülermerkmale betrachtet werden, die Einfluss auf die Benotung nehmen und so die beschriebenen Effekte medieren können. Als motivationale Merkmale werden Anstrengungsbereitschaft und Gewissenhaftigkeit der Schülerinnen und Schüler im Unterricht herangezogen. Beide Merkmale sind Selbstberichte der Schülerinnen und Schüler.

Betrachtet man Anteile aufgeklärter Varianz durch individuelle Schülermerkmale an der Durchschnittsnote,

so zeigt sich, dass die motivationalen Schülermerkmale mit 13,5 Prozent für Gewissenhaftigkeit im Unterricht und 8,1 Prozent für Anstrengungsbereitschaft substantielle Varianz an der Durchschnittsnote erklären. Beide Merkmale korrelieren mit $r = 0,67$. Dies bedeutet, dass beide Merkmale miteinander kovariieren, aber inhaltlich hinreichend verschieden sind.

Wenn die beschriebenen Effekte der sozialen Herkunft und des Geschlechts auf die Benotung durch die motivationalen Merkmale mediiert werden sollen, müssten sich diese Merkmale zwischen den verschiedenen sozialen Herkunftsgruppen unterscheiden. Diese bivariate Zusammenhangsstruktur zwischen den motivationalen Merkmalen und der sozialen Herkunft soll in einem ersten Schritt untersucht werden.

Für den sozialen Hintergrund soll aus Gründen der besseren Veranschaulichung der elterliche allgemeinbildende Schulabschluss herangezogen werden. Hier wird zwischen drei Abschlussniveaus unterschieden: Eltern, die als höchsten Schulabschluss einen Hauptschulabschluss haben, Eltern, die die Mittlere Reife abgelegt haben und Eltern mit Abitur. Die Ergebnisse sind in Abbildung 2 zusammenfassend dargestellt. Beide Merkmale weisen die gleiche Zusammenhangsstruktur auf. Kinder, deren Eltern maximal einen Hauptschulabschluss besitzen, weisen in beiden Skalen die niedrigsten Werte auf und Kinder, deren Eltern eine Hochschulzugangsberechtigung besitzen, die höchsten Werte. Die Unterschiede zwischen den drei Bildungsgruppen sind insgesamt nur gering.

(→ **ABBILDUNG 2**)

Tabelle 8: Korrelationen der motivationalen Merkmale mit Schulnoten

	Anstrengungsbereitschaft	Gewissenhaftigkeit im Unterricht
Durchschnittsnote	-0,38	-0,30
Deutschnote	-0,35	-0,26
Mathematiknote	-0,31	-0,28
Sachkundenote	-0,33	-0,24

Tabelle 9: Korrelationen der motivationalen Merkmale mit Ergebnissen aus standardisierten Leistungstests

	Anstrengungsbereitschaft	Gewissenhaftigkeit im Unterricht
Durchschnittsleistung	0,23	0,17
Mathematikleistung	0,19	0,15
Deutschleistung	0,21	0,16
Naturwissenschaftsleistung	0,17	0,12

Wie hängen die motivationalen Merkmale der Schülerinnen und Schüler mit den Schulnoten zusammen? Die Korrelationen beider Merkmale mit Schulnoten sind signifikant und weisen in die erwartete Richtung. In der Tendenz finden sich bessere Schulnoten, wenn Kinder gewissenhafter sind und eine höhere Anstrengungsbereitschaft aufweisen. Die Korrelationen mit der Durchschnittsnote liegen bei der Skala Gewissenhaftigkeit bei $r = -0,38$ und bei der Skala Anstrengungsbereitschaft bei $r = -0,30$. Für beide Konstrukte konnten also mittlere Korrelationen mit den Schulnoten identifiziert werden. Die für die Durchschnittsnote berechneten Korrelationen lassen sich in leicht abgeschwächter Form auch für die Einzelnoten in den Domänen Deutsch, Mathematik und Sachkunde nachweisen (vgl. Tabelle 8).

(→ TABELLE 8)

Neben den Schulnoten sind die Ergebnisse der standardisierten Leistungstests entscheidende Leistungsindikatoren. Die Korrelationen mit den Testleistungen weisen eine vergleichbare Struktur auf, wie sie bereits bei den Schulnoten berichtet wurde, nur auf geringerem Niveau. So beträgt die Korrelation zwischen der Durchschnittsleistung in den Domänen Deutsch, Mathematik und Naturwissenschaft und der Skala Gewissenhaft $r = 0,23$ und mit der Skala Anstrengungsbereitschaft $r = 0,17$.

(→ TABELLE 9)

In der multivariaten Analyse zeigt sich, dass Anstrengungsbereitschaft und Gewissenhaftigkeit im Unterricht zusammen ca. 14 Prozent der Variabilität der Schulnoten erklären (vgl. M1, Tabelle 10). Modell 2 stellt noch einmal

das bereits beschriebene Modell ohne Berücksichtigung der motivationalen Merkmale dar. Wenn die in diesem Modell erkennbaren sozialen Herkunftseffekte auf eine differenziell ausgeprägte Gewissenhaftigkeit und/oder Anstrengungsbereitschaft zurückzuführen sind, sollten sich die Effekte von M2 zu M3 spürbar verringern (partielle Mediation) oder nicht mehr das Signifikanzkriterium erreichen (vollständige Mediation). In Modell 3 ist zu erkennen, dass sich die Koeffizienten für den sozialen Hintergrund nur geringfügig verkleinern.

(→ TABELLE 10)

In den Tabellen 11, 12 und 13 sind die zentralen Befunde für die Fachnoten in den Fächern Mathematik, Deutsch und Sachkunde zusammengefasst. Die domänenspezifische Fachleistung erklärt in Mathematik knapp 39 Prozent der Variabilität der Mathematiknote und bindet mit knapp 22 Prozent in Sachkunde deutlich weniger Varianz. Bezogen auf den sozialen Hintergrund lassen sich keine domänenspezifischen Unterschiede erkennen.

(→ TABELLEN 11–13)

3.2.4. RESÜMEE

Für die These, dass die unterschiedlich strenge Benotung von Kindern aus verschiedenen sozialen Herkunftsgruppen auf sozialschichtspezifische Ausprägungen motivationaler Merkmale wie Gewissenhaftigkeit oder Anstrengungsbereitschaft zurückzuführen sind, konnten in dieser Teilstudie keine empirischen Belege gefunden werden. Zumindest die beiden von uns untersuchten Merkmale erklären also nicht die mit dem sozialen Hintergrund assoziierten Notendivergenzen.

Tabelle 10: Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum Migrationsstatus, zum sozio-kulturellen familiären Hintergrund, zentraler Leistungsindikatoren sowie motivationaler Merkmale

	M1	M2	M3
	<i>b</i>	<i>b</i>	<i>b</i>
Kombinierter Leistungswert ⁺		0,43	0,40
Kognitive Grundfähigkeit (KFT) ⁺		0,09	0,09
Geschlecht (<i>Ref. Mädchen</i>)			
Jungen		-0,10	-0,06
Sozioökonomischer Status (ISEI) ⁺		0,06	0,05
Ausbildung, schulisch (<i>Ref. Abitur</i>)			
Hauptschulabschluss		-0,29	-0,27
Mittlere Reife		-0,09	-0,09
Fachhochschulreife		-0,08	-0,08
Ausbildung, beruflich (<i>Ref. Uni-abschluss</i>)			
nicht universitär Abschlüsse		-0,01	0,01
Bücherbesitz (<i>Ref. 26-100</i>)			
weniger als 25		-0,08	-0,08
101 bis 250		0,03	0,03
251 bis 500		0,03	0,02
Mehr als 500		0,04	0,04
Kulturelle Aktivitäten Eltern ⁺		0,04	0,03
Migration (<i>Ref. maximal ein Elternteil in Deutschland geboren</i>)			
beide Eltern im Ausland geboren		0,00	-0,01
Gewissenhaftigkeit im Unterricht ⁺	0,25		0,13
Anstrengungsbereitschaft ⁺	0,06		0,05
<i>R</i> ²	13,90	55,40	59,40

Anmerkungen: ⁺ Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$;
 ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie);
b = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

Tabelle 11: Domänenspezifische Befunde für die Mathematiknote

	M1	M2	M3	M4	M5	M6	M7
	<i>b</i>						
Mathematikleistung ⁺	0,59	0,41	0,37	0,36	0,35	0,35	0,33
Deutschleistung ⁺		0,22	0,21	0,22	0,20	0,20	0,18
Naturwissenschaftsleistung ⁺		0,09	0,07	0,07	0,05	0,05	0,04
Kognitive Grundfähigkeit (KFT) ⁺⁺			0,12	0,13	0,12	0,12	0,12
Geschlecht (Ref. Mädchen)							
Jungen				0,11	0,11	0,11	0,15
Sozioökonomischer Status (ISEI) ⁺					0,12	0,12	0,11
Migration (Ref. Ref. maximal ein Elternteil in Deutschland geboren.)							
beide Eltern im Ausland geboren						0,05	0,04
Gewissenhaftigkeit im Unterricht ⁺							0,11
Anstrengungsbereitschaft ⁺							0,08
<i>R</i> ²	38,80	43,80	45,00	45,30	46,60	46,70	50,00

Anmerkungen: ⁺ Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; SES = sozioökonomischer Status; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

Tabelle 12: Domänenspezifische Befunde für die Deutschnote

	M1	M2	M3	M4	M5	M6	M7
	<i>b</i>						
Mathematikleistung ⁺		0,21	0,18	0,20	0,19	0,19	0,17
Deutschleistung ⁺	0,55	0,35	0,34	0,31	0,29	0,29	0,27
Naturwissenschaftsleistung ⁺		0,11	0,10	0,12	0,10	0,09	0,09
Kognitive Grundfähigkeit (KFT) ⁺			0,10	0,08	0,07	0,07	0,07
Geschlecht (Ref. Mädchen)							
Jungen				-0,28	-0,28	-0,28	-0,24
Sozioökonomischer Status (ISEI) ⁺					0,13	0,12	0,11
Migration (Ref. Ref. maximal ein Elternteil in Deutschland geboren.)							
beide Eltern im Ausland geboren						-0,06	-0,07
Gewissenhaftigkeit im Unterricht ⁺							0,14
Anstrengungsbereitschaft ⁺							0,03
<i>R</i> ²	35,60	42,30	43,20	45,60	47,30	47,40	50,40

Anmerkungen: ⁺ Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; SES = sozioökonomischer Status; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

Tabelle 13: Domänenspezifische Befunde für die Sachkundenote

	M1	M2	M3	M4	M5	M6	M7
	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Mathematikleistung ⁺		0,22	0,19	0,20	0,18	0,18	0,17
Deutschleistung ⁺		0,20	0,19	0,18	0,16	0,16	0,14
Naturwissenschaftsleistung ⁺	0,40	0,16	0,15	0,16	0,14	0,13	0,12
Kognitive Grundfähigkeit (KFT) ⁺			0,10	0,09	0,09	0,09	0,09
Geschlecht (<i>Ref. Mädchen</i>)							
Jungen				-0,11	-0,11	-0,11	-0,07
Sozioökonomischer Status (ISEI) ⁺							
Migration (<i>Ref. Ref. maximal ein Elternteil in Deutschland geboren.</i>)					0,13	0,13	0,11
beide Eltern im Ausland geboren						-0,12	-0,13
Gewissenhaftigkeit im Unterricht ⁺							0,15
Anstrengungsbereitschaft ⁺							0,03
<i>R</i>²	21,8	33,3	34,4	34,8	36,9	37,1	40,6

Anmerkungen: ⁺ Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; SES = sozioökonomischer Status; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; R^2 = Multipler Determinationskoeffizient.

3.3. Teilstudie 3: Quantifizierung primärer und sekundärer Herkunftseffekte der Benotung auf die Empfehlung und den Übergang auf das Gymnasium

Frage 3: Welchen Einfluss haben sozialschichtspezifische Benotungspraxen der Lehrkräfte in der Grundschule auf den Erhalt einer Gymnasialempfehlung und den Übergang auf ein Gymnasium? Wie viel Prozent des sozialen Herkunftseffektes auf die Empfehlung und den Übergang lassen sich auf sozialschichtspezifische Benotungspraxen der Lehrkräfte zurückführen?

3.3.1. HINTERGRUND

Für die Analyse sozialer Ungleichheiten im Bildungssystem hat sich das inzwischen sehr prominente theoretische Modell von Boudon aus den 1970er Jahren bewährt (vgl. Kapitel 2). Boudon unterscheidet zwischen primären und sekundären Effekten der sozialen Herkunft. Für beide Effekte gibt es in der jüngeren Bildungsforschung empirische Evidenz. Welcher Effekt – der primäre oder der sekundäre – ist jedoch stärker? Um die einzelnen Effekte berechnen und quantifizieren zu können, muss das Basismodell an das deutsche Bildungssystem angepasst und um Faktoren erweitert werden, die für den Übergang von der Grundschule in die weiterführenden Schulen besonders bedeutsam sind.

Maaz und Nagy (2009) schlugen in diesem Sinne eine Konzeptualisierung von Herkunftseffekten vor, die zwischen statusabhängigen Beurteilungen der Schülerinnen und Schüler und statusabhängigen Bildungsentscheidungen (Übergangsverhalten) unterscheidet. *Statusabhängige Schülerbeurteilungen* zeichnen sich durch systematische Differenzen in der Beurteilung von Schülerinnen und Schülern in Abhängigkeit des familiären sozioökonomischen Status aus, wie sie sich in den vergebenen Noten und Schullaufbahneempfehlungen widerspiegeln. Diese Einflüsse erfassen somit einen institutionellen Aspekt von Herkunftseffekten, da diese ohne das aktive Eingreifen der Eltern zustande kommen. *Statusabhängige Bildungsentscheidungen* hingegen beziehen sich auf die Einflüsse der Sozialschicht auf das tatsächliche Übergangsverhalten. Diese Effekte sind somit auf das Entscheidungsverhalten der Eltern zurückzuführen. Statusabhängige Beurteilungen und Bildungsentscheidungen dürfen nicht unabhängig voneinander betrachtet werden, da sich Schülerbeurteilungen in Form von Noten und Schullaufbahn-

empfehlungen direkt auf den tatsächlichen Übergang in die Sekundarstufe I auswirken können. Die bisherige Forschung hat viele Belege dafür geliefert, dass das tatsächliche Übergangsverhalten relativ eng mit der Schullaufbahneempfehlung und den Schulnoten verknüpft ist (u. a. Ditton, 2007).

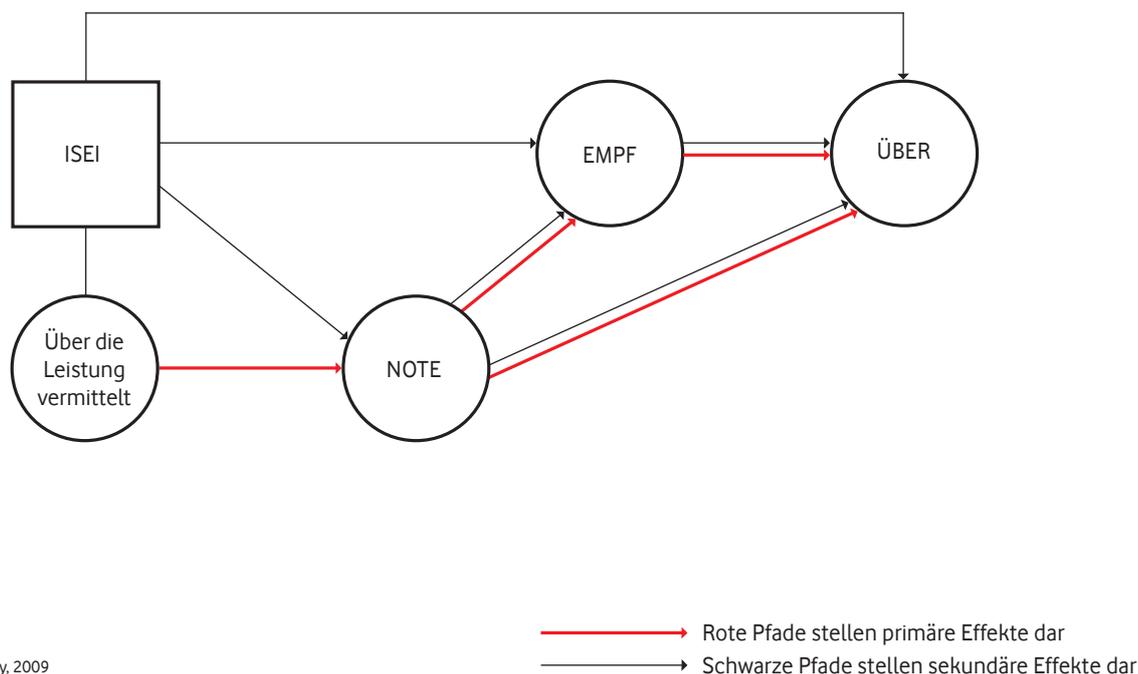
Maaz und Nagy (2009) haben diese Konkretisierung erstmals für den Übergang von der Grundschule in die Sekundarstufe I vorgenommen. Die Definition primärer Herkunftseffekte muss um Einflüsse der sozialen Herkunft, die sich auf die Kompetenzentwicklung der Schülerinnen und Schüler auswirken und infolgedessen in ihren Zensuren, Übergangsempfehlungen und Schulformwahlen niederschlagen, konkretisiert werden. Die sekundären Herkunftseffekte werden dagegen als diejenigen Einflüsse des sozialen Hintergrunds definiert, die losgelöst von der Schulleistung entstehen und zum Beispiel aus unterschiedlichen Bildungserwartungen und einem unterschiedlichen Entscheidungsverhalten Angehöriger verschiedener Sozialschichten resultieren. Sekundäre Effekte stehen besonders in der Kritik, da sie – anders als die primären Effekte, bei denen dies strittig ist – nicht mit Vorstellungen einer leistungsbezogenen Verteilungsgerechtigkeit vereinbar sind.

Die Wirkung primärer und sekundärer Herkunftseffekte ist in Abbildung 3 zusammenfassend dargestellt. Die durchgezogenen Linien stehen für primäre Effekte, das heißt Effekte der sozialen Herkunft auf die Leistung, die wiederum auf die Schulnoten, die Empfehlungsvergabe und den schließlich realisierten Übergang wirkt. Die gestrichelten Linien beziehen sich dagegen auf sekundäre Herkunftseffekte, die unabhängig von der Leistung auf die Noten- und Empfehlungsvergabe und die letztlich getroffene Übergangsentscheidung einwirken. In der Abbildung wird deutlich, dass primäre Herkunftseffekte nur indirekt (vermittelt über die Leistung) auf die Noten- und Empfehlungsvergabe sowie den Übergang wirken, während sekundäre sowohl indirekt als auch direkt Einfluss nehmen können.

(→ **ABBILDUNG 3**)

Die Abbildung 3 zur Wirkungsweise primärer und sekundärer Herkunftseffekte stellt aber auch die zentrale Bedeutung der Schülerbeurteilung in Form von Schulnoten dar. Lassen sich direkte Effekte der sozialen Herkunft auf die Notenvergabe nach Kontrolle der Leistungsmerkmale nachweisen, so wirken diese Effekte als

Abbildung 3: Darstellung der Wirkungsweise primärer und sekundärer Herkunftseffekte beim Übergang in die Sekundarstufe I



indirekte sekundäre Herkunftseffekte sowohl auf die Empfehlung als auch auf den eigentlichen Übergang. Im Folgenden soll untersucht werden, wie groß ein bereits bei der Benotung wirkender sekundärer Herkunftseffekt am Gesamtherkunftseffekt bei der Empfehlungsvergabe und beim Übergang ist. Hierfür kann auf die Arbeit von Maaz und Nagy (2009) zurückgegriffen werden, in der eine Quantifizierung der verschiedenen Herkunftseffekte vorgenommen wurde. Die Empfehlungsvergabe wird aus der sozialen Herkunft der Kinder, den Ergebnissen aus den Leistungstests sowie den Schulnoten vorhergesagt. Mit diesem sparsamen Modell können insgesamt 86 Prozent der Variabilität der Empfehlungsvergabe erklärt werden. Für die Analyse des Übergangs wird die Empfehlung mit in das Modell aufgenommen. Mit dem relativ sparsamen Modell können 74 Prozent der Varianz einer als latent kontinuierlich konzipierten dichotomen Übergangsvariable erklärt werden.

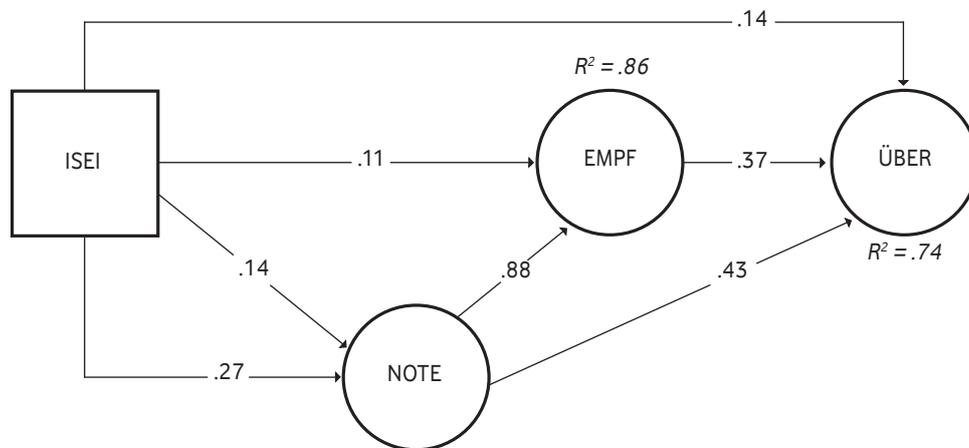
Fasst man den aktuellen Forschungsstand zusammen, liegen mittlerweile zahlreiche Arbeiten vor, in denen bedeutsame Herkunftseffekte beim Übergang am Ende der Grundschule nachgewiesen wurden. Die Ergebnisse zeigen zusammenfassend, dass Kinder aus sozial weniger begünstigten Familien im Vergleich zu Kindern aus sozial

privilegierten Elternhäusern (1) über niedrigere schulische Kompetenzen verfügen, (2) bei gleichen Leistungen in standardisierten Leistungstests von den Lehrkräften schlechter bewertet werden, (3) auch unter Kontrolle der Schulleistungen und Noten geringere Chancen auf den Erhalt einer Gymnasialempfehlung haben und (4) Eltern ihr Kind schließlich bei Kontrolle von Leistungen seltener auf ein Gymnasium schicken.

Folgt man den Befunden von Maaz und Nagy, so entsteht ein Großteil des am Übergang sichtbaren Herkunftseffekts nicht in der Übergangssituation, sondern ist viel früher angelegt. In dieser Teilstudie sollen diese Effekte quantifiziert werden. Wir gehen der Frage nach, wie groß die Anteile der einzelnen Herkunftseffekte am Gesamtherkunftseffekt auf die Empfehlungsvergabe und den Übergang auf ein Gymnasium ist, der auf Einflüsse der Notengebung zurückzuführen ist. Dabei kann die Unterscheidung zwischen primären und sekundären Herkunftseffekten berücksichtigt werden. Primäre Herkunftseffekte auf die Benotung nehmen ihren Ursprung in sozialschichtspezifischen Leistungsunterschieden und sekundäre Effekte äußern sich in herkunftsabhängigen Benotungspraxen der Lehrkräfte bei gleichen Leistungen der Schülerinnen und Schüler.

Abbildung 4:

Wirkungsweise primärer und sekundärer Herkunftseffekte beim Übergang in die Sekundarstufe I



3.3.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Die Daten der nachfolgenden Analysen entstammen dem am Berliner Max-Planck-Institut für Bildungsforschung durchgeführten Projekt „Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (ÜBERGANG) (Maaz, Baumert, Gresch & McElvany, 2010; Becker et al., 2010). Weitere Informationen zur Studienanlage finden sich im Anhang 1.

Die einzelnen Indikatoren des zu testenden Modells wurden entsprechend nationaler und internationaler Standards operationalisiert. Dabei entsprechen die Operationalisierungen denen in der Teilstudie 1. Zusätzlich wurde die Schullaufbahnempfehlung berücksichtigt.

Schullaufbahnempfehlung. Die Empfehlungen der abgebenden Grundschule basieren auf Angaben der Schulen und wurden zu einem Zeitpunkt erfasst, als die Empfehlung bereits ausgestellt wurde. Für die empirischen Analysen wurde eine Empfehlungsvariable definiert, die zwischen einer Gymnasialempfehlung und einer Empfehlung für eine andere Schulform bzw. für einen anderen Bildungsgang diskriminiert.

Die Spezifikation und Schätzung von Herkunftseffekten geschah auf Grundlage pfadanalytischer Verfahren. Zur Schätzung des Pfadmodells wurde das Statistikprogramm Mplus 5.2 (Muthén & Muthén, 1998-2010) herangezogen. Dieses Programm hat den Vorteil, dass es die Verwendung dichotomer abhängiger und vermittelnder Variablen erlaubt. Das spezifizierte Modell kombiniert lineare Regressionen (bei kontinuierlichen Outcomes) mit logistischen Regressionen (bei dichotomen Outcomes). Da-

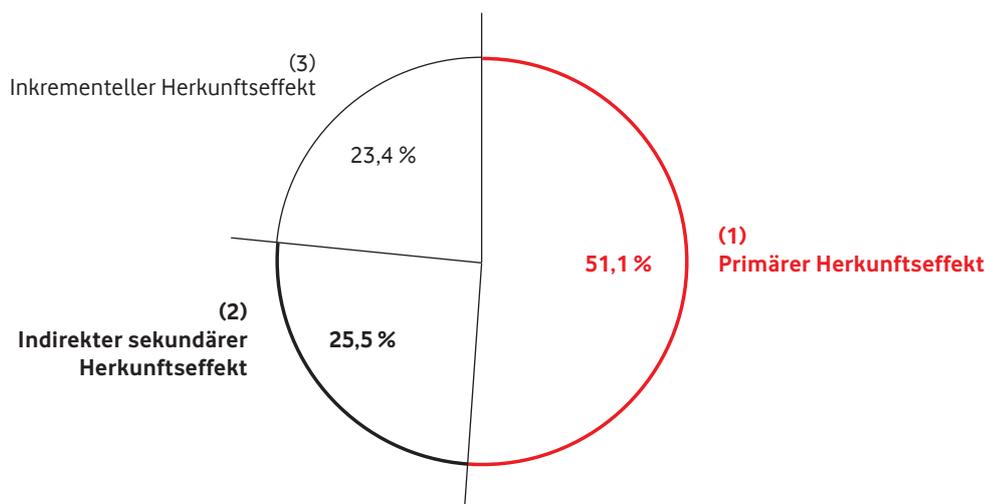
mit ist es möglich, die Parameter eines überidentifizierten simultanen Gleichungssystems analog zur linearen Pfadanalyse zu schätzen. Gleichwohl stellt die Verwendung von kategorialen Variablen besondere Herausforderungen an die Spezifikation der postulierten Wirkmechanismen und der Evaluation direkter und indirekter Effekte. Die Schätzung der Modellparameter fand auf Grundlage einer robusten Maximum Likelihood Schätzung (MLR) mittels eines integrativen Expectation-Maximization Algorithmus statt. Die Schätzung berücksichtigte die geclusterte Datenstruktur und wurde auf Grundlage aller vorhandenen Werte (Full Information MLR) durchgeführt (für weitere Informationen siehe Maaz & Nagy, 2009).

3.3.3. ERGEBNISSE

In Übereinstimmung mit bereits vorliegenden Forschungsbefunden zeigten die Analysen zunächst, dass die soziale Herkunft mit der Benotung, der Vergabe der Schullaufbahnempfehlungen und der schließlich getroffenen Übergangentscheidung assoziiert waren. Dabei fiel der absolute Einfluss der sozialen Stellung auf die Benotung am geringsten und auf die vollzogene Übergangentscheidung am höchsten aus. Für alle drei betrachteten Aspekte fanden sich sowohl primäre als auch sekundäre Effekte der sozialen Herkunft. Bei der Leistungsbewertung war der Anteil des primären Effekts größer als der des sekundären. Bei der Empfehlungsvergabe fielen beide Effekte etwa gleich groß aus. In Bezug auf das realisierte Übergangsverhalten überwog der sekundäre Effekt.

(→ ABBILDUNG 4)

Abbildung 5:
Effektzerlegung des sozialen Herkunftseffektes auf die Übergangsempfehlung



Im Folgenden soll der Einfluss der sozialen Herkunft auf die Benotung und deren Konsequenzen für den weiteren Übergangsprozess genauer betrachtet werden. Der Gesamteffekt der sozialen Herkunft soll hierfür seiner Entstehung nach in unterschiedliche Teileffekte zerlegt werden. Dabei werden für die Empfehlung drei Teileffekte unterschieden. (1) Der primäre Herkunftseffekt ist der Einfluss der familiären Herkunft auf die schulische Performanz der Kinder. (2) Der indirekte sekundäre Herkunftseffekt wird bei der Notenvergabe angelegt und wird über die Noten auf die Empfehlung weitergeleitet. Kinder aus unterschiedlichen Herkunftsfamilien, die die gleichen Ergebnisse im Leistungstest erzielten, werden unterschiedlich bewertet. (3) Ein inkrementeller Effekt der sozialen Herkunft auf die Empfehlungsvergabe bei Konstanzhaltung aller Leistungs- und Benotungsvariablen.

Die Ergebnisse der Effektzerlegung zeigen, dass der primäre Herkunftseffekt am größten ist. Mehr als die Hälfte des sozialen Herkunftseffektes sind auf die Empfehlungsvergabe auf sozialschichtspezifische Unterschiede in der schulischen Leistung der Kinder zurückzuführen. Diese Unterschiede können während der Schulzeit entstehen, sie können aber auch bereits zum Schuleintritt bestanden haben und wurden während der Grundschulzeit nicht abgebaut. Der Anteil des Herkunftseffektes, der auf sozialschichtabhängige Benotungspraxen zurückzuführen ist, beträgt ca. 25 Prozent. Und schließlich setzt sich der Herkunftseffekt auf die Empfehlung zu ca. 23 Prozent aus sozialschichtabhängigen Empfehlungen zusammen. Diese Befunde machen deutlich, dass der soziale Her-

kunftseffekt auf die Empfehlungsvergabe um die Hälfte reduziert werden könnte, wenn sozialschichtabhängige Benotungs- und Empfehlungspraxen ausgeschlossen werden könnten.

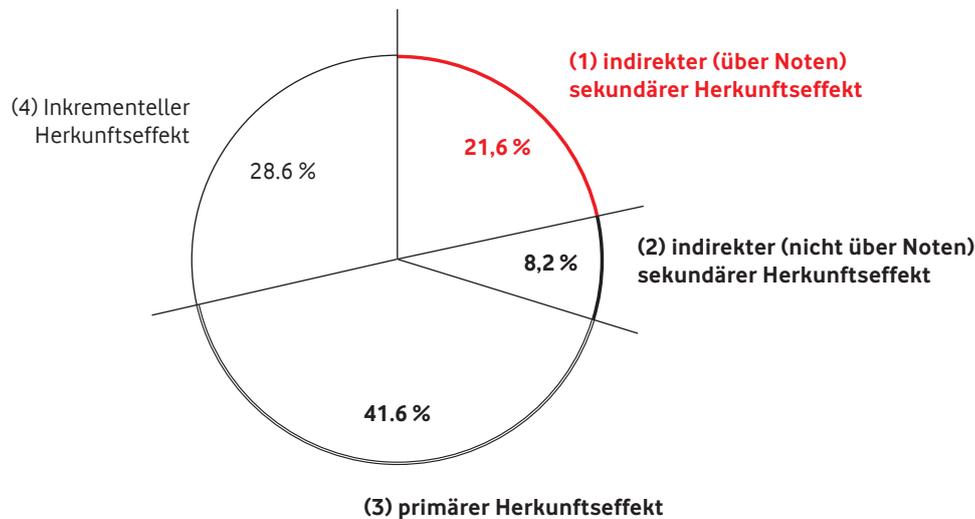
(→ ABBILDUNG 5)

Für die Zerlegung des sozialen Herkunftseffektes beim Übergang wurden vier Teileffekte identifiziert. (1) Die Einflüsse der sozialen Herkunft, die als indirekte sekundäre Herkunftseffekte auf die Empfehlung wirkten, werden auch bis zum Übergang weitertransportiert. (2) Der auf die Empfehlung als inkrementell bezeichneter Effekt schreibt sich ebenfalls fort und wirkt als ein weiterer sekundärer Herkunftseffekt auf den Übergang. (3) Wie bei der Empfehlung wird auch der primäre Effekt weitertransportiert, einmal über die Empfehlung und einmal an der Empfehlung vorbei. Dieser Effekt soll jedoch für die vorliegende Fragestellung nicht weiter differenziert betrachtet werden. (4) Schließlich gibt es einen inkrementellen – direkten sekundären – Effekt, der direkt auf das Entscheidungsverhalten einwirkt.

Der primäre Herkunftseffekt, der seinen Anfang in den sozialschichtabhängigen Kompetenzen hat, bindet mit ca. 42 Prozent den größten Teil des Gesamteffektes. Unter den sekundären Effekten hat der inkrementelle Effekt, also der Effekt, der direkt im Entscheidungsgeschehen zum Tragen kommt, den größten Anteil (ca. 29 %). 22 Prozent des gesamten Herkunftseffektes gehen auf sozialschichtabhängige Benotungen bei gleichen Kompetenzen zurück und ca. 8 Prozent resultieren aus einer sozial-

Abbildung 6:

Effektzerlegung des sozialen Herkunftseffektes auf den Übergang



schichtabhängigen Vergabepraxis der Empfehlungen bei Konstanzhaltung aller Leistungsindikatoren.

(→ **ABBILDUNG 6**)

3.3.4. RESÜMEE

Die Effektzerlegung verdeutlicht damit einen oft vernachlässigten, aber sehr bedeutsamen Befund. Der Anteil des sozialen Herkunftseffektes, der unmittelbar in der Übergangssituation zum Tragen kommt, macht nur ca. ein Viertel (28,6 %) des gesamten Herkunftseffektes aus. Dieser Befund ist daher so bedeutsam, weil er Ansatzpunkte für mögliche Interventionen geben kann. Würden alle Maßnahmen mit dem Ziel, den Herkunftseffekt an dieser entscheidenden Übergangsschwelle zu verringern, ausschließlich den Übergangsprozess fokussieren, blieben ca. drei Viertel des Gesamteffektes von den Maßnahmen unberührt, weil sie nicht in der Übergangssituation entstehen, sondern weit vorher angelegt sind. Auch reicht eine ausschließliche Thematisierung der sekundären Herkunftseffekte nicht aus, da auch bei Ausschalten des sekundären Herkunftseffektes noch ca. 40 Prozent bestehen bleiben. Ungefähr ein Drittel des gesamten Herkunftseffektes ist auf Benotungs- und Empfehlungsvergabe zurückzuführen, da bei Kontrolle der Leistung die Kinder in Abhängigkeit ihrer sozialen Herkunft unterschiedlich benotet werden und unterschiedliche Maßstäbe für den Erhalt einer Gymnasialempfehlung zum Tragen kommen.

3.4. Teilstudie 4:

Simulation der Neutralisation von primären und sekundären Effekten der Notengebung – Wie groß wäre der Anstieg der Gymnasialquote für Kinder aus Arbeiterfamilien?

Frage 4: Wie würde sich die Übergangssquote für Kinder aus sozial weniger begünstigten Familien auf das Gymnasium verändern, wenn primäre und/oder sekundäre Effekte, die auf die Notengebung wirken, ausgeschaltet wären?

3.4.1. HINTERGRUND

Soziale Disparitäten der Bildungsbeteiligung können als ein Ergebnis des Zusammenwirkens von primären und sekundären Effekten der Sozialschichtzugehörigkeit betrachtet werden, die einerseits über die gezeigte Schulleistung der Schülerinnen und Schüler überwiegend die Erfolgswahrscheinlichkeit der Bildungsinvestition bestimmen (primäre Effekte) und andererseits auf variierende Kosten- und Nutzenabwägungen (sekundäre Effekte) zwischen den Sozialschichten zurückzuführen sind.

Die Unterscheidung der Mechanismen, die zu Bildungsungleichheiten führen sowie ihre theoretische Spezifikation und empirische Identifikation besitzen neben einem akademischen Interesse auch eine unmittelbare Praxisrelevanz, weil Interventionen entwickelt werden

können, die je nach Gewichtung unterschiedlich ausfallen und unterschiedliche Wirkung zeigen können. In diesem Zusammenhang ist ein sich gerade entwickelnder Strang empirischer Studien zu erkennen, der sich mit den Auswirkungen von Kompensationsstrategien primärer und sekundärer Herkunftseffekte beschäftigt (R. Becker, 2009; Becker & Schubert, 2011; Müller-Benedict, 2007, 2008). In den entsprechenden Studien wird der Frage nachgegangen, wie sich die Kompensation von primären und sekundären Herkunftseffekten auf die Gymnasialbeteiligung in der Sekundarstufe I auswirkt. Nach einem von Boudon vorgeschlagenen Verfahren entspricht die Elimination des primären Herkunftseffektes einer hypothetischen Situation, in der für die Arbeiterklasse die gleiche Leistungsverteilung angenommen wird wie für die Dienstleistungsklasse, bei Beibehalten des sekundären Herkunftseffektes (der sich in der klassen- und leistungsspezifischen Übergangsquote widerspiegelt). Dem gegenüber entspricht einer Elimination des sekundären Herkunftseffektes eine hypothetische Situation, in der die klassenspezifischen Leistungsverteilungen erhalten bleiben und die leistungsbedingten Übergangsquoten der Arbeiterklasse denen der Dienstleistungsklasse angeglichen werden.

Müller-Benedict (2007) legte als einer der ersten eine Analyse zur Kompensation von primären und sekundären Herkunftseffekten vor. Für die empirische Analyse wurden Daten der PISA-2000-Studie herangezogen, die allerdings nur begrenzt geeignet sind, das Ausmaß primärer und sekundärer Herkunftseffekte korrekt zu spezifizieren (Baumert & Schümer, 2001; Maaz & Nagy, 2009). Die soziale Herkunft wurde über eine Recodierung des EGP-Klassenschemas durch eine Ober- und eine Unterschicht abgebildet. Die Grundwahrscheinlichkeit des Gymnasialbesuchs lag bei Kindern aus der Oberschicht bei 66 Prozent, bei Kindern aus der Unterschicht bei 30,3 Prozent (vgl. Müller-Benedict, 2007, S. 623). Der für die Identifikation der Herkunftseffekte notwendige Leistungsindikator wurde als Index aus der Leseleistung und Schulnoten abgebildet und in drei Kategorien (gute, mittlere und schlechte) eingeteilt. Der primäre Herkunftseffekt äußert sich in dem von Müller-Benedict verwendeten Vorgehen in der unterschiedlichen Zugehörigkeit zu einer leistungspotenten Gruppe. Betrachtet man die Verteilung der Kinder aus den beiden Herkunftsgruppen auf die unterschiedlichen Leistungsniveaus, zeigt sich der bereits gut dokumentierte Befund, wonach Kinder aus der Unterschicht ein deutlich geringeres Leistungspotential aufweisen als Kinder aus der Oberschicht. Während 44 Prozent der Oberschichtkinder der Leistungsgruppe gut angehören, sind es bei den Kin-

dern der Unterschicht 22 Prozent. Diese ungleiche Verteilung bezeichnet Müller-Benedict (S. 625) als eine „klare Demonstration des primären sozialen Effektes“. Für den Nachweis des sekundären Herkunftseffektes wurden die Anteile der Kinder aus der guten Leistungsgruppe berechnet, die ein Gymnasium besuchen. Während Kinder aus der Oberschicht aus dieser Leistungsgruppe zu 86 Prozent ein Gymnasium besuchen, sind es bei den Unterschichtkindern 63 Prozent. Nach einer Quantifizierung der Effekte durch die Berechnung von Übergangswahrscheinlichkeiten kommt Müller-Benedict zu dem Schluss, dass der primäre und der sekundäre Herkunftseffekt in den PISA-Daten vergleichbar groß ausfielen.

Durch die Manipulation der Leistungsverteilungen lässt sich die Änderung der oben berichteten Grundwahrscheinlichkeiten des Gymnasialbesuchs (66 % für Kinder aus der Oberschicht und 30 Prozent für Kinder aus der Unterschicht) für die beiden Herkunftsgruppen berechnen. Nach diesem Verfahren wird der primäre Herkunftseffekt durch das Einsetzen der Leistungsverteilung der Oberschicht für die Unterschicht unter Beibehaltung des schichtspezifischen Übertrittsverhaltens „simuliert“. Somit wird in der Unterschicht die gleiche Leistungsverteilung wie in der Oberschicht angenommen. Nach diesem Verfahren würden 41,6 Prozent der Kinder aus der Unterschicht das Gymnasium besuchen. Durch die Neutralisation des primären Herkunftseffektes könnten die Unterschichten ihren Anteil am Gymnasium um 11,3 Prozentpunkte steigern. Nach dem gleichen Verfahren kann die Wirkung des sekundären Herkunftseffektes „simuliert“ werden. In diesem Fall werden für die Unterschicht die gleichen Übergangswahrscheinlichkeiten wie bei der Oberschicht angenommen, bei Beibehaltung der herkunftsspezifischen Leistungsverteilung. Jetzt könnten 54,6 Prozent der Unterschichtkinder ein Gymnasium besuchen. Damit würde sich der Anteil der Unterschichtkinder durch eine Neutralisation des sekundären Herkunftseffektes um 24,3 Prozentpunkte erhöhen.

R. Becker (2009b) kommt in einer Replikation mit Daten des Konstanzer Panels „Bildungsverläufe in Arbeiterfamilien“ aus dem Jahr 1983 zu ähnlichen Ergebnissen. Im Unterschied zu Müller-Benedict kann R. Becker auf Daten der 5. Jahrgangsstufe zurückgreifen und damit Effekte der differentiellen Leistungsentwicklung sowie der Mobilitätsprozesse in der Sekundarstufe I ausblenden. Seine Ergebnisse stimmen weitgehend mit denen von Müller-Benedict überein. Eine Neutralisation des primären Herkunftseffektes führt zu einer Steigerung der Gymnasialbeteiligung der Arbeiterschicht um 15 Prozentpunkte und eine

Tabelle 14: Bedingte Wahrscheinlichkeiten eine gute Note zu bekommen und auf das Gymnasium zu wechseln, getrennt für Schülerinnen und Schüler der Dienstleistungs- und Arbeiterklasse für das Noten-Cutoff-Kriterium 2,33

		Klasse							
		Dienstleistungsklasse (N = 2653)							
Schulleistung		hoch				niedrig			
N		1660				993			
%		62,6				37,4			
Note		hoch		niedrig		hoch		niedrig	
N		1054		606		154		839	
%		63,5		36,5		15,5		84,4	
% Klasse		39,7		22,8		5,8		31,6	
Gymnasialübergang		Ja	nein	ja	nein	ja	nein	ja	nein
N		945	109	257	349	109	45	171	668
%		89,7	10,3	42,4	57,6	70,8	29,2	20,4	79,6
% Klasse		35,6	4,1	9,7	13,2	4,1	1,7	6,4	25,2

Neutralisation des sekundären Herkunftseffektes um 25 Prozentpunkte. Zu vergleichbaren Ergebnissen kommen auch Maaz, Schörder und Gresch (2010) mit den Daten der TIMSS-Übergangsstudie.

Die vorliegenden Befunde sollen in dieser Teilstudie um die Benotungskomponente erweitern werden. Es gilt die Frage zu beantworten, welche Effekte die Neutralisation des primären und des sekundären Noteneffekts auf die Gymnasialübergangsquote hat.

3.4.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Die Daten der nachfolgenden Analysen entstammen dem am Berliner Max-Planck-Institut für Bildungsforschung durchgeführten Projekt „Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (ÜBERGANG) (Maaz, Baumert, Gresch & McElvany, 2010; Becker, et al., 2010). Weitere Informationen zur Studienanlage finden sich im Anhang 1.

Für die Analyse wurde folgende Konstrukte operationalisiert.

Soziale Klasse. Die soziale Herkunft wurde über das EGP-Klassenschema erfasst. Um Zuordnungsprobleme zu vermeiden und einen möglichst klaren sekundären Herkunftseffekt zu erhalten, wurden nur die beiden höchsten und niedrigsten Klassen miteinander verglichen, während die beiden mittleren Klassen und die Gruppe der Nicht-Erwerbstätigen unberücksichtigt blieben. Konkret wurde für jeden Schüler und jede Schülerin zunächst die höchste soziale Klasse der Familie ermittelt, indem die Klassen des

Vaters und der Mutter miteinander verglichen und die jeweils höhere Klasse übernommen wurde. Dieser Wert wurde dann weiter zusammengefasst, indem die beiden Dienstleistungsklassen zu einer gemeinsamen Kategorie „Dienstleistungsklasse“ zusammengezogen und mit einer Kategorie „Arbeiterklasse“ (die sich aus den beiden Klassen der Facharbeiter und der ungelernten Arbeiter zusammensetzte) verglichen wurde.

Schulleistung. Die Schulleistung der Schülerinnen und Schüler wurde mittels standardisierter Leistungstests für die Bereiche Mathematik, Naturwissenschaft und Deutsch erfasst (für eine genaue Beschreibung der eingesetzten Instrumente vgl. M. Becker et al., 2010). Da differentielle Effekte der drei Domänen für die nachfolgenden Analysen nicht relevant sind (vgl. hierzu Maaz & Nagy, 2009), wurden die drei Testscores zu einem Gesamtscore zusammengefasst, der als ein Maß der allgemeinen, fachübergreifenden Schulleistung interpretiert werden kann. Dieser Gesamtscore war erwartungsgemäß hinreichend reliabel ($\alpha = 0,80$). Analog zu Müller-Benedict und Becker wurde der kontinuierliche Leistungsscore dichotomisiert. Dabei wurden Schülerinnen und Schüler mit einem Gesamtscore über oder gleich dem Stichprobenmittelwert der Gruppe mit „hoher“ Schulleistung zugeordnet. Schülerinnen und Schüler, deren Gesamtscore unter dem Durchschnitt lag, wurden hingegen der Gruppe mit „niedriger“ Schulleistung zugewiesen.

Gymnasialübergang. Die Informationen zum Übergang in die Sekundarstufe I entstammten einem Elternfragebogen, der am Ende der 4. Klasse eingesetzt wurde. Bei feh-

Arbeiterklasse (N=1073)

hoch				niedrig			
274				799			
25,5				74,5			
hoch		niedrig		hoch		niedrig	
111		163		36		763	
40,5		59,5		4,5		9,5	
10,3		15,2		3,4		71,1	
ja	nein	ja	nein	ja	nein	ja	nein
87	24	34	129	28	8	57	706
78,4	21,6	20,9	79,1	77,8	22,2	7,5	92,5
8,1	2,2	3,2	12,0	2,6	0,7	5,3	65,8

lenden Angaben wurde zudem auf den Schülerfragebogen zurückgegriffen.

Noten. Die Noten der Schülerinnen und Schüler in den Fächern Mathematik, Deutsch und Sachkunde basieren auf Angaben der Schule. Aus diesen drei Noten wurde eine Durchschnittsnote gebildet. Um ein Kriterium für „gute“ und „schlechte“ Noten zu erhalten, wurde der Durchschnittswert 2,33 als Cut-Off-Kriterium festgelegt. Der Wert korrespondiert mit dem Notendurchschnitt, der in Bayern für den Erhalt einer Gymnasialempfehlung erforderlich ist.

3.4.3. ERGEBNISSE

Zunächst zeigt sich der bekannte Befunde, wonach Kinder aus einer Dienstklassenfamilie (55,8%) häufiger auf ein Gymnasium wechseln als Kinder aus einer Arbeiterfamilie (19,2%). Die Größe der unterschiedlichen Gymnasialbeteiligungsquoten stimmt ziemlich gut mit den Ergebnissen anderer Schulleistungs- und Übergangsstudien überein (vgl. z. B. Baumert & Schümer, 2001; Ditton, 2007; Jonkmann, Neumann, Maaz & Gresch, 2010; Knigge & Leucht, 2010;) und ist auch international vergleichbar (vgl. z. B. Erikson, Goldthorpe, Jackson, Yaish & Cox, 2005). Darüber hinaus belegen die Ergebnisse eine deutliche Leistungsdifferenz zwischen den beiden sozialen Klassen: Während in der Dienstleistungsklasse 63 Prozent eine hohe Schulleistung haben, sind es in der Arbeiterklasse lediglich 26 Prozent. Die Größe dieses Unterschiedes stimmt ebenfalls mit dem nationalen und internationalen Befundmuster gut überein.

In Tabelle 14 wird das Übergangsverhalten in der TIMSS-Übergangsstudie in Abhängigkeit von der Schulleistung und der Schulnoten der Schülerinnen und Schüler dargestellt. Sie zeigt sowohl die absoluten Fallzahlen als auch die (bedingten) Wahrscheinlichkeiten der Dienstleistungs- und Arbeiterklasse, a) ein hohes oder ein niedriges Leistungsniveau zu haben, b) in Abhängigkeit von seiner Klasse und seinem Leistungsniveau einen guten oder schlechten Notendurchschnitt zu erzielen und c) in Abhängigkeit von seiner Herkunftsklasse, seinem Leistungsniveau und der Benotung auf das Gymnasium zu wechseln. Die Tabelle ist komplex, deswegen besprechen wir die Ergebnisse schrittweise.

Zunächst ist in Tabelle 14 die bedingte Wahrscheinlichkeit abgetragen, dass eine Schülerin oder ein Schüler aus der Dienstleistungs- beziehungsweise Arbeiterklasse ein hohes oder niedriges Leistungsniveau hat. Die Ergebnisse wiederholen das Befundmuster, das sich bereits bei der deskriptiven Betrachtung der Ergebnisse gezeigt hat: Die Wahrscheinlichkeit, dass ein Kind aus der Dienstleistungsklasse ein hohes Leistungsniveau erreicht, beträgt ca. 63 Prozent, während es für ein Kind aus der Arbeiterklasse lediglich ca. 26 Prozent sind. Die Differenz der beiden Werte $(63 - 26) = 37\%$ kann als primärer Herkunftseffekt interpretiert werden, da er sich direkt auf schichtspezifische Unterschiede im Leistungsniveau der Schülerinnen und Schüler bezieht.

(→ TABELLE 14)

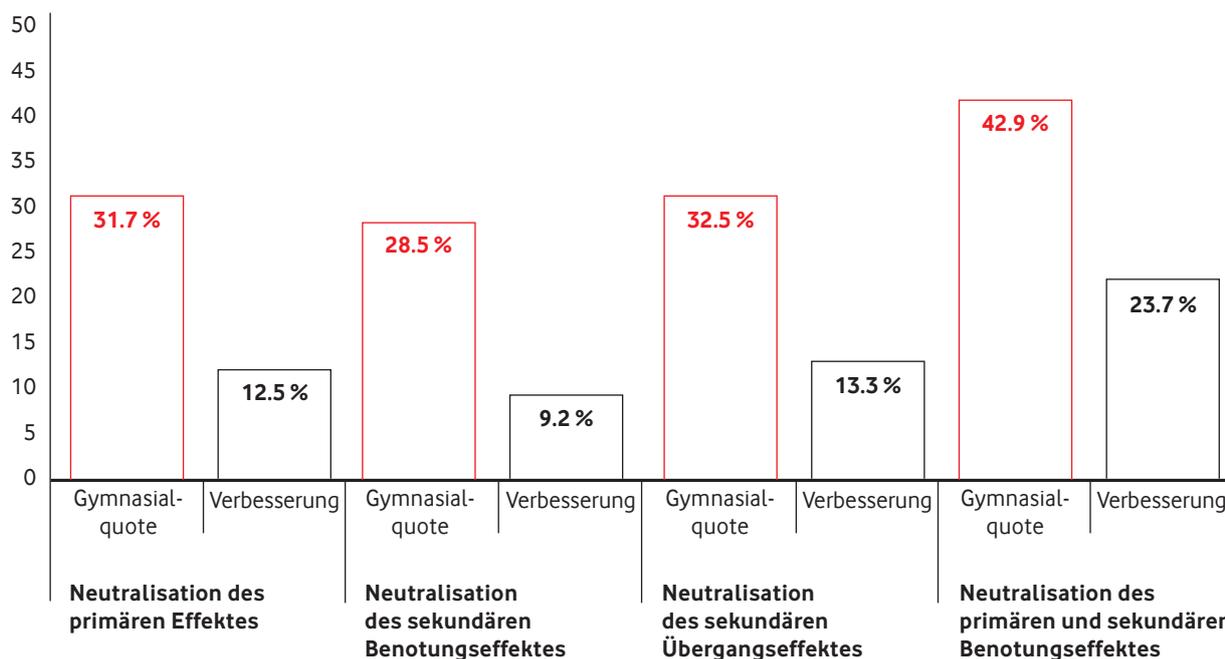
In der nächsten Zeile von Tabelle 14 finden sich die bedingten Wahrscheinlichkeiten, dass ein Schüler aus der Dienstleistungs- und Arbeiterklasse und einem hohen oder niedrigen Leistungsniveau einen guten Notendurchschnitt erzielt. Diese Trennung ist notwendig, um einen Indikator für die sekundären herkunftsbedingten Effekte bei der Schulbenotung bilden zu können. Denn da es sich um bedingte Wahrscheinlichkeiten handelt, können sie als schichtspezifische Benotungen unter Kontrolle des jeweiligen Leistungsniveaus der Schülerinnen und Schüler interpretiert werden. Unterschiede in diesen Wahrscheinlichkeiten stehen also für herkunftsbedingte Unterschiede in der Benotungspraxis, die unabhängig vom Leistungsniveau der Schülerinnen und Schüler existieren. Die Daten aus der TIMSS-Übergangsstudie zeigen, dass diese sekundären Herkunftseffekte substantiell sind. Prozentual ausgedrückt betragen die sekundären Herkunftseffekte 23 Prozent (63,5 % – 40,5 %) für die Schüler mit hoher Schulleistung und 11 Prozent (15,5% – 4,5 %) für die Schüler mit niedrigem Leistungsniveau. Auf den ersten Blick scheint also der primäre Effekt (37 %) stärker ausgeprägt zu sein als die sekundären. Allerdings ist darauf hinzuweisen, dass die Größe beider Effekte nicht unmittelbar miteinander vergleichbar ist, da sie sich auf unterschiedliche Sachverhalte beziehen: Während der primäre Herkunftseffekt die absoluten Unterschiede der beiden Klassen in ihrem Leistungsniveau angibt, bezieht sich der sekundäre Herkunftseffekt lediglich auf die klassenspezifischen Unterschiede in den bedingten Wahrscheinlichkeiten, bei einer bestimmten Leistung auf das Gymnasium zu wechseln.

Besonders die geringe Quote von Schülerinnen und Schülern aus der Arbeiterklasse mit hohem Leistungsniveau und weniger guter Note, unterstreicht die soziale Ungleichheit, die in Deutschland für die Benotungspraxis einzelner Klassen existiert. Dieses Gefälle ist in der Zeile „% Klasse“ von Tabelle 14 noch einmal zusammenfassend dargestellt. Die Werte geben für jede Klasse die Wahrscheinlichkeit an, eine der vier Kombinationsmöglichkeiten von Leistungsniveau und Note aufzuweisen (hoch/gute Leistungen, hoch/schlechte Leistungen, niedrig/gute Leistungen, niedrig/schlechte Leistungen). Die Werte summieren sich deshalb in jeder Klasse zu jeweils 100 Prozent auf. In der Dienstleistungsklasse ist die Wahrscheinlichkeit, ein hohes Leistungsniveau zu haben und gleichzeitig eine gute Note zu erhalten, vergleichsweise hoch ausgeprägt: 40 Prozent aller Schülerinnen und Schü-

ler der Dienstleistungsklasse fallen in diese Kategorie. In der Arbeiterklasse sind dies hingegen nur 11 Prozent. Die Gruppe der Schülerinnen und Schüler, die trotz hohem Leistungsniveau schlechtere Noten bekommen, ist in der Dienstleistungsklasse (23 %) größer als in der Arbeiterklasse (16 %). Die Gruppe der Schülerinnen und Schülern, die trotz niedriger Schulleistung gute Noten erzielen, ist in beiden Klassen nicht besonders groß. In der Dienstleistungsklasse sind dies ca. 6 Prozent aller Schülerinnen und Schüler, während es in der Arbeiterklasse 4 Prozent sind. Starke Unterschiede zwischen den beiden Klassen sind hingegen wieder bezüglich der Gruppe von Schülerinnen und Schülern festzustellen, die ein niedriges Leistungsniveau aufweisen und keine guten Noten bekommen: In der Dienstleistungsklasse sind dies 32 Prozent, während in der Arbeiterklasse fast drei Drittel aller Schülerinnen und Schüler in diese Kategorie fallen (71 %).

Beim Gymnasialübergang werden die zuvor erkennbaren Unterschiede zwischen den beiden Herkunftsgruppen noch einmal besonders deutlich. In der Dienstleistungsklasse ist die Gruppe mit guten Leistungen und guten Noten, die auch auf ein Gymnasium wechselt die größte Gruppe (36 %). In der Arbeiterklasse ist die mit 8 Prozent deutlich kleiner. Die Gruppe, die bei guten Leistungen und guten Noten nicht auf das Gymnasium wechselt, ist in der Dienstklasse mit 4 Prozent vergleichbar klein, während sie mit 12 Prozent in der Arbeiterklasse sogar größer ist, als diejenigen, die auf ein Gymnasium wechseln. Haben die Kinder gute Leistungen, aber schlechtere Noten, realisieren von den Kindern aus der Dienstklasse 10 Prozent den Gymnasialübergang, in der Arbeiterklasse nur ca. 3 Prozent. Diese Befunde lassen sich auch für Kinder, die der Gruppe mit den niedrigen Leistungen angehören übertragen. Die Gymnasialübergangsquote ist in der Dienstklasse höher als in der Arbeiterklasse: Für diejenigen mit guten Noten, beträgt sie 6 Prozent in der Dienstklasse und 3 Prozent in der Arbeiterklasse und für Kinder mit schlechteren Noten 6 Prozent in der Dienstklasse und 5 Prozent in der Arbeiterklasse. Starke Unterschiede zwischen den beiden Herkunftsklassen lassen sich wieder bezüglich der Gruppe von Schülerinnen und Schülern feststellen, die keine guten Leistungen und keine guten Noten haben und auch nicht auf das Gymnasium übergegangen sind. In der Dienstklasse ist diese Gruppe 25 Prozent groß und in der Arbeiterklasse 66 Prozent.

Um die Auswirkungen der Neutralisation von primären und sekundären Herkunftseffekten zu bestimmen, wird

Abbildung 7:**Hypothetische Gymnasialquote von Kindern aus der Arbeiterklasse und Differenz zur empirischen Quote**

in einem weiteren Schritt auf ein Boudon vorgeschlagenes Verfahren zurückgegriffen, das auch in den Analysen von Müller-Benedict, Becker und Maaz und Kollegen zur Anwendung kam. Wie bereits beschrieben, entspricht die Elimination des primären Herkunftseffektes einer hypothetischen Situation, in der die Arbeiterklasse die gleiche Leistungsverteilung hat wie die Dienstleistungsklasse. Die existierenden sekundären Herkunftseffekte, die sich in der klassen- und leistungsspezifischen Übergangsquote widerspiegeln, werden hingegen nicht verändert. Als Erweiterung zu den bereits vorliegenden Analysen kann bei gegebener Leistungsverteilung auch die Verteilung der Notengebung simuliert werden, in dem in der Arbeiterklasse die gleiche Notenverteilung wie in der Dienstklasse angenommen wird. Theoretisch wird damit der sekundäre Herkunftseffekt auf die Notengebung eliminiert. Schließlich entspricht einer Elimination des sekundären Herkunftseffektes des Übergangsverhaltens eine hypothetische Situation, in der die klassenspezifischen Leistungsverteilungen und/oder Notenverteilungen erhalten bleiben, die leistungsbedingten Übergangsquoten der Arbeiterklasse aber denen der Dienstleistungsklasse angeglichen werden. Durch die getrennte Manipulation beider Effekte lässt sich abschätzen, welche Maßnahmen zur Reduzierung sozialer Ungleichheit am wirkungsvollsten sind. Dies setzt allerdings ein Kriterium voraus, mit dem der jeweilige Erfolg sinnvoll beurteilt werden kann. Ein intuitiv einleuchtender Indikator ist dabei die Gymnasialquote der Arbeiter-

klasse, das heißt, der Anteil von Schülerinnen und Schülern aus dieser Klasse, die das Gymnasium besuchen. Zur Erinnerung: In den vorliegenden Daten liegt dieser Wert bei 19,2 Prozent – jeder Prozentpunkt, der hiervon nach oben abweicht, trägt zur Reduktion der Bildungsbenachteiligung der Arbeiterklasse bei.

Folgt man dieser Logik und setzt auf der Grundlage der vorliegenden Daten für die Arbeiterklasse die gleiche Leistungswahrscheinlichkeiten an wie für die Dienstleistungsklasse, dann ergeben sich folgende Werte: Die Gymnasialquote der Arbeiterklasse setzt sich zusammen aus der Gruppe der Schülerinnen und Schüler, die mit hohem Leistungsniveau das Gymnasium besuchen, und der Gruppe mit niedriger Schulleistung, die trotzdem auf das Gymnasium geht. Die Gymnasialquote der Arbeiterklasse 31,7 Prozent. Im Vergleich mit der realen Gymnasialquote von 19,2 Prozent erbringt die Elimination des primären Herkunftseffektes also eine Verbesserung von 12,5 Prozentpunkten.

Die Elimination des sekundären Herkunftseffektes auf die Notengebung lässt sich nach dem gleichen Vorgehen simulieren. Hierfür werden die herkunftsspezifischen Leistungsverteilungen und das Übergangsverhalten beibehalten, aber die Notenverteilung aus der Dienstklasse übernommen. Unter diesen Rahmenbedingungen würde die Gymnasialquote der Arbeiterklasse 28,5 Prozent betragen, was einer Verbesserung von 9,2 Prozentpunkten entspricht.

Des Weiteren lässt sich ausrechnen, wie sehr sich die Gymnasialquote der Arbeiterklasse erhöhen würde, wenn sich der sekundäre Herkunftseffekt des Entscheidungsverhaltens beseitigen ließe. Hierfür werden die Leistungsverteilung und die Notenverteilung der Arbeiterklasse beibehalten, aber die Übergangsquoten der Dienstleistungsklasse übernommen. Entsprechend des obigen Vorgehens beträgt die Gymnasialquote für diese hypothetische Situation 32,5 Prozent. Die Elimination der sekundären Herkunftseffekte erhöht die Gymnasialquote im Vergleich zur realen um 13,3 Prozentpunkte und um 0,8 Prozentpunkte mehr als die Neutralisation des primären Herkunftseffektes.

Primäre Herkunftseffekte und sekundäre Benotungseffekte lassen sich auch parallel ausschalten. Hierfür wurden die Leistungsverteilung und die Notenverteilung aus der Dienstklasse übernommen, aber die Übergangsquoten der Arbeiterklasse beibehalten. Entsprechend dieser hypothetischen Situation würde die Gymnasialquote in der Arbeiterklasse bei 42,9 Prozent liegen. Dies entspricht einer Verbesserung von 23,7 Prozent.

(→ **ABBILDUNG 7**)

3.4.4. RESÜMEE

Die vorgestellten Analysen konnten sehr deutlich zeigen, dass sich eine substanzielle Vergrößerung der Gymnasialquote nicht durch die ausschließliche Fokussierung auf primäre oder sekundäre Effekte erreichen lässt, vielmehr müssten primäre und sekundäre Disparitäten gleichzeitig minimiert werden. Entsprechende Bemühungen sollten sich auf eine Verbesserung der Leistung sowie eine „objektive“ Leistungsbewertung konzentrieren. Hier lassen sich Ansätze entwickeln, die auch in die Breite getragen werden können. Das elterliche Entscheidungsverhalten verspricht möglicherweise die größten Effekte, allerdings ist das Elternverhalten nur begrenzt „manipulierbar“. Hinzu kommt hier der rechtliche Rahmen, der den Einfluss des Staates bei der Wahl eines Bildungsganges oder einer Schulform regelt. Das sogenannte Recht der „positiven“ Auslese steht ausschließlich den Eltern zu. Damit ist gemeint, dass sich Eltern zum Beispiel entgegen einer Empfehlung für einen anspruchsvolleren Bildungsgang für ihr Kind entscheiden können. In die Entscheidung der Eltern darf der Staat korrigierend, in der Form der „negativen“ Auslese, eingreifen, wenn eine mangelnde Eignung des Kindes für die gewählte Schulform festgestellt wird. Die „positive“ Auslese durch den Staat, also

die Realisierung eines anspruchsvolleren Bildungsgangs (z.B. Gymnasium), wenn Eltern für ihr Kind einen weniger anspruchsvollen Bildungsgang (z.B. Realschule) wählen, ist unzulässig. Dabei wird auch in Kauf genommen, dass Eltern eine aus Sicht der Schule im Einzelfall „falsche“ Entscheidung treffen können (vgl. Füssel, Gresch, Baumert & Maaz, 2010).

3.5. Teilstudie 5: Konsequenzen der Simulation von primären und sekundären Effekten der Notengebung für den Übergang auf das Gymnasium

Frage 5: Worauf ist der in der Teilstudie 4 dokumentierte Anstieg der Gymnasialquote zurückzuführen? Wer „gewinnt“ und wer „verliert“?

3.5.1. HINTERGRUND

In der Studie von Maaz, Schröder und Gresch (2010) wurden analog zu Müller-Benedict und Becker die Veränderung der Gymnasialquote für Kinder aus Arbeiterfamilien berechnet. Zusätzlich untersuchten sie die mit der Neutralisation der primären und sekundären Effekte verbundenen Konsequenzen. Bezogen auf die Steigerung der Gymnasialquote der Kinder aus Arbeiterfamilien berichten sie in ihren Analysen vergleichbare Werte wie Müller-Benedict und Becker. Bezogen auf die Konsequenzen der Neutralisation der primären und sekundären Effekte zeigte sich, dass eine alleinige Neutralisation des primären Herkunftseffektes vor allem zu einer extremen Erhöhung (um 20,3 Prozentpunkte) der Kategorie der Schülerinnen und Schüler führt, die trotz guter Leistungen nicht das Gymnasium besuchen. Die Neutralisation des sekundären Herkunftseffektes bewirkt unter anderem, dass 2,1 Prozent der Schülerinnen und Schüler, deren Leistungsniveau eigentlich zu gering für den Gymnasialbesuch ist, ebenfalls in das Gymnasium übergehen. Diese Ergebnisse zeigen, dass Bemühungen, allein den primären oder den sekundären Herkunftseffekt zu minimieren, wahrscheinlich äußerst ineffizient bleiben.

Dieser Befund soll für diese Teilstudie aufgegriffen werden, indem die Analysen von Maaz, Schröder und Gresch um die Notenkomponente erweitert werden. In dieser Teilstudie wird untersucht, welche Konsequenzen mit einem Anstieg der Gymnasialquote verbunden sind.

Tabelle 15: Neutralisation des primären Effektes (der Leistungsverteilung), Note und Entscheidung bleiben herkunftsspezifisch

Arbeiterklasse (N = 1073)								
Schulleistung	hoch				niedrig			
N	672				401			
%	62,6				37,4			
Note	hoch		niedrig		hoch		niedrig	
N	272		400		18		383	
%	40,5		59,5		4,5		95,5	
% Klasse	25,3		37,3		1,7		35,7	
Gymnasialübergang	Ja	nein	ja	nein	ja	nein	ja	nein
N	213	59	84	316	14	4	29	354
%	78,4	21,6	20,9	79,1	77,8	22,2	7,5	92,5
% Klasse _{hyp}	19,9	5,5	7,8	29,5	1,3	0,4	2,7	33,0
% Klasse _{emp}	8,1	2,2	3,2	12,0	2,6	0,7	5,3	65,8
Δ % Klasse	+11,8	+3,3	+4,6	+17,5	-1,3	-0,3	-2,6	-32,8

3.5.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Die Daten der nachfolgenden Analysen entstammen dem am Berliner Max-Planck-Institut für Bildungsforschung durchgeführten Projekt „Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (ÜBERGANG) (Maaz, Baumert, Gresch & McElvany, 2010; Becker, et al., 2010). Weitere Informationen zur Studienanlage finden sich im Anhang 1.

Für die empirische Analyse wurden die gleichen Konstrukte operationalisiert, wie sie in Teilstudie 4 verwendet wurden.

3.5.3. ERGEBNISSE

In dieser Teilstudie soll genauer zu betrachten, wie das in der Teilstudie 4 berichtete Befundmuster einer höheren Gymnasialbeteiligung der Arbeiterkinder zustande kommt. Hierfür sei auf die Tabellen 15 bis 18 verwiesen, welche die Auswirkung der Neutralisation des primären und sekundären Herkunftseffektes auf die verschiedenen Leistungsgruppen in der Arbeiterklasse direkt angeben. Die Werte können direkt aus Tabelle 14 ermittelt werden. In jedem dieser Abschnitte finden sich wieder die Anzahl der Fälle und der Prozentsatz für jede der vier Kombinationsmöglichkeiten von Leistungsniveau und Schulbesuch in der Arbeiterklasse. Zusätzlich ist für jede dieser Kategorien angegeben, inwieweit sich durch die Neutralisation des entsprechenden Effektes ihre Besetzung verändern würde (Δ%).

(→ TABELLE 15)

Betrachtet werden soll als nächstes, welche Auswirkung die Elimination des primären Herkunftseffektes auf die verschiedenen Kategorien des Gymnasialübergangs hat (vgl. Tabelle 15). Hier ist zunächst festzustellen, dass eine Reduktion des primären Herkunftseffektes die in der Arbeiterklasse extrem hoch besetzte Kategorie der Schülerinnen und Schüler, die wegen ihres niedrigen Leistungsniveaus und schlechteren Noten nicht auf das Gymnasium übergehen, fast halbiert (auf 33,0 %). Gleichzeitig werden die Kategorien der Schülerinnen und Schüler, die trotz niedriger Schulleistung mit guten Noten (-1,3 Prozentpunkte) bzw. mit schlechten Noten (-2,6 Prozentpunkte) auf das Gymnasium übergehen, leicht reduziert. Einen deutlichen Zuwachs um 11,8 Prozentpunkte kann für die Kinder verbucht werden, die aus der Arbeiterklasse mit guten Leistungen und guten Noten kommen. Kinder mit guten Leistungen und schlechten Noten weisen ebenfalls eine leicht höhere Gymnasialquote auf (+4,6 Prozentpunkte).

Erschreckend ist jedoch, dass eine alleinige Neutralisation des primären Herkunftseffektes vor allem zu einer starken Erhöhung (um 20,8 Prozentpunkte) der Kategorie der Schülerinnen und Schüler führt, die trotz guter Leistungen nicht das Gymnasium besuchen: 3,3 Prozent mit guten Noten und 17,5 Prozent mit schlechten Noten (vgl. Tabelle 15).

Eine Neutralisation des sekundären Effektes auf die Benotung führt zu einem Anstieg der Gymnasialquote von Kindern mit guten Leistungen und guten Noten (um 4,6 Prozentpunkte) (vgl. Tabelle 16). Dieser Anstieg ist jedoch an eine Verringerung der Gymnasialquote gekoppelt,

Tabelle 16: Neutralisation des sekundären Noteneffektes (der Benotungspraxis), Leistung und Entscheidung bleiben herkunftsspezifisch

Arbeiterklasse (N= 1073)								
Schulleistung	hoch				niedrig			
N	274				799			
%	25,5				74,5			
Note	hoch		niedrig		hoch		niedrig	
N	174		100		125		674	
%	63,5		36,5		15,5		84,4	
% Klasse	16,2		9,3		11,6		62,3	
Gymnasialübergang	Ja	nein	ja	nein	ja	nein	ja	nein
N	136	38	21	79	97	28	51	623
%	78,4	21,6	20,9	79,1	77,8	22,2	7,5	92,5
% Klasse _{hyp}	12,7	3,5	2,0	7,4	9,0	2,6	4,8	58,1
% Klasse _{emp}	8,1	2,2	3,2	12,0	2,6	0,7	5,3	65,8
Δ % Klasse	+4,6	+1,3	-1,2	-4,6	+6,4	+1,9	-0,5	-7,7

wenn bei guten Leistungen schlechtere Noten vorliegen (-4,6 Prozentpunkte). Liegen gute Leistungen aber schlechte Noten vor, erhöht sich die Gymnasialquote um 1,3 Prozentpunkte und sie verringert sich leicht für diejenigen, die trotz guter Leistungen keine guten Noten bekommen haben (-1,2 Prozentpunkte). Für Schülerinnen und Schüler mit niedrigen Leistungen kann eine um 6,4 Prozentpunkte höhere Gymnasialquote erwartet werden. Dieser Vorteil wird aber real um 1,9 Prozentpunkte reduziert, da sich diese Gruppe, die nicht auf das Gymnasium übergeht, ebenfalls vergrößert. Die Verkleinerung der Kategorie, die mit niedrigen Leistungen und schlechteren Noten nicht auf da Gymnasium übergeht, wird im Vergleich zur Eliminierung primärer Effekte nur um 7,7 Prozentpunkte reduziert.

(→ TABELLE 16)

In der Tabelle 17 werden die Ergebnisse für die Eliminierung des sekundären Herkunftseffektes dargestellt. Hier zeigt sich zunächst, dass diese Maßnahme in der Tat dazu führt, dass 4,4 Prozent der Schülerinnen und Schüler mit guten Leistungen zusätzlich das Gymnasium besuchen würden: 1,2 Prozent mit guten Noten und 3,2 Prozent mit schlechteren Noten. Die Kehrseite der Medaille ist allerdings, dass gleichzeitig auch 9,2 Prozent der Schülerinnen und Schüler, deren Leistungsniveau eigentlich zu gering für den Gymnasialbesuch ist und deren Noten auch weniger gut sind, ebenfalls in das Gymnasium wechseln würden. Daran wird auch deutlich, dass die vermeintlich leicht höhere Gymnasialquote durch die Neutralisation des

sekundären Herkunftseffektes vor allem dadurch zustande kommt, dass nun mehr Kinder aus der Arbeiterklasse auf das Gymnasium geschickt werden – unabhängig davon, ob sie die notwendigen kognitiven Voraussetzungen entwickeln konnten. Ohne diesen Anteil ist die Effektivität der Neutralisation der sekundären Herkunftseffekte hingegen eher gering. Auch hier zeigt sich also, dass eine einseitige Konzentration auf die Neutralisation des sekundären Herkunftseffektes allein kaum hilfreich sein dürfte, ohne gleichzeitig auch das schulische Leistungsniveau der Arbeiterklasse zu fördern.

(→ TABELLE 17)

Abschließend veranschaulicht Tabelle 18, welche Konsequenzen eine Elimination des primären Herkunftseffektes und des sekundären Benotungseffektes hat. Der Anstieg der Gymnasialquote in der Arbeiterklasse um knapp 24 Prozent basiert im Wesentlichen durch eine deutliche Reduktion der Kategorie niedrige Leistungen, schlechte Noten und kein Gymnasialübergang. Diese Gruppe konnte um 36,6 Prozent verringert werden. Einen deutlichen Anstieg würde die Gruppe derjenigen, die mit guten Leistungen und guten Noten auf das Gymnasium übergehen, verzeichnen (um 23,1 Prozentpunkte). Gleichzeitig zeigt sich aber auch ein unerwünschter Nebeneffekt, denn die Gruppe der leistungsstarken Schülerinnen und Schüler mit guten Noten, die nicht auf das Gymnasium übergehen, würde sich auch um 6,4 Prozent vergrößern. Ebenfalls unerfreulich wäre der Anstieg um 6,1 Prozent der Kategorie gute Leistungen, aber schlechte Noten und kein Gym-

Tabelle 17: Neutralisation des sekundären Übergangseffektes (des Entscheidungsverhaltens), Leistung und Note bleiben herkunftsspezifisch

		Arbeiterklasse (N = 1073)							
Schulleistung	hoch				niedrig				
	N	274				799			
%	25,5				74,5				
Note	hoch		niedrig		hoch		niedrig		
	N	111	163	36	763	%	40,5	59,5	4,5
% Klasse	10,3	15,2	3,4	71,1					
Gymnasialübergang	Ja	nein	ja	nein	ja	nein	ja	nein	
	N	100	11	69	94	25	11	156	607
%	89,7	10,3	42,4	57,6	70,8	29,2	20,4	79,6	
% Klasse _{hyp}	9,3	1,0	6,4	8,8	2,3	1,0	14,5	56,6	
% Klasse _{emp}	8,1	2,2	3,2	12,0	2,6	0,7	5,3	65,8	
Δ % Klasse	+1,2	-1,2	+3,2	-3,2	-0,3	+0,3	+9,2	-9,2	

nasialübergang. Knapp drei Prozent der Arbeiterkinder würden bei niedrigen Leistungen und schlechten Noten zusätzlich auf das Gymnasium wechseln (2,9 Prozent).

(→ TABELLE 18)

3.5.4. RESÜMEE

Die Analysen dieser Teilstudie haben deutlich zeigen können, dass sich durch die Eliminierung der unterschiedlichen Herkunftseffekte zwar die Gymnasialquote erhöhen würde, dies aber mit unerwünschten Nebeneffekten verbunden sein kann. Ein solcher Effekt wäre zum Beispiel eine Vergrößerung der Gruppe, die die notwendigen kognitiven Voraussetzungen für den Gymnasialbesuch nicht erreicht, aufgrund der vorgenommenen Maßnahmen nun aber das Gymnasium besucht. Die Befunde zu den Konsequenzen machen schließlich auf einen weiteren wichtigen Punkt aufmerksam.

Maßnahmen zur Reduzierung sozialer Ungleichheitseffekte beim Übergang in die Sekundarstufe I lassen sich nicht global für alle Schülerinnen und Schüler implementieren. Hier geht es vielmehr darum, adressatengerecht zu intervenieren. Auf diese Weise erhöht sich die Chance, die unerwünschten Nebeneffekte, die in dieser Studie beschrieben wurden, zu minimieren.

Tabelle 18: Neutralisation des primären Effektes (Leistung) und des sekundären Benotungseffektes bei Beibehaltung und des sekundären Entscheidungseffekte

Arbeiterklasse (N= 1073)								
Schulleistung	hoch				niedrig			
N	672				401			
%	62,6				37,4			
Note	hoch		niedrig		hoch		niedrig	
N	427		245		62		338	
%	63,5		36,5		15,5		84,4	
% Klasse	39,8		22,9		5,8		31,5	
Gymnasialübergang	Ja	nein	ja	nein	ja	nein	ja	nein
N	335	92	51	194	48	14	25	313
%	78,4	21,6	20,9	79,1	77,8	22,2	7,5	92,5
% Klasse _{hyp}	31,2	8,6	4,8	18,1	4,5	1,3	2,4	29,2
% Klasse _{emp}	8,1	2,2	3,2	12	2,6	0,7	5,3	65,8
Δ % Klasse	+23,1	+6,4	+1,6	+6,1	+1,9	+0,6	-2,9	-36,6

3.6. Teilstudie 6: Bewertungsunterschiede am Ende der gymnasialen Oberstufe

Frage 6: In welchem Maß wird die Notenvergabe am Ende der gymnasialen Oberstufe neben leistungsbezogenen Merkmalen auch durch Merkmale der familiären Herkunft beeinflusst? Inwieweit lassen sich diesbezüglich Unterschiede bei der Leistungsbewertung im Fachunterricht (Fachnoten) und in der schriftlichen Zentralabiturprüfung (Prüfungsnoten) feststellen?

3.6.1. HINTERGRUND

Während zur Frage, in welchem Maß in die Leistungsbewertung neben leistungs- und lernbezogenen Kriterien (z.B. Kompetenzniveau, kognitive Grundfähigkeiten, Mitarbeit und Anstrengung im Unterricht) auch leistungs-fremde Faktoren der familiären Herkunft einfließen, für den Bereich der Grundschule inzwischen eine Reihe von Untersuchungen und empirischen Befunden vorliegen (siehe Teilstudien 1 bis 5), finden sich im deutschsprachigen Raum kaum Forschungsarbeiten, die sich dieser Thematik für die Mittelstufe und die gymnasiale Oberstufe gewidmet haben. Dabei ist die Rolle der in den Abschlusszeugnissen für den mittleren Schulabschluss und das Abitur enthaltenen Noten aufgrund ihrer Selektionsfunktion für den Zugang zu stark nachgefragten Berufs- und Ausbildungsplätzen kaum zu überschätzen.

Zwar sind verschiedene Arbeiten der Frage nachgegangen, inwieweit in der Mittelstufe und in der Sekundarstufe

II mit Bewertungsunterschieden zwischen Bundesländern (vgl. Baumert et al., 2003a; Baumert & Watermann, 2000; Neumann, Nagy, Trautwein & Lüdtke, 2009) und nominal gleichwertigen Schulformen (z.B. allgemein bildendes Gymnasium, berufliches Gymnasium, Oberstufe an Gesamtschule, vgl. Köller, Baumert & Schnabel, 1999; Watermann, Nagy & Köller, 2004) zu rechnen ist. In welchem Maß neben dem Kompetenzniveau die soziale Herkunft oder der Migrationsstatus der Schülerinnen und Schüler einen Einfluss auf die Notenvergabe ausübt, wurde dabei jedoch bislang kaum untersucht (vgl. aber z.B. Hochweber, 2010). Aus diesem Grund werden im Folgenden Analysen präsentiert, die Bewertungsunterschieden am Ende der gymnasialen Oberstufe nachgehen und den Schwerpunkt auf die Rolle familiärer Hintergrundmerkmale legen. Eine weitere Besonderheit der Analysen besteht darin, dass neben den im Rahmen des regulären Fachunterrichts vergebenen Noten auch die Prüfungsnoten aus dem schriftlichen Zentralabitur betrachtet werden.

Im Zentrum der nachfolgenden Analysen steht die Frage, in welchem Maß am Ende der gymnasialen Oberstufe mit Bewertungsunterschieden zwischen Schülerinnen und Schülern unterschiedlicher familiärer Herkunft zu rechnen ist. Konkret soll untersucht werden, inwieweit sich bei gleicher individueller Schülerleistung (gemessen über standardisierte Leistungstests in den Fächern Mathematik und Englisch) Effekte der sozialen Herkunft (Sozioökonomischer Status und Bildungshintergrund der Eltern sowie kulturelles Kapital) und des Migrationsstatus auf die Leistungsbewertung finden. Dabei soll auch die

Frage adressiert werden, ob eventuell vorzufindende Einflüsse der familiären Herkunft zwischen den Fachnoten und den Prüfungsnoten aus dem schriftlichen Zentralabitur differieren. Hier ließe sich mutmaßen, dass der Einfluss leistungsfremder Einflüsse bei den Prüfungsnoten in dem von uns untersuchten Land Baden-Württemberg etwas niedriger ausfallen sollte, da die schriftlichen Prüfungsaufgaben hier zentral administriert werden und die Korrektur in einem teilweise anonymisierten Verfahren (Zweitkorrektur durch eine schulexterne Lehrkraft) erfolgt. Allerdings ist zu berücksichtigen, dass die Erstkorrektur der Prüfungsaufgaben jeweils durch den Kurslehrer der Schülerinnen und Schüler erfolgt, der auch für die Vergabe der Fachnoten zuständig ist. Insofern lassen sich keine klaren Erwartungen hinsichtlich differierender Einflüsse der familiären Herkunft auf die Fach- und Prüfungsnoten formulieren.

Bei den nachfolgend präsentierten Analysen ist ferner zu berücksichtigen, dass es sich bei der Schülerschaft der gymnasialen Oberstufe um eine hinsichtlich des familiären Hintergrundes stark positiv selektierte Schülergruppe handelt (Maaz, Chang & Köller, 2004). Daraus resultiert gleichzeitig eine deutlich eingeschränkte Varianz hinsichtlich der familiären Hintergrundmerkmale, was unter Umständen geringere Effekte der familiären Herkunft auf die Benotung erwarten lässt, als etwa im Fall der in den vorherigen Kapiteln präsentierten Befunde für die Grundschule, in der das gesamte Spektrum der familiären Herkunft eines Altersjahrgangs abgedeckt wird.

3.6.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Die Daten der vorliegenden Untersuchung entstammen der im Jahr 2002 in Baden-Württemberg durchgeführten Studie „Transformation des Sekundarschulsystems und akademische Karrieren“ (TOSCA; vgl. Köller et al., 2004; vgl. auch Anhang 3). Auch wenn inzwischen das Oberstufensystem in Baden-Württemberg reformiert wurde (vgl. Trautwein, Neumann, Nagy, Lüdtke & Maaz, 2010), gibt es wenig Grund für die Annahme, dass die im Folgenden berichteten Ergebnisse heute keine Gültigkeit mehr besitzen würden. In der TOSCA-Studie wurden die Bildungserträge von Abiturienten in den Bereichen Mathematik und Englisch am Ende der 13. Jahrgangsstufe untersucht. Zu diesem Zweck wurde eine repräsentative Stichprobe von $N = 3.526$ Schülerinnen und Schülern aus 90 allgemeinbildenden Gymnasien gezogen. Pro Schule wurden bis zu 40 Schülerinnen und Schüler in die Untersuchung einbezogen. 35,6 Prozent der Schülerinnen

und Schüler belegten den Mathematikleistungskurs, 64,4 Prozent besuchten den Grundkurs. Das Fach Englisch haben 39,5 Prozent im Leistungskurs und 45,8 Prozent im Grundkurs belegt.

Die Leistungen in Mathematik wurden mit dem Test zur voruniversitären Mathematik aus der Third International Mathematics and Science Study (TIMSS/III; vgl. Baumert, Bos & Lehmann 2000a, 2000b; Mullis et al., 1998; Waterman & Klieme, 2002) erfasst. Der Test umfasste insgesamt 68 Aufgaben aus den Stoffgebieten Zahlen/Gleichungen/Funktionen, Analysis, Geometrie, Aussagenlogik/Beweise und Wahrscheinlichkeitsrechnung/Statistik. Die Testkonzeption des TIMSS-Oberstufentests sah ausdrücklich einen starken Bezug zum Oberstufen-curriculum der Teilnehmerländer vor. Die Lehrplan- und Unterrichtsvalidität des Tests wurde in TIMSS mittels Lehrplananalysen, Experten- und Fachleiterbefragungen geprüft (vgl. zur Testvalidierung in Deutschland Klieme, 2000). Es zeigte sich, dass die Testaufgaben sowohl die Lehrpläne der Grund- und Leistungskurse als auch die Unterrichtsinhalte (realisiertes Curriculum) in hohem Maße widerspiegeln. Die Validitätsbelege konnten in der TOSCA-Untersuchung auf Grundlage erneuter Experten- und Fachleiterbefragungen repliziert werden (Waterman, Nagy & Köller, 2004).

Zur Erfassung der Leistungen in Englisch diente eine validierte Kurzform des Test of English as a Foreign Language (TOEFL, vgl. Köller & Trautwein, 2004). Der vom Educational Testing Service (ETS) in Princeton, New Jersey, entwickelte TOEFL, umfasst die Bereiche Hörverstehen, Grammatik und Orthografie sowie Wortschatz und Leseverständnis und wird insbesondere von amerikanischen Universitäten dazu verwendet, die Englisch-Fähigkeiten von Studienbewerbern, deren Muttersprache nicht Englisch ist, auf ein ausreichendes Niveau hin zu überprüfen. Anders als der TIMSS-Test zur voruniversitären Mathematik erhebt der TOEFL von seiner Konzeption her keinen direkten Anspruch auf curriculare Validität bezüglich der Inhalte des Englisch-Oberstufenunterrichts. Mit Bezug auf die baden-württembergischen Lehrpläne sowie die bundesweit verankerten Prüfungsvorgaben für das Fach Englisch kann jedoch davon ausgegangen werden, dass die Förderung der mit dem TOEFL erfassten rezeptiv-kommunikativen Fremdsprachenkompetenzen in den Zielkorridor des Englischunterrichts in der gymnasialen Oberstufe fällt (Köller & Trautwein, 2004), so dass auch hier eine hinreichende curriculare Validität der Testinhalte gegeben sein sollte.

Tabelle 19: Vorhersage der Fachnote in Mathematik (Schulhalbjahr 13/1) durch individuelle Schülermerkmale im Grundkurs

	R ²					
	Bivariat	Bivariat	M1	M2	M3	M4
	<i>b</i>		<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Intercept Fachnote Mathematik			7,79	7,75	7,60	7,61
Testleistung Mathematik ⁺	2,37	25,2	2,57	2,55	2,53	2,53
Kognitive Grundfähigkeit (KFT) ⁺	0,17	2,8	-0,14	-0,14	-0,13	-0,13
Geschlecht (<i>Ref. Jungen</i>)						
Mädchen	0,95	1,6	1,47	1,48	1,44	1,44
Sozioökonomischer Status (ISEI) ⁺	0,29	0,6		0,20	0,14	0,13
Ausbildung, schulisch (<i>Ref. kein Abitur</i>)						
mindestens ein Elternteil Abitur	0,69	0,9			0,25	0,26
Bücherbesitz (<i>Ref. mehr als 500</i>)						
bis 50	-1,98				-0,46	-0,39
51 bis 100	-0,61				0,30	0,34
101 bis 250	-0,63				0,12	0,13
251 bis 500	-0,27	1,3			0,06	0,07
Kulturelle Aktivitäten Eltern ⁺	0,28	0,6			0,09	0,09
Migrationshintergrund (<i>Ref. maximal ein Elternteil im Ausland geboren</i>)						
beide Eltern im Ausland geboren	-0,83	0,4				-0,28
R ²			29,1	29,4	29,7	29,8

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; R² = Multipler Determinationskoeffizient.

Neben den Fachleistungen in Mathematik und Englisch wurden auch die kognitiven Grundfähigkeiten (KFT Subtest Figuren analogien, Heller & Perleth, 2000) und Merkmale des familiären Hintergrundes der Schülerinnen und Schüler sowie die Fach- und Prüfungsnoten in den beiden untersuchten Fächern erhoben. Als sozialstrukturelles Merkmal der familiären Herkunft diente der höchste sozioökonomische Status in der Familie (ISEI). Der elterliche Bildungsabschluss (mindestens ein Elternteil Abitur), die Anzahl der Bücher zu Hause sowie eine Skala zur kulturellen Praxis (z.B. Häufigkeit und Opern- und Theaterbesuchen) dienten als Indikatoren des kulturellen Kapitals der Familie. Der Migrationsstatus ging in dichotomer Form in die Analysen ein (0 = maximal ein Elternteil im Ausland geboren, 1 = beide Elternteile im Ausland geboren). Alle kontinuierlichen Vorhersagevariablen (Testleistung, KFT, ISEI und kulturelle Praxis) wurden für die Analysen z-standardisiert (Mittelwert = 0, Standardabweichung = 1).

Die Angaben zu den Fach- und Prüfungsnoten entstammen den Schülerakten und sind auf der Metrik des

Credit-Systems zur Berechnung der Abiturgesamtnote abgetragen. Sie umfassen einen Wertebereich von 0 bis 15 Punkten, wobei höhere Werte bessere Noten indizieren. Die 15-Punkte-Skala lässt sich auch in die herkömmliche Notenskala von 1 bis 6 überführen: *Note 1* (13-15 Punkte) = „sehr gut“, *Note 2* (10-12 Punkte) = „gut“. *Note 3* (7-9 Punkte) = „befriedigend“, *Note 4* (4-6 Punkte) = „ausreichend“, *Note 5* (1-3 Punkte) = „mangelhaft“, *Note 6* (0 Punkte) = „ungenügend“. Aufgrund der feineren Abstufung wird in den nachfolgenden Analysen die Original-Punkte-Skala, die von 0 bis 15 Punkte reicht, verwendet.

Zur Untersuchung der Fragestellungen wurden mehrere Regressionsmodelle spezifiziert, in denen die Fach- und Prüfungsnoten durch die individuelle Testleistung, die kognitiven Grundfähigkeiten, das Geschlecht und die Merkmale des familiären Hintergrundes vorhergesagt wurden. Sofern sich auch nach Kontrolle der Testleistungen, der kognitiven Grundfähigkeiten und des Geschlechts Effekte der familiären Herkunft zeigen und man annimmt, dass der in Teilstudie 2 präsentierte Befund, wonach Indikato-

ren für schulischen Fleiß Effekte der familiären Herkunft nicht (vollständig) erklären können, auch für die gymnasiale Oberstufe ihre Gültigkeit behält, wäre dies ein Indiz dafür, dass Schülerinnen und Schüler auch bei gleichem Leistungsniveau in Abhängigkeit ihrer familiären Herkunft unterschiedlich benotet werden.

Die Analysen wurden mit dem Programmpaket Mplus 6.1 (Muthén & Muthén, 1998-2010) durchgeführt, das über die Analyseoption *type = complex* eine korrekte Schätzung der Standardfehler für die Regressionskoeffizienten für die geschachtelte Datenstruktur (Schüler geschachtelt innerhalb von Schulen) ermöglicht. Fehlende Werte wurden durch das *Multiple Imputation-Verfahren* (Rubin, 1987; Schafer & Graham, 2002) ersetzt. Die Analysen basieren auf dem kombinierten Ergebnis von fünf imputierten Datensätzen.

3.6.3. ERGEBNISSE

In den Tabellen 1a bis 4b sind die Ergebnisse aus den Regressionsanalysen zur Vorhersage der Fach- und Prüfungsnoten in den Fächern Mathematik und Englisch getrennt für das jeweilige Kursniveau (Grundkurs/Leistungskurs) ausgewiesen. Für den Englisch-Grundkurs beschränken wir uns auf Befunde für die Fachnoten, da nur 13,2 Prozent der Grundkursschülerinnen und -schüler auch eine schriftliche Abiturprüfung absolvierten. Da es sich bei diesen Schülerinnen und Schülern um eine leistungsmäßig besonders starke Teilgruppe handelt, wären die Ergebnisse dieser Gruppe nicht auf die Gesamtheit der Grundkursschülerinnen und -schüler übertragbar. Hingegen haben in Mathematik 90 Prozent der Grundkursschülerinnen und -schüler eine schriftliche Abiturprüfung abgelegt. Für die Leistungskursschülerinnen und -schüler war im Jahr 2002 die schriftliche Prüfung in beiden Fächern verpflichtend.

Vorhersagemodelle Mathematik

Die Tabellen 19 und 20 enthalten die Ergebnisse für den *Mathematikgrundkurs*. In Tabelle 1a finden sich zunächst die Befunde für die Fachnoten. In den beiden ersten Spalten sind die Angaben zu den bivariaten Zusammenhängen zwischen den verwendeten Prädiktoren und der Fachnote ausgewiesen. Spalte 1 enthält die Regressionkoeffizienten und Spalte 2 die durch die einzelnen Prädiktoren jeweils erklärte Varianz in der abhängigen Variablen. Wie aus Tabelle 19 hervorgeht, leistete die Mathematiktestleistung mit 25,2 Prozent in erwartbarer Weise den größten Erklärungsbeitrag. Unterschieden sich die Testleistungen von

zwei Schülern um eine Standardabweichung, ging dies mit durchschnittlichen Notenunterschieden von 2,37 Punkten auf der 15-Punkte-Metrik einher. Die aufgeklärte Varianz von 25,2 Prozent entspricht im bivariaten Modell einer Korrelation zwischen Testleistung und Fachnote in Höhe von $r = 0,50$ ($r = \text{Quadratwurzel aus } R^2$).

Auch für alle anderen Prädiktoren ergaben sich auf bivariater Ebene statistisch signifikante Zusammenhänge, deren Erklärungswert für die Unterschiede in den Fachnoten im Vergleich mit der Testleistung jedoch deutlich niedriger ausfiel. Schülerinnen erhielten im Mittel um knapp einen Punkt höhere Fachnoten als Schüler (siehe genauer Teilstudie 9). Die kognitiven Grundfähigkeiten korrelierten zu $r = 0,17$, der sozioökonomische Status und die kulturelle Praxis lediglich zu $r = 0,08$ mit den Fachnoten. Schülerinnen und Schüler, von denen wenigstens ein Elternteil über das Abitur verfügte, erzielten im Durchschnitt um 0,69 Punkte bessere Noten, und Schülerinnen und Schüler, deren Eltern beide im Ausland geboren wurden, erhielten um 0,83 Punkte schlechtere Noten als die Schülerinnen und Schüler der jeweiligen Referenzkategorien. Allerdings ist bei diesen Angaben zu berücksichtigen, dass es sich nur um die bivariaten Zusammenhänge ohne Berücksichtigung des tatsächlichen Leistungsniveaus der Schülerinnen und Schüler handelt. Die simultane Betrachtung der verschiedenen Prädiktoren erfolgte in den Analysemodellen M1 bis M4.

(→ TABELLE 19)

Im Analysemodell M1 wurden neben den Testleistungen auch die kognitiven Grundfähigkeiten und das Geschlecht mit aufgenommen. Wie zu erkennen ist, erhielten Mädchen bei gleichen Testleistungen im Durchschnitt eine um 1,47 Punkte bessere Fachnote (vgl. auch Teilstudie 10). Die kognitiven Grundfähigkeiten übten keinen spezifischen, über die Testleistungen und das Geschlecht hinausgehenden prädiktiven Einfluss auf die Fachnote aus. Durch die Hinzunahme beider Variablen stieg die erklärte Varianz in den Fachnoten um knapp 4 Prozent auf 29,1 Prozent. In Modell M2 wurde als erstes Merkmal der familiären Herkunft der sozioökonomische Status (ISEI) mit aufgenommen. Es zeigte sich ein statistisch signifikanter, von seinem Ausmaß her betrachtet jedoch eher als gering einzustufender Effekt. Bei ansonsten gleichen Ausprägungen auf den Prädiktorvariablen war ein Unterschied von einer Standardabweichung im sozioökonomischen Status mit einem Notenunterschied von 0,2 Punkten auf der 15-Punkte-Metrik verbunden. Entspre-

Tabelle 20: Vorhersage der schriftlichen Prüfungsnote in Mathematik durch individuelle Schülermerkmale im Grundkurs

	R ²					
	Bivariat	Bivariat	M1	M2	M3	M4
	<i>b</i>		<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Intercept Prüfungsnote Mathematik			8,44	8,42	8,32	8,34
Testleistung Mathematik ⁺	3,21	37,1	3,41	3,40	3,37	3,37
Kognitive Grundfähigkeit (KFT) ⁺	0,90	4,5	-0,16	-0,16	-0,14	-0,14
Geschlecht (Ref. Jungen)						
Mädchen	0,65	0,6	1,37	1,38	1,31	1,31
Sozioökonomischer Status (ISEI) ⁺	0,35	0,7		0,22	0,14	0,13
Ausbildung, schulisch (Ref. kein Abitur)						
mindestens ein Elternteil Abitur	0,89	1,2			0,23	0,24
Bücherbesitz (Ref. mehr als 500)						
bis 50	-2,24				-0,16	-0,09
51 bis 100	-0,90				0,29	0,33
101 bis 250	-1,10				-0,11	-0,10
251 bis 500	-0,44	1,8			0,04	0,05
Kulturelle Aktivitäten Eltern ⁺	0,38	0,8			0,20	0,20
Migrationshintergrund (Ref. maximal ein Elternteil im Ausland geboren)						
beide Eltern im Ausland geboren	-0,99	0,5				-0,30
R ²			39,8	40,1	40,6	40,6

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; R² = Multipler Determinationskoeffizient.

chend stieg die aufgeklärte Varianz durch Hinzunahme des sozioökonomischen Status lediglich um 0,3 Prozent auf 29,4 Prozent.

(→ TABELLE 20)

In Modell M3 wurden zusätzlich zum sozioökonomischen Status die Indikatoren des kulturellen Kapitals der Familie berücksichtigt. Hierbei zeigten sich keine statistisch signifikante Effekte. Gleichzeitig wurde nun der Effekt des sozioökonomischen Status nicht mehr signifikant, was als möglicher Hinweis auf die mediiierende Funktion der Indikatoren des kulturellen Kapitals für den Einfluss des sozioökonomischen Status gewertet werden kann. In Modell M4 wurde schließlich der Migrationsstatus mit einbezogen. Auch hier war nach Kontrolle der anderen Individualmerkmale kein statistisch signifikanter Effekt feststellbar. Insgesamt deuten die Befunde für die Fachnoten im Mathematikgrundkurs somit nur auf einen sehr geringen Einfluss familiärer Herkunftsmerkmale auf die Leistungsbewertung hin.

Tabelle 20 enthält die entsprechenden Befunde für die schriftlichen Prüfungsnoten, die mit einer Korrelation in Höhe von $r = 0,72$ substantiell mit den Fachnoten kovariierten. Gleichwohl gilt zu bedenken, dass eine Korrelation von $r = 0,72$ etwa 52 Prozent geteilter Varianz entspricht und somit prinzipiell betrachtet immer noch Spielräume für abweichende Ergebnismuster zwischen Fach- und Prüfungsnoten vorhanden sind. Dies zeigt sich in den bivariaten Analysen am ehesten für den Zusammenhang der Prüfungsnoten mit den Testleistungen, der deutlich stärker ausgeprägt ist als im Falle der Fachnoten. Ansonsten ließen sich auf bivariater Ebene, aber auch für die weiteren Analysemodelle, kaum Unterschiede zu den Ergebnissen für die Fachnoten ausmachen. Auch bei den Prüfungsnoten finden sich Bewertungsvorteile zugunsten der Mädchen und geringe Effekte des sozioökonomischen Status (vgl. Modell M2). Letztere wurden jedoch nicht mehr signifikant, sobald die Indikatoren des kulturellen Kapitals mit berücksichtigt wurden (vgl. Modell M3), unter denen wiederum nur für die kulturelle Praxis kleine

Tabelle 21: Vorhersage der Fachnote in Mathematik (Schulhalbjahr 13/1) durch individuelle Schülermerkmale im Leistungskurs

	R ²					
	Bivariat	Bivariat	M1	M2	M3	M4
	<i>b</i>		<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Intercept Fachnote Mathematik			7,11	7,13	6,99	7,01
Testleistung Mathematik ⁺	2,19	33,2	2,39	2,38	2,36	2,34
Kognitive Grundfähigkeit (KFT) ⁺	0,73	4,4	-0,21	-0,20	-0,20	-0,19
Geschlecht (<i>Ref. Jungen</i>)						
Mädchen	0,10	0,00	0,93	0,94	0,88	0,87
Sozioökonomischer Status (ISEI) ⁺	0,39	1,4		0,31	0,21	0,18
Ausbildung, schulisch (<i>Ref. kein Abitur</i>)						
mindestens ein Elternteil Abitur	0,87	1,7			0,28	0,34
Bücherbesitz (<i>Ref. mehr als 500</i>)						
bis 50	-1,60				0,14	0,25
51 bis 100	-1,11				-0,05	0,03
101 bis 250	-0,92				-0,07	-0,05
251 bis 500	-0,23	2,3			0,10	0,10
Kulturelle Aktivitäten Eltern ⁺	0,33	1,0			0,21	0,21
Migrationshintergrund (<i>Ref. maximal ein Elternteil im Ausland geboren</i>)						
beide Eltern im Ausland geboren	-1,42	1,6				-0,60
R ²			35,2	36,1	36,9	37,2

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; R² = Multipler Determinationskoeffizient.

statistisch absicherbare Effekte feststellbar waren. Die Individualvariablen klärten insgesamt 40,6 Prozent der Unterschiede in den Prüfungsnoten und damit deutlich mehr Varianz als im Falle der Fachnoten auf, wobei der erhöhte Erklärungsbeitrag in erster Linie auf den engeren Zusammenhang der Testleistungen mit den Prüfungsnoten zurückzuführen ist.

Die Ergebnisse für den *Mathematikleistungskurs* sind in den Tabellen 21 und 22 dargestellt. Die bivariaten Zusammenhänge der Vorhersagevariablen zu den Fachnoten (vgl. Tabelle 21) bewegten sich überwiegend in einem Bereich wie im Mathematikgrundkurs, wobei der Erklärungsanteil der Testleistungen im Leistungskurs etwas höher ausfiel. Interessant ist zudem, dass sich abweichend von den bivariaten Befunden für den Grundkurs kein statistisch signifikanter Unterschied in den Fachnoten zwischen den Geschlechtern fand (vgl. dazu auch Teilstudie 10). Die Analysemodelle M1 bis M4 replizierten im Wesentlichen die Ergebnisse des Mathematikgrundkurses. Nach Kontrolle der Testleistungen fanden sich Notenvorteile zugunsten

der Schülerinnen gegenüber den Schülern in Höhe von etwa einem Notenpunkt. Weiterhin zeigte sich ein geringer, aber statistisch signifikanter Effekt des sozioökonomischen Status, der aber verschwand, sobald die Indikatoren des kulturellen Kapitals mit einbezogen wurden, wobei unter letzteren erneut ein geringer Effekt für die kulturelle Praxis beobachtet werden konnte. Erwähnenswert ist zudem, dass der negative Effekt des Migrationsstatus das Signifikanzniveau von 5 Prozent nur knapp verfehlte ($p = .074$) und somit eine Tendenz erkennbar ist, wonach Schülerinnen und Schüler, deren Eltern beide nicht in Deutschland geboren sind, auch bei gleichen Testleistungen und vergleichbaren sozialen Hintergrundmerkmalen etwas niedrigere Noten erhielten als Schülerinnen und Schüler, von denen maximal ein oder kein Elternteil im Ausland geboren wurde. Interessant ist, dass sich dieser Befund nur für die Fachnoten, jedoch nicht für Prüfungsnoten (vgl. Tabelle 22) zeigte. Hier ließen sich im finalen Modell M 4 keinerlei statistisch signifikante Effekte der familiären Herkunft mehr nachweisen. (→ TABELLE 21 UND 22)

Tabelle 22: Vorhersage der Prüfungsnote in Mathematik durch individuelle Schülermerkmale im Leistungskurs

	R ²					
	Bivariat	Bivariat	M1	M2	M3	M4
	<i>b</i>		<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Intercept Prüfungsnote Mathematik			7,52	7,53	7,19	7,20
Testleistung Mathematik ⁺	2,60	37,7	2,74	2,73	2,72	2,72
Kognitive Grundfähigkeit (KFT) ⁺	0,96	6,2	-0,11	-0,11	-0,11	-0,10
Geschlecht (Ref. Jungen)						
Mädchen	-0,29	0,1	0,70	0,71	0,65	0,65
Sozioökonomischer Status (ISEI) ⁺	0,27	0,5		0,20	0,13	0,12
Ausbildung, schulisch (Ref. kein Abitur)						
mindestens ein Elternteil Abitur	0,76	1,1			0,27	0,30
Bücherbesitz (Ref. mehr als 500)						
bis 50	-1,49				0,41	0,46
51 bis 100	-1,00				0,14	0,17
101 bis 250	-0,68				0,12	0,14
251 bis 500	0,16	1,8			0,50	0,50
Kulturelle Aktivitäten Eltern ⁺	0,22	0,4			0,17	0,21
Migrationshintergrund (Ref. maximal ein Elternteil im Ausland geboren)						
beide Eltern im Ausland geboren	-1,18	0,9				-0,26
R ²			38,6	38,9	39,6	39,7

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; R² = Multipler Determinationskoeffizient.

Vorhersagemodelle Englisch

Die Ergebnisse zur Vorhersage der Fachnoten im *Englischgrundkurs* finden sich in Tabelle 23. Bei der Betrachtung der bivariaten Zusammenhänge zeigte sich mit 26,8 Prozent erklärter Varianz zunächst erneut der mit Abstand größte Erklärungsbeitrag der Testleistungen. Aber auch für den Großteil der familiären Hintergrundmerkmale konnten auf bivariater Ebene statistisch signifikante Zusammenhänge beobachtet werden. Diese waren jedoch nicht mehr nachweisbar, sobald die Testleistungen, die kognitiven Grundfähigkeiten und das Geschlecht der Schülerinnen und Schüler simultan in den Analysen berücksichtigt wurden (vgl. Modelle M2 bis M4). Im Englischgrundkurs fanden sich somit keine Hinweise auf Bewertungsunterschiede in Abhängigkeit der familiären Herkunft.

(→ TABELLE 23)

Die Tabellen 24 und 25 weisen die Ergebnisse für die Fach- und Prüfungsnoten im Englischleistungskurs aus.

Wie Modell M2 für die Fachnoten (vgl. Tabelle 24) entnommen werden kann, resultierte für den Englischleistungskurs ähnlich wie in Mathematik ein spezifischer Effekt des sozioökonomischen Status, der von seinem Ausmaß her jedoch eher als gering einzustufen ist. Wie Modell M3 zeigt, verschwand der Effekt des sozioökonomischen Status nach Aufnahme der Indikatoren des kulturellen Kapitals. Hier ließen sich statistisch signifikante Effekte des elterlichen Bildungshintergrundes und der kulturellen Praxis feststellen. Spezifische Effekte des Migrationsstatus (vgl. M4) waren hingegen nicht nachweisbar. Das beschriebene Befundmuster ergab sich im Wesentlichen auch für die Prüfungsnoten im Englischleistungskurs (vgl. Tabelle 25), die in Höhe von $r = .78$ mit den Fachnoten korrelierten. Auch hier verblieben nach Kontrolle der Testleistungen und der kognitiven Grundfähigkeiten statistisch signifikante Effekte des elterlichen Bildungshintergrundes und der kulturellen Praxis, allerdings erneut von eher als gering einzustufendem Ausmaß.

(→ TABELLE 24 UND 25)

Tabelle 23: Vorhersage der Fachnote in Englisch (Schulhalbjahr 13/1) durch individuelle Schülermerkmale im Grundkurs

	R ²					
	Bivariat	Bivariat	M1	M2	M3	M4
	<i>b</i>		<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Intercept Fachnote Englisch			8,68	8,68	8,70	8,70
Testleistung Englisch ⁺	1,65	26,8	1,73	1,72	1,69	1,69
Kognitive Grundfähigkeit (KFT) ⁺	0,38	1,6	0,02	0,02	0,02	0,02
Geschlecht (<i>Ref. Jungen</i>)						
Mädchen	0,27	0,2	0,89	0,89	0,86	0,86
Sozioökonomischer Status (ISEI) ⁺	0,25	0,7		0,06	-0,02	-0,02
Ausbildung, schulisch (<i>Ref. kein Abitur</i>)						
mindestens ein Elternteil Abitur	0,66	1,3			0,15	0,16
Bücherbesitz (<i>Ref. mehr als 500</i>)						
bis 50	-1,19				-0,16	-0,14
51 bis 100	-1,10				-0,26	-0,25
101 bis 250	-1,07				-0,39	-0,38
251 bis 500	-0,21	2,8			0,05	0,05
Kulturelle Aktivitäten Eltern ⁺	0,19	0,4			0,03	0,03
Migrationshintergrund (<i>Ref. maximal ein Elternteil im Ausland geboren</i>)						
beide Eltern im Ausland geboren	-0,49	0,2				-0,10
R ²			29,0	29,0	29,5	29,5

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; R² = Multipler Determinationskoeffizient.

3.6.4. RESÜMEE

Die vorstehenden Befunde zur Leistungsbewertung am Ende der gymnasialen Oberstufe lassen sich dahingehend zusammenfassen, dass in den beiden untersuchten Fächern Mathematik und Englisch nach Berücksichtigung des Leistungsniveaus der Schülerinnen und Schüler zwar statistisch signifikante, vom Ausmaß her betrachtet jedoch vergleichsweise geringe Effekte der familiären Herkunft beobachtet werden konnten. Mit Ausnahme der Fachnoten im Grundkurs Englisch fand sich durchgängig ein Effekt des sozioökonomischen Status, der jedoch nach zusätzlicher Kontrolle der Indikatoren des kulturellen Kapitals nicht mehr statistisch signifikant wurde. Unter den Indikatoren des kulturellen Kapitals zeigten sich die konsistentesten Effekte für die kulturelle Praxis im Elternhaus, wobei die vorliegende Untersuchung keine Aussagen über die diesem Effekt zugrunde liegenden kausalen Wirkprozesse erlaubt. Auch mit Blick auf den Migrationsstatus fanden sich kaum Hinweise auf eine Benachteiligung von Schülerinnen und Schülern mit Mig-

rationshintergrund. Lediglich im Falle der Fachnoten im Mathematikleistungskurs fand sich ein in diese Richtung weisender Effekt, der jedoch das statistische Signifikanzkriterium von $p < 0,05$ knapp verfehlte.

Hervorzuheben ist weiterhin, dass sich bezüglich des beschriebenen Ergebnismusters kaum Unterschiede zwischen den Fach- und Prüfungsnoten zeigten, was im vorliegenden Fall als eher positives Resultat betrachtet werden sollte, da sich für die Fachnoten wie ausgeführt kaum Hinweise auf Nachteile bei der Leistungsbewertung in Abhängigkeit des familiären Hintergrundes feststellen ließen. Auffällig waren die Befunde für das Geschlecht, die in Teilstudie 9 detaillierter betrachtet werden sollen.

Einschränkend ist zu sagen, dass in der vorliegenden Studie neben den eingesetzten Leistungstests keine motivationalen Schülermerkmale wie Anstrengungsbereitschaft, Mitarbeit im Unterricht, Gewissenhaftigkeit und ähnliche unterrichtsbezogene Aspekte einbezogen wurden, von denen bekannt ist, dass sie ebenfalls von Bedeutung für die Notengebung sind (vgl. Teilstudie 2)

Tabelle 24: Vorhersage der Fachnote in Englisch (Schulhalbjahr 13/1) durch individuelle Schülermerkmale im Leistungskurs

	R ²					
	Bivariat	Bivariat	M1	M2	M3	M4
	<i>b</i>		<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Intercept Fachnote Englisch			7,92	7,91	8,09	8,08
Testleistung Englisch ⁺	1,70	30,9	1,79	1,77	1,67	1,68
Kognitive Grundfähigkeit (KFT) ⁺	0,27	0,9	-0,10	-0,10	-0,08	-0,08
Geschlecht (Ref. Jungen)						
Mädchen	0,76	1,8	1,07	1,08	0,98	0,98
Sozioökonomischer Status (ISEI) ⁺	0,36	1,6		0,23	0,07	0,09
Ausbildung, schulisch (Ref. kein Abitur)						
mindestens ein Elternteil Abitur	0,88	2,4			0,30	0,29
Bücherbesitz (Ref. mehr als 500)						
bis 50	-2,40				-0,41	-0,45
51 bis 100	-1,75				-0,50	-0,52
101 bis 250	-1,35				-0,34	-0,35
251 bis 500	-0,88	6,2			-0,33	-0,34
Kulturelle Aktivitäten Eltern ⁺	0,59	4,3			0,24	0,24
Migrationshintergrund (Ref. maximal ein Elternteil im Ausland geboren)						
beide Eltern im Ausland geboren	-0,61	0,4				0,18
R ²			34,6	35,2	36,9	36,9

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; R² = Multipler Determinationskoeffizient.

und somit möglicherweise Hinweise auf die den Effekten der familiären Herkunft und des Geschlechts zugrundeliegenden Prozessen geben können. Ferner ist darauf hinzuweisen, dass sich unsere Befunde nur auf die Fächer Mathematik und Englisch beziehen. Um die Ergebnisse auf eine breitere empirische Basis zu stellen, bedarf es weiterer Untersuchungen in anderen Fächern unter Einbezug weiterer Bundesländer.

3.7. Teilstudie 7: Übertrittsempfehlungen aus Sicht der Lehrkräfte: Pro Klasse im Durchschnitt eine Gymnasialempfehlung zu viel

Frage 7: Wie oft geben Lehrkräfte nach eigenen Angaben eine „falsche“ Gymnasialempfehlung? Wie sehr hängen „falsche“ Gymnasialempfehlungen mit dem sozialen Hintergrund der Schülerinnen und Schüler zusammen?

3.7.1. HINTERGRUND

Lehrkräfte empfinden es oftmals als belastend, am Ende der Grundschulzeit eine Übertrittsempfehlungen auszusprechen (McElvany, 2010), da Übertrittsempfehlungen eine wichtige Rolle in der weiteren schulischen Laufbahn der Schülerinnen und Schüler zukommt und die diagnostischen Kompetenzen der Lehrkräfte sowie die zur Verfügung stehenden Informationen für eine solche diagnostische Entscheidung im Grunde nicht ausreichend sind (vgl. Kapitel 2 sowie Schrader, 2006; Tent, 2006). Entsprechend muss auch angenommen werden, dass mit einer substanziellen Anzahl von in psychometrischer Hinsicht „falschen“ Übertrittsempfehlungen gerechnet werden muss. Dass sie eine in diesem Sinne „falsche“ Übertrittsempfehlung aussprechen, ist den Lehrkräften dabei in vielen Fällen nicht bewusst.

Die Situation liegt bei der im Folgenden untersuchten Frage anders. Untersucht wird, wie häufig Lehrkräfte nach eigenen Angaben eine Gymnasialempfehlung aussprechen, obwohl sie eine andere Empfehlung für angemessen

Tabelle 25: Vorhersage der Prüfungsnote in Englisch durch individuelle Schülermerkmale im Leistungskurs

	R ²					
	Bivariat	Bivariat	M1	M2	M3	M4
	<i>b</i>		<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Intercept Prüfungsnote Englisch			7,59	7,59	7,61	7,61
Testleistung Englisch ⁺	2,03	32,3	2,14	2,12	2,05	2,05
Kognitive Grundfähigkeit (KFT) ⁺	0,27	0,7	-0,16	-0,16	-0,13	-0,13
Geschlecht (<i>Ref. Jungen</i>)						
Mädchen	0,68	1,1	1,12	1,13	1,03	1,03
Sozioökonomischer Status (ISEI) ⁺	0,39	1,4		0,21	0,07	0,08
Ausbildung, schulisch (<i>Ref. kein Abitur</i>)						
mindestens ein Elternteil Abitur	1,01	2,4			0,39	0,39
Bücherbesitz (<i>Ref. mehr als 500</i>)						
bis 50	-2,47				-0,11	-0,12
51 bis 100	-1,63				-0,12	-0,12
101 bis 250	-1,42				-0,20	-0,21
251 bis 500	-0,97	4,7			-0,27	-0,27
Kulturelle Aktivitäten Eltern ⁺	0,62	3,7			0,24	0,24
Migrationshintergrund (<i>Ref. maximal ein Elternteil im Ausland geboren</i>)						
beide Eltern im Ausland geboren	-0,79	0,5				0,02
R ²			35,5	36,0	37,1	37,2

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; R² = Multipler Determinationskoeffizient.

sen halten, bzw. wie oft sie von einer Gymnasialempfehlung absehen, obwohl sie diese für richtig halten. Anders gesagt: Wie häufig spielen sachfremde Einflüsse bei den Übergangsempfehlungen eine Rolle?

Wie kann es zu Übertrittsempfehlungen bei den Schülerinnen und Schülern kommen, die nicht im Einklang mit den Überzeugungen der Lehrkräfte über die „richtige“ Empfehlung stehen? Es lassen sich grob zumindest drei unterschiedliche Arten unterscheiden. Erstens nehmen in der Grundschule viele Eltern einen Einfluss auf die von ihren Kindern tatsächlich gezeigte Schulleistung, indem sie ihre Kinder sehr intensiv beim Lernen unterstützen, ihnen die Wahrnehmung zusätzlicher Lernangebote ermöglichen bzw. sie bei „schwachen“ Schulleistungen unter Druck setzen, mehr für die Schule zu tun (vgl. Wild, 2004; zu Grenzen der Förderung, siehe Dumont et al., in press). In Bezug auf Kinder, in deren „guter“ Schulleistung sich der Einfluss der Eltern sehr stark abzeichnet bzw. bei denen die Lehrkräfte wegen des Einsatzes und Drucks der Eltern eine kognitive oder psychische Überforderung im

Gymnasium erwarten, könnten Lehrkräfte – trotz objektiv guter Schulleistungen – zu der Auffassung kommen, dass der Gymnasialbesuch „eigentlich“ nicht zu empfehlen sei.

Zweitens weist anekdotische Evidenz immer wieder darauf hin, dass viele Eltern – neben der Unterstützung ihrer Kinder beim Lernen – in relativ eindeutiger Art und Weise einen direkten Einfluss auf die Übertrittsempfehlungen zu nehmen versuchen, wenn die Schulleistungen ihres Kindes keine Gewähr für eine Gymnasialempfehlung gibt. Die entsprechenden Versuche können sich dabei u.a. auf die übertrittsrelevanten Schulnoten beziehen. So können Eltern auf „mildernde“ Umstände hinweisen, die dafür gesorgt haben, dass ihr Kind nicht seine maximale Leistungsfähigkeit zeigen konnte. Zudem können sie in Vorbereitung auf die Übertrittsentscheidungen Eigenschaften des Kindes hervorheben, die es besonders auszeichnen und denen eine leistungsförderliche Wirkung in der Sekundarschulzeit zugesprochen wird. Außerdem können Eltern ihre Bereitschaft signalisieren, ihr Kind

während der Sekundarschulzeit intensiv zu unterstützen, sollte es zu Leistungsschwierigkeiten kommen. Trifft die Lehrkraft in einer solchen Situation eine Empfehlung für das Gymnasium, so ist die Entscheidung zwar falsch (da die Leistungen nicht den Anforderungen an eine Gymnasialempfehlung entsprachen), aber vertretbar in dem Sinne, dass neben den gezeigten Leistungen auch das Leistungspotenzial bei Übertrittsentscheidungen einbezogen werden soll; entsprechende Gymnasialempfehlungen mögen aus subjektiver Sicht der Lehrkraft deshalb auch mehr oder weniger plausibel und gerechtfertigt sein.

Drittens gibt es auch anekdotische Evidenz dafür, dass Eltern den Entscheidungsprozess auf eindeutig unangemessene Art und Weise zu beeinflussen versuchen. Solche Beeinflussungsmaßnahmen können beispielsweise mit Hinweis auf Rolle der Eltern in der Schule/der Gemeinde bzw. persönliche Beziehungen oder Beziehungsnetzwerke erfolgen, und sie können direkt bei der entsprechenden Lehrkraft oder bei anderen einflussreichen Personen (beispielsweise beim Schulleiter) ansetzen. In diesem Falle trifft die Lehrkraft ggf. eine Empfehlung, die bewusst und explizit in Kontrast zur Überzeugung der Lehrkraft in Hinblick auf die eigentlich „richtige“ Empfehlung ausfällt.

In dieser Teilstudie soll untersucht werden, wie häufig die von den Lehrkräften vergebene Übertrittsempfehlung von einer Empfehlung abweicht, die nach dem eigenen Dafürhalten der Lehrkräfte die richtige Empfehlung wäre. Zudem soll geprüft werden, ob eine „zu gute“ Empfehlung mit dem sozialen Hintergrund der Schülerinnen und Schüler assoziiert ist.

3.7.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Die Daten der nachfolgenden Analysen entstammen dem am Berliner Max-Planck-Institut für Bildungsforschung durchgeführten Projekt „Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (ÜBERGANG) (Maaz, Baumert, Gresch & McElvany, 2010; Becker et al., 2010). Weitere Informationen zur Studienanlage finden sich im Anhang 1.

Für die Analysen wurden drei Informationen herangezogen. Erstens wurde der soziale Hintergrund der Schülerinnen und Schüler in Form der bereits vorgestellten ISEI-Variable verwendet. Zweitens wurden die Lehrkräfte nach der tatsächlich vergebenen Übertrittsempfehlung für jeden einzelnen Schüler/jede einzelne Schülerin ihrer Klasse gefragt. Drittens wurden die Lehrkräfte auch um die Angabe gebeten, welche Übertrittsempfehlung sie persönlich als angemessen empfinden würden. Aus den beiden letztgenannten Informationen wurde eine dichotome Variable gebildet, die bei allen Schülerinnen und

Schülern den Wert 1 annimmt, die eine Gymnasialempfehlung erhielten, obwohl die Lehrkraft eigentlich eine andere Empfehlung als angemessen empfand. Bei allen anderen Schülerinnen und Schüler wurde der Wert auf 0 festgesetzt.

3.7.3. ERGEBNISSE

Nach Angaben der Lehrkräfte erhielt etwas mehr als jeder 20. Schüler/jede 20. Schülerin (5,6 % der Schülerinnen und Schüler) eine Gymnasialempfehlung, obwohl sie eine Gymnasialempfehlung persönlich nicht als angemessen empfanden. Somit kann man grob von einem Schüler/einer Schülerin pro Klasse sprechen, die eine in diesem Sinne „zu gute“ Empfehlung erhielt. Unterteilt man die Stichprobe in Bundesländer mit bzw. ohne bindender Empfehlung, so zeigt sich, dass in den Bundesländern mit bindender Empfehlung die „zu guten“ Empfehlungen mit insgesamt 6,6 Prozent etwas häufiger vorkommen als in den Bundesländern ohne bindende Empfehlung, wo diese Zahl 3,4 Prozent betrug. Der entsprechende statistische Vergleich ist statistisch signifikant.

In einem zweiten Schritt wurde überprüft, inwieweit die neu gebildete Empfehlungsstatus-Variable mit dem sozialen Hintergrund kovariiert. Anders gesagt: Gibt es einen Zusammenhang zwischen sozialem Hintergrund und dem Bericht der Lehrkraft, dass ein Schüler bzw. eine Schülerin eine Gymnasialempfehlung erhalten habe, obwohl diese nicht angemessen sei? Für die Gesamtstichprobe lag Kendalls τ , ein Maß für die Assoziation von ordinal skalierten Variablen, bei $\tau = 0,05$ ($p < 0,01$). In getrennte Analysen fand sich für die Bundesländer mit bindender Empfehlung ebenfalls ein statistisch signifikanter Wert, $\tau = 0,06$ ($p < 0,01$), während für die Bundesländer mit nicht-bindender Empfehlung keine statistisch signifikante Assoziation ($\tau = 0,03$, *ns*) resultierte.

3.7.4. RESÜMEE

Schülerinnen und Schüler erhielten in 5,6 Prozent aller Fälle eine Übertrittsempfehlung für das Gymnasium, obwohl ihre Lehrkräfte eine weniger positive Empfehlung für angemessen hielten. Die Abweichung von der angemessenen Empfehlung ist dabei schwach, aber statistisch signifikant mit dem sozialen Hintergrund der Schülerinnen und Schüler assoziiert.

Die berichteten Befunde machen deutlich, dass auch Lehrkräften bewusst ist, dass ein gewisses Maß an „Abweichung“ von den eigentlich angemessenen Empfehlungen auftaucht. Tiefergehenden Analysen muss es vorbehalten bleiben, zu prüfen, welche Art von Elternverhalten sowie weiteren Ursachen in besonders großer Zahl zu einer „zu guten“ Übertrittsempfehlung beitragen.

Tabelle 26: Von Lehrkräften eingeschätzte Wichtigkeit unterschiedlicher Bereiche für den schulischen Erfolg eines Kindes

Skala		Beispielitem	Anzahl Items	Cronbachs α
		Für den schulischen Erfolg eines Kindes ist es wichtig, dass es...		
AF	Akademische Begabung und schulische Fähigkeiten	...gute sprachliche Fähigkeiten hat.	8	0,81
AD	Anstrengung und Durchhalten	...bereit ist, sich in der Schule anzustrengen.	5	0,79
BL	Psychische Belastbarkeit	...nicht überängstlich ist.	5	0,78
MO	Motivation	...Freude daran hat, sich neues Wissen anzueignen.	4	0,75
SB	Selbstbeherrschung	...Emotionen auch kontrollieren kann.	5	0,79
SO	Soziales Verhalten	...ein gutes Sozialverhalten hat.	4	0,85
ST	Sekundärtugend	...diszipliniert ist.	4	0,76
SV	Schulrelevantes Verhalten	...gut im Unterricht mitarbeitet.	5	0,70
UM	Familiäres / soziales Umfeld	...im Hinblick auf Bildungsfragen in einem unterstützenden sozialen Umfeld lebt.	3	0,80

3.8. Teilstudie 8: Wird von Lehrkräften das familiäre Umfeld bei der Übertrittsentscheidung (zu) wichtig genommen?

Frage 8: Für wie wichtig erachten Lehrkräfte das familiäre und soziale Umfeld für den schulischen Erfolg und hat diese Einschätzung einen Effekt auf die Übertrittschancen der Schülerinnen und Schüler?

3.8.1. HINTERGRUND

Der soziale Hintergrund von Schülerinnen und Schülern ist ein wichtiger Prädiktor für ihren Bildungserfolg (vgl. Baumert et al., 2011), der sich über primäre und sekundäre Disparitäten auf den Bildungserfolg auswirkt (Kap. 2). Ein günstiger familiärer Hintergrund ist während der Primar- und Sekundarschulzeit mit hohen Schulleistungen ebenso assoziiert wie mit einer besonders großen Chance auf den Besuch attraktiver Bildungsgänge, und einem günstigen familiären Hintergrund wird auch eine Schutzwirkung gegen Leistungsprobleme nach dem Übertritt in die weiterführenden Schulen nachgesagt (z.B. Baumert, Maaz & Trautwein, 2009; Hillmert & Jakob, 2010). Die Bildungsforschung ist sich jedoch darin einig, dass der Effekt des sozialen Hintergrunds unerwünscht groß ausfällt und reduziert werden sollte.

Hinsichtlich des Wissens über die (tatsächliche) Bedeutung des sozialen Hintergrunds ergibt sich ein Dilemma: Einerseits ist es prinzipiell wünschenswert, dass Lehrkräfte über unterschiedliche Einflussfaktoren des Schulerfolgs Bescheid wissen. Auch kann es potenziell zur Erhöhung der Prognoseleistung von Übertrittsentscheidungen beitragen, wenn Lehrkräfte in ihre Empfehlung

alle wichtigen Faktoren einbeziehen. Andererseits muss es aus Hinsicht sozialer Gerechtigkeit als unerwünscht gelten, wenn bei gleicher individueller Leistung der soziale Hintergrund darüber entscheidet, wie die Übertrittsentscheidung ausfällt, auch wenn dieser möglicherweise einen Prognosebeitrag leistet. Lehrkräfte stehen damit in einem gewissen Spannungsfeld zwischen schulischer Wirklichkeit mit persistierenden Effekten sozialer Ungleichheiten (zu denen sie durch ihre Übertrittsempfehlung dann auch beitragen) und den Zielen „gerechterer“ Lernumgebungen, in denen die Übertrittsentscheidungen ohne Berücksichtigung des sozialen Hintergrund getroffen werden und Schülerinnen und Schüler mit weniger günstigem familiären/sozialem Hintergrund eine spezifische Förderung in den weiterführenden Schulen erfahren (vgl. Baumert et al., 2011).

In dieser Teilstudie soll anhand der Aussagen der Lehrkräfte geprüft werden, für wie wichtig sie das familiäre/soziale Umfeld für den Lernerfolg der Schülerinnen und Schüler einschätzen. Zudem soll geprüft werden, in welchem Zusammenhang diese Einschätzung mit weiteren übertrittsrelevanten Variablen steht.

3.8.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Die Daten der nachfolgenden Analysen entstammen dem am Berliner Max-Planck-Institut für Bildungsforschung durchgeführten Projekt „Der Übergang von der Grundschule in die weiterführende Schule – Leistungs-gerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (ÜBERGANG) (Maaz, Baumert, Gresch & McElvany, 2010; Becker, et al., 2010). Weitere Informationen zur Studienanlage finden sich im Anhang 1.

(→ TABELLE 26)

Tabelle 27: Mittelwerte für die von den Lehrkräften eingeschätzte Wichtigkeit unterschiedlicher Merkmale

	<i>M</i>	<i>SD</i>
Akademische Begabung und schulische Fähigkeiten	4,46	0,54
Anstrengung und Durchhalten	4,87	0,54
Psychische Belastbarkeit	4,36	0,55
Motivation	4,98	0,58
Selbstbeherrschung	3,77	0,64
Soziales Verhalten	3,78	0,72
Sekundärtugend	4,12	0,61
Schulrelevantes Verhalten	4,57	0,52
Familiäres / soziales Umfeld	4,22	0,80

Verwendet wurden Einschätzungen von Lehrkräften zu für den Schulerfolg wichtigen Aspekten, die in Tabelle 26 dargestellt sind. Den Lehrkräften stand zur Einschätzung der Wichtigkeit der jeweiligen Aspekte eine Antwortskala zur Verfügung, die von 1 bis 6 reichte, wobei ein höherer Wert eine höhere Wichtigkeit bedeutete. Alle Skalen wiesen akzeptable bis gute psychometrische Kennwerte auf. Ein besonderer Schwerpunkt liegt im Folgenden auf der „Umfeld-Skala“ (letzte Zeile der Tabelle).

Zudem wurden drei weitere Variablen verwendet, die auf dem Bericht der Lehrkräfte basierten. Erstens wurden die Lehrkräfte gefragt, ob sie bisweilen die Schulnoten eines Kindes an die vergebene Übertrittsempfehlung anpassen: „Noten spielen eine wichtige Rolle bei der Frage des Übergangs. Passen Sie Ihre Notengebung an, um die bevorstehende Übergangsempfehlung/-entscheidung eindeutiger zu machen?“ Zweitens wurden die Lehrkräfte danach gefragt, wie stark sie schwache Leistungen auf mangelnde Anstrengung der Kinder attribuierten: „Wenn Sie einmal an die leistungsschwächeren Schülerinnen und Schüler Ihrer Klasse denken, woran könnten deren Misserfolge liegen?“ Aus Antworten zu vier Items (beispielweise „zu geringe häusliche Anstrengung“ und „mangelnder Fleiß“) wurde eine Skala gebildet. Drittens wurden die Lehrkräfte selbst nach ihrem sozialen Hintergrund gefragt, so dass für jede Lehrkraft der sozioökonomische Hintergrund (ISEI-Variable) bestimmt werden konnte.

3.8.3. ERGEBNISSE

Wie der Tabelle 27 zu entnehmen ist, maßen Lehrkräfte im Mittel dem familiären/schulischen Umfeld der Kinder eine nicht unwichtige, vielleicht aber auch keine herausragende Bedeutung zu. (Da die in Tabelle 27 darge-

stellten unterschiedlichen Merkmale mit jeweils spezifischen Fragen erhoben wurden, wäre es nicht statthaft, die Mittelwerte direkt miteinander zu vergleichen.) Auffällig ist die relativ hohe Standardabweichung (SD) bei der Einschätzung der Wichtigkeit des familiären/sozialen Umfelds. Diese hohe Standardabweichung bedeutet, dass die Antworten der unterschiedlichen Lehrerinnen und Lehrer relativ stark streuten. Während manche Lehrkräfte diesem Merkmal eine sehr wichtige Rolle zumaßen, betrachteten es andere als weniger wichtig. Eine statistische Analyse, in der geprüft wurde, ob sich die Wichtigkeit des genannten Indikators über Länder mit bzw. ohne bindender Übertrittsempfehlung unterscheidet, erbrachte ein nicht-signifikantes Ergebnis.

(→ TABELLE 27)

Im nächsten Schritt wurde geprüft, ob es einen Zusammenhang zwischen dem eigenen familiären Hintergrund der Lehrkraft, der Bedeutung, die die Lehrkräfte dem familiären/sozialen Umfeld der Schülerinnen und Schüler zusprechen, sowie der Anpassung von Noten an die Übertrittsempfehlung und Attribuierungsmuster für schulischen Misserfolg gibt. Die Ergebnisse dieser Analyse sind in Tabelle 28 berichtet.

(→ TABELLE 28)

Wie die Tabelle 28 zeigt, muss bei der Betrachtung des Zusammenhangs zwischen Bundesländern mit bzw. ohne bindender Übertrittsempfehlung unterschieden werden. In Bundesländern mit bindender Empfehlung fand sich ein statistisch signifikanter Zusammenhang der Lehrkräfte-Einschätzung der Wichtigkeit des familiären/sozialen Umfelds mit dem sozialen Hintergrund der Lehrkräfte,

Tabelle 28: Korrelationen der untersuchten Lehrervariablen in Bundesländern mit bindender Übergangsempfehlung (unterhalb der Diagonalen, untere Dreiecksmatrix) vs. in Bundesländern ohne bindende Übergangsempfehlung (oberhalb der Diagonalen, obere Dreiecksmatrix)

	UM	AN	MA	HISEI
(UM) Familiäres/soziales Umfeld	1,00	0,06	0,13	-0,03
(AN) Anpassung der Notengebung an Übergangsempfehlung	0,20	1,00	-0,13	-0,18
(MA) Attribution schwacher Leistungen auf mangelnde Anstrengung	0,26	0,02	1,00	0,09
(HISEI) Sozioökonomischer Status	0,21	-0,10	0,02	1,00

Anmerkungen: HISEI = Höchster Wert in der Familie des Lehrers für den *International Socioeconomic Index of Occupational Status*; Rot gedruckte Koeffizienten sind statistisch signifikant bei $p < .05$

mit der Anpassung der Noten an die Übergangsempfehlung sowie mit der Attribution schwacher Leistungen auf mangelnde Anstrengung. Lehrkräfte, die selbst einer sozial eher begünstigten Familien entstammten, bewerteten in diesen Ländern also auch die soziale Herkunft ihrer Schüler als wichtiger für den Schulerfolg, und sie tendierten eher dazu, die Schulnoten der Kinder an „ihre“ Übertrittsempfehlung anzupassen und schulischen Misserfolg auf fehlende Anstrengung in der Schule und bei den Hausaufgaben zu attribuieren. In Ländern ohne bindende Übertrittsempfehlung fielen die Zusammenhänge nicht signifikant aus.

In einem abschließenden analytischen Schritt wurden die Auffassungen der Lehrkräfte zur Wichtigkeit des familiären/sozialen Umfelds mit den Schulleistungen und dem sozialen Hintergrund der Schülerinnen und Schüler in Verbindung gebracht. In den oben berichteten Teilstudien fanden sich Belege dafür, dass die soziale Herkunft der Schülerinnen und Schüler – auch bei vergleichbarer Schulleistung – mit der Übertrittsempfehlung assoziiert ist. Es wurde nun zusätzlich geprüft, ob diese Assoziation besonders eng in solchen Klassen ausfällt, bei denen die Lehrkraft dem familiären/sozialen Umfeld der Schülerinnen und Schüler eine besonders große Bedeutung zumisst. Technisch gesprochen, wurde hierbei eine *cross-level-Interaktion* zwischen der Lehrkräfteeinschätzung zur Wichtigkeit des familiären/sozialen Umfeld und dem sozioökonomischen Hintergrund der Schülerinnen und Schüler getestet. Die entsprechende Analyse dokumentierte jedoch keine statistisch signifikante Prädiktionskraft für den Interaktionsterm.

3.8.4. RESÜMEE

Erwartungsgemäß geben Lehrkräfte an, dass sie dem familiären/sozialen Umfeld von Schülerinnen und Schülern eine bedeutsame Rolle für den Schulerfolg beimessen. Die Auffassung der Lehrkräfte steht hierbei im Einklang mit wissenschaftlichen Befunden, die auf den substanziellen Zusammenhang von Herkunft und Bildungserfolg verweisen. Problematisch kann eine entsprechende Einstellung jedoch deshalb sein, weil sie potenziell zu einer Stabilisierung sozialer Disparitäten beitragen kann. In Hinblick auf die Reduzierung sozialer Disparitäten wird es daher auch darauf ankommen, schulische Angebote so zu gestalten, dass auch bei fehlender familiärer/sozialer Unterstützung der Schulerfolg von leistungsfähigen Schülerinnen und Schülern nicht gefährdet ist (vgl. Baumert et al., 2011). Auffällig bei der Einschätzung der Wichtigkeit des familiären/sozialen Umfelds war die relativ große Streubreite in den Antworten der Lehrkräfte.

Weitere Analysen wiesen darauf hin, dass in Ländern mit verbindlicher Übertrittsempfehlung ein sicherlich nicht erwünschter Zusammenhang zwischen eigenem familiären Hintergrund der Lehrkräfte, der Bewertung der Wichtigkeit des sozialen Umfelds sowie der Anpassung der Noten an die Übertrittsempfehlung sowie der Tendenz zur Attribution von Misserfolg auf fehlende Anstrengung bestand.

Abschließend wurde untersucht, ob sich Hinweise darauf finden lassen, dass die Einschätzung der Wichtigkeit des familiären/sozialen Umfelds mit der sozialen Herkunft der Schülerinnen und Schüler statistisch interagiert. Es fanden sich – nach Kontrolle der jeweiligen Haupteffekte – jedoch keine empirischen Belege dafür, dass die Einschätzung der Lehrkräfte zur Bedeutung des familiären/

sozialen Umfeldes – nach statistischer Kontrolle von Schulleistung und sozialem Hintergrund der Schülerinnen und Schüler – für zusätzliche sekundäre Disparitäten sorgt.

Insgesamt stehen die Befunde in Einklang mit Forderungen (vgl. Baumert et al., 2011), über entsprechende Fördergarantien in den weiterführenden Schulen die Voraussetzungen dafür zu schaffen, dass Aspekte des familiären und sozialen Hintergrunds eine zunehmend geringere Rolle bei den Übertrittsentscheidungen spielen.

3.9. Teilstudie 9: Persistierende Geschlechterdisparitäten: Verhaltensunterschiede und die Wirkung von Stereotypen

Frage 9: Werden Mädchen am Ende der Grundschule und am Ende der gymnasialen Oberstufe besser bewertet als Jungen?

3.9.1. HINTERGRUND

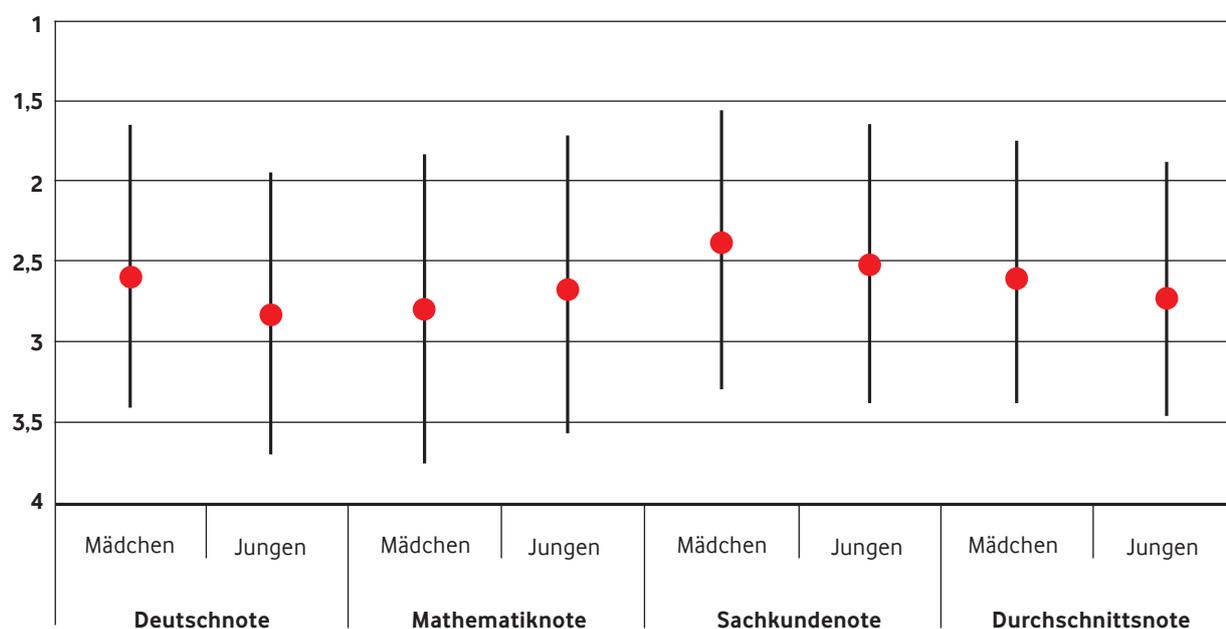
Bei Schulleistungen und schulischen Bewertungen lassen sich anhaltende und in Teilen größer werdende Geschlechterunterschiede finden. Geschlechterunterschiede lassen sich unter anderem für die Leistungen in standardisierten Leistungstests dokumentieren. So zeigen sich in Deutschland beispielsweise in PISA substanzielle Unterschiede in der Lesekompetenz zugunsten der Mädchen – diese lagen in PISA 2009 rund 40 Punkte vor den Jungen (Naumann, Artelt, Schneider & Stanat, 2010) – sowie etwas weniger deutliche Unterschiede zugunsten der Jungen in Mathematik, die sich in PISA 2009 auf 16 Punkte beliefen (Frey, Heinze, Mildner, Hochweber & Asseburg, 2010). Während die Unterschiede im Bereich des Lesens (und Schreibens) zugunsten der Mädchen recht stabil über unterschiedliche Testverfahren sind, fallen die Unterschiede in Mathematik zugunsten der Jungen vermutlich umso geringer aus, je „schulnäher“ der verwendete Leistungstest ist (vgl. Else-Quest et al., 2010). Einige Aufmerksamkeit hat zudem in jüngerer Vergangenheit die Beobachtung gefunden, dass Mädchen häufiger als Jungen eine Empfehlung für höhere Bildungslaufbahnen erhalten und bei Absolventen mit Abitur überrepräsentiert sind (im Überblick Hannover & Kessels, 2011).

Nach Kontrolle der Leistung in standardisierten Schulleistungstests erhalten Mädchen in den meisten Fächern bessere Schulnoten als Jungen. Die Disparitäten bei den

Übertrittsempfehlungen scheinen komplett durch die besseren Schulnoten in der Grundschule erklärbar zu sein und keinen „zusätzlichen“ Geschlechtereffekt bei den Empfehlungen zu repräsentieren (vgl. Ditton, 2007). In den deutschsprachigen Ländern dürfte die Überrepräsentation der Mädchen in höheren Bildungsgängen aufgrund des Schereneffekts zwischen Schulformen (vgl. Baumert, Stanat & Watermann, 2006) die Vorteile für die Mädchen in standardisierten Lesefähigkeitstests erhöhen sowie den Leistungsvorsprung der Jungen in Mathematikleistungstests reduzieren.

Welche Ursachen haben die besseren Noten der Mädchen? Inzwischen gibt es gute Belege dafür, dass Mädchen insgesamt in Hinblick auf wichtige lernrelevante Kompetenzen im Bereich der Selbststeuerung, Motivation und des Sozialverhaltens von sich selbst und anderen positiver eingeschätzt werden als Jungen. Diese Merkmale fließen auch in die – nach Kontrolle der Leistung in standardisierten Schulleistungstests – besseren Schulnoten der Mädchen ein (zusammenfassend Hannover & Kessels, 2011). Allerdings gibt es auch Hinweise darauf, dass Geschlechterunterschiede bei der Anstrengungsbereitschaft über die Schulfächer hinweg nicht einheitlich sind, sondern zwischen eher geisteswissenschaftlich und eher naturwissenschaftlich geprägten Fächern variieren (vgl. Trautwein & Lüdtke, 2009). Ursache hierfür scheinen Geschlechterunterschiede in Erwartungs- und Wertüberzeugungen zu sein, die den üblichen Geschlechterstereotypen folgen (vgl. Watt & Eccles, 2008). Dass Geschlechterstereotypen über entsprechende Anstrengungsinvestitionen einen Bruttoeffekt auf Geschlechterdisparitäten in Leistungstests, Bewertungen durch Lehrkräfte und bei beruflichen Wahlentscheidungen haben, gilt als gesichert (im Überblick Watt & Eccles, 2008). Wie groß die zusätzlichen „direkten“ Effekte von Geschlechterstereotypen bei der Beurteilung durch Lehrkräfte ausfallen, ist hingegen noch umstritten. Jungen scheinen noch immer in Hinblick auf allgemeine kognitive Fähigkeiten besser bewertet zu werden. So attribuierten in einer Studie von Tiedemann (1995) Lehrkräfte gute Leistungen bei Schülerinnen stärker auf Anstrengung und weniger stark auf Fähigkeiten als bei Jungen. In einer Studie mit dem Deutschfreiburger Übertrittsdatensatz von Trautwein und Baeriswyl (2007) wurde Jungen (bei gleicher Testleistung) von den Lehrkräften eine höhere kognitive Leistungsfähigkeit attestiert. Ob und in welchem Maße die in der großen Mehrzahl der Studien gefundenen Bewertungsvorteile für Mädchen bei der Motivation und

Abbildung 8: Mittelwerte und Standardabweichungen von Schulnoten in der Grundschule differenziert nach Geschlecht



Anmerkung: Die roten fett gedruckten Punkte bilden den Mittelwert ab, die vertikalen Linien beschreiben den Bereich, in dem rund zwei Drittel der jeweiligen Gruppe liegen (eine Standardabweichung unterhalb und oberhalb des Mittelwerts).

dem schulbezogenen Verhalten neben tatsächlichen Unterschieden auch entsprechende Stereotype reflektieren, lässt sich bislang schwer abschätzen.

In dieser Teilstudie untersuchen wir Geschlechtereffekte bei der Notenvergabe an zwei wichtigen Gelenkstellen von Bildungsbiographien: Am Ende der Grundschulzeit entscheiden Schulnoten, welche Schulform die Schülerinnen und Schüler in der Sekundarstufe I besuchen. Am Ende der gymnasialen Oberstufe stellen Schulnoten ein zentrales Selektionskriterium für den Zugang zu spezifischen Studienangeboten dar. Zusätzlich soll für den Grundschulbereich untersucht werden, ob Geschlechterunterschiede mit geschlechtsspezifischen Motivationen zusammenhängen.

3.9.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Die Daten der nachfolgenden Analysen entstammen dem am Berliner Max-Planck-Institut für Bildungsforschung durchgeführten Projekt „Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (ÜBERGANG) (Maaz, Baumert, Gresch & McElvany, 2010; Becker et al., 2010). Weitere Informationen zur Studienanlage finden sich im Anhang 1.

Die Daten der vorliegenden Untersuchung zur gymnasialen Oberstufe entstammen der im Jahr 2002 in Baden-Württemberg durchgeführten Studie „Transformation des Sekundarschulsystems und akademische Karrieren“ (TOSCA; vgl. Köller et al., 2004; vgl. auch Teilstudie 7).

3.9.3. ERGEBNISSE

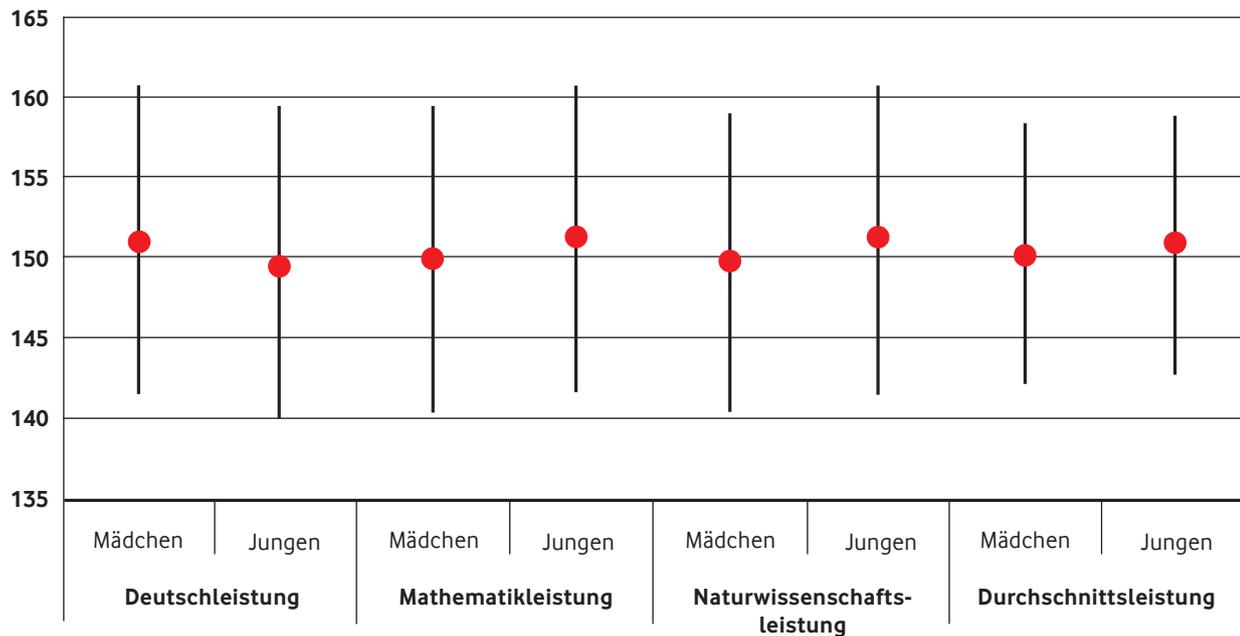
Geschlechtereffekte am Ende der Grundschulzeit

Um den Einfluss des Geschlechts der Schülerinnen und Schüler zu analysieren, sollen die in der Teilstudie 1 für den Grundschulbereich berichteten Befunde noch einmal differenziert betrachtet werden. In der deskriptiven Analyse fallen dabei zwei Befunde auf. Betrachtet man die Schulnoten von Jungen und Mädchen am Ende der Grundschulzeit, zeigen sich, bezogen auf den Notendurchschnitt, leichte Vorteile zugunsten der Mädchen. Bei den Fachnoten haben die Jungen in Mathematik leicht bessere Noten als die Mädchen. In Sachkunde und Deutsch sind wieder die Mädchen besser (vgl. Abbildung 8).

(→ **ABBILDUNG 8**)

Dieser Notenvorteil der Mädchen spiegelt sich nicht in den Leistungen, gemessen mit standardisierten Tests, wider. Zwar erzielen die Mädchen im Deutschttest bessere

Abbildung 9: Mittelwerte und Standardabweichungen der Leistungstests differenziert nach Geschlecht



Anmerkung: Die roten fett gedruckten Punkte bilden den Mittelwert ab, die vertikalen Linien beschreiben den Bereich, in dem rund zwei Drittel der jeweiligen Gruppe liegen (eine Standardabweichung unterhalb und oberhalb des Mittelwerts).

Ergebnisse als die Jungen, in Naturwissenschaft und Mathematik schneiden die Jungen leicht besser ab (vgl. auch Bonsen, Lintorf & Bos, 2008). Betrachtet man den Durchschnittstestwert aus den drei Testdomänen, erreichen die Jungen einen leicht besseren Testwert als die Mädchen.

(→ **ABBILDUNG 9**)

In Tabelle 29 sind die zentralen Befunde der ersten Teilstudie noch einmal zusammenfassend dargestellt. Bei Hinzunahme des Geschlechts in das Analysemodell zeigt sich der bereits in anderen Studien dokumentierte Befund, dass Mädchen bei gleichen schulischen Leistungen und kognitiven Grundfähigkeiten besser benotet werden als Jungen. Dieser Effekt ist insgesamt statistisch signifikant, führt aber nur zu einer geringfügigen Verbesserung der Erklärungsgüte des Modells (M3).

(→ **TABELLE 29**)

Wie in der Gesamtanalyse aus Teilstudie 2, soll untersucht werden, ob motivationale Merkmale für den beschriebenen Geschlechtereffekt verantwortlich sind. Betrachtet man die Mittelwerte der beiden motivationalen Merkmale Anstrengungsbereitschaft und Gewissenhaftigkeit, zeigen sich Unterschiede. Demnach beschreiben sich

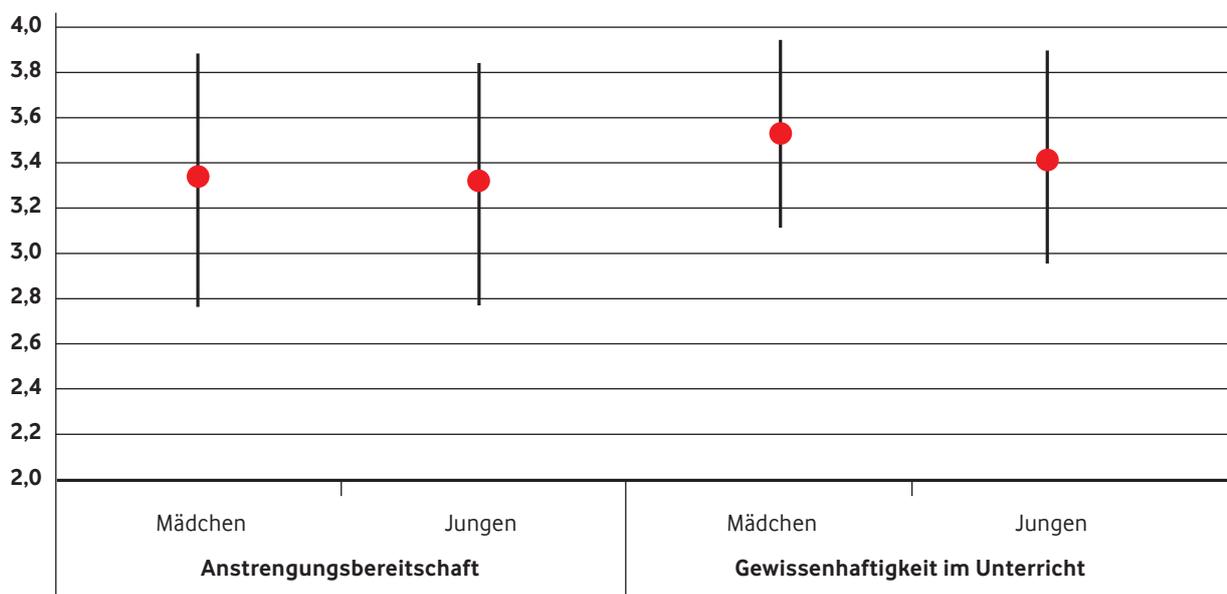
Mädchen als gewissenhafter und anstrengungsbereiter als Jungen (vgl. Abbildung 10). Die Geschlechterunterschiede in den motivationalen Merkmalen sind statistisch signifikant. Allerdings ist der Unterschied lediglich bei der Skala Gewissenhaftigkeit im Unterricht auch von praktischer Bedeutsamkeit ($d = 0,27$), bei der Skala Anstrengungsbereitschaft ist der Effekt weniger bedeutsam ($d = 0,067$).

(→ **ABBILDUNG 10**)

Wie in der Teilstudie 1 soll nun überprüft werden, ob der Geschlechtereffekt auf unterschiedliche Ausprägungen in den motivationalen Merkmalen zurückzuführen ist. Wenn dies der Fall ist, sollte sich der Geschlechtereffekt von M1 zu M2 spürbar verringern (partielle Mediation) oder nicht mehr das Signifikanzkriterium erreichen (vollständige Mediation). Der Tabelle 30 ist zu entnehmen, dass für den Geschlechtereffekt eine partielle Mediation nachgewiesen werden konnte. Hier reduziert sich der Effekt bei Hinzunahme der motivationalen Merkmale ins Modell von $b = -0,093$ auf $b = -0,056$. Zumindest ein Teil der geschlechterspezifischen Bewertung ist also vermutlich darauf zurückzuführen, dass Mädchen gewissenhafter und anstrengungsbereiter sind als Jungen.

(→ **TABELLE 30**)

Abbildung 10: Mittelwerte und Standardabweichungen für die Schülermerkmale Gewissenhaftigkeit im Unterricht und Anstrengungsbereitschaft differenziert nach Geschlecht



Anmerkung: Die roten fett gedruckten Punkte bilden den Mittelwert ab, die vertikalen Linien beschreiben den Bereich, in dem rund zwei Drittel der jeweiligen Gruppe liegen (eine Standardabweichung unterhalb und oberhalb des Mittelwerts).

Geschlechtereffekte am Ende der gymnasialen Oberstufe

Die Tabellen 31 und 32 geben zunächst einen deskriptiven Überblick über die von den Schülerinnen und Schülern in Mathematik und Englisch erreichten Testleistungen sowie die in den beiden Fächern vergebenen Fach- und Prüfungsnoten. Auf die Darstellung der Befunde der Prüfungsnoten für den Englisch Grundkurs wird aufgrund der geringen Stichprobengröße und der eingeschränkten Übertragbarkeit der Ergebnisse dieser Gruppe wie in Teilstudie 7 verzichtet.

Tabelle 31 enthält die Ergebnisse für Mathematik. Wie zu erkennen ist, erzielten die Schüler sowohl im Grund- als auch im Leistungskurs höhere Testleistungen als die Schülerinnen. Der Unterschied zwischen den Gruppen betrug im Grundkurs $d = 0,24$ Standardabweichungen und im Leistungskurs $d = 0,38$ Standardabweichungen. Die Leistungsunterschiede können als substanziell eingestuft werden und fielen auch statistisch signifikant aus. Wie Tabelle 31 weiterhin entnommen werden kann, spiegelten sich die Leistungsunterschiede nicht in den Fach- und Prüfungsnoten der Schülerinnen und Schüler wieder. So erhielten die Schülerinnen im Grundkurs bei geringerer mittlerer Leistung im Mittel höhere Noten als die Jungen, wobei der Notenunterschied bei den Fachnoten größer

ausfiel als bei den Prüfungsnoten. Im Mathematikleistungskurs waren die vorhandenen Notenunterschiede zwischen Schülerinnen und Schülern nicht statistisch signifikant, und dies, obwohl die Testleistungen der Schülerinnen deutlich unter den Leistungen der Schüler lagen.

Ein ganz ähnliches Befundmuster zeigte sich in Englisch. Auch hier fielen die Testleistungen der Schüler durchweg höher aus als die Leistungen der Schülerinnen. Der Leistungsunterschied betrug im Grundkurs $d = 0,35$ Standardabweichungen und im Leistungskurs $d = 0,19$ Standardabweichungen. Erneut korrespondierten die Leistungsunterschiede nicht mit den Fach- und Prüfungsnoten, die für die Schülerinnen im Mittel besser ausfielen als für die Schüler.

(→ TABELLEN 31 UND 32)

Die Vorhersagemodelle für die Fach- und Prüfungsnoten in den Tabellen 33 und 34 geben Auskunft darüber, wie stark die Notenunterschiede zwischen Schülerinnen und Schülern bei gleichen Testleistungen ausfielen. Im Mathematikgrundkurs (vgl. Tabelle 33) erhielten Schülerinnen bei vergleichbaren Testleistungen eine um etwa 1,5 Punkte höhere Fachnote als die Schüler. Der Bewertungsvorteil der Schülerinnen blieb unverändert, wenn zusätz-

Tabelle 29: Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum sozio-kulturellen familiären Hintergrund und zentrale Leistungsindikatoren

	M1	M2	M3	M4
	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Kombinierter Leistungswert ⁺	0,50	0,51	0,44	0,43
Kognitive Grundfähigkeit (KFT) ⁺	0,11	0,11	0,09	0,09
Geschlecht (Ref. Mädchen)				
Jungen		-0,11	-0,10	-0,10
Sozioökonomischer Status (ISEI) ⁺				
Ausbildung, schulisch (Ref. Abitur)			0,08	0,06
Hauptschulabschluss			-0,33	-0,29
Mittlere Reife			-0,11	-0,09
Fachhochschulreife			-0,09	-0,08
Ausbildung, beruflich (Ref. Uni-Abschluss)				
nicht universitär Abschlüsse			0,00	-0,01
Bücherbesitz (Ref. 26-100)				
weniger als 25				-0,08
101 bis 250				0,03
251 bis 500				0,03
Mehr als 500				0,04
Kulturelle Aktivitäten Eltern ⁺				
				0,04
<i>R</i> ²	51,00	51,40	55,00	55,40

Anmerkungen: ⁺ Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; SES = sozioökonomischer Status; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

lich zu den Testleistungen die kognitiven Fähigkeiten mit berücksichtigt wurden (vgl. Modell M2). Im Leistungskurs (vgl. Tabelle 34) waren die Bewertungsunterschiede zwischen den Schülerinnen und Schülern etwas geringer ausgeprägt als im Grundkurs, aber dennoch substanziell. Bei den Prüfungsnoten fiel der Bewertungsvorsprung der Schülerinnen der Tendenz nach niedriger aus als bei den Fachnoten.

Vergleichbare Ergebnisse ergaben sich für Englisch. Der Bewertungsvorteil zugunsten der Mädchen bewegte sich in einem Rahmen von 0,88 bis 1,14 Punkten auf der 15er-Notenmetrik und unterschied sich kaum für die Fach- und die Prüfungsnoten. Die Befunde blieben stabil, wenn zusätzlich zu den Testleistungen auch die kognitiven Grundfähigkeiten mit in die Modelle aufgenommen wurden.

(→ TABELLEN 33–36)

3.9.4. RESÜMEE

Für die Grundschule konnten Bewertungsunterschiede zwischen Mädchen und Jungen festgestellt werden. Während die Ergebnisse aus standardisierten Leistungstests leichte Vorteile der Jungen aufzeigten, erreichten die Mädchen bessere oder gleich gute Schulnoten. Allerdings differieren hier die Befunde domänenspezifisch. Bei gleichen Leistungen in den Tests wurden Jungen strenger bewertet als Mädchen. Darüber hinaus zeigte sich, dass der Geschlechtereffekt zumindest partiell durch motivationale Merkmale vermittelt war, in der Art, dass Mädchen sich als gewissenhafter und anstrengungsbereiter als Jungen beschreiben und zumindest ein Teil des Geschlechtereffekts auf motivationale Merkmale zurückzuführen war. Allerdings sollte hierbei auch kritisch bedacht werden, dass die entsprechenden Skalen Selbstberichte der Schülerinnen und Schüler darstellten und sich auch bei diesen Effekte von Geschlechterstereotypen niederschlagen könnten.

Tabelle 30: Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum Migrationsstatus, zum sozio-kulturellen familiären Hintergrund, zentraler Leistungsindikatoren sowie motivationaler Merkmale

	M1	M2
	<i>b</i>	<i>b</i>
Kombinierter Leistungswert [†]	0,43	0,40
Kognitive Grundfähigkeit (KFT) [†]	0,09	0,09
Geschlecht (<i>Ref. Mädchen</i>)		
Jungen	-0,10	-0,06
Sozioökonomischer Status (ISEI) [†]	0,06	0,05
Ausbildung, schulisch (<i>Ref. Abitur</i>)		
Hauptschulabschluss	-0,29	-0,27
Mittlere Reife	-0,09	-0,09
Fachhochschulreife	-0,08	-0,08
Ausbildung, beruflich (<i>Ref. Uni-abschluss</i>)		
nicht universitär Abschlüsse	-0,01	0,01
Bücherbesitz (<i>Ref. 26-100</i>)		
weniger als 25	-0,08	-0,08
101 bis 250	0,03	0,03
251 bis 500	0,03	0,02
Mehr als 500	0,04	0,04
Kulturelle Aktivitäten Eltern [†]	0,04	0,03
Migration (<i>Ref. Beide Eltern in Dt geboren</i>)		
beide Eltern im Ausland geboren	0,00	-0,01
Gewissenhaftigkeit im Unterricht [†]		0,13
Anstrengungsbereitschaft [†]		0,05
<i>R</i> ²	55,40	59,40

Anmerkungen: [†] Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; SES = sozioökonomischer Status; ISEI = International Socio-Economic Index of Occupational Statuts, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

Tabelle 31: Testleistungen, Fachnote Schulhalbjahr 13/1 und Prüfungsnote in Mathematik getrennt nach Grund- und Leistungskurs (Standardabweichung in Klammern)

	Grundkurs		Leistungskurs	
	Schüler	Schülerinnen	Schüler	Schülerinnen
Testleistung	475,19 (61,72)	460,49 (59,16)	576,79 (68,22)	550,96 (67,14)
Fachnote (13/1)	7,00 (3,66)	7,94 (3,63)	9,29 (3,39)	9,41 (3,09)
Prüfungsnote	7,40 (4,22)	8,08 (4,03)	10,11 (3,67)	9,84 (3,59)

Tabelle 32: Testleistungen, Fachnote Schulhalbjahr 13/1 und Prüfungsnote in Englisch getrennt nach Grund- und Leistungskurs (Standardabweichung in Klammern)

	Grundkurs		Leistungskurs	
	Schüler	Schülerinnen	Schüler	Schülerinnen
Testleistung	524,80 (47,78)	508,45 (45,08)	551,41 (46,40)	542,48 (46,72)
Fachnote (13/1)	8,68 (2,91)	8,95 (2,93)	8,79 (2,84)	9,57 (2,74)
Prüfungsnote	—	—	8,71 (3,22)	9,42 (3,24)

Die Ergebnisse zu den geschlechtsbezogenen Bewertungsunterschieden am Ende der gymnasialen Oberstufe zeigten erstmals auf Grundlage einer repräsentativen Datenbasis, dass auch beim Abitur mit zum Teil beträchtlichen Bewertungsunterschieden zwischen Schülerinnen und Schülern zu rechnen ist. So entspricht der im Mathematikgrundkurs für die Fachnoten festgestellte Bewertungsvorteil zugunsten der Schülerinnen ca. 1,5 Punkte, überführt in die herkömmliche Notenmetrik von 1 bis 6 immer noch 0,3 Notenstufen. Die Tatsache, dass beim Wettbewerb um stark nachgefragte Studienplätze zum Teil die zweite Nachkommastelle bei den Abiturnoten den Ausschlag für eine Zusage oder Ablehnung geben kann, unterstreicht die praktische Bedeutsamkeit der vorgefundenen Bewertungsunterschiede. Die Geschlechtsunterschiede in der Benotung fielen in Mathematik bei den Prüfungsnoten etwas niedriger aus als bei den Fachnoten, in Englisch zeigten sich diesbezüglich kaum Unterschiede. Insgesamt scheinen jedoch die Prüfungsnoten aus den Zentralabiturprüfungen ähnlich anfällig für geschlechtsbezogene Bewertungsunterschiede wie die Fachnoten.

Einschränkungen unserer Befunde ergeben sich zum einen dahingehend, als dass sich die Ergebnisse nur auf die von uns untersuchten Fächer beziehen und nicht ohne Weiteres auf andere Fächer übertragbar sind. Zum anderen liefert die Untersuchung keine Informationen über die den Bewertungsunterschieden zugrunde liegenden Prozessen. Existieren beispielsweise notenrelevante Unterschiede – etwa in Bezug auf die Beteiligung am Unterrichtsgeschehen und die Sorgfalt in der Aufgabenerledigung – die durch die Leistungstests nicht mit erfasst wurden und evtl. einen Teil der Bewertungsunterschiede zwischen den Schülerinnen und Schülern erklären könnten? Im Falle der Mathematik könnten zudem Stereotype der Lehrkräfte hinsichtlich der Leistungserwartung an Jungen und Mädchen eine Rolle spielen, wonach Mädchen in Mathematik üblicherweise eine geringere Leistungsfähigkeit zugeschrieben wird und entsprechend bei gleichen Leistungen ein Bewertungsbonus zugunsten der Schülerinnen vergeben wird. Hier besteht nach wie vor erheblicher Forschungsbedarf, insbesondere mit Blick auf die Leistungsbewertung in der gymnasialen Oberstufe.

Tabelle 33: Vorhersage der Fach- und Prüfungsnote in Mathematik durch Testleistung, Geschlecht und kognitive Grundfähigkeit im Grundkurs

	Fachnote (13/1)		Prüfungsnote	
	M1	M2	M1	M2
Intercept Fachnote Mathematik	7,78	7,79	8,43	8,44
Testleistung Mathematik ⁺	2,50	2,56	3,34	3,41
Geschlecht (<i>Ref. Jungen</i>)				
Mädchen	1,48	1,47	1,38	1,37
Kognitive Grundfähigkeit (KFT) ⁺		-0,14		-0,16
R^2	29,0	29,1	39,7	39,8

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; R^2 = Multipler Determinationskoeffizient.

Tabelle 34: Vorhersage der Fach- und Prüfungsnote in Mathematik durch Testleistung, Geschlecht und kognitive Grundfähigkeit im Leistungskurs

	Fachnote (13/1)		Prüfungsnote	
	M1	M2	M1	M2
Intercept Fachnote Mathematik	7,14	7,11	7,54	7,52
Testleistung Mathematik ⁺	2,29	2,39	2,68	2,74
Geschlecht (<i>Ref. Jungen</i>)				
Mädchen	0,90	0,93	0,68	0,70
Kognitive Grundfähigkeit (KFT) ⁺		-0,21		-0,11
R^2	35,0	35,2	38,5	38,6

Anmerkungen: ⁺ Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; R^2 = Multipler Determinationskoeffizient.

Tabelle 35: Vorhersage der Fachnote in Englisch durch Testleistung, Geschlecht und kognitive Grundfähigkeit im Grundkurs

	Fachnote (13/1)	
	M1	M2
Intercept Fachnote Englisch	8,68	8,68
Testleistung Englisch [†]	1,74	1,73
Geschlecht (<i>Ref. Jungen</i>)		
Mädchen	0,88	0,89
Kognitive Grundfähigkeit (KFT) [†]		0,02
R^2	29,0	29,0

Anmerkungen: [†] Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; R^2 = Multipler Determinationskoeffizient.

Tabelle 36: Vorhersage der Fach- und Prüfungsnote in Englisch durch Testleistung, Geschlecht und kognitive Grundfähigkeit im Leistungskurs

	Fachnote (13/1)		Prüfungsnote	
	M1	M2	M1	M2
Intercept Fachnote Englisch	7,93	7,92	7,62	7,60
Testleistung Englisch [†]	1,76	1,79	2,10	2,14
Geschlecht (<i>Ref. Jungen</i>)				
Mädchen	1,08	1,07	1,14	1,12
Kognitive Grundfähigkeit (KFT) [†]		-0,10		-0,16
R^2	34,5	34,6	35,3	35,5

Anmerkungen: [†] Variable wurde z-standardisiert (M = 0, SD = 1); **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < 0,05$; R^2 = Multipler Determinationskoeffizient.

3.10. Teilstudie 10: Wirkungen primärer und sekundärer Herkunftseffekte sowie von Merkmalen der Lehrperson auf die Benotung schulischer Leistungen

Frage 10: Welchen Einfluss haben sozialschichtspezifische Merkmale der Schülerinnen und Schüler auf die Benotungspraxen der Lehrpersonen? Inwiefern ist bei der Benotung ein klassenspezifischer Referenzgruppeneffekt festzustellen? Beeinflussen Merkmale der Lehrperson, wie Geschlecht und Unterrichtserfahrung, die Notengebung?

3.10.1. HINTERGRUND

Studien jüngerer Datums (u. a. Becker, 2010; Ditton und Krüsken, 2010; Ehmke und Jude, 2010; Klein et al. 2010; Maaz, Baumert, Trautwein, 2010; Maaz und Nagy, 2010) haben immer noch Hinweise auf eine systematische Bevorteilung bzw. Benachteiligung bestimmter Schülergruppen geliefert, die in Konflikt mit meritokratischen Grundsätzen stehen. So fanden sich Belege dafür, dass Effekte des sozialen Hintergrunds sowie der mittleren Leistungsstärke einer Schulklasse wirksam werden (vgl. Baumert et al., 2009). Ebenfalls konnte gezeigt werden, dass neben Referenzgruppeneffekten (bei gleicher individueller Leistung fiel die Übertrittsempfehlung in leistungsstarken Klassen niedriger aus) auch Unterschiede in substantieller Größe zwischen einzelnen Schulen und Lehrkräften auftreten (Baeriswyl, Wandeler & Trautwein, 2011).

Soziale Gerechtigkeit bei der Beurteilung von Schulleistungen und individuellen Fähigkeiten gehört aber zum Grundrecht in einem demokratisch geordneten Staat. Die Vergabe von Schulungs-, Bildungs- und Berufschancen müssen unabhängig sein von sozialer Herkunft, sozioökonomischem Hintergrund der Eltern sowie von deren Kultur- oder Bildungsnähe. Jede Demokratie ist bemüht, diese Chancengleichheit herzustellen. Trotz diesen Bemühungen der Staatsschule um Bildungs- und Chancengerechtigkeit zeigen diese neuesten Studien, dass Herkunftseffekte stabil sind. Familiären Sozialisationsprozessen vor und während der Schulzeit kommt in Hinblick auf soziale Disparitäten eine bedeutsame Rolle zu, da das heimische Lernumfeld und die Unterstützung durch Eltern einen beträchtlichen Anteil der individuellen Lerngeschichte ausmachen.

PISA 2000 hat gezeigt, dass das schweizerische Schulsystem mit diesen sozialen Disparitäten nicht besonders gut umgehen kann. Schülerinnen und Schüler mit einem

ISEI-Index im untersten Quartil hatten eine der größten Wahrscheinlichkeiten aller untersuchten OECD-Länder, auch in der Gesamtskala der Leseleistung im untersten Quartil zu liegen (OECD, 2001, S. 327). Ramseier und Brühwiler (2003) kamen nach einer Detailanalyse der PISA Daten zum Schluss, dass die soziale Herkunft bzw. die Bildungsnähe der Eltern für die Lese- und Mathematikleistung von großer Bedeutung ist. Dieser Effekt bleibt auch bei Kontrolle der kognitiven Grundfähigkeiten und der kulturellen Herkunft als ein eigenständiger Effekt bestehen (Ramseier & Brühwiler, 2003, S. 39). Nach ihnen ist dieser Effekt besonders stark durch den Schultyp vermittelt. Dies könnte ein Hinweis dafür sein, dass soziale Herkunftseffekte beim Übertritt von der Primar- in die Sekundarstufe I entstehen.

Moser und Berweger (2005) stellten zudem sehr große Unterschiede dieser Effekte zwischen den kantonalen Schulsystemen fest. Große soziale Herkunftseffekte gingen mit hoch selektiven Aufnahmeverfahren einher. „Die Ergebnisse im Kanton Freiburg, Jura und Wallis bestätigen zudem, dass die Maximierung der durchschnittlichen Mathematikkompetenzen und die Verminderung schulischer Segregation zwei Anliegen sind, die sich auch in der Schweiz nicht ausschließen“ (Moser und Berweger, 2005, S. 117). Baeriswyl et al. (2006) stellten für den deutschsprachigen Kantonsteil Freiburg zwar einen signifikanten, aber vergleichsweise geringen sekundären Effekt des sozioökonomischen Status der Eltern auf den Zuweisungsentscheid für eine Abteilung der Sekundarstufe I fest. Sie schrieben diesen „Teilerfolg“ einem multikriterialen Übertrittsverfahren zu. Verlaufsstudien der PISA-Untersuchungen von 2000 bis 2009 (Ehmke und Jude, 2010) zeigen, dass sich das Problem mit dem Umgang sozialer Disparitäten in der Schweiz etwas gemildert darstellt, aber noch keinesfalls zufriedenstellend bewältigt worden ist.

Wenig untersucht ist die Wirkung der sozialen Herkunft auf die Notengebung im Schulalltag. Viele Übertrittsverfahren berücksichtigen die Fachnoten oder die Notendurchschnitte als Entscheidungselement für die Zuweisung in die gegliederte Sekundarstufe I.

Roeder et al. (1986) berichten eine geringe Bedeutung der sozialen Herkunft der Schülerinnen und Schüler für die Zensurengebung in der Schule. „Gemessen am Bildungsstatus beider Eltern erklärt sie nur etwa ein halbes Prozent der Variation der Schulzensuren in jedem der drei Fächer (geschätzt als Quadrat der Produkt-Moment-Korrelation zum zweiten Testzeitpunkt am Ende des 7.

Schuljahrs).“ (S. 46). Ebenso berichten sie zwar eine beträchtliche Varianz der Variable „Sozialstatus der Schüler“, jedoch nur eine geringe Aufklärungskraft der Fachnoten in einem Regressionsmodell. Diese Daten erklären Roeder et al. (1986) mit einer Art „survival of the fittest“: „da die Stichprobe im Hinblick auf den Sozialstatus der Schüler durchaus eine beträchtliche Varianz hat, ... ist dieses Ergebnis wohl so zu verstehen, dass die wesentlichen schichtspezifischen Einflüsse auf die Schullaufbahn sich in der Grundschule und wahrscheinlich in den ersten beiden Schuljahren im Gymnasium geltend gemacht haben. Schüler aus Familien mit niedrigem Bildungsstatus, die das 7. Schuljahr in der höheren Schule erreicht haben, unterscheiden sich in ihren Leistungen in den drei Hauptfächern nicht mehr von ihren Mitschülern.“ (S. 46). Der bei Roeder et al. dokumentierte geringe Einfluss des sozioökonomischen Hintergrundes ist präziser ausgedrückt ein kleiner sekundärer Einfluss des sozioökonomischen Hintergrundes. Ihre Argumentation läuft darauf hinaus, dass durch die Selektion durch das Schulsystem die primären Effekte bereits ihre volle Wirkung ausgeübt haben, da die Leistungsschwachen aus tieferen Schichten bereits weg selektioniert wurden, ohne dass die Spannweite an sozioökonomischen Hintergründen extrem eingeschränkt wurde. Der Studie mangelt es an einem objektiveren Leistungsindikator als es die Noten sind, von daher bleibt es unklar, inwiefern primäre und sekundäre Effekte (vgl. dazu Kap. 2.3) auftreten. Allein auf Grund der Note gibt es nach ihnen keine Bevorteilung von sozialen Schichten (Roeder et al., 1986).

In dieser letzten Teilstudie untersuchen wir nun die mögliche Wirkung des sozioökonomischen und kulturellen Hintergrundes auf die Benotung. Dabei werden primäre und sekundäre Herkunftseffekte getrennt. Weiter wird ein möglicher Bezugsgruppeneffekt für die Klassenleistung bei einem Fachleistungstest untersucht. Zudem beschäftigt die Frage, inwiefern Berufsmerkmale der Lehrperson sekundäre Herkunftseffekte beeinflussen. Durch die getrennte Berücksichtigung der Fachbereiche Mathematik und Deutsch wird auch die Stabilität der Effekte überprüft.

Dabei werden primäre Herkunftseffekte auf die Noten, von direkten Einflüssen des sozialen Hintergrundes auf die Noten, getrennt. Diese direkten Einflüsse des Hintergrundes auf die Noten können als Teil der sekundären Herkunftseffekte gesehen werden (vgl. die indirekten sekundären Herkunftseffekte in Teilstudie 3). Weiter wird

untersucht ob das Leistungsniveau der Klasse einen Effekt auf Noten der einzelnen Schülerinnen und Schüler hat. Zudem beschäftigt die Frage, inwiefern Berufsmerkmale der Lehrperson sekundäre Herkunftseffekte beeinflussen.

3.10.2. DATENGRUNDLAGE UND ANALYSESTRATEGIE

Die Daten der nachfolgenden Analysen entstammen der Deutschfreiburger Studie (deutschsprachiger Kantonsteil Freiburg/Schweiz) zum Übertrittsverfahren von der Primarschule in die Sekundarstufe I (vgl. Baeriswyl et al. 2006; Zum Schweizer Schulsystem vgl. Anhang 4). Jährlich werden standardmäßig von jedem Kind die Zeugnisnoten des ersten Semesters der sechsten Primarklasse, die Einschätzung der Lern- und Arbeitsmotivation, sowie der kognitiven Fähigkeiten durch die Lehrperson, die Angaben zur Inanspruchnahme der Schuldienste (z.B. Logopädie), die globale Zuweisungsempfehlung der Lehrperson und der Eltern, sowie die Ergebnisse der Leistungsprüfung in Deutsch und Mathematik erhoben.

Die Übertrittsjahrgänge 2009 und 2010 wurden zum regulären Verfahren zusätzlich ausführlich befragt. Von den Schülerinnen und Schülern wurden Migrationshintergrund, gesprochene Sprache in der Familie, kulturelle und sozioökonomische Hintergrundsvariablen, ihr Verhalten bei den Hausaufgaben, aber auch die wahrgenommene Vorbereitungsgüte auf die Prüfung, die wahrgenommene Gerechtigkeit in der Klasse und bei der Vorbereitung erfragt. Die Befragungsinstrumente blieben während den beiden Jahren mit Ausnahme von zwei ergänzenden Fragen unverändert. Gleichzeitig zur Elternbefragung wurde den Lehrpersonen ein zweiter Fragebogen zur retrospektiven Einschätzung des Übertrittsverfahrens abgegeben.

(→ TABELLE 37)

Sozioökonomische Stellung der Eltern. Zur Beschreibung der sozioökonomischen Stellung dient der auf den elterlichen Berufsangaben basierende Internationale Sozioökonomische Index (ISEI), den Ganzeboom, DeGraaf, Treiman & De Leeuw (1992) entwickelt haben. Für die Analysen wurde der höchste sozioökonomische Index von Vater bzw. Mutter gewählt. Für die Erfassung der kulturellen Praxis und die Anzahl der Bücher wurden die Items der PISA-Studie verwendet.

Familiensprache: Die Familiensprache wurde bei den Schülerinnen und Schülern erfragt. Bei der Familiensprache wurde für die vorliegenden Analysen zwischen Landessprachen (Deutsch, Französisch, Italienisch) und an-

Tabelle 37: Deskriptive Angaben zu den Erhebungen der Übertrittsjahrgänge 2009 und 2010

	2009	2010
Anzahl Schülerinnen und Schüler	824	861
davon weiblich	425 (51,6%)	436 (50,6%)
davon männlich	399 (48,4%)	425 (49,4%)
Anzahl Lehrpersonen	50	49
davon weiblich	34 (68%)	33 (67%)
davon männlich	16 (32%)	16 (33%)

deren Sprachen unterschieden. Familiensprache Deutsch und Landessprachen wurden mit 1 kodiert; 0 entspricht einer Familiensprache, die nicht gleich der Unterrichtssprache ist.

Noten: Die von den Lehrpersonen vergebenen Noten wurden von den Lehrpersonen im Rahmen der Standarddaten zum Übertrittsverfahren für jedes Kind berichtet. Es sind dies die Zeugnisnoten am Ende des ersten Semesters der sechsten Klasse. Dabei handelt es sich um eine in Halbnotenschritten abgestufte Skala von 6 bis 1: sehr gut (6), sehr gut – gut (5,5), gut (5), gut-genügend (4,5), genügend (4), ungenügend (3,5), (3), (2,5), (2), (1,5), (1). Mittelwerte der Noten für die Fächer Mathematik, Deutsch, Französisch und Musik werden in Tabelle 38 berichtet. Die Unterschiede zwischen den beiden Jahren erwiesen sich als statistisch nicht signifikant. Für die Noten wurde eine jahrgangswise Standardisierung vorgenommen. Dadurch werden auch kleinere Unterschiede in der Notengebung berücksichtigt und eine Interpretation der Befunde in der Metrik standardisierter Regressionskoeffizienten ermöglicht.

Standardisierter Leistungstest: Im Rahmen des Übertrittsverfahrens wird ein standardisierter Mathematik- und Deutschttest administriert. In beiden Testteilen (Deutsch und Mathematik) können maximal 50 Punkte erzielt werden. Der Mathematiktest besteht aus den vier Subtests Grundoperationen, Textaufgaben, Brüche sowie Geometrie und der Deutschttest aus den Subtests Textschaffen, Textverständnis und Rechtschreibung. Die Aufgabeninhalte beziehen sich auf den Lehrplan der fünften und einen Teil der sechsten Klasse. Die Prüfungsaufgaben werden von Lehrpersonen der Sekundarstufe I und von nicht am Übertritt beteiligten Primarlehrpersonen auf der Basis der verbindlichen Lehrpläne erarbeitet. Zudem werden die Tests in einigen Klassen eines anderen Kantons, in welchem dieselben Lehrpläne gelten, validiert.

Die Teilnahme am Leistungstest ist in der Regel für alle Schülerinnen und Schüler verpflichtend. Dispensiert werden heilpädagogisch begleitete Schülerinnen und Schüler mit großen Entwicklungsrückständen. Der Leistungstest findet in allen Schulen am gleichen Tag statt. Er wird in den Klassenzimmern der Primarschule durchgeführt und von einer Lehrperson der Sekundarstufe I überwacht. Die Prüfungskorrekturen werden ebenfalls von Lehrpersonen der Sekundarstufe I durchgeführt, um die Güte der Korrekturen zu garantieren. Die Testergebnisse werden für ganz Deutschfreiburg zentral ausgewertet.

Für die vorliegende Studie wurden die Prüfungswerte in Mathematik und Deutsch als standardisierte Leistungsindikatoren verwendet. In Mathematik lag der Mittelwert im Jahr 2009 bei $M = 30,80$ ($SD = 8,01$) und 2010 bei $M = 27,95$ ($SD = 7,96$). In der Deutschprüfung lagen die Werte 2009 bei $M = 34,91$ ($SD = 7,90$) und 2010 bei $M = 32,83$ ($SD = 8,12$) (vgl. Tabelle 38). Da der Leistungstest in jedem Jahr neu konstruiert wird, kommt es zu einer natürlichen Schwankung. Diese fiel für die beobachteten Jahrgänge in der Mathematikprüfung ($F(1, 1669) = 53,22$, $p < 0,001$, partielles $\eta^2 = 0,03$) wie auch in der Deutschprüfung ($F(1, 1671) = 28,07$, $p < 0,001$, partielles $\eta^2 = 0,02$) statistisch signifikant aus. Um diese, für die vorliegende Fragestellung unerwünschte Variation auszuschalten und darüber hinaus einfacher interpretierbare Koeffizienten in der Mehrebenenanalyse zu erhalten, wurden die Leistungswerte jahrgangswise z-standardisiert. Der Testwert im Leistungstest wurde sowohl als Individualvariable als auch – in pro Klasse aggregierter und nicht erneut standardisierter Form – als Klassenvariable verwendet.

Lehrermerkmale: Neben dem Geschlecht, der Anzahl Unterrichtsjahre und der Anzahl der begleiteten Übertrittsverfahren wurde auch die Selbstwirksamkeitsüberzeugung der Lehrpersonen erfasst (*Cronbach's alpha* = 0,72). (→ TABELLE 38)

Tabelle 38: Vergleich der Noten- und Prüfungsmittelwerte

		2009	2010	F	sig
Mathematiknote	M	4,99	4,96	0,687	0,407
	SD	0,61	0,62		
Deutschnote	M	4,92	4,87	3,364	0,067
	SD	0,51	0,52		
Französisch	M	5,11	5,08	0,947	0,331
	SD	0,60	0,63		
Musik	M	5,33	5,36	1,313	0,252
	SD	0,51	0,53		
Mathe Prüfung	M	30,80	27,95	53,215	0,000
	SD	8,01	7,96		
Deutsch Prüfung	M	34,91	32,83	28,07	0,000
	SD	7,90	8,12		

In Untersuchungen mit Schulklassen ist die hierarchische Struktur der gewonnenen Daten zu berücksichtigen, um mögliche Fehlschlüsse bei der Analyse zu vermeiden (vgl. Raudenbush & Bryk, 2002). Ein adäquates Auswertungsverfahren für Daten mit hierarchischer Struktur stellt die Mehrebenenanalyse dar, eine allgemeine Form der Regressionsanalyse, die es erlaubt, simultan Prädiktoren auf mehreren Ebenen (z.B. Individualebene, Klassenebene) zu berücksichtigen und Standardfehler der Regressionskoeffizienten konservativ zu schätzen. Alle berichteten Modelle sind sogenannte *random intercept* Modelle, die über das *full information maximum likelihood* Verfahren mit dem Programmpaket HLM 6 (Raudenbush, Bryk, Cheong & Congdon, 2004) geschätzt wurden. Das verwendete Softwarepaket für Mehrebenenanalysen (HLM 6.04, Raudenbush, Bryk, Cheong & Congdon, 2004) bietet allein unstandardisierte Regressionskoeffizienten in den Ergebnisausdrucken an, was die Interpretation der Regressionsgewichte aufgrund der oftmals arbiträren Metriken der Prädiktoren und Kriterien erschwert. Um die spätere Interpretation etwas zu vereinfachen, wurden im vorliegenden Fall alle metrischen Variablen standardisiert ($M = 0$, $SD = 1$). Durch diese sogenannte Y-Standardisierung lassen sich die Effekte der Prädiktorvariablen einfacher interpretieren. Die Regressionskoeffizienten zeigen dann an, um welchen Anteil einer Standardabweichung die abhängigen Variablen bei der Zunahme/Abnahme um eine Einheit bei den Prädiktorvariablen zunehmen/abnehmen. Die Daten der beiden Jahrgänge wurden jahrgangsweise standardisiert und dann in jahrgangsübergreifenden Modellen in den Fachbereichen Mathematik und Deutsch überprüft. Die Noten in Musik und Französisch wurden nur in der Analyse zur Zerlegung der Varianzen verwen-

det und werden wegen fehlender Leistungswerte in den weiteren Fragestellungen nicht mehr berücksichtigt.

3.10.3. ERGEBNISSE

Varianzen auf Klassenebene:

Gibt es fachspezifische Unterschiede?

In den ersten Regressionsmodellen, den sogenannten Nullmodellen, wurden noch keine Prädiktoren berücksichtigt. Diese Modelle geben Auskunft darüber, wie sich die Gesamtvarianz der Noten in den Fächern Mathematik, Deutsch, Musik und Französisch auf die Individual- und Klassenebene verteilte (vgl. Tabelle 39). Es zeigte sich, dass in Mathematik und Deutsch jeweils 94 und 95 Prozent der Varianz in den Noten auf Schülerebene lagen und etwa 5-6 Prozent auf der Klassenebene. Bei den Musiknoten war die Varianz auf Klassenebene mit jeweils 19 Prozent deutlich stärker ausgeprägt, während die Klassenvarianz in Französisch mit 2-5 Prozent geringer ausfiel.

(→ TABELLE 39)

Die Klassenvarianzanteile der Mathematik-, Deutsch- und Französischnoten fielen, außer für Französisch 2010, ähnlich groß aus. Vor allem unterschied sich ihr Anteil zu jenem der Musiknote. Auffallend ist, dass die Fächer mit engeren Vorgaben durch den Lehrplan und das Lehrbuch wesentlich weniger Klassenvarianz aufweisen.

Eine erste Erkenntnis aus dem Vergleich der Varianzerlegungen ist, dass der Anteil der Klassenvarianz auch stark vom Fach abhängig ist. Außer in Französisch sind die jeweiligen Varianzanteile in den beiden Kohorten sehr ähnlich. Diese geringen Schwankungen sind ein Indiz dafür, dass die Größenordnung dieser fachspezifischen Variation auf Klassenebene relativ stabil ist.

Tabelle 39: Varianzkomponenten der Noten in verschiedenen Fächern

Jahr	Mathematik		Deutsch		Musik		Französisch	
	2009	2010	2009	2010	2009	2010	2009	2010
Klassenebene	0,05	0,06	0,06	0,05	0,19	0,19	0,05	0,02
Schülerebene	0,95	0,94	0,94	0,95	0,81	0,81	0,95	0,98

Einfluss des sozialen Hintergrundes (primärer und sekundärer Effekt)

In den nächsten Modellserien wurde die Beziehung des sozialen Hintergrundes mit den Noten in den Fächern Mathematik und Deutsch analysiert (vgl. Tabelle 40). Im Modell 1 wurde jeweils eine Reihe von Indikatoren des sozialen Hintergrundes als Prädiktorvariablen berücksichtigt. Dabei handelte es sich um den totalen Effekt des sozialen Hintergrundes (d.h. inkrementeller primärer und sekundärer Effekt). Die Landessprache, der sozioökonomische Status und der Bücherbestand spielten sowohl in Mathematik als auch in Deutsch eine Rolle, während die kulturellen Aktivitäten der Eltern keinen zusätzlichen Einfluss hatten. Beim Migrationsstatus zeigte sich, dass für die Mathematiknote nur Kinder mit einer im Ausland geborenen Mutter eine tiefere Mathematiknote hatten, während in Deutsch die Geburt von Vater, Mutter oder beiden Elternteilen im Ausland einen Zusammenhang mit tieferen Noten hatte. Während 11 Prozent der Varianz der Mathematiknote durch den sozialen Hintergrund erklärt werden konnte, waren es bei der Deutschnote 17 Prozent. Dieser Unterschied deutet darauf hin, dass die Deutschnote enger mit dem Bücherstand und dem Migrationshintergrund zusammenhing als die Mathematiknote.

Im Modell 2 wurde zusätzlich das Individualmerkmal Geschlecht berücksichtigt. Geschlechtseffekte traten in Mathematik keine auf, in Deutsch hatten die Mädchen bessere Noten. Die Berücksichtigung des Geschlechts hatte die Effekte der Indikatoren des sozialen Hintergrundes nicht bedeutend verändert.

Im Modell 3 wurde die individuelle Testleistung je in Mathematik und Deutsch kontrolliert. Dieses standardisierte Leistungsmaß dient als „neutraler“ Maßstab für die

Leistung. So können Kinder aus verschiedenen Klassen miteinander verglichen werden und man kann vergleichen ob Kinder mit den gleichen Testwerten auch gleiche Noten erhalten. In der Tat bestand in beiden Fächern ein enger Zusammenhang zwischen Testleistung und den Noten ($B = 0,85$ und $0,84$). Einem Leistungsvorsprung von einer Standardabweichung in den Prüfungen entsprach also durchschnittlich eine um etwa 0,85 Standardabweichungen besser ausgefallene Note (also rund vier Fünftel einer Standardabweichung).

Zudem gibt dieses Modell 3 Auskunft darüber, wie stark die Indikatoren des sozialen Hintergrundes über die individuellen Testleistungen hinaus mit den durch die Lehrperson vergebenen Noten assoziiert waren. Unter der Annahme, dass alle primären Einflüsse des sozialen Hintergrundes auf die Note vollständig in der Testleistung absorbiert sind, wären die übrig bleibenden Effekte des sozialen Hintergrundes sogenannte sekundäre Effekte. Ebenfalls wären alle primären Effekte der Landessprache und des Geschlechtes auf die Leistungsfähigkeit über die Testleistung mediiert.

Die Berücksichtigung der Testleistung erhöhte die Varianzaufklärung der Note des Individuums um fast 40 bzw. 50 Prozentwertpunkte. Durch die Kontrolle der Leistung wurden die Effekte des sozialen Hintergrundes stark reduziert. Ein Großteil der Effekte des sozialen Hintergrundes auf die Notengebung waren also primäre Effekte. Der Effekt der Landessprache wurde ausgeschaltet. Der Effekt des sozioökonomischen Status wurde drastisch reduziert, es blieb aber noch ein kleiner signifikanter Effekt. In Mathematik und Deutsch verblieb nur ein Effekt der Anzahl Bücher, nämlich die Gruppe „101-200 Bücher“ hatte höhere Noten als die Referenzgruppe. Der negative Effekt des

Migrationsstatus der Mutter wurde in Mathematik stark vermindert, blieb aber bestehen. In Deutsch wurden die Effekte des Migrationsstatus ebenfalls reduziert, aber nur der Effekt der beiden im Ausland geborenen Elternteile wurde ausgeschaltet. Kinder mit einer im Ausland geborenen Mutter oder mit einem im Ausland geborenen Vater hatten schlechtere Deutschnoten.

In Bezug auf das Geschlecht der Schülerinnen und Schüler war auffällig, dass bei gleichen Testleistungen die Jungen eine um mehr als eine Zehntel Standardabweichung schlechtere Mathematiknote erhalten. Die besseren Mathematiknoten der Mädchen können also nicht vollständig durch eine höhere Leistungsfähigkeit erklärt werden. Zu berücksichtigen ist, dass in den berichteten Analysen keine Anstrengungsindikatoren einbezogen wurden.

(→ TABELLE 40)

Einfluss der Leistungsstärke der Klasse

In der nächsten Modellserie (vgl. Tabelle 41) wurde nun zusätzlich zu den bisherigen auf der Schülerschichtebene berücksichtigten Variablen die durchschnittliche Testleistung auf Klassenebene eingefügt. Mit diesem Modell lässt sich prüfen, inwiefern bei der Benotung Referenzgruppeneffekte auftreten. D.h. inwiefern ist die Note eines Individuums durch die Leistungsfähigkeit der Klasse beeinflusst?

Der Referenzgruppeneffekt trat in Mathematik und Deutsch auf. In Klassen, deren mittlere Testleistung in Mathematik um eine Standardabweichung über dem Mittelwert lag (in der Metrik der individuellen Schülerleistung), erhielten die Schüler – nach Kontrolle ihrer individuellen Leistungen – eine im Mittel um .31 bis .37 Standardabweichung schlechtere Note. Dieser Befund ist in Übereinstimmung mit ähnlich gelagerten Ergebnissen einer Reihe von Studien der jüngeren Vergangenheit. Die Berücksichtigung der Leistungsstärke der Klasse erhöhte die Varianzaufklärung der Note des Individuums zwar nur um maximal 1%, die Höhe des Regressionskoeffizienten zeigte jedoch auf, dass der Klassenkontext bei der Notengebung als Referenzgruppe eine bedeutsame Rolle spielt.

(→ TABELLE 41)

Einfluss von Lehrpersonenmerkmalen

In der nächsten Modellserie (vgl. Tabelle 41) wurden nun auf der Klassenebene zusätzlich zur durchschnittlichen Testleistung auch folgende Lehrpersonenmerkmale eingefügt: das Geschlecht der Lehrperson, die Anzahl der

Jahre Unterrichtserfahrung, die Anzahl begleiteter Übertrittsverfahren und die Selbstwirksamkeit der Lehrperson. Keines der Lehrpersonenmerkmale hatte einen statistisch signifikanten Effekt.

3.10.4. RESÜMEE

Eine erste Erkenntnis aus dem Vergleich der Varianzzerlegungen war, dass der Anteil an Klassenvarianz stark vom Fach abhängig war. Fächer mit engerer Bindung an den Lehrplan und an das Lehrbuch wiesen eine wesentlich geringere Varianz zwischen den Klassen auf. Die Note im Fach Musik hingegen war sehr stark von der Lehrperson, respektive von der Klasse, abhängig.

Der primäre Herkunftseffekt war fachübergreifend stabil und äußerst bedeutsam. Der sozio-ökonomische Hintergrund des Kindes wirkte sich vorwiegend auf die objektiv erfassten Leistungen in Mathematik und Deutsch aus. Dazu hatten die Landessprache, der Bücherbestand und der Migrationsstatus bedeutsame Einflüsse, während die elternberichteten kulturellen Aktivitäten sich nicht als zusätzliche prädiktive Faktoren für die Noten erwiesen. Die besseren oder schlechteren Noten von Schülerinnen und Schüler in Abhängigkeit des sozialen Hintergrundes hingen demnach mit tatsächlich höheren oder tieferen Leistungen zusammen. So erbrachten zum Beispiel Schülerinnen und Schüler aus einem sozioökonomisch tieferen Familienmilieu tiefere Leistungen im Mathematik- und Deutschtest. Der größte Teil der Effekte des sozialen Hintergrundes auf die Notengebung waren primäre Effekte (Effekte verbunden mit der Leistungsfähigkeit).

Die sekundären Effekte des sozialen Hintergrundes, also ein Einfluss der Herkunft bei gleicher Leistung, waren eher schwach ausgeprägt und in den Fächern Mathematik und Deutsch ähnlich. Die Landessprache spielte keine direkte Rolle mehr, jedoch hatte der sozioökonomische Status und der Migrationsstatus direkte Effekte. Beim Migrationsstatus war überraschend, dass zwischen jenem des Vaters und dem der Mutter differenziert werden muss. Wurde die Mutter im Ausland geboren, waren die Noten sowohl in Deutsch wie in Mathematik schlechter. Der Migrationshintergrund des Vaters zeigte sich jedoch nur für die Deutschnote. Diese Differenzierung könnte eine Spekulation zulassen, dass der Migrationsstatus der Mutter mit (geringeren oder fehlenden) Hilfestellungen bei den Hausaufgaben etwas zu tun haben könnte, die sich in beiden Fächern negativ auswirkt. Der Migrationshintergrund des Vaters dürfte stärker mit der zu Hause gesprochenen

Tabelle 40: Vorhersage der Mathematik- und Deutschnoten durch individuelle Schülermerkmale zum sozio-kulturellen familiären Hintergrund, Geschlecht und Testleistung

	Mathematik			Deutsch		
	M1	M2	M3	M1	M2	M3
	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Landessprache (=1; 0=nicht)	0,32	0,32	0,14	0,28	0,30	0,06
Sozioökonomischer Status (ISEI) [†]	0,22	0,22	0,04	0,22	0,22	0,07
Kulturelle Aktivitäten	0,02	0,02	0,02	0,04	0,03	0,01
Bücherbesitz (Ref. 26-100)						
weniger als 25	-0,26	-0,27	0,02	-0,29	-0,26	-0,03
101 bis 200	0,25	0,25	0,12	0,36	0,35	0,17
201 bis 500	0,21	0,21	0,05	0,27	0,26	0,09
Mehr als 500	0,20	0,20	0,05	0,28	0,28	0,07
Migrationsstatus (Ref. beide Eltern)						
Vater im Ausland geboren	-0,22	-0,23	-0,04	-0,33	-0,28	-0,16
Mutter im Ausland geboren	-0,42	-0,42	-0,19	-0,25	-0,26	-0,15
Beide Elternteile im Ausland gegeboren	-0,07	-0,07	-0,01	-0,24	-0,23	-0,11
Geschlecht Kind (männl.=1, weibl.=0)		0,02	-0,12		-0,37	-0,06
Testleistung Mathematik / Deutsch			0,85			0,84
<i>R</i> ²	0,11	0,11	0,60	0,17	0,21	0,66

Anmerkungen: [†] Variable wurde z-standardisiert; **Rot** gedruckte Koeffizienten sind statistisch signifikant bei $p < .05$; SES = sozioökonomischer Status; ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie); *b* = Regressionskoeffizient; *R*² = Multipler Determinationskoeffizient.

Sprache assoziiert sein, was sich auf die Deutschleistung des Kindes auswirken könnte. Diese Differenzierung müsste näher untersucht werden. Die Anzahl der Bücher spielte vor allem für die Gruppe „101-200 Bücher“ eine Rolle. Die Jungen blieben jedoch auch unter Kontrolle aller anderen Variablen im Vergleich zu den Mädchen im Fach Mathematik leicht benachteiligt.

Die klassenspezifischen Referenzgruppeneffekte waren in beiden Fächern vorhanden. Um zu untersuchen, ob die Unterschiede zwischen den Jahren bzw. Klassen durch die Zusammensetzung des Lehrkörpers beeinflusst werden könnte, wurden zusätzlich Merkmale der Lehrperson auf ihren prädiktiven Gehalt hin getestet. Die Ergebnisse zeigten, dass die berücksichtigten Lehrpersonenmerkmale keine Zusammenhänge mit der Note aufwiesen.

Tabelle 41: Vorhersage der Mathematik- und Deutschnote 2009 und 2010 durch individuelle Schülermerkmale zum soziokulturellen familiären Hintergrund, Leistungsstärke der Klasse und durch Merkmale der Lehrperson

	Mathematik		Deutsch	
	M4	M5	M4	M5
	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
Landesprache (=1; 0=nicht)	0,14	0,14	0,06	0,06
Sozioökonomischer Status (ISEI) ⁺	0,04	0,04	0,07	0,07
Kulturelle Aktivitäten (Eltern)	0,02	0,02	0,01	0,01
Bücherbesitz (Ref. 26-100)				
weniger als 25	0,02	0,02	-0,02	-0,02
101 bis 200	0,12	0,12	0,17	0,17
201 bis 500	0,04	0,05	0,08	0,08
Mehr als 500	0,05	0,05	0,06	0,06
Migrationsstatus (Ref. beide Eltern)				
Vater im Ausland geboren	-0,05	-0,05	-0,17	-0,17
Mutter im Ausland geboren	-0,19	-0,19	-0,15	-0,15
Beide Elternteile im Ausland geboren	-0,01	-0,01	-0,12	-0,12
Geschlecht Kind (männl.=1; weibl.=0)	-0,12	-0,12	-0,06	-0,06
Testleistung Mathematik / Deutsch	0,87	0,87	0,84	0,84
Mittlere Testleistung Mathematik /Deutsch	-0,31	-0,31	-0,31	-0,30
Geschlecht Lehrperson (m=1; w=0)		0,02		0,00
Anzahl Jahre Unterrichtserfahrung		0,05		0,01
Anzahl begleitete Übertrittsverfahren		-0,02		0,01
Selbstwirksamkeit Lehrpersonen		-0,04		0,00
<i>R</i> ²	0,61	0,61	0,67	0,67

4. Resümee, Ausblick und Einschränkungen

Die Befunde der vorliegenden Expertise konnten zeigen, dass Leistungsbewertung in Form von Schulnoten nicht ausschließlich leistungsrelevanten Kriterien folgt. Vielmehr zeigten sich auch Einflüsse leistungsfremder Merkmale. Hierzu zählen neben Referenzgruppeneffekten insbesondere der soziale Hintergrund der Schülerinnen und Schüler sowie das Geschlecht. In den Teilstudien konnten Effekte dieser Merkmale sowohl am Beginn der Bildungsbiografie, in der Grundschule als auch zu einem späteren Zeitpunkt, am Ende der gymnasialen Oberstufe festgestellt werden. Demzufolge erhielten Kinder aus sozial weniger begünstigten Familien bei gleichen Leistungen in einem standardisierten Leistungstest schlechtere Noten als Kinder aus sozial begünstigten Familien. Geschlechterunterschiede zeigten sich zu Gunsten der Mädchen, die bei gleichen Leistungen in standardisierten Leistungstests bessere Noten erhielten als die Jungen.

Bei den beschriebenen Zusammenhängen zwischen sozialer Herkunft und Schulnoten und Geschlecht und Schulnoten handelt es sich um einen „Bruttoeffekt“, bei dem unterstellt wird, dass die Ergebnisse in einem standardisierten Leistungstest ein valides Abbild der Gesamtleistung darstellen. Dies ist jedoch nicht der Fall, da standardisierte Leistungstests immer nur einen Ausschnitt der benotungsrelevanten Fähigkeiten und Verhaltensweisen erfassen. Leistungen im mündlichen Bereich, wie beispielsweise die Mitarbeit im Unterricht und bei den Hausaufgaben, fließen ebenfalls in die Notenvergabe mit ein und könnten für die beschriebenen Effekte verantwortlich sein.

Für die These, dass die unterschiedlich strenge Benotung von Kindern aus verschiedenen sozialen Herkunftsgruppen auf sozialschichtspezifische Ausprägungen motivationaler Merkmale wie Gewissenhaftigkeit oder Anstrengungsbereitschaft zurückzuführen sind, konnten in dieser Teilstudie keine empirischen Belege gefunden werden. Für den Geschlechtereffekt konnte zumindest eine partielle

Mediation nachgewiesen werden. Offen muss bleiben, ob mit den hier verwendeten Indikatoren von Gewissenhaftigkeit und Anstrengungsbereitschaft die motivationalen Merkmale hinreichend vollständig erfasst werden konnten oder ob bei Einbezug weiterer Merkmale eine Verstärkung der Mediationskraft zu beobachten wäre.

Unerwünschte leistungsfremde Einflüsse auf die Notenvergabe sind insbesondere dann potenziell problematisch, wenn die Noten ein wichtiges Selektionskriterium für den weiteren Bildungsweg darstellen. Dies ist beispielsweise am Ende der Grundschule der Fall, da für den Übergang auf das Gymnasium die Noten von besonderer Wichtigkeit sind. Die Befunde aus den Effektzerlegungen konnten zeigen, dass Interventionen oder bildungspolitische Maßnahmen, die ausschließlich die Übergangssituation in den Blick nehmen zu kurz greifen, da ein Großteil des Herkunftseffektes viel früher angelegt ist. Selbst wenn alle sekundären Herkunftseffekte vermieden werden könnten, würde Ungleichheit weiterbestehen, da Leistungsunterschiede, die zum Beispiel schon zum Beginn der Schullaufbahn bestehen und im Bildungssystem nicht abgebaut werden könnten, immer zu primären Herkunftseffekten führen werden. Effektive Maßnahmen zur Verringerung sozialer Herkunftseffekte können am effektivsten durch die Abmilderung des primären und des sekundären Herkunftseffektes erreicht werden. Darüber hinaus konnten die vorgestellten Analysen zeigen, dass mit einer ausschließlichen Neutralisation insbesondere des sekundären Herkunftseffektes unerwünschte Nebeneffekte auftreten, die dazu führen würden, dass der Anteil der Kinder, die nicht die Leistungsvoraussetzungen für das Gymnasium mitbringen, trotzdem aber auf ein Gymnasium wechseln, ansteigt.

Für den Übergang auf ein Gymnasium ist die von der Grundschule ausgesprochene Übertrittsempfehlung besonders relevant. Schülerinnen und Schüler erhielten in 5.6 Prozent aller Fälle eine Übertrittsempfehlung für das

Gymnasium, obwohl ihre Lehrkräfte eine weniger positive Empfehlung für angemessen hielten. Diese Abweichung von der als angemessen wahrgenommenen Empfehlung ist mit dem sozialen Hintergrund der Schülerinnen und Schüler assoziiert.

Auch in der gymnasialen Oberstufe zeigen sich leichte Einflüsse leistungsfremder Merkmale. Besonders hervorzuheben ist, dass sich kaum Unterschiede zwischen den Fach- und Prüfungsnoten zeigten. Dieser Befund ist eher als erfreuliches Resultat zu betrachten, da sich für die Fachnoten kaum Hinweise auf Nachteile bei der Leistungsbewertung in Abhängigkeit des familiären Hintergrundes feststellen ließen.

Einschränkend muss betont werden, dass in den vorliegenden Studien neben den eingesetzten Leistungstests keine oder nur sehr begrenzt motivationale Schülermerkmale wie Anstrengungsbereitschaft, Mitarbeit im Unterricht, Gewissenhaftigkeit und ähnliche unterrichtsbezogene Aspekte einbezogen werden konnten, von denen bekannt ist, dass sie ebenfalls von Bedeutung für die Notengebung sind und somit möglicherweise Hinweise auf die den Effekten der familiären Herkunft und des Geschlechts zu Grunde liegenden Prozessen geben können. Ferner ist darauf hinzuweisen, dass sich unsere Befunde nur auf ausgewählte Fächer beziehen. Um die Ergebnisse auf eine breitere empirische Basis zu stellen, bedarf es weiterer Untersuchungen in anderen Fächern unter Einbezug mehrerer Bundesländer. In diesen Studien müssten auch die motivationalen Schülermerkmale differenzierter erfasst werden, um theoretisch vorstellbare Vermittlungshypothesen zu testen. Des Weiteren machen die Befunde der Expertise darauf aufmerksam, dass es differenzierterer Analysen bedarf, um an die Prozesse und Mechanismen zu gelangen, warum Lehrkräfte für Kinder bei gleichen Leistungen unterschiedliche Bewertungsmaßstäbe ansetzen.

Anhang 1:

Beschreibung der TIMSS-Übergangsstudie

1. Einleitung

Die folgende Beschreibung der TIMSS-Übergangsstudie folgt im Wesentlichen der Beschreibung von Becker et al. (2010). Dabei werden nur die für diese Expertise relevanten Informationen aufgeführt. Für weiterführende Informationen sei daher auf Becker et al. (2010) verwiesen.

Bei der TIMSS-Übergangsstudie „Der Übergang von der Grundschule in die weiterführende Schule – Leistungserechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (Maaz et al., 2010) handelt es sich um ein Kooperationsprojekt des Max-Planck-Instituts für Bildungsforschung (MPIB), des Instituts für Schulqualität (IFS) und des Instituts für Qualitätsentwicklung im Bildungswesen (IQB). Die „TIMSS-Übergangsstudie“ ist an die, in Deutschland auf einer repräsentativen Stichprobe 4. Klassen beruhende, internationale Vergleichsstudie *Trends in International Mathematics and Science Study*, 2007 (TIMSS) angegliedert (Bos et al., 2008; Martin, Mullis & Foy, 2008; Mullis, Martin & Foy, 2008).

2. Anlage der Untersuchung

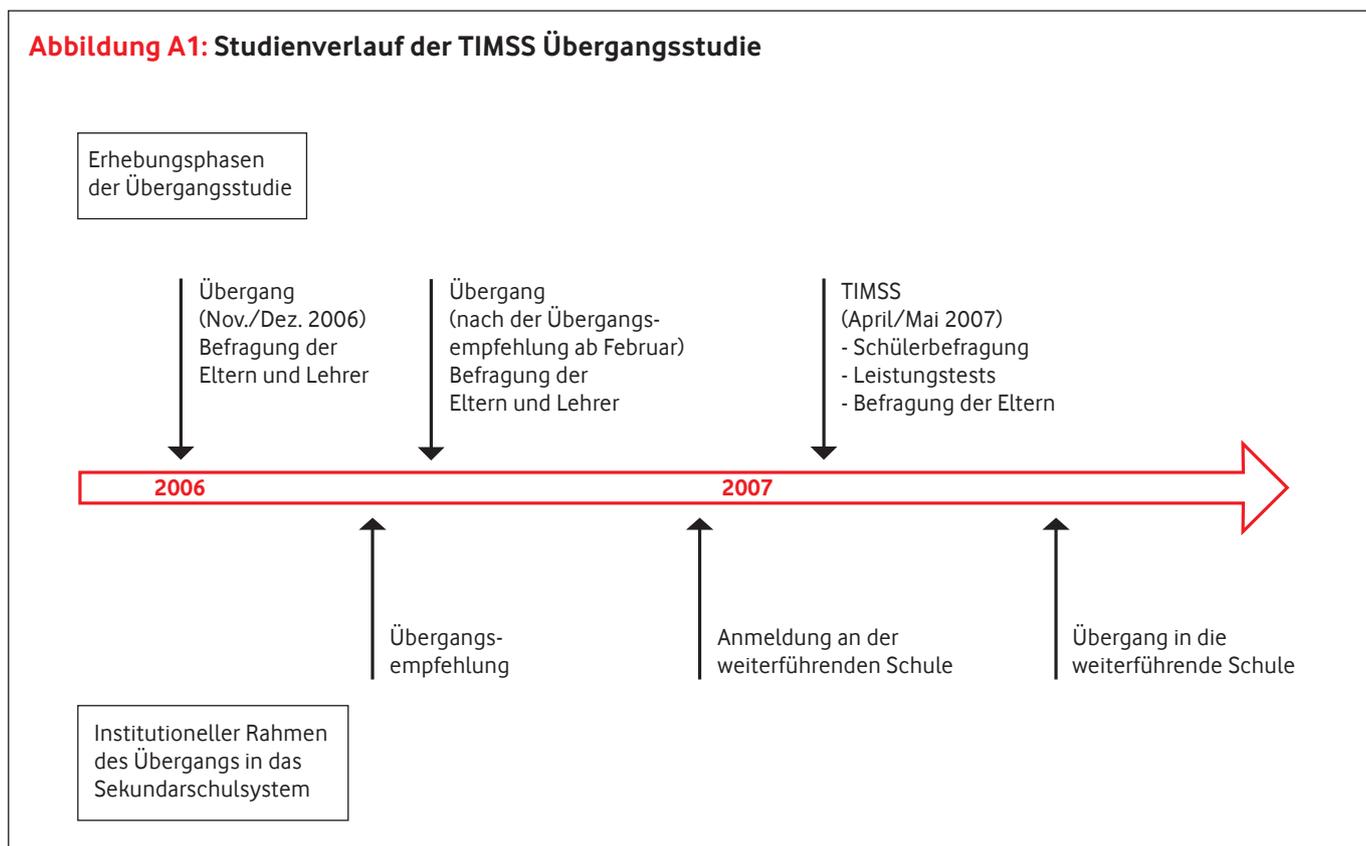
Die Untersuchungspopulation der TIMSS-Übergangsstudie ist weitgehend deckungsgleich mit derjenigen der TIMS-Studie. Für die TIMSS-Übergangsstudie wurden alle diejenigen Schülerinnen und Schüler als Grundgesamtheit definiert, die im Schuljahr 2006/07 in einem Bundesland lebten, in dem der Übergang in die Sekundarstufe nach der 4. Klasse erfolgt, und eine 4. Klasse auf einer Regelschule besuchten. Entsprechend dieser Setzung wurden, im Unterschied zu TIMSS, Berlin, Brandenburg und Mecklenburg-Vorpommern nicht in die Population einbezogen, da in diesen Bundesländern der Übergang von der Primar- in die Sekundarstufe erst nach der 6. Klasse stattfindet. Aufgrund ihrer gesonderten Übergangssituation wurden Schülerinnen und Schüler, die zum Zeitpunkt der Erhebung keine Regelschule, sondern eine Förder- oder Sonderschule besuchten, ebenfalls nicht als Teil der Grundgesamtheit der TIMSS-Übergangsstudie definiert.

Studiendesign und zeitlicher Ablauf

Der Zeitraum der Datenerhebung für die TIMSS-Übergangsstudie erstreckte sich vom November 2006 bis zum September/Oktober 2008. Abbildung 1 stellt die verschiedenen Messzeitpunkte der TIMSS-Übergangsstudie dar und ordnet diese zeitlich in den Übergangsprozess ein. Oberhalb des Zeitstrahls sind die einzelnen Erhebungen der Studie mit Informationen über die Untersuchungseinheiten abgetragen, während unterhalb des Zeitstrahls die einzelnen Phasen/Stationen des Übergangsprozesses dargestellt sind. (→ **ABBILDUNG 1**)

Schülerinnen und Schüler der 4. Klasse, deren Eltern sowie die jeweiligen Klassenlehrerinnen und Klassenlehrer wurden an bis zu drei Zeitpunkten vor dem Übergang befragt. Die erste Befragung der Eltern und Lehrkräfte fand im November/Dezember 2006 statt, bevor die teilnehmenden Schülerinnen und Schüler das Halbjahreszeugnis der 4. Klasse erhalten hatten. Die zweite Befragung – ebenfalls der Eltern und Lehrkräfte – erfolgte direkt nach Erhalt der Übergangsempfehlung bzw. der Beantragung des Übertrittszeugnisses im Februar/März 2007. Die dritte Erhebung wurde im Rahmen der TIMSS-Testung zu einem Zeitpunkt durchgeführt, zu dem die meisten Eltern ihr Kind bereits auf der angestrebten Schule bzw. Schulform angemeldet hatten. Zu diesem Zeitpunkt wurden Leistungstests bei den Schülerinnen und Schülern erhoben sowie Lehrkräfte, Eltern und die Schülerinnen und Schüler selbst zu ihrem schulischen und psychosozialen Hintergrund befragt.

Parallel zur Datenerhebung in Form von Leistungstests und Fragebögen wurden zusätzlich Basisinformationen wie Noten, Geschlecht, Alter oder die erhaltene Übergangsempfehlung der teilnehmenden Schülerinnen und Schüler von ihren Schulen/Lehrkräften erfasst. Diese Informationen wurden für jedes Kind der teilnehmenden Klasse – unabhängig der eigenen Teilnahme zu einem der Erhebungszeitpunkte – erhoben.

Abbildung A1: Studienverlauf der TIMSS Übergangsstudie

2.3 UNTERSUCHUNGSINSTRUMENTE

Leistungstests: Testkonzeption und -statistik

In der TIMSS-Übergangsstudie kamen unterschiedliche Test- und Erhebungsinstrumente zum Einsatz.

Im Folgenden soll auf die wesentlichen Teststatistiken der Leistungstests eingegangen werden. Eine ausführlichere Darstellung der Testkonzeption und Teststatistik findet sich darüber hinaus bei Bensen et al. (2008) sowie bei Olson, Martin und Mullis (2008) und Granzer, Köller und Bremerich-Vos (2009).

Der *Mathematiktest* bestand aus 179 Items im Multiple-Choice- (96) und Kurzantwortformat (83). Die Aufteilung nach den mathematischen Inhaltsbereichen Arithmetik (52 %), Geometrie/Messen (34 %) und Daten (15 %) entspricht im Wesentlichen ihrer Bedeutungs-samkeit in den Schulbüchern. Darüber hinaus lassen sich die Items nach den kognitiven Anforderungsbereichen *Reproduzieren* (39 %), *Anwenden* (39 %) und *Problemlösen* (22 %) unterteilen. Die Reliabilität des Mathematik-Leistungstests, erhoben über die interne Konsistenz, lag in Deutschland bei einem Cronbachs Alpha von $\alpha = .83$ (vgl. Bensen et al., 2008).

Der Aufgabenpool des Leistungstests *Naturwissenschaften* umfasste 174 Testaufgaben, von denen 93 im Multiple-Choice-Format und 81 in einem kurzen offenen Antwortformat vorgegeben wurden. In Anlehnung an die

schulischen Fächer können die Aufgaben inhaltlich den Bereichen Biologie (43 %), Physik (37 %) und Chemie (21 %) zugeordnet werden. Darüber hinaus lassen sich die Aufgaben nach dem kognitiven Anforderungsprofil *Reproduzieren* (44 %), *Anwenden* (26 %) und *Problemlösen* (20 %) aufteilen. Die Reliabilität des naturwissenschaftlichen Leistungstests lag in Deutschland bei einem Cronbachs Alpha von $\alpha = .80$ (vgl. Bensen et al., 2008).

Die Erhebung der *Deutschleistung* fand im Rahmen der Normierung 2007 der BiSta statt (vgl. Granzer et al., 2009). Insgesamt wurden 446 Items, die theoretisch vier verschiedenen Kompetenzbereichen *Lesen*, *Hören*, *Sprachgebrauch* und *Rechtsschreibung* zugeordnet werden können, eingesetzt. Das *Multi-Matrix-Sampling*, analog zum Vorgehen in den Mathematik- und Naturwissenschaftstests in TIMSS, erlaubte, dass nicht alle Schülerinnen und Schüler alle Aufgaben bzw. Aufgabenbereiche bearbeiten mussten. Vielmehr bearbeiteten die Kinder Aufgaben aus zwei bis drei Kompetenzbereichen. Trotz der Zuordenbarkeit der Items zu vier unterschiedlichen Subdomänen konnte ein eindimensionales Rasch-Modell angepasst werden. Somit lässt sich die Deutschleistung ebenfalls wie die Testleistung in Mathematik und Sachkunde mit einem gemeinsamen Wert (*composite score*) für die individuelle Zielperson abbilden. Für die TIMSS-Übergangsstudie ergab sich insgesamt ein Mittelwert von $M = 0.06$ und eine

Standardabweichung von $SD = 1.20$. Die Werte wurden analog zum Vorgehen in TIMSS auf die Metrik $M = 150$ und $SD = 10$ standardisiert. Die Standardisierung erfolgte anhand der TIMSS-Stichprobe ohne die zusätzlich gezo-genen Migrantenschulen. Die WLE-Reliabilität des Tests betrug $r = .81$.

Als Indikator der *kognitiven Leistungsfähigkeit* wurde der figurale Subtest N2 des Kognitiven Fähigkeits-Tests für 4. Klassen (KFT 4–12+R; Heller & Perleth, 2000) verwendet. Bei diesem Test sollen Analogien zwischen zwei Figurenpaaren gebildet werden. Vorgegeben wird jeweils ein Paar und eine Hälfte eines zweiten Paares, zu der dann aus fünf möglichen Antworten diejenige ausgesucht werden muss, die für das zweite Paar die gleiche Beziehung herstellt, wie sie durch das erste Paar vorgegeben ist. Die Tests wurden in zwei parallelisierten Versionen A und B administriert. Die Reliabilität dieser Tests liegt für Version A bei Cronbachs $\alpha = .92$ und für Version B bei $\alpha = .93$ (vgl. Heller & Perleth, 2000).

Fragebögen

Zur Erfassung relevanter Hintergrundmerkmale und psychosozialer Angaben wurden für Schülerinnen und Schüler, Eltern und Lehrkräfte *Fragebögen* verwendet. Die Fragebögen wurden in deutscher Sprache vorgegeben. Zusätzlich wurden den Eltern bei Bedarf Fragebögen in den Sprachen Russisch und Türkisch zur Verfügung gestellt.

Im Schülerfragebogen (TIMSS) wurden die Schülerinnen und Schüler um Auskunft über ihre eigenen Einstellungen zu Schule und ihren Eltern gebeten. Der Schülerfragebogen wurde durch die TIMSS-Übergangsstudie um übergangsrelevante Aspekte erweitert. Die in den postalischen Nachbefragungen eingesetzten Schülerfragebögen erfassten primär Aspekte der psychosozialen Bewältigung des Übergangs. Über den Elternfragebogen wurden neben übergangsrelevanten Angaben der Eltern zu ihrem Kind weitere Merkmale wie etwa Ansichten und Einstellungen zur Schule und der demografische Hintergrund erfasst. Der Lehrerfragebogen erfasste schließlich Angaben zur Demografie, Ausbildung sowie individuelle Einstellungen der Lehrerschaft zur Schule und zum Übergang. Zudem wurden in der zweiten Befragung der Lehrer individuelle Einschätzungen der einzelnen Schülerinnen und Schüler erhoben.

Alle Fragebögen, die im Rahmen der Studie zum Einsatz kamen, wurden im Vorfeld der Untersuchung pilotiert. Insgesamt gab es drei Pilotstudien, die im Januar, Mai und Herbst des Jahres 2006 stattfanden.

Schülerteilnahmeliste (Trackingdaten)

Über die Leistungstests hinaus wurden Geschlecht, Alter, Noten aus dem Halbjahreszeugnis sowie die Empfehlung für die Sekundarschule in einer Schülerteilnahmeliste unabhängig der individuellen Teilnahme erhoben. Diese Informationen werden im Folgenden auch als Trackingdaten bezeichnet. Mit Ausnahme von zwei Bundesländern (Sachsen und Sachsen-Anhalt¹) ist die Schülerteilnahmeliste um Angaben zur Umgangssprache und dem Status, ob ein Kind aus einer (Spät-)Aussiedlerfamilie stammte, erweitert.

Übergang in die Sekundarschulen

Der tatsächliche Übergang der Kinder von den Grund- in die Sekundarschulen wurde über Angaben der Eltern zum Zeitpunkt der dritten Erhebung mit der Frage „Welche Schulform wird Ihr Kind im nächsten Schuljahr besuchen?“ erfasst. Bei fehlenden Angaben wurde auf Schülerangaben zurückgegriffen (ebenfalls Welle 3 – Wortlaut der Frage: „Auf welche Schulform wirst du nach den Sommerferien gehen?“). Die Übereinstimmung zwischen Eltern- und Schülerangaben lag bei 92,0 Prozent, bei einer dichotomen Differenzierung zwischen Gymnasium und anderen Schulformen, wie sie den meisten der folgenden Analysen zugrunde gelegt wurde, sogar bei 98,5 Prozent.² Restliche fehlende Angaben (betrifft 12,2 % aller Schülerinnen und Schüler) wurden nach dem unten beschriebenen Verfahren imputiert.

Übergangsempfehlung für die Sekundarschulen

Die Übergangsempfehlung für die Sekundarschule wurde über die Lehrkräfte zu einem Zeitpunkt erfasst, als die Empfehlung bereits ausgestellt wurde (Welle 3). Unterschieden wird zwischen „Gymnasialempfehlung“ und „Empfehlung für eine andere Schulform“.

Es gab zum damaligen Zeitpunkt der Erhebungen drei Bundesländer, in denen nicht alle Schülerinnen und Schüler eine Empfehlung erhielten: So musste in Bayern das Übertrittszeugnis (entspricht der Empfehlung) von den Eltern beantragt werden. Wurde kein Antrag gestellt, zählte dies automatisch als „Empfehlung für die Hauptschule“. In Hessen wurde von den Schulen zunächst der Wunsch der Eltern erfasst. Nur wenn der Besuch der Realschule oder des Gymnasiums (bzw. entsprechende Bildungsgänge an der Kooperativen Gesamtschule) angestrebt wurde, erteilte die Klassenkonferenz eine Empfehlung. In Thüringen dagegen konnten die Eltern ihr Kind bei

entsprechenden Noten auch ohne Empfehlung auf dem Gymnasium anmelden. Somit hatten fehlende Angaben zur Übergangsempfehlung in diesen drei Bundesländern eine andere Bedeutung als in Bundesländern, in denen jeder Schüler bzw. jede Schülerin automatisch eine Empfehlung erhielt. Fehlende Werte wurden durch Angaben der Eltern aufgefüllt, wodurch für knapp 99 Prozent Angaben zur erhaltenen Empfehlung zur Verfügung standen.³ Die restlichen 1 Prozent wurden über das unten beschriebene Verfahren imputiert (vgl. letzter Abschnitt des vorliegenden Kapitels zum „Umgang mit fehlenden Werten“).

3.1 REALISIERTE STICHPROBE

In TIMSS waren strenge Auflagen zur Teilnahme sowohl auf Schüler- als auch auf Schulebene vorgesehen, die in internationaler Absprache vereinbart und eingehalten werden mussten (vgl. Bensen et al., 2008). Die Ausschöpfungsquote der Stichprobe musste insgesamt auf Schüler- und Schulebene mehr als 95 Prozent betragen (unter Einschluss von nachrückenden Schulen). In Deutschland betrug die Beteiligungsquote insgesamt 96 Prozent (vgl. Bensen et al., 2008).

Ebenso wie bei vergleichenden Analysen, wie sie in TIMSS vorgesehen sind, kommt in einer längsschnittlichen Studie wie der vorliegenden TIMSS-Übergangsstudie, die sich insgesamt über fünf Erhebungswellen erstreckt, dem Datenausfall eine zentrale Bedeutung zu. Je geringer der Datenausfall ist, und je geringer die Unterschiede zwischen denjenigen, die sich an der Studie durchweg beteiligen, von denjenigen sind, die nur selektiv an der Untersuchung teilnehmen, desto aussagekräftiger sind die Kennwerte im Hinblick auf Effizienz und Konsistenz. Für die TIMSS-Übergangsstudie stehen verschiedene Hintergrundinformationen (z. B. Geschlecht, Noten oder die Übergangsempfehlung) nahezu vollständig zur Verfügung, da sie von den Klassenlehrern unabhängig von der Teilnahme der Schüler oder Eltern über die Schülerteilnahmeliste mitgeteilt wurden. Sie ermöglichen zu testen, inwiefern es zu möglicherweise systematischem Datenausfall kam und statistische Korrekturverfahren indiziert sind (vgl. Abschnitt „Umgang mit fehlenden Werten“).

Die Tabellen 1 bis 3 stellen vergleichend die jeweils spezifischen Teilnahmequoten sowie die zentralen Merkmale der Schülerinnen und Schüler der erreichten Personen dar. Werte werden getrennt ausgewiesen für die Stichprobe, die sich an die TIMSS-Stichprobe anlehnt (im Folgenden als TIMSS-Übergangsstudie-Stichprobe

bezeichnet), und für diejenigen, die ein *oversampling* von Klassen mit hohem Anteil von Schülerinnen und Schülern mit Migrationshintergrund vorsah (im Folgenden als Migranten-Stichprobe bezeichnet).

Wie sich mit einem Blick auf die Tabellen 1 bis 3 erkennen lässt, nahmen die Beteiligungsquoten im Laufe der Erhebungen vor dem Übergang ab und lagen zwischen 99 und 70 Prozent. Tabelle 1 stellt die Unterschiede zwischen der anvisierten und erreichten Schülerstichprobe dar. Es zeigte sich, dass für die meisten Schülerinnen und Schüler (99 % bzw. 96,2 %) Angaben aus dem Schülerfragebogen vorlagen. Entsprechend sind praktisch keinerlei Abweichungen zwischen der anvisierten Stichprobe und denjenigen, die an der Schülerbefragung teilnahmen, zu erkennen. Geschlechteranteil, Noten und Gymnasialempfehlung unterscheiden sich nicht zwischen den Stichproben (Cohens $d \leq 0.03$; Cohen, 1988).

(→ TABELLE 1)

Die Beteiligung aufseiten der Eltern an den einzelnen Wellen lag zwischen 85 und 71 Prozent (vgl. Tab. 2), bei einer etwas höheren Teilnahme in jeweils früheren Wellen. Hinsichtlich der Unterschiede zwischen Schülerinnen und Schülern der Basisstichprobe und denjenigen, deren Eltern tatsächlich an einer der Wellen teilnahmen, sind die teilnehmenden Gruppen tendenziell positiver ausgewählt in Hinblick auf Noten und Gymnasialempfehlung, eine Tendenz, die sich von Welle 1 zu Welle 3 leicht verstärkte. Jedoch erscheinen diese inhaltlich wenig bedeutsam (Cohens $d \leq 0.11$; Cohen, 1988).

(→ TABELLE 2)

¹ Für diese beiden Bundesländer lagen entsprechende Informationen auf Klassenebene vor, das heißt, es war jeweils bekannt, wie viele Schüler in dieser Klasse Spätaussiedlerstatus oder einen bestimmten sprachlichen Hintergrund aufweisen.

² Gesamtschulen wurden dabei als „andere Schulform“ klassifiziert, da auch der Besuch des Gymnasialzweiges einer Kooperativen Gesamtschule in der Regel eine Entscheidung gegen die Schulform „Gymnasium“ darstellt. Wie zusätzliche Auswertungen zeigten, sind die zentralen Analyseergebnisse im Wesentlichen deckungsgleich, unabhängig von der Zuordnung der Schülerinnen und Schüler der Gesamtschulen.

³ Für Bayern und Hessen wurde Schülerinnen und Schüler, für die nach Angaben der Lehrkräfte keine Empfehlung beantragt wurde, eine Empfehlung einer „anderen Schulform“ zugeordnet. In Thüringen wurde Schülern, die anhand ihrer Noten und entsprechend der Regelungen eindeutig als „Gymnasiasten“ eingestuft werden konnten, eine „Gymnasialempfehlung“ zugeordnet.

Tabelle 1: Angaben zu Schülerinnen und Schülern nach Teilnahmestatus der Schülerinnen und Schüler für die TIMSS-Übergangsstudie-Stichprobe und die Migranten-Stichprobe

	Anvisierte Stichprobe		Teilnahme am SFB	
	TIMSS-Ü	TIMSS-MHG	TIMSS-Ü	TIMSS-MHG
Anteil männlicher Schüler	50,9%	50,6%	50,9%	50,8%
Halbjahresnote Deutsch; M (<i>SD</i>)	2,72 (0,89)	2,75 (0,90)	2,72 (0,89)	2,73 (0,89)
Halbjahresnote Mathematik; M (<i>SD</i>)	2,74 (0,96)	2,77 (0,97)	2,75 (0,96)	2,76 (0,97)
Halbjahresnote Sachkunde; M (<i>SD</i>)	2,48 (0,87)	2,52 (0,89)	2,48 (0,87)	2,50 (0,88)
Anteil Gymnasialempfehlung	42,0%	40,8	42,0%	41,2%
<i>N</i>	5174	5712	5122	5499
Anteil an Gesamtstichprobe	100%	100%	99,0%	96,3%

Anmerkung: Angaben jeweils für valide Fälle auf den jeweiligen Variablen. *M* = Mittelwert; *SD* = Standardabweichung; TIMSS-Ü: TIMSS-Übergangsstudie-Stichprobe; TIMSS-MHG: Migranten-Stichprobe.

Tabelle 2: Angaben zu Schülerinnen und Schülern nach Teilnahmestatus der Eltern an Welle 1, 2 oder 3 für die TIMSS-Übergangsstudie-Stichprobe und die Migranten-Stichprobe

	Anvisierte Stichprobe		Teilnahme an Welle 1		Teilnahme an Welle 2		Teilnahme an Welle 3		Teilnahme zu mindestens einem Messzeitpunkt	
	TIMSS-Ü	TIMSS-MHG	TIMSS-Ü	TIMSS-MHG	TIMSS-Ü	TIMSS-MHG	TIMSS-Ü	TIMSS-MHG	TIMSS-Ü	TIMSS-MHG
Anteil männlicher Schüler	50,9%	50,6%	50,3%	50,4%	49,6%	49,7%	50,4%	50,5%	50,6%	50,6%
Halbjahresnote Deutsch; M (<i>SD</i>)	2,72 (0,89)	2,75 (0,90)	2,66 (0,88)	2,70 (0,89)	2,62 (0,87)	2,65 (0,88)	2,62 (0,87)	2,65 (0,88)	2,69 (0,88)	2,72 (0,89)
Halbjahresnote Mathematik; M (<i>SD</i>)	2,74 (0,96)	2,77 (0,97)	2,69 (0,94)	2,71 (0,95)	2,65 (0,93)	2,67 (0,95)	2,65 (0,93)	2,66 (0,94)	2,71 (0,95)	2,74 (0,96)
Halbjahresnote Heimat- und Sachunterricht; M (<i>SD</i>)	2,48 (0,87)	2,52 (0,89)	2,43 (0,85)	2,47 (0,97)	2,40 (0,84)	2,43 (0,86)	2,39 (0,84)	2,42 (0,86)	2,45 (0,86)	2,49 (0,88)
Anteil Gymnasialempfehlung	42,0%	40,8	43,3%	42,0%	45,2%	44,1%	45,9%	44,9%	42,0%	0,8%
<i>N</i>	5174	5712	4450	4889	4019	4366	3754	4046	4768	5242
Anteil an Gesamtstichprobe	100%	100%	85,9%	85,5%	77,6%	76,3%	72,5%	70,7%	92,1%	91,7%

Anmerkung: Angaben jeweils für valide Fälle auf den jeweiligen Variablen. *M* = Mittelwert; *SD* = Standardabweichung; TIMSS-Ü: TIMSS-Übergangsstudie-Stichprobe; TIMSS-MHG: Migranten-Stichprobe.

Tabelle 3: Angaben zu Schülerinnen und Schülern nach Teilnahmestatus der Lehrkräfte für die TIMSS-Übergangsstudie-Stichprobe und die Migranten-Stichprobe

	Anvisierte Stichprobe		Teilnahme am LFB Welle 1		Teilnahme am LFB Welle 2	
	TIMSS-Ü	TIMSS-MHG	TIMSS-Ü	TIMSS-MHG	TIMSS-Ü	TIMSS-MHG
Anteil männlicher Schüler	50,9%	50,6%	51,2%	51,2%	50,2%	50,2%
Halbjahresnote Deutsch; <i>M (SD)</i>	2,72 (0,89)	2,75 (0,90)	2,71 (0,89)	2,74 (0,90)	2,64 (0,87)	2,67 (0,88)
Halbjahresnote Mathematik; <i>M (SD)</i>	2,74 (0,96)	2,77 (0,97)	2,75 (0,96)	2,76 (0,97)	2,68 (0,94)	2,69 (0,95)
Halbjahresnote Sachkunde; <i>M (SD)</i>	2,48 (0,87)	2,52 (0,89)	2,49 (0,87)	2,52 (0,89)	2,41 (0,85)	2,44 (0,87)
Anteil Gymnasial- empfehlung	42,0%	40,8	42,1%	41,0%	44,8%	43,9%
N	5174	5712	4868	5330	3801	4130
Anteil an Gesamtstichprobe	100%	100%	94,1%	93,3%	73,4%	72,2%

Anmerkung: Angaben jeweils für valide Fälle auf den jeweiligen Variablen. *M* = Mittelwert; *SD* = Standardabweichung; TIMSS-Ü: TIMSS-Übergangsstudie-Stichprobe; TIMSS-MHG: Migranten-Stichprobe.

Vollständig parallel hierzu ist das Antwortmuster zu sehen, wenn die Stichproben hinsichtlich der Teilnahmehäufigkeit von den Eltern verglichen werden. Insgesamt haben rund 92 Prozent der Eltern an mindestens einer Erhebung teilgenommen (vgl. Tab. 2, vorletzte sowie letzte Spalte). Auch diese Unterschiede zwischen Basis- und ausgeschöpfter Stichprobe sind aus inhaltlicher Perspektive als gering einzustufen (Cohens $d \leq 0,03$; Cohen, 1988). Da die zentralen Analysevariablen der folgenden inhaltlichen Kapitel in der Regel den Elternangaben entnommen sind, stellen die Stichprobenangaben, wie sie in den letzten beiden Spalten der Tabelle 2 dargestellt sind, diejenigen Fälle dar, die den nachfolgenden Analysen zugrunde gelegt wurden. Für die reguläre TIMSS-Übergangsstudien-Stichprobe umfasste dies 4768 Schülerinnen und Schüler und für die erweiterte Migranten-Stichprobe 5242 Fälle. (→ TABELLE 3)

Der Stichprobenausfall aufseiten der Lehrkräfte ist in Tabelle 3 zusammengefasst. Für die TIMSS-Übergangsstudien-Stichprobe beteiligten sich 214 von 226 Lehrern (94,7 %) bzw. für die Migranten-Stichprobe 238 von 253 Lehrern (94,1 %). Es resultierten auf der Schülerebene Angaben zu 94,1 bzw. 93,3 Prozent der einzelnen Schülerinnen und Schüler. Zu Welle 2, in der zu jeder Schüle-

rin bzw. jedem Schüler der betreffenden Klasse Angaben gemacht werden sollten, lag die Beteiligung auf Ebene der Lehrerinnen und Lehrer ähnlich hoch mit 211 Lehrkräften (93,4 %) in der TIMSS-Übergangsstudie-Stichprobe und 234 (92,5 %) in der Migranten-Stichprobe. Da jedoch drei (TIMSS-Übergangsstudie) bzw. vier Lehrkräfte (Migranten-Stichprobe) keine Individualeinschätzungen abgaben sowie Lehrkräfte nur diejenigen Schülerinnen und Schüler beurteilten, für die eine Einverständniserklärung der Eltern vorlag, fiel der Anteil von Angaben auf Schülerebene geringer aus als zur ersten Welle der Lehrerbefragung. Somit sind von 73,4 Prozent der Kinder aus der TIMSS-Übergangsstudie-Stichprobe bzw. 72,3 Prozent der Kinder der Migranten-Stichprobe Daten verfügbar. Der Unterschied zur Basisstichprobe erscheint relativ gering, wenngleich tendenziell mehr Informationen zu Kindern mit besseren Noten und einer höheren Wahrscheinlichkeit einer Gymnasialempfehlung vorliegen (Cohens $d \leq 0,09$; Cohen, 1988).

Anhang 2: Beschreibung der ELEMENT-Studie

Für die Analysen zum sechsjährigen Grundschulsystem konnte auf Daten der in Berlin durchgeführten Untersuchung „Erhebungen zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin“ (ELEMENT) zurückgegriffen werden. In dieser längsschnittlich angelegten Untersuchung wurde die Entwicklung der Deutsch- und Mathematikleistungen der Schülerinnen und Schüler des Primarbereichs von der vierten bis zur sechsten Klasse untersucht (vgl. Lehmann & Lenkeit, 2008). In Jahrgangsstufe 6 wurden zudem die Englischleistungen querschnittlich erhoben. Die Leistungstests wurden durch Schüler-, Eltern- und Lehrerbefragungen ergänzt.

Anders als in den meisten Bundesländern wechselt der Großteil der Schülerinnen und Schüler in Berlin erst nach der 6. Jahrgangsstufe in die weiterführenden Schulen der gegliederten Sekundarstufe I. Für besonders leistungsstarke Schülerinnen und Schüler besteht allerdings die Möglichkeit, bereits nach der vierten Klasse auf ein grundständiges Gymnasium zu wechseln. Für den Jahrgang der ELEMENT-Untersuchung lag der Anteil der vorzeitigen Übergänger an das Gymnasium bei rund sieben Prozent der Gesamtschülerschaft.

In der ELEMENT-Untersuchung wurde sowohl die Leistungsentwicklung der vorzeitigen Übergänger (Vollerhebung mit $N = 1.724$ Schülerinnen und Schülern zur ersten Erhebungswelle) als auch der regulären Grundschülerschaft untersucht, für die ein repräsentatives Sample von 3.293 Schülerinnen und Schülern aus 71 Grundschulen mit 140 Klassen (Bruttostichprobe erste Erhebungswelle) gezogen wurde. Da im vorliegenden Beitrag nur der Übergang nach der sechsten Klasse in die weiterführenden Schulen untersucht wird, beziehen sich alle Auswertungen und Analysen ausschließlich auf das Sample der regulären Grundschülerschaft zum dritten Erhebungszeitpunkt in der sechsten Klassenstufe ($N = 3.008$). Allerdings handelt es sich dabei nur um einen Teil der Grundschulstichprobe, denn erst eine Anschlussuntersuchung in Jahrgangsstufe 8

(ELEMENT 8), die unter Federführung des Max-Planck-Instituts für Bildungsforschung durchgeführt wurde, machte die im Zentrum der vorliegenden Untersuchung stehenden Lernkompetenzeinschätzungen aus den Bildungsgangsempfehlungen zugänglich und überführte sie in den Datensatz. Um die ELEMENT-Schülerinnen und Schüler in die Nachuntersuchung einbeziehen zu können, bedurfte es zunächst dreier zentraler Voraussetzungen. Zum einen mussten die Zuordnungslisten von Schülername und Schüler-ID, die aus datenschutzrechtlichen Gründen in den ehemaligen Grundschulen der Schülerinnen und Schüler verwahrt wurden, noch vorhanden sein, um die nacherhobenen Daten mit den bereits vorhandenen Schülerdaten verbinden zu können. Zum anderen musste herausgefunden werden, an welche Schulen die Schülerinnen und Schüler nach dem Verlassen der Grundschule gegangen waren, da die Bildungsgangsempfehlungen samt Lernkompetenzeinschätzungen Bestandteil der Schülerakten sind, die von der Grundschule an die weiterführende Schule weitergegeben werden. Und schließlich bedurfte es der Bereitschaft der Grund- und der weiterführenden Schulen, die verfügbaren Informationen bereitzustellen.

Letztlich lagen von insgesamt 976 Schülerinnen und Schülern aus 83 Klassen (50 % Mädchen) und damit von etwa 30 Prozent des ursprünglichen Grundschulsamples verwertbare Angaben aus den Bildungsgangsempfehlungen vor. Da nicht ohne Weiteres von einem zufälligen Ausfall von Grund- bzw. Sekundarschulen ausgegangen werden kann, stellt sich die Frage, wie repräsentativ die für unsere Analysen herangezogene Stichprobe ist. Wie den Tabellen 1 und 2 entnommen werden kann, unterscheidet sich die Gruppe der Schülerinnen und Schüler mit vorhandenen Lernkompetenzeinschätzungen nur verhältnismäßig gering von der Gruppe ohne vorliegende Lernkompetenzeinschätzungen, auch wenn die vorhandenen Unterschiede ausnahmslos statistisch signifikant ausfielen. Vergleicht man die Teilgruppe der Schülerinnen und Schüler mit vorhandenen Lernkompetenzein-

Tabelle 1: Mittelwerte und Standardabweichungen zentraler Leistungsindikatoren und des sozioökonomischen Status für die Gruppen mit und ohne Bildungsgangempfehlung und die Gesamtstichprobe

	Bildungsgangempfehlung liegt vor (Analysestichprobe, N = 976)		Bildungsgangempfehlung liegt nicht vor (N = 2.032)		Gesamtstichprobe (N = 3.008)	
	M	SD	M	SD	M	SD
Lesetest	110,82	13,2	109,44	12,27	109,93	12,52
Mathematiktest	116,87	15,6	113,91	14,86	114,86	15,18
Kognitive Grundfähigkeit ¹⁾	26,38	10,09	25,28	9,87	25,64	10,00
Deutschnote	2,75	0,94	2,85	0,88	2,82	0,90
Mathematiknote	2,91	1,11	3,02	1,05	2,99	1,05
Fremdsprachnote	2,93	1,03	3,02	1,11	2,98	1,11
ISEI ²⁾	47,89	16,28	45,80	15,38	46,50	15,69

Anmerkungen: M = Mittelwert; SD = Standardabweichung

1) Kognitive Grundfähigkeit = Kognitiver Fähigkeitstest, KFT.

2) ISEI = International Socio-Economic Index of Occupational Status, (Sozioökonomischer Status, höchster Wert in der Familie).

schätzungen mit der gesamten Grundschulstichprobe, reduzieren sich die Unterschiede nochmals. Hinsichtlich der Leistungsindikatoren für die sechste Klasse (Leistungstests und Noten) fallen die Unterschiede gegenüber dem Gesamtsample bis auf den Mathematiktest ($d = 0,14$) geringer als ein zehntel Standardabweichung aus. Dasselbe gilt für den sozioökonomischen Status, für den sich ebenfalls leicht höhere Werte als im Gesamtsample finden. Die in Tabelle 2 aufgeführten prozentualen Angaben zur beruflichen Ausbildung der Eltern, dem Migrationsstatus und der ausgesprochenen Übergangsempfehlung machen ebenfalls deutlich, dass bei der von uns herangezogenen Teilstichprobe erwartungsgemäß nicht von einem repräsentativen Sample Berliner Schülerinnen und Schüler ausgegangen werden kann, wenngleich die Unterschiede zwischen den „Teilnehmern“ und „Nicht-Teilnehmern“ relativ gering ausfielen.

Tabelle 2: Prozentuale Verteilung zentraler Hintergrundindikatoren und der vergebenen Übergangsempfehlung für die Gruppen mit und ohne Bildungsgangempfehlung und die Gesamtstichprobe

	Bildungsgangempfehlung liegt vor (Analysestichprobe, N = 976)	Bildungsgangempfehlung liegt nicht vor (N = 2.032)	Gesamtstichprobe (N = 3.008)
Berufliche Ausbildung			
ohne Abschluss	11,8 %	13,6 %	13,0 %
Lehre	27,6 %	25,8 %	26,4 %
Fachschulabschluss	29,4 %	31,5 %	30,8 %
FH-Abschluss	11,7 %	10,7 %	11,0 %
Uni-Abschluss	19,6 %	18,4 %	18,8 %
Migrationshintergrund			
ein Elternteil im Ausland geb.	13,2 %	12,1 %	12,4 %
beide Elternteile im Ausland geb.	25,0 %	28,4 %	27,3 %
Übergangsempfehlung			
Hauptschule	18,2 %	19,2 %	18,9 %
Realschule	40,4 %	46,0 %	44,1 %
Gymnasium	41,4 %	34,8 %	37,0 %

Anhang 3:

Beschreibung der TOSCA-Studie

Die folgende Beschreibung der TOSCA-Studie folgt im Wesentlichen den Beschreibungen von Maaz (2006) und Köller, Watermann, Trautwein & Lüdtke (2004), auf die gleichzeitig für weitere Ausführungen verwiesen sein soll.

1. Anlage der Studie

Die TOSCA-Studie (Transformation des Sekundarschulsystems und akademische Karrieren; vgl. Köller, Watermann, Trautwein & Lüdtke, 2004) ist ein interdisziplinäres Forschungsprojekt, das in Kooperation zwischen dem Max-Planck-Institut für Bildungsforschung und dem Institut zur Qualitätsentwicklung im Bildungswesen im Jahr 2002 begründet wurde und durch die Universität Tübingen weitergeführt wird. Die TOSCA-Studie ist als Längsschnittpunkt konzipiert, in dem der Lebens- und Bildungsweg von Abiturienten über einen Zeitraum von zehn Jahren verfolgt werden soll. Die Frage, wie das Zusammenspiel zwischen institutionellen Gelegenheitsstrukturen und Einflüssen (z.B. Schulformen, Zusammensetzung der Schülerschaft), familiärem Hintergrund und den individuellen Ressourcen der Schülerinnen und Schüler die weitere schulische (akademische) und persönliche Entwicklung der jungen Erwachsenen beeinflusst, steht im Mittelpunkt der Untersuchung. Dabei liegt ein besonderer Fokus auf den Vergleich der Schülerinnen und Schüler der allgemeinbildenden und beruflichen Gymnasien.

Die erste Datenerhebung der TOSCA-Studie fand im Frühjahr 2002 in den 13. Jahrgangsstufen der baden-württembergischen Oberstufe in 90 allgemeinbildende und 59 beruflichen Gymnasien statt. Zum diesem ersten Erhebungszeitpunkt wurden Leistungstests in den Fächern Mathematik und Englisch durchgeführt. Ferner wurden die Schülerinnen und Schüler gebeten, einen umfangreichen Schülerfragebogen auszufüllen. Insbesondere soziokulturelle Merkmale wurden ebenfalls von den Eltern in Form eines Elternfragebogens erfasst. Über die Schulen konnten weitere Informationen der Schülerinnen und Schüler, wie das Geschlecht und die Abiturnote, erhoben werden.

Die in Teilstudie xx vorgestellten Analysen beziehen sich ausschließlich auf die Schülerschaft der allgemeinbildenden Gymnasien zu diesem Zeitpunkt der Datenerhebung.

2. Stichprobenausfälle und Selektivitätsanalysen

Die Qualität einer Stichprobe wurde in bisherigen Schulleistungsuntersuchungen meistens auf der Basis ihres Ausschöpfungsgrads beurteilt, die jedoch jedoch nur Aussagen über das Ausmaß des Stichprobenausfalls, nicht über die selektive Stichprobenverzerrung gibt. Um dieses Problem zu überwinden, wurden in der TOSCA-Studie mittels an der Schule vorhandenen Schülerlisten zentrale Informationen auch über die Schülerinnen und Schüler erhoben, die in der gezogenen Stichprobe enthalten waren, nicht aber an der Datenerhebung teilnahmen. Aus wenigen Schulen waren diese Informationen nicht erhältlich, sodass letztendlich N=5809 Schülerinnen und Schüler in die Selektivitätsanalysen eingingen.

(→ TABELLE 1)

Tabelle 1 gibt einen Überblick über die gewonnenen Ergebnisse. Die Mittelwerte zeigen, dass die Gruppe der Teilnehmer bezüglich der Abiturnote, den Abiturprüfungsfächern sowie den Kurspunkten im Halbjahreszeugnis der 13. Jahrgangsstufe leicht positiv selektiert ist, dass die teilnehmenden Schülerinnen und Schüler häufiger einem Mathematikleistungskurs angehören als die nichtteilnehmenden und dass Frauen seltener der Gruppe der Nichtteilnehmer angehören als Männer. Obwohl die Ergebnisse implizieren, dass die statistischen Analysen auf der Basis einer realisierten Stichprobe von N = 4730 Schülerinnen und Schüler zu einer leichten Überschätzung der Leistungen in den Populationen führen könnte, konnte Lüdtke u.a. (2004) zeigen, dass die dargestellten Unterschiede der Selektivitätsanalyse bereits durch die durch die vorgenommene Gewichtung der Daten kompensiert werden konnten.

Tabelle 1: Mittelwerte (M) und Standardabweichungen (SD) zentraler Informationen für die Nichtteilnehmer und Teilnehmer

	Personen der angestrebten Stichprobe, die nicht an der Erhebung teilgenommen haben			Personen der angestrebten Stichprobe, die an der Erhebung teilgenommen haben		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Frauenanteil	1067	0,49	0,50	4730	0,54	0,50
Abiturnote	891	2,58	0,62	4451	2,42	0,64
<i>Abiturprüfung</i>						
1. Fach	915	8,56	3,26	4512	8,98	3,35
2. Fach	912	8,58	3,42	4508	9,22	3,35
3. Fach	915	7,09	3,58	4514	7,92	3,49
4. Fach	907	9,33	3,51	4494	9,88	3,31
<i>Leistungskurs</i>						
Mathematik	921	0,28	0,45	4557	0,35	0,48
Englisch	922	0,37	0,48	4558	0,35	0,48
<i>Noten (Punkte) im Halbjahreszeugnis</i>						
Mathematik	1038	7,07	3,60	4641	8,31	3,50
Englisch	901	8,07	2,91	4102	8,72	2,80

Geschlecht: 0 = männlich, 1 = weiblich; *N* = Anzahl der Personen

3. Leistungsbezogene Indikatoren

Mathematikleistungen. Die Mathematikleistungen im Fachunterricht der gymnasialen Oberstufe wurden mit Originaltestheften aus der Third International Mathematics and Science Study (TIMSS; vgl. Baumert, Bos & Lehmann, 2000a, 2000b) erhoben. Der Test umfasst insgesamt 68 Aufgaben mit unterschiedlichen Aufgabenformaten (multiple-choice, offene kurze Antworten, offene ausführliche Antworten), die unterschiedliche Stoffgebiete und kognitive Anforderungen abdecken (zu Details siehe Klieme, 2000). Um die Schülerinnen und Schüler nicht über Gebühr zu belasten und gleichzeitig die Stoffgebiete breit erfassen zu können, wurden vier Testheftversionen mit unterschiedlichen Aufgaben vorgegeben, die über ein Set gemeinsamer Aufgaben miteinander verknüpft waren (Multi Matrix Sampling). Einschätzungen der Aufgaben durch Lehrplanexperten und Mathematikfachleiter in den untersuchten Schulen belegen die zufriedenstellende Lehrplanvalidität sowohl für allgemeinbildende als auch für die unterschiedlichen Richtungen beruflicher Gymnasien (zu Details siehe Watermann, Nagy & Köller, 2004). Die Werte für die interne Konsistenz (Kuder-Richardson-Formel 20) der jeweiligen Testheftversionen betragen 0,80. Für die statistischen Analysen wurde ein auf der Basis

der Item Response Theory skaliertes Leistungswert verwendet. Der Test wurde auf die TIMSS-Metrik normiert. Der internationale Mittelwert der Mathematikleistung wurde auf 500 Punkte festgelegt und die Standardabweichung beträgt 100 Punkte. Die Korrelation zwischen der Mathematiknote im letzten Zeugnis und der Testleistung beträgt über alle Schulen und Kursniveaus hinweg $r = .47$ ($p < .001$) und kann als zufriedenstellender Validitätshinweis gewertet werden.

Englischleistungstest. Die Englischleistung wurde mit einer Kurzversion des Test of English as a Foreign Language (TOEFL) erhoben, der vom Educational Testing Service (ETS) in Princeton (vgl. ETS, 1997) entwickelt wurde. Die Kurzversion umfasst 71 Items zum Hör- und Leseverständnis sowie zu Grammatik- und Orthografiekenntnissen (vgl. Köller & Trautwein, 2004). Die interne Konsistenz (Kuder-Richardson-Formel 20) des Gesamttests lag bei 0,92. Mit der Englischnote im letzten Zeugnis ergaben sich in Grund- und Leistungskursen Korrelationen um $r = .50$ ($p < .001$).

Kognitive Grundfähigkeiten. Zur Erfassung der kognitiven Grundfähigkeiten wurden zwei Untertests aus dem Kognitiven Fähigkeitstest für 4. bis 12. Klassen, Revision

(KFT 4–12 + R; Heller & Perleth, 2000) verwendet. Der KFT 4–12 + R folgt dem Paradigma der psychometrischen Intelligenzmessung (Waldmann, 1996), bei dem weniger die zu Grunde liegenden Denkprozesse, sondern die Produkte dieser Denkprozesse im Vordergrund des Interesses stehen. Der Test steht in der faktorenanalytischen Tradition und rekurriert hier am ehesten auf das Modell mehrerer gemeinsamer Faktoren von Thurstone (1938; siehe im Überblick Conrad, 1983). Der KFT 4–12 + R ist allerdings auch zur Untersuchung kognitiver Prozesse geeignet, wie sie in modernen kognitionspsychologisch ausgerichteten Theorien intellektueller Fähigkeiten untersucht werden (vgl. Waldmann, 1996). Berücksichtigt wurden die Subskalen „Figurenanalogien“ (Untertest N2) und „Wortanalogien“ (Untertest V4). Der Untertest N2 erfasst die allgemeine Denkfähigkeit, der Untertest V4 erfasst neben allgemeiner Denkfähigkeit auch verbale Fähigkeiten. Im Sinne prozessorientierter Theorien (Waldmann, 1996) sind beide Untertests geeignet, induktives Denken zu erfassen (vgl. auch Klauer, 1993). Der Untertest N2 besteht aus 25 figuralen Items, der Test V4 aus 20 Items. Auf der Basis aller 45 Items wurden aus beiden Tests Fähigkeitswerte bestimmt. Dazu wurden mithilfe der ConQuest-Software (Wu, Adams & Wilson, 2000) individuelle Werte auf der Grundlage von Item-Response-Modellen berechnet (vgl. Lüdtke u.a., 2004) und auf einer Skala mit einem Mittelwert von $M = 250$ und einer Standardabweichung von $SD = 50$ (in der ungewichteten Stichprobe) abgebildet. Dieses Gesamtmaß war hoch reliabel (Schätzung nach Rost, 1996; $RTT = 0,91$). Als Indikator für die Validität wurde die Korrelation mit der durchschnittlichen Abiturnote und den Leistungen im Mathematiktest bestimmt. Hier ergaben sich Koeffizienten von $r = 0,30$ ($p < 0,01$; Abiturnote) und $r = 0,51$ ($p < 0,01$; Mathematiktest) in erwarteter Höhe.

Anhang 4:

Das Bildungssystem in der Schweiz

Die folgende Beschreibung des Schweizer Schulsystems und des Übergangsverfahrens im deutschsprachigen Kanton Freiburg/Schweiz folgt im Wesentlichen den Ausführungen von Maaz, Neumann und Trautwein (2008) und Baeriswyl, Wandeler, Trautwein und Oswald (2006).

1. Das Bildungssystem

Ähnlich wie Deutschland verfügt die Schweiz über ein föderales Schulsystem mit unterschiedlichen Differenzierungsformen. Die obligatorische Schulpflicht beträgt neun Jahre und umfasst die Primarschule und die Sekundarstufe I. Die Grundschule besteht zumeist aus den Jahrgangsstufen 1 bis 6. In der Mehrzahl der Kantone schließt sich daran die in zwei oder drei verschiedene Bildungsgänge („Schultypen“) gegliederte Sekundarstufe I an. Sie differenziert sich in einen Bildungsgang mit Grundansprüchen, der zumeist als „Realschule“ bezeichnet wird und auf einfache Berufslehren vorbereiten soll, sowie ein oder zwei Bildungsgänge mit erweiterten Ansprüchen („Sekundarschule“ und „Progymnasium“), deren Absolventen anspruchsvollere Berufslehren aufnehmen bzw. in eine Maturitätsschule (Gymnasium) oder die Fachmittelschule überwechseln. Die verschiedenen Bildungsgänge werden oft gemeinsam *innerhalb* sogenannter „Orientierungsschulen“ geführt. Neben der fachübergreifenden Leistungsdifferenzierung in Form der verschiedenen Schultypen finden sich in einigen Kantonen aber auch integrative und kooperative Modelle mit Stammklassen und Niveauunterricht in den Hauptfächern (Deutsch, Fremdsprache und Mathematik).

Die in Teilstudie 10 berichteten Analysen beziehen sich auf den deutschsprachigen Kantonsteil Freiburg. Der deutschsprachige Teil des Kantons Freiburg führt die Sekundarstufe I in Form einer Orientierungsschule, in der sich alle drei Abteilungen im gleichen Gebäude mit teilweise gemeinsamen Aktivitäten (bspw. bei Projektwochen) befinden. Nach der obligatorischen Schulzeit von neun Jahren sind Eintritte ins Gymnasium oder in eine Berufsausbildung möglich.

2. Das Übergangsverfahren im deutschsprachigen Kanton Freiburg/Schweiz

Für die deutschsprachigen Schulen im Kanton Freiburg wurde Mitte der 1990er-Jahre ein Übergangsmodell von der Grundschule in die Sekundarschule entwickelt. Dessen Hauptmerkmale sind eine Standardisierung von Leistungsbeurteilungen durch zentrale Leistungstests, eine Berücksichtigung von Motivation und Arbeitsverhalten der Kinder bei der Übergangsempfehlung, eine intensive Einbindung der Eltern bei der Übergangentscheidung durch eine obligatorische Beratung sowie eine vertikale Öffnung des Zugangs zur Sekundarstufe II.

Im Einzelnen umfasst der Übergangsprozess drei Hauptelemente: Die *Übertrittsempfehlung* durch die Lehrperson, die Übertrittsempfehlung durch die Eltern sowie die Leistungen der Schülerinnen und Schüler in einem standardisierten Leistungstest. Die *Übertrittsempfehlung der Primarlehrperson* berücksichtigt zum einen die Schülerleistungen anhand der Zeugnisnoten des ersten Semesters der sechsten Klasse, zum anderen Beobachtungs- und Beurteilungskriterien der kognitiven Fähigkeiten und des Lern- und Arbeitsverhalten des Kindes. Auf dieser Basis formuliert die Lehrperson eine Empfehlung für die Zuweisung zu einer der drei Abteilungen. Die *Übertrittsempfehlung der Eltern* ist das Ergebnis des kommunikativen Prozesses zwischen Lehrperson und Eltern. In einem für die Lehrkraft verpflichtenden Gespräch wird den Eltern die Empfehlung der Schule offengelegt, worauf hin die Eltern formal ihre eigene Empfehlung einreichen. Den Eltern ist dabei freigestellt, bei ihrer Empfehlung von der Empfehlung der Lehrperson abzuweichen. Das dritte Element des Übergangsverfahrens ist eine *Vergleichsprüfung* für alle Schülerinnen und Schüler des Kantons in den Bereichen Deutsch und Mathematik mit einem standardisierten Leistungstest, der von allen Schülerinnen und Schülern am selben Tag bearbeitet wird. Die Auswertung erfolgt durch Lehrkräfte aus dem Sekundarschulbereich. Für jede Schülerin und jeden Schüler wird das Gesamtprüfungsergebnis als Prozentrang festgehalten und mit der Lehrerempfehlung verglichen.

Die Ergebnisse der Vergleichsprüfungen werden, wie die Übertrittsempfehlungen von Lehrperson und Eltern, zentral festgehalten.

Im Zuweisungsprozess wird abschließend die Übereinstimmung zwischen Lehrerempfehlung mit dem Prüfungsergebnis überprüft. Bei Übereinstimmung zwischen Lehrerempfehlung und Prüfungsleistung folgt eine direkte Zuweisung zu der von der Lehrperson empfohlenen Abteilung. Bei einer Diskrepanz zwischen Empfehlung und Prüfungsergebnis treffen die Primarschule und Eltern gemeinsam auf der Basis von Beurteilungs- und Prüfungsdokumenten eine Zuweisungsentscheid. Nach dem Übergang in die Orientierungsschule kann zu jeder Zeit innerhalb des ersten Semesters auf Antrag der Eltern oder der Lehrpersonen der Sekundarstufe I ein Abteilungswechsel vorgenommen werden.

Abbildungs- und Tabellenverzeichnis

ABBILDUNGEN

Abbildung 1: Ein vereinfachtes Arbeitsmodell zu möglichen Einflussfaktoren auf die Leistungsbewertung.....	11
Abbildung 2: Mittelwerte und Standardabweichungen für die Schülermerkmale Gewissenhaftigkeit im Unterricht und Anstrengungsbereitschaft differenziert nach dem Bildungsgrad der Eltern.....	31
Abbildung 3: Darstellung der Wirkungsweise primärer und sekundärer Herkunftseffekte beim Übergang in die Sekundarstufe I.....	37
Abbildung 4: Wirkungsweise primärer und sekundärer Herkunftseffekte beim Übergang in die Sekundarstufe I.....	38
Abbildung 5: Effektzerlegung des sozialen Herkunftseffektes auf die Übergangsempfehlung.....	39
Abbildung 6: Effektzerlegung des sozialen Herkunftseffektes auf den Übergang.....	40
Abbildung 7: Hypothetische Gymnasialquote von Kindern aus der Arbeiterklasse und Differenz zur empirischen Quote.....	45
Abbildung 8: Mittelwerte und Standardabweichungen von Schulnoten in der Grundschule differenziert nach Geschlecht.....	65
Abbildung 9: Mittelwerte und Standardabweichungen der Leistungstests differenziert nach Geschlecht.....	66
Abbildung 10: Mittelwerte und Standardabweichungen für die Schülermerkmale Gewissenhaftigkeit im Unterricht und Anstrengungsbereitschaft differenziert nach Geschlecht.....	67

TABELLEN

Tabelle 1:	Referenznorm für Schulzeugnisnoten.....	16
Tabelle 2:	Vorhersage der Durchschnittsnote in der Bildungsgangempfehlung durch individuelle Schülermerkmale des sozio-kulturellen familiären Hintergrund und zentraler Leistungsindikatoren (ELEMENT)	24
Tabelle 3:	Korrelationsmuster alle leistungsbezogener Schülermerkmale	25
Tabelle 4:	Anteile aufgeklärter Varianz durch Schülermerkmale an der Durchschnittsnote auf Individualebene	27
Tabelle 5:	Anteile aufgeklärter Varianz durch Schülermerkmale an den Fachnoten in Mathematik, Deutsch und Sachkunde auf Individualebene.....	28
Tabelle 6:	Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum sozio-kulturellen familiären Hintergrund und zentrale Leistungsindikatoren	29
Tabelle 7:	Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum Migrationsstatus, zum sozio-kulturellen familiären Hintergrund und zentraler Leistungsindikatoren	30
Tabelle 8:	Korrelationen der motivationalen Merkmale mit Schulnoten.....	32
Tabelle 9:	Korrelationen der motivationalen Merkmale mit Ergebnissen aus standardisierten Leistungstests	32
Tabelle 10:	Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum Migrationsstatus, zum sozio-kulturellen familiären Hintergrund, zentraler Leistungsindikatoren sowie motivationaler Merkmale	33
Tabelle 11:	Domänenspezifische Befunde für die Mathematiknote	34
Tabelle 12:	Domänenspezifische Befunde für die Deutschnote	34
Tabelle 13:	Domänenspezifische Befunde für die Sachkundenote.....	35
Tabelle 14:	Bedingte Wahrscheinlichkeiten eine gute Note zu bekommen und auf das Gymnasium zu wechseln, getrennt für Schülerinnen und Schüler der Dienstleistungs- und Arbeiterklasse für das Noten-Cutoff-Kriterium 2,33	42
Tabelle 15:	Neutralisation des primären Effektes (der Leistungsverteilung), Note und Entscheidung bleiben herkunftsspezifisch	47
Tabelle 16:	Neutralisation des sekundären Noteneffektes (der Benotungspraxis), Leistung und Entscheidung bleiben herkunftsspezifisch	48
Tabelle 17:	Neutralisation des sekundären Übergangseffektes (des Entscheidungsverhaltens), Leistung und Note bleiben herkunftsspezifisch	49

Tabelle 18:	Neutralisation des primären Effektes (Leistung) und des sekundären Benotungseffektes bei Beibehaltung und des sekundären Entscheidungseffekte.....	50
Tabelle 19:	Vorhersage der Fachnote in Mathematik (Schulhalbjahr 13/1) durch individuelle Schülermerkmale im Grundkurs	52
Tabelle 20:	Vorhersage der schriftlichen Prüfungsnote in Mathematik durch individuelle Schülermerkmale im Grundkurs	54
Tabelle 21:	Vorhersage der Fachnote in Mathematik (Schulhalbjahr 13/1) durch individuelle Schülermerkmale im Leistungskurs.....	55
Tabelle 22:	Vorhersage der Prüfungsnote in Mathematik durch individuelle Schülermerkmale im Leistungskurs	56
Tabelle 23:	Vorhersage der Fachnote in Englisch (Schulhalbjahr 13/1) durch individuelle Schülermerkmale im Grundkurs	57
Tabelle 24:	Vorhersage der Fachnote in Englisch (Schulhalbjahr 13/1) durch individuelle Schülermerkmale im Leistungskurs.....	58
Tabelle 25:	Vorhersage der Prüfungsnote in Englisch durch individuelle Schülermerkmale im Leistungskurs	59
Tabelle 26:	Von Lehrkräften eingeschätzte Wichtigkeit unterschiedlicher Bereiche für den schulischen Erfolg eines Kindes	61
Tabelle 27:	Mittelwerte für die von den Lehrkräften eingeschätzte Wichtigkeit unterschiedlicher Merkmale	62
Tabelle 28:	Korrelationen der untersuchten Lehrervariablen in Bundesländern mit bindender Übergangsempfehlung (unterhalb der Diagonalen, untere Dreiecksmatrix) vs. in Bundesländern ohne bindende Übergangsempfehlung (oberhalb der Diagonalen, obere Dreiecksmatrix)	63
Tabelle 29:	Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum sozio-kulturellen familiären Hintergrund und zentrale Leistungsindikatoren	68
Tabelle 30:	Vorhersage der Durchschnittsnote durch individuelle Schülermerkmale zum Migrationsstatus, zum sozio-kulturellen familiären Hintergrund, zentraler Leistungsindikatoren sowie motivationaler Merkmale	69
Tabelle 31:	Testleistungen, Fachnote Schulhalbjahr 13/1 und Prüfungsnote in Mathematik getrennt nach Grund- und Leistungskurs (Standardabweichung in Klammern)	70
Tabelle 32:	Testleistungen, Fachnote Schulhalbjahr 13/1 und Prüfungsnote in Englisch getrennt nach Grund- und Leistungskurs (Standardabweichung in Klammern)	70
Tabelle 33:	Vorhersage der Fach- und Prüfungsnote in Mathematik durch Testleistung, Geschlecht und kognitive Grundfähigkeit im Grundkurs	71

Tabelle 34:	Vorhersage der Fach- und Prüfungsnote in Mathematik durch Testleistung, Geschlecht und kognitive Grundfähigkeit im Leistungskurs	71
Tabelle 35:	Vorhersage der Fachnote in Englisch durch Testleistung, Geschlecht und kognitive Grundfähigkeit im Grundkurs.	72
Tabelle 36:	Vorhersage der Fach- und Prüfungsnote in Englisch durch Testleistung, Geschlecht und kognitive Grundfähigkeit im Leistungskurs	72
Tabelle 37:	Deskriptive Angaben zu den Erhebungen der Übertrittsjahrgänge 2009 und 2010	75
Tabelle 38:	Vergleich der Noten- und Prüfungsmittelwerte	76
Tabelle 39:	Varianzkomponenten der Noten in verschiedenen Fächern	77
Tabelle 40:	Vorhersage der Mathematik- und Deutschnoten durch individuelle Schülermerkmale zum sozio-kulturellen familiären Hintergrund, Geschlecht und Testleistung	79
Tabelle 41:	Vorhersage der Mathematik- und Deutschnote 2009 und 2010 durch individuelle Schülermerkmale zum soziokulturellen familiären Hintergrund, Leistungsstärke der Klasse und durch Merkmale der Lehrperson	80

Literaturverzeichnis

Achermann, E. (1966). *Methodik des Volksschulunterrichtes*. Dritte, teilweise umgearbeitete Auflage. Hochdorf: Martinusverlag.

Aebli, H. (1973). *Psychologische Didaktik. Didaktische Auswertung der Psychologie von Jean Piaget*. Stuttgart: Klett.

Aebli, H. (1978). *Grundformen des Lehrens*. Stuttgart: Klett.

Aebli, H. (1987). *Grundlagen des Lehrens*. Stuttgart: Klett.

Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer.

Arnold, K.-H., Bos, W., Richert, P. & Stubbe, T. (2010). Der Übergang von der Grundschule in die Sekundarstufe: Schullaufbahnpräferenzen von Lehrkräften und Eltern im Ländervergleich. In: W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, I. Tarelli & R. Valtin (Hrsg.). *IGLU-E 2006. – Die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens*. (S.13-32). Münster: Waxmann.

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.

Baeriswyl, F. Wandeler, C. Trautwein, T. & Oswald, K. (2006). Leistungstest, Offenheit von Bildungsgängen und obliigatorische Beratung der Eltern. Reduziert das Deutschfreiburger Übergangsmodell die Effekte des sozialen Hintergrunds bei Übergangentscheidungen? *Zeitschrift für Erziehungswissenschaft*, 9, 373-392.

Baeriswyl, F., Wandeler, C. & Christ, O. (2008). Die Übertrittsempfehlung – zufällig oder zuverlässig? Analyse der Determinanten und Konstanz von Lehrerempfehlungen bei Schulübertritten während sieben Jahren. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30, 549-578.

Baeriswyl, F., Wandeler, C. & Trautwein, U. (2011): Auf einer anderen Schule oder bei einer anderen Lehrkraft hätte es für's Gymnasium gereicht: Eine Untersuchung zur Bedeutung von Schulen und Lehrkräften für die Übertrittsempfehlung. *Zeitschrift für pädagogische Psychologie*, 25, 39-47.

Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. (S. 323-407) Opladen: Leske + Budrich.

Baumert, J., Artelt, C., Ditton, H., Fend, H., Hasselhorn, M., Macher, I. et al. (2011). *Empfehlungen des Expertenrats „Herkunft und Bildungserfolg“*. Stuttgart: Ministerium für Kultus, Jugend und Sport Baden-Württemberg.

Baumert, J., Bos, W., & Lehmann, R. (Hrsg.). (2000a). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd. 1. Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske + Budrich.

Baumert, J., Bos, W., & Lehmann, R. (Hrsg.). (2000b). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd. 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe*. Opladen: Leske + Budrich.

- Baumert, J., Trautwein, U., & Artelt, C.** (2003). 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., Trautwein, U., & Artelt, C.** (2003a). 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. & Watermann, R.** (2000). Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe. In J. Baumert, W. Bos, & R. Lehmann (Hrsg.), *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 317–372). Opladen: Leske + Budrich.
- Baumert, J., Watermann, R., & Schümer, G.** (2003b). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. *Zeitschrift für Erziehungswissenschaft*, 6, 46-71.
- Becker, M.** (2009). *Kognitive Leistungsentwicklung in differenziellen Lernumwelten: Effekte des gegliederten Sekundarschulsystems in Deutschland*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Becker, M., Gresch, C., Baumert, J., Watermann, R., Schnitger, D. & Maaz, K.** (2010). Durchführung, Daten und Methoden. 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. 107-121). Bonn & Berlin: BMBF.
- Becker, R. & Schubert, F.** (2011). Die Rolle von primären und sekundären Herkunftseffekten für Bildungschancen von Migranten im deutschen Bildungssystem. In R. Becker (Hrsg.), *Integration durch Bildung. Bildungserwerb von jungen Migranten in Deutschland* (S. 161-194). Wiesbaden: Westdeutscher Verlag.
- Becker, R.** (2000). Klassenlage und Bildungsentscheidungen. Eine empirische Anwendung der Wert-Erwartungstheorie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 52, 450–474.
- Becker, R.** (2003). Educational expansion and persistent inequalities of education. Utilizing subjective expected utility theory to explain increasing participation rates in upper secondary school in the Federal Republic of Germany. *European Sociological Review*, 19, 1-24.
- Becker, R.** (2009). Entstehung und Reproduktion dauerhafter Bildungsungleichheiten. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (S. 85–129). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bellenberg, G., & Klemm, K.** (1998). Von der Einschulung bis zum Abitur. Zur Rekonstruktion von Schullaufbahnen in Nordrhein-Westfalen. *Zeitschrift für Erziehungswissenschaft*, 1, 577-596.
- Biermann, R.** (1976). *Schulische Selektion in der Diskussion*. Bad Heilbrunn: Klinghardt.
- Birkel, C. & Birkel, P.** (2002): Wie einzig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht*, 49, 219-224.
- Bonsen, M., Lintorf, K. A., Bos, W., & Frey, K.** (2008). TIMSS 2007: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS: Dokumentation der Erhebungsinstrumente zur Trends in International Mathematics and Science Study* (S. 19–48). Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.).** (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- 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.), *IGLU Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 191-220). Münster: Waxmann.

Bos, W., Bensen, M., Baumert, J., Prenzel, M., Selter, C., & Walther, G. (Hrsg.). (2008). *TIMSS 2007: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.

Boudon, R. (1974). *Education, opportunity, and social inequality: Changing prospects in Western society*. New York: Wiley.

Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 183-189). Göttingen: Schwarz.

Breen, R., & Goldthorpe, J. H. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society*, 9, 275-305.

Ceci, S. (1991). How Much Does Schooling Influence General Intelligence and Its Cognitive Components? A Reassessment of the Evidence. *Developmental Psychology*, 27, 703-722.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Coleman, J. S. (1988). Social capital and the creation of human capital. *American Journal of Sociology*, 94, 95-120.

Coleman, J. S. (1996). Der Verlust sozialen Kapitals und seine Auswirkungen auf die Schule. In A. Leschinsky (Hrsg.), *Die Institutionalisierung von Lehren und Lernen*. Weinheim: Beltz. S. 99-105.

Conrad, W. (1983). Intelligenzdiagnostik. In K.-J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik* (S. 104-201). Göttingen: Hogrefe (Enzyklopädie der Psychologie, B/II/2).

Ditton, H. (1992). *Ungleichheit und Mobilität durch Bildung. Theorie und empirische Untersuchung über sozial-räumliche Aspekte von Bildungsentscheidungen*. Weinheim: Juventa.

Ditton, H. (2005). Der Beitrag von Familie und Schule zur Reproduktion von Bildungsungleichheit. In H. G. Holtappels & K. Höhmann (Hrsg.), *Schulentwicklung und Schulwirksamkeit. Systemsteuerung, Bildungschancen und Entwicklung der Schule. 30 Jahre Institut für Schulentwicklungsforschung* (S. 121-130). Weinheim: Juventa.

Ditton, H. (2007). *Kompetenzaufbau und Laufbahnen im Schulsystem. Ergebnisse einer Längsschnittuntersuchung an Grundschulen*. Münster: Waxmann.

Ditton, H., & Krüsken, J. (2006). Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 9, 348-372.

Ditton, H., Krüsken, J., & Schauenberg, M. (2005). Bildungsungleichheit – der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft*, 8, 285-303.

Dumont, H., Trautwein, U., Lüdtke, O., Neumann, M., Niggli, A. & Schnyder, I. (in press). Does parental homework involvement mediate the relationship between family background and educational outcomes? *Contemporary Educational Psychology*.

Ehmke, T., Hohensee, F., Heidemeier, H., & Prenzel, M. (2004). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 225-254). Münster: Waxmann.

Ehmke, T., Hohensee, F., Siegle, T., & Prenzel, M. (2006). Soziale Herkunft, elterliche Unterstützungsprozesse und Kompetenzentwicklung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 225-248). Münster: Waxmann.

Erikson, R. & Goldthorpe, J. H. (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.

Erikson, R., Goldthorpe, J. H., Jackson, M., Yaish, M., & Cox, D. R. (2005). On class differentials in educational attainment. *Proceedings of the National Academy of Sciences*, 102, 9730-9733.

Erikson, R., Goldthorpe, J.H. & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, 30, 341 – 415.

Fend, H. (1969). *Sozialisation und Erziehung*. Weinheim und Basel: Beltz.

- Fend, H.** (1974). *Gesellschaftliche Beziehungen schulischer Sozialisation*. Weinheim und Basel: Beltz.
- Flammer, A.** (1971). Zur Definition der Notenskala. *Schweizerische Zeitschrift für Psychologie*, 30, 204–218.
- Flammer, A., Perrig-Chiello, P. & Rüegg, T.** (1983). *Zeugnisnoten vor dem Übertritt in die Sekundarstufe. Dokumentation und Vorschläge*. Freiburg/Schweiz: Universitätsverlag.
- Füssel, H.-P., Gresch, C., Baumert, J. & Maaz, K.** (2010). Der institutionelle Kontext von Übergangsentscheidungen. Rechtliche Regelungen und die Schulformwahl am Ende der Grundschulzeit. In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Hrsg.): *Der Übergang von der Grundschule in die weiterführende Schule*. (S.87-106). Bonn: Bundesministerium für Bildung u. Forschung.
- Ganzeboom, H. B. G. & Treiman, D. J.** (1996). Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Social Science Research*, 25, 201-239.
- Ganzeboom, H. B. G., De Graaf, P. M., Treiman, D. J. & De Leeuw, J.** (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1-56.
- Granzer, D., Köller, O., & Bremerich-Vos, A.** (2009). *Bildungsstandards Deutsch und Mathematik: Leistungsmessung in der Grundschule*. Weinheim: Beltz.
- Gresch, C., & Becker, M.** (2010). Sozial- und leistungsbedingte Disparitäten im Übergangsverhalten bei türkischstämmigen Kindern und Kindern aus (Spät-) Aussiedlerfamilien. 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. 181-200). Bonn: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung.
- Hausen, C., Jungblut, J. M., Müller, W., Pollak, R., & Wirth, H.** (2006). *Validation of ESeC: The effect of coding procedures and occupational aggregation level. Deutsche ESeC-Validierungsstudie*. Mannheim: MZES and ZUMA.
- Heller, K. A. & Perleth, C.** (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Beltz.
- Heller, W.** (1986). *Primarschule Schweiz. 22 Thesen zur Entwicklung der Primarschule*. Bern: EDK.
- Hillmert, S. & Jacob, M.** (2010). Selections and social selectivity on the academic track: A life-course analysis of educational attainment in Germany. *Research in Social Stratification and Mobility*, 28, 59-76.
- Hochweber, J.** (2010). *Was erfassen Mathematiknoten? Korrelate von Mathematik-Zeugnisensuren auf Schüler- und Schulklassenebene in Primar- und Sekundarstufe*. Münster u.a.: Waxmann.
- Ingenkamp, K.** (1969). *Zur Problematik der Jahrgangsklasse: Eine empirische Untersuchung*. Weinheim; Berlin; Basel: Beltz
- Ingenkamp, K.** (1971). *Zur Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz.
- Ingenkamp, K.** (1993). Der Prognosewert von Zensuren, Lehrergutachten, Aufnahmeprüfungen und Tests während der Grundschulzeit für den Sekundarschulerfolg. In R. Olechowski & E. Persy (Hrsg.), *Frühe schulische Auslese* (S.68-85). Frankfurt, Main: Lang.
- Jäger, R. S.** (2004). *Von der Beobachtung zur Notengebung*. Landau: Empirische Pädagogik.
- Jonkmann, K., Maaz, K., Neumann, M. & Gresch, C.** (2010). Übergangsquoten und Zusammenhänge zu familiärem Hintergrund und schulischen Leistungen: Deskriptive Befunde. 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. 123-149). Bonn & Berlin: BMBF
- Jungbauer-Gans, M.** (2004). Einfluss des sozialen und kulturellen Kapitals auf die Lesekompetenz: Ein Vergleich der PISA 2000-Daten aus Deutschland, Frankreich und der Schweiz. *Zeitschrift für Soziologie*, 33, 375-397.
- Jungbauer-Gans, M.** (2006). Kulturelles Kapital und Mathematikleistungen – eine Analyse der PISA 2003-Daten für Deutschland. In W. Georg (Hrsg.), *Soziale Ungleichheit im Bildungssystem: Eine empirisch-theoretische Bestandsaufnahme* (S. 175-198). Konstanz: UVK Verlagsgesellschaft.
- Klauer, K.-J.** (Hrsg.). (1993). *Kognitives Training*. Göttingen: Hogrefe.

Klieme, E. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht. Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte. In J. Baumert, W. Bis & R. Lehmann (Hrsg.): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. (S. 57-128). Opladen: Leske u. Budrich.

Knigge, M. & Leucht, M. (2010). Soziale Disparitäten im Spracherwerb. In O. Köller, M. Knigge & B. Tesch (Hrsg.) Sprachliche Kompetenzen im Ländervergleich. Waxmann: Münster.

Köller, O., Baumert, J. & Schnabel, K. (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, 385-422.

Köller, O. & Trautwein, U. (2004). Englischleistungen von Schülerinnen und Schülern an 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. 285-326). Opladen: Leske + Budrich.

Köller, O., Watermann, R., Trautwein, U. & Lüdtke, O. (2004). *Wege zur Hochschulreife in Baden-Württemberg. TOSCA – eine Untersuchung an allgemein bildenden und beruflichen Gymnasien*. Opladen: Leske + Budrich.

Kristen, C. & Dollmann, J. (2009). Sekundäre Effekte der ethnischen Herkunft? Kinder aus türkischen Familien am ersten Bildungsübergang. *Zeitschrift für Erziehungswissenschaft*, 12, 205–229.

Kropf, M., Gresch, C. & Maaz, K. (2010). Überblick über die rechtlichen Regelungen des Übergangs in den beteiligten Ländern. 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. 399-429). Bonn u.a.: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung.

Lehmann, R. & Lenkeit, J. (2008). *Erhebung zum Lese- und Mathematikverständnis. Entwicklung in den Jahrgangsstufen 4 bis 6 in Berlin: Abschlussbericht über die Untersuchungen 2003, 2004 und 2005 an Berliner Grundschulen und grundständigen Gymnasien*. Berlin: Humboldt Universität zu Berlin.

Lehmann, R. H., Peek, R. & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.

Lehmann, R., Gänsfuß, R. & Peek, R. (1999). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen – Klassenstufe 7*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.

Lüdtke, O., Köller, O., Bundt, B., Gomolka, J. & Watermann, R. (2004). Durchführung und methodische Grundlagen der TOSCA-Studie. 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. 121-151). Opladen: Leske + Budrich.

Maaz, K. (2006). *Soziale Herkunft und Hochschulzugang. Effekte institutioneller Öffnung im Bildungssystem*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Maaz, K. & Nagy, G. (2009). Der Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems: Definition, Spezifikation und Quantifizierung primärer und sekundärer Herkunftseffekte. *Zeitschrift für Erziehungswissenschaft, Sonderheft 12-2009*, 153-182.

Maaz, K. & Watermann, R. (2007). Reproduktion oder Mobilität? Zur Wirkung familiärer Prozessmerkmale auf die Studienintention am Ende der gymnasialen Oberstufe. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 27, 285–303.

Maaz, K., Baumert, J. & Cortina, K. S. (2008). Soziale und regionale Ungleichheit im deutschen Bildungssystem. 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. 205–243). Reinbek: Rowohlt.

Maaz, K., Baumert, J. & Trautwein, U. (2009). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo entsteht und vergrößert sich soziale Ungleichheit? *Zeitschrift für Erziehungswissenschaft, Sonderheft 12-2009*, 11-46.

Maaz, K., Baumert, J. & Trautwein, U. (2009). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo entsteht und vergrößert sich soziale Ungleichheit? *Zeitschrift für Erziehungswissenschaft, Sonderheft 12-2009*, 11-46.

Maaz, K., Baumert, J., Gresch, C. & McElvany, M. (Hrsg.). (2010). *Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten*. Bonn & Berlin: BMBF.

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 Baden-Württemberg. TOSCA - Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien* (S. 153–203). Opladen: Leske + Budrich.

Maaz, K., Gresch, C., McElvany, N., Jonkmann, K. & Baumert, J. (2010). Theoretische Konzepte für die Analyse von Bildungsübergängen: Adaptation ausgewählter Ansätze für den Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems. 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. 65-85). Bonn & Berlin: BMBF.

Maaz, K., Neumann, M. & Trautwein, U. (2009). Schulsysteme im deutschsprachigen Raum. In S. Blömeke, T. Bohl, L. Haag, G. Lang-Wojtasik & W. Sacher (Hrsg.), *Handbuch Schule. Theorie – Organisation – Entwicklung* (S. 171–179). Bad Heilbrunn: Klinkhardt.

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, 519-548.

Maaz, K., Schroeder, S. & Gresch, C. (2010). Primäre und sekundäre soziale Herkunftseffekte beim Übergang in die Sekundarstufe I. Neutralisation sozialer Herkunftseffekte und Konsequenzen auf das Übergangsverhalten. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert*. (S. 285-310). Münster: Waxmann.

Maaz, K., Trautwein, U. & Dumont, H. (2011). *Definition und Verteilung von Schulen mit benachteiligter Schülerschaft*. Expertise im Auftrag der Bertelsmann Stiftung.

Maaz, K., Trautwein, U., Gresch, C. & Lüdtke, O. (2009). Intercoder-Reliabilität bei der Berufscodierung nach der ISCO-88 und Validität des sozioökonomischen Status. *Zeitschrift für Erziehungswissenschaft*, 12, 281-301.

Maier, U. (2007). Systematische Lehrereffekte bei Übergangsquoten auf weiterführende Schulen. *Zeitschrift für Erziehungswissenschaft*, 10, 271–284.

Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007: International science report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mayer, K.-U. & Aisenbrey, S. (2007). Variations on a theme: trends in social mobility in (West) Germany for cohorts born between 1919 and 1971. In S. Scherer, R. Pollack, G. Otto & M. Gangl (Hrsg.), *From origin to destination. Trends and mechanisms in social stratification research*. (S.125-154). Frankfurt, Main: Campus Verlag.

McElvany, N. (2010). Die Übergangsempfehlung von der Grundschule auf die weiterführende Schule im Erleben der Lehrkräfte. . 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. 295-311). Bonn & Berlin: BMBF.

Merkens, H., & Wessel, A. (2002). *Zur Genese von Bildungsentscheidungen. Eine empirische Studie in Berlin und Brandenburg*. Baltmannsweiler: Schneider Verlag Hohengehren.

Milek, A.; Lüdtke, O.; Trautwein, U.; Maaz, K. & Stubbe, T. C. (2009). Wie konsistent sind Referenzgruppeneffekte bei der Vergabe von Schulformempfehlungen? Bundeslandspezifische Analysen von Daten der IGLU-Studie. *Zeitschrift für Erziehungswissenschaft, Sonderheft 12-2009*, 282-307.

Mischo, C. & Rheinberg, F. (1995). Erziehungsziele von Lehrern und individuelle Bezugsnormen der Leistungsbewertung. *Zeitschrift für Pädagogische Psychologie*, 9, 139-151.

Müller, W., Wirth, H., Bauer, G., Pollak, R. & Weiss, F. (2006). ESeC – Kurzbericht zur Validierung und Operationalisierung einer europäischen sozioökonomischen Klassifikation. *ZUMA-Nachrichten* 59, Jg. 30, 111 – 119.

Müller-Benedict, V. (2007). Wodurch kann die soziale Ungleichheit des Schulerfolgs am stärksten verringert werden? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59, 615-639.

Müller-Benedict, V. (2008). Strukturveränderungen oder Fördermaßnahmen? Analyse einer unpopulären Alternative mit Hilfe einer Simulationsstudie. *Die Deutsche Schule*, 100, 412-424.

Mullis, I. et al. (1998). *Mathematics and science achievement in the final year of secondary school. IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, Mass.: TIMSS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007: International mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Muthén, L.K. & Muthén, B. (1998-2010). *Mplus User's Guide. Version 6*. Los Angeles, CA: Muthén & Muthén.

Neumann, M., Nagy, G., Trautwein, U. & Lüdtke, O. (2009). Vergleichbarkeit von Abiturleistungen: Leistungs- und Bewertungsunterschiede zwischen Hamburger und Baden-Württemberger Abiturienten und die Rolle zentraler Abiturprüfungen. *Zeitschrift für Erziehungswissenschaft*, 12, 691-714

Neumann, M., Milek, A., Maaz, K. & Gresch, C. (2010). Zum Einfluss der Klassenzusammensetzung auf den Über-

gang 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, Referat Bildungsforschung.

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (2008). *TIMSS 2007: Technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Pietsch, M. (2007). Schulformwahl in Hamburger Schülerfamilien und die Konsequenzen für die Sekundarstufe 1. In W. Bos, C. Grölich & M. Pietsch (Hrsg.), *KESS 4 - Lehr- und Lernbedingungen in Hamburger Grundschulen* (S. 127-165). Münster: Waxmann.

Rheinberg, F. (2001). Bezugsnorm und schulische Leistungsbeurteilung. In Weinert, F. E. (Hrsg.) *Leistungsmessungen in Schulen*. Weinheim und Basel: Beltz.

Rost, J. (1996). *Lehrbuch Testtheorie*. Bern: Huber.

Rubin, D. B. (1987). *Multiple imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7 (2), 147-177.

Schelsky, H. (1957). *Schule und Erziehung in der industriellen Gesellschaft*. Würzburg: Werkbund-Verlag.

Schnabel, K. U., Alfeld, C., Eccles, J. S., Köller, O., & Baumert, J. (2002). Parental influence on students' educational choices in the United States and Germany: Different ramifications – same effect? *Journal of Vocational Behavior*, 60, 178-198.

Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In D.H. Rost(Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. überarb. u. erw. Aufl., S. 95-100). Weinheim: Beltz.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland – KMK. (Hrsg.). (2009). *Das Bildungswesen in der Bundesrepublik Deutschland 2009. Darstellung der Kompetenzen, Strukturen und bildungspolitischen Entwicklungen für den Informationsaustausch in Europa*. <http://www.kmk.org/fileadmin/doc/>

Dokumentation/Bildungswesen_pdfs/dossier_dt_ebook.pdf (18.11.2011).

Stocké, V. (2007). Explaining Educational Decision and Effects of Families' Social Class Position: An Empirical Test of the Breen–Goldthorpe Model of Educational Attainment. *European Sociological Review*, 23, 505-519.

Stubbe, T. C. & Bos, W. (2008). Schullaufbahnpfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik*, 22 (1), 49-63.

Tent, L. (2006). Zensuren. In D.H. Rost(Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. überarb. u. erw. Aufl., S. 873-880). Weinheim: Beltz.

Tent, L. (1969). *Die Auslese von Schülern für weiterführende Schulen*. Göttingen: Hogrefe.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, 1, 9–121.

Tiedemann, J., & Billmann-Mahecha, E. (2007). Zum Einfluss von Migration und Schulklassenzugehörigkeit auf die Übergangsempfehlung für die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 10, 108-120.

Trautwein, U., & Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind: Referenzgruppeneffekte bei Übergangentscheidungen. *Zeitschrift für Pädagogische Psychologie*, 21, 119-133.

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, 648-665.

Treiman, D. J. (1977). *Occupational prestige in comparative perspective*. New York: Academic Press

Vögeli-Manotvani, U. (1999). *Mehr fördern, weniger auslesen. Zur Entwicklung der schulischen Beurteilung in der Schweiz. Trendbericht der SKBF (Schweizerische Koordinationsstelle für Bildungsforschung)*. Aarau: SKBF.

Waldmann, M. R. (1996). Kognitionspsychologische Theorien von Begabung und Expertise. In F. E. Weinert (Hrsg.), *Psychologie des Lernens und der Instruktion* (S. 445-476). Göttingen: Hogrefe (Enzyklopädie der Psychologie, D/II/2).

Walter, G., Selter, C., Bonsel, M. & Bos, W. (2008). Mathematische Kompetenz im internationalen Vergleich: Testkonzeption und Ergebnisse. In W. Bos, M. Bonsel, J. Baumert, M. Prenzel, G. Walther, & C. Selter (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 49-85). Münster: Waxmann.

Watermann, R. & Klieme, E. (2002). Reporting Results of Large-Scale Assessment in Psychologically and Educationally Meaningful Terms - Construct Validation and Proficiency Scaling in TIMSS. *European Journal of Psychological Assessment*, 18, 190-203.

Watermann, R., Nagy, G. & Köller, O. (2004). Mathematische Leistungen 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 u. Budrich.

Wild, E. (2004). Häusliches Lernen. Forschungsdesiderate und Forschungsperspektiven. In D. Lenzen, J. Baumert, R. Qatermann & U. Trautwein (Hrsg.), *PISA und die Konsequenzen für die erziehungswissenschaftliche Forschung* (S. 37-64). Wiesbaden: VS Verlag für Sozialwissenschaften.

Wittwer, J., Saß, S. & Prenzel, M. (2008). Naturwissenschaftliche Kompetenz im internationalen Vergleich: Testkonzeption und Ergebnisse. In W. Bos, M. Bonsel, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenz von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 87-124). Münster: Waxmann.

Wu, M., Adams, R. J. & Wilson, M. (2000). *ACER ConQuest. Generalised item response modelling software manual*. Melbourne: ACER.

Ziegenspeck, J. (1978). Zensur und Zeugnis. In: Klauer, K.,J. (Hrsg.): *Handbuch der Pädagogischen Diagnostik* (Bd. 3, S. 621-632). Düsseldorf: Schwann.

Ziegenspeck, J. (1999). *Handbuch Zensur und Zeugnis in der Schule.: Historischer Rückblick, allgemeine Problematik, empirische Befunde und bildungspolitische Implikationen. Ein Studien-und Arbeitsbuch*. Klinghardt.

Über die Vodafone Stiftung

Erkennen. Fördern. Bewegen.

Die Vodafone Stiftung ist eine der großen unternehmensverbundenen Stiftungen in Deutschland und Mitglied einer weltweiten Stiftungsfamilie. Als eigenständige gemeinnützige Institution fördert und initiiert sie Projekte insbesondere mit Bildungsbezug. Ziel ist es, Impulse für den gesellschaftlichen Fortschritt zu geben, die Entwicklungen einer aktiven Bürgergesellschaft zu unterstützen und gesellschaftspolitische Verantwortung zu übernehmen. Dabei geht es der Stiftung vor allem darum, benachteiligten Kindern und Jugendlichen sozialen Aufstieg zu ermöglichen. Das Förderprofil der Stiftung steht unter dem Leitmotiv: „Erkennen. Fördern. Bewegen.“

Ansprechpartner



Dr. David Deißner

Leiter Programmbereich Think Tank,
Bildungsforschung und Integration

Vodafone Stiftung Deutschland gGmbH
Büro Berlin
Pariser Platz 6a
10117 Berlin

Telefon: +49 (0)30 20 61 76-19
Telefax: +49 (0)30 20 61 76-29
Mobil: +49 (0)173 205 00 75
E-Mail: david.deissner@vodafone.com

Impressum

Herausgeber:

Vodafone Stiftung Deutschland gemeinnützige GmbH
Am Seestern 1
40547 Düsseldorf
www.vodafone-stiftung.de

Verantwortlich:

Dr. Mark Speich
Dr. David Deißner (Redaktionelle Leitung)

Autoren:

Prof. Dr. Kai Maaz, Universität Potsdam
Prof. Dr. Franz Baeriswyl, Universität Freiburg (Schweiz)
Prof. Dr. Ulrich Trautwein, Universität Tübingen

Koordination:

Danyal Alaybeyoglu

Gestaltung:

trafodesign GmbH, Düsseldorf

Druck:

Druckstudio GmbH, Düsseldorf

Erscheinungsdatum der Studie:

14. Dezember 2011



Vodafone
Stiftung
Deutschland