

Soziologische Studien

Martin Neugebauer

Kommt es auf die Lehrkraft an?

Empirische Studien zur Selektion in das Lehramt und zu
Lehrereffekten bei der Entstehung ungleicher Bildungschancen

D 180 (Diss. Universität Mannheim)

Shaker Verlag
Aachen 2015

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Copyright Shaker Verlag 2015

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8440-3465-3

ISSN 1433-3546

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen

Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9

Internet: www.shaker.de • E-Mail: info@shaker.de

Vorwort

Dieses Buch ist eine leicht überarbeitete und gekürzte Fassung meiner publikationsbasierten Dissertation, die 2013 an der Fakultät für Sozialwissenschaften der Universität Mannheim angenommen wurde. Zu Beginn des Jahres 2015 sind nun alle empirischen Kapitel als eigenständig begutachtete Fachbeiträge erschienen. Mit dem vorliegenden Buch möchte ich die Gelegenheit wahrnehmen, die Einzelbeiträge zusammen mit einem Rahmenpapier als Gesamtwerk zu publizieren.

In der Themensetzung spiegelt sich mein Bestreben wider, aktuelle, bildungspolitisch relevante Fragen aufzugreifen und mit den Mitteln der empirischen Sozialforschung zu deren Klärung beizutragen. Wie sollte die Bewerberauswahl für das Lehramt gestaltet werden? Brauchen wir mehr Männer im Lehramt, um die vermeintlichen Bildungsmisserfolge der Jungen zu reduzieren? Sollte die Grundschulempfehlung verbindlich gestaltet sein, weil sich dadurch soziale Ungleichheiten im Zugang zum Gymnasium reduzieren lassen? Die Lektüre der Dissertationsschrift wird verdeutlichen, dass der Anspruch einer „wissenschaftlichen Klärung“ praktisch relevanter – und damit zumeist auch politisch drängender – Fragen ein schwieriges Unterfangen ist. Die gewonnenen Erkenntnisse sind mit vielerlei Einschränkungen behaftet, die die Kraft darauf basierender Handlungsempfehlungen begrenzt. Bei vielen Fragen, etwa zur Bedeutsamkeit von Merkmalen der Lehramtsanwärter für den Lernfortschritt von Schulkindern, befinden wir uns zudem erst am Anfang des Erkenntnisprozesses. Trotz dieser Schwierigkeiten glaube ich an die grundsätzliche Notwendigkeit, sich als Sozialwissenschaftler gesellschaftlich relevanten Fragen zu widmen, diese nach wissenschaftlichen Standards zu analysieren und – wenn es der kumulierte Wissensstand bei einer bestimmten Frage zulässt – fundierte Einschätzungen auf dem besten Stand des Wissens abzugeben.

Ich möchte einer Reihe von Menschen danken, die es mir ermöglicht haben, mich in den letzten Jahren praktisch relevanten sozialwissenschaftlichen Fragestellungen zu widmen. Mein herzlicher Dank gilt zuallererst meinem Doktorvater Walter Müller. Er hat mir nicht nur viele Freiräume zum wissenschaftlichen Arbeiten gelassen, sondern sie mit ebenso großem Engagement begleitet. Walter strahlt eine Begeisterung für wissenschaftliches Denken aus, die für mich sehr inspirierend ist. Irena Kogan danke ich für ihre Bereitschaft, sich als Zweitgutachterin mit dieser Arbeit intensiv und kritisch auseinanderzusetzen. Ich habe enorm vom persönlichen Austausch mit Kollegen und Freunden profitiert. Vor allem Steffen Schindler und Andreas Landmann möchte ich für zahlreiche Anregungen und hilfreiche Diskussionen herzlich danken. Ebenfalls danken möchte ich Markus Klein, Felix Weiss, David Reimer, Oliver Dickhäuser,

Susanne Hirth und Volker Stocké. Volker gebührt mein Dank auch, weil er mich für die Bildungssoziologie begeistert hat, während ich als Hilfskraft in seinem Forschungsprojekt arbeiten durfte. Jessica Walter, Florian Götsche, Maria Gerth und Juliane Stahl haben über die Jahre als wissenschaftliche Hilfskräfte bei mir gearbeitet. Für ihren hervorragenden Einsatz danke ich ihnen herzlich und hoffe, dass ich etwas von der Begeisterung weitergeben konnte.

Meinen Eltern verdanke ich so viel, ihnen ist diese Arbeit gewidmet. Zu guter Letzt danke ich von ganzem Herzen Franziska. Für ihre Unterstützung in kritischen Phasen, ihre Geduld in stressigen Phasen, ihre Liebe in sämtlichen Phasen.

Mannheim, im Februar 2015

Martin Neugebauer

Inhaltsverzeichnis

| | |
|--|----------|
| Kapitel 1 Kommt es auf die Lehrkraft an? - ein Überblick | 7 |
| 1.1 Einleitung | 7 |
| 1.2 Der Einfluss von Lehrkräften auf Lernerfolge von Schulkindern..... | 10 |
| 1.2.1 Wie groß sind Lehrereffekte? | 10 |
| 1.2.2 Welche Lehrermerkmale erzeugen Lernzuwächse bei Schülern? | 15 |
| 1.2.3 Fazit | 22 |
| 1.3 Die Genese professioneller Kompetenz und die Bedeutung von Eingangsvoraussetzungen..... | 22 |
| 1.4 Fragestellungen und Befunde – Teil 1..... | 25 |
| 1.5 Lehrereffekte und Bildungsungleichheiten – eine Systematik..... | 27 |
| 1.5.1 Lehrkräfte und primäre Disparitäten | 29 |
| 1.5.2 Lehrkräfte und sekundäre Disparitäten..... | 33 |
| 1.6 Fragestellungen und Befunde – Teil 2..... | 36 |
| 1.7 Literatur | 39 |

TEIL 1: *Selektion in das Lehramt*

| | |
|---|-----------|
| Kapitel 2 Wer entscheidet sich für ein Lehramtsstudium – und warum? Eine empirische Überprüfung der These von der Negativselektion in den Lehrerberuf..... | 49 |
| 2.1 Einleitung | 49 |
| 2.2 Theoretische Bezüge und empirische Evidenz..... | 51 |
| 2.2.1 Leistungsbezogene Eingangsmerkmale..... | 52 |
| 2.2.2 Motivationale Eingangsmerkmale..... | 53 |
| 2.2.3 Soziodemographische Eingangsmerkmale | 55 |
| 2.2.4 Zusammenfassende Hypothesen..... | 56 |
| 2.3 Methoden..... | 57 |
| 2.3.1 Stichprobe..... | 57 |
| 2.3.2 Studienentscheidung..... | 58 |
| 2.3.3 Eingangsmerkmale | 59 |
| 2.3.4 Statistische Analysen..... | 62 |
| 2.4 Ergebnisse..... | 63 |

| | | |
|--|---|-----------|
| 2.4.1 | Unterschiede zwischen den nicht-gymnasialen Lehrämtern | 63 |
| 2.4.2 | Überprüfung der vermuteten Gruppenunterschiede | 65 |
| 2.5 | Diskussion | 71 |
| 2.5.1 | Eingangsmerkmale von Lehramtsstudierenden im Vergleich | 71 |
| 2.5.2 | Einschränkungen der Studie | 73 |
| 2.5.3 | Bedeutung der Befunde | 74 |
| 2.6 | Appendix | 77 |
| 2.7 | Literatur | 82 |
| Kapitel 3 Who chooses teaching under different labor market conditions? Evidence from West Germany, 1980 – 2009 | | 87 |
| 3.1 | Introduction | 87 |
| 3.1.1 | Academic aptitude and motivation among teacher candidates | 88 |
| 3.1.2 | Changes over time in response to labor market shifts? | 89 |
| 3.2 | Methods | 92 |
| 3.2.1 | Data to analyze labor market conditions | 92 |
| 3.2.2 | Data to analyze characteristics of applicants to the profession | 93 |
| 3.3 | Results | 96 |
| 3.3.1 | The changing attractiveness of the teacher labor market | 96 |
| 3.3.2 | Changing pre-training characteristics of future teachers | 102 |
| 3.4 | Summary and discussion | 108 |
| 3.5 | Appendix | 111 |
| 3.6 | Literature | 117 |

TEIL 2: *Lehrkräfte und Bildungsungleichheiten*

| | | |
|---|---|------------|
| Kapitel 4 Unmasking the myth of the same-sex teacher advantage | | 121 |
| 4.1 | Introduction | 121 |
| 4.2 | Theoretical considerations and empirical evidence | 125 |
| 4.3 | Data and variables | 128 |
| 4.4 | Results | 132 |
| 4.5 | Robustness checks | 134 |
| 4.5.1 | Does selection mask a same-sex teacher effect? | 134 |
| 4.5.2 | Do subgroup differences mask a same-sex teacher effect? | 140 |

| | |
|---|------------|
| 4.6 Summary and conclusion | 142 |
| 4.7 Appendix | 146 |
| 4.8 Literature | 148 |
| Kapitel 5 Werden Jungen von Lehrerinnen bei den Übergangsempfehlungen für das Gymnasium benachteiligt? Eine Analyse auf Basis der IGLU-Daten..... | 153 |
| 5.1 Einleitung | 153 |
| 5.2 Die „Krise“ der Jungen, die Feminisierung des Lehrerberufs und die Bedeutung der Grundschulempfehlung | 156 |
| 5.2.1 Gibt es eine „Krise der Jungen“?..... | 156 |
| 5.2.2 Die Feminisierung der Lehrerschaft als Ursache?..... | 157 |
| 5.2.3 Die Rolle der Grundschulempfehlung..... | 159 |
| 5.3 Forschungsstand zur Feminisierungsthese | 160 |
| 5.4 Daten und Operationalisierungen | 162 |
| 5.5 Ergebnisse..... | 165 |
| 5.5.1 Werden Jungen bei der Gymnasialempfehlung von weiblichen Lehrkräften benachteiligt?..... | 166 |
| 5.5.2 Bewerten Lehrer strenger als Lehrerinnen?..... | 169 |
| 5.5.3 Wie wichtig ist die Zusammensetzung der Klassenkonferenz? | 170 |
| 5.5.4 Warum werden Jungen seltener für das Gymnasium empfohlen? | 171 |
| 5.6 Zusammenfassung und Diskussion | 172 |
| 5.7 Literatur..... | 175 |
| Kapitel 6 Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte. | 181 |
| 6.1 Einleitung | 181 |
| 6.2 Forschungsstand: Relative Bedeutung primärer und sekundärer Effekte | 183 |
| 6.3 Freier Elternwille versus verbindliche Grundschulempfehlung – substantielle Länderunterschiede..... | 183 |
| 6.4 Daten und Operationalisierungen..... | 185 |
| 6.4.1 Datenbasis..... | 185 |
| 6.4.2 Operationalisierungen..... | 186 |
| 6.5 Ergebnisse..... | 187 |

| | |
|--|------------|
| 6.5.1 Soziale Herkunft und schulische Leistungen beim Übergang auf das Gymnasium..... | 187 |
| 6.5.2 Effektdekomposition..... | 189 |
| 6.5.3 Lehrer oder Eltern: Der Einfluss der Entscheidungshoheit auf die relative Bedeutung der sekundären Effekte..... | 191 |
| 6.6 Zusammenfassung und Diskussion | 192 |
| 6.7 Appendix | 196 |
| 6.8 Literatur..... | 198 |
| Kapitel 7 Ausblick | 201 |
| 7.1 Selektion in das Lehramt..... | 202 |
| 7.2 Lehrkräfte und Bildungsungleichheiten | 203 |
| 7.3 Literatur..... | 206 |

Kapitel 1 Kommt es auf die Lehrkraft an? - ein Überblick

1.1 Einleitung

Mit der Veröffentlichung der PISA-2000-Ergebnisse vor rund 12 Jahren (Deutsches PISA-Konsortium 2001) rückten zwei Problembereiche des deutschen Bildungssystems schlagartig in das öffentliche Blickfeld. Einerseits das *durchschnittlich schlechte* Abschneiden deutscher Schülerinnen und Schüler¹ im internationalen Leistungsvergleich. Andererseits die überdurchschnittlich *ungleiche Verteilung* des Bildungserfolges entlang soziodemographischer Merkmale, insbesondere der sozialen Herkunft. Die breite, zum Teil aufgeregte Rezeption der PISA-Befunde führte unweigerlich dazu, dass auch Lehrkräfte verstärkt in den Fokus rückten (Terhart 2008) – schließlich sind sie die zentralen Akteure im Schulsystem. Ihnen obliegt die Kernaufgabe der Schule (vgl. Fend 1981, S. 13-49; Parsons 1959): die Vermittlung von Wissen, Fähigkeiten und Werten an Schüler. Ebenso üben sie über die Vergabe von Noten, Zeugnissen und Übergangsempfehlungen eine zweite zentrale Funktion von Schule aus: Die Bewertung und Selektion, mittels derer der Zugang zu weiterführender Bildung und beruflichen Positionen reguliert wird. Lehrer sind nicht nur die zentralen Akteure, sondern gleichzeitig der größte Kostenfaktor. Die Personalkosten für Schulen machen rund 80 Prozent des gesamten öffentlichen Budgets für Bildung aus (Klemm 2008).

Es verwundert daher kaum, dass Lehrkräften eine besondere Bedeutung für die Qualität des Bildungssystems, ebenso wie für dessen Problembereiche, zugemessen wird. Die intuitive Annahme der Bedeutsamkeit von Lehrkräften kann inzwischen als empirisch belegt gelten (vgl. Abschnitt 1.2.1). Lehrkräfte unterscheiden sich – wenig überraschend – in ihrer Kompetenz, Lernzuwächse bei Schulkindern zu erzeugen. Sie beeinflussen damit das Ausmaß der beiden eingangs genannten Problembereiche. Schafft ein Land es nicht, kompetente Lehrkräfte einzustellen, wird die durchschnittliche Qualität des Unterrichts und damit der Lerngelegenheiten von Schülern, *ceteris paribus*, niedriger ausfallen. Ist die Variabilität der Lehrerkompetenz in einem Land groß und verteilen sich diese in systematischer Weise ungleich über die verschiedenen Schulen, steigt die Wahrscheinlichkeit, ungleiche Bildungschancen zu generieren. Auf der Suche nach Möglichkeiten zur Verbesserung der beiden genannten Problembereiche ist es daher naheliegend, sich auf den Lehrer zu konzentrieren. Damit soll jedoch nicht suggeriert

¹ Im Sinne eines besseren Leseflusses wird im Folgenden nur die männliche Geschlechtsform verwendet. Selbstverständlich sind immer beide Geschlechter gemeint.

werden, die deutschen Lehrer seien an der hinlänglich beklagten Bildungsmisere „schuld“. Vielmehr möchte ich auf Basis von empirischen Studien einige Aspekte zur Bedeutsamkeit von Lehrkräften für die beiden Problembereiche genauer analysieren. Dabei werden auch einige typische Vorurteile über Lehrkräfte auf ihren Wahrheitsgehalt geprüft. Obwohl in den vergangenen Jahren eine Intensivierung der Forschung zum Lehrerberuf zu verzeichnen ist, muss der empirisch gesicherte Erkenntnisstand als lückenhaft beschrieben werden (Terhart 2012a). Mit den in dieser publikationsbasierten Dissertation versammelten Aufsätzen sollen einige dieser Lücken angegangen werden.

Ausgehend von der Tatsache, dass die Qualität der Lehrkräfte auch von der Rekrutierung geeigneter Bewerber abhängt, beschäftigen sich die beiden Arbeiten im ersten Teil der Dissertation mit Fragen der (Selbst)selektion in das Lehramt. Wer entscheidet sich für ein Lehramtsstudium? Welche leistungsbezogenen, motivationalen und soziodemographischen Merkmale bringen Lehramtsaspiranten mit ins Studium und inwiefern unterscheiden sie sich von anderen Studierenden? Wie haben sich die Eingangsmerkmale von Lehramtsanwärtern im Laufe der vergangenen 30 Jahre gewandelt? Inwiefern können veränderte Arbeitsmarktbedingungen auf dem Teilarbeitsmarkt Schule eine Rolle dabei spielen, für wen der Lehrerberuf attraktiv ist? Diese Fragen werden empirisch analysiert und vor dem Hintergrund der Bedeutsamkeit von Eingangsmerkmalen für den Kompetenzerwerb von Lehrkräften – und damit langfristig für den Lernerfolg von Schulkindern – diskutiert. Sie liefern damit Hintergrundwissen zur Beantwortung der Frage, wie die Auswahl zukünftiger Lehramtsanwärter gestaltet werden sollte. Nicht zuletzt ist empirisches Wissen zur Wahl des Lehramtes auch vor dem Hintergrund aktueller Debatten zur Deckung des zukünftigen Lehrerbedarfs relevant (KMK 2013). Die im zweiten Teil der Dissertation versammelten Aufsätze knüpfen an den zweiten Problembereich – die ungleichen Bildungserfolge – an. Fragen nach der Entstehung ungleicher Bildungschancen stehen im Zentrum soziologischer Bildungsforschung. Erstaunlicherweise wird die Rolle von Lehrkräften dabei kaum thematisiert. Entsprechend spärlich ist der empirische Forschungsstand zu Lehrereffekten und Bildungsungleichheiten. Dabei gibt es eine ganze Reihe möglicher Wirkungsweisen, wie die Systematisierung in Abschnitt 1.6 zeigt. In zwei Aufsätzen geht es um geschlechtliche Bildungsungleichheiten. Die Arbeiten untersuchen, inwiefern die Feminisierung des Lehrerberufs für die Bildungsmisserfolge der Jungen verglichen mit den Mädchen (mit)verantwortlich sein kann. Die Studien liefern Antworten auf die Frage, inwiefern die Rekrutierung von mehr männlichen Lehrkräften zu einem Abbau der Nachteile der Jungen beitragen kann. Ein weiterer Aufsatz thematisiert soziale Bildungsungleichheiten

am Übergang auf das Gymnasium. Dabei geht es um die Rolle der von Grundschullehrern vergebenen Schulformempfehlung für die Sekundarstufe. Während in der Mehrzahl der Bundesländer die Schulformempfehlung eine unverbindliche Empfehlung im eigentlichen Wortsinn darstellt und somit die Entscheidungshoheit bei den Eltern liegt, hat sie in einigen einen verbindlichen Charakter – und verschiebt damit die Entscheidungsgewalt zu Gunsten der Lehrkräfte. Der Aufsatz analysiert inwiefern diese institutionelle Variation primäre und sekundäre Herkunftseffekte beeinflusst und damit zu einer größeren oder geringeren sozialen Bildungsungleichheit am Übergang auf das Gymnasium beiträgt. Zum Abschluss der Dissertationsschrift gebe ich einen kurzen Ausblick auf zukünftige Forschungsfragen. Ethnische Bildungsungleichheiten, obwohl ebenso relevant, werden in dieser Dissertation nicht behandelt.

Den empirischen Beiträgen vorgelagert ist dieses einleitende Kapitel. Es ist etwas breiter aufgespannt und vermittelt einen Überblick zum Forschungsgebiet, der für das Verständnis und die Einordnung der nachfolgenden Kapitel notwendig ist. In einem Literaturüberblick stelle ich einleitend den Forschungsstand dar zum Einfluss von Lehrkräften auf Lernzuwächse von Schulkindern (Abschnitt 1.2). Darin wird zum einen der Frage nachgegangen, wie groß Lehreffekte sind. Zum anderen wird dargestellt, welche konkreten, beobachtbaren Lehrermerkmale in Zusammenhang stehen mit Lernzuwächsen bei Schülern. Abschnitt 1.3 widmet sich der Frage, wie Lehrerkompetenzen entstehen – und damit auch verändert werden können. Es wird deutlich werden, dass die Kompetenz von Lehrern auch davon abhängt, welche Personen sich für das Lehramtsstudium und den damit verbundenen Beruf entscheiden. Damit ist das notwendige Hintergrundwissen für die empirischen Arbeiten in Kapitel 2 und 3 aufbereitet, die sich mit den Eingangsvoraussetzungen von Lehramtsanwärtern befassen. Die zentralen Ergebnisse dieser Arbeiten werden in Abschnitt 1.4 vorgestellt. Mit Blick auf die im zweiten Teil versammelten Aufsätze entwickle ich in Abschnitt 1.5 eine Systematik, wie Lehrkräfte die Entstehung ungleicher Bildungschancen beeinflussen können. Die Systematik hat zwei Zielsetzungen. Zum einen dient sie als Struktur, um die im zweiten Teil der Arbeit versammelten Arbeiten einzuordnen (Kapitel 4, 5 und 6). Deren zentrale Ergebnisse werden in Abschnitt 1.6 vorgestellt. Zum anderen zeigt sie die Forschungslücken auf und skizziert damit in Ansätzen bereits ein zukünftiges Forschungsprogramm zur Rolle von Lehrkräften bei der Entstehung ungleicher Bildungserfolge.

1.2 Der Einfluss von Lehrkräften auf Lernerfolge von Schulkindern²

Der Lernerfolg von Schülern ist von einer Vielzahl an Einflussfaktoren abhängig. Dabei zeigt sich immer wieder, dass die größte Variation im Lernerfolg von Faktoren außerhalb der Schule abhängt: insbesondere von den individuellen Fähigkeiten und der familiären Herkunft der Schüler (Coleman et al. 1966; Hosenfeld et al. 2001; Rowan et al. 2002; Scheerens und Bosker 1997). Solche Faktoren sind von Seiten des Bildungssystems nur schwer zu beeinflussen (OECD 2005). Von den schulischen Faktoren, die möglicherweise besser zu beeinflussen sind, kommt der Kompetenz der Lehrkräfte nach weitverbreiteter Meinung die größte Bedeutsamkeit zu (Darling-Hammond 2000; Hattie 2009; Schleicher 2011). Aber wie groß sind Lehrereffekte tatsächlich? Welchen Spielraum bieten Interventionen, die bei den Lehrkräften ansetzen? Die empirische Forschung zu den Determinanten von Lernerfolg hat die Frage nach dem Einfluss von Lehrkräften von zwei Seiten in Angriff genommen. Zum einen wird der *Gesamtbeitrag von Lehrkräften* bei der Erzeugung schulischer Lernerfolge, unabhängig von beobachtbaren Lehrermerkmalen, abgeschätzt (Abschnitt 1.2.1). Zum anderen wird untersucht, inwiefern *konkrete Merkmale oder Verhaltensweisen von Lehrkräften* schulischen Lernerfolg fördern (Abschnitt 1.2.2).³

1.2.1 Wie groß sind Lehrereffekte?

Eine analytische Strategie um die Größe von Lehrereffekten abzuschätzen sind aus der Bildungsökonomie stammende *Value-Added* Studien (im Überblick: Hanushek und Rivkin 2010a). Dabei wird versucht, anhand umfassender (bislang fast ausschließlich amerikanischer) Paneldatensätze den Einfluss von Lehrkräften auf die *Lernzuwächse* von Schülern zu berechnen, gemessen über Veränderungen in standardisierten Schulleistungstests. Durch die längsschnittliche Analyse von Zuwächsen wird versucht, den konfundierenden Einfluss von vorgelagerten Faktoren, die das Lernniveau bis zu Beginn eines Schuljahres beeinflusst haben, weitgehend zu eliminieren. Die Berechnung des von einer Lehrkraft erzeugten „Mehrwertes“

² Die Integration dieses Literaturüberblicks weicht etwas vom üblichen Aufbau solcher Einleitungskapitel ab, da er die empirischen Arbeiten nicht im engeren Sinne „rahmt“, sondern als eigenständiges Kapitel ergänzt.

³ Schulischer Lernerfolg kann neben leistungsbezogenen Kriterien auch an motivationalen, emotionalen oder persönlichkeitsbezogenen Kriterien aufseiten der Schüler festgemacht werden. Wie die überwiegende Mehrheit der Studien beziehe ich mich in dieser Arbeit jedoch auf leistungsbezogene Kriterien, vor allem den *Lernzuwachs* auf Seiten der Schüler.

erfolgt dabei unabhängig von konkret messbaren Merkmalen der Lehrkräfte. Eine ‚value-added‘ Modellierung ist möglich, wenn mindestens zwei Messzeitpunkte vorliegen, Leistungsmessungen für mehrere Schulklassen und Schulen vorhanden sind und wenn bestimmte Lehrkräfte diesen Schulklassen zugeordnet werden können. Der Grundgedanke der ‚Value-added‘-Modelle (VAM) besteht darin, den Lernzuwachs eines Schülers innerhalb eines Schuljahres bei einem durchschnittlichen Lehrer auf Basis seines vorherigen Lernerfolges und anderer Prädiktoren (etwa der soziale Herkunft oder der Klassenkomposition) vorherzusagen. Im nächsten Schritt wird der durchschnittlich erwartete Lernzuwachs verglichen mit dem tatsächlichen Lernzuwachs. Die Differenz wird für jeden Schüler berechnet. Der Durchschnitt dieser Abweichungen innerhalb einer Klasse ist der Beitrag, der auf eine einzelne Lehrkraft zurückgeführt werden kann, wenn für andere Fehlerkomponenten kontrolliert wird. Es zeigt sich, dass es systematische Variationen gibt. So erzeugen manche Lehrer überdurchschnittliche Lernzuwächse, während andere unterdurchschnittliche Lernzuwächse erzeugen. Die Lehrer lassen sich so in eine Rangfolge bringen, nach der Größe der von ihnen erzeugten Lernzuwächse. Es gibt eine ganze Reihe an Spezifikationen, mit denen dieser Grundgedanke umgesetzt wird. Sie unterscheiden sich in der Art der Modellierung, der Messung der vorgelagerten Lernleistung, sowie der Anzahl und Auswahl der mit den Lernzuwächsen korrelierten Kovariaten (kontrolliert werden muss vor allem für Merkmale der Klassenkomposition und -größe, sowie die nicht-zufällige Zuweisung von Schülern zu Lehrern; vgl. McCaffrey et al. 2004 für einen Überblick). In den letzten Jahren sind die ökonometrischen Verfahren zunehmend komplexer geworden, zudem sind die zu Grunde liegenden Datensätze – zumeist administrative Daten der Schulbezirke – immer umfangreicher geworden. Beispielsweise beruhen die in Chetty et al. (2011) berichteten Ergebnisse auf Daten von 2.5 Millionen Schülern. Mit Hilfe der VAM lässt sich folgende Frage beantworten: Wenn eine Schulklasse Lehrer A anstatt Lehrer B hätte, wie stark würden sich die Testscores der Schulkinder in dieser Schulklasse im Durchschnitt unterscheiden? Um ein einheitliches Maß für die Effektstärke zu erhalten, wird üblicherweise der Vergleich zwischen Lehrern vorgenommen, die eine Standardabweichung auseinander liegen, also beispielsweise ein durchschnittlicher Lehrer mit einem Lehrer, der eine Standardabweichung über dem Durchschnitt liegt. Tabelle 1 gibt einen Überblick, über die Größe der so ermittelten Effekte. Die Ergebnisse für Lesen bedeuten, dass ein Schüler innerhalb eines Schuljahres seine Schulleistungen um rund 0.14 Standardabweichungen verbessern würde (gemittelter Wert über alle Studien), wenn er, bei ansonsten identischen Voraussetzungen, anstelle eines durchschnittlichen kompetenten Lehrers einen Lehrer hätte, der innerhalb der Kompetenzverteilung um eine Standardabweichung höher läge.

Tabelle 1: Standardabweichung von Lehrereffekten („value-added“) gemessen in Standardabweichungen von Lernzuwächsen bei Schülern innerhalb eines Schuljahres

| Studie | Land | Value-Added | |
|-----------------------------|------|-------------|------------|
| | | Lesen | Mathematik |
| Rockoff (2004) | USA | 0.10 | 0.11 |
| Nye et al. (2004) | USA | 0.26 | 0.36 |
| Rivkin et al. (2005) | USA | 0.10 | 0.11 |
| Aaronson et al. (2007) | USA | - | 0.13 |
| Kane et al. (2008) | USA | 0.08 | 0.11 |
| Jacob und Lefgren (2008) | USA | 0.12 | 0.26 |
| Kane und Staiger (2008) | USA | 0.18 | 0.22 |
| Koedel und Betts (2007) | USA | - | 0.23 |
| Rothstein (2010) | USA | 0.11 | 0.15 |
| Hanushek und Rivkin (2010b) | USA | - | 0.11 |
| Papay (2011) | USA | 0.05-0.21 | - |
| Chetty et al. (2011) | USA | 0.08 | 0.12 |
| Corcoran et al. (2011) | USA | 0.13-0.26 | 0.15-0.25 |

Um aktuelle Studien erweiterte Tabelle aus Hanushek und Rivkin (2010a).

Zur Einordnung der Effektstärke ist es hilfreich, die Standardabweichungen auf eine bekannte Skala umzurechnen. Umgerechnet auf die IGLU-Skala aus der deutschen IGLU-Erhebung von 2001 ($M = 539$, $SD = 67$) entsprechen 0.14 Standardabweichungen etwa 9 Punkten. Der Lernzuwachs eines Schuljahres von Klasse 3 zu Klasse 4 in einer deutschen Grundschule entspricht 43 Punkten (Bos et al. 2004a, S. 56). Der Unterschied zwischen einem durchschnittlichen und einem um eine Standardabweichung kompetenteren Lehrer korrespondiert also in etwa mit dem Lernzuwachs von 2-3 Monaten. In Mathematik sind die Effekte tendenziell stärker. Ein in der Größe vergleichbarer Effekt ist die Differenz der Lesekompetenz zwischen Jungen und Mädchen am Ende der 4. Jahrgangsstufe. Im Jahr 2001 betrug diese 13 Punkte auf der IGLU-Skala, im IGLU Zyklus 2006 7 Punkte. Diese Effektstärken mögen zunächst klein erscheinen. Allerdings ist zu berücksichtigen, dass sich die Effekte kumulieren, wenn ein Schüler mehrere Jahre von einem kompetenten Lehrer unterrichtet wird (wie es in deutschen Grundschulen üblich ist). Auch wenn die amerikanischen Befunde nicht ohne weiteres auf Deutschland übertragen werden können, liefern die zur Verfügung stehenden Befunde und die (etwas freimütige) Umrechnung in greifbare Größen empirische Belege für die These, dass Lehrkräfte deutlich in ihrer Kompetenz variieren. Ob ein Schüler von einer mehr oder weniger kompetenten Lehrkraft unterrichtet wird, trägt demnach entscheidend zum langfristigen Lernerfolg eines Schülers bei. Vergleichbare Analysen mit deutschen Daten wären wünschenswert, liegen aber meines Wissens bislang nicht vor. Eine Ausnahme ist die COACTIV-Studie, die anschließend an PISA 2003 15-jährige Schüler ($n = 4353$) an deutschen Sekundarschulen ein Jahr später wiederholt

im Fach Mathematik testet (Baumert et al. 2010). In dieser Analyse werden etwas andere Modelle spezifiziert als in den amerikanischen ‚value-added‘ Studien, die Befunde lassen sich jedoch in vergleichbare Effektgrößen umrechnen. Baumert et al. (2010) modellieren den Lernerfolg am Ende der 10. Klasse auf Basis von Mehrebenenmodellen. Unter Kontrolle der zentralen individuellen Merkmale (mathematisches Vorwissen in Klasse 9, Intelligenz, Leseleistung, soziale Herkunft und Migrationshintergrund) können die Autoren 64% der Varianz der Mathematikleistungen am Ende der 10. Klasse vorhersagen. In einem so spezifizierten Modell verbleiben noch 4.6 % Residualvarianz der Mathematikleistungen, die durch Variablen der Klassenebene und damit der Lehrkraft erklärt werden können. Wenn man davon ausgeht, dass es sich bei dieser Residualvarianz vor allem um den Effekt der (unbeobachteten) Lehrkompetenz handelt, entspricht die Quadratwurzel aus dem Anteil erklärter Varianz von $\Delta R^2=0.046$, also $\Delta R=0.21$, in etwa der Standardabweichung des Lehrereffektes gemessen in Standardabweichungen von Lernzuwächsen der Schüler (vgl. Nye et al. 2004, S. 240). Diese Effektgröße ist vergleichbar mit den Schätzern der amerikanischen Studien (siehe Tabelle 1).

Die Abschätzung von Lehrereffekten mittels VAM sieht sich einer Reihe von Kritikpunkten ausgesetzt (Papay 2011; Corcoran et al. 2011; Rothstein 2010). So ergeben sich je nach Modell-Spezifikation andere ‚value-added scores‘, auch wenn diese in der Regel stark korreliert sind. Zudem werden manche Lehrer in einem Jahr als besonders effektiv eingestuft, im Folgejahr jedoch nicht. Das legt nahe, dass dieselbe Lehrkraft nicht alle Schulklassen gleich effektiv unterrichten kann. Damit stellt sich die Frage, inwiefern es sinnvoll ist, einen ‚value-added‘ pro Lehrer zu berechnen und als stabiles Personenmerkmal zu interpretieren. Begegnet wird diesem Problem, indem mehrere ‚value-added scores‘ für dieselbe Lehrkraft auf Basis multipler Schülerkohorten berechnet werden, um so den permanenten ‚value-added‘ von variierenden Einflüssen zu separieren (Rokoff 2004). Kritisiert wird ebenfalls die Instabilität hinsichtlich der verwendeten Tests. Corcoran et al. (2011) weisen nach, dass der ‚value-added score‘ eines Lehrers bei identischen Schülern auf Basis zweier unterschiedlicher Leseleistungs-Tests lediglich eine Korrelation von 0.5 (0.59 in Mathematik) aufweist. Dabei spielen zum einen unterschiedliche Testitems eine Rolle. Zum anderen beeinflusst die Tagesform einzelner Schüler die Präzision der Messung ihres tatsächlichen Wissens, ebenso wie externe Umstände am Tag der Testung, die nie vollständig kontrolliert werden können (z. B. der ‚bellende Hund vor dem Klassenzimmer am Tag des Tests‘). Darüber hinaus wird vermutet, dass auch die Bedeutung des Testes bei der Leistungsbewertung der Lehrer eine Rolle spielt. Es besteht Grund zur Sorge, dass Lehrer systematisch die Leistungen in sogenannten ‚high-stake‘ Tests beeinflussen, an die

in einigen Schulbezirken eine leistungsbezogene Vergütungskomponente gekoppelt ist (Stichwort: ‚teaching to the test‘). Trotz dieser Einwände, vor allem was die Einordnung einzelner Lehrkräfte und eine darauf basierende Entlohnung angeht, liefern die ‚value-added‘-Studien nach überwiegender Meinung stabile Schätzer (Hanushek und Rivkin 2010a; McCaffrey et al. 2004) der quantitativen Bedeutung, die Lehrkräften bei der Genese von Lernzuwächsen innerhalb eines Schuljahres zukommt.

Die zitierten Studien, insbesondere die von Chetty und Kollegen (2011) durchgeführte Untersuchung, haben eine Reihe weiterer interessanter Befunde zu Tage gefördert. Sie zeigen, dass sich die Zuweisung zu einem effektiven Lehrer in einem Schuljahr auch auf den Schulerfolg in den Folgejahren auswirkt (vgl. auch Kane und Staiger 2008). Zudem liefern sie Hinweise, dass Schüler, die von Lehrkräften mit hohem ‚value-added‘ unterrichtet wurden, eine höhere Wahrscheinlichkeit haben, eine Hochschule zu besuchen, mehr Geld zu verdienen und seltener als Teenager schwanger zu werden. Eine kürzlich vorgelegte Studie von Jennings und DiPrete (2010) erweitert die ‚value-added‘-Forschung, indem der Beitrag von Kindergartenpädagogen auf *soziale* Fertigkeiten und Verhaltensweisen von Kindern abgeschätzt wird.⁴ Demnach leisten unterschiedliche Pädagogen einen deutlich unterschiedlichen Beitrag bei der Förderung solcher Merkmale. Zudem erleben Schüler mit größeren sozialen Fertigkeiten größere Lernzuwächse hinsichtlich ihrer akademischen Fertigkeiten. Interessanterweise sind es nicht dieselben Pädagogen, die gleichsam soziale Fertigkeiten und akademische Fertigkeiten fördern. Damit wird, so die Autoren, die weitverbreitete Meinung angezweifelt, dass es „gute“ und „schlechte“ Pädagogen gibt. Die Frage sollte vielmehr lauten: gut worin? (Jennings und DiPrete 2010, S. 156).

Insgesamt informieren die bildungsökonomischen ‚value-added‘-Studien, welchen Beitrag Lehrer bei der Erzeugung von Lernzuwächsen haben. Sie stecken damit wie eine ‚black-box‘ den Spielraum ab, indem konkrete, beobachtbare Lehrermerkmale eine Rolle spielen können. Über den Einfluss solcher beobachtbarer Merkmale sagen diese Studien jedoch nichts aus. Darüber informiert der folgende Abschnitt.

⁴ Jennings und DiPrete (2010) sprechen von *social and behavioral skills* und meinen damit Merkmale wie eine positive Lernorientierung, einen angemessenen Kommunikationsstil und die Abwesenheit von Verhalten, das den Unterricht stört.

1.2.2 Welche Lehrermerkmale erzeugen Lernzuwächse bei Schülern?

Obwohl inzwischen ein breiter Konsens zur Bedeutsamkeit von Lehrkräften herrscht, bleibt vielfach unklar, welche konkreten Eigenschaften von Lehrkräften für die Genese von Leistungszuwächsen verantwortlich sind. Damit ist ein grundlegendes Dilemma der Lehrerforschung angesprochen. Wie sollen Fragen nach den Ausbildungsinhalten von Lehramtsanwärtern, ebenso wie nach der Auswahl geeigneter Bewerber beantwortet werden, wenn unklar ist, welche Eigenschaften eine Lehrkraft „effektiv“ oder „kompetent“ machen im Sinne von „dem Lernerfolg der Schüler zuträglich“? Ebenfalls geklärt werden muss die Frage danach, wie dem Lernerfolg der Schüler zuträglich Eigenschaften entstehen – und damit: beeinflusst werden können? Der folgende Abschnitt gibt einen Überblick über die erste der beiden Fragen, in Abschnitt 1.3 wird dann die Frage nach der Entstehung der diskutierten Lehrermerkmale aufgegriffen.

Hinsichtlich der Lehrermerkmale wird in der Schuleffektivitätsforschung häufig zwischen *Merkmale der Lehrkraft* (sogenannte ‚teacher‘ oder ‚input‘ Variablen wie z. B. das Fachwissen) und *Merkmale des Lehrerhandelns im Unterricht* (sogenannte ‚teaching‘ oder ‚process‘ Variablen wie z. B. kognitiv aktivierender Unterricht) unterschieden (vgl. z. B. Goe 2007). Die folgende Übersicht beschränkt sich auf Merkmale der Lehrkraft, von denen angenommen werden kann, dass sie ein kompetentes Lehrerhandeln überhaupt erst ermöglichen. Eine aktuelle Übersicht zur Effektivität bestimmter Handlungen und Interaktionsformen von Lehrkräften findet sich unter anderem bei Seidel und Shavelson (2007).⁵ Ich lege in diesem Abschnitt den Schwerpunkt auf Studien, die den Lernerfolg längsschnittlich und unter Kontrolle des Vorwissens messen, da nur solche Studien zwischen (dem Klassenbesuch zeitlich vorgelagerten) Selektionseffekten und eigentlichen Wirkungen der Lehrer im Klassenraum unterscheiden können.

Lehrerkompetenzmodelle bilden die konzeptionelle Grundlage zur Strukturierung der Merkmale, die eine Lehrkraft befähigen, Lernzuwächse zu erzeugen. Aufbauend auf früheren Systematisierungen (Bromme 1997; Shulman 1987) hat die Arbeitsgruppe COACTIV um Jürgen Baumert und Mareike Kunter ein innerhalb der deutschsprachigen Lehrerforschung einflussreiches Modell vorgeschlagen, das hier als Ordnungsrahmen für die Strukturierung der

⁵ Die systematische Analyse zwischen Merkmalen des Lehrerhandelns und Schülerleistungen firmiert unter dem Begriff Prozess-Produkt-Paradigma. Im sogenannten Experten-Paradigma werden Studien dieser Tradition ergänzt um Analysen von lernwirksamen Merkmalen der Lehrkraft, wie etwa das professionelle Wissen (vgl. z. B. Helmke 2012, S. 45ff.). Auf letzteres Paradigma beziehen sich die hier diskutierten Befunde.

empirischen Befunde verwendet werden soll (Baumert und Kunter 2006, 2011).⁶ Um einen Unterricht zu inszenieren, der *in gewisser Regelmäßigkeit kognitiv herausfordernde Lerngelegenheiten* schafft, den Schülern sorgfältige *Lernunterstützung* bietet und *Ordnung und Strukturen* herstellt, die Lernen unterstützen, sind laut dem Modell eine Reihe persönlicher Voraussetzungen notwendig, die eine erfolgreiche Bewältigung des Unterrichts erlauben. Diese „Aspekte professioneller Lehrerkompetenz“ umfassen

- a) professionelles Wissen (Fachwissen, fachdidaktisches Wissen und fachübergreifendes pädagogisches Wissen)
- b) motivationale Orientierungen (vor allem intrinsische Motivation und Enthusiasmus),
- c) eine ausgeprägte Selbstregulationsfähigkeit (d.h. Belastungsregulation) ebenso wie bestimmte
- d) Werthaltungen und Überzeugungen.

1.2.2.1 Professionelles Wissen

Im Zentrum des COACTIV-Modells wie auch anderer Kompetenzmodelle steht das professionelle Wissen. Ihm soll hier der größte Raum zugestanden werden, auch weil die empirischen Befunde zu dieser Dimension am überzeugendsten sind. Theoretisch wird unterschieden zwischen Fachwissen (einem tiefen fachlichen Verständnis der zu vermittelnden Inhalte), fachdidaktischem Wissen (Wissen, um den Schülern die fachlichen Inhalte vermitteln zu können) und fachübergreifendem pädagogisch-psychologischen Wissen (Wissen über die Gestaltung von Lehr-Lern-Prozessen, über Klassenführung, über Entwicklung und Lernen, sowie über Diagnostik und Leistungsbeurteilung) (Baumert und Kunter 2011).⁷

Verschiedene Studien können die zentrale Bedeutung des professionellen Wissens belegen. Auf Basis der bereits angesprochenen COACTIV-Studie, die anknüpfend an PISA 2003 die Lernzuwächse deutscher Schüler (n = 4353) von der neunten zur zehnten Klasse in Mathematik über 2 Messzeitpunkte längsschnittlich misst, prüfen Baumert et al. (2010) den Einfluss von Fach- und fachdidaktischem Wissen der Lehrer (n = 181), mittels eines etwa zweistündigen

⁶ Das Modell hat sich auch deshalb als leistungsfähig erwiesen, weil das COACTIV-Team empirische Belege für die Vorhersagekraft der von ihr postulierten Kompetenzdimensionen gleich mitliefert, wie die folgenden Ausführungen zeigen.

⁷ Als weitere nicht vertiefend behandelte Wissensdimensionen weist das COACTIV-Modell Beratungswissen (Wissen über professionelle Kommunikation mit Schülern und Eltern) und Organisationswissen (Wissen über die Funktionslogik des Bildungssystems) aus.

‚paper-pencil‘-Tests für die teilnehmenden Lehrer. Es zeigt sich, dass beide Wissensdimensionen den Lernerfolg am Ende der zehnten Klasse vorhersagen, wobei der stärkere Effekt vom fachdidaktischen Wissen ausgeht. So lassen sich 39 % der Zwischen-Klassen-Varianz im Lernerfolg auf Unterschiede im fachdidaktischen Wissen der Lehrkraft zurückführen. Getrennte Analysen nach Schulform zeigen zudem, dass besonders Schüler auf niedrigeren Sekundarschulformen vom Wissen der Lehrkräfte profitieren. Anlass zur Sorge gibt dieser Befund deshalb, weil gerade Lehrkräfte auf den niedrigeren Sekundarschulformen über deutlich geringeres Wissen verfügen. Der Mittelwertsunterschied zwischen Lehrern mit Gymnasiallehrausbildung und Lehrern mit einer Ausbildung für nicht-gymnasiale Sekundarschulformen beträgt 1.26 Standardabweichungen beim Fachwissen und 0.43 Standardabweichungen beim fachdidaktischen Wissen. Die Ergebnisse aus Deutschland ergänzen internationale Studien, die ebenfalls die Bedeutung von mathematischem Wissen der Lehrer auf Lernzuwächse der Schüler identifizieren konnten. Im Gegensatz zu Baumert et al. (2010) wird in diesen Studien allerdings nicht zwischen fach- und fachdidaktischem Mathematikwissen unterschieden. Stattdessen werden die Wissensfacetten aufgrund ihrer hohen Korrelation als ein Konstrukt operationalisiert.⁸ Hill et al. (2005) zeigen, dass die Lernzuwächse von amerikanischen Grundschulkindern größer sind, wenn ihre Lehrer über höheres Wissen verfügen. Unter Verwendung des gleichen Wissenstests für Lehrer bestätigen Rockoff et al. (2011) den Befund mit Daten von Grund- und Mittelschullehrern aus New York, allerdings sind die Zusammenhänge zwischen dem Wissen der Lehrer und Lernzuwächsen der Schüler eher schwach. Ein um eine Standardabweichung höheres Fachwissen erhöht den Lernzuwachs um 0.03 Standardabweichungen. Der Wissenstest sagt Lernzuwächse dennoch besser voraus als allgemeine kognitive Fähigkeiten, Big Five Persönlichkeitsmerkmale, oder Selbstwirksamkeitsüberzeugungen der Lehrer. Metzler und Wößmann (2012) finden den gleichen Effekt bei peruanischen Sechstklässlern. Demnach erhöht ein um eine Standardabweichung höheres Wissen die Mathematikleistungen um 0.09 Standardabweichungen. Für Lesekompetenzen sind die Effekte jedoch deutlich kleiner und können nicht zufallskritisch abgesichert werden. Einen statistisch signifikanten Effekt des professionellen Lehrerwissens auf die Leseleistung findet Kelcey (2011) anhand einer Stichprobe von Grundschulkindern aus Problem-Schulbezirken in Michigan. Demnach erhöht ein um eine Standardabweichung höheres Wissen das Leseverständnis der Grundschüler um 0.07 Standardabweichungen, was in etwa dem Lernzuwachs von 2 zusätzlichen Schulwochen entspricht (Kelcey 2011, S. 480). Kein signifikanter Zusammenhang zeigt

⁸ Die Betonung in diesen Tests liegt eher auf dem fachdidaktischen Wissen, dem Wissen also, das Lehrkräfte benötigen, um curriculare Lerninhalte effektiv und verständlich zu vermitteln.

sich jedoch bei der Analyse der Lernzuwächse bei einem Wortanalyse-Test. Die Verallgemeinerbarkeit auf andere Grundschulen ist aufgrund der Besonderheiten der Stichprobe nicht gewährleistet.

Als dritte Wissensdimension wird häufig das fachübergreifende pädagogische Wissen genannt (Baumert und Kunter 2011; König und Seifert 2012). Zwar deuten einige Befunde darauf hin, dass das Ausmaß an pädagogischem Wissen Lernzuwächse positiv beeinflusst, weil Lehrpersonen mit einer pädagogischen Ausbildung eine effektivere Klassenführung zeigen, Schülerleistungen adäquater einschätzen, angemesseneres Schüler-Feedback geben und flexibler und anspruchsvoller unterrichten (vgl. Lipowsky 2006, S. 53). Studien, die das pädagogische Wissen direkt messen und in Bezug setzen zu Lernzuwächsen auf Schülerseite liegen aber meines Wissens bislang nicht vor.

Neben Studien, die das theoretisch wirksame Professionswissen direkt messen, gibt es eine Vielzahl an amerikanischen Studien, die den Einfluss solcher Lehrereigenschaften auf Lernzuwächse geprüft haben, die in den großen administrativen Datensätzen zur Verfügung stehen. Bei den analysierten Variablen wird implizit angenommen, dass sie das Ausmaß an Wissen, vermittelt über die Ausbildung, approximieren. Zu den Proxyvariablen gehören der Studienabschluss (Bachelor vs. Master), die Anzahl der besuchten fachwissenschaftlichen Kurse, das Prestige oder die Selektivität der besuchten Universität, die Art des absolvierten Lehrervorbereitungskurses⁹, und die Berufserfahrung (vgl. die Übersichtsartikel von Goe 2007; Lipowsky 2006; Wayne und Youngs 2003; aktuellere Studien sind Boyd et al. 2009; Clotfelter et al. 2007; Harris und Sass 2011). Die Befunde lassen sich wie folgt zusammenfassen: Für die Erfahrung – als Proxy für informell ‚on-the-job‘ erworbenes Professionswissen – finden sich überwiegend positive Effekte für die ersten 2-5 Jahre der Berufstätigkeit. Zusätzliche Erfahrung erbringt keine zusätzlichen Lernzuwächse. Für die restlichen Variablen zeigen sich mal positive und mal nicht signifikante Effekte; im Falle des Masterabschlusses sogar überraschend einige negative Effekte. Insgesamt sind die Effektstärken klein und variieren nach Schulstufe und untersuchtem Fach. Für andere Fächer als Mathematik und Lesen liegen praktisch keine Studien vor. Für Deutschland liegen gar keine Befunde vor. Insgesamt muss konstatiert werden, dass sich die indirekte und wenig theoriegeleitete Messung verfügbarer Proxyvariablen als wenig fruchtbar erwiesen hat. Entsprechend fassen Rockoff et al. (2011, S. 44) zusammen: „Like the

⁹ Jeder US-Bundesstaat hält unterschiedliche Ausbildungsgänge und Zertifikate bereit für Personen, die Lehrer werden wollen. Diese unterscheiden sich zumeist nach Fach und Schulart. „Fully certified teachers“ müssen in der Regel ein Mindestmaß an fachbezogenem Wissen nachweisen und ein Lehrerausbildungsprogramm besucht haben (vgl. Lipowsky 2006).

well-known story of a man looking for his keys under a street light – not because he dropped them nearby, but because that is where he can see – researchers’ lack of success in predicting new teacher performance may be driven by a narrow focus on commonly available data.” Die Tatsache, dass sich bei direkter Messung von Professionswissen stärkere und konsistentere Effekte finden als bei der indirekten Messung über Proxy-Variablen, deutet auf die Wichtigkeit einer präzisen Operationalisierung und eines theoriegeleiteten Vorgehens hin. Fast durchgehend zeigen sich stärkere und konsistentere Befunde in Mathematik. Nye et al. (2004, S. 247) vermuten, dass Mathematik vor allem in der Schule gelernt wird und daher unmittelbarer von der Lehrkraft beeinflusst wird, während Lesen zu einem großen Teil auch außerhalb der Schule erlernt wird.

1.2.2.2 Motivationale Orientierungen

Das Thema Motivation findet sich in vielen Arbeiten zum Lehrerberuf. Dahinter steckt die Annahme, der Lehrerberuf erfordere “immer wieder hohe Konzentration, Aufmerksamkeit, die Fähigkeit, mit Misserfolgen umzugehen, und die Bereitschaft, langfristig engagiert zu bleiben und sich immer wieder neuen Situationen auszusetzen und entsprechende Lerngelegenheiten zu nutzen” (Kunter 2011a, S. 259). Zur erfolgreichen Erfüllung dieser Anforderungen ist neben (kognitiven) Fähigkeiten auch eine entsprechende (motivationale) Bereitschaft notwendig. Die psychologische Motivationsforschung unterscheidet eine Vielzahl an motivationalen Konstrukten. Insgesamt werden diese als Bedingungen für die Aufnahme, Aufrechterhaltung, sowie Qualität von Handlungen angesehen (Kunter 2011a; Pintrich 2003). Häufig thematisierte motivationale Konstrukte innerhalb der Lehrerforschung sind Studien- bzw. Berufswahlmotive (Watt und Richardson 2007), Enthusiasmus bzw. intrinsische Motivation beim Unterrichten (Kunter et al. 2008) sowie Zielorientierungen (Butler 2007).¹⁰ Allerdings gibt es bislang fast keine Studien, die den Zusammenhang solcher Konstrukte mit den Lernzuwächsen von Schülern prüfen. Eine Ausnahme ist die Analyse von Kunter (2011a) auf Basis der COACTIV-Daten (vgl. auch Kunter et al. in Druck). Demnach zeigt sich ein positiver Effekt des Enthusiasmus zu Unterrichten (nicht jedoch des Fachenthusiasmus) auf Lernzuwächse deutscher Sekundarschüler. Ebenso zeigt sich eine positive Wirkung, wenn die Motivation der Schüler als

¹⁰ Zielorientierungen beschreiben dauerhafte Zielüberzeugungen, die in Leistungssituationen handlungsleitend werden. Zielorientierungstheorien unterscheiden grob zwischen Lernzielen, die darauf abzielen, eigene Kompetenzen zu erweitern, Leistungszielen, die darauf abzielen, anderen gegenüber hohe Kompetenzen zu demonstrieren und Arbeitsvermeidung, die darauf abzielt, den Arbeitsaufwand größtmöglich zu reduzieren (Köller und Schiefele 2010).

abhängige Variable modelliert wird. Die Ergebnisse fügen sich ein in Befunde aus korrelativen und experimentellen Studien, wonach Schüler von enthusiastischen bzw. intrinsisch motivierten Lehrkräften höhere Motivation und mehr Interesse aufweisen (Babad 2007; Frenzel et al. 2009; Kunter et al. 2011b; Kunter et al. 2008; Roth et al. 2007; Wild et al. 1992). Insgesamt sprechen die Befunde für die Bedeutsamkeit motivationaler Orientierungen, der Zusammenhang mit Lernzuwächsen bei Schülern ist allerdings nicht eindeutig geklärt.

1.2.2.3 Selbstregulationsfähigkeit

Belastung und Beanspruchung ist immer wieder ein Thema innerhalb der Forschung zum Lehrerberuf (Rothland 2013). Deshalb wird innerhalb der Kompetenzmodelle von Lehrern verstärkt auf die Bedeutsamkeit einer guten Selbstregulationsfähigkeit hingewiesen. Gemeint ist damit die Fähigkeit ein angemessenes berufliches Engagement an den Tag zu legen, sich aber *gleichzeitig* von beruflichen Belangen distanzieren zu können und Ressourcen zu schonen, um den Anforderungen über viele Jahre und ohne gesundheitliche Einschränkungen begegnen zu können (Klusmann 2011). Zahlen aus der Lehrgesundheitsforschung zeigen, dass nicht alle Lehrkräfte erfolgreich mit den beruflichen Belastungen umgehen können. Der Anteil der vorzeitigen Dienstunfähigkeit an den jährlichen Ruhestandseintritten liegt bei Lehrern in Deutschland seit Jahren zwischen 50 und 60 Prozent, wobei der gesundheitlich begründete Berufsausstieg durchschnittlich zehn Jahre vor der Regelaltersgrenze von 65 Jahren erfolgt (Weber et al. 2004). Als Ursache werden häufig psychische Beeinträchtigungen genannt (Hillert und Christ 2004). Basierend auf Arbeiten von Uwe Schaarschmidt und Kollegen (Schaarschmidt 2005; Schaarschmidt und Fischer 2003) können bei Lehrkräften anhand der Dimensionen „Widerstandsfähigkeit“ und „Engagement“ vier Typen unterschieden werden, die sich in ihrer Selbstregulationsfähigkeit unterscheiden (Gesundheitstyp, Schontyp, Risikotyp A, Risikotyp B, wobei sich der Gesundheitstyp durch eine hohe Ausprägung auf beiden Dimensionen auszeichnet). Studien, die den Zusammenhang zwischen Selbstregulationsfähigkeit und Lernzuwächsen bei Schülern nachweisen, liegen nicht vor. Hinsichtlich anderer schulischer Ergebnisvariablen zeigen sich jedoch Zusammenhänge, die insgesamt für die Relevanz dieses Konstruktes sprechen. So kann auf Basis der COACTIV-Studie nachgewiesen werden, dass sich Lehrkräfte entsprechend ihrem selbstregulativen Stil in ihrem Unterrichtsverhalten unterscheiden. Lehrkräfte die sich durch ein hohes Engagement *und* eine hohe Widerstandsfähigkeit auszeichnen (der Gesundheitstyp), führen aus Sicht ihrer Schüler einen kognitiv aktivierenden Unterricht durch, verfolgen ein angemesseneres Tempo und leisten mehr konstruktive

Unterstützung als die anderen Selbstregulationstypen. Vermittelt über das Unterrichtsgeschehen wirkt sich dieser Typ auch auf die Motivation der Schüler für das Unterrichtsfach Mathematik aus (Klusmann et al. 2008). Neben schülerbezogenen Ergebnisvariablen werden innerhalb der Forschung zur Selbstregulationsfähigkeit von Lehrern auch lehrerbezogene Ergebnisvariablen betrachtet, speziell das berufliche Wohlbefinden. Auch hier zeigen sich die Vorzüge des Gesundheitstyps. Er weist sowohl bei der emotionalen Erschöpfung als auch bei der Arbeitszufriedenheit günstigere Werte auf als die übrigen Typen (Klusmann 2011). Beachtenswert ist der Befund, dass auf Basis der Potsdamer Lehrerstudie nur 18 Prozent der Lehrer dem wünschenswerten Gesundheitstyp zugeordnet werden (Schaarschmidt 2005), in den Daten der COACTIV-Erhebung sind es zwischen 24 und 31 Prozent (Klusmann 2011). Die Mehrzahl der Lehrer ist demnach entweder wenig widerstandsfähig oder wenig engagiert, oder beides.

1.2.2.4 Werthaltungen und Überzeugungen

Das COACTIV-Modell geht davon aus, dass ergänzend zu den genannten Dimensionen bestimmte subjektive Werthaltungen und Überzeugungen handlungssteuernde Faktoren darstellen, die das Unterrichten und damit die Lernzuwächse der Schüler beeinflussen (vgl. auch Calderhead 1996; Pajares 1992). Die Ergebnisse zeigen, „dass Überzeugungen über die Natur des mathematischen Wissens und Überzeugungen über das Lehren und Lernen von Mathematik in charakteristischen Überzeugungssyndromen zusammenfallen“ (Kunter und Baumert 2011, S. 348). Lehrkräfte mit stärker ausgeprägter „transmissiver Überzeugung“ (der Lehrer sendet, der Schüler empfängt) gestalten ihren Unterricht weniger kognitiv aktivierend als Lehrer, die den „Prozess des Wissensaufbaus als einen Diskurs zwischen Lehrern und Schülern“ verstehen. Entsprechend erzeugen letztere vergleichsweise bessere Lernergebnisse bei ihren Schülern (Dubberke et al. 2008; Kunter et al. in Druck; Voss et al. 2011). In meinen Augen ist das Konzept der Überzeugungen und Werthaltungen unscharf. Ebenso gut könnte man sie als (mehr oder weniger) korrektes pädagogisches Wissen konzipieren, darüber, wie Unterricht zu gestalten ist, um den Lernerfolg von Schülern zu unterstützen. Weniger zuträgliche Überzeugungen entsprächen nach einer solchen Konzeption lediglich einem geringeren Ausmaß an pädagogischem Wissen, das sich in der Folge auf weniger zuträgliches Unterrichtshandeln auswirkt.

1.2.3 Fazit

Zusammenfassend lässt sich festhalten, dass sich insgesamt das Konzept eines multidimensionalen Kompetenzportfolios zu bestätigen scheint. Allerdings wurden im Hinblick auf die Frage, welche Kompetenzen tatsächlich *Lernzuwächse* bei Schülern erzeugen, mit Ausnahme des professionellen Wissens kaum wirklich belastbare Zusammenhänge gefunden. Die gefundenen Zusammenhänge hinsichtlich motivationaler Orientierungen, Selbstregulationsfähigkeit und Überzeugungen beziehen sich fast ausschließlich auf eine Datengrundlage (COACTIV). Weitere längsschnittliche Studien sind notwendig, um die gefundenen Zusammenhänge abzusichern.

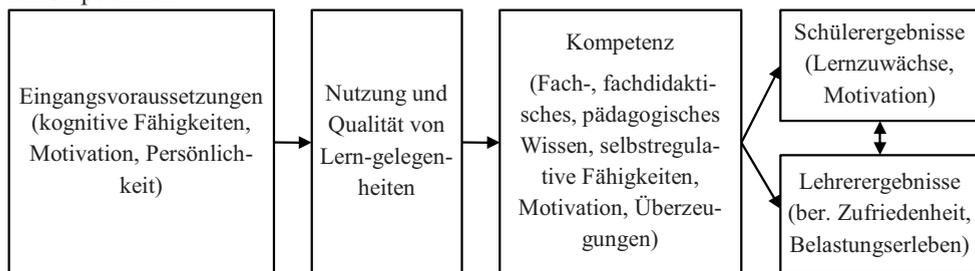
Die Bildungsökonomie verwendet alle Kraft darauf, den Einfluss des Lehrers auf das Produkt (den Lernzuwachs) möglichst genau zu bestimmen (vgl. Abschnitt 1.2.1), vernachlässigt aber ein theoriegeleitetes Vorgehen, das die wirksamen Kompetenzdimensionen ebenso wie deren Messung angeht. Diesem Desiderat begegnen die Pädagogische Psychologie und die Erziehungswissenschaft (vgl. Abschnitt 1.2.2), vernachlässigen aber bislang weitgehend die ökonometrisch belastbare Modellierung der Beiträge dieser Variablen zu den *Lernzuwächsen* der Schüler (mit der COACTIV-Studie als positive Ausnahme). Am vielversprechendsten ist es sicherlich, die beiden Forschungsstränge stärker zu vernetzen.

1.3 Die Genese professioneller Kompetenz und die Bedeutung von Eingangsvoraussetzungen

Nachdem zumindest in Ansätzen geklärt wurde, welche Kompetenzen Lernzuwächse erzeugen, schließt sich die Frage an, wie diese entstehen. Diese Frage betrifft den Kern aller Diskussionen um Lehrerqualität. „Von Erkenntnissen darüber, warum unterschiedliche Ausprägungen in der professionellen Kompetenz zu beobachten sind, erwartet man sich vor allem Hinweise darauf, wie es gelingen kann, Lehrkräfte besser auszubilden bzw. möglicherweise geeignetere Personen für den Lehrerberuf zu gewinnen.“ (Kunter et al. 2011a, S. 55-56). Das plausibelste Modell zur Genese von Lehrerkompetenz ist aus der Unterrichtsforschung entlehnt. Die gängige Denkfigur zur Erklärung von Schülerlernen ist das sogenannte Angebot-Nutzungs-Modell, das von Helmut Fend (1981) in die Pädagogik eingebracht wurde (vgl. Helmke 2012). Demnach stellt der Unterricht ein Lernangebot dar, dass von den Schülern in Abhängigkeit von ihren individuellen Voraussetzungen (Intelligenz, Motivation) unterschiedlich stark genutzt wird. Aus dem Zusammenwirken der Qualität und Intensität der unterrichtlichen Lerngelegenheiten und den

individuellen Voraussetzungen lässt sich der Lernerfolg erklären.¹¹ Die gleiche Denkfigur liegt aktuellen Modellen des Kompetenzerwerbs von Lehrkräften zugrunde (Blömeke et al. 2011; Kunter et al. 2011a; Terhart 2012b). Auch hier sollten folglich die individuellen Eingangsvoraussetzungen der Lehramtsanwärter das Ausmaß der Nutzung der Lerngelegenheiten beeinflussen. Die Quantität und Qualität der Lerngelegenheiten – universitäre Ausbildung, Referendariat, Weiterbildung ebenso wie informelle Lerngelegenheiten – erzeugen wiederum im Zusammenspiel mit der Intensität der Nutzung professionelle Lehrkompetenz. Dahinter steckt die Annahme, dass Kompetenzen beeinflussbar und veränderbar sind (Kunter et al. 2011a). Für die Wissensdimension ist dieser Umstand zweifelsfrei zutreffend. Inwieweit auch motivationale Orientierungen und die Fähigkeiten zur Selbstregulation erlernbar sind, oder inwieweit diese Kompetenzfacetten eher stabile Personenmerkmale sind, ist nicht abschließend geklärt.¹² In jedem Falle kommt es auf beides an, Eingangsvoraussetzungen ebenso wie die Qualität der Aus- bzw. Weiterbildung. Dem theoretischen Modell zufolge bestimmt dann die so erzeugte professionelle Kompetenz effektives Unterrichtshandeln und in der Folge Lernzuwächse bei Schülern (vgl. Abbildung 1).

Abbildung 1: Die Bedeutung der Eingangsvoraussetzung für die Genese von Lehrer- und Schülerkompetenzen



Anmerkung: Modell einer Wirkungskette in Anlehnung an COACTIV (Kunter et al. 2011a, S. 59)

¹¹ Die Idee wurde zuvor schon in sehr ähnlicher Form von den Soziologen Sørensen und Hallinan (1977) formuliert.

¹² Kunter (2011a) zeigt, dass sich im Verlaufe eines Jahres über zwei Messzeitpunkte beim Fachenthusiasmus eine Stabilität von $r=.72$, beim Unterrichtsenthusiasmus von $r=.61$ ergibt. Sie interpretiert dieses Ergebnis als Beleg für die Veränderbarkeit und Kontextabhängigkeit des Enthusiasmus. Genauso könnte man die Korrelationen als stark interpretieren und argumentieren, dass abgesehen von zufälligen Schwankungen aufgrund der Tagesform oder aktuellen Unterrichtsinhalte die Korrelationen für eine deutliche Stabilität sprechen. Lehrer, die im Vorjahr über einen hohen Fachenthusiasmus verfügen, tun dies mit großer Wahrscheinlichkeit auch im Folgejahr. Der Nachweis über die Veränderbarkeit und Modifizierbarkeit der Selbstregulationsfähigkeit steht meines Wissens noch aus.

Schülerergebnisse stellen das zentrale Erfolgskriterium dar, aber auch Lehrerergebnisse, wie etwa die berufliche Zufriedenheit oder das Belastungserleben, können als Erfolgskriterien betrachtet und als Folge des Kompetenzausmaßes verstanden werden. Nach dem Modell können Unterschiede im Kompetenzausmaß *erstens* die Folge von unterschiedlichen Eingangsvoraussetzungen sein und *zweitens* von unterschiedlich guter Lehrerausbildung (wenn beispielsweise die Ausbildung für das gymnasiale Lehramt systematisch mehr Kompetenzen vermittelt als die Ausbildung für nicht-gymnasiale Lehramtsstudiengänge). Welcher dieser beiden Faktoren die größere Rolle spielt, ist nicht leicht zu beantworten; die empirische Trennung der beiden konfundierten Einflussgrößen ist alles andere als trivial, weil

- die Selektion in die unterschiedlichen Lehramtsstudiengänge systematisch mit beobachteten und unbeobachteten Eingangsmerkmalen variiert und weil
- die beiden Einflussbereiche nach dem Modell interaktiv miteinander verbunden sind (Eingangsvoraussetzungen * Lerngelegenheiten = Kompetenzen).

Es gibt jedoch gute Gründe davon auszugehen, dass den individuellen Voraussetzungen eine bedeutsame Rolle bei der Entstehung von Lehrerkompetenzen zukommt. Lernen ist keine passive Routine, sondern gelingt nur über einen aktiven, von kognitiven ebenso wie von motivationalen Voraussetzungen beeinflussten Lernprozess. Dabei erleichtern größere Lern- und Leistungsfähigkeit die Einpassung neuen Wissens (z. B. Klauer und Leutner 2010). Zudem hängt das Ausmaß an investierter Energie für die Lehrerausbildung – und damit der Wissenszuwachs – auch damit zusammen, welche Motivation eine Person zur Erreichung eines Ziels hat (Kunter 2011b). Persönlichkeitsmerkmale oder motivationalen Orientierungen beeinflussen zudem, welche Lerngelegenheiten (z. B. Art der Lehramtsausbildung) sich ein Lehramtsanwärter sucht, oder wie regelmäßig und intensiv er oder sie an Fortbildungsmaßnahmen teilnimmt.

Inzwischen liegen eine ganze Reihe empirischer Belege vor, die einen Zusammenhang von kognitiven und motivationalen Eingangsmerkmalen und dem Ausmaß an erworbenen Kompetenzen von (angehenden) Lehrkräften belegen (Blömeke und Buchholtz 2011; Blömeke et al. 2012; Kleickmann und Anders 2011; Kleickmann et al. 2013; König et al. 2012; Kunina-Habenicht et al. 2013). Diese Zusammenhänge werden in Kapitel 2, Abschnitt 5.3 sowie Kapitel 3, Abschnitt 1 ausführlicher dargestellt. Geht man entsprechend davon aus, dass die Eingangsmerkmale entscheidend zur Kompetenzgenese von Lehrern beitragen, stellt sich die Frage, über welche Eingangsvoraussetzungen (zukünftige) Lehrer verfügen.

1.4 Fragestellungen und Befunde – Teil 1

Obwohl die Bedeutsamkeit der Eingangsmerkmale für die Genese der Lehrerkompetenz kaum zu bezweifeln ist, ist der empirische Kenntnisstand zu den Eingangsmerkmalen von Lehramtsstudierenden im Vergleich zu anderen Studierenden sehr begrenzt. Zwar gibt es eine Vielzahl an Studien zu diesem Thema (für eine Übersicht vgl. Rothland 2011b,a), allerdings basieren diese fast ausschließlich auf kleinen Gelegenheitsstichproben ungeklärter Repräsentativität und beschränken sich in der Regel auf Lehramtsstudierende. Durch das Fehlen von Vergleichsgruppen ist es kaum möglich, die Eigenständigkeit bestimmter Merkmale bei der Wahl des Lehramtsstudiums zu identifizieren. Vor diesem Hintergrund untersucht der Beitrag in *Kapitel 2* die leistungsbezogenen, motivationalen und soziodemographischen Eingangsmerkmale von Lehramtsstudierenden unterschiedlicher Schulformen im Vergleich zu anderen Studierenden. Als Datengrundlage dienen die HIS-Studienberechtigtenpanels der Jahre 1999 und 2002. Die Ergebnisse zeigen, dass sich Gymnasiallehramtsstudierende im Hinblick auf Abiturleistungen und fachliche Studienwahlmotivation nicht von anderen Universitätsstudierenden unterscheiden. Allerdings sind den Gymnasiallehramtsstudierenden wissenschaftliche Interessen, Karrieremotive und das Ziel, überdurchschnittliche Kompetenzen erwerben zu wollen, weniger wichtig als anderen Universitätsstudierenden. Ausgeprägter sind hingegen soziale Interessen, der Wunsch nach beruflicher Sicherheit und private Vereinbarkeitsziele. Die genannten Unterschiede gelten ebenfalls für Studienanfänger des Real-, Grund-/Haupt- sowie Sonderschullehramtes. Darüber hinaus haben die nicht-gymnasialen Lehramtsstudienanfänger schlechtere Abiturleistungen und geringere fachliche sowie wissenschaftliche Studienwahlmotivationen als ihre Kommilitonen im Gymnasiallehramt. Des Weiteren zeigt sich, dass insbesondere nicht-gymnasiale Lehramtsstudiengänge vermehrt weibliche Studienberechtigte und solche aus bildungsfernen sozialen Schichten anziehen. Problematisch erscheint der Befund der niedrigeren Abiturleistungen und fachlichen Motivationen von nicht-gymnasialen Lehramtsstudierenden. Die vorgestellten Analysen legen nahe, die Rekrutierungspraxis in die Lehrerausbildung zu überprüfen. Die Studie macht deutlich, dass bei der Debatte um Eignungsprüfungen (Rothland und Terhart 2011) insbesondere auf die nicht-gymnasialen Lehramter geachtet werden sollte. Die ungleiche Selbstselektion in die verschiedenen Lehramtsstudiengänge ist in ihren langfristigen Folgen möglicherweise auch für die Genese von Bildungsungleichheiten bedeutsam. So deutet sich eine zusätzliche Benachteiligung der Kinder auf niedrigeren Schulstufen an, nicht nur weil die Ausbildung für diese Lehramter weniger

Kompetenzen vermittelt (Baumert et al. 2010), sondern auch weil die Studierenden die vorhandenen Lerngelegenheiten aufgrund ihrer unvorteilhaften Eingangsmerkmale womöglich nicht optimal nutzen. Dieser Punkt wird in Abschnitt 1.5 aufgegriffen.

Während in Kapitel 2 die Eingangsmerkmale von Personen untersucht werden, die um das Jahr 2000 ein Studium begonnen haben, werden in **Kapitel 3** die Eingangsmerkmale verschiedener Kohorten im Zeitverlauf untersucht. Die Arbeit knüpft an Studien aus den USA und Großbritannien an, die darauf hindeuten, dass sich die leistungsbezogenen Eingangsmerkmale von Lehrern im Vergleich mit anderen Akademikern im Laufe der letzten Dekaden verschlechtert haben (Bacolod 2007; Corcoran et al. 2004; Nickell und Quintini 2002). Für Deutschland liegen zu dieser Frage bislang keinerlei Erkenntnisse vor. Auf Basis der Konstanzer Studierendensurveys von 1982 bis 2009 wird nachgezeichnet, inwiefern sich die Abiturleistungen, ebenso wie die fachlichen und sicherheitsbezogenen Studienwahlmotive im Laufe der letzten 30 Jahre gewandelt haben. Wiederum wird den Lehramtsstudierenden eine Vergleichsgruppe von anderen Studierenden gegenübergestellt, um die charakteristischen Merkmale von Lehramtsanwärtern aufzuzeigen. Darüber hinaus beschreibt die Studie auf Basis der Mikrozensen von 1980 bis 2009 die Arbeitsmarktaussichten (Einkommen und Arbeitslosigkeitsrisiken) von Lehramtsanwärtern im Vergleich zu anderen akademischen Berufen. Auch dazu liegen bislang keine Befunde vor. Es wird herausgearbeitet, wie ein Wandel in den Arbeitsmarktaussichten Personen mit unterschiedlichen Leistungsvoraussetzungen oder Motivlagen anziehen kann. So wird vermutet, dass sich leistungsstärkere Studienberechtigte bei fallenden Lehrereinkommen eher für alternative berufliche Wege entscheiden, wo sie ihr Leistungspotenzial in entsprechend hohe Einkommen umsetzen können. Weiterhin wird angenommen, dass der Lehrerberuf bei fallendem Arbeitslosigkeitsrisiko auch für solche Personen attraktiv wird, die sich durch eine besonders hohe berufliche Sicherheitsmotivation auszeichnen. Die Ergebnisse lassen sich wie folgt zusammenfassen: Die Lehrereinkommen sind in Deutschland, im Gegensatz zu den USA oder Großbritannien, nicht gefallen – weder absolut noch relativ zu den Einkommen anderer Akademiker. Deutsche Lehrer erzielen vergleichbare Einkommen wie andere Akademiker; Frauen im Lehramt verdienen sogar mehr als andere Akademikerinnen. Entsprechend ist auch kein Abfall in den Abiturleistungen von Lehramtsanwärtern zu verzeichnen. Die Gesamtgruppe der Lehramtsanwärter ist über die Zeit hinweg zwischen 0.17 und 0.27 Standardabweichungen schlechter als andere Studierende an westdeutschen Universitäten (Eine Differenzierung nach Art des Lehramtes ist mit den verfügbaren Daten leider nicht möglich). Hinsichtlich der Arbeitslosigkeitsrisiken zeigen sich deutliche Schwankungen auf dem Teilarbeitsmarkt Schule.

Für Berufsanfänger im Lehramt steigt das Arbeitslosigkeitsrisiko bis Mitte der 1980er stark an, bevor es ab Mitte der 1990er deutlich unter das Niveau anderer Akademiker zurückfällt. Die Studienwahlmotive korrespondieren mit diesen Schwankungen. Die Wahrscheinlichkeit das Lehramt zu wählen, ist für Personen mit einer hohen beruflichen Sicherheitsmotivation sehr gering in Zeiten eines hohen Arbeitslosigkeitsrisikos. Sie steigt aber deutlich ab Mitte der 1990er, als die Arbeitsmarktaussichten deutlich vielversprechender sind. Im Gegensatz dazu war der Anteil an Personen mit einer hohen fachlichen Motivation bis Mitte der 1990er im Lehramt höher als in anderen Studienfächern, vermutlich weil Personen mit niedriger Fachmotivation zu dieser Zeit in andere Studienfächer ausgewichen sind. Das Kapitel zeigt am Beispiel des Lehramtes, wie Arbeitsmarktaussichten die Studienfachwahl beeinflussen können. Dadurch ergibt sich für den Lehrerberuf je nach Arbeitsmarktlage ein unterschiedlicher Pool an Bewerbern, der hinsichtlich leistungsbezogener und motivationaler Eingangsmerkmale als mehr oder weniger geeignet charakterisiert werden kann. Dieser Aspekt sollte bei der zukünftigen Abschätzung und Planung des zyklisch schwankenden Lehrbedarfs berücksichtigt werden.

1.5 Lehrereffekte und Bildungsungleichheiten – eine Systematik

Der zweite Teil der kumulativen Dissertation widmet sich dem Einfluss von Lehrern bei der Entstehung ungleicher Bildungschancen. Fragen nach der Entstehung und dem Wandel von *sozialen, geschlechtlichen* und *ethnischen* Bildungsungleichheiten stehen im Zentrum soziologischer Bildungsforschung. Erstaunlicherweise wird die Rolle von Lehrkräften dabei kaum thematisiert. Betrachtet man die soziologische Forschung, *kommt es auf den Lehrer offenbar nicht an*. Eine Literaturrecherche¹³ nach dem Begriff „Teacher“ in Titel, Abstract oder Schlüsselwörtern von Aufsätzen, die zwischen 1956 und 2013 in den zentralen Zeitschriften *American Journal of Sociology*, *American Sociological Review* und *European Sociological Review* publiziert wurden, ergab, dass nur 10 von 6036 Aufsätzen in den genannten Textteilen den Begriff „Teacher“ erwähnen. Die gleichen Suchkriterien lieferten unter den deutschsprachigen Journalen *Kölner Zeitschrift für Soziologie und Sozialpsychologie* und *Zeitschrift für Soziologie* 20

¹³ Die Recherche erfolgte über den Social Science Citation Index (<http://apps.webofknowledge.com>). Buchbesprechungen wurden ausgeschlossen. Aufsätze, die sich auf Hochschullehre oder „law teachers“ bezogen, ebenfalls.

Treffer unter insgesamt 2719 Aufsätzen.¹⁴ Entsprechend spärlich und unstrukturiert ist der empirische Forschungsstand zu Lehrereffekten und Bildungsungleichheiten. Dabei gibt es eine ganze Reihe möglicher Wirkungsweisen, die im Folgenden benannt werden. Die Wirkungsweisen sollen dabei so systematisiert werden, dass sie anschlussfähig an die soziologische Ungleichheitsforschung sind. Durch dieses Vorgehen ergibt sich einerseits eine Struktur, um die im zweiten Teil der Arbeit versammelten Arbeiten einzuordnen. Andererseits ergibt sich ein logisches Ordnungssystem, welches auf einige Forschungsdesiderata hindeutet. Dadurch wird in Ansätzen bereits ein mögliches zukünftiges Forschungsprogramm sichtbar.

Die Ursachen für die Entstehung ungleicher Bildungschancen werden innerhalb der Bildungssoziologie in Anlehnung an die klassische Arbeit von Boudon (1974) in primäre und sekundäre Effekte unterteilt. Wie ich zeigen werde, eignet sich diese Unterscheidung auch für die Strukturierung von Lehrereffekten. Primäre Effekte beschreiben das Phänomen, dass Schüler aus höheren Sozialschichten¹⁵ bessere Schulleistungen aufweisen als Schüler aus niedrigeren Sozialschichten, gemessen etwa über Kompetenztests oder Schulnoten. Viele Faktoren können dazu führen, dass es zu primären Effekten kommt, etwa unterschiedliche Anregungs- und Unterstützungsbedingungen im Elternhaus oder Intelligenzunterschiede. Sekundäre Effekte beschreiben das Phänomen, das Personen aus unterschiedlichen Sozialschichten mit *gleichen* Schulleistungen unterschiedlich ambitionierte Bildungsentscheidungen treffen (z. B. Entscheidungen über die weiterführende Schulwahl oder über die Aufnahme eines Studiums). Als Erklärung für dieses Phänomen werden herkunftsabhängige Erfolgserwartungen und Kosten-Nutzen-Abwägungen angeführt (Breen und Goldthorpe 1997; Erikson und Jonsson 1996; Esser 1999, S. 265ff.). Beide Effekte führen im Zusammenspiel dazu, dass Personen unterschiedlicher Herkunft unterschiedliche Bildungsabschlüsse erreichen. Primäre und sekundäre Effekte beziehen sich auf die Entstehung sozialer Bildungsungleichheiten, allerdings wird das Konzept in jüngerer Zeit ebenso zur Erklärung geschlechtlicher (Becker und Müller 2011) und ethnischer (Dollmann 2010) Ungleichheiten herangezogen. Wenn ich im

¹⁴ Über die Auswahl der Suchbegriffe und der Zeitschriften mag man streiten; so finden sich etwa in der Zeitschrift *Sociology of Education* deutlich mehr Einträge. Der generelle Befund, dass Lehrer von der soziologischen Bildungsforschung bisher wenig Beachtung finden, bleibt.

¹⁵ Begriffe wie Sozialschicht, Klasse oder sozioökonomischer Status werden in dieser theoretischen Diskussion weitgehend synonym verwendet.

Folgenden von primären und sekundären Effekten spreche, meine ich damit soziale, geschlechtliche und ethnische Disparitäten gleichermaßen. Übergreifend spreche ich im Folgenden von primären und sekundären *soziodemographischen* Disparitäten.¹⁶

Das Konzept der primären und sekundären Effekte eignet sich deshalb als Ordnungsrahmen, weil Lehrkräfte das Ausmaß beider Effekte beeinflussen können. Ich stelle im Folgenden dar, wie Lehrkräfte primäre und sekundäre Disparitäten beeinflussen könnten. Anschließend diskutiere ich die Plausibilität der Wirkungsweisen vor dem Hintergrund empirischer Befunde und daraus resultierende Konsequenzen für Ansatzpunkte zur Reduktion der Disparitäten.

1.5.1 Lehrkräfte und primäre Disparitäten

Lehrer können das Ausmaß primärer Effekte – soziodemographische Unterschiede in schulischen Leistungen also – beeinflussen. Das ist nicht weiter verwunderlich, sind sie es doch, die für die Vermittlung von Kompetenzen und die Benotung derselben hauptverantwortlich sind. Folgende Wirkungsweisen sind möglich:

- a. *Lehrkräfte können sich darin unterscheiden, wie erfolgreich sie benachteiligten Schülergruppen schulische Fähigkeiten vermitteln.* Damit würde das Ausmaß an Leistungsdisparitäten zwischen bestimmten Schülergruppen (z. B. Schulkinder aus niedrigen versus aus hohen Sozialschichten) in Klasse *a* größer oder kleiner ausfallen als in Klasse *b*. Gäbe es mehr Lehrkräfte, die benachteiligte Schülergruppen erfolgreicher unterrichteten, würden sich die primären Disparitäten insgesamt reduzieren. Solche primären Disparitäten in schulischen Kompetenzen könnten einerseits durch selektive *Lehrererwartungen* zustande kommen. Gemeint sind damit Erwartungen hinsichtlich der Leistungspotenziale bestimmter Schülergruppen, die zu (unbewussten) Verhaltensänderungen der Lehrer führen und sich dadurch im Stile von sich selbst erfüllenden Prophezeiungen auf die tatsächlichen Leistungen der Schüler auswirken (Rosenthal und Jacobson 1968). Die empirische Evidenz für die Existenz solcher Erwartungseffekte ist gemischt (Jussim et al. 2009); angesichts der schwachen Effektstärken ist eher nicht davon auszugehen, dass Lehrererwartungen soziodemographische Leistungsunterschiede nachhaltig zementieren oder auflösen können. Andererseits könnten primäre Disparitäten zustande kommen, weil

¹⁶ Das Ausmaß an Ungleichheiten unterscheidet sich je nach Dimension, zudem sind die einzelnen Dimensionen konfundiert. So lassen sich etwa ethnische Ungleichheiten in erheblichem Maße auf soziale Ungleichheiten zurückführen. Dies soll hier nicht weiter thematisiert werden.

Lehrkräfte über ein unterschiedliches Ausmaß an *pädagogischem Wissen* zum Umgang mit schwer zu unterrichtenden Schülern oder heterogenen Schulklassen verfügen. Aktuelle Kompetenzmodelle enthalten die Dimension „pädagogisches Wissen“ (vgl. Abschnitt 1.3.1), wovon eine Sub-Dimension wiederum das Wissen im „Umgang mit Heterogenität“ abbildet (König und Blömeke 2009). Allerdings ist bislang nicht untersucht worden, inwiefern diese Kompetenz mit soziodemographischen Leistungsdisparitäten von Schülern in Zusammenhang steht. Es kann vermutet werden, dass ein kompetenter Umgang mit Schülerheterogenität systematische Leistungsunterschiede zwischen Schülern verringert (vgl. auch Hofer 2009).

Sollte das Wissen zum Umgang mit heterogenen Klassen oder sozio-strukturell benachteiligten Schülern tatsächlich primäre Disparitäten reduzieren, ergäben sich unmittelbare Implikationen für die Lehrerbildung. Die pädagogische Ausbildung müsste dann in dieser Hinsicht erweitert werden, etwa durch die Entwicklung von speziellen Unterrichtsmethoden zur besseren Unterrichtung von benachteiligten Schülergruppen.

- b. *Lehrkräfte können sich in der Benotung der schulischen Fähigkeiten von benachteiligten Gruppen unterscheiden.*¹⁷ Beurteilungsdifferenzen können durch *Diskriminierung* zustande kommen. Damit ist gemeint, dass Merkmale, die nicht mit den Leistungen der Schüler in Verbindung stehen, die Beurteilung derselben beeinflussen (Arrow 1973, S. 3). Vor allem im Zusammenhang mit ethnischen Bildungsungleichheiten wird häufig von Diskriminierung gesprochen (vgl. Kristen 2006), dieser Mechanismus kann aber gleichsam andere soziodemographische Dimensionen betreffen. Warum aber sollten Lehrkräfte Schüler bestimmter Herkunftsgruppen diskriminieren? Ein Grund könnte sein, dass bestimmte Lehrer einen ‚taste for discrimination‘ hegen (Becker 1971), wobei unklar bleibt, woher solche Präferenzen rühren. Sie würden klar dem „normativen Selbstverständnis des Lehrerberufs“ widersprechen (Wiese 1982, S. 52). Als weitere Möglichkeit lassen sich Informationsdefizite anführen. Im Falle unzureichender Informationen werden sichtbare askriptive Merkmale der Schüler als Beurteilungsgrundlage herangezogen. Dabei wird von Annahmen über die im Durchschnitt erbrachten Leistungen dieser Gruppe auf die Leistung des Individuums geschlossen. Dieser Mechanismus wird als ‚statistical discrimination‘ bezeichnet

¹⁷ Der Literatur zu primären und sekundären Effekten folgend, konzipiere ich das Ausmaß an ungleichen Schulnoten als primäre Effekte, auch wenn in den Noten zu einem Teil antizipatorische sekundäre Entscheidungskomponenten enthalten sind (Erikson und Rudolphi 2010; Jackson et al. 2007; Neugebauer 2010; Schindler und Reimer 2010).

(Arrow 1973; Phelps 1972). Bei der Benotung im Schulkontext ist allerdings kaum von gravierenden Informationsdefiziten auszugehen. Durch den täglichen Kontakt und die regelmäßige Leistungsüberprüfung hat wohl kaum jemand mehr Informationen über einen individuellen Schüler, als sein Lehrer.¹⁸ Eine statistische Diskriminierung in der Notenvergabe ist deshalb nicht zu erwarten. Entsprechend gibt es kaum überzeugende Nachweise für eine diskriminierende Benotungspraxis (Becker und Beck 2012; Farkas 2003; Kristen 2006). Das heißt jedoch nicht, dass soziodemographische Merkmale unter Kontrolle der schulischen Leistungen keinen Einfluss haben müssen auf die Notenvergabe. Bei der Benotung werden neben schulischen Leistungen auch *die mündliche Mitarbeit, Fleiß, oder das Sozialverhalten* berücksichtigt, die mit den soziodemographischen Merkmalen der Schüler kovariieren können. Zudem ist es möglich, dass gegen Ende der Grundschulzeit Eltern aus höheren Sozialschichten versuchen, die Notengebung durch Interventionen in ihrem Sinne zu beeinflussen (Neugebauer 2010).

Wenn die Berücksichtigung von Fleiß oder Sozialverhalten zur Benachteiligung bestimmter Schülergruppen führt, ergeben sich zwei logische Ansatzpunkte zur Reduktion der Notendisparitäten. Zum einen können die Richtlinien zur Benotungspraxis geändert werden, so dass entsprechende Merkmale bei der Benotung ignoriert werden (allerdings dürften sich kaum Befürworter für eine solche Änderung finden). Zum anderen können (theoretisch) die Verhaltensweisen von benachteiligten Schülern geändert werden. Wie das (praktisch) gelingen kann, vermag ich nicht zu sagen.

- c. *c.* ist ein Spezialfall von *a.* und *b.* So wird häufig vermutet, dass *Kompetenzen bzw. Benotungen von benachteiligten Schülergruppen beeinflusst werden von der Ähnlichkeit zwischen soziodemographischen Lehrer- und Schülermerkmalen.* Das Ausmaß an primären Disparitäten hänge dann von der Passung („congruence“, nach Alexander et al. 1987) von Lehrern und Schülern ab. Schüler werden in dem Maße benachteiligt, in dem sie mit „unpassenden“ Lehrkräften konfrontiert sind. Da es innerhalb der Lehrerpopulation im Vergleich mit der Schülerpopulation mehr Frauen, weniger Personen mit Migrationshintergrund und (vermutlich) mehr Personen aus höheren Sozialschichten gibt, sind bestimmte

¹⁸ Aus diesem Grund läuft die experimentelle Studie von Sprietsma (in Druck) an der schulischen Realität vorbei. Die Autorin legt Lehrkräften Essays vor und weist einen kleinen negativen Effekt von türkischen Namen im Vergleich zu deutschen Namen bei der Benotung nach. Kein Lehrer verfügt bei tatsächlichen Beurteilungen ausschließlich über den Namen eines Kindes und ein einzelnes Essay.

Schülergruppen häufiger mit „unähnlichen“ Lehrkräften konfrontiert – und damit möglicherweise benachteiligt. Dahinter steckt die Annahme, dass Lehrkräfte aufgrund ihrer eigenen (geschlechtlichen, ethnischen oder sozialen) Herkunft (quasi von Natur aus) mit den notwendigen Fähigkeiten oder Neigungen ausgestattet wurden, die entsprechende Schülergruppe erfolgreich zu unterrichten. Damit kämen die gleichen theoretischen Mechanismen zum Tragen, die in *a* und *b* diskutiert wurden. Zusätzliche Erklärungsmechanismen setzen an den Schülern an. So wird vermutet, dass Schulkinder sich besser von „ähnlichen“ Lehrkräften unterrichten lassen, etwa weil sie beim Lernen auf „ähnliche“ Rollenvorbilder angewiesen sind, oder weil sie sich in Anwesenheit von „ähnlichen“ Lehrkräften unterrichtskonformer verhalten (vgl. Neugebauer et al. 2011, S. 671-673). Empirisch finden sich für die USA einige Belege für die These hinsichtlich der ethnischen sowie der sozialen Passung (Alexander et al. 1987; Dee 2004; Ehrenberg et al. 1995; McGrady und Reynolds 2013). Für Deutschland lässt sich an dieser Stelle ein Forschungsdesiderat ausmachen, da entsprechende Untersuchungen bislang ausstehen. Hinsichtlich der Passung von Lehrergeschlecht und Schülergeschlecht finden sich in aktuellen Studien aus vielen Ländern hingegen klare Belege, die der Passungsthese widersprechen (Ammermüller und Dolton 2006; Cho 2012; Coenen und Van Klaveren 2013; Sokal et al. 2007; Winters et al. 2013; für einen gegenteiligen Befund siehe Dee 2007). Zur Situation in Deutschland habe ich drei Studien durchgeführt, wovon zwei Teil dieser Dissertation sind. Die Ergebnisse werden in Abschnitt 1.7 zusammengefasst.

Sollten benachteiligte Schülergruppen tatsächlich bessere Lernergebnisse erzielen bei der Unterrichtung durch soziodemographisch ähnliche Lehrkräfte, hätte das Implikationen für die Rekrutierung zukünftiger Lehrer. Wären die Schulleistungen von Kindern mit Migrationshintergrund besser, wenn sie von Lehrkräften mit dem gleichen Migrationshintergrund unterrichtet würden? Könnten soziale Bildungsungleichheiten verringert werden, wenn Kinder aus der Arbeiterklasse verstärkt von Lehrkräften aus der Arbeiterklasse unterrichtet würden? Brauchen wir, um geschlechtliche Bildungsungleichheiten zu reduzieren, mehr männliche Lehrkräfte? Diese Fragen sind bislang nicht abschließend geklärt.

- d. Bekanntlich kovariert die Verteilung von Schülern auf weiterführende Schulen mit der Sozialschichtzugehörigkeit, dem Migrationshintergrund und dem Geschlecht der Schüler. Wenn es mehr oder weniger kompetente Lehrkräfte gibt, und wenn sich diese systematisch

selektieren, so dass *die weniger kompetenten Lehrkräfte überzufällig häufig die benachteiligten Schüler unterrichten*, beeinflusst das das Ausmaß an primären Bildungsungleichheiten zwischen den Schulformen. Mathematiklehrkräfte an Gymnasien verfügen über ein höheres Fachwissen und fachdidaktisches Wissen als Mathematiklehrkräfte auf anderen Schulformen (Baumert et al. 2010). Der Wissensumfang hat, wie in Abschnitt 1.2.2.1 berichtet, unmittelbare Auswirkungen auf den Lernzuwachs der Schüler. Zudem zeigt sich, dass besonders Schüler auf niedrigeren Sekundarschulformen vom Wissen der Lehrkräfte profitieren (Baumert et al. 2010, S. 165). Damit trägt die Verteilung von kompetenten Lehrkräften dazu bei, dass Schulformen „differenzielle Lern- und Entwicklungsmilieus“ (Baumert et al. 2006) darstellen, die bereits bestehende Ungleichheiten der Schüler verstärken. Nochmals verstärkt werden könnte ein solcher Scheren-Effekt, wenn die problematischere Klassenkomposition auf den niedrigeren Sekundarschulformen dazu führt, dass die ohnehin weniger kompetenten Lehrkräfte aufgrund einer höheren Stressbelastung häufiger krankheitsbedingt fehlen und Unterricht ausfällt.

Im Sinne einer Ungleichheitsreduktion müsste es gelingen, die kompetentesten Lehrkräfte auf den Schulformen unterrichten zu lassen, auf denen die benachteiligten Schüler überrepräsentiert sind; insbesondere trifft das auf Hauptschulen zu. Damit das gelingt, erscheint es erstens sinnvoll, im Rahmen der nicht-gymnasialen Lehrerausbildung mehr Wissen zu vermitteln (Baumert et al. 2010). Da das Ausmaß an Wissenserwerb auch von den Eingangsmerkmalen abhängt, wäre es zudem zielführend, mehr geeignete Kandidaten für die Ausbildung auf den nicht-gymnasialen Lehramtsstudiengängen zu rekrutieren (vgl. Kapitel 2).

1.5.2 Lehrkräfte und sekundäre Disparitäten

Sekundäre Effekte beruhen auf Unterschieden im Entscheidungsverhalten von Schülern oder deren Familien. Selbst bei gleichen Schulleistungen variieren Bildungsentscheidungen nach sozialer Herkunft, Geschlecht oder Migrationshintergrund. In Deutschland treffen Familien Bildungsentscheidungen nicht alleine. Lehrer und Schulleiter beeinflussen den schulischen Werdegang durch Beratungsgespräche und Empfehlungen maßgeblich mit. Das prominenteste Beispiel ist die Vergabe der Schullaufbahneempfehlung am Ende der Grundschule (z. B. Bos et al. 2004b; Wiese 1982). Lehrer beeinflussen aber auch andere schulische Entscheidungen, etwa zum Einschulungszeitpunkt (Kratzmann und Schneider 2009), zur Klassenwiederholung

(Krohne et al. 2004) oder zur Zurückstufung in eine niedrigere Schulform. Alle genannten Entscheidungen sind Maßnahmen, möglichst homogene Lerngruppen herzustellen (Krohne et al. 2004). Über zwei Wirkungsweisen können Lehrer dabei das Ausmaß an sekundären Disparitäten beeinflussen.

- a. *Über das Ausmaß, das Lehrkräfte leistungsfremden, soziodemographischen Merkmalen bei ihrer Empfehlungsvergabe einräumen.* Während der Forschungsstand zu Klassenwiederholungen oder dem Einschulungszeitpunkt dürftig ist, gibt es eine Vielzahl an Studien, die sich mit der Grundschulempfehlung befassen. Dabei zeigt sich konsistent, dass die soziale Herkunft auch bei gleicher Leistung die Bildungsempfehlungen der Lehrkräfte beeinflusst (Bos et al. 2004b; Ditton und Krüskens 2006; Lehmann et al. 1997; Schneider 2011). Im Hinblick auf das Geschlecht finden sich unter Kontrolle der Leseleistungen geringe signifikante Effekte zu Ungunsten der Jungen (Aktionsrat Bildung 2009), nicht jedoch bei zusätzlicher Kontrolle der Schulnoten oder des Lern- und Arbeitsverhaltens (vgl. Kapitel 5). Hinsichtlich des Migrationshintergrundes berichten Arnold et al. (2007), dass Kinder mit Migrationshintergrund unter Kontrolle der Leseleistung etwas geringere Chancen für eine Gymnasialempfehlung haben, während Lehmann et al. (1997) bei vergleichbaren Leistungen von höheren Chancen berichten. In den meisten Untersuchungen lassen sich bei vergleichbaren Leistungen jedoch keine voneinander abweichenden Chancen feststellen (Kristen 2006; Schneider 2011; Tiedemann und Billmann-Mahecha 2007; Wagner et al. 2010). Wodurch lässt sich der verbleibende Effekt der sozialen Herkunft erklären? Zum einen können die bereits bei der Notenvergabe angesprochenen Diskriminierungsmechanismen greifen (vgl. Abschnitt 1.5.1 b). Aus den gleichen Gründen wie oben dargelegt, sind jedoch sowohl ‚taste for discrimination‘ als auch ‚statistical discrimination‘ unwahrscheinlich. Die verbleibenden Herkunftseffekte sind vielmehr darauf zurückzuführen, dass Lehrkräfte weitere, mit der sozialen Herkunft korrelierende Faktoren bei der Empfehlungsvergabe berücksichtigen, wie das antizipierte Unterstützungsverhalten im Elternhaus bei Schulproblemen, oder das Lern- und Arbeitsverhalten (vgl. Ditton und Krüskens 2006; Neugebauer 2010; Schneider 2011). Während es sich bei der Notenvergabe um ein *retrospektives* Urteil bezüglich bereits gezeigter Leistungen handelt, soll bei der Empfehlungsvergabe *prospektiv* zukünftiges Lern- und Erfolgspotenzial berücksichtigt werden. Dadurch ergibt sich ein relativ großer pädagogischer Entscheidungsspielraum, in dem soziale Herkunftseffekte wirksam werden können. Eine in diesem Zusammenhang aufschlussreiche Untersuchung hat Thorsten Schneider (2011) vorgelegt, u. a. weil er das Lern-

und Arbeitsverhalten ebenso wie das durch die Lehrkräfte wahrgenommene Unterstützungsverhalten der Eltern direkt misst. Auf Basis der BiKS-Daten zeigt er, dass beide Konstrukte bei der Empfehlungsvergabe bedeutsam sind. Der Zusammenhang von sozioökonomischem Elternstatus und Empfehlungsvergabe lässt sich durch die Berücksichtigung dieser Variablen allerdings nicht vollständig erklären. Vielmehr zeigt sich, dass Eltern offenbar ihren schichtspezifisch variierenden Schulwunsch durchsetzen können, so dass er sich in der Empfehlungsvergabe widerspiegelt. Je höher der soziale Status der Eltern, desto eher schlägt sich die elterliche Bildungsaspiration in der Empfehlung nieder. Inwiefern Lehrer ihre Empfehlung auf Druck der Eltern hin anpassen, ist jedoch nicht abschließend geklärt.

Wie könnte die Beeinflussung der Lehrkräfte durch die Eltern reduziert werden? Schneider (2011, S. 391) schlägt vor, bundesweit standardisierte Tests in der zweiten Hälfte der Grundschulzeit einzuführen, die bei der Empfehlungsvergabe ergänzend berücksichtigt werden sollen. Lehrkräfte hätten damit ein Instrument, die eigene Einschätzung zu überprüfen, zudem könnte ein Testergebnis als Argumentationshilfe gegenüber Eltern mit überhöhten Bildungsaspirationen dienen. Befunde aus den Niederlanden, in denen ein solches System implementiert ist, stützen seine Argumentation. Inwiefern sich durch ein solches System der psychische Druck auf die Kinder von solchen Eltern erhöhen würde, müsste allerdings vorab in Pilotprojekten getestet werden.

- b. *Über das Gewicht, das der Lehrerempfehlung bei der Aushandlung der tatsächlich getroffenen Bildungsentscheidung zukommt.* Im Folgenden beziehe ich mich wiederum auf die Grundschulempfehlung, da das Gewicht der Lehrerempfehlung bei anderen Entscheidungen (Klassenwiederholungen, Einschulungszeitpunkt, Querversetzungen) bislang kaum untersucht wurde. Obwohl Lehrkräfte sich bei der Grundschulempfehlung offenbar von der sozioökonomischen Situation im Elternhaus beeinflussen lassen, spricht vieles dafür, dass Herkunftseffekte die Lehrerempfehlung weniger stark beeinflussen, als den elterlichen Bildungswunsch (Ditton et al. 2005). Faktoren, wie das Motiv des familiären Statuserhalts oder schichtspezifische Kosten-Nutzen-Erwägungen, sollten bei der Lehrerempfehlung keine Rolle spielen; vielmehr sollte diese in hohem Maße an leistungsrelevanten Kriterien orientiert sein. Entsprechend ist davon auszugehen, dass Herkunftseffekte weniger stark auftreten, wenn der Lehrerempfehlung bei der Aushandlung der tatsächlichen Bildungsentscheidung ein stärkeres Gewicht zukommt als dem elterlichen Bildungswunsch. In dem Fall

wäre es für statushöhere Familien nicht mehr ohne weiteres möglich, hohe Bildungswünsche auch bei schlechten Schulleistungen des Kindes durchzusetzen. Mein Beitrag in Kapitel 6 untersucht diese These; die Ergebnisse und entsprechende Implikationen werden in Abschnitt 1.6 zusammengefasst.

1.6 Fragestellungen und Befunde – Teil 2

Die Systematisierung im vorherigen Abschnitt hat aufgezeigt, über welche Wirkungsweisen Lehrkräfte primäre und sekundäre Disparitäten beeinflussen können. Der empirische Kenntnisstand zu diesen Wirkungsweisen weist allerdings noch viele weiße Flecken auf. Die im zweiten Teil der Dissertation versammelten Beiträge knüpfen daran an. Sie haben zum Ziel, auf der Basis empirischer Analysen über den Einfluss von Lehrkräften bei der Entstehung ungleicher Bildungschancen zu informieren. Dabei können nur einige der angesprochenen Wirkungsweisen überprüft werden. Die weiteren skizzierten Wirkungsweisen, jeweils mit Blick auf soziale, ethnische und geschlechtsspezifische Disparitäten, bleiben zukünftigen Forschungsbemühungen überlassen.

Anknüpfend an die Passungsthese (Abschnitt 1.5.1 c) untersucht der Beitrag in *Kapitel 4*, inwiefern Schulkinder von einer gleichgeschlechtlichen Lehrkraft profitieren im Hinblick auf Lernerfolg und Schulnoten. Mehrere Autoren vermuten, dass Jungen besser lernen oder besser benotet werden, wenn sie von einem Mann unterrichtet werden (BBC-News 2000; Dee 2007; Diefenbach und Klein 2002). Weil das angesichts der massiven Feminisierung des Lehrerberufs immer seltener der Fall ist, wäre es kaum verwunderlich, so die These, dass die Jungen in vielen schulischen Aspekten hinter den Mädchen zurückfallen. Diese Argumentation wird zum Teil von der Politik aufgegriffen. „Das schlechtere Abschneiden von Jungen liegt unter anderem daran, dass Kindergärten und Schulen weiblich dominiert sind“ (Bundesfamilienministerin Kristina Schröder in einem Interview in „Die Zeit“ vom 22.04.2010). Der Beitrag testet auf Basis von Individualdaten aus der IGLU-E Erhebung von 2001, ob Schulkinder von gleichgeschlechtlichen Lehrkräften profitieren. Als abhängige Variablen werden Leistungstests in Deutsch, Mathematik und Sachkunde herangezogen, ebenso wie Schulnoten in den drei Fächern. Es findet sich praktisch kein Beleg für die Passungsthese, weder für Jungen noch für Mädchen. Ein Beitrag auf Basis der aktuelleren IGLU Daten von 2006, der nicht Teil dieser Dissertation ist, repliziert diese Ergebnisse (Gerth und Neugebauer in Druck). Insgesamt deuten

die Befunde darauf hin, dass die Rekrutierung von Männern in den Schuldienst kein geeignetes Mittel darstellt, um die Bildungsmisserfolge der Jungen zu kompensieren.

Der Beitrag in **Kapitel 5** ergänzt die Analysen aus dem vorherigen Beitrag um einen wichtigen Aspekt. Er analysiert, inwiefern Jungen bei der Vergabe der Grundschulempfehlung von weiblichen Lehrkräften benachteiligt werden. Dazu wird zunächst ermittelt, ob die für die Empfehlungsvergabe verantwortliche Klassenkonferenz weiblich, männlich, oder gemischtgeschlechtlich zusammengesetzt ist. Als Datengrundlage dient wiederum die IGLU-E Erhebung von 2001. Der Beitrag findet keinen Beleg für die Passungsthese: Jungen werden nicht von weiblichen Klassenkonferenzen benachteiligt, wenn es um die Vergabe einer Gymnasialempfehlung geht. Die Analysen zeigen darüber hinaus, dass männliche Klassenkonferenzen, verglichen mit weiblichen Klassenkonferenzen seltener Gymnasialempfehlungen aussprechen – und zwar für Mädchen und Jungen gleichermaßen. Lehrer, so scheint es, urteilen insgesamt strenger als Lehrerinnen. Darüber hinaus finden sich Hinweise, dass Jungen bei den Gymnasialempfehlungen gar nicht benachteiligt werden, wenn man die Faktoren berücksichtigt, die laut den Vorgaben der Kultusministerkonferenz in die Empfehlung einfließen sollen. Jungen, die die gleiche Lernbereitschaft an den Tag legen wie Mädchen, werden sogar signifikant häufiger für das Gymnasium empfohlen.

Der Beitrag in **Kapitel 6** befasst sich mit dem Gewicht der Lehrerempfehlung bei der Aushandlung der Entscheidung über den Besuch der weiterführenden Schulform (und knüpft damit an den in Abschnitt 1.5.2 b diskutierten Mechanismus an). Während in der Mehrzahl der Bundesländer die Vergabe der Grundschulempfehlung eine „Empfehlung“ im eigentlichen Wortsinn darstellt, hat sie in einigen Ländern einen verbindlichen Charakter. Wenn bestimmte Leistungskriterien aus Sicht der Lehrer nicht erfüllt sind, wird der Zugang zum Gymnasium verwehrt. In dem Beitrag postuliere ich, dass in diesem Fall sekundären Effekten ein geringeres Gewicht zukommen sollte, da hier der Spielraum für die Durchsetzung schichtverzerrter Elternpräferenzen zugunsten einer durch Lehrer induzierten, meritokratischeren Schulwahl eingeschränkt wird. Auf Basis bundesweiter Längsschnittdaten des DJI Kinderpanels der Jahre 2002 bis 2005 wird zunächst das relative Gewicht von primären und sekundären Effekten bestimmt. In einem zweiten Schritt wird dann gezeigt, dass der relative Erklärungsbeitrag von sekundären Effekten größer ist in Ländern, in denen die Eltern frei entscheiden können und geringer, wenn die elterliche Wahlfreiheit durch eine verbindliche Grundschulempfehlung eingeschränkt wird.

Allerdings ist die Reduktion sekundärer Effekte nur moderat und kann nicht zufallskritisch abgesichert werden. Zudem ist die Gesamtungleichheit in Ländern mit verbindlicher Empfehlung *nicht* signifikant kleiner. Insofern deuten die Ergebnisse zwar – theoriekonform – auf eine ungleichheitsreduzierende Wirkung der verbindlichen Empfehlung bei den sekundären Effekten hin (vgl. dazu auch die zeitgleich bzw. später erschienenen Beiträge von Dollmann 2011; Gresch et al. 2010). Die Belege reichen aber aus meiner Sicht nicht aus, um daraus Empfehlungen für die Gestaltung der Verbindlichkeit der Grundschulempfehlung abzuleiten.

1.7 Literatur

- Aaronson, Daniel, Lisa Barrow und William Sander. 2007. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics* 25: 95-135.
- Aktionsrat Bildung (Hrsg.). 2009. *Geschlechterdifferenzen im Bildungssystem. Jahresgutachten 2009*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Alexander, Karl L., Doris R. Entwisle und Maxine S. Thompson. 1987. School Performance, Status Relations, and the Structure of Sentiment: Bringing the Teacher Back In *American Sociological Review* 52: 665-682.
- Ammermüller, Andreas, und Peter Dolton. 2006. Pupil-Teacher Gender Interaction Effects on Scholastic Outcomes in England and the USA. *ZEW Discussion Paper* 06-060: 1-32.
- Arnold, Karl-Heinz, Wilfried Bos, Peggy Richert und Tobias C. Stubbe. 2007. Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In *IGLU 2006 - Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Hrsg. Wilfried Bos, Sabine Hornberg, Karl-Heinz Arnold, Gabriele Faust, Lilian Fried, Eva Maria Lankes, Knut Schwippert und Renate Valtin, 271-298. Münster: Waxmann.
- Arrow, Kenneth J. 1973. The Theory of Discrimination. In *Discrimination in Labor Markets*, Hrsg. Orley Ashenfelter, Albert Rees, 3-33. Princeton: Princeton University Press.
- Babad, Elisha. 2007. Teachers' Nonverbal Behavior and its Effects on Students. In *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, Hrsg. Raymond P. Perry, John C. Smart, 201-261. New York: Springer.
- Bacolod, Marigee P. 2007. Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality. *The Review of Economics and Statistics* 89: 737-751.
- Baumert, Jürgen, und Mareike Kunter. 2006. Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft* 9: 469-520.
- Baumert, Jürgen, und Mareike Kunter. 2011. Das Kompetenzmodell von COACTIV. In *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, Hrsg. Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand, 29-53. Münster: Waxmann.
- Baumert, Jürgen, Petra Stanat und Rainer Watermann. 2006. Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000*, Hrsg. Jürgen Baumert, Petra Stanat und Rainer Watermann, 95-188. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baumert, Jürgen, Mareike Kunter, Werner Blum, Martin Brunner, Tamar Voss, Alexander Jordan, Uta Klusmann, Stefan Krauss, Michael Neubrand und Yi-Miau Tsai. 2010. Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal* 47: 133-180.
- BBC-News. 2000. Male Teachers for 'Role Models'. http://news.bbc.co.uk/2/hi/uk_news/education/893313.stm (Zugegriffen: 08.07.2010).
- Becker, Gary S. 1971. *The Economics of Discrimination*. Chicago: University of Chicago Press

- Becker, Rolf, und Walter Müller. 2011. Bildungsungleichheiten nach Geschlecht und Herkunft im Wandel. In *Geschlechtsspezifische Bildungsungleichheiten*, Hrsg. Andreas Hadjar, 55-75. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, Rolf, und Michael Beck. 2012. Herkunftseffekte oder statistische Diskriminierung von Migrantenkindern in der Primarstufe? In *Soziologische Bildungsforschung. Sonderheft 52 der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Hrsg. Rolf Becker, Heike Solga, 137-163. Wiesbaden: Springer.
- Blömeke, Sigrid, und Christiane Buchholtz. 2011. Familiäre und kognitive Bedingungen des Wissenserwerbs in Deutsch, Englisch und Mathematik. In *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen – Erste Ergebnisse aus TEDS-LT*, Hrsg. Sigrid Blömeke, Albert Bremerich-Vos, Helga Haudeck, Gabriele Kaiser, Rainer Lehmann, Günter Nold, Knut Schwippert und Heiner Willenberg. Münster: Waxmann.
- Blömeke, Sigrid, Gabriele Kaiser und Martina Döhrmann. 2011. Bedingungsfaktoren des fachbezogenen Kompetenzerwerbs von Lehrkräften. Zum Einfluss von Ausbildungs-, Persönlichkeits- und Kompositionsmerkmalen in der Mathematiklehrausbildung für die Sekundarstufe I. *Zeitschrift für Pädagogik* 57. Beiheft: 77-103.
- Blömeke, Sigrid, Ute Suhl, Gabriele Kaiser und Martina Döhrmann. 2012. Family Background, Entry Selectivity and Opportunities to Learn: What Matters in Primary Teacher Education? An International Comparison of Fifteen Countries. *Teaching and Teacher Education* 28: 44-55.
- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther. 2004a. *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Bos, Wilfried, Andreas Voss, Eva-Maria Lankes, Knut Schwippert, Oliver Thiel und Renate Valtin. 2004b. Schullaufbahneempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe. In *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*, Hrsg. Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther, 191-228. Münster: Waxmann.
- Boudon, Raymond. 1974. *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*. New York: Wiley.
- Boyd, Donald J., Pamela L. Grossman, Hamilton Lankford, Susanna Loeb und James Wyckoff. 2009. Teacher Preparation and Student Achievement. *Educational Evaluation and Policy Analysis* 31: 416-440.
- Breen, Richard, und John H. Goldthorpe. 1997. Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society* 9: 275-305.
- Bromme, Rainer. 1997. Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In *Psychologie des Unterrichts und der Schule*. Pädagogische Psychologie, Band 3, Hrsg. Franz E. Weinert, 177-212. Göttingen: Hogrefe.
- Butler, Ruth. 2007. Teachers' Achievement Goal Orientations and Associations with Teachers' Help Seeking: Examination of a Novel Approach to Teacher Motivation. *Journal of Educational Psychology* 99: 241-252.

- Calderhead, James. 1996. Teachers: Beliefs and Knowledge. In *Handbook of Educational Psychology*, Hrsg. David C. Berliner, Robert C. Calfee, 709-725. New York: Macmillan.
- Chetty, Raj, John N. Friedman und Jonah E. Rockoff. 2011. The Long-term Impacts of Teachers: Teacher Value-added and Student Outcomes in Adulthood. *NBER Working Paper* 17699.
- Cho, Insook. 2012. The Effect of Teacher–Student Gender Matching: Evidence from OECD Countries. *Economics Of Education Review* 31: 54-67.
- Clotfelter, Charles T., Helen F. Ladd und Jacob L. Vigdor. 2007. Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects. *Economics Of Education Review* 26: 673-682.
- Coenen, Johan, und Chris Van Klaveren. 2013. Better Test Scores with a Same-gender Teacher? *Tier Working Paper Series* 13/03.
- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld und Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, D.C: United States Government Printing Office.
- Corcoran, Sean P, Jennifer L Jennings und Andrew A Beveridge. 2011. Teacher effectiveness on high-and low-stakes tests. Paper originally delivered to the SREE Conference Washington, DC. https://files.nyu.edu/sc129/public/papers/corcoran_jennings_beveridge_2011_wkg_teacher_effects.pdf (Zugegriffen: 01.06.2013).
- Corcoran, Sean P., William N. Evans und Robert M. Schwab. 2004. Women, the Labor Market, and the Declining Relative Quality of Teachers. *Journal of Policy Analysis and Management* 23: 449-470.
- Darling-Hammond, Linda. 2000. Teacher Quality and Student Achievement: A Review of State Policy Evidence. *Education Policy Analysis Archive* 8: 1-44.
- Dee, Thomas S. 2004. Teachers, Race, and Student Achievement in a Randomized Experiment. *Review of Economics and Statistics* 86: 195-210.
- Dee, Thomas S. 2007. Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources* 42: 528-554.
- Deutsches PISA-Konsortium (Hrsg.). 2001. *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske+Budrich.
- Diefenbach, Heike, und Michael Klein. 2002. "Bringing Boys Back In". Soziale Ungleichheit zwischen den Geschlechtern im Bildungssystem zuungunsten von Jungen am Beispiel der Sekundarschulabschlüsse. *Zeitschrift für Pädagogik* 48: 938-958.
- Ditton, Hartmut, und Jan Krüsken. 2006. Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft* 9: 348-372.
- Ditton, Hartmut, Jan Krüsken und Magdalena Schauenberg. 2005. Bildungsungleichheit - der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft* 8: 285-304.
- Dollmann, Jörg. 2010. *Türkischstämmige Kinder am ersten Bildungsübergang. Primäre und sekundäre Herkunftseffekte*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Dollmann, Jörg. 2011. Verbindliche und unverbindliche Grundschulempfehlungen und soziale Ungleichheiten am ersten Bildungsübergang. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 63: 595-621.

- Dubberke, Thamar, Mareike Kunter, Nele McElvany, Martin Brunner und Jürgen Baumert. 2008. Lerntheoretische Überzeugungen von Mathematiklehrkräften. *Zeitschrift für Pädagogische Psychologie* 22: 193-206.
- Ehrenberg, Ronald G., Daniel G. Goldhaber und Dominic J. Brewer. 1995. Do Teachers' Race, Gender and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review* 48: 547-561
- Erikson, Robert, und Jan O. Jonsson. 1996. Explaining Class Inequality in Education: The Swedish Test Case. In *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, Hrsg. Robert Erikson, Jan O. Jonsson, 1-63. Boulder: Westview Press.
- Erikson, Robert, und Frida Rudolph. 2010. Change in Social Selection to Upper Secondary School - Primary and Secondary Effects in Sweden. *European Sociological Review* 26: 291-305.
- Esser, Hartmut. 1999. *Soziologie: Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt a. M.: Campus.
- Farkas, George. 2003. Racial Disparities and Discrimination in Education: What do we know, how do we know it, and what do we need to know? *The Teachers College Record* 105: 1119-1146.
- Fend, Helmut. 1981. *Theorie der Schule*, 2. Auflage. München: Urban & Schwarzenberg.
- Frenzel, Anne C., Thomas Goetz, Oliver Lüdtke, Reinhard Pekrun und Rosemary E. Sutton. 2009. Emotional Transmission in the Classroom: Exploring the Relationship Between Teacher and Student Enjoyment. *Journal of Educational Psychology* 101: 705.
- Gerth, Maria, und Martin Neugebauer. in Druck. Die Feminisierung des Bildungswesens und der Schulerfolg von Jungen. In *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen*, Hrsg. Rolf Becker, Alexander Schulze. Wiesbaden: Springer VS.
- Goe, Laura. 2007. *The Link Between Teacher Quality and Student Outcomes: A Research Synthesis*. Washington D.C.: National Comprehensive Center for Teacher Quality.
- Gresch, Cornelia, Jürgen Baumert und Kai Maaz. 2010. Empfehlungsstatus, Übergangsempfehlung und der Wechsel in die Sekundarstufe I: Bildungsentscheidungen und soziale Ungleichheit. In *Bildungsentscheidungen. Sonderheft 12 der Zeitschrift für Erziehungswissenschaft*, Hrsg. Jürgen Baumert, Kai Maaz und Ulrich Trautwein, 230-256. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hanushek, Eric A., und Steven G. Rivkin. 2010a. Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review* 100: 267-271.
- Hanushek, Eric A., und Steven G. Rivkin. 2010b. Constrained Job Matching: Does Teacher Job Search Harm Disadvantaged Urban Schools? *NBER Working Paper* 15816.
- Harris, Douglas N., und Tim R. Sass. 2011. Teacher Training, Teacher Quality and Student Achievement. *Journal of Public Economics* 95: 798-812.
- Hattie, John. 2009. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London: Routledge.
- Helmke, Andreas. 2012. *Unterrichtsqualität und Lehrerprofessionalität - Diagnose, Evaluation und Verbesserung des Unterrichts*, 4. Auflage. Seelze: Klett-Kallmeyer.

- Hill, Heather C., Brian Rowan und Deborah Loewenberg Ball. 2005. Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal* 42: 371-406.
- Hillert, Andreas, und Oliver Christ. 2004. *Psychosomatische Erkrankungen bei Lehrerinnen und Lehrern*. Stuttgart: Schattauer.
- Hofer, Manfred. 2009. Kompetenz im Umgang mit Schülerheterogenität als Beitrag zur Bildungsgerechtigkeit. In *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung*, Hrsg. Olga Zlatkin-Troitschanskaia, Klaus Beck, Detlef Sembill, Reinhold Nickolaus und Regina Mulder, 141-152. Weinheim: Beltz Verlag.
- Hosenfeld, Ingmar, Andreas Helmke, Arnd Ridder und Friedrich-Wilhelm Schrader. 2001. Eine mehrbenenanalytische Betrachtung von Schul- und Klasseneffekten. *Empirische Pädagogik* 15: 513-534.
- Jackson, Michelle, Robert Erikson, John H. Goldthorpe und Meir Yaish. 2007. Primary and Secondary Effects in Class Differentials in Educational Attainment: The Transition to A-Level Courses in England and Wales. *Acta Sociologica* 50: 211-229.
- Jacob, Brian A., und Lars Lefgren. 2008. Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education. *Journal of Labor Economics* 26: 101-136.
- Jennings, Jennifer L., und Thomas A. DiPrete. 2010. Teacher Effects on Social and Behavioral Skills in Early Elementary School. *Sociology of Education* 83: 135-159.
- Jussim, Lee, Stacy L. Robustelli und Thomas R. Cain. 2009. Teacher Expectations and Self-fulfilling Prophecies. In *Handbook of Motivation at School*, Hrsg. Kathryn R. Wentzel, Allan Wigfield, 349-380. New York: Routledge.
- Kane, Thomas J., und Douglas O. Staiger. 2008. Estimating the Impacts on Student Achievement: An Experimental Evaluation. *NBER Working Paper* 14607.
- Kane, Thomas J., Jonah E. Rockoff und Douglas O. Staiger. 2008. What Does Certification Tell Us About Teacher Effectiveness? Evidence From New York City. *Economics of Education Review* 27: 615-631.
- Kelcey, Ben. 2011. Assessing the Effects of Teachers' Reading Knowledge on Students' Achievement using Multilevel Propensity Score Stratification. *Educational Evaluation and Policy Analysis* 33: 458-482.
- Klauer, Karl Josef, und Detlev Leutner. 2010. Intelligenz und Begabung. In *Handwörterbuch Pädagogische Psychologie*, Hrsg. Detlef H. Rost, 304-311. Weinheim: Beltz.
- Kleickmann, Thilo, und Yvonne Anders. 2011. Lernen an der Universität. In *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, Hrsg. Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand, 305-315. Münster: Waxmann.
- Kleickmann, Thilo, Dirk Richter, Mareike Kunter, Jürgen Elsner, Michael Besser, Stefan Krauss und Jürgen Baumert. 2013. Teachers' Content Knowledge and Pedagogical Content Knowledge: The Role of Structural Differences in Teacher Education. *Journal of Teacher Education* 64: 90-106.
- Klemm, Klaus. 2008. Bildungsausgaben: Woher sie kommen, wohin sie fließen. In *Das Bildungswesen in der Bundesrepublik Deutschland*, Hrsg. Kai S. Cortina, Jürgen

- Baumert, Achim Leschinsky, Karl Ullrich Mayer und Luitgard Trommer, 245-280. Reinbek bei Hamburg: Rowohlt.
- Klusmann, Uta. 2011. Allgemeine berufliche Motivation und Selbstregulation. In *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, Hrsg. Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand, 277-294. Münster: Waxmann.
- Klusmann, Uta, Mareike Kunter, Ulrich Trautwein, Oliver Lüdtke und Jürgen Baumert. 2008. Teachers' Occupational Well-Being and Quality of Instruction: The Important Role of Self-Regulatory Patterns. *Journal of Educational Psychology* 100: 702-715.
- KMK. 2013. Lehrereinstellungsbedarf und -angebot in der Bundesrepublik Deutschland. Modellrechnung 2012 - 2025. Statistische Veröffentlichungen der Kultusministerkonferenz. Bonn. Dokumentation Nr. 201 - Juni 2013.
- Koedel, Cory, und Julian R. Betts. 2007. *Re-Examining the Role of Teacher Quality In the Educational Production Function*. San Diego: University of California.
- Köller, Olaf, und Ulrich Schiefele. 2010. Zielorientierung. In *Handwörterbuch Pädagogische Psychologie*, Hrsg. Detlef H. Rost, 959-965. Weinheim: Beltz.
- König, Johannes, und Sigrid Blömeke. 2009. Pädagogisches Wissen von angehenden Lehrkräften. Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerbildung. *Zeitschrift für Erziehungswissenschaft* 12: 499-527.
- König, Johannes, und Andreas Seifert (Hrsg.). 2012. *Lehramtsstudierende erwerben pädagogisches Professionswissen: Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung*. Münster: Waxmann.
- König, Johannes, Sarantis Tachtsoglou und Andreas Seifert. 2012. Individuelle Voraussetzungen, Lerngelegenheiten und der Erwerb von pädagogischem Professionswissen. In *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung*, Hrsg. Johannes König, Andreas Seifert, 243-283. Münster: Waxmann.
- Kratzmann, Jens, und Thorsten Schneider. 2009. Soziale Ungleichheiten beim Schulstart. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 61: 211-234.
- Kristen, Cornelia. 2006. Ethnische Diskriminierung in der Grundschule. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58: 79-97.
- Krohne, Julia A., Ulrich Meier und Klaus-Jürgen Tillmann. 2004. Sitzenbleiben, Geschlecht und Migration - Klassenwiederholungen im Spiegel der PISA-Daten. *Zeitschrift für Pädagogik* 50: 373-391.
- Kunina-Habenicht, Olga, Franziska Schulze-Stocker, Mareike Kunter, Jürgen Baumert, Detlev Leutner, Doris Förster, Hendrik Lohse-Bossenz und Ewald Terhart. 2013. Die Bedeutung der Lerngelegenheiten im Lehramtsstudium und deren individuelle Nutzung für den Aufbau des bildungswissenschaftlichen Wissens. *Zeitschrift für Pädagogik* 59: 1-23.
- Kunter, Mareike. 2011a. Motivation als Teil der professionellen Kompetenz - Forschungsbefunde zum Enthusiasmus von Lehrkräften. In *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, Hrsg. Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand, 259-275. Münster: Waxmann.

- Kunter, Mareike. 2011b. Forschung zur Lehrermotivation. In *Handbuch der Forschung zum Lehrerberuf*, Hrsg. Ewald Terhart, Hedda Bennewitz und Martin Rothland, 527-539. Münster: Waxmann.
- Kunter, Mareike, und Jürgen Baumert. 2011. Das COACTIV-Forschungsprogramm zur Untersuchung professioneller Kompetenz von Lehrkräften - Zusammenfassung und Diskussion. In *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, Hrsg. Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand, 345-366. Münster: Waxmann.
- Kunter, Mareike, Thilo Kleickmann, Uta Klusmann und Dirk Richter. 2011a. Die Entwicklung professioneller Kompetenzen von Lehrkräften. In *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, Hrsg. Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand, 55-68. Münster: Waxmann.
- Kunter, Mareike, Anne Frenzel, Gabriel Nagy, Jürgen Baumert und Reinhard Pekrun. 2011b. Teacher Enthusiasm: Dimensionality and Context Specificity. *Contemporary Educational Psychology* 36: 289-301.
- Kunter, Mareike, Yi-Miau Tsai, Uta Klusmann, Martin Brunner, Stefan Krauss und Jürgen Baumert. 2008. Students' and Mathematics Teachers' Perceptions of Teacher Enthusiasm and Instruction. *Learning and Instruction* 18: 468-482.
- Kunter, Mareike, Uta Klusmann, Jürgen Baumert, Dirk Richter, Tamar Voss und Axinja Hachfeld. in Druck. Professional Competence of Teachers: Effects on Instructional Quality and Student Development. *Journal of Educational Psychology*.
- Lehmann, Rainer H., Rainer Peek und Rüdiger Gänsfuß. 1997. Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern, die im Schuljahr 1996/97 eine fünfte Klasse an Hamburger Schulen besuchten. Hamburg: Behörde für Schule, Jugend und Berufsbildung.
- Lipowsky, Frank. 2006. Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. *Zeitschrift für Pädagogik* 51. Beiheft: 47-65.
- McCaffrey, Daniel F., J.R. Lockwood, Daniel M. Koretz und Laura S. Hamilton. 2004. Models for Value-added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics* 29: 67-101.
- McGrady, Patrick B., und John R. Reynolds. 2013. Racial Mismatch in the Classroom: Beyond Black-white Differences. *Sociology of Education* 86: 3-17.
- Metzler, Johannes, und Ludger Wößmann. 2012. The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation. *Journal of Development Economics* 99: 486-496.
- Neugebauer, Martin. 2010. Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte. *Zeitschrift für Soziologie* 39: 202-214.
- Neugebauer, Martin, Marcel Helbig und Andreas Landmann. 2011. Unmasking the Myth of the Same-Sex Teacher Advantage. *European Sociological Review* 27: 669-689.

- Nickell, Stephen, und Glenda Quintini. 2002. The Consequences of the Decline in Public Sector Pay in Britain: A Little Bit of Evidence. *The Economic Journal* 112: F107-F118.
- Nye, Barbara, Spyros Konstantopoulos und Larry V. Hedges. 2004. How Large Are Teacher Effects? *Educational Evaluation and Policy Analysis* 26: 237-257.
- OECD. 2005. *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: OECD Publishing.
- Pajares, M. Frank. 1992. Teachers' Beliefs and Educational Research: Cleaning up a Messy Construct. *Review of Educational Research* 62: 307-332.
- Papay, John P. 2011. Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates Across Outcome Measures. *American Educational Research Journal* 48: 163-193.
- Parsons, Talcott. 1959. The School Class as System: Some of Its Functions in American Society. *Harvard Educational Review* 29: 297 – 318.
- Phelps, Edmund S. 1972. The Statistical Theory of Racism and Sexism. *The American Economic Review* 62: 659-661.
- Pintrich, Paul R. 2003. A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology* 95: 667.
- Rivkin, Steven G., Eric A. Hanushek und John F. Kain. 2005. Teachers, Schools, and Academic Achievement. *Econometrica* 73: 417-458.
- Rockoff, Jonah E. 2004. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review* 94: 247-252.
- Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane und Douglas O. Staiger. 2011. Can You Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy* 6: 43-74.
- Rosenthal, Robert, und Lenore Jacobson. 1968. Pygmalion in the Classroom. *The Urban Review* 3: 16-20.
- Roth, Guy, Avi Assor, Yaniv Kanat-Maymon und Haya Kaplan. 2007. Autonomous Motivation for Teaching: How Self-Determined Teaching May Lead to Self-Determined Learning. *Journal of Educational Psychology* 99: 761.
- Rothland, Martin. 2011a. Wer entscheidet sich für den Lehrerberuf? Forschung zum soziodemographischen Profil sowie zu Persönlichkeits- und Leistungsmerkmalen angehender Lehrkräfte. In *Handbuch der Forschung zum Lehrerberuf*, Hrsg. Ewald Terhart, Hedda Bennewitz und Martin Rothland, 243-267. Münster: Waxmann.
- Rothland, Martin. 2011b. Warum entscheiden sich Studierende für den Lehrerberuf? Interessen, Orientierungen und Berufswahlmotive angehender Lehrkräften im Spiegel der empirischen Forschung. In *Handbuch der Forschung zum Lehrerberuf*, Hrsg. Ewald Terhart, Hedda Bennewitz und Martin Rothland, 268-295. Münster: Waxmann.
- Rothland, Martin. 2013. Belastung und Beanspruchung im Lehrerberuf und die Modellierung professioneller Kompetenz von Lehrerinnen und Lehrern. In *Belastung und Beanspruchung im Lehrerberuf*, Hrsg. Martin Rothland, 7-20. Wiesbaden: Springer.
- Rothland, Martin, und Ewald Terhart. 2011. Eignungsabklärung angehender Lehrerinnen und Lehrer. Einführung in den Thementeil. *Zeitschrift für Pädagogik* 57: 635-638.
- Rothstein, Jesse. 2010. Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *The Quarterly Journal of Economics* 125: 175-214.

- Rowan, Brian, Richard Correnti und Robert J. Miller. 2002. What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools. *The Teachers College Record* 104: 1525-1567.
- Schaarschmidt, Uwe. 2005. *Halbtagsjobber? Psychische Gesundheit im Lehrerberuf – Analyse eines veränderungsbedürftigen Zustandes*, 2. Auflage. Weinheim: Beltz.
- Schaarschmidt, Uwe, und Andreas W. Fischer. 2003. *AVEM – Arbeitsbezogenes Verhaltens- und Erlebensmuster*, 2. Auflage. Frankfurt a.M.: Swets & Zeitlinger.
- Scheerens, Jaap, und Roel J. Bosker. 1997. *The Foundations of Educational Effectiveness*. Oxford: Pergamon.
- Schindler, Steffen, und David Reimer. 2010. Primäre und sekundäre Effekte der sozialen Herkunft beim Übergang in die Hochschulbildung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62: 623-653.
- Schleicher, Andreas. 2011. *Building a High-Quality Teaching Profession: Lessons from around the World*. Paris: OECD Publishing.
- Schneider, Thorsten. 2011. Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschulempfehlung. *Zeitschrift für Erziehungswissenschaft* 14: 371-396.
- Seidel, Tina, und Richard J. Shavelson. 2007. Teaching Effectiveness Research in the past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research* 77: 454-499.
- Shulman, Lee S. 1987. Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review* 57: 1-22.
- Sokal, Laura, Herb Katz, Les Chaszewski und Cecilia Wojcik. 2007. Good-bye, Mr. Chips: Male Teacher Shortages and Boys' Reading Achievement. *Sex Roles* 56: 651-659.
- Sørensen, Aage B., und Maureen T. Hallinan. 1977. A Reconceptualization of School Effects. *Sociology of Education* 50: 522-535.
- Sprietsma, Maresa. in Druck. Discrimination in Grading: Experimental Evidence from Primary School Teachers. *Empirical Economics*: 1-16.
- Terhart, Ewald. 2008. Die Lehrerbildung. In *Das Bildungswesen in der Bundesrepublik Deutschland*, Hrsg. Kai S. Cortina, Jürgen Baumert, Achim Leschinsky, Karl Ullrich Mayer und Luitgard Trommer, 745-772. Reinbek bei Hamburg: Rowohlt.
- Terhart, Ewald. 2012a. Was wissen wir über Lehrerinnen und Lehrer? *Pädagogik* 64: 43-47.
- Terhart, Ewald. 2012b. Wie wirkt Lehrerbildung? Forschungsprobleme und Gestaltungsfragen. *Zeitschrift für Bildungsforschung* 2: 3-21.
- Tiedemann, Joachim, und Elfriede Billmann-Mahecha. 2007. Zum Einfluss von Migration und Schulklassenzugehörigkeit auf die Übergangsempfehlung für die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft* 10: 108-120.
- Voss, Tamar, Thilo Kleickmann, Mareike Kunter und Axinja Hachfeld. 2011. Überzeugungen von Mathematiklehrkräften. In *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, Hrsg. Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand, 235-257. Münster: Waxmann.
- Wagner, Wolfgang, Andreas Helmke und Friedrich-Wilhelm Schrader. 2010. Die Rekonstruktion der Übergangsempfehlung für die Sekundarstufe I und der Wahl des

- Bildungsgang auf der Basis des Migrationsstatus, der sozialen Herkunft, der Schulleistung und schulklassenspezifischer Merkmale. In *Bildungsentscheidungen. Sonderheft (Band 12) der Zeitschrift für Erziehungswissenschaft*, Hrsg. Jürgen Baumert, Kai Maaz und Ulrich Trautwein, 183-204. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Watt, Helen, und Paul W. Richardson. 2007. Motivational Factors influencing teaching as a Career Choice: Development and Validation of the FIT-Choice Scale. *The Journal of Experimental Education* 75: 167-202.
- Wayne, Andrew J., und Peter Youngs. 2003. Teacher Characteristics and Student Achievement Gains: A Review. *Review of Educational Research* 73: 89-122.
- Weber, Andreas, Dieter Weltle und Peter Lederer. 2004. Frühinvalidität im Lehrerberuf: Sozial- und arbeitsmedizinische Aspekte. *Deutsches Ärzteblatt* 101: 850-859.
- Wiese, Wilhelm. 1982. Elternstatus, Lehrerempfehlung und Schullaufbahn: Eine empirische Analyse des Einflusses des Grundschullehrers auf die Bildungslaufbahn des Schülers. *Zeitschrift für Soziologie* 11: 49-63.
- Wild, T. Cameron, Michael E. Enzle und Wendy L. Hawkins. 1992. Effects of Perceived Extrinsic versus Intrinsic Teacher Motivation on Student Reactions to Skill Acquisition. *Personality and Social Psychology Bulletin* 18: 245-251.
- Winters, Marcus A., Robert C. Haight, Thomas T. Swaim und Kathy Pickering. 2013. The Effect of Same-Gender Teacher Assignment on Student Achievement in the Elementary and Secondary Grades: Evidence from Panel Data. *Economics of Education Review* 34: 69–75.

Kapitel 2 Wer entscheidet sich für ein Lehramtsstudium – und warum? Eine empirische Überprüfung der These von der Negativselektion in den Lehrerberuf

Erschienen als: Neugebauer, Martin. 2013. Wer entscheidet sich für ein Lehramtsstudium – und warum? Eine empirische Überprüfung der These von der Negativselektion in den Lehrerberuf. *Zeitschrift für Erziehungswissenschaft* 16: 157-184. Abdruck mit Erlaubnis von Springer Science+Business Media.

Abstract: In öffentlichen Diskussionen wird häufig die Annahme formuliert, dass sich Personen mit ungünstigen kognitiven und motivationalen Merkmalen für den Lehrerberuf entscheiden. Vor diesem Hintergrund untersucht der vorliegende Beitrag anhand einer deutschlandweiten Stichprobe von 14815 Studienanfängern die leistungsbezogenen, motivationalen und soziodemographischen Eingangsmerkmale von Lehramtsstudierenden unterschiedlicher Schulformen im Vergleich zu anderen Studierenden. Die Ergebnisse zeigen, dass sich Gymnasiallehramtsstudierende im Hinblick auf Abiturleistungen und fachliche Studienwahlmotivation nicht von anderen Universitätsstudierenden unterscheiden. Allerdings sind den Gymnasiallehramtsstudierenden wissenschaftliche Interessen, Karrieremotive und das Ziel, überdurchschnittliche Kompetenzen erwerben zu wollen, weniger wichtig als anderen Universitätsstudierenden. Ausgeprägter sind hingegen soziale Interessen, der Wunsch nach beruflicher Sicherheit und private Vereinbarkeitsziele. Die genannten Unterschiede gelten ebenfalls für Studienanfänger des Real-, Grund-/Haupt- sowie Sonderschullehramtes. Darüber hinaus haben die nicht-gymnasialen Lehramtsstudienanfänger schlechtere Abiturleistungen und geringere fachliche sowie wissenschaftliche Studienwahlmotivationen als ihre Kommilitonen im Gymnasiallehramt. Des Weiteren zeigt sich, dass insbesondere nicht-gymnasiale Lehramtsstudiengänge vermehrt weibliche Studienberechtigte und solche aus bildungsfernen sozialen Schichten anziehen.

2.1 Einleitung

Für die meisten Bildungssysteme dieser Welt ist die Rekrutierung kompetenter Lehrer¹ ein zentrales Anliegen (OECD 2005). Schließlich hängt der Lernfortschritt eines Schülers in großem Maße von der Kompetenz seiner Lehrkräfte ab (Hanushek und Rivkin 2010). Spätestens seit dem „PISA-Schock“ sind diese wieder vermehrt ins bildungspolitische Interesse gerückt. Das schlechte Abschneiden deutscher Schüler in internationalen Schulleistungstudien sei, so eine populäre Annahme, auch einem Mangel an kompetenten Lehrern geschuldet

¹ Im Sinne eines besseren Leseflusses wird auf die Verwendung beider Geschlechtsformen der Substantive verzichtet.

(McKinsey&Company 2007). In der öffentlichen Diskussion werden vermeintliche Kompetenzdefizite von Lehrern häufig damit begründet, dass sich die falschen Personen für diesen Berufsweg entscheiden (für eine Darstellung der Lehrer in den Printmedien vgl. Blömeke 2005). Gängigen Alltagstheorien nach zu urteilen, mangle es Lehramtsstudierenden mitunter an kognitiven und/oder motivationalen Lernvoraussetzungen, die für ein wissenschaftliches Studium notwendig seien. Solche durch negative Selbstselektion induzierten ungünstigen Voraussetzungen angehender Lehrer, so die These, tragen dazu bei, dass diese ihren späteren beruflichen Anforderungen nicht gerecht werden (Rauin 2007). Dahinter steckt die Annahme, dass der Kompetenzerwerb angehender Lehrkräfte nicht alleine von der Ausbildungsqualität abhängt, sondern auch entscheidend davon, wie Studierende vorhandene Lerngelegenheiten auf Basis ihrer individuellen Dispositionen nutzen (zur Debatte um die Eignungsabklärung von Lehramtskandidaten vgl. Rothland und Terhart 2011). Vor diesem Hintergrund stellt sich für Wissenschaft und Politik gleichermaßen die Frage nach den Eingangsvoraussetzungen von zukünftigen Lehrern. Mit anderen Worten: Wer entscheidet sich für ein Lehramtsstudium – und warum?²

Angesichts der Hartnäckigkeit der Vorurteile ist es erstaunlich, wie unbefriedigend der empirische Kenntnisstand zu den leistungsbezogenen, motivationalen und soziodemographischen Eingangsmerkmalen von Lehramtsstudierenden ist. Zwar gibt es eine Vielzahl an Studien zu diesem Thema (für eine Übersicht im deutschsprachigen Raum vgl. Rothland 2011a, b), allerdings basieren diese *erstens* fast ausschließlich auf kleinen Gelegenheitsstichproben ungeklärter Repräsentativität, *zweitens* sind sie geographisch und/ oder fachspezifisch begrenzt. Die amerikanische Übersicht von Brookhart und Freeman (1992) zeigt beispielsweise, dass von 44 Studien nur drei an mehr als einem Campus durchgeführt wurden. Ähnlich ist die Situation im deutschsprachigen Raum. *Drittens* – und dieser Punkt ist besonders kritisch – beschränken sich die meisten Studien auf Lehramtsstudierende; durch das Fehlen von Vergleichsgruppen ist es kaum möglich, die Eigenständigkeit bestimmter Merkmale bei der Wahl des Lehramtsstudiums zu identifizieren. Das zeigt sich beispielsweise bei den relativ häufig untersuchten Studienwahlmotiven. Sie gelten neben kognitiven Merkmalen als bedeutsame Eingangsmerkmale, da bestimmte Motivkonstellationen Auswirkungen auf Lernstrategien und Wissenserwerb im Studium haben (vgl. dazu Abschnitt 2.5.3). Ob sich Lehramtsanwärter beispielsweise aus besonders fachinteressierten Studienanfängern rekrutieren, können diese Studien nicht klären.

² Die Entscheidung für das Lehramtsstudium ist der ‚kritische Übergang‘ auf dem Weg zum Lehrerberuf. 2009 handelte es sich bei lediglich 5.9 % aller Einstellungen in den öffentlichen Schuldiensten um Seiteneinsteiger (KMK 2010).

Entsprechend konstatiert Rothland, dass „bisherige Studien – wenn überhaupt – eher explorativen Charakter“ haben (2011a, S. 284). Positive Ausnahmen bilden eine ältere Untersuchung von Gold und Giesen (1993) sowie eine kürzlich vorgelegte Studie von Klusmann et al. (2009) auf Basis einer baden-württembergischen Stichprobe. Dabei werden kognitive Fähigkeiten, berufliche Interessen, Persönlichkeits- und soziodemographische Merkmale angehender Lehramtsstudierender mit denen anderer Studierender verglichen; Studienwahlmotive werden jedoch nicht analysiert. Die beiden Studien stellen die bislang umfassendsten Untersuchungen zu den Eingangsmerkmalen von Lehramtsstudierenden in Deutschland dar. Allerdings basieren die Befunde auf Studienberechtigtenkohorten aus den Siebzigerjahren (Gold und Giesen 1993) oder auf Daten aus nur einem Bundesland (Klusmann et al. 2009). Aufgrund der Besonderheiten der baden-württembergischen Lehrerausbildung (nicht-gymnasiale Lehramtsstudierende werden an Pädagogischen Hochschulen ausgebildet) lassen sich die Befunde nicht ohne weiteres auf Gesamtdeutschland übertragen. Zudem schränken die Fallzahlen die Möglichkeiten differenzierterer Analysen nach Art des Lehramtes ein.

Ziel des vorliegenden Beitrags ist es, die bisherige Befundlage zu ergänzen, indem *erstens* auf Basis einer breiten, deutschlandweiten Erhebung leistungsbezogene und soziodemographische Eingangsmerkmale von Lehramtsstudierenden mit anderen Studierenden verglichen werden. *Zweitens* werden zusätzlich Studienwahlmotive und langfristige Berufs- und Lebensziele analysiert, für welche vergleichende Analysen bislang fehlen. *Drittens* können Lehramtsstudierende differenzierter als bisher nach angestrebter Schulform betrachtet werden, wodurch mögliche Unterschiede zwischen den Lehrämtern aufgespürt werden können. Der Beitrag ist wie folgt strukturiert: In Abschnitt 2.2 wird zunächst theoretisch hergeleitet, welche Eingangsmerkmale bei Lehramtsstudierenden im Vergleich zu anderen Studierenden zu erwarten sind. Dazu wird auch der aktuelle Forschungsstand rezipiert. In Abschnitt 2.3 werden die Methoden vorgestellt, bevor in Abschnitt 2.4 die Ergebnisse präsentiert werden. Abschnitt 2.5 fasst die Befunde zusammen und bewertet die eingangs gestellte Frage, ob angesichts der Befunde tatsächlich von einer Negativselektion in das Lehramt gesprochen werden kann.

2.2 Theoretische Bezüge und empirische Evidenz

Warum entscheidet sich ein Studienberechtigter angesichts zahlreicher Alternativen für ein Lehramtsstudium? Erwartungswerttheorien beschreiben Wahlentscheidungen als Folge eines subjektiv rationalen Abwägungsprozesses, der direkt mit den *Erfolgserwartungen* und dem *subjektiven Wert* der Entscheidungsalternativen verbunden ist (z. B. Wigfield und Eccles 2000;

Esser 1999, 247 ff.). Der Wert einer Alternative ergibt sich dabei nicht ausschließlich aus intrinsischen Aspekten, wie Spaß oder Interesse an einer Tätigkeit. Ebenso sollten antizipierte Nutzen- (z. B. berufliche Sicherheit) oder Kostenaspekte (z. B. Aufwand für ein Studium) eine Rolle spielen. Bezogen auf die Wahl des Lehramtsstudiums bedeutet das: Ein Studienberechtigter sollte sich für ein Lehramtsstudium entscheiden, wenn er/sie damit verbundene Tätigkeiten und Berufsmerkmale als subjektiv wertvoller erachtet als diejenigen der Alternativen und wenn er/sie zudem glaubt, den Beruf und das dafür notwendige Studium erfolgreich bewältigen zu können. Einen weiteren theoretischen Ansatz bieten Studien- und Berufswahltheorien (z. B. Holland 1997), die das Wahlverhalten im Wesentlichen als ein Streben nach optimaler Passung (Kongruenz) beschreiben. Eine zentrale Annahme dieser Theorien besagt, dass Abiturienten solche nachschulischen Werdegänge suchen und auswählen, die einerseits zu ihren Interessen und andererseits zu ihren Fähigkeiten passen (für empirische Prüfungen dieser Theorie vgl. Bergman 1992; Smart et al. 2000). Die Vorhersagen hinsichtlich der Studienwahl, die sich aus solchen kongruenztheoretischen Überlegungen ableiten, sind dabei durchaus kompatibel mit Erwartungswerttheorien. Schließlich sollte die subjektive Erfolgserwartung mit der Passung des antizipierten Anforderungsniveaus zu den eigenen Fähigkeiten steigen; der subjektive intrinsische Wert sollte höher bewertet werden, wenn Studiengang und damit verbundene Berufe den eigenen Interessen entsprechen.

Der vorliegende Beitrag knüpft an diese theoretischen Überlegungen an, um daraus abzuleiten, für wen ein Lehramtsstudium attraktiv sein sollte und für wen nicht – und was das für Konsequenzen für die Eingangsmerkmale von Lehramtsstudierenden hat. Das Lehramt kennzeichnet die Besonderheit, dass mit der Studienwahl wie in nur wenigen anderen Studiengängen ein festes Berufsbild verbunden ist. Entsprechend sollten antizipierte Berufsmerkmale (mehr als Studienmerkmale) bei der Bewertung der Entscheidungsalternative Lehramt eine wichtige Rolle spielen.

2.2.1 Leistungsbezogene Eingangsmerkmale

In Bezug auf leistungsbezogene Eingangsmerkmale könnte vermutet werden, dass die eingangs erwähnten negativen Alltagstheorien – unabhängig von ihrem Wahrheitsgehalt – dazu beitragen, dass das Lehramt als vermeintlich „leichte Alternative“ zu einem anderen universitären Studium wahrgenommen wird und in der Folge Studienberechtigte anzieht, die sich ein wissenschaftliches Fachstudium nicht zutrauen (vgl. auch Spinath et al. 2005, S. 187). Dabei scheint

die Differenzierung zwischen dem Lehramt für Gymnasien und anderen Lehramtsstudiengängen bedeutsam, denn das Gymnasiallehramtsstudium wird im Vergleich als anspruchsvoller und wissenschaftlicher wahrgenommen (Hitpass und Mock 1972).

Zu dieser Einschätzung passt der Befund von Pohlmann und Möller (2010), wonach Studierende des Grund- oder Sek.-I-Lehramts häufiger als Gymnasiallehramtsstudierende angeben, das Studium gewählt zu haben, weil es leichter ist als andere Studiengänge. Empirisch zeigen zwei Studien, dass kognitive Eingangsmerkmale (operationalisiert über Abiturnoten, Intelligenz- und Wissenstests) von Gymnasiallehramtsstudierenden positiver und in etwa vergleichbar mit Diplomstudierenden an Universitäten sind, während die übrigen Lehramtsstudierenden weniger gute Abiturnoten und Testleistungen aufweisen (Gold und Giesen 1993; Klusmann et al. 2009). Diese Befunde stehen teilweise im Widerspruch zu den Ergebnissen von Spinath et al. (2005), die die Gruppe der nicht-gymnasialen Lehramtsstudierenden differenzierter analysieren. Spinath et al. (2005) schlussfolgern, dass Studierende mit dem Berufsziel Grund- oder Sonderschullehrer ähnliche Lern- und Leistungsvoraussetzungen aufweisen wie Studierende des Gymnasiallehramts. Allerdings werden in der Studie nur etwa 50 Studierende pro Gruppe betrachtet, die zudem an einer einzigen Universität rekrutiert wurden.

2.2.2 Motivationale Eingangsmerkmale

Der Lehrerberuf ist eine in hohem Maße soziale Tätigkeit, bei der zwischenmenschliche Interaktion und Wissensvermittlung im Vordergrund stehen. Entsprechend lässt sich in Bezug auf motivationale Merkmale ableiten, dass der Beruf für Personen mit ausgeprägten sozialen und pädagogischen Interessen besonders attraktiv sein sollte. Bei ausgeprägten fachlichen oder wissenschaftlichen Interessen sollten sich Studienberechtigte hingegen eher für ein reines fachwissenschaftliches Studium entscheiden, bei dem die pädagogischen Studienanteile zugunsten fachlicher Anteile wegfallen.³ Auch hier scheint die Differenzierung unterschiedlicher Lehrämter bedeutsam. So ist der fachliche Anteil bei Gymnasiallehrern höher, während an anderen Schulformen den pädagogischen Anteilen ein stärkeres Gewicht zukommt. Entsprechend wäre zu erwarten, dass Studierende des Lehramts Gymnasium höhere fachliche und wissenschaftliche Interessen und Studierende in anderen Lehramtsstudiengängen höhere soziale oder

³ Persönlichkeitsfacetten nach der Theorie Hollands (1997) (bspw. eine investigative Orientierung) werden üblicherweise über Interesseninventare erfasst (bspw.: Es interessiert mich „gar nicht – sehr stark“, mich „mit unerforschten Dingen zu beschäftigen“). Sie sollten aber ebenso in den hier verwendeten Studienmotiven und Berufs-/Lebenszielen zum Ausdruck kommen (bspw.: Bei der Studienwahl ist das Motiv „wissenschaftliches Arbeiten“ für mich „bedeutungslos – sehr bedeutend“).

pädagogische Interessen mitbringen. Alle Lehrämter zeichnen sich auch dank des öffentlichen Dienstes durch eine große Arbeitsplatzsicherheit bei relativ günstiger Einkommenssituation aus (Weishaupt und Huth 2012). Zudem bietet der Lehrerberuf aufgrund der Urlaubs- und Arbeitszeiten wie kaum ein anderer die Möglichkeit, familiäre und berufliche Interessen zu verbinden. Folglich sollte der Lehrerberuf besonders attraktiv für Personen sein, denen berufliche Sicherheit wichtig ist und die besonderen Wert auf eine gute Vereinbarkeit von Familie und Beruf legen. Keine attraktive Alternative sollte der Lehrerberuf hingegen für Personen darstellen, die beruflich aufsteigen wollen und die für höhere Anstrengungen entsprechend entlohnt werden wollen: Die Karrieremöglichkeiten sind eingeschränkt, zudem führen höhere Anstrengungen, anders als in der freien Wirtschaft, nicht unmittelbar zu finanziellen Verbesserungen oder beruflichen Aufstiegen. Im Gegensatz zu Karrierezielen sind Ziele, die sich darauf beziehen, die eigenen Kompetenzen zu erweitern, im Lehrerberuf durchaus zu verwirklichen. Um in ihrem Beruf erfolgreich zu sein, müssen sich Lehrkräfte ebenso wie andere Akademiker beständig weiterbilden und ihre berufsbezogenen Kompetenzen erweitern (vgl. Nitsche et al. 2013). Entsprechend ist eher nicht davon auszugehen, dass angehende Lehrkräfte andere Kompetenzentwicklungsziele aufweisen als entsprechende Vergleichsgruppen.

Empirisch wurden Studienmotive und berufliche Interessen von angehenden Lehrkräften häufig untersucht (im Überblick: Rothland 2011a). Dabei hat sich die hohe Bedeutsamkeit von pädagogischen und sozialen Motivationen ausnahmslos bestätigt. Als weitere wichtige Gründe werden fachliches Interesse oder eine gute Übereinstimmung mit eigenen Fähigkeiten genannt (vgl. Pohlmann und Möller 2010). Extrinsische Aspekte, wie die Vereinbarkeit von Familie und Beruf oder die berufliche Sicherheit, werden eher nachrangig genannt. Bestätigt hat sich die Vermutung, dass pädagogische Motive für angehende Grund- und Sonderschullehrer wichtiger sind als für angehende Gymnasiallehrer, die wiederum häufiger fachliche Motive nennen (vgl. Pohlmann und Möller 2010; Retelsdorf und Möller 2012; Ulich 1998). Die Aussagekraft dieser Studien ist allerdings begrenzt. Zum einen werden in der Regel bereits im Studium immatrikulierte Studierende befragt, sodass Tendenzen zum entscheidungskonformen Antworten nicht ausgeschlossen werden können. Zum anderen beschränken sich sämtliche Studien auf Lehramtsstudierende; durch das Fehlen von Vergleichsgruppen ist es kaum möglich, die Eigenständigkeit bestimmter Merkmale und Motive bei der Wahl des Lehramtsstudiums zu identifizieren. So sollte etwa eine gute Übereinstimmung mit den eigenen Fähigkeiten für alle Studienanfänger gleichermaßen bedeutsam sein; schließlich ist nach Holland (1997) davon auszugehen, dass Studienberechtigte stets solche Studienumwelten anstreben, die zu den eigenen Fähigkeiten passen. Hinsichtlich langfristiger, handlungssteuernder Berufs- und Lebensziele

von Lehrkräften gibt es kaum empirische Studien. Eine Ausnahme ist die Untersuchung von Abele (2011), in der Personen verglichen werden, die Mathematik entweder mit dem Abschlussziel Diplom oder dem Abschlussziel Gymnasiallehreramt studiert haben. Die Ergebnisse deuten unter anderem darauf hin, dass Lehramtsstudierende nach ihrem Studium seltener Karriereziele verfolgen und seltener „schwierige und herausfordernde Aufgaben bearbeiten“ wollen. Dagegen geben sie häufiger an, eine Arbeit anzustreben, die „gut mit familiären/privaten Bindungen vereinbar ist“ (Abele 2011, S. 680 f.). Letzteres Ziel ist allerdings nur für Frauen mit Staatsexamen signifikant ausgeprägter. Inwiefern sich die berichteten Unterschiede auch zu Beginn des Studiums oder für nicht-gymnasiale Lehrämter zeigen, ist eine offene Frage.

2.2.3 Soziodemographische Eingangsmerkmale

Aus den oben angestellten Überlegungen lassen sich auch Unterschiede hinsichtlich soziodemographischer Personenmerkmale ableiten. Da pädagogische Interessen ebenso wie familiäre Vereinbarkeitsziele bei Frauen vermutlich stärker ausgeprägt sind, sollte der Lehrerberuf für sie besonders attraktiv sein. Insbesondere die Schulstufen mit den höheren Pädagogikanteilen sollten demnach von Frauen präferiert werden. Zudem sind im Staatsdienst die in vielen Berufen virulenten geschlechtlichen Einkommensunterschiede nach Kontrolle der Teilzeitarbeit aufgehoben (z. B. Europäische Kommission 2011), woraus sich ein zusätzlicher Anreiz für Frauen ableitet. Mit Blick auf die soziale Herkunft kann vermutet werden, dass der Lehrerberuf ein sozialer Aufstiegsberuf ist. Kaum ein anderer akademischer Beruf dürfte Studienberechtigten aus nicht-akademischen Elternhäusern (aus Schülersicht) so vertraut sein wie der des Lehrers. Die negativere Selbsteinschätzung der bildungsfernen Schichten, einen akademischen Beruf und das dafür notwendige Studium erfolgreich bewältigen zu können (vgl. Becker und Hecken 2008), könnte durch die eigenen Erfahrungen als Schüler bei der Bewertung des Lehrerberufs aufgehoben sein. Die in Aussicht stehende berufliche Sicherheit dürfte einen zusätzlichen Anreiz für Studienberechtigte aus bildungsfernen Schichten darstellen, die in Ermangelung elterlicher Ressourcen eher auf stabile Beschäftigungsverhältnisse angewiesen sind.

Zahlen zur Geschlechterverteilung im Lehrerberuf bestätigen dessen besondere Attraktivität für Frauen. Sie zeigen darüber hinaus, dass der Lehrerinnenanteil unter den hauptberuflichen Lehrkräften in der Grundschule mit 88 % am höchsten und im Gymnasium mit 55 % am niedrigsten ausfällt (Statistisches Bundesamt 2011, S. 135). Empirische Studien, die die soziale Herkunft von Lehrern oder Lehramtsstudierenden vergleichend analysieren, gibt es kaum; entsprechend dünn ist die Evidenz. Mehrheitlich deutet sich jedoch an, dass die soziale

Herkunft von nicht-gymnasialen Lehrern tendenziell weniger exklusiv ist als die von Personen in anderen Studiengängen bzw. akademischen Berufen (vgl. Klusmann et al. 2009; Rothland 2011b).

2.2.4 Zusammenfassende Hypothesen

Zusammenfassend leiten sich aus der bisherigen Diskussion Hypothesen hinsichtlich leistungsbezogener, motivationaler und soziodemographischer Unterschiede zwischen a) Lehramtsstudierenden und anderen Universitätsstudierenden sowie b) zwischen verschiedenen Arten des Lehramtes ab. Die erwarteten Gruppenunterschiede sind in der folgenden Darstellung schematisch zusammengefasst. Da die Hypothesen in der Regel nur zwischen Gymnasiallehramtsstudierenden und anderen Lehramtsstudierenden insgesamt unterscheiden, werden diese hier als LA RGHS („Lehramt Real-, Grund-, Haupt- und Sonderschule“) zusammengefasst.

Bevor allerdings die in Tabelle 1 skizzierten Gruppenunterschiede geprüft werden, soll zunächst untersucht werden, ob die Zusammenfassung der nicht-gymnasialen Lehrämter zu einer Gruppe gerechtfertigt ist. *Inwieweit unterscheiden sich die nicht-gymnasialen Lehramtsstudierenden bei differenzierter Betrachtung?* Bislang ist diese Frage nicht eindeutig zu beantworten. Das liegt in erster Linie daran, dass bisherige Studien nichtgymnasiale Lehramtsstudierende a priori zusammenfassen – in der Regel weil die Stichprobengrößen für differenziertere Analysen nicht ausreichen (z. B. Klusmann et al. 2009). Die Studien, die einen differenzierteren Vergleich vornehmen, kommen zu uneinheitlichen Schlüssen. Auf der einen Seite weisen Spinath et al. (2005) auf Unterschiede zwischen Grund- und Sonderschullehramtsstudierenden einerseits sowie Sek-I-Lehramtsstudierenden andererseits hin. Allerdings lassen sich diese Unterschiede bis auf ein Merkmal statistisch nicht absichern. Auf der anderen Seite berichten Retelsdorf und Möller (2012, S. 9, Tabelle 1) zwar von zum Teil signifikanten Unterschieden hinsichtlich leistungsbezogener und motivationaler Merkmale, die sie aufgrund ihrer im Vergleich zu Spinath et al. (2005) deutlich größeren Stichprobe ausmachen können. Allerdings sind die Unterschiede nach der Effektstärkenklassifikation von Cohen (1988) als eher klein zu bezeichnen.

Tabelle 1: Schematische Darstellung der vermuteten Gruppenunterschiede

| | |
|---|------------------------|
| <i>Leistungsbezogene Eingangsmerkmale</i> | Uni = LA Gym > LA RGHS |
| <i>Motivationale Eingangsmerkmale</i> | |
| soziales Interesse | LA RGHS > LA Gym > Uni |
| fachliches Interesse | Uni > LA Gym > LA RGHS |
| wissenschaftliches Interesse | Uni > LA Gym > LA RGHS |
| berufliche Sicherheit | LA Gym = LA RGHS > Uni |
| Fähigkeitsüberzeugung | LA Gym = LA RGHS = Uni |
| private/familiäre Vereinbarkeitsziele | LA Gym = LA RGHS > Uni |
| Karriereziele | Uni > LA Gym = LA RGHS |
| Kompetenzziele | Uni = LA Gym = LA RGHS |
| <i>Soziodemographische Eingangsmerkmale</i> | |
| weiblich | LA RGHS > LA Gym > Uni |
| bildungsferne Herkunftsfamilie | LA RGHS = LA Gym > Uni |

2.3 Methoden

2.3.1 Stichprobe

Zur Überprüfung der Hypothesen wird auf die vom HIS-Institut für Hochschulforschung durchgeführten Studienberechtigtenbefragungen der Schulentlassjahrgänge 1999 und 2002 zurückgegriffen (Heine et al. 2004). Im Rahmen dieser deutschlandweiten Erhebung wurden die nachschulischen Werdegänge von Hochschulzugangsberechtigten untersucht. Die beiden Jahrgänge (1999 und 2002) wurden für die Analysen gepoolt, um eine möglichst große Stichprobe zur Verfügung zu haben. Die postalischen Befragungen fanden jeweils ein ½ Jahr und 3½ Jahre nach Erwerb der Hochschulzugangsberechtigung statt. Zum ersten Erhebungszeitpunkt beteiligten sich in beiden Jahren zusammengenommen 20337 Personen mit allgemeiner Hochschulzugangsberechtigung. In die Analysen gingen die Daten aller 14815 Personen ein, die zum ersten Erhebungszeitpunkt (½ Jahr nach dem Abitur) angegeben hatten, nach einer Übergangszeit (z. B. Wehrdienst) ein Studium aufnehmen zu wollen, bzw. bereits ein Studium aufgenommen hatten. Zur besseren Einordnung der Stichprobe sei darauf verwiesen, dass ein Vergleich zentraler Randverteilungen mit der amtlichen Statistik weitgehende Ähnlichkeiten aufzeigt: Für ein Lehramtsstudium entscheiden sich 12.6 % (13.5 % laut amtlichen Daten), 46.0 % aller Studienanfänger sind männlich (46.6 % laut amtlichen Daten, jeweils Durchschnittswerte der Jahre 1999 und 2002).

Wie alle postalischen Befragungen weisen auch die HIS-Befragungen niedrige Rücklaufquoten auf. Nach Angaben von HIS (persönliche Korrespondenz) betrug die Rücklaufquote

(jeweils für die erste Welle) 1999 26 % und 2002 22 %.⁴ Da es keine Anhaltspunkte gibt, dass die Ausschöpfung systematisch zwischen den verschiedenen Studienberechtigten Gruppen variiert, ist der Einfluss auf die vergleichende Analyse der Eingangsmerkmale trotz niedriger Quote als gering einzuschätzen. Als Sensitivitätskontrolle für mögliche Verzerrungen durch Panelmortalität wurden die Analysen auf die Personen beschränkt, die auch am zweiten Messzeitpunkt (3½ Jahre nach dem Abitur) teilnahmen und zu diesem Zeitpunkt ein Studium aufgenommen hatten (63 % des Analysesamples).⁵ Da sich die Befunde gleichen, werden im Folgenden lediglich die Analysen mit Informationen des ersten Messzeitpunktes berichtet.

2.3.2 Studienentscheidung

Die verwendete deutschlandweite Zufallsstichprobe ermöglicht es, angehende Lehramtsstudierende mit Personen zu vergleichen, die diese Wahl auch hätten treffen können, sich aber für ein anderes Studium entschieden haben. Um Studierende möglichst differenziert zu betrachten und gleichzeitig der bundesweiten Vielfalt der Lehramtsstudiengänge Rechnung zu tragen (so gibt es in einigen Bundesländern zusammengefasste Studiengänge für das Grund- und Hauptschullehramt), wird unterschieden zwischen Lehramtsstudierenden für das Gymnasium, für das Grund- und Hauptschullehramt, für die Realschule sowie für Berufs- und Sonderschulen (eine weitere Differenzierung ist aufgrund der Antwortskala nicht möglich). Als primäre Vergleichsgruppe dienen Studierende an Universitäten (Dipl., Mag., Staatsexamen ohne Lehramt, Bachelor). Um alle Studienanfängergruppen vollständig abzubilden, werden als weitere Vergleichsgruppe Studierende an anderen tertiären Einrichtungen (Fachhochschulen, Berufsakademien, Verwaltungsfachhochschulen) berücksichtigt. Zur Fokussierung von Hypothesen und Ergebnisdarstellung wird in diesem Beitrag aber vor allem auf den Vergleich der verschiedenen Lehramtsstudierenden mit den Studierenden an Universitäten eingegangen.

⁴ Die bereinigte Rücklaufquote liegt höher, da zum einen nicht alle Personen, welche den Fragebogen erhalten haben, Teil der Grundgesamtheit waren. Zum anderen handelt es sich teilweise um stichprobenneutrale Ausfälle, die auf die durch Umzüge geprägte Lebensphase und damit zusammenhängende schwere Erreichbarkeit der Studienberechtigten zurückzuführen ist. Ausgeschlossen sind Personen, die zum ersten Erhebungszeitpunkt über 30 Jahre alt waren sowie Personen, die nur über eine Fachhochschulreife/fachgebundene (Fach-)Hochschulreife verfügten.

⁵ Die Sensitivitätskontrollen sind im Appendix, Tabelle A1 abgebildet.

2.3.3 Eingangsmerkmale

Zur Erfassung der leistungsbezogenen Eingangsmerkmale steht die *Abiturgesamtnote* (1 = sehr gut bis 4 = ausreichend) zur Verfügung ($M = 2.29$, $SD = 0.63$). Zwar sind Abiturnoten über Länder hinweg schwer vergleichbar (Köller et al. 2004), sie stellen jedoch ein kompaktes Maß dar, in dem sich Merkmale wie allgemeine kognitive Fähigkeiten, Wissen sowie Leistungsmotivation konzentrieren (Mayr 2010, S. 76). Folglich erweisen sie sich immer wieder als valider und starker Prädiktor für den Studien- und Berufserfolg (z. B. Trapmann et al. 2007) und als erklärungskräftigster Leistungsindikator bei der Studienwahl (Klusmann et al. 2009).

Motivationale Eingangsmerkmale werden über zwei Skalen erfasst. Zum einen wurden Studienberechtigte nach ihren langfristigen Berufs- und Lebenszielen gefragt, zum anderen nach ihren Studienwahlmotiven.

Berufs- und Lebensziele beschreiben, welche langfristigen Ziele Personen situationsübergreifend bevorzugt verfolgen. Neben einer motivationalen Qualität haben Ziele eine volitionale Komponente (die Bereitschaft, die gesetzten Ziele auch zu erreichen), was sie bedeutsam für die Handlungssteuerung und damit auch für die Vorhersage von Studienerfolg macht (Abele 2011; Kunter 2011). Zielorientierungstheorien unterscheiden grob zwischen Lernzielen, die darauf abzielen, eigene Kompetenzen zu erweitern, und Leistungszielen, die darauf abzielen, anderen gegenüber hohe Kompetenzen zu demonstrieren.⁶ In Anlehnung werden in den HIS-Studien „Kompetenzziele“ und „Karriereziele“ unterschieden. Kompetenzziele beziehen sich darauf, neue Kompetenzen zu erwerben und das Leistungsvermögen auszuschöpfen. Karriereziele beziehen sich dagegen auf den Zweck, eine leitende Position zu erlangen und viel Geld zu verdienen. Da eine Person jedoch nicht nur berufliche Ziele hat, werden zusätzlich „private Ziele“ unterschieden, die sich auf familiäre und Freizeitaktivitäten richten (vgl. auch Abele 2011). Erfasst wurden die Ziele mittels einer aus 15 Items bestehenden Skala, auf der die Studienberechtigten angeben sollten, wie stark sie verschiedene Berufs- und Lebensziele verfolgen. Das Antwortformat reichte von 1 („sehr stark“) bis 5 („überhaupt nicht“). Eine Hauptkomponentenanalyse mit orthogonaler Rotation wies nach dem Ausschluss von Items mit Nebenladungen > 0.40 auf die Drei-Faktoren-Lösung hin. Die drei Dimensionen sind anhand

⁶ Kompetenzdemonstration kann dabei nochmals unterteilt werden in „Kompetenz demonstrieren“ und „Inkompetenz verbergen“.

eines Beispielitems in Tabelle 2 dargestellt. Sie erklären 65 % der Varianz der Einzelitems und weisen bis auf die Dimension „Private Ziele“ eine hinreichende interne Konsistenz auf.⁷

Tabelle 2: Berufs- und Lebensziele

| Faktor | Beispielitem | Items | α |
|--|--|-------|----------|
| | „Mir geht es darum, ...“ | | |
| Kompetenzziele (Fokus: beruflich etwas leisten) | „fundierte, ausbaufähige Kompetenzen zu erwerben“ | 3 | 0.61 |
| Karriereziele (Fokus: Status demonstrieren) | „eine leitende Funktion einzunehmen“ | 5 | 0.87 |
| Private Ziele (Fokus: Freizeit haben) | „das Leben zu genießen und genügend Freizeit zu haben“ | 2 | 0.55 |

Studienwahlmotive. Auch wenn bislang empirische Belege rar sind, gelten Studienwahlmotive als wichtige Prädiktoren für Studienzufriedenheit und -abbruch (Heublein et al. 2010; Küsting und Lipowsky 2011). In jüngerer Zeit wurden auf Basis des Erwartungs-Wert-Modells von Eccles Instrumente zur Erfassung der Motivation für die Wahl eines Lehramtsstudiums (FEMOLA) entwickelt (Pohlmann und Möller 2010; Watt und Richardson 2007). Diese sind jedoch nicht ideal, um Studienmotive *vergleichend mit anderen Studienanfängern* zu analysieren. Dafür fehlen wichtige Dimensionen wie beispielsweise Statusmotive, die für viele Studieninteressierte relevant sind (Heine et al. 2004). Zudem sind die Items in FEMOLA so formuliert, dass sie nur von Lehramtsstudierenden sinnvoll beantwortet werden können. Aus diesem Grund wird hier der Ansatz verfolgt, die grundlegenden in FEMOLA entwickelten Motivkomponenten aufzugreifen und um weitere Dimensionen zu ergänzen, die sich in unselegierten Studienberechtigtenstichproben als bedeutsam erwiesen haben. Anders als in FEMOLA sind die Items so formuliert, dass sie auch von Nicht-Lehramtsstudierenden sinnvoll beantwortet werden können. So gibt es etwa für die Dimension „Fähigkeitsüberzeugung“ anstelle des FEMOLA-Items „Ich denke, dass ich eine gute Lehrerin/ein guter Lehrer sein werde“ das fachunspezifische Item „gute Übereinstimmung mit eigener Leistungsfähigkeit“. Studienwahlmotive wurden über eine Liste aus 25 Items erfasst. Die Liste war mit der Frage eingeleitet „Welche Bedeutung haben die folgenden Gründe und Motive für den von Ihnen gewählten nachschulischen Werdegang?“. Es standen sechs Antwortalternativen zur Verfügung (1 = bedeutungslos bis 6 = sehr bedeutend). Dabei kann zwischen verschiedenen intrinsischen wertbezogenen Motivationen, extrinsischen wertbezogenen Motivationen und erwartungsbezogenen Motivationen unterschieden werden (vgl. Tabelle 3).

⁷ Sensitivitätsanalysen, bei denen die beiden Items der Dimension als Einzelitems einfließen, führten zu den gleichen Befunden.

Tabelle 3: Studienwahlmotive

| Motivationskomponente | Item | Items | α |
|-------------------------------------|---|-------|----------|
| | „Welche Bedeutung haben die folgenden Gründe und Motive für den von Ihnen gewählten nachschulischen Werdegang?“ | | |
| <i>Wertkomponente (intrinsisch)</i> | | | |
| - soziales Interesse | „anderen Menschen helfen“ | 2 | 0.82 |
| - fachliches Interesse | „Interesse am vermittelten Sachwissen“ | 1 | - |
| - wissenschaftl. Interesse | „Interesse an wissenschaftlicher Arbeit“ | 1 | - |
| <i>Wertkomponente (extrinsisch)</i> | | | |
| - berufliche Sicherheit | „Wunsch nach sicherer beruflicher Zukunft“ | 1 | - |
| - Familie | „familiäre Gründe“ | 1 | - |
| - Karriere/Status | „in leitende Position gelangen“ | 2 | 0.82 |
| - soziale Einflüsse | „Rat von Verwandten, Freunden, Bekannten“ | 1 | - |
| <i>Erwartungskomponente</i> | | | |
| - Fähigkeitsüberzeugung | „gute Übereinstimmung mit eigener Leistungsfähigkeit“ | 1 | - |

Obwohl in Bezug auf erwartungsbezogene Motivationen sowie soziale Einflüsse („Rat von Verwandten, Freunden, Bekannten“) theoretisch keine Gruppenunterschiede erwartet werden, werden diese der Vollständigkeit halber mit aufgenommen. Für die Erwartungskomponente steht dabei lediglich der Indikator „Fähigkeitsüberzeugung“ zur Verfügung; die in FEMOLA zusätzlich eingesetzte Dimension „geringe Schwierigkeit des gewählten Studiums“ kann mit den vorliegenden Daten nicht abgebildet werden. Anstelle der FEMOLA-Dimension „pädagogisches Interesse“ wird nach „sozialem Interesse“ gefragt. Diese Dimension ist natürlich nicht gleichbedeutend, sie hat aber den Vorteil, dass sie lehramtsunspezifischer und damit besser für einen Vergleich mit Nicht-Lehramtsstudierenden geeignet ist. Die extrinsische Motivkomponente wird differenziert abgebildet, indem zwischen Sicherheitsmotiven und familiären Nützlichkeitsaspekten unterschieden wird. Ergänzt werden die Motive aus den genannten Gründen um die intrinsische Dimension „Interesse an wissenschaftlicher Arbeit“ und die extrinsische Dimension „Karriere/Status“. Es sei darauf hingewiesen, dass sich die genannten Motivdimensionen leider größtenteils nur mit Einzelitems abbilden lassen (da die HIS-Studien eine Vielzahl von Themen im Blick haben, ist der Einsatz von Verfahren ausgeschlossen, in denen latente Dimensionen mit umfangreichen Itembatterien abgebildet werden). In Tabelle 4 sind die bivariaten Korrelationskoeffizienten der leistungsbezogenen und motivationalen Prädiktoren dargestellt. Es zeigen sich durchweg erwartbare Zusammenhänge (z. B. $r = 0.27$ zwischen dem Motiv „fachliches Interesse“ und dem Berufsziel „fundierte, ausbaufähige Kompetenzen zu erwerben“), was das Vertrauen in die Validität der verwendeten Prädiktoren stärkt.

Für die Analysen wurden die motivationalen Prädiktoren z-standardisiert ($M = 0$, $SD = 1$). Als soziodemographische Merkmale werden das *Geschlecht* sowie die *soziale Herkunft* der

Studienberechtigten berücksichtigt. Die soziale Herkunft wird erfasst über den höchsten Bildungsabschluss der Eltern (höchstens Hauptschulabschluss, Realschulabschluss/Abitur, Hochschulabschluss).

Tabelle 4: Bivariate Korrelationskoeffizienten (Pearson) der Eingangsmerkmale

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. |
|----------------------|--------|--------|--------|--------|-------|--------|-------|-------|--------|-----|-----|-----|
| 1. Abiturnote | 1 | | | | | | | | | | | |
| 2. soziales Int. | 0.03* | 1 | | | | | | | | | | |
| 3. fachliches Int. | -0.16* | 0.03* | 1 | | | | | | | | | |
| 4. wiss. Int. | -0.20* | | 0.28* | 1 | | | | | | | | |
| 5. ber. Sicherheit | 0.08* | 0.03* | | 0.02* | 1 | | | | | | | |
| 6. Familie | 0.10* | 0.18* | -0.12* | -0.04* | 0.28* | 1 | | | | | | |
| 7. Karriere/Status | | | 0.02 | 0.10* | 0.43* | 0.12* | 1 | | | | | |
| 8. Fähigkeitsüberz. | -0.12* | 0.09* | 0.23* | 0.13* | 0.11* | 0.03* | 0.09* | 1 | | | | |
| 9. soziale Einflüsse | | 0.07* | -0.07* | -0.04* | 0.17* | 0.29* | 0.11* | 0.04* | 1 | | | |
| 10. Komp. ziele | -0.20* | 0.09* | 0.27* | 0.23* | 0.02 | -0.06* | 0.11* | 0.30* | -0.03* | 1 | | |
| 11. Karriereziele | 0.05* | -0.14* | -0.04* | 0.05* | 0.41* | 0.11* | 0.77* | 0.06* | 0.11* | | 1 | |
| 12. Private Ziele | 0.02* | 0.09* | -0.02* | -0.03 | 0.16* | 0.22* | | 0.03* | 0.16* | | | 1 |

Dargestellt sind nur Koeffizienten, die auf dem 5%-Niveau bzw. 1%-Niveau (mit * versehen) signifikant sind.

2.3.4 Statistische Analysen

Das erste Forschungsanliegen besteht darin, explorativ zu prüfen, ob sich die nicht-gymnasialen Lehrämter (Studienanfänger mit dem Berufsziel Real-, Grund-/Haupt- sowie Sonderschullehramt) hinsichtlich ihrer Eingangsmerkmale unterscheiden. Dazu werden zunächst nur diese drei Gruppen betrachtet. Für jedes Eingangsmerkmal wurde eine einfaktorielle Varianzanalyse durchgeführt. Angesichts der großen Stichprobe sowie der Vielzahl an durchgeführten Paarvergleichen wurde für alle bivariaten Auswertungen die Irrtumswahrscheinlichkeit auf $\alpha = 0.01$ festgesetzt. Um der Kumulierung des α -Fehlers aufgrund der multiplen Paarvergleiche entgegenzuwirken, wurde die gewählte Irrtumswahrscheinlichkeit von 1% pro Merkmal Bonferroni-adjustiert.

Das zweite Forschungsanliegen ist der Vergleich der gymnasialen mit der Gruppe der nicht-gymnasialen Lehramtsstudierenden sowie der Vergleich dieser Gruppen mit den Nicht-Lehramtsstudierenden. Im Gegensatz zum ersten Forschungsanliegen wurden für diese Vergleiche ordinale Unterschiedshypothesen entwickelt (siehe Tabelle 1). Anstatt die vermuteten Gruppenunterschiede indirekt durch herkömmliche Varianzanalysen zu prüfen, folgte ich dem Vorschlag von Furr und Rosenthal (2003) und teste die vermuteten Unterschiede direkt mithilfe

geplanter Kontraste. Diese Sonderform der Varianzanalyse erlaubt es, Alternativhypothesen präziser zu prüfen, zudem werden vorhandene Unterschiede mit einer größeren Wahrscheinlichkeit aufgedeckt. Die vermuteten Unterschiede wurden je Eingangsmerkmal paarweise geprüft (LA Gym vs. LA RGHS, LA Gym vs. Uni, LA RGHS vs. Uni). Der Gruppe FH wurde dabei jeweils das Kontrastgewicht 0 zugewiesen, weil für diese Gruppe keine expliziten Hypothesen hergeleitet wurden. Für die Merkmale, für die kein Gruppenunterschied postuliert wurde („Fähigkeitsüberzeugung“, „Rat von Anderen“ und „Kompetenzziele“), wurden Varianzanalysen ohne a priori Kontraste berechnet. Um zu prüfen, ob die Gruppenunterschiede bei simultaner Betrachtung bestehen bleiben und um die relative Wichtigkeit der Prädiktoren abzuschätzen, wurden zudem Personenunterschiede zwischen den möglichen Studiengängen mittels multinomialer logistischer Regressionen analysiert. Da die Studienoptionen zwischen den Bundesländern variieren (z. B. durch die Lehramtsausbildung an Pädagogischen Hochschulen in Baden-Württemberg), ist es möglich, dass ein unbeobachteter Einfluss des Wohnortes auf die Studienentscheidung besteht und sich die Studienberechtigten hinsichtlich ihrer Studienwahl innerhalb eines Landes ähnlicher sind als die Studienberechtigten zwischen den Ländern. Um diesem Umstand Rechnung zu tragen, wurden robuste Standardfehler unter Berücksichtigung der Bundeslandcluster geschätzt.

Mit Blick auf die Item-Nonresponse hat sich die Einsicht durchgesetzt, dass die Schätzung fehlender Werte durch multiple Imputationen gegenüber dem paar- oder fallweisen Ausschluss vorzuziehen ist (Allison 2001). Um die Unsicherheit aufgrund fehlender Werte zu berücksichtigen, wurden mithilfe des Stata Programms ICE (Royston 2004) für jeden fehlenden Wert fünf plausible Werte erzeugt, die in den anschließenden Analysen nach dem von Rubin (1987) vorgeschlagenen Verfahren kombiniert wurden. In allen Analysen wurden vom HIS entwickelte Gewichte verwendet, um die Repräsentativität im Hinblick auf bekannte Merkmale der Hochschulzugangsberechtigten (Geschlecht, Art der Schule, Art der Hochschulreife und Bundesland) zu maximieren.

2.4 Ergebnisse

2.4.1 Unterschiede zwischen den nicht-gymnasialen Lehrämtern

Zunächst widmen wir uns der Frage, inwiefern sich die drei Gruppen nicht-gymnasialer Lehramtsstudierender unterscheiden. In Tabelle 5 sind neben den Gruppenmittelwerten der Eingangsmerkmale die Ergebnisse der Varianzanalysen dargestellt.

Tabelle 5: Ergebnisse der Varianzanalysen sowie Gruppenmittel-/Anteilswerte der Eingangsmerkmale (Standardfehler in Klammern) von nicht-gymnasialen Lehramtsstudierenden

| | LA Real | LA GH | LA Sonder | ANOVA | |
|----------------------------------|---------------------------|-----------------------------|-----------------------------|---------|--------------|
| | n=215 | n=587 | n=226 | F | η^2 |
| Abiturnote | 2.68 _a (0.03) | 2.63 _{a,b} (0.02) | 2.52 _b (0.03) | 5.86** | 0.01 |
| Studienwahlmotive | | | | | |
| <i>Wertkomponente (intr.)</i> | | | | | |
| - soziales Interesse | 0.68 _a (0.05) | 0.81 _a (0.03) | 1.05 (0.04) | 18.58** | 0.03 |
| - fachliches Interesse | -0.29 _a (0.08) | -0.39 _a (0.05) | -0.25 _a (0.07) | 1.67 | 0.00 |
| - wiss. Interesse | -0.46 _a (0.06) | -0.74 _b (0.04) | -0.70 _{a,b} (0.06) | 6.99** | 0.02 |
| <i>Wertkomponente (extr.)</i> | | | | | |
| - berufliche Sicherheit | 0.18 _a (0.06) | -0.00 _{a,b} (0.04) | -0.09 _b (0.06) | 5.33* | 0.01 |
| - Familie | 0.34 _a (0.07) | 0.43 _a (0.04) | 0.18 _a (0.07) | 4.05 | 0.01 |
| - Karriere/Status | -0.46 _a (0.06) | -0.70 _b (0.04) | -0.71 _{a,b} (0.06) | 5.51* | 0.01 |
| - Rat von Anderen | 0.10 _a (0.07) | 0.18 _a (0.04) | 0.20 _a (0.07) | 0.74 | 0.00 |
| <i>Fähigkeitsüberzeugung</i> | 0.02 _a (0.07) | -0.05 _a (0.04) | -0.01 _a (0.06) | 0.43 | 0.00 |
| Berufs-/Lebensziele | | | | | |
| <i>Kompetenzziele</i> | -0.32 _a (0.07) | -0.51 _a (0.04) | -0.32 _a (0.07) | 4.38 | 0.01 |
| <i>Karriereziele</i> | -0.48 (0.06) | -0.79 _a (0.04) | -0.75 _a (0.06) | 9.74** | 0.02 |
| <i>Private Ziele</i> | 0.19 _a (0.06) | 0.18 _a (0.04) | 0.17 _a (0.06) | 0.01 | 0.00 |
| Soziodemographie | | | | Chi2 | Cramers V |
| <i>weiblich (in %)</i> | 81 (0.02) | 92 (0.01) | 85 (0.02) | 24.22** | 0.15 |
| <i>Bildung der Eltern (in %)</i> | | | | 7.46 | 0.06 |
| - Hochschulabschluss | 39 (0.03) | 49 (0.02) | 49 (0.03) | | |
| - Abiur/Realschulabschl. | 45 (0.03) | 37 (0.02) | 38 (0.03) | | |
| - Hauptschul-/kein Aschl. | 16 (0.02) | 14 (0.01) | 13 (0.02) | | |

* $p < 0.01$, ** $p < 0.001$; gemittelt über 5 multiple imputierte Datensätze. Mittelwerte mit gleichen Subskripts unterscheiden sich in Paarvergleichen nicht signifikant voneinander ($\alpha' = 0.01/3 = 0.003$, Bonferroni-adjustiert).

Zunächst zeigt ein Blick auf die F-Werte (Chi2-Werte bei nominalskalierten Variablen), dass sich die Mittelwerte der drei Gruppen ‚insgesamt‘ bei sieben von 14 Merkmalen nicht signifikant (1 %-Niveau) unterscheiden. Auf den übrigen sieben Merkmalen ist der Anteil der aufgeklärten Varianz jeweils klein ($\eta^2 \leq 0.03$). Eine Durchsicht der Paarvergleiche zeigt darüber hinaus, dass sich bei diesen Merkmalen mindestens zwei der drei Gruppen nicht signifikant

voneinander unterscheiden. Trotz der großen Ähnlichkeiten zwischen diesen Lehramtsstudierenden zeigen sich aber auch signifikante Unterschiede. Studierende des Sonderschullehramtes weisen bessere Abiturnoten auf als Realschullehramtsstudierende, allerdings handelt es sich nach der Effektstärken-Klassifikation von Cohen (1988) um eine eher kleine Differenz ($d = 0.31$). Darüber hinaus weisen Studierende des Sonderschullehramtes ein höheres soziales Interesse auf als die übrigen nicht-gymnasialen Lehramtsstudierenden ($d = 0.46$). Studierende des Realschullehramtes haben dagegen höhere Karriereziele ($d = 0.26$). Auch das wissenschaftliche Interesse ist ihnen etwas wichtiger als den anderen nicht-gymnasialen Lehrämtern. Der Frauenanteil ist im Grund- und Hauptschullehramtsstudium (92 %) etwas höher als im Sonder- (85 %) und im Realschullehramtsstudium (81 %). Der Anteil an Studienanfängern aus Akademikerelternhäusern ist mit 39 % im Realschullehramt am geringsten, die Differenz zu den anderen nicht-gymnasialen Lehrämtern lässt sich aber nicht zufallskritisch absichern. In der Summe überwiegen, trotz vereinzelter Unterschiede, die Gemeinsamkeiten zwischen den nicht-gymnasialen Lehrämtern.

2.4.2 Überprüfung der vermuteten Gruppenunterschiede

Als nächstes fokussieren wir uns auf den Vergleich der gymnasialen mit den nicht-gymnasialen Lehramtsstudierenden sowie den Vergleich dieser Gruppen mit den Nicht-Lehramtsstudierenden. Aus Gründen der Übersichtlichkeit werden die Lehramtsstudierenden der Realschule, der Grund- und Hauptschule sowie der Sonderschule zusammengefasst (im Folgenden: RGHS-Lehramtsstudierende). Dieses Vorgehen erscheint zulässig, da sich die Gruppen – wie im vorangegangenen Abschnitt berichtet – gleichen bzw. nur leicht unterscheiden.

2.4.2.1 Leistungsbezogene Eingangsmerkmale

In Tabelle 6 sind die leistungsbezogenen Eingangsmerkmale, operationalisiert über die Abiturdurchschnittsnoten, dargestellt.

Die besten Abiturnoten haben Nicht-Lehramtsstudierende an Universitäten (2.21) und Studierende des Gymnasiallehramts (2.25), die sich, wie vermutet, nicht signifikant voneinander unterscheiden ($t(14811) = 1.77, p = 0.08$). Nichtlehramtsstudierende an sonstigen Hochschulen (FH/BA/VerwFH) erzielen hingegen schlechtere Abiturnoten (2.42). Die schlechtesten Noten weisen Studienanfänger auf, die sich für das Real-, Sonder- oder Grund/Hauptschullehramt entscheiden (zusammengenommen 2.62). Der Unterschied zwischen

dieser Gruppe und den Gymnasiallehramtsstudierenden erweist sich als mittlerer bis starker Effekt ($t(14811) = 12.65, p = 0.00, d = 0.59$), ebenso wie der Unterschied zur Vergleichsgruppe Uni ($t(14811) = 19.83, p = 0.00, d = 0.65$).

Tabelle 6: Abiturdurchschnittsnoten (Standardfehler in Klammern) nach gewähltem Studiengang

| | LA Gym <i>n</i> = 843 | LA RGHS <i>n</i> = 1028 | Uni <i>n</i> = 9813 | FH <i>n</i> = 3131 | Kontrast | |
|------------------------------|--------------------------|----------------------------|------------------------|-----------------------|------------------------------|----------|
| | | | | | <i>t</i> _{contrast} | <i>d</i> |
| Abiturnote | 2.25 (0.02) | 2.62 (0.02) | 2.21 (0.01) | 2.42 (0.01) | | |
| K: LA Gym > LA RGHS | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | $\lambda = 0$ | 12.64** | 0.59 |
| K: Uni = LA Gym ^a | $\lambda = 1$ | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 0$ | 1.77 | 0.06 |
| K: Uni > LA RGHS | $\lambda = 0$ | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | 19.83** | 0.65 |

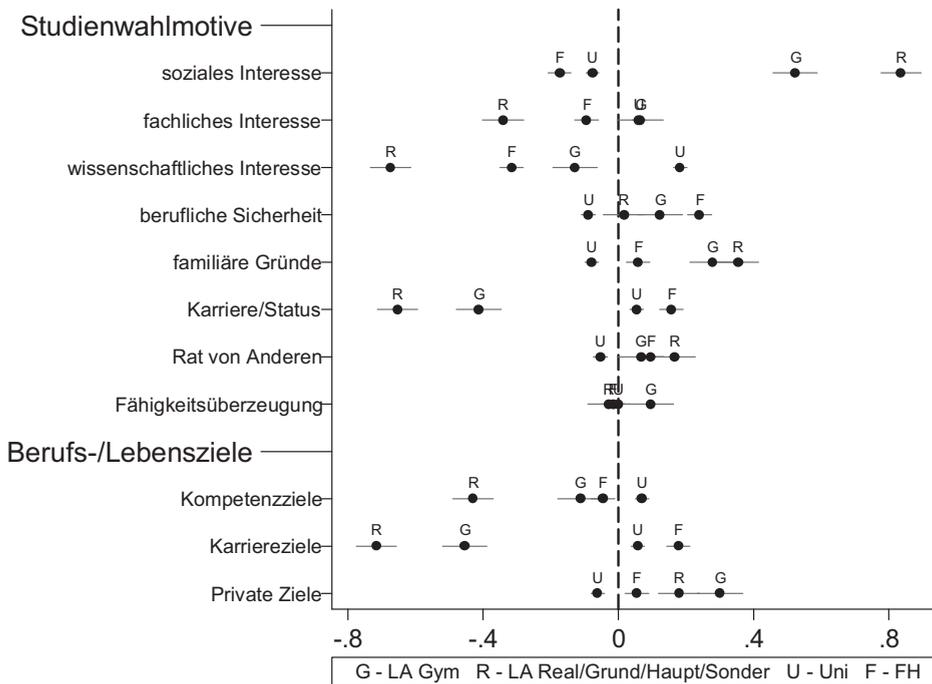
* $p < 0.01$, ** $p < 0.001$; gemittelt über 5 multiple imputierte Datensätze. λ = Kontrastgewichte. Für die geplanten paarweisen Kontraste wird als Effektstärkemaß Cohen's *d* berichtet. ^a Wenn inhaltlich ‚kein Unterschied‘ postuliert wird, wird der Kontrast analytisch als Unterschiedshypothese modelliert.

2.4.2.2 Motivationale Eingangsmerkmale

Um sich bei der Vielzahl an Vergleichen ein Gesamtbild der wesentlichen Befunde machen zu können, werden diese graphisch veranschaulicht. Abbildung 1 zeigt die z-standardisierten Gruppenmittelwerte und 95 % Konfidenzintervalle für die verschiedenen Merkmale. Die gestrichelte Vertikallinie zeigt die Stichprobenmittel an, die Einheit der x-Achse sind Standardabweichungen. Theoretisch wurde angenommen, dass RGHS-Lehramtsstudierende das größte soziale Interesse aufweisen sollten, gefolgt von Gymnasiallehramtsstudierenden und anderen Uni-Studierenden. Dieses Muster bestätigt sich in den Daten: Die Gruppe RGHS liegt 0.84 Standardabweichungen und die Gruppe Gym immerhin 0.52 Standardabweichungen über dem Stichprobenmittel. Alle drei Kontraste sind signifikant, sowohl zwischen LA Gym und LA RGHS ($t(14811) = 6.96, p = 0.00, d = 0.32$), zwischen LA Gym und Uni ($t(14.811) = 17.19, p = 0.00, d = 0.62$) als auch zwischen LA RGHS und Uni ($t(14.811) = 28.85, p = 0.00, d = 0.94$). Bei der Differenz zwischen RGHS und der Vergleichsgruppe Uni handelt es sich um einen starken Effekt ($d = 0.94$), noch größer ist die Distanz der Lehramter zur Vergleichsgruppe FH. Starke Unterschiede gibt es darüber hinaus – wie erwartet – bei den Motiven „wissenschaftliches Interesse“ sowie „Karriere/Status“, die von Lehramtsstudierenden insgesamt jeweils deutlich seltener angegeben werden; innerhalb der Lehramtsstudierenden am seltensten von den RGHS-Lehramtsstudierenden.⁸

⁸ Aus Platzgründen werden nicht alle Varianz- und Kontrastanalysen hier berichtet. Sie stehen jedoch im Appendix, Tabelle A2 zur Verfügung.

Abbildung 1: Mittelwerte und 95% Konfidenzintervalle der Studienwahlmotive und Berufs-/Lebensziele nach gewählttem Studiengang



Die Vermutung, dass Studienberechtigte mit ausgeprägtem fachlichen Interesse ein universitäres Fachstudium über ein Gymnasiallehramtsstudium präferieren sollten, bestätigt sich nicht ($t(14.811) = -0.16, p = 0.88$). Dagegen zeigen sich signifikante Unterschiede zur Gruppe RGHS (Kontrast Uni > LA RGHS $t(14.811) = 12.19, p = 0.00$), die dieses Motiv deutlich seltener ($d = 0.40$) nennen. Lehramtsstudierende insgesamt wählen ihr Studium erwartungsgemäß häufiger aus „familiären Gründen“ und aus „Gründen der beruflichen Sicherheit“. Der Vergleichsgruppe FH ist die berufliche Sicherheit allerdings noch wichtiger als den Lehramtsstudierenden. Erwartungsgemäß weitgehend unabhängig vom gewählten Studiengang scheint die Motivdimension „Fähigkeitsüberzeugung“ zu sein (,overall‘ ANOVA $F(3;14811) = 3.01, p = 0.03, \eta^2 = 0.00$). Ein signifikanter Unterschied findet sich im Vergleich der Gruppen Gym-Uni, allerdings handelt es sich um einen kleinen Effekt ($t(14811) = 2.63, p = 0.01, d = 0.09$). Ebenfalls gering (jedoch signifikant) sind die Differenzen bei der Nennung des Motivs „auf Rat von Verwandten, Freunden, Bekannten“ ($F(3;14811) = 29.27, p = 0.00, \eta^2 = 0.01$). Die Lehramtsstudierenden insgesamt sowie die Gruppe FH haben dieses Motiv etwas häufiger angegeben als die Gruppe Uni.

Hinsichtlich der langfristigen *Berufs- und Lebensziele* unterscheiden sich die Gruppen entgegen der Vermutung bei der Dimension Kompetenzziele signifikant voneinander ($F(3;14811) = 86.01, p = 0.00, \eta^2 = 0.02$). So geben Studierende in der Vergleichsgruppe Uni überdurchschnittlich stark an, langfristig fundierte Kompetenzen erwerben zu wollen. Weniger ausgeprägt ist dieses Ziel bei Gymnasiallehramtsstudierenden (und FH Studierenden). Am wenigsten stark werden Kompetenzziele von RGHS-Lehramtsstudierenden verfolgt, die fast eine halbe Standardabweichung unter dem Stichprobenmittel liegen (Kontrast Uni > RGHS $t(14811) = 15.32, p = 0.00, d = 0.50$). Wie theoretisch vermutet, sind Karriereziele bei Lehramtsstudierenden insgesamt deutlich geringer ausgeprägt als bei den Vergleichsgruppen. Die Gruppe Gym liegt hier fast eine halbe Standardabweichung unter dem Stichprobenmittel, die Gruppe RGHS sogar 0.72 Standardabweichungen (Kontrast Uni > RGHS $t(14811) = 24.25, p = 0.00, d = 0.80$). Hingegen sind ihnen private Ziele („das Leben genießen und genügend Freizeit haben“; „mich intensiv um Familie und Partnerschaft kümmern“) wichtiger als anderen Studierenden.

2.4.2.3 Soziodemographische Eingangsmerkmale

In Tabelle 7 sind die soziodemographischen Gruppenunterschiede dargestellt. Die Geschlechterkomposition entspricht dem bekannten Muster und unterscheidet sich signifikant zwischen den Studiengängen ($\text{Chi}^2(3) = 562.92, p = 0.00$).

Tabelle 7: Geschlecht und soziale Herkunft (Standardfehler in Klammern) nach gewähltem Studiengang

| | LA Gym | LA RGHS | Uni | FH | Kontrast |
|----------------------------------|----------------|----------------|----------------|---------------|----------|
| <i>weiblich (in %)</i> | 62 (0.02) | 88 (0.02) | 51 (0.00) | 50 (0.01) | z |
| K: LA RGHS > LA Gym | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | $\lambda = 0$ | 12.50** |
| K: LA Gym > Uni | $\lambda = 1$ | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 0$ | 6.30** |
| K: LA RGHS > Uni | $\lambda = 0$ | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | 19.93** |
| <i>Bildung der Eltern (in %)</i> | | | | | |
| - Hochschulabschluss | 55 (0.02) | 47 (0.02) | 62 (0.00) | 49 (0.01) | |
| K: LA Gym = LA RGHS ^a | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 3.32** |
| K: Uni > LA Gym | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 1$ | $\lambda = 0$ | 4.21** |
| K: Uni > LA RGHS | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | 9.29** |
| - Abitur/Realschulabschluss | 34 (0.02) | 39 (0.02) | 28 (0.00) | 36 (0.01) | |
| - Hauptschul-/kein Abschl. | 11 (0.01) | 14 (0.01) | 10 (0.00) | 15 (0.01) | |

* $p < 0.01$, ** $p < 0.001$; gemittelt über 5 multiple imputierte Datensätze. λ = Kontrastgewichte. Für die soziale Herkunft wurde nur der Anteil der Eltern mit Hochschulabschluss kontrastiert. ^a Wenn inhaltlich ‚kein Unterschied‘ postuliert wird, wird der Kontrast analytisch als Unterschiedshypothese modelliert

Der Frauenanteil ist in den Lehramtsstudiengängen deutlich höher als in den Vergleichsgruppen Uni und FH. Wenn sich junge Männer für den Lehrerberuf entscheiden, dann wählen sie bevorzugt das Gymnasiallehramt, wo der Männeranteil immerhin bei 38 % liegt. Hinsichtlich der sozialen Herkunft zeigen sich signifikante Unterschiede zwischen allen Gruppen ($\chi^2(10) = 222, p = 0.00$). Der Anteil der Akademikerkinder ist am höchsten bei Studierenden, die einen universitären Abschluss (außer Lehramt) anstreben (62 %), gefolgt von Gymnasiallehramts studierenden (55 %, Kontrast Uni > LA Gym $z = 4.21, p = 0.00$). In den übrigen Lehramtsstudiengängen (47 %) sowie an Fach- und sonstigen Hochschulen (49 %) studieren weniger Personen aus Akademikerfamilien (Kontrast LA Gym > LA RGHS $z = 3.32, p = 0.00$). Dafür ist der Anteil der Studierenden, deren Eltern höchstens über einen Hauptschulabschluss verfügen, in den Gruppen LA RGHS und FH mit 14 bzw. 15 % am höchsten.

2.4.2.4 Multivariate Analysen

Um zu prüfen, ob die Gruppenunterschiede bei simultaner Betrachtung bestehen bleiben und um die relative Wichtigkeit der Prädiktoren abzuschätzen, werden im Folgenden Personenunterschiede zwischen den möglichen Studiengängen mittels multinomialer logistischer Regressionen analysiert. Die in Tabelle 8 dargestellten Odds-Ratios beschreiben die Chance, den jeweiligen Studiengang im Vergleich zur Referenzkategorie Uni zu wählen, wenn der entsprechende Prädiktor um eine Einheit steigt (Koeffizienten kleiner 1 beschreiben eine geringere Chance und Koeffizienten größer 1 eine höhere Chance). Um aufzuzeigen, wie sich die beiden Lehramtsstudierendengruppen untereinander unterscheiden, sind in der letzten Spalte der Tabelle 8 Koeffizienten des gleichen Modells dargestellt, mit dem Unterschied, dass als Vergleichsgruppe nun die Gymnasiallehramtsstudierenden gewählt wurden.

Die Ergebnisse zeigen, dass sich Männer und Frauen unter Kontrolle der übrigen Prädiktoren bei der Wahl zwischen einem Gymnasiallehramtsstudium oder einem anderen universitären Studium nicht signifikant voneinander unterscheiden. Allerdings haben Frauen eine 3.86-fach höhere Chance als Männer, sich für ein RGHS-Lehramtsstudium statt für ein anderes universitäres Studium zu entscheiden; diese Geschlechterdifferenz wird nicht durch die Kontrolle der Studienmotive und der übrigen Prädiktoren aufgefangen. Bezüglich der sozialen Herkunft zeigt sich, dass Personen, die sich für ein Lehramtsstudium oder für ein FH-Studium entscheiden, auch bei ansonsten gleichen Charakteristika aus niedrigeren Bildungsschichten stammen als Personen, die sich für ein universitäres Studium entscheiden. Die Differenz zwischen den Gruppen LA RGHS und LA Gym lässt sich jedoch nicht mehr zufallskritisch (auf

dem 5 % Niveau) absichern. Die Unterschiede hinsichtlich der Abiturnoten, die sich bereits in den bivariaten Vergleichen zeigten, bleiben bei der simultanen Berücksichtigung aller Prädiktoren bestehen. Eine um eine Standardabweichung höhere (d. h. schlechtere) Abiturnote verdoppelt bei ansonsten gleichen Merkmalen nahezu die Chance, ein RGHS anstelle eines gymnasialen Lehramtsstudiums zu wählen.

Tabelle 8: Multinomiale logistische Regression der Studienwahl (Odds-Ratios)

| | LA Gym | | LA RGHS | | FH | | LA RGHS | |
|--------------------------------|---------|-----------|---------|-----------|---------|-----------|------------|-----------|
| | vs. Uni | | vs. Uni | | vs. Uni | | vs. LA Gym | |
| | OR | (S.E.) | OR | (S.E.) | OR | (S.E.) | OR | (S.E.) |
| Frau (Mann = Ref.) | 1.03 | (0.11) | 3.86 | (0.42)*** | 0.94 | (0.04) | 3.74 | (0.05)*** |
| Eltern: höchstens Hauptsch. | 1.36 | (0.14)** | 1.66 | (0.34)* | 1.70 | (0.15)*** | 1.22 | (0.20) |
| Eltern: Abitur/Realschule | 1.39 | (0.15)** | 1.66 | (0.19)*** | 1.50 | (0.09)*** | 1.19 | (0.18) |
| Eltern: Hochschule (Ref.) | | | | | | | | |
| Abiturnote (z-standardisiert) | 0.97 | (0.05) | 1.86 | (0.09)*** | 1.25 | (0.07)*** | 1.91 | (0.07)*** |
| <i>Studienwahlmotive</i> | | | | | | | | |
| (intr) soziales Interesse | 1.91 | (0.11)*** | 2.99 | (0.18)*** | 0.89 | (0.02)*** | 1.56 | (0.11)*** |
| (intr) fachliches Interesse | 1.10 | (0.07) | 0.88 | (0.05)* | 1.01 | (0.02) | 0.80 | (0.08)* |
| (intr) wiss. Interesse | 0.75 | (0.04)*** | 0.56 | (0.02)*** | 0.61 | (0.02)*** | 0.74 | (0.03)*** |
| (extr) berufliche Sicherheit | 1.63 | (0.08)*** | 1.66 | (0.12)*** | 1.45 | (0.07)*** | 1.02 | (0.07) |
| (extr) familiäre Gründe | 1.22 | (0.06)*** | 1.23 | (0.04)*** | 1.02 | (0.02) | 1.00 | (0.06) |
| (extr) Karriere/Status | 0.67 | (0.03)*** | 0.65 | (0.05)*** | 1.03 | (0.04) | 0.97 | (0.10) |
| (extr) Rat von Anderen | 1.02 | (0.03) | 1.19 | (0.05)*** | 1.11 | (0.03)*** | 1.16 | (0.06)*** |
| Fähigkeitsüberzeugung | 1.14 | (0.05)** | 1.13 | (0.04)*** | 1.06 | (0.03)* | 0.99 | (0.05) |
| <i>Berufs- und Lebensziele</i> | | | | | | | | |
| Kompetenzziele | 0.77 | (0.04)*** | 0.69 | (0.03)*** | 1.01 | (0.03) | 0.89 | (0.07) |
| Karriereziele | 0.67 | (0.05)*** | 0.54 | (0.05)*** | 0.91 | (0.04) | 0.81 | (0.09) |
| Private Ziele | 1.31 | (0.04)*** | 1.17 | (0.05)*** | 1.05 | (0.02)* | 0.89 | (0.04)* |
| Nagelkerke R ² | | | | | 0.31 | | | |
| N | | | | | 14815 | | | |

* p<0.05, ** p<0.01, *** p<0.001; gemittelt über 5 multiple imputierte Datensätze, robuste Standardfehler. Referenzkategorie: UNI. In der letzten Spalte bildet LA Gym die Referenzkategorie.

Hinsichtlich der Studienmotive sowie der Berufs- und Lebensziele decken sich die Regressionsergebnisse ebenfalls weitgehend mit den bivariaten Befunden: Die Prädiktoren behalten bei gleichzeitiger Berücksichtigung ihre Erklärungskraft und auch die relativen Effektstärken entsprechen dem Muster der Mittelwertdifferenzen (da die metrischen Variablen z-standardisiert sind, ist es möglich die Koeffizienten innerhalb des Modells miteinander zu vergleichen). Im Unterschied zu den bivariaten Befunden unterscheiden sich LA RGHS und LA Gym Studierende bei gemeinsamer Betrachtung aller Prädiktoren nicht mehr signifikant bei den Studienwahlmotiven „Karriere/Status“ sowie den „Kompetenz- und „Karrierezielen“.

2.5 Diskussion

Im vorliegenden Beitrag wurde zunächst theoretisch hergeleitet, warum sich ein Studienberechtigter für ein Lehramtsstudium entscheiden sollte – und was das für Konsequenzen für die Eingangsmerkmale von Lehramtsstudierenden hat. Anschließend wurden die leistungsbezogenen, motivationalen und soziodemographischen Eingangsmerkmale verschiedener Lehramtsstudierender im Vergleich zu anderen Studierenden empirisch untersucht.

2.5.1 Eingangsmerkmale von Lehramtsstudierenden im Vergleich

Hinsichtlich leistungsbezogener Eingangsmerkmale kann die Studie die regional beschränkten bzw. älteren Befunde von Klusmann et al. (2009) sowie Gold und Giesen (1993) anhand einer deutschlandweiten Stichprobe bestätigen. Während es keine generelle Negativselektion im Hinblick auf Abiturleistungen von Lehramtsstudierenden gibt, zeigt sich eine negative Selektion zu Ungunsten der nicht-gymnasialen Lehramter, welche das Gros der Lehramtsstudierenden ausmachen. Verglichen mit den genannten Studien konnten nicht-gymnasiale Lehramtsstudierende differenzierter nach angestrebter Schulform betrachtet werden. Dabei wurde deutlich, dass die Unterschiede in den Abiturnoten von Real-, Sonder-, sowie Grund- und Hauptschullehramtsstudierenden nicht signifikant sind bzw. nach der Effektstärkenklassifikation von Cohen als eher klein zu bezeichnen sind. Nicht bestätigt werden konnte daher die Vermutung von Spinath et al. (2005), wonach Primar- und Sonderschullehramtsstudierende ähnliche Lern- und Leistungsvoraussetzungen aufweisen wie Studierende mit dem Berufsziel Gymnasiallehrer – zumindest hinsichtlich der hier analysierten Abiturnoten. Eher scheint es eine Notenabstufung in der Form Gym < Sonder < Grund/Haupt < Real zu geben. Allerdings mussten bei der vorliegenden Befragung das Grund- und Hauptschullehramt sowie das Sonder- und Berufsschullehramt zusammengefasst werden, was eine trennscharfe Prüfung der Vermutung von Spinath et al. (2005) verhindert. Es ist durchaus möglich, dass die Hauptschullehramtsstudierenden den Gruppendurchschnitt der Grund- und Hauptschullehramtsstudierenden drücken. Darauf deuten die Befunde von Retelsdorf und Möller (2012) hin, die für Gymnasiallehramtsstudierende den gleichen Abiturdurchschnittsnotenwert finden wie die vorliegende Studie (2.25), während sie für die (isolierte) Gruppe der Grundschullehramtsstudierenden eine mittlere Abiturnote von 2.5 ausweisen.

Ergänzend zum bisherigen Wissensstand konnten erstmals Studienwahlmotive und Berufs-/ Lebensziele von Lehramtsstudierenden denen externer Vergleichsgruppen

gegenübergestellt werden. Erst dieser Vergleich ermöglicht es, abzuschätzen, welche Motive und Ziele charakteristisch für das Lehramt sind und welche von Lehramtsstudierenden häufig genannten nur ein generelles Antwortverhalten von Studienanfängern widerspiegeln. Zunächst zeigte der differenziertere Blick auf die unterschiedlichen nicht-gymnasialen Lehrämter, dass die Unterschiede nicht signifikant sind bzw. nach der Effektstärkenklassifikation von Cohen als eher klein zu bezeichnen sind. Zum gleichen Ergebnis kommen (implizit) Retelsdorf und Möller (2012, S. 9, Tabelle 1). Fasst man die nicht-gymnasialen Lehrämter daher zusammen und vergleicht die Ergebnisse mit den vermuteten Mustern aus Abschn. 2.2.4 (vgl. Tabelle 9), zeigt sich, dass sich die Hypothesen weitgehend bestätigen lassen: Ein ausgeprägtes soziales Interesse, hohe familiäre Interessen, der Wunsch nach beruflicher Sicherheit sowie geringe Karriere- oder Statusmotive sind charakteristisch für Lehramtsstudierende im Vergleich zu anderen Studierenden. Passend dazu liegen die langfristigen Berufs- und Lebensziele eher im privaten (familiären und freizeitlichen) Bereich. Ziele, die sich darauf beziehen, beruflich viel zu leisten oder Status zu demonstrieren, sind bei Lehramtsstudierenden insgesamt hingegen unterdurchschnittlich stark ausgeprägt. Die Gruppe der nicht-gymnasialen Lehramtsstudierenden zeichnet sich zudem durch besonders hohe soziale Interessen und besonders schlechte Abiturnoten aus. Darüber hinaus sind das geringere fachliche und wissenschaftliche Interesse charakteristische Merkmale; Kompetenz- und Karriereziele sind niedriger als bei allen anderen Gruppen.

Nicht antizipiert wurde die Tatsache, dass Gymnasiallehramtsstudierende sich im Hinblick auf fachliches Interesse nicht von anderen Uni-Studierenden unterscheiden. Keine oder nur geringe Unterschiede zeigen sich beim Motiv „Fähigkeitsüberzeugung“. Die starke Bedeutung der „Fähigkeitsüberzeugung“ als Motivation bei Lehramtsstudierenden (vgl. z. B. Pohlmann und Möller 2010) spiegelt demnach eher ein generelles Antwortverhalten aller Studienanfänger wider und ist kein charakteristisches Motiv von Lehramtsstudierenden. Hinsichtlich soziodemographischer Merkmale bestätigt sich die ausgeprägte Attraktivität von – insbesondere nicht-gymnasialen – Lehramtsstudiengängen für Frauen. Lehramtsstudiengänge aber auch Studiengänge an Fach- oder anderen nicht-universitären Hochschulen ziehen zudem vermehrt Studienberechtigte aus bildungsfernen sozialen Schichten an. Es zeigt sich somit, dass die Lehramtsstudiengänge durchaus soziale Aufstiegskanäle darstellen; insbesondere gilt das für nicht-gymnasiale Lehramtsstudiengänge.

Tabelle 9: Schematische Darstellung der Ergebnisse

| | | |
|---|------------------------|---------------|
| <i>Leistungsbezogene Eingangsmerkmale</i> | Uni = LA Gym > LA RGHS | (beibehalten) |
| <i>Motivationale Eingangsmerkmale</i> | | |
| soziales Interesse | LA RGHS > LA Gym > Uni | (beibehalten) |
| fachliches Interesse | Uni = LA Gym > LA RGHS | (teilweise) |
| wissenschaftliches Interesse | Uni > LA Gym > LA RGHS | (beibehalten) |
| berufliche Sicherheit | LA Gym = LA RGHS > Uni | (beibehalten) |
| Fähigkeitsüberzeugung | LA Gym = LA RGHS = Uni | (beibehalten) |
| private/famil. Vereinbarkeitsziele | LA Gym = LA RGHS > Uni | (beibehalten) |
| Karriereziele ^a | Uni > LA Gym ≥ LA RGHS | (teilweise) |
| Kompetenzziele ^a | Uni > LA Gym ≥ LA RGHS | (nein) |
| <i>Soziodemographische Eingangsmerkmale</i> | | |
| weiblich | LA RGHS > LA Gym > Uni | (beibehalten) |
| bildungsferne Herkunftsfamilie ^a | LA RGHS > LA Gym > Uni | (teilweise) |

Grau unterlegte Zeichen markieren Abweichungen von den postulierten Unterschieden (vgl. Tabelle 1). ^a Bei diesen Merkmalen konnte die Differenz LA RGHS – LA Gym bei gleichzeitiger Berücksichtigung aller Merkmale (Mlogit) nicht zufallskritisch abgesichert werden.

2.5.2 Einschränkungen der Studie

Zwar erweitert die Studie durch ihre differenzierten Vergleiche, ihre deutschlandweit generalisierbaren Aussagen und die Analyse von Studienwahlmotiven sowie Berufs- und Lebenszielen die bisherige Befundlage. Allerdings ist zu berücksichtigen, dass die Aussagekraft limitiert ist durch den Kranz an zur Verfügung stehenden Variablen sowie deren Messqualität (latente motivationale Merkmale lassen sich mit den HIS-Daten in der Regel nur mit Einzelitems abbilden). Darüber hinaus stellt die niedrige Rücklaufquote ein nicht zu unterschätzendes Problem dar. Da es keine Hinweise dafür gibt, dass die Ausfälle systematisch zwischen den Wahlalternativen der abhängigen Variable variieren, hält sich die Verzerrung im Hinblick auf Gruppenvergleiche aber in Grenzen. Mit den vorliegenden Daten ist es nicht möglich zu prüfen, welche der Studierenden das Studium auch tatsächlich beenden und in den Lehrerberuf einmünden. So könnte vermutet werden, dass Differenzen zwischen Lehramtsstudierenden und anderen Studierenden durch selektives Wechselverhalten im Laufe der Zeit verschwinden. Dagegen sprechen allerdings Hinweise, wonach Studienabbrecher und Berufswechsler des Lehramtes im Vergleich zu den verbliebenen Personen höhere Intelligenzwerte und bessere Noten aufweisen (Guarino et al. 2006, S. 186). Es wäre sicherlich wünschenswert, mit aktuelleren Daten zu prüfen, inwiefern die Befunde auch für Studierende in reformierten (konsekutiven) Studiengängen gelten. Obwohl direkte Kohortenvergleiche fehlen, lassen Ergebnisse auf Basis späterer Kohorten

vermuten, dass sowohl der Wunsch Lehrer zu werden als auch die Motivstruktur von Lehramtsstudierenden weitgehend gleich geblieben sind (Bauer et al. 2011; Rothland 2011a), sodass die Befunde auch für aktuellere Studienberechtigte gelten sollten.

2.5.3 Bedeutung der Befunde

Was bedeuten nun die Befunde und kann man tatsächlich von einer Negativselektion in das Lehramtsstudium sprechen? Zunächst ist festzustellen, dass einige der hier analysierten Eingangsmerkmale weder etwas Positives noch etwas Negatives über die Frage der Eignung als Lehrer aussagen. Zudem muss konstatiert werden, dass der kausale Zusammenhang zwischen den Eingangsmerkmalen der Lehramtsstudierenden, der Ausbildungsqualität, der Qualität des Unterrichtens – und schließlich – den Auswirkungen auf Lernprozesse der Schüler indirekt und komplex ist (vgl. Terhart 2008); man wird daher kaum direkte Effekte der Eingangsmerkmale auf Lernprozesse der Schüler nachweisen können. Zerlegt man jedoch die komplexe Kausalkette und betrachtet weniger distale Erfolgskriterien – etwa das im Studium angeeignete Wissen – erweisen sich einige der hier betrachteten leistungsbezogenen und motivationalen Eingangsmerkmale durchaus als bedeutsam. Das gilt insbesondere für Abiturleistungen, für fachlich-intrinsische Studienwahlmotive sowie damit korrelierende Kompetenzziele. Dahinter steckt die Annahme, dass der Kompetenzerwerb angehender Lehrkräfte nicht alleine von der Ausbildungsqualität abhängt, sondern auch entscheidend davon, wie Studierende vorhandene Lerngelegenheiten auf Basis ihrer individuellen Dispositionen nutzen. *Erstens* erleichtern größere Lern- und Leistungsfähigkeiten die Einpassung neuen Wissens (z. B. Klauer und Leutner 2010). *Zweitens* hängt das Ausmaß an investierter Energie für das Studium – und damit der Wissenszuwachs – auch damit zusammen, welche Motivation eine Person zur Erreichung eines Ziels hat (vgl. Kunter 2011). *Drittens* ist das Ziel, auf das sich die Investitionen richten, entscheidend: Wer Kompetenzen erweitern will, wird eher in den Erwerb von berufsrelevantem Wissen investieren als jemand, der Aufwand vermeiden will oder der in erster Linie außerberufliche Lebensziele verfolgt (vgl. Malmberg 2006). Auch wenn (vor allem) weitere längsschnittliche Forschung notwendig ist, die den Einfluss bestimmter Eingangsmerkmale auf langfristigen Studien- und Berufserfolg von Lehrern misst, stützt die bislang verfügbare Evidenz die vermuteten Zusammenhänge. Gute Abiturleistungen erweisen sich immer wieder als starker Prädiktor für den Studienerfolg und berufsrelevantes Wissen, sowohl fachübergreifend (z. B. Trapmann et al. 2007) als auch lehramtsspezifisch (Blömeke 2009; Blömeke und Buchholtz 2011; Mayr 2010). Hinsichtlich der Studienwahlmotive zeigt sich, dass intrinsische – vor

allem fachlich-intrinsische – Studienwahlmotivationen positiv mit Lernstrategien, Studienzufriedenheit (Künsting und Lipowsky 2011), Examensnoten (Blömeke 2009; Mayr 2009) sowie fach- und fachdidaktischem Wissen (Blömeke et al. 2012) zusammenhängen. Negative Zusammenhänge zeigen sich hingegen mit dem Belastungserleben und der Abbruchintention im Studium (Blömeke 2009; Heublein et al. 2010). Hinsichtlich langfristiger Berufs- und Lebensziele ist der empirische Kenntnisstand unbefriedigend. Es gibt jedoch Hinweise, dass hohe private Ziele (die sich auf familiäre Vereinbarkeit beziehen) mit einer niedrigeren Arbeitszufriedenheit und höherem Belastungserleben von Lehrern nach zehnjähriger Berufspraxis einhergehen (Abele 2011). Betrachtet man nun im nächsten Schritt der Kausalkette den Einfluss der angeeigneten fach- und fachdidaktischen Kompetenz (Baumert et al. 2010) oder motivationaler Lehrermerkmale (intrinsische Motivation, Enthusiasmus beim Unterrichten) (Kunter 2011) auf die Unterrichtsqualität bzw. die Lernentwicklung der Schüler, zeigen sich wiederum positive Zusammenhänge.

Vor diesem Hintergrund erscheinen die Eingangsmerkmale der Gymnasiallehramtsstudierenden überwiegend wünschenswert, wenn auch die ausgeprägten privaten Ziele und niedrigen Kompetenzziele den Gesamteindruck trüben. Ein problematischeres Profil weisen die RGHS-Lehramtsstudierenden auf, deren unterdurchschnittliche Abiturnoten, fachliche Studienwahlmotive und Kompetenzziele bedenklich sind. Die vorgestellten Analysen legen nahe, die bisherige Rekrutierungspraxis in die Lehrerbildung zu überprüfen (zur Debatte um Rekrutierung und Eignungsabklärung angehender Lehrkräfte vgl. Rothland und Terhart 2011). Bislang ist der Zugang zum Lehramtsstudium recht offen geregelt. Problematisch an dieser Regelung ist die Tatsache, dass Lehramtsanwärter oft erst während oder nach der universitären Ausbildungsphase eine Rückmeldung über ihre (Nicht-)Eignung erhalten. Eine umfassende Rückmeldung über die persönliche Eignung hinsichtlich des angestrebten Berufs *vor* Studienbeginn – etwa in Form eines mehrwöchigen begleiteten Pflichtpraktikums – wäre sicherlich sinnvoll. Obwohl sich hier in den letzten Jahren viel getan hat, gilt für die Mehrzahl der Lehramtsstudierenden, dass eine Rückmeldung vor Studienbeginn nicht stattfindet. In keinem Bundesland ist ein Orientierungspraktikum verpflichtende Zulassungsvoraussetzung zum Studium (siehe Appendix, Tabelle A3). Inwieweit eine solche Rückmeldung durch formale Zulassungsbeschränkungen für Personen mit ungeeigneten Eingangsmerkmalen ergänzt werden soll und kann, sollte zumindest weiter diskutiert werden. Wenn sich – wie der vorliegende Beitrag nahelegt – bereits vor Studienbeginn Eingangsmerkmale identifizieren lassen, die nicht wünschenswert sind, wäre es dann nicht sinnvoll, den Zugang für Personen mit diesen Merkmalen stärker zu beschränken? Angesichts eines sich abzeichnenden Lehrermangels stellt sich

dabei die doppelte Herausforderung, qualitativ *und* quantitativ in ausreichendem Maße geeignete Bewerber zu rekrutieren. Die Sorge, dass Zulassungsbeschränkungen die Anzahl der Lehramtsstudierenden weiter reduziert (Rothland und Terhart 2011, S. 638), ist berechtigt. Zwar kann eine selektivere Studierendenauswahl auch eine positive Signalwirkung auf Studienberechtigte mit wünschenswerten Lernvoraussetzungen haben, die sich andernfalls z. B. für ein Medizin- oder Psychologiestudium entscheiden. Dennoch scheint mir eine Strategie, die ausschließlich auf frühzeitige Eignungsrückmeldung und Zulassungsbeschränkung abzielt, zu kurz zu greifen. Ergänzt werden müsste sie um eine Qualitätsinitiative mit dem Ziel, mehr geeignete Bewerber zu rekrutieren. Wie das gelingen soll, ist nicht leicht zu beantworten.

Die vorliegende Studie macht deutlich, dass bei der Debatte um Eignungsprüfungen insbesondere auf die nicht-gymnasialen Lehrämter geachtet werden sollte. Die ungleiche Selbstselektion in die verschiedenen Lehramtsstudiengänge ist in ihren langfristigen Folgen möglicherweise auch für die Genese von Bildungsungleichheiten bedeutsam. So deutet sich eine zusätzliche Benachteiligung der Kinder auf niedrigeren Schulstufen an, nicht nur weil die Ausbildung für diese Lehrämter weniger Kompetenzen vermittelt (Baumert et al. 2010), sondern auch weil die Studierenden die vorhandenen Lerngelegenheiten aufgrund ihrer unvorteilhaften Eingangsmerkmale womöglich nicht optimal nutzen.

Danksagung und Hinweis

Ich danke Oliver Dickhäuser sowie zwei anonymen Gutachtern für wertvolle Kommentare. Finanzielle Unterstützung erfolgte durch das BMBF im Rahmenprogramm für empirische Bildungsforschung (FKZ: 01JG1058).

2.6 Appendix

Als Sensitivitätskontrolle für mögliche Verzerrungen durch Panelmortalität wurden zunächst Analysen gerechnet, in denen alle Personen berücksichtigt wurden, die zum ersten Erhebungszeitpunkt (1/2 Jahr nach dem Abitur) angegeben hatten, nach einer Übergangszeit (z. B. Wehrdienst) ein Studium aufnehmen zu wollen, bzw. bereits ein Studium aufgenommen hatten (Samplezuschnitt 1). Zum Vergleich wurden daran anschließend die gleichen Analysen gerechnet mit Personen, die auch am zweiten Messzeitpunkt teilnahmen und zu diesem Zeitpunkt ein Studium aufgenommen hatten (Samplezuschnitt 2). Multinomiale logistische Analysen auf die Studienentscheidung mit beiden Samplezuschnitten sind in Tabelle II abgebildet. Als weitere Kontrolle wurde ein dritter Samplezuschnitt gewählt. Dort wurden ebenfalls Informationen zur Studienentscheidung aus der zweiten Welle verwendet. Darüber hinaus wurden aber, falls Personen sich nicht an der 2. Erhebung beteiligten, jedoch bereits zum Zeitpunkt der 1. Erhebung (im Dezember nach dem Erhalt des Abiturs) ein Studium aufgenommen hatten, die Information der 1. Welle herangezogen (Samplezuschnitt 3). Dadurch wurden einige zusätzliche Fälle gewonnen. Zur Überprüfung, ob dieser Samplezuschnitt zu inhaltlich anderen Ergebnissen führt, wurde die aktuelle Modellspezifikation ebenfalls für dieses Sample gerechnet. Es zeigt sich, dass sich die Befunde über die drei Samplezuschnitte hinweg gleichen. Insofern kann getrost auf die Informationen des ersten Messzeitpunktes zurückgegriffen werden. Die Verwendung der Informationen des zweiten Messzeitpunktes reduziert das Analysesample und führt damit zu ineffizienteren Schätzern, ändert aber nichts an den inhaltlichen Ergebnissen.

Tabelle A1: Multinomiale logistische Regression der Studienwahl (Odds Ratios) für Samplezuschnitt 1, (Samplezuschnitt 2 in runden Klammern) und [Samplezuschnitt 3 in eckigen Klammern].

| | LA Gym vs. <i>Uni</i> | LA RGHS vs. <i>Uni</i> | FH vs. <i>Uni</i> | LA Gym vs. <i>Uni</i> | LA RGHS vs. <i>Uni</i> | FH vs. <i>Uni</i> |
|--|-----------------------------------|-----------------------------------|-----------------------------------|------------------------------|-----------------------------------|-----------------------------------|
| Frau (Mann = Ref.) | 1.61*** (1.61***) [1.64***] | 7.87*** (7.19***) [7.39***] | 0.99 (1.09*) [1.02] | 1.03 (1.08) [1.07] | 3.86*** (3.62***) [3.84***] | 0.94 (1.05) [0.99] |
| Eltern: höchstens Haupt. | 1.29** (1.33*) [1.19] | 1.53** (1.75**) [1.67**] | 1.62*** (1.80***) [1.64***] | 1.36** (1.42*) [1.26] | 1.66* (1.96**) [1.81*] | 1.70*** (1.86***) [1.71***] |
| Eltern: Abitur/Realsch. (Hochschule = Ref.) | 1.36*** (1.34**) [1.23*] | 1.54*** (1.68***) [1.66***] | 1.48*** (1.50***) [1.51***] | 1.39** (1.35*) [1.24*] | 1.66*** (1.81***) [1.79***] | 1.50*** (1.53***) [1.53***] |
| Abiturnote | 1.06 (1.06) [1.06] | 2.05*** (2.05***) [2.05***] | 1.37*** (1.39***) [1.39***] | 0.97 (0.98) [0.99] | 1.86*** (1.85***) [1.90***] | 1.25*** (1.26***) [1.26***] |

| <i>Studienmotive</i> | | | |
|---------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| soziales Interesse | 1.91*** (1.80***) [2.00***] | 2.99*** (2.95***) [2.99***] | 0.89*** (0.91**) [0.90**] |
| fachliches Interesse | 1.10 (1.02) [1.01] | 0.88* (0.84*) [0.85*] | 1.01 (1.00) [1.02] |
| wissenschaftliches Interesse | 0.75*** (0.77***) [0.77***] | 0.56*** (0.60***) [0.59***] | 0.61*** (0.65***) [0.64***] |
| berufliche Sicherheit | 1.63*** (1.67***) [1.67***] | 1.66*** (1.68***) [1.67***] | 1.45*** (1.52***) [1.56***] |
| familiäre Gründe | 1.22*** (1.25***) [1.22***] | 1.23*** (1.28***) [1.25***] | 1.02 (1.03) [1.02] |
| Karriere/Status | 0.67*** (0.66***) [0.67***] | 0.65*** (0.73***) [0.68***] | 1.03 (1.02) [1.04] |
| Fähigkeitsüberzeugung | 1.14** (1.18***) [1.21***] | 1.13*** (1.09) [1.15***] | 1.06* (1.01) [1.04] |
| Rat von Anderen | 1.02 (1.03) [1.02] | 1.19*** (1.18*) [1.17**] | 1.11*** (1.10*) [1.10**] |
| <i>Berufs- und Lebensziele</i> | | | |
| Kompetenzziele | 0.77*** (0.81***) [0.76***] | 0.69*** (0.67***) [0.67***] | 1.01 (0.99) [0.96] |
| Karriereziele | 0.67*** (0.68***) [0.64***] | 0.54*** (0.50***) [0.51***] | 0.91 (0.90) [0.90*] |
| Private Ziele | 1.31*** (1.28***) [1.33***] | 1.17*** (1.11) [1.21***] | 1.05* (1.07**) [1.07**] |
| Nagelkerke R ² | .11 (.11) [.11] | .31 (.31) [.32] | |
| N | 14815 (9345) [11771] | 14815 (9345) [11771] | |

* p<0.05; ** p<0.01. *** p<0.001, Robuste Standardfehler, Referenzkategorie: UNI, metrische Variablen sind z-standardisiert. Gemittelt über 5 multiple imputierte Datensätze.

Tabelle A2: Ergebnisse der Varianz-/ Kontrastanalysen sowie Gruppenmittelwerte (Standardfehler in Klammern) der motivationalen Eingangsmerkmale

| | LA Gym n=843 | LA RGHS n=1028 | Uni n=9813 | FH n=3131 | F/t _{contrast} | η^2/d |
|----------------------------------|-----------------|-------------------|----------------|---------------|-------------------------|------------|
| Studienwahlmotive | | | | | | |
| <i>Wertkomponente (intr.)</i> | | | | | | |
| - soziales Interesse | 0.52 (0.03) | 0.84 (0.03) | -0.08 (0.01) | -0.17 (0.02) | 389.35** | 0.07 |
| K: LA RGHS > LA Gym | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | $\lambda = 0$ | 6.96** | 0.32 |
| K: LA Gym > Uni | $\lambda = 1$ | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 0$ | 17.19** | 0.62 |
| K: LA RGHS > Uni | $\lambda = 0$ | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | 28.85** | 0.94 |
| - fachliches Interesse | 0.07 (0.03) | -0.34 (0.03) | 0.06 (0.01) | -0.09 (0.02) | 61.98** | 0.01 |
| K: LA Gym > LA RGHS | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 8.79** | 0.41 |
| K: Uni > LA Gym | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 1$ | $\lambda = 0$ | -0.16 | |
| K: Uni > LA RGHS | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | 12.19** | 0.40 |
| - wiss. Interesse | -0.13 (0.03) | -0.67 (0.03) | 0.18 (0.01) | -0.32 (0.02) | 390.38** | 0.07 |
| K: LA Gym > LA RGHS | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 12.14** | 0.56 |
| K: Uni > LA Gym | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 1$ | $\lambda = 0$ | 8.93** | 0.32 |
| K: Uni > LA RGHS | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | 26.28** | 0.87 |
| <i>Wertkomponente (extr.)</i> | | | | | | |
| - berufliche Sicherheit | 0.12 (0.03) | 0.02 (0.03) | -0.09 (0.01) | 0.24 (0.02) | 90.27** | 0.02 |
| K: LA Gym = LA RGHS ^a | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 2.31 | |
| K: LA Gym > Uni | $\lambda = 1$ | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 0$ | 5.96** | 0.21 |
| K: LA RGHS > Uni | $\lambda = 0$ | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | 3.21* | 0.10 |
| - Familie | 0.28 (0.03) | 0.35 (0.03) | -0.08 (0.01) | 0.06 (0.02) | 87.48** | 0.02 |
| K: LA Gym = LA RGHS ^a | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | $\lambda = 0$ | 1.64 | |
| K: LA Gym > Uni | $\lambda = 1$ | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 0$ | 10.02** | 0.36 |
| K: LA RGHS > Uni | $\lambda = 0$ | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | 13.20** | 0.43 |
| - Karriere/Status | -0.41 (0.03) | -0.65 (0.03) | 0.05 (0.01) | 0.16 (0.02) | 238.45** | 0.05 |
| K: LA Gym = LA RGHS ^a | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 5.28** | 0.25 |
| K: Uni > LA Gym | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 1$ | $\lambda = 0$ | 13.26** | 0.48 |
| K: Uni > LA RGHS | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | 22.05** | 0.72 |
| - Rat von Anderen ^b | 0.07 (0.03) | 0.17 (0.03) | -0.05 (0.01) | 0.10 (0.02) | 29.27** | 0.01 |
| K: LA RGHS > LA Gym | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | $\lambda = 0$ | 2.13 | |
| K: LA Gym > Uni | $\lambda = 1$ | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 0$ | 3.35* | 0.12 |
| K: LA RGHS > Uni | $\lambda = 0$ | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | 6.69** | 0.22 |
| <i>Fähigkeitsüberzeugung</i> | | | | | | |
| K: LA Gym = LA RGHS ^a | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 2.64* | 0.12 |
| K: Uni = LA Gym ^a | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 1$ | $\lambda = 0$ | -2.63* | 0.09 |
| K: Uni = LA RGHS ^a | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | 0.85 | |
| Berufs-/Lebensziele | | | | | | |

| | | | | | | |
|----------------------------------|----------------|----------------|----------------|---------------|----------|------|
| <i>Kompetenzziele</i> | -0.11 (0.03) | -0.43 (0.03) | 0.07 (0.01) | -0.05 (0.02) | 86.01** | 0.02 |
| K: LA Gym = LA RGHS ^a | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 6.85** | 0.32 |
| K: Uni = LA Gym ^a | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 1$ | $\lambda = 0$ | 5.11** | 0.18 |
| K: Uni = LA RGHS ^a | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | 15.32** | 0.50 |
| <i>Karriereziele</i> | -0.46 (0.03) | -0.72 (0.03) | 0.06 (0.01) | 0.18 (0.02) | 293.83** | 0.06 |
| K: LA Gym = LA RGHS ^a | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 5.77** | 0.27 |
| K: Uni > LA Gym | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 1$ | $\lambda = 0$ | 14.67** | 0.53 |
| K: Uni > LA RGHS | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 1$ | $\lambda = 0$ | 24.25** | 0.80 |
| <i>Private Ziele</i> | 0.30 (0.03) | 0.18 (0.03) | -0.06 (0.01) | 0.05 (0.02) | 52.25** | 0.01 |
| K: LA Gym = LA RGHS ^a | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | $\lambda = 0$ | 2.60* | 0.12 |
| K: LA Gym > Uni | $\lambda = 1$ | $\lambda = 0$ | $\lambda = -1$ | $\lambda = 0$ | 10.10** | 0.36 |
| K: LA RGHS > Uni | $\lambda = 0$ | $\lambda = 1$ | $\lambda = -1$ | $\lambda = 0$ | 7.39** | 0.24 |

* $p < 0.01$, ** $p < 0.001$; gemittelt über 5 multiple imputierte Datensätze. λ = Kontrastgewichte. In der ersten Zeile je Merkmal ist ergänzend das Ergebnis einer ‚overall‘ ANOVA abgebildet. Die Teststatistiken sind F für die ANOVA und t für die Kontraste. Gruppenmittelwerte und Standardfehler entsprechen den graphisch veranschaulichten Werten aus Abbildung 1. Als Effektstärkemaße werden η^2 (für die ANOVAs) und Cohen's d (für die geplanten paarweisen Kontraste) berichtet. ^a Wenn inhaltlich ‚kein Unterschied‘ postuliert wird, wird der Kontrast analytisch als Unterschiedshypothese modelliert. ^b Für diese Dimension wurde keine Unterschiedshypothese entwickelt, dennoch werden aus Gründen der Konsistenz Paarvergleiche zwischen den Gruppen durchgeführt, nachdem die ANOVA insgesamt auf signifikante Unterschiede hingewiesen hat.

Tabelle A3: Regelungen der Bundesländer zum Orientierungspraktikum vor Studienbeginn

| Bundesland | Praktikum vor Studienbeginn vorgesehen ^a , Nachholfristen | Praktikum vor Studienbeginn <i>nicht</i> vorgesehen | Einführung | Länge |
|------------|---|---|--------------|----------------|
| BW | Nur gym. Lehramt, bis zum Beginn des 3. Semesters nachzuholen | | WS 2010/11 | 2 Wochen |
| BY | Spätestens vor Beginn des schulpädagogischen Blockpraktikums, d.h. möglichst nach dem ersten, spätestens nach dem 5. Semester | | WS 2003/04 | 3 bis 4 Wochen |
| BE | | X | | |
| BB | | X | | |
| HB | | X | | |
| HH | | X | | |
| HE | Zwingende Voraussetzung zum Beginn der schulpraktischen Studien (i.d.R. ab dem 2. Semester) ^b | | WS 2005/06 | mind. 4 Wochen |
| MV | | X | | |
| NI | | X | | |
| NW | Spätestens vor Aufnahme des (nach-universitären) Vorbereitungsdienstes nachzuweisen ^c | | WS 2011/2012 | 20 Tage |
| RP | | X | | |
| SL | | X | | |
| SN | | X | | |
| ST | | X | | |
| SH | | X | | |
| TH | Keine landesweite Regelung, nur an einzelnen Hochschulen (z. B. Uni Jena) | | | |

Stand September 2012. Quellen: Lehramtsprüfungsordnungen bzw. Lehrerbildungsgesetze der Länder, persönliche Korrespondenz mit den zuständigen Landesbehörden. Die Angaben beziehen sich auf die Lehramtsstudiengänge für Grund-, Haupt-, Realschule und Gymnasium, nicht auf die Lehramtsstudiengänge für Sonderpädagogik oder berufliche Schulen.

^a In keinem Bundesland ist ein Orientierungspraktikum vor Studienbeginn zwingende Zulassungsvoraussetzung.

^b dient aber nicht der Eignungsabklärung oder Auseinandersetzung mit schulischen Anforderungen, sondern dazu, Erfahrungen in außerschulischen pädagogischen Berufsfeldern zu sammeln.

^c heißt in Nordrhein-Westfalen ‚Eignungspraktikum‘.

2.7 Literatur

- Abele, Andrea E. 2011. Prädiktoren des Berufserfolgs von Lehrkräften. Befunde der Langzeitstudie MATHE. *Zeitschrift für Pädagogik* 57: 674-694.
- Allison, Paul D. 2001. *Missing Data*. Thousand Oaks, CA: Sage.
- Bauer, Johannes, Uta Diercks, Jan Retelsdorf, Tabea Sporer, Friederike Zimmermann, Olaf Köller, Jens Möller und Manfred Prenzel. 2011. Spannungsfeld Polyvalenz in der Lehrerbildung: Wie polyvalent sind Lehramtsstudiengänge und was bedeutet dies für die Berufswahlsicherheit der Studierenden? *Zeitschrift für Erziehungswissenschaft* 14: 629-649.
- Baumert, Jürgen, Mareike Kunter, Werner Blum, Martin Brunner, Thamar Voss, Alexander Jordan, Uta Klusmann, Stefan Krauss, Michael Neubrand und Yi-Miau Tsai. 2010. Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal* 47: 133-180.
- Becker, Rolf, und Anna E. Hecken. 2008. Warum werden Arbeiterkinder vom Studium an Universitäten abgelenkt? Eine empirische Überprüfung der „Ablenkungsthese“ von Müller und Pollak (2007) und ihrer Erweiterung durch Hillmert und Jacob (2003). *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 60: 7-33.
- Bergmann, Christian. 1992. Schulisch-berufliche Interessen als Determinanten der Studien- bzw. Berufswahl und-bewältigung: Eine Überprüfung des Modells von Holland. In *Interesse, Lernen, Leistung. Neuere Ansätze der pädagogisch-psychologischen Interessenforschung*, Hrsg. Andreas Krapp, Manfred Prenzel, 195-220. Münster: Aschendorff.
- Blömeke, Sigrid. 2005. Das Lehrerbild in Printmedien. *Die Deutsche Schule* 97: 24-39.
- Blömeke, Sigrid. 2009. Ausbildungs- und Berufserfolg im Lehramtsstudium im Vergleich zum Diplom-Studium – Zur prognostischen Validität kognitiver und psycho-motivationaler Auswahlkriterien. *Zeitschrift für Erziehungswissenschaft* 12: 82-110.
- Blömeke, Sigrid, und Christiane Buchholtz. 2011. Familiäre und kognitive Bedingungen des Wissenserwerbs in Deutsch, Englisch und Mathematik. In *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen – Erste Ergebnisse aus TEDS-LT*, Hrsg. Sigrid Blömeke, Albert Bremerich-Vos, Helga Haudeck, Gabriele Kaiser, Rainer Lehmann, Günter Nold, Knut Schwippert und Heiner Willenberg. Münster: Waxmann.
- Blömeke, Sigrid, Ute Suhl, Gabriele Kaiser und Martina Döhrmann. 2012. Family background, entry selectivity and opportunities to learn: What matters in primary teacher education? An international comparison of fifteen countries. *Teaching and Teacher Education* 28: 44-55.
- Brookhart, Susan M., und Donald J. Freeman. 1992. Characteristics of Entering Teacher Candidates. *Review of Educational Research* 62: 37-60.
- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Hillsdale, NJ: Lawrence Erlbaum.
- Esser, Hartmut. 1999. *Soziologie: Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt a. M.: Campus.

- Europäische Kommission. 2011. *Report on Progress on Equality between Women and Men in 2010: The Gender Balance in Business Leadership*. Luxembourg: Publications Office of the European Union.
- Furr, R. Michael, und Robert Rosenthal. 2003. Evaluating Theories Efficiently: The Nuts and Bolts of Contrast Analysis. *Understanding Statistics 2*: 33-67.
- Gold, Andreas, und Heinz Giesen. 1993. Leistungsvoraussetzungen und Studienbedingungen bei Studierenden verschiedener Lehrämter. *Psychologie in Erziehung und Unterricht 40*: 111–124.
- Guarino, Cassandra M., Lucrecia Santibañez und Glenn A. Daley. 2006. Teacher Recruitment and Retention: A Review of the Recent Empirical Literature. *Review of Educational Research 76*: 173-208.
- Hanushek, Eric A., und Steven G. Rivkin. 2010. Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review 100*: 267-271.
- Heine, Christoph, Heike Spangenberg und Dieter Sommer. 2004. Studienberechtigte 2002 ein halbes Jahr nach Schulabgang. Ergebnisse einer Vorabauswertung der ersten Befragung der Studienberechtigten 2002 und Vergleich mit den Studienberechtigten 1990, 1994, 1996 und 1999. *HIS Kurzinformation*. Hannover: HIS.
- Heublein, Ulrich, Christopher Hutzsch, Jochen Schreiber, Dieter Sommer und Georg Besuch. 2010. Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen. Ergebnisse einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08. *HIS: Forum Hochschule*. Hannover: HIS.
- Hitpass, Josef, und Albert Mock. 1972. *Das Image der Universität: studentische Perspektiven*. Düsseldorf: Bertelsmann.
- Holland, John L. 1997. *Making vocational Choices: A Theory of Vocational Personalities and Work Environments*. 3rd Edition. Odessa, FL: Psychological Assessment Resources.
- Klauer, Karl Josef, und Detlev Leutner. 2010. Intelligenz und Begabung. In *Handwörterbuch Pädagogische Psychologie*, Hrsg. Detlef H. Rost, 304-311. Weinheim: Beltz.
- Klusmann, Uta, Ulrich Trautwein, Oliver Lüdtke, Mareike Kunter und Jürgen Baumert. 2009. Eingangsvoraussetzungen beim Studienbeginn - Werden die Lehramtskandidaten unterschätzt? *Zeitschrift für Pädagogische Psychologie 23*: 265-278.
- KMK. 2010. Das Bildungswesen in der Bundesrepublik Deutschland 2009. Darstellung der Kompetenzen, Strukturen und bildungspolitischen Entwicklungen für den Informationsaustausch in Europa. Bonn: KMK.
- Köller, Olaf, Rainer Waterman, Ulrich Trautwein und Oliver Lüdtke (Hrsg.). 2004. *Wege zur Hochschulreife in Baden-Württemberg. TOSCA - Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien*. Opladen: Leske+Budrich.
- Künsting, Josef, und Frank Lipowsky. 2011. Studienwahlmotivation und Persönlichkeitseigenschaften als Prädiktoren für Zufriedenheit und Strategienutzung im Lehramtsstudium. *Zeitschrift für Pädagogische Psychologie 25*: 105-114.
- Kunter, Mareike. 2011. Forschung zur Lehrermotivation. In *Handbuch der Forschung zum Lehrerberuf*, Hrsg. Ewald Terhart, Hedda Bennewitz und Martin Rothland, 527-539. Münster: Waxmann.
- Malmberg, Lars-Erik. 2006. Goal-Oriented and Teacher Motivation among Teacher Applicants and Student Teachers. *Teaching and Teacher Education 22*: 58-76.

- Mayr, Johannes. 2009. *Studien- und Berufswahlmotive von Lehrpersonen. Wie sie entstehen, wie sie sich verändern und was sie bewirken*. Tagung der Sektion „Empirische Bildungsforschung“ der DGfE. Landau.
- Mayr, Johannes. 2010. Selektieren und/oder qualifizieren? Empirische Befunde zu guten Lehrpersonen. In *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung*, Hrsg. Jürgen Abel, Gabriele Faust, 73-89. Münster: Waxmann.
- McKinsey&Company. 2007. How the World's Best-Performing School Systems Come Out On Top. www.mckinsey.com/App_Media/Reports/SSO/Worlds_School_Systems_Final.pdf (Zugegriffen: 12. Juli 2011).
- Nitsche, Sebastian, Oliver Dickhäuser, Markus Dresel und Michaela S. Fasching. 2013. Zielorientierungen von Lehrkräften als Prädiktoren lernrelevanten Verhaltens im Lehrerberuf. *Zeitschrift für Pädagogische Psychologie* 27: 95-103.
- OECD. 2005. *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: OECD Publishing.
- Pohlmann, Britta, und Jens Möller. 2010. Fragebogen zur Erfassung der Motivation für die Wahl des Lehramtsstudiums (FEMOLA). *Zeitschrift für Pädagogische Psychologie* 24: 73-84.
- Rauin, Udo. 2007. Im Studium wenig engagiert – im Beruf schnell überfordert. *Forschung aktuell* 3: 60-64.
- Retelsdorf, Jan, und Jens Möller. 2012. Grundschule oder Gymnasium? Zur Motivation ein Lehramt zu studieren. *Zeitschrift für Pädagogische Psychologie* 26: 5-17.
- Rothland, Martin. 2011a. Warum entscheiden sich Studierende für den Lehrerberuf? Interessen, Orientierungen und Berufswahlmotive angehender Lehrkräfte im Spiegel der empirischen Forschung. In *Handbuch der Forschung zum Lehrerberuf*, Hrsg. Ewald Terhart, Hedda Bennewitz und Martin Rothland, 268-295. Münster: Waxmann.
- Rothland, Martin. 2011b. Wer entscheidet sich für den Lehrerberuf? Forschung zum soziodemographischen Profil sowie zu Persönlichkeits- und Leistungsmerkmalen angehender Lehrkräfte. In *Handbuch der Forschung zum Lehrerberuf*, Hrsg. Ewald Terhart, Hedda Bennewitz und Martin Rothland, 243-267. Münster: Waxmann.
- Rothland, Martin, und Ewald Terhart. 2011. Eignungsabklärung angehender Lehrerinnen und Lehrer. Einführung in den Thementeil. *Zeitschrift für Pädagogik* 57: 635-638.
- Royston, Patrick 2004. Multiple Imputation of Missing Values. *Stata Journal* 4: 227-241.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Smart, John C., Kenneth A. Feldmann und Corinna A. Ethington. 2000. *Academic Disciplines. Holland's Theory and the Study of College Students and Faculty*. Nashville: Vanderbilt University Press.
- Spinath, Birgit, Stefanie van Ophuysen und Elke Heise. 2005. Individuelle Voraussetzungen von Studierenden zu Studienbeginn: Sind Lehramtsstudierende so schlecht wie ihr Ruf? *Psychologie in Erziehung und Unterricht* 52: 186–197.
- Statistisches Bundesamt. 2011. *Statistisches Jahrbuch 2011*. Wiesbaden: Statistisches Bundesamt.
- Terhart, Ewald. 2008. Die Lehrerbildung. In *Das Bildungswesen in der Bundesrepublik Deutschland*, Hrsg. Kai S. Cortina, Jürgen Baumert, Achim Leschinsky, Karl Ullrich Mayer und Luitgard Trommer, 745-772. Reinbek bei Hamburg: Rowohlt.

-
- Trapmann, Sabrina, Benedikt Hell, Sonja Weigand und Heinz Schuler. 2007. Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie* 21: 11-27.
- Ulich, Klaus. 1998. Berufswahlmotive angehender LehrerInnen. *Die Deutsche Schule* 90: 64-78.
- Watt, Helen, und Paul W. Richardson. 2007. Motivational Factors influencing teaching as a Career Choice: Development and Validation of the FIT-Choice Scale. *The Journal of Experimental Education* 75: 167-202.
- Weishaupt, Horst, und Radoslaw Huth. 2012. *Systematisierung der Lehrerforschung und Verbesserung ihrer Datenbasis*. Bildungsforschung Band 36. Berlin: BMBF.
- Wigfield, Allan, und Jacquelynne S. Eccles. 2000. Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology* 25: 68-81.

Kapitel 3 Who chooses teaching under different labor market conditions? Evidence from West Germany, 1980 – 2009

Erschienen als: Neugebauer, Martin. 2015. Who chooses teaching under different labor market conditions? Evidence from West Germany, 1980 – 2009. *Teaching and Teacher Education* 45: 137-148. Abdruck mit Erlaubnis von Elsevier.

Abstract: To shed light on this question, this paper analyzes trends in labor market conditions and trends in teacher graduates' characteristics over three decades. Based on representative data from West Germany, it shows that teacher earnings have not declined over time. Likewise, there is no decline in academic aptitude among teacher graduates. Unemployment risks, on the other hand, have fluctuated greatly. In times of high unemployment risks, the likelihood to enter teaching was low, especially for persons with pronounced extrinsic job-security motivations. This changed markedly in recent years, when employment prospects became auspicious.

3.1 Introduction

Teacher labor markets in many countries have experienced major changes over the past decades. Several countries report that teacher pay has not kept up with earnings in other occupations (U.S.: Flyer and Rosen 1997; Hanushek and Rivkin 2007; AUS: Leigh and Ryan 2008; U.K.: Dolton 2005; Nickell and Quintini 2002). What is more, many countries face cyclical waves of teacher shortages followed by surpluses (Watt et al. 2012, p. 791). This has left different cohorts of new teachers with strikingly different career prospects. This paper assesses whether such fluctuations influence self-selection processes into teaching. Do labor market conditions influence who enters the profession? This question is timely in light of current debates on teacher quality and pressing teacher shortages in many countries, where a major challenge is to optimize the pool of individuals from which teacher candidates are drawn (OECD 2005; Schleicher 2011; UNESCO 2013). To inform teacher education systems about qualitative characteristics of the applicant pool of teacher candidates under different labor market conditions, this study analyzes the teacher labor market in West Germany over the past three decades. It then traces changes in academic aptitude and selected career choice motivations of student teacher graduates over the same time-period. So far, little is known on the interplay between labor markets and self-selection into teaching. Findings from the current study will help to fill this knowledge gap. First, it contributes to the literature by describing, for the first time, trends in earnings and unemployment risks of teachers relative to other tertiary-educated workers in West Germany. Second,

while some studies analyze pre-training characteristics of student teachers relative to other university students (Gold and Giesen 1993; Klusmann et al. 2009; Neugebauer 2013), this study is the first to describe trends in such characteristics over several cohorts. It focuses on a limited set of characteristics – academic aptitude, as well as intrinsic (subject-interest) and extrinsic (job security) career choice motivations. These characteristics are chosen, because they are available in different cohorts, are correlated with the study success of student teachers (see below), and can arguably be influenced by labor market conditions. On a more general take, the paper shows how expected labor market returns may influence career choices.

3.1.1 Academic aptitude and motivation among teacher candidates

The recruitment of talented teacher personnel is a central goal for most educational systems (OECD 2005; Schleicher 2011). Such recruitment goals are based on the notion that “talented” persons can be identified prior to in-service teaching. It is argued that academic aptitude and certain motivations constitute important preconditions for how well and motivated a person learns during teacher education (e.g. Blömeke et al. 2012; Klusmann 2013; Mayr 2010).

The available empirical evidence supports these conjectures, with some limitations. Results from several recent studies in Germany provide evidence that academic aptitude, measured in terms of grade point averages (GPA, henceforth) obtained in upper secondary graduation certificates (the Abitur), influences the amount of professional knowledge that student teachers accumulate by the end of their study course, net of teacher training effects (Blömeke and Buchholtz 2011; Blömeke et al. 2012; Kleickmann and Anders 2011; Kleickmann et al. 2013, p. 103; König et al. 2012; Kunina-Habenicht et al. 2013). In addition to GPA, teaching motivations are considered to be related to both positive and negative outcome variables among beginning teachers (see Watt & Richardson, 2007). However, research on the predictive value of motivations is still in its fledging stages. What we know so far is that especially intrinsic motivation (i.e. doing something because it is inherently interesting or enjoyable) results in high-quality learning (Ryan and Deci 2000). In teacher research, intrinsic reasons to choose teaching are often distinguished into altruistic-pedagogical and subject-related motives (Brookhart and Freeman 1992; Watt and Richardson 2007). While altruistic-pedagogical motives (example item: “I love working with children”) are mostly unrelated to accumulated teacher knowledge, several studies find that subject-related motivation (example item “I am interested in mathematics”) is a relevant determinant for the amount of accumulated professional knowledge (Blömeke et al. 2012; Kleickmann et al. 2013, p. 103; König and Rothland 2013). Furthermore, subject-interest

correlates positively with learning strategies, study satisfaction (Künsting and Lipowsky 2011), and college grades (Blömeke 2009), and negatively with college dropout intention (Blömeke 2009; Heublein et al. 2010). The effects of extrinsic motivation (i.e. doing something because it leads to a separable outcome) are less clear, and correlations are typically weak. According to Watt and Richardson (2007) extrinsic motivations (also named ‘personal utility values’) such as job security, transferability, and time for family, relate negatively to later planned persistence and career choice satisfaction. Among student teachers, the strongest extrinsic career choice motivation is “job security” (König and Rothland 2013; Watt et al. 2012). Blömeke et al. (2012) find negative effects of such motivations (example item “I seek the long-term security associated with being a teacher”) on different professional knowledge dimensions among future primary school teachers. However, they cannot replicate this finding among secondary school teachers at the end of their teacher education program (Blömeke et al. 2011). One reason for the indefinite associations might be that persons with high extrinsic motivations can have high intrinsic motivations at the same time. If a person is characterized by high (extrinsic) job security motivations and high (intrinsic) subject-related motivations, no negative outcomes are to be expected. In turn, if a person's motivational profile is dominated by security but not subject-related motivations, chances are that this person will invest less energy and accumulate less professional competence during teacher training.

3.1.2 Changes over time in response to labor market shifts?

In sum, the recruitment of personnel with high academic aptitude and teaching motivations that include high subject interests is one important cornerstone for raising professional teacher competence. Have these characteristics remained stable over the past decades or have persons with different characteristics been attracted to teaching, because the job environment has changed? Clearly, a teacher's job environment is rather stable across time, because most features related to the task of teaching itself are relatively stable e teachers transfer knowledge, work with children and youth, and manage complex classroom environments. For this reason, one would assume rather similar characteristics of different cohorts of teacher trainees. However, some features related to the external conditions of the profession can change, such as earnings and employment prospects. Have these external conditions changed, and has this had an effect on who self-selects into the profession?

In recent years, debates on teacher recruitment have intensified as “some countries [...] express concern about the quality and motivation of a proportion of teacher trainees” (OECD

2005, p. 29). These concerns are mainly based on studies by economists, showing that the academic aptitude of future teachers has been declining relative to other college-educated workers over the past decades in the U.S., U.K., and Australia (Bacolod 2007; Corcoran et al. 2004; Hoxby and Leigh 2004; Lakdawalla 2006; Leigh and Ryan 2008; Nickell and Quintini 2002; Stoddard 2003). The most frequently mentioned explanation for declining aptitude is related to shifts in earning opportunities outside of teaching. Over the course of the past decades, women especially have increasingly gained access to alternative occupations with attractive wages. Based on the underlying argument that individuals aim to maximize personal income, and that high aptitude individuals have the highest propensity to achieve high earnings, it is believed that increased labor market opportunities outside of teaching have detracted high-ability individuals – especially women – from teaching. Bacolod (2007) tests this hypothesis with U.S. data and finds indeed that where teacher wages became relatively less attractive, both men and women were less likely to choose teaching as their occupation, especially those with high academic aptitude. Recent calculations show that teachers' wages in Germany are relatively good, both, in comparison with other tertiary-educated workers and in comparison with other countries (OECD 2012, pp. 456-466; Weishaupt and Huth 2012, pp. 97-98). But how have they developed over the past decades? If the theoretical argument holds, one would expect the same link in Germany, i.e. declining academic aptitude in case of relative earnings decline.

The mentioned studies have focused on earnings, but earnings are clearly not the only external return dimension influencing career choice, or the choice of associated fields of study. A recent study from Germany provides evidence that higher graduate unemployment in a given discipline reduces the share of students opting for the associated field of study, with a time lag of several years (Reisz and Stock 2013). This is especially true for teacher studies, where employment opportunities are restricted to a certain employment sector. As mentioned in the previous section, the strongest extrinsic career choice motivation among student teachers is “job security”. How do individuals with high job security motivations react to changing unemployment risks? It can be assumed that persons who are extrinsically motivated by a desire to obtain a secure job are unlikely to choose teaching in times where unemployment risks are higher in teaching than in other occupations. On the other hand, one would not expect any effects on exclusively intrinsically motivated individuals, who choose teaching for reason related to the teaching experience itself (e.g. interest in the subject taught, working with children). Despite decades of research on teachers' career choice motivations (e.g. Brookhart and Freeman 1992; Rothland 2011; Tudhope 1944; Watt et al. 2012; Young 1995), only one study, to my knowledge,

traces changes in motivations over cohorts (Jantzen 1981). The author finds that extrinsic motivations, such as adequate income, retirement provisions and tenure laws were declining in importance between the mid-1940s and the late 1970s among pre-service teachers in the U.S., while intrinsic pedagogical motives such as an “interest in dealing with children” became more relevant over time. Interestingly, extrinsic motivations decreased and intrinsic motivations increased in times of teacher surpluses in the 1970s; prior to this time, between World War II and into the 1960s, teacher shortages were acute. While Jantzen (1981) does not directly discuss potential labor market influences on these motivational shifts, his findings are in line with my conjecture, that career prospects may influence the motivational composition of teacher candidates. Note that Jantzen's (1981) findings are based on non-random samples of student teachers at 1 to 3 institutions in California. In addition and like almost all studies in the field, Jantzen does not compare the motivations of teachers to motivations of students who choose not to become teachers. Hence, the reported changes may simply reflect general motivational shifts among entire student cohorts, and not describe genuine characteristics of teacher candidates. To sum up, this study aims to answer the following research questions:

- How have teacher labor market conditions (earnings and unemployment risks) changed over the past 30 years?
- How have academic aptitude and motivational characteristics of applicants to the profession evolved over the same time-period, possibly in response to changing labor market returns?

More specifically, I assess whether there is a link between academic aptitude and earnings, and a link between unemployment risks and extrinsic ‘job security’ motivation. I contrast the developments in teaching with those outside of teaching, to disentangle general trends from typical trends among teachers. Such a comparative perspective is also important because the propensity to choose teaching will be influenced by the earnings and employment prospects of teachers relative to those of other occupations requiring similar levels of qualifications. The German system of teacher education requires candidates to study teaching at a university (phase 1, duration 3.5 - 5 years) followed by a preparatory training at teacher seminars and schools (phase 2, duration 1.5 - 2 years). Thus, unlike U.S. students, those in Germany typically choose their career path before university enrollment. Hence, a meaningful comparison group for teacher candidates are students or graduates in other fields of study. For labor market analyses, a meaningful comparison group to teachers are persons with tertiary education degrees, who work in other fields.

3.2 Methods

3.2.1 Data to analyze labor market conditions

To document trends in hiring numbers of new teachers, I draw on official statistical data. In addition, to study temporal developments in earnings and unemployment risks of teachers relative to other occupations on the basis of individual-level data, I draw on data from the German Microcensus (Federal Statistical Office 2013). The Microcensus is the official annual labor force survey covering 1 percent of all German households. I use Scientific Use Files containing a 70 percent sample of the originally interviewed individuals from 15 different years (1980, 1985, 1987, 1989, 1991, 1993, 1995, 1996, 2000, and 2004 - 2009), in which information on field of study (including teacher studies) is available. The availability of field of study information is an important requirement to calculate unemployment rates for labor market entrants, because many of them have not yet worked in any job, which could otherwise serve as the basis to compare teachers and non-teachers. Major advantages of the Microcensus are a high continuity in survey questionnaires, large sample sizes, and compulsory response to most questions, contributing to a very low level of non-response. Because I am interested in long-term developments, I concentrate all analyses on the West German population (East German data is available only since 1991). I further restrict the sample to persons with tertiary degrees, to compare teachers to other persons with tertiary degrees, which I believe are the most informative comparison group. I exclude persons younger than 26 and all those who are still in any kind of full-time education or training, as well as those older than 56, to mitigate early retirement issues.

For the persons remaining, I identify the current occupation. I identify all teachers working in the general education system (i.e. in elementary, secondary, or special education schools). I exclude ambiguous cases and persons who work in pre-school, higher education, or as instructors outside of the general education system, such as driving school teachers or private tutors. I also record respondents' field of study and their (un)employment status. Status of unemployment is measured according to the ILO convention as a binary variable differentiating between employed and unemployed. I exclude those who have not been seeking a job in the last four weeks (inactive), and those still in education, vocational programs or military service (ILO 2003). Finally, personal net monthly income is measured. The availability of net income data is a precondition for any meaningful comparison between teachers and other occupations, because the civil servant status of teachers implies that deductions from gross income vary substantially between teachers and other occupations. Because net monthly income is influenced by working

hours, age, citizenship, marital status, other sources of income, and federal state, I control for these factors in the empirical models. The income information is given in intervals (e.g. 24 income categories in the year 2000). I take midpoints of the categories, and multiply the first and last income categories by 0.9 and 1.1, respectively. Net monthly earnings are adjusted by Consumer Price Index (base year = 2005). Income prior to 2002 is converted into Euros. In the analysis, I trim off the top and bottom 2.5 percent of the income distribution to minimize the bias caused by outliers. Overall, the sample comprises 295,467 persons. The data is summarized in Table 1.

Table 1: Data to analyze labor market conditions, by survey year.

| year | N | % teachers | % females | mean income (in EUR) | unemp. rates (in %) |
|-------|---------|------------|-----------|-------------------------|------------------------|
| 1980 | 11 721 | 22 | 29 | 2 585 | 1 |
| 1985 | 14 006 | 23 | 28 | 2 320 | 4 |
| 1987 | 15 172 | 20 | 29 | 2 371 | 4 |
| 1989 | 15 298 | 19 | 33 | 2 380 | 4 |
| 1991 | 16 531 | 19 | 34 | 2 587 | 3 |
| 1993 | 17 627 | 17 | 35 | 2 532 | 4 |
| 1995 | 19 790 | 15 | 36 | 2 457 | 5 |
| 1996 | 20 641 | 13 | 37 | 2 364 | 4 |
| 2000 | 19 448 | 12 | 39 | 2 637 | 3 |
| 2004 | 21 269 | 11 | 41 | 2 629 | 4 |
| 2005 | 23 283 | 10 | 42 | 2 558 | 4 |
| 2006 | 25 005 | 10 | 42 | 2 552 | 4 |
| 2007 | 23 145 | 11 | 40 | 2 633 | 3 |
| 2008 | 24 724 | 11 | 41 | 2 582 | 3 |
| 2009 | 27 807 | 10 | 44 | 2 534 | 3 |
| Total | 295 467 | 14 | 38 | 2 524 | 4 |

Notes: German Microcensus, Scientific-Use-Files 1980-2009; individuals aged 26-56 with tertiary education. Net monthly income is CPI-adjusted (base year = 2005) and converted into Euros prior to 2002. Status of unemployment is measured according to the ILO conventions. The table illustrates how females have been underrepresented among tertiary degree holders, but have been catching up in recent years. It also shows that the share of teachers fluctuated, which will be discussed further in the results section.

3.2.2 Data to analyze characteristics of applicants to the profession

It is rather difficult to find data that captures characteristics of different cohorts of teacher candidates over a thirty-year time-period with identical measures of academic aptitude and

motivational indicators. Especially data that includes comprehensive measurement of motivations such as the FIT-Choice scale is unavailable.¹ The most suitable data source is the repeated series of Konstanz Student Surveys (Konstanz Research Group on Higher Education 2013). Starting in the winter term 1982/83, a representative sample of tertiary students was interviewed on a wide range of topics concerning their student life. Since then, 10 further surveys have been carried out in intervals of two to three years, the most recent one in 2009/10. A two-stage sampling procedure was used. First, there was a structured selection of institutions of higher education, categorized by state, date of foundation and subject offerings. Within each institution, a random student sample was interviewed through mail surveys. Across surveys, the selection of institutions and the sampling strategy within institutions has remained largely constant. Likewise, the core program of the questionnaire has remained unaltered over time, including questions on demographics, high school grades, choice of field of study, and study motives. In order to compare teacher candidates to other students over a long period of time, I restrict the sample to West German students who have obtained the 'Abitur', which is the unrestricted higher education entrance qualification.²

While the high comparability of surveys makes them a good source for time trend analysis, the data is not without imperfections. One drawback of the data, as it is common for mail surveys, is that response rates dropped from 41 % in 1982/83 to 28 % in 2009/10. However, a comparison with official statistics shows that the marginal distributions of available variables match, except for gender. Because female students are overrepresented by about 5 percent in the more recent surveys, I constructed redressment weights to match the official gender distribution for each year. Another limitation to the Konstanz surveys is that they are cross-sectional, and thus not well suited to determine which students actually end up as teachers. To mitigate student dropout and mobility issues, and to proxy actual teacher candidates more closely, I restrict my sample to students reporting to finish their university education within the next 3 semesters (1/3 of the sample), assuming that systematic dropout does not occur shortly before students finish their degree. This assumption seems plausible, given that most dropouts and

¹ The FIT-Choice (Factors Influencing Teaching Choice) scale assesses the primary motivations of teachers to teach, and was demonstrated to be psychometrically sound (see Watt and Richardson 2007 for details).

² Students from universities of applied sciences are excluded, because 50.5 % of these students have restricted higher education entrance qualifications ('fachgebundene' and 'Fachhochschulreife'). Grades from these qualifications are not comparable to 'Abitur' GPA, which I will use as a measure of academic aptitude. In addition, students with a restricted entrance qualification are usually not able to enter university-based teacher programs. I also exclude students older than 35 (2 percent of the sample). These students do not typically enter any academic occupation: Because of free university tuition, some are officially enrolled to obtain student benefits (such as reduced public transport fees), but never intend to finish a degree. Others are pensioners who study for amusement.

transitions to other fields of study occur in the first 2 semesters (Heublein et al. 2003, p. 42). While far from perfect, this approach promises to be a reliable strategy to identify the applicant pool of future teachers and non-teachers as good as possible in lack of longitudinal data. According to their anticipated graduation year, I re-group the students into 7 graduate cohorts. The grouping is somewhat arbitrary. I ran several sensitivity checks varying the time cut-points for the groups and found out that results do not differ substantially. I ended up choosing groups which are sufficiently large and where the timing reflects different labor market situations. In each cohort, I can identify between 257 and 565 student teachers and between 1 161 and 3 553 other students, who will graduate from their university-based training within the next 3 semesters (see Table 2). I refer to these “soon-to-graduate-students” simply as “graduates” in the remainder of the paper. A third drawback is the fact that different types of teacher studies (e.g. teachers for primary and upper secondary school) cannot be differentiated except for the first two cohorts, which is why I am forced to analyze them jointly (but see the tentative findings mentioned in ‘Summary and Discussion’).

Academic aptitude is measured through grade point averages (GPA) obtained in the upper secondary graduation certificate (the Abitur). ‘Abitur’ GPA is measured in 31 categories running from 1.0 to 4.0. For the analyses, GPA is reverse-coded so that higher values indicate better performance. For each student I calculate GPA *relative* to other students of the same graduate-cohort (using both standardized mean GPA and centile rankings). This measure gives the position of a graduate relative to other graduates of the same cohort. Unstandardized grades reveal moderate grade improvement over time (see Table 2). In case this would imply an absolute rise in academic aptitude over cohorts, a relative decline in teacher student grades could in fact mask an absolute rise. However, this is unlikely to have occurred. If anything, educational expansion has led to an increasing intake of less talented students (Walker and Zhu 2005). The grade improvement much rather reflects inflation of grading practices, which is removed by standardization.

To measure *motivations* of graduates, I am restricted to using single-item indicators. I focus on the two previously discussed career choice motivations, intrinsic subject-related interest (= the motivational dimension which is most predictive for learning during teacher training) and extrinsic job security motivation (= the strongest extrinsic motivational dimension). On a 7-point likert item, students indicated how important subject-interest (‘Fachinteresse’) was when choosing their study program. On another 7-point likert item, they indicated how important a secure workplace (‘berufliche Sicherheit’) was. On average, students rated subject-related interest higher ($M=4.85$, $SD=1.33$) than job security ($M=2.80$, $SD=1.89$), which is in

line with other studies (e.g. König and Rothland 2012). I standardize the motivations similarly to ‘Abitur’ GPA, i.e. values reflect the importance of the motive relative to all other graduates of the same cohort. Sample characteristics are summarized in Table 2.

Table 2: Data to analyze characteristics of applicants to the profession, by year of graduation

| year | N | % teachers | % females | unstd. mean GPA | unstd. subj.- interest | unstd. job se- curity |
|---------|--------|------------|-----------|--------------------|---------------------------|--------------------------|
| 1982-83 | 1 528 | 25 | 42 | 2.47 | 4.60 | 2.49 |
| 1984-87 | 3 570 | 16 | 41 | 2.55 | 4.85 | 2.48 |
| 1989-94 | 3 982 | 11 | 42 | 2.66 | 4.87 | 2.73 |
| 1995-98 | 2 052 | 15 | 46 | 2.72 | 4.88 | 2.71 |
| 2000-03 | 1 859 | 14 | 48 | 2.77 | 4.80 | 2.82 |
| 2004-06 | 1 609 | 15 | 52 | 2.72 | 4.89 | 3.06 |
| 2007-10 | 2 412 | 17 | 53 | 2.71 | 4.95 | 3.46 |
| Total | 17 012 | 15 | 45 | 2.65 | 4.85 | 2.80 |

Notes: Konstanz Student Surveys, 1982-2009; students graduating within 3 semesters, own re-grouping. ‘Abitur’ GPA is reverse-coded so that higher numbers indicate better GPA. Unstandardized measures show absolute shifts in graduate characteristics. In the analyses, I standardize measures per cohort to measure characteristics of teacher graduates relative to other graduates from that year.

3.3 Results

To make results more tangible, I rely on graphical representations of the major findings. Further in-depth analyses are provided in an appendix. Within the teacher workforce, the most important differentiation, both in terms of earnings and in terms of applicants' characteristics lies between teachers at different school types (i.e. upper secondary vs. other school types). While I focus my analysis on the teacher labor market as a whole, I point to differences between school types wherever the data allows me to do so.

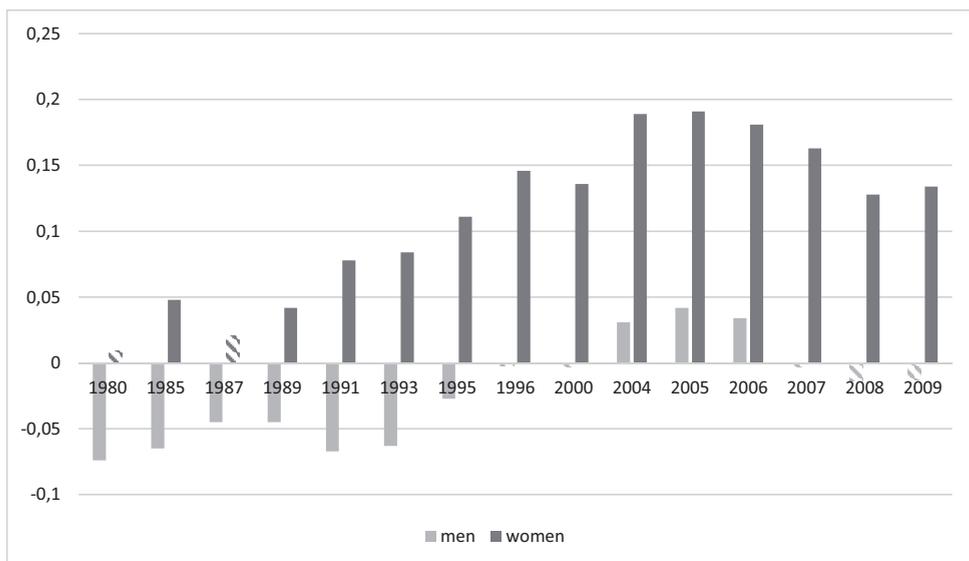
3.3.1 The changing attractiveness of the teacher labor market

The first research question of this study asks how the teacher labor market has changed over the past 30 years.

3.3.1.1 Earnings

I start by describing trends in relative teacher earnings. To compare earnings of teachers against those of other university-educated workers, I ran a series of 30 Ordinary Least Squares (OLS)-regressions, two for each year (one for men, one for women), with log monthly income (5%-trimmed) as dependent variable and working as a teacher vs. working in another occupation as independent variable. Income is log-transformed to reach a better model fit, and so that income differentials between teachers and non-teachers can be interpreted as percentage point differences (Wooldridge 2009, pp. 189-192). I control for potential confounders, such as working hours, age, citizenship, marital status, and federal state. Estimations based on employed persons, aged 26-56, with tertiary degrees who make a living primarily from their own employment and who work at least 10 hours per week. Figure 1 shows bars representing the coefficients of the independent variable from these regression models (control variable coefficients are shown in the supplementary material). Bars can be interpreted as the percentage point difference in monthly earnings of teachers compared to other university-educated workers.

Figure 1: Effect of working as a teacher compared to any other occupation for tertiary-educated workers on personal log net monthly income, by gender.



Notes: Bars with diagonal lines indicate non-significant differences ($p > 0.05$) between teachers and the comparison group. All models include controls for age, age², marital status, citizenship, working hours (in 5 hour steps), and federal state. The full set of covariates can be seen in table A2 (appendix), exemplarily for 2009. Estimations based on employed persons, aged 26-56, with tertiary degrees who make a living primarily from their own employment and who work at least 10 hours per week.

Source: *Microcensus*, different years, own calculations.

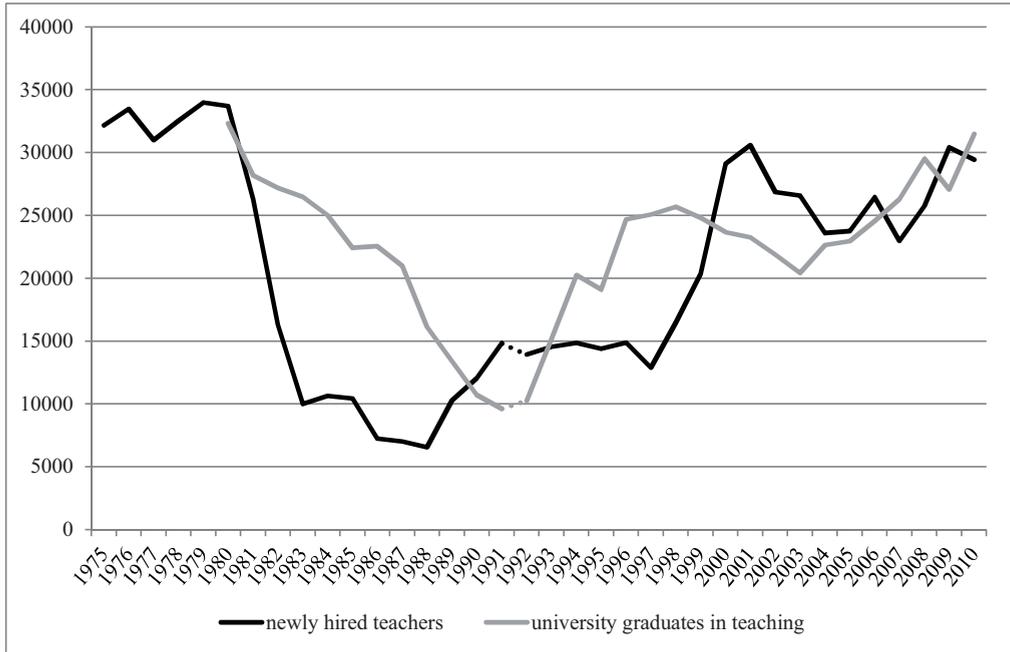
Bars with diagonal lines indicate non-significant differences between the two groups. Between 1980 and 1995, male teachers earned between 7 and 2 percent less than their academic peers in other occupations, with narrowing differences. After that, earnings disadvantages disappear or are too small to pass the test of significance. Between 2004 and 2006, they even experienced a slight earnings advantage of 3-4 percent. For women, teaching has been a financially rewarding occupation in every year of the observation period. Compared to other female academics, their earnings are between 2 and 19 percentage points higher (The female advantage over males is, of course, due to the fact that the female comparison group earns less than the male comparison group). Only in 1980, teaching and non-teaching females do not differ significantly in their log monthly earnings. Between 1991 and 2005, female teachers' earnings increase relative to other tertiary-educated females, from 8 percent to 19 percent, after which the advantage diminishes to about 13 percent. I also looked at differences between school types. On average, upper secondary teachers log net monthly earnings are about 5 percent higher compared to teachers at the lower secondary or elementary level. Somewhat surprisingly, female teachers earn less than their male counterparts, even when controlling for school type and the mentioned control variables. Nevertheless, gender earning gaps in teaching are smaller in comparison to other employment sectors. In sum, earnings in teaching compare favorable to those of tertiary-educated non-teachers, especially for women. In addition, no fall in relative teacher earnings is visible over the past three decades. If anything, teacher earnings rise over time relative to other academic occupations. This picture stands in sharp contrast to the U.K. (Dolton 2005) or U.S. (Bacolod 2007; Hanushek and Rivkin 2007), where relative teacher earnings are lower and have been declining over the past decades. In addition to the presented results, I analyzed absolute earnings trends and dispersions, which showed that earnings have not been declining in absolute terms either, and that earnings dispersions are greater in the comparison group. These additional analyses are displayed in the supplementary material section.

Of course, earnings are not the only important labor market return, which may influence the choice of profession. Out of many factors, such as working conditions, personal satisfaction or social prestige, risk of unemployment stands out as a decisive element in study or occupational decisions (Reisz and Stock 2013). I am going to examine this element in the following section.

3.3.1.2 Supply, demand, and risk of unemployment

Between 1960 and 1980, an unprecedented expansion of new teacher positions occurred, because exceptionally large student cohorts (the “baby-boomers”) entered the education system, and because these students remained in the system longer than previous cohorts as a result of the educational expansion. This increased demand was met by an increasing number of high school graduates opting for teacher studies at university (see Lundgreen 2013 chapt. 4 for more details). During the peak in 1974, 34 percent of all first-year students studied to become a teacher; among female first-year students the share rose up to striking 52 percent (Federal Statistical Office 1975, own calculations). However, when these students entered the teacher labor market roughly 7 years later, not all of them were hired; teacher supply surpassed demand. Available teacher positions had been filled between 1960 and 1980, and only very few new positions were created after 1980, because student numbers dropped after the baby bust in the second half of the 1960s. From Lundgreen (2013, p. 114) it can be inferred that this is true for all school types. In response, but with a time lag, student teachers numbers dropped drastically (to reach a minimum of 7 percent in 1986). In accordance, the number of university graduates in teaching dropped, but some graduated from teaching despite poor hiring prospects. Especially for those entering the labor market in the mid-1980s, career prospects were poor. Teachers born between 1930 and 1950 and hired between 1960 and 1980 dominate the age structure in the teacher labor force; “like a wave” (Lundgreen 2013, p. 110), they shift through the work force until their retirement between 1995 and 2015, and determine the hiring opportunities of the following teacher generations. Figure 2 illustrates the economic cycle in the teacher labor market: massive recruitment in the 1970s; hiring peak in 1980 followed by a dramatic drop until 1988, from 33 987 to 6 559. After that, a slight bettering and stagnation is visible until the mid-1990s, followed by an economic pick-up after 1997, when those hired between 1960 and 1980 (birth cohorts 1930 to 1950) were starting to retire. In anticipation of the slightly improving labor market prospects, the number of persons graduating from teaching began to rise again in the early 1990s. However, their number quickly exceeded those of the newly hired until the turn of the century. Thus, while their labor market prospects were somewhat better compared to the mid-1980s, the labor market situation was still difficult. Only for graduates from the late-1990s onwards, career prospects seriously turned for the better. If a person graduated in 1999 and finished the preparatory training two years later, he/she would enter the labor market at a time where hiring numbers were beginning to exceed the supply.

Figure 2: Number of university graduates in teaching (1980-2010), and newly hired teachers (1975-2010)



Notes: Figures include East Germany as of 1992.

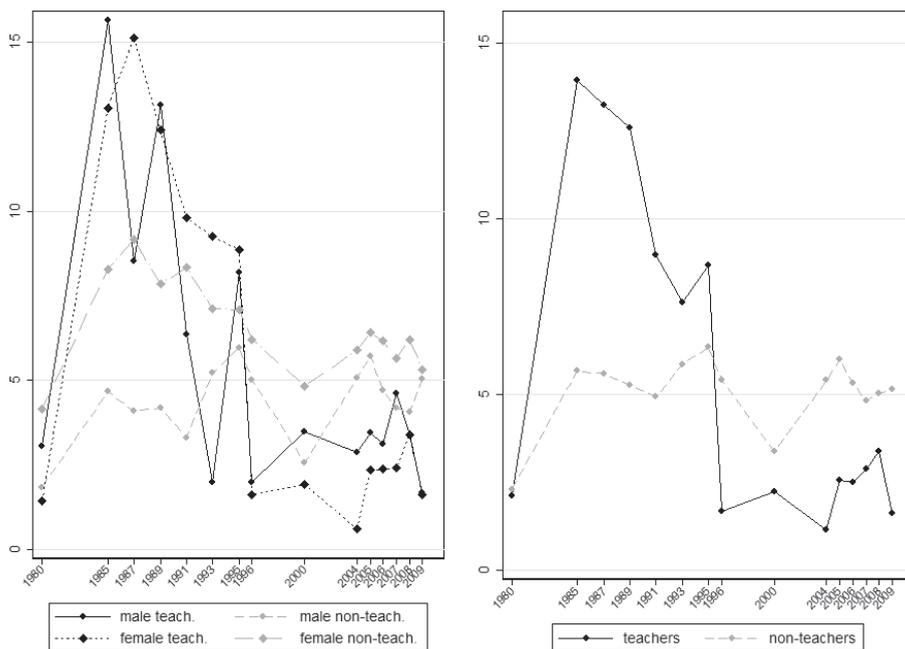
Sources: Standing Conference of the Ministers of Education and Cultural Affairs (2003, 2011) and Lundgreen (2013, p. 134), own depiction.

What are the implications of these changing hiring practices for the relative unemployment risks of new teachers? To answer this question, I calculate unemployment rates from Microcensus data for those who studied teaching and those who studied something else. I restrict the sample to persons aged 26-32 to approximate labor market entrants. Such an approximation becomes necessary, because the Microcensus provides only information on current or last occupation, not on first significant job or time point of labor market entry. The availability of field of study information is an important requirement to calculate unemployment rates for labor market entrants, because many of them have not yet worked in any job, which could otherwise serve as the basis to compare teachers and non-teachers.

Figure 3 shows the unemployment rates for labor market entrants who studied teaching, and for all other fields of study. The left panel of figure 4 shows the unemployment rates separately for men and women. Estimates for male teachers are fluctuating due to small sample sizes, and should be treated with caution. For example, in 1993 only 56 men aged 26-32 who studied teaching can be identified. If we ignore these random fluctuations, the overall trend looks similar for men and women. A clearer picture emerges, when I combine both genders, as

can be seen in the right panel. While unemployment rates in the comparison group is low and rather stable over time, unemployment rates in teaching increased substantially from 2 percent in 1980 to 14 percent in 1986. They remained high compared to other fields of study until 1995. In 1996 rates dropped to 2 percent, a proportion significantly lower than the proportion of unemployed in the comparison group (5 percent, $F(1;4396) = 16.58, p \leq 0.00$). In the following years, unemployment rates for persons with teaching degrees were 1/2 to 1/3 compared to persons with other tertiary degrees. Unemployment rates are in line with what one would suspect from looking at the hiring practices in teaching, as depicted in Figure 2.

Figure 3: Unemployment rates (in %) for labor market entrants with teaching degrees and labor market entrants with other tertiary degrees, 1980-2009.



Notes: Estimates based on persons aged 26-32 who are working or actively looking for a position. Estimates for male teachers should be treated with caution due to small sample sizes.
Source: Microcensus, different years, own calculations.

Overall, going into teaching was a very risky decision in the 1980s and mid-1990s. Only after that time, it became an occupation with smooth school-to-work transition and a low risk of being unemployed.

3.3.2 Changing pre-training characteristics of future teachers

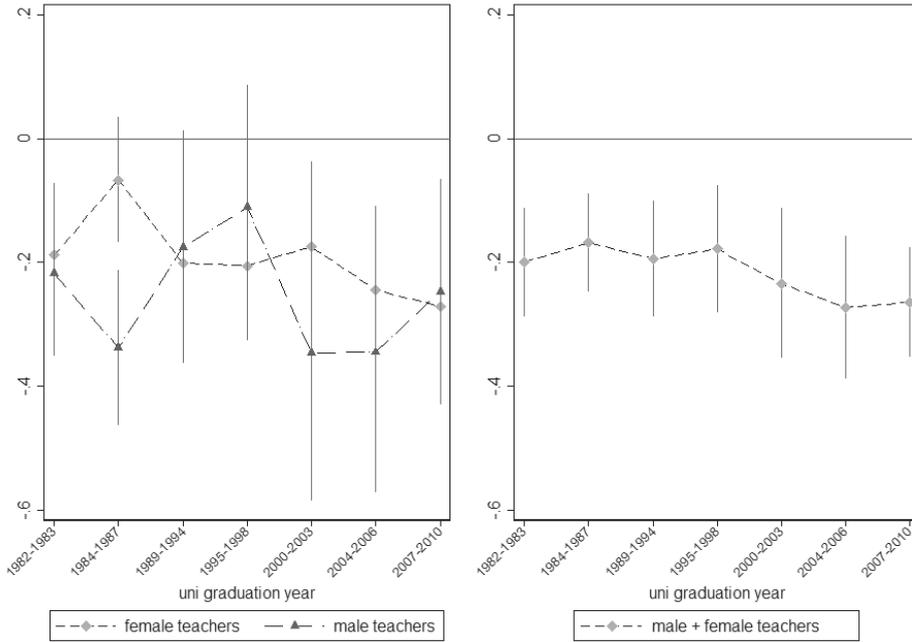
I now turn to the second research question of the paper: How have academic aptitude and motivational characteristics of applicants to the profession evolved over time, possibly in response to changing labor market returns?

3.3.2.1 *Academic aptitude*

As mentioned in section 3.1.2, increased earnings opportunities outside of teaching are believed to have detracted high-ability individuals from teaching in the U.S., Australia, and the U.K.. In West Germany, as I have shown above, relative earnings in teaching are stable over time, and even increased for females. Thus, I do not expect a detraction of high-ability individuals from teaching. Much rather, if the theoretical argument holds, academic ability of teachers should be stable over time, or even increase among female teacher graduates.

Figure 4 shows ‘Abitur’ GPA of student teachers relative to all other students in their graduate-cohort. In each graduate-cohort, GPA has been standardized in a way that the comparison group – non-teaching graduates – has a mean of 0 and a standard deviation of 1, indicated by a horizontal line at 0. The left panel of Figure 4 shows the relative GPA of future teachers separate by gender. A negative trend is visible among female teacher candidates, relative to other graduates, between 1984-1987 and 2007-2010, but the drop across cohorts is not statistically significant. For men, no clear picture emerges, which may simply be a result of small case numbers (between 237 and 62 male future teachers can be identified across cohorts). When men and women are combined, results are more clear-cut (see the right panel of Figure 4). In all years, future teachers’ GPA is, on average, between 0.17 and 0.27 SD below those of other graduates. This corresponds to 0.11 - 0.17 ‘Notenpunkte’ on the German grading scale and has to be considered a rather small difference. From the figure, it looks like relative academic aptitude is declining over time; however, the decline is not significant for any pairwise comparison of future teacher cohort means. The same picture emerges when I use cohort-specific percentile rankings instead of standardized grades. The average percentile rank of those graduating from teaching fell from 46 to 43 over the observation period. In all years, teachers rank lower than other graduates, but they do not decline significantly over time.

Figure 4: ‘Abitur’ GPA of teacher graduates relative to other graduates



Source: Konstanz student surveys, different years, own calculations. Whiskers are 95%-confidence intervals.

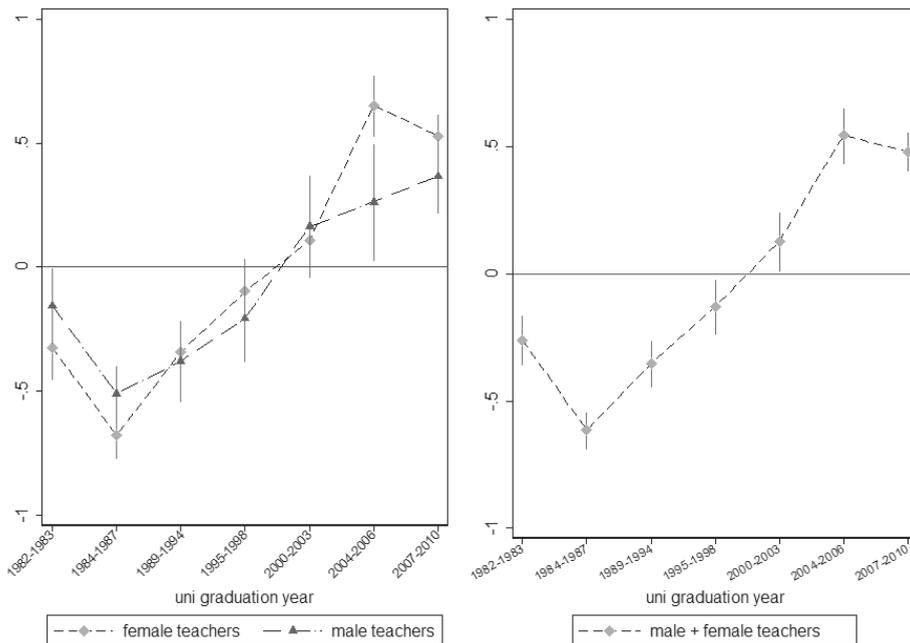
It may well be, that looking at average effects hides important changes at the tails of the distribution, as shown by Corcoran et al. (2004). It would be informative to know if entry into the teaching profession changed for the best and the worst students (i.e. over the surface of the aptitude measure) within each cohort over time. To this end, I estimate 7 logit models (one for each cohort) where the probability of choosing teaching as a career (vs. some alternative) is assumed to be a function of GPA. Instead of entering GPA linearly into the model, I enter it as a set of cohort-standardized quintile dummies. Results (presented in the Appendix) show that students from the tails of the academic aptitude distribution are relatively unaffected by labor market changes. Unlike the U.S., the probability to enter teaching for females or males coming from the top of the distribution does not fall over time.

3.3.2.2 *Motivations for choosing teaching as a career*

In a next step, I turn to motivations for choosing teaching as a career. Who evaded into other fields of study during times of high unemployment, and who remained? Who was motivated to choose teaching, when labor market prospects turned for the better from the late-1990s onwards?

Figure 5 shows the mean *extrinsic job security motive* of student teachers relative to all other students in their graduate-cohort. Again, the comparison group is fixed at 0, indicated by a horizontal line. Trends for both genders are similar (see left panel of Figure 5). While any causal inference is unfeasible in this descriptive study, the parallelism between objective unemployment risks and reported career choice motivation is striking. During the 1980s, when the teacher labor market was stringent, very few teacher candidates report that they chose teaching for job-security reasons. Teacher candidates graduating between 1984 and 1987 had a job security motivation of 0.61 SD below the comparison group. Persons with high job security motivations evaded most strongly into other careers in the mid-1980s, and only few remained, who did not care too much about job security issues. When labor market prospects turned for the better from the late 1990s onwards, more people were attracted into teaching, and the number of graduates rose markedly, as shown in Figure 2, above. This had noticeable effects on the motivational composition of the applicant pool. Those with high extrinsic job security motivations became more likely to choose teaching over alternative careers, which lifted the mean substantially, to reach a high of 0.54 above the cohort-mean in the 2004-2006 graduate cohort.

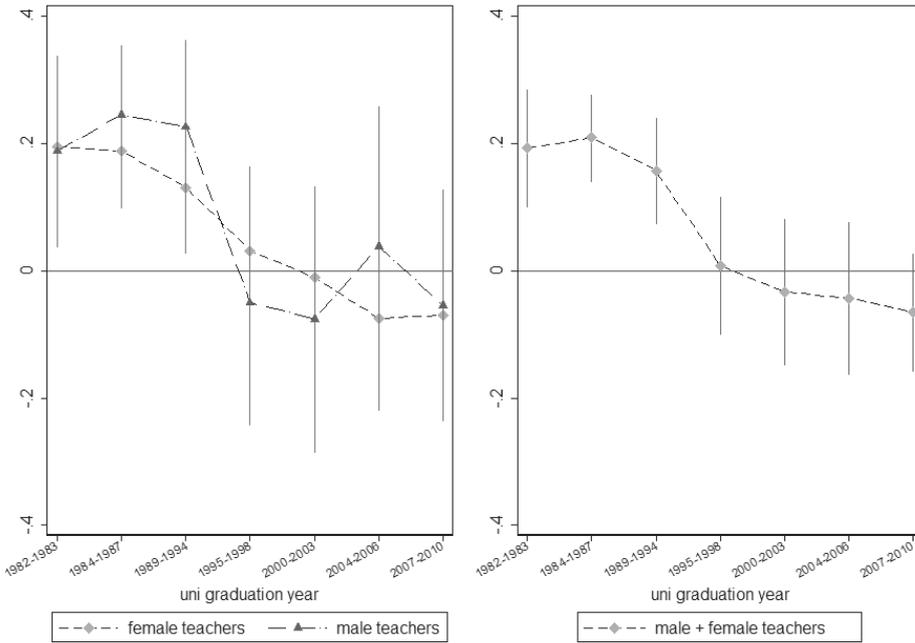
Figure 5: Extrinsic job security motivations of teacher graduates relative to other graduates



Source: Konstanz student surveys, different years, own calculations. Whiskers are 95%-confidence intervals.

Are persons with high intrinsic motives unaffected by labor market shifts? In Figure 6, I display the importance of *intrinsic subject-related motivation* among teacher graduates, again relative to all other graduates. The first thing to notice is that the intrinsic measure is less responsive to labor market shifts (note the different y-axis scaling in Figures 5 and 6). Interestingly, teacher candidates had higher intrinsic subject-related motivations than their peers in other fields of study between 1982 and 1994, when labor market prospects were poor. Differences are rather small, however, with teacher candidates lying 0.21 SD above the comparison group in the 1984-1987 graduate-cohort. After that, teacher candidates relative subject-interest levels fell slightly below the comparison group, but non-significantly so (e.g. 2007-2010: diff.= 0.065, $t = 1.19$, $p=0.23$). The most obvious explanation for this trend is that the inflowing new teacher candidates with high job-security motivations have somewhat lower subject-interests than previous cohorts. This is in fact the case: pairwise correlations between job-security and subject-interest motivations are between -0.09 and -0.19 for all graduate-cohorts.

Figure 6: Intrinsic subject-related interest of teacher graduates relative to other graduates



Source: Konstanz student surveys, different years, own calculations. Whiskers are 95%-confidence intervals.

3.3.2.3 Joint consideration

The presented findings can be modeled and jointly tested in a regression framework. To this end, I estimate 7 logit models (one for each cohort) where the probability of choosing teaching as a career (vs. some alternative) is assumed to be a function of GPA, extrinsic job security motivation, and intrinsic subject-interest motivation. In the models, I control for confounders, i.e. variables that are potentially related to GPA or motivations as well as the career choice. These variables include parental education (a dummy indicating whether at least one parent has a tertiary degree), age, and federal state fixed effects. Table 3 shows coefficients from a joint model for males and females, controlling for gender. To make effects comparable across cohorts, I report average marginal effects (AME) which give the average percentage unit effect of a variable over all studied individuals. In the case of the dummy variable gender, they give the

average percentage unit difference in the probability to choose teaching between females and males (the reference category).³

Table 3: The probability to choose teaching (vs. some alternative) as a function of GPA, gender, and motivational profile

| | 1982-83 | 1984-87 | 1989-94 | 1995-98 | 2000-03 | 2004-06 | 2007-10 |
|------------------------|----------------------|----------------------|----------------------|---------------------|----------------------|---------------------|---------------------|
| GPA | -0.044*** (0.011) | -0.020** (0.006) | -0.014** (0.005) | -0.019* (0.009) | -0.021* (0.009) | -0.024** (0.009) | -0.023** (0.008) |
| (extr.) job security | -0.030** (0.011) | -0.075*** (0.007) | -0.023*** (0.005) | -0.010 (0.008) | 0.016+ (0.008) | 0.067*** (0.009) | 0.076*** (0.008) |
| (intr.) subj.-interest | 0.032** (0.011) | 0.014* (0.007) | 0.009+ (0.005) | -0.005 (0.008) | -0.004 (0.008) | 0.003 (0.009) | 0.002 (0.007) |
| female | 0.175*** (0.023) | 0.106*** (0.013) | 0.111*** (0.011) | 0.145*** (0.017) | 0.088*** (0.016) | 0.108*** (0.017) | 0.108*** (0.015) |
| age | -0.018*** (0.005) | -0.003 (0.003) | -0.009*** (0.002) | -0.007+ (0.003) | -0.002 (0.003) | 0.001 (0.004) | 0.008** (0.003) |
| parental educ. | -0.039 (0.024) | -0.032* (0.014) | -0.020+ (0.011) | -0.021 (0.016) | -0.064*** (0.016) | 0.011 (0.018) | -0.018 (0.015) |
| pseudo R^2 | 0.100 | 0.125 | 0.106 | 0.068 | 0.072 | 0.137 | 0.108 |
| N | 1528 | 3570 | 3982 | 2052 | 1859 | 1609 | 2412 |

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Standard errors in parentheses. Displayed are AMEs. All models include federal state dummies (not shown). parental educ. = at least one parent has a tertiary degree.

Source: Konstanz Student Surveys, different years, own calculations.

The results confirm the findings reported in the descriptive figures. Higher, i.e. better GPA is associated with a lower probability to graduate from teaching as opposed to another field of study in all cohorts. But again, there seems to be no clear trend over time. The strongest difference in AMEs is visible between the 1982-83 cohort (-0.044, 95 % C.I. [-0.066, -0.022]) and the 1989-94 cohort (-0.014, 95 % C.I. [-0.024, -0.0039]), but even here the confidence intervals overlap. Higher extrinsic job-security motivation is associated with a lower probability to graduate from teaching until 1989-94, after which the effect is reversed. A higher intrinsic subject-interest motivation is associated with a higher probability of graduating from teaching in the first three cohorts, after which group differences vanish. Note that females have a higher probability than males to choose teaching in all cohorts. Interestingly, however, females' compared

³ STATA's *margins, dydx* command was used for these calculations. Unlike OR or LnOR, AMEs circumvent scaling issues in non-linear models and make estimates comparable across cohorts.

to males' probability to choose teaching fell over cohorts. This does not mean that the share of females in teaching has been declining; quite the opposite is true (Neugebauer 2011, p. 236). The coefficients indicate that while the share of females in teaching has been rising, the share of females in alternative university study programs (the comparison group) has been rising more rapidly over the past three decades. This reduces the likelihood to enter teaching for females, conditional on entering university.

3.4 Summary and discussion

It is well known that career choices are influenced not only by features of the occupational tasks itself, but also by external conditions of the occupation, such as earnings and employment prospects. Despite recognition that such external conditions are fluctuating in many countries (Schleicher 2011), and that many countries face teacher shortages in the years to come (UNESCO 2013), little research exists on the impact of labor market conditions on who is motivated to choose teaching as a career. This study helps to fill this research gap by exploring the interplay between labor markets and self-selection processes. In light of current debates on teacher quality, it aims to contribute to the understanding of qualitative characteristics of the applicant pool of teacher candidates under different labor market conditions; conditions, which can arguably be influenced by policy makers. The study examines – for the first time – trends in earnings and unemployment risks in the teaching profession in West Germany relative to other academic occupations over the past three decades. It is also the first to describe trends in cognitive and motivational characteristics of teacher candidates relative to other university graduates, in order to examine whether the pool of teacher candidates has changed, possibly in response to shifts in labor market returns.

Several economists from the U.S., Australia, and the U.K. find that increased earnings opportunities outside of teaching have detracted high-ability individuals – especially women – from teaching. In contrast to these studies, I find no earnings decline over the past three decades in West Germany, especially not for women, who have even gained an earnings advantage over other university-educated women. Likewise, academic aptitude of teacher candidates compared to other university graduates has not declined. Overall, teacher graduates score between 0.17 and 0.27 SD below the other university graduates, a significant but relatively small difference. In all cohorts, the probability to choose teaching is lower for persons scoring at the top quintile than for persons scoring at the bottom quintile. Why this is the case remains unclear, and calls for more research. One reason could be the higher earnings dispersion outside of teaching

(Leigh and Ryan 2008). Another reason could be greater prospects for promotion or a higher occupational prestige outside of teaching, which is attractive for persons with high GPAs.

So far, economic studies have concentrated on earnings and academic aptitude. This study complements such studies by acknowledging the fact that unemployment risks are another major labor market condition that affects career choice. The study shows that student teachers react very sensitively to fluctuations of employment prospects. Not only is the number of applicants substantially lower in times of high unemployment risks. Labor market prospects also seem to influence who self-selects into teacher training in terms of motivational characteristics. The study shows that intrinsic motivations draw individuals into the teaching profession, even when labor market conditions are difficult (for a similar finding from Hong Kong see Wong et al. 2014). However, the likelihood of extrinsically motivated persons to enter teaching is very low in times of high unemployment risks, but increases substantially after the mid-1990s, when career prospects are auspicious. Because newly attracted security-affine students have, on average, relatively low intrinsic motivations at the same time, the share of persons with high intrinsic motivations decreases somewhat over time. While causality is hard to establish, the close connection between labor market shifts and motivational shifts strongly suggests that labor market conditions influence who is attracted by the profession. This finding is consistent with Jantzen's (1981) U.S. study. While Jantzen does not discuss labor market influences, he finds that intrinsic motivations were lower and extrinsic motivations were higher in times of teacher shortages, compared to the 1970s where the U.S. experienced teacher surpluses. This suggests that a link between labor market prospects and career choice motivations might be cross-national phenomenon; further research is needed to clarify this.

There are some limitations to this study. First, persons are surveyed shortly before graduating from university, and it remains unclear, whether they actually work as teachers in the future. Because there are no repeated longitudinal data sources, which follow different cohorts of teacher candidates into their job, this is the best possible solution. Second, motivations are self-reported. I cannot rule out the possibility that graduates report motivations in compliance with their already taken career decision. For example, it is possible, that respondents understate the importance of job security in times of high unemployment risks. Third, differences between upper secondary teacher candidates and other teacher candidates cannot be properly investigated with the available data, except for the first two cohorts (1982-83 and 1984-87), where the questionnaires permit such a separation. Studies with more recent cohorts have shown that students aiming to become upper secondary school teachers have better GPAs and higher subject-interest motivations than their peers who are aiming to become a teacher at the primary or lower

secondary level (Neugebauer 2013). I ran separate analyses for the first two cohorts. The findings are in line with Neugebauer (2013) and suggest that differences are rather stable across time. Those aiming to become a teacher at the lower secondary or primary level have worse ‘Abitur’ GPAs (0.44 SD) and lower subject-related motivations (0.24 SD) than those aiming to become an upper secondary teacher. Fourth, persons are attracted to teaching by other external rewards than earnings and employment security, such as occupational status, or general work environment. I was unable to measure these facets, and it is possible that they have changed over time and influenced the types of people attracted by the profession. Fifth, I focused my description on a limited set of indicators – academic aptitude and two career choice motivations. I chose these indicators, because they are available in different cohorts, are correlated with the study success of student teachers, and can arguably be influenced by labor market conditions. Certainly, there are other personal characteristics, and I do not know if they have changed over time in any systematic way. This calls for more research to investigate, for example, if altruistic and other intrinsic or extrinsic motivations have changed over time.

In Germany, the ‘wave’ of teachers hired between 1960 and 1980 (see section 3.3.1.2) is retiring, which has opened up excellent hiring opportunities for new teachers in recent years. This has attracted persons with above-average job security motivations and below-average subject-interests. Whether this is problematic remains ambiguous, and calls for more longitudinal research on the predictive value of career choice motivations. So far, existing research in the field suggests that this might be detrimental to producing teachers who are fully committed to the profession (e.g. Watt and Richardson 2007).

Who are going to be the people filling vacant teaching positions in the future – and under what conditions? Currently, it looks like teacher labor market conditions are worsening again for new teachers in West Germany (but not in East Germany). Student numbers are declining, which is why about one third of the retiring teachers does *not* have to be replaced. In their most recent forecast until 2025, the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (2013) estimates an oversupply of teachers in West Germany, especially for upper secondary schools. Against the overall trend, more teachers are needed for special types of students (vocational and special needs students). Nevertheless, persons who primarily seek job security will most likely evade into alternative occupations in the years to come. A recurring excess supply of new teacher candidates could be an opportunity, to make entry into the profession more selective. In the long run, however, policy makers must aim at interrupting the cyclical waves of teacher surpluses followed by shortages, which are apparent in many countries. This study illustrates how such waves can

have negative consequences, not only for student teachers in excess cohorts, but possibly, also for future generations of students, because in times of teacher shortage candidates may be less committed to teaching than in times of teacher surplus.

Acknowledgements

I am grateful to Walter Müller, Irena Kogan, and Markus Klein for comments on an earlier draft. Markus Klein has also been a valuable help in preparing the Microcensus data. I would like to thank the Research Group on Higher Education at the University of Konstanz for access to the Konstanz survey data, and Juliane Stahl for excellent research assistance. Financial support granted by the Federal Ministry of Education and Research (BMBF, grant number 01JG1058) is gratefully acknowledged.

3.5 Appendix

A1. Additional earnings analysis

In the paper, I do not show control variables from the earnings regressions. I do this here, exemplarily for the most recent observation year.

Table A1: Effect of working as a teacher compared to any other occupation for tertiary-educated workers on personal log net monthly earnings, by gender (year 2009, control variables shown)

| | men | | women | |
|------------------------|---------------------|---------|---------------------|---------|
| | β | (S.E.) | β | (S.E.) |
| teacher | -0.013 | (0.014) | 0.134*** | (0.010) |
| age | 0.082*** | (0.004) | 0.074*** | (0.005) |
| age ² | -0.001*** | (0.000) | -0.001*** | (0.000) |
| married | 0.216*** | (0.007) | -0.017* | (0.009) |
| foreign citizen | -0.123*** | (0.012) | -0.085*** | (0.015) |
| working hours | 0.077*** | (0.002) | 0.099*** | (0.002) |
| <i>federal states:</i> | | | | |
| Schleswig-Holst | -0.001 | (0.018) | -0.034 | (0.022) |
| Hamburg | -0.015 | (0.018) | -0.016 | (0.020) |
| Lower Saxony | -0.023 ⁺ | (0.012) | -0.019 | (0.015) |
| Bremen | -0.059 ⁺ | (0.034) | -0.066 ⁺ | (0.038) |
| North Rhine-Westph. | ref. | | ref. | |
| Hesse | 0.032** | (0.012) | 0.026 ⁺ | (0.015) |
| Rhineland-Palatinate | 0.005 | (0.015) | 0.050** | (0.019) |
| Baden-Württemberg | 0.048*** | (0.010) | 0.020 | (0.013) |
| Bavaria | 0.067*** | (0.009) | 0.035** | (0.012) |
| Saarland | -0.106*** | (0.031) | 0.010 | (0.041) |
| Constant | 5.168*** | (0.083) | 5.097*** | (0.096) |
| R ² | 0.30 | | 0.27 | |
| N | 13818 | | 9106 | |

Notes: ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; standard errors in parenthesis. Working hours measured in 5 hour steps from 10-85 hours. Estimations based on employed persons, aged 26-56, with tertiary degrees who make a living primarily from their own employment and who work at least 10 hours per week.

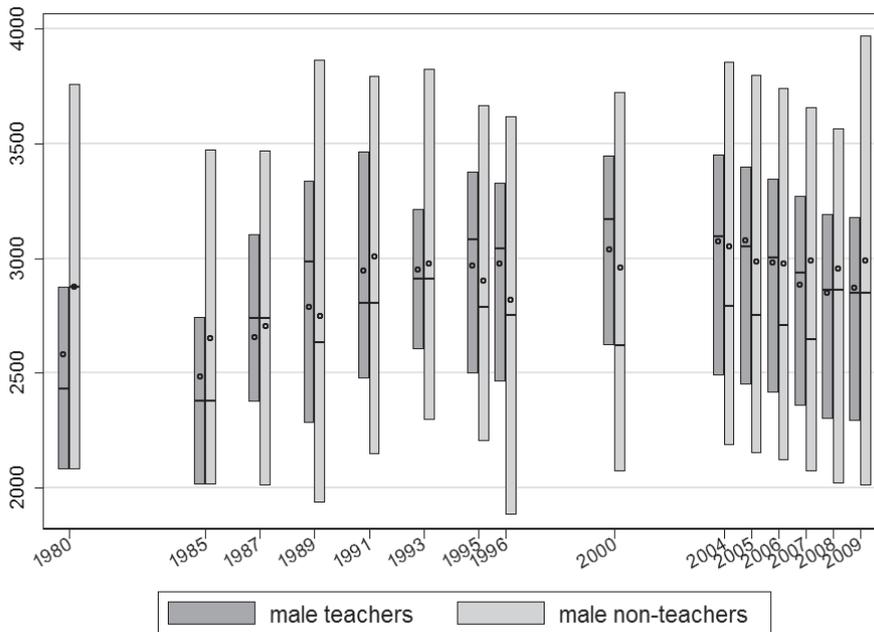
Source: Microcensus, different years, own calculations.

In the paper, I show relative earnings advantages/disadvantages of teachers. Here, I display *absolute* personal net monthly earnings (means and medians), as well as earnings *dispersions* (interquartile range), and earnings of part-time employees.

I begin with males working full-time. Again, the sample is restricted to persons aged 26-56 who make a living primarily from their own employment. The plotted boxes show the median incomes and the range between the upper and lower quartiles. Circles depict mean earnings. In 2009, the latest year for which data is available, median CPI-adjusted net monthly earnings for full-time male teachers as well as for the comparison group are 2850 Euros. This result is in line with OECD statistics, who, for 2010, report equal salaries of teachers and other full-time workers with tertiary degrees. Mean earnings differ by 118 Euros in favor of the comparison group, which is driven by a higher earnings dispersion outside of teaching, especially towards higher earnings. This becomes more evident when looking at the quartiles. Persons at the 75th percentile of the non-teacher earnings distribution earn almost 800 Euros more than

persons at the 75th percentile of the teacher earnings distribution. At the same time, non-teachers at the 25th percentile earn 281 Euros less than teachers scoring at the same percentile in their distribution. In short, male teachers in 2009 earn, on average, about as much as their tertiary educated counterparts outside of teaching, but those outside of teaching have a higher chance to achieve really high or low earnings.

Figure A1: Male personal net monthly earnings, 1980-2009



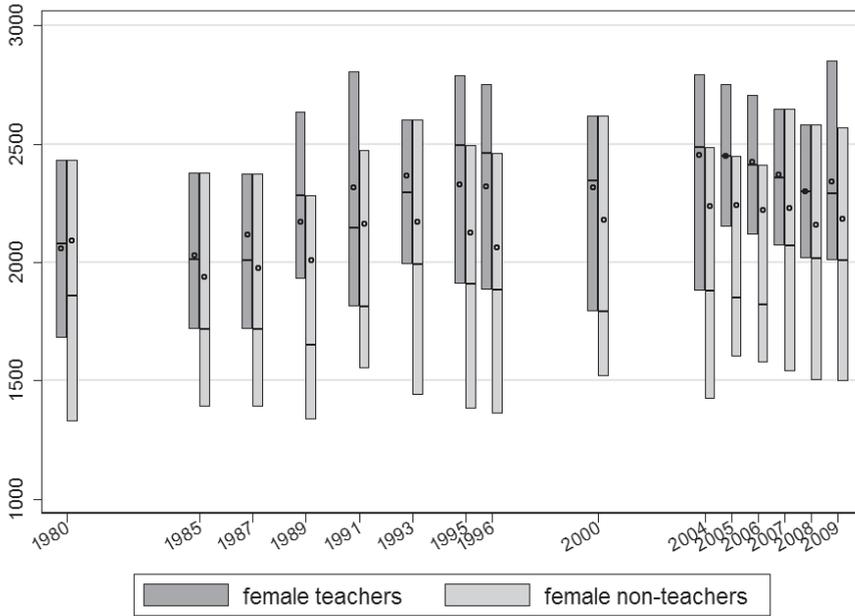
Notes: Estimations based on full-time employed males, aged 26-56, with tertiary degrees who make a living primarily from their own employment. Boxes show the median income and the range between the 25th- and 75th-percentile. Circles depict mean income (5%-trimmed). Income is CPI-adjusted (base year = 2005) and converted into Euros prior to 2002. The somewhat odd-looking monotone decline between 2004 and 2009 is caused by the fact that earnings are measured in intervals, and that they are CPI-adjusted. Those at the 25th- and 75th-percentile fall into the same intervals between 2004 and 2009, and CPI-adjustment reduces these values at a constant rate. For example, teachers at the 75th percentile fall into a category with a mean of 3400 Euros in 2005. They also fall into the 3400 category in 2006, but CPI-adjustment reduces this value by a factor of 0.984 (=3346 Euros).

Source: *Microcensus*, different years, own calculations.

But how have teachers' earnings developed over the past three decades, relative to earnings outside of teaching? The earnings distributions look more similar than different over the course of time. In all years except 1980, male teachers' median earnings are at least as high as those of the comparison group, and are higher between 1995 and 2007. Non-teachers at the 75th percentile earn more than teachers at the 75th percentile in all years, and less at the 25th percentile except for 1980 and 1985. No relative earnings decline in teaching is apparent. Averaging over

the observation period, male teachers' median earnings are 2 percent higher than those of the comparison group. For females, the same measures are depicted in Figure A2.

Figure A2: Female personal net monthly earnings, 1980-2009



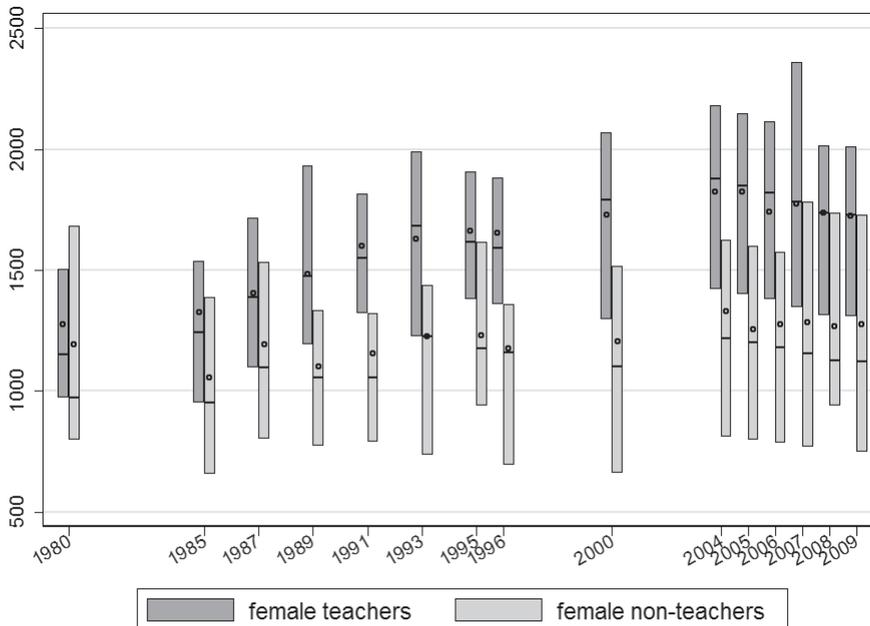
Notes: Boxes show the median income and the range between the 25th- and 75th percentile. Circles depict mean income (5%-trimmed). Estimations based on full-time employed females, aged 26-56, with tertiary degrees who make a living primarily from their own employment. Income is CPI-adjusted (base year = 2005) and converted into Euros prior to 2002.

Source: Microcensus, different years, own calculations.

In all years except for 1980, both measures of central tendency show a clear advantage of female teachers compared to tertiary educated females outside of teaching. For example, in 2009, mean earnings are 2339 Euros for female teachers and 2184 Euros for the comparison group; median earnings are 2290 Euros and 2009 Euros, respectively. On average, female teachers median earnings are 22 percent higher than those of tertiary-educated women working full-time outside of teaching. Unlike men, women at the 75th percentile of the non-teaching earnings distribution never earn more than the women at the 75th percentile in teaching. Over time, female teachers seem to gain some further advantage over other females. A comparison across both figures shows that while male teachers earn more than female teachers (partially because more of them work in higher secondary schools where earnings are higher), the gender earnings gap in school is smaller than outside of the teaching profession. For women working part-time, teaching is

even more profitable relative to non-teaching occupations and earning advantages are increasing over the years, until 2004. Female teachers' median earnings were 18 percent higher in 1980, but 54 percent higher in 2009. Part-time is defined as working between 10 and 34 hours per week. Over the years, between 16 and 47 per cent of female teachers work in part-time. The increase in earnings advantages for female part-time teachers is *not* caused by an increase in working hours among part-time teachers. On average, mean working hours are 22.0 for teachers and 22.8 for non-teachers. Among males, between 4-14 per cent work in part-time, most of them at the beginning of their career, when they are unable to get hold of a full-time position. The limited number of observations precludes any meaningful assessment of part-time male teacher earnings.

Figure A3: Female personal net monthly earnings working *part-time*, 1980-2009



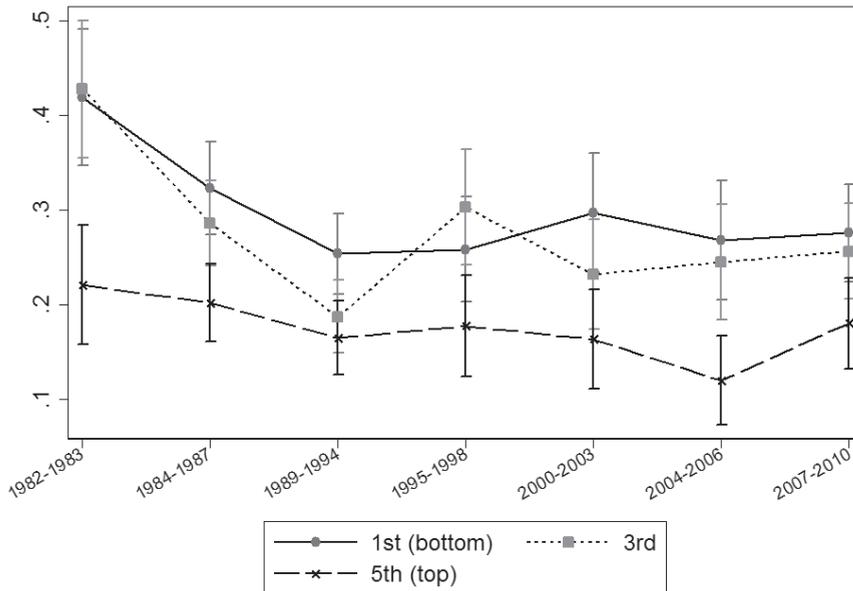
Notes: Boxes show the median income and the range between the 25- and 75 percentile. Circles depict mean income (5%-trimmed). Income is CPI-adjusted (base year = 2005) and converted into Euros prior to 2002. Estimations based on part-time employed females, aged 26-56, with tertiary degrees who make a living primarily from their own employment. *Source:* *Microcensus*, different years, own calculations.

A2 Entry into teaching for the best and worst students

It may well be, that looking at average effects hides important changes at the tails of the distribution, as shown by Corcoran et al. (2004). It would be informative to know if entry into the

teaching profession changed for the best and the worst students (i.e. over the surface of the aptitude measure) within each cohort over time. To this end, I estimate 7 logit models (one for each cohort) where the probability of choosing teaching as a career (vs. some alternative) is assumed to be a function of GPA. Instead of entering GPA linearly into the model, I enter it as a set of cohort-standardized quintile dummies. To make results more tangible, I calculate predicted probabilities of choosing teaching by performance quintile. In Figure A4, I show these probabilities for females coming from the 1st (bottom), 3rd and 5th (top) quintile, with other covariates set to their mean and parental education background fixed to non-tertiary (the pattern looks similar for males, but probability levels are, of course, substantially lower). In all cohorts, females from the bottom 20 percent of the GPA distribution have a higher probability to choose teaching than those from the top 20 percent. During the recession on the teacher labor market in the 1980s, probabilities to choose teaching decreased for all groups, but most strongly for those coming from the middle 20 percent of the aptitude distribution. The best female students seem to be relatively unaffected by labor market changes. Thus, unlike the U.S., the probability to enter teaching for females coming from the top of the aptitude distribution does not fall over time.

Figure A4: Predicted probabilities to choose teaching for females, by GPA quintile



Notes: whiskers show 95 % confidence intervals.
Source: Konstanz Student Surveys, different years, own calculations.

3.6 Literature

- Bacolod, Marigee P. 2007. Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality. *The Review of Economics and Statistics* 89: 737-751.
- Blömeke, Sigrid. 2009. Ausbildungs- und Berufserfolg im Lehramtsstudium im Vergleich zum Diplom-Studium – Zur prognostischen Validität kognitiver und psycho-motivationaler Auswahlkriterien. *Zeitschrift für Erziehungswissenschaft* 12: 82-110.
- Blömeke, Sigrid, and Christiane Buchholtz. 2011. Familiäre und kognitive Bedingungen des Wissenserwerbs in Deutsch, Englisch und Mathematik. In *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen – Erste Ergebnisse aus TEDS-LT*, Eds. Sigrid Blömeke, Albert Bremerich-Vos, Helga Haudeck, Gabriele Kaiser, Rainer Lehmann, Günter Nold, Knut Schwippert and Heiner Willenberg. Münster: Waxmann.
- Blömeke, Sigrid, Gabriele Kaiser and Martina Döhrmann. 2011. Bedingungsfaktoren des fachbezogenen Kompetenzerwerbs von Lehrkräften. Zum Einfluss von Ausbildungs-, Persönlichkeits- und Kompositionsmerkmalen in der Mathematiklehrerausbildung für die Sekundarstufe I. *Zeitschrift für Pädagogik* 57. Beiheft: 77-103.
- Blömeke, Sigrid, Ute Suhl, Gabriele Kaiser and Martina Döhrmann. 2012. Family Background, Entry Selectivity and Opportunities to Learn: What Matters in Primary Teacher Education? An International Comparison of Fifteen Countries. *Teaching and Teacher Education* 28: 44-55.
- Brookhart, Susan M., and Donald J. Freeman. 1992. Characteristics of Entering Teacher Candidates. *Review of Educational Research* 62: 37-60.
- Butz, Marcus. 2001. Lohnt sich Bildung noch? Ein Vergleich der bildungsspezifischen Nettoeinkommen 1982 und 1995 In *Die Erwerbsgesellschaft: Neue Ungleichheiten und Unsicherheiten*, Eds. Peter A. Berger, Dirk Konietzka, 95-117. Opladen: Leske + Budrich.
- Corcoran, Sean P., William N. Evans and Robert M. Schwab. 2004. Women, the Labor Market, and the Declining Relative Quality of Teachers. *Journal of Policy Analysis and Management* 23: 449-470.
- Dolton, Peter J. 2005. *The Labour Market for Teachers: A Policy Perspective*. UK: Office of Manpower Economics, School Teachers' Review Body.
- Federal Statistical Office. 1975. Fachserie 11. Reihe 4.1. Bildung und Kultur. Studierende an Hochschulen. Wintersemester 1974/1975. Wiesbaden: Statistisches Bundesamt.
- Federal Statistical Office. 2013. Microcensus. https://www.destatis.de/EN/Meta/abisz/Mikrozensus_e.html (accessed: 14.05.2013).
- Flyer, Fredrick, and Sherwin Rosen. 1997. The New Economics of Teachers and Education. *Journal of Labor Economics* 15: 104-139.
- Gold, Andreas, and Heinz Giesen. 1993. Leistungsvoraussetzungen und Studienbedingungen bei Studierenden verschiedener Lehrämter. *Psychologie in Erziehung und Unterricht* 40: 111-124.
- Hanushek, Eric A., and Steven G. Rivkin. 2007. Pay, Working Conditions, and Teacher Quality *The Future of Children* 17: 69-86
- Heublein, Ulrich, Christopher Hutzsch, Jochen Schreiber, Dieter Sommer and Georg Besuch. 2010. *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen*.

- Ergebnisse einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08*. HIS: Forum Hochschule. Hannover: HIS.
- Heublein, Ulrich, Heike Spangenberg and Dieter Sommer. 2003. *Ursachen des Studienabbruchs. Analyse 2002*. Hannover: HIS.
- Hoxby, Caroline M., and Andrew Leigh. 2004. Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States. *American Economic Review* 94: 236-240.
- ILO (International Labour Office). 2003. *International Training Compendium on Labour Statistics*. Module 1. Turin: ILO.
- Jantzen, J. Marc. 1981. Why College Students Choose to Teach: A Longitudinal Study. *Journal of Teacher Education* 32: 45-49.
- Kleickmann, Thilo, and Yvonne Anders. 2011. Lernen an der Universität. In *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, Eds. Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss and Michael Neubrand, 305-315. Münster: Waxmann.
- Kleickmann, Thilo, Dirk Richter, Mareike Kunter, Jürgen Elsner, Michael Besser, Stefan Krauss and Jürgen Baumert. 2013. Teachers' Content Knowledge and Pedagogical Content Knowledge: The Role of Structural Differences in Teacher Education. *Journal of Teacher Education* 64: 90-106.
- Klusmann, Uta. 2013. Individual Characteristics of Prospective Teachers. In *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers*, Eds. Mareike Kunter, Jürgen Baumert, Werner Blum, Ute Klusmann, Stefan Krauss and Michael Neubrand, 311-319. New York: Springer.
- Klusmann, Uta, Ulrich Trautwein, Oliver Lüdtke, Mareike Kunter and Jürgen Baumert. 2009. Eingangsvoraussetzungen beim Studienbeginn - Werden die Lehramtskandidaten unterschätzt? *Zeitschrift für Pädagogische Psychologie* 23: 265-278.
- König, Johannes, and Martin Rothland. 2012. Motivations for choosing teaching as a career: effects on general pedagogical knowledge during initial teacher education. *Asia-Pacific Journal of Teacher Education* 40: 289-315.
- König, Johannes, and Martin Rothland. 2013. Pädagogisches Wissen und berufsspezifische Motivation am Anfang der Lehrerausbildung. Zum Verhältnis von kognitiven und nicht-kognitiven Eingangsmerkmalen von Lehramtsstudierenden. *Zeitschrift für Pädagogik* 59: 1-23.
- König, Johannes, Sarantis Tachtsoglou and Andreas Seifert. 2012. Individuelle Voraussetzungen, Lerngelegenheiten und der Erwerb von pädagogischem Professionswissen. In *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung*, Eds. Johannes König, Andreas Seifert, 243-283. Münster: Waxmann.
- Konstanz Research Group on Higher Education. 2013. Student Survey in Germany 1983 to 2010. <http://cms.uni-konstanz.de/en/ag-hochschulforschung/student-survey-in-germany> (accessed: 20.11.2012).
- Kunina-Habenicht, Olga, Franziska Schulze-Stocker, Mareike Kunter, Jürgen Baumert, Detlev Leutner, Doris Förster, Hendrik Lohse-Bossenz and Ewald Terhart. 2013. Die Bedeutung

- der Lerngelegenheiten im Lehramtsstudium und deren individuelle Nutzung für den Aufbau des bildungswissenschaftlichen Wissens. *Zeitschrift für Pädagogik* 59: 1-23.
- Künsting, Josef, and Frank Lipowsky. 2011. Studienwahlmotivation und Persönlichkeitseigenschaften als Prädiktoren für Zufriedenheit und Strategienutzung im Lehramtsstudium. *Zeitschrift für Pädagogische Psychologie* 25: 105-114.
- Lakdawalla, Darius. 2006. The Economics of Teacher Quality. *Journal of Law and Economics* 49: 285-329.
- Leigh, Andrew, and Chris Ryan. 2008. How and Why Has Teacher Quality Changed in Australia? *The Australian Economic Review* 41: 141-159.
- Lundgreen, Peter. 2013. *Datenhandbuch zur deutschen Bildungsgeschichte. Band XI: Die Lehrer an den Schulen in der Bundesrepublik Deutschland 1949–2009*. Göttingen: Vandenhoeck & Ruprecht.
- Mayr, Johannes. 2010. Selektieren und/oder qualifizieren? Empirische Befunde zu guten Lehrpersonen. In *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung*, Eds. Jürgen Abel, Gabriele Faust, 73-89. Münster: Waxmann.
- Neugebauer, Martin. 2011. Werden Jungen von Lehrerinnen bei den Übergangsempfehlungen für das Gymnasium benachteiligt? Eine Analyse auf Basis der IGLU-Daten. In *Geschlechtsspezifische Bildungsungleichheiten*, Eds. Andreas Hadjar, 235-260. Wiesbaden: VS.
- Neugebauer, Martin. 2013. Wer entscheidet sich für ein Lehramtsstudium – und warum? Eine empirische Überprüfung der These von der Negativselektion in den Lehrerberuf. *Zeitschrift für Erziehungswissenschaft* 16: 157-184.
- Nickell, Stephen, and Glenda Quintini. 2002. The Consequences of the Decline in Public Sector Pay in Britain: A Little Bit of Evidence. *The Economic Journal* 112: F107-F118.
- OECD. 2005. *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: OECD Publishing.
- OECD. 2012. *Education at a Glance: OECD Indicators*. Paris: OECD Publishing.
- Reisz, Robert D., and Manfred Stock. 2013. Hochschulexpansion, Wandel der Fächerproportionen und Akademikerarbeitslosigkeit in Deutschland. *Zeitschrift für Erziehungswissenschaft* 16: 137-156.
- Rothland, Martin. 2011. Warum entscheiden sich Studierende für den Lehrerberuf? Interessen, Orientierungen und Berufswahlmotive angehender Lehrkräften im Spiegel der empirischen Forschung. In *Handbuch der Forschung zum Lehrerberuf*, Eds. Ewald Terhart, Hedda Bennewitz and Martin Rothland, 268-295. Münster: Waxmann.
- Ryan, Richard M., and Edward L. Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25: 54-67.
- Schleicher, Andreas. 2011. *Building a High-Quality Teaching Profession: Lessons from Around the World*. Paris: OECD Publishing.
- Standing Conference of the Ministers of Education and Cultural Affairs. 2003. *Einstellung von Lehrkräften 2002*. Statistische Veröffentlichungen der Kultusministerkonferenz (Nr. 166).
- Standing Conference of the Ministers of Education and Cultural Affairs. 2011. *Einstellung von Lehrkräften 2010*. Statistische Veröffentlichungen der Kultusministerkonferenz (Nr. 193).
- Standing Conference of the Ministers of Education and Cultural Affairs. 2013. *Lehrereinstellungsbedarf und -angebot in der Bundesrepublik Deutschland*.

- Modellrechnung 2012 - 2025*. Statistische Veröffentlichungen der Kultusministerkonferenz (Nr. 201).
- Stoddard, Christiana. 2003. Why has the Number of Teachers per Student Risen While Teacher Quality has Declined? The Role of Changes in the Labor Market for Women. *Journal of Urban Economics* 53: 458–481.
- Tudhope, William B. 1944. Motives for the Choice of the Teaching Profession by Training College Students. *British Journal of Educational Psychology* 14: 129-141.
- UNESCO. 2013. A Teacher for Every Child: Projecting Global Teacher Needs from 2015 to 2030. Montreal: UNESCO Institute for Statistics.
- Walker, Ian, and Yu Zhu. 2005. The College Wage Premium, Overeducation and the Expansion of Higher Education in the UK. *IZA Discussion Paper* No. 1627.
- Watt, Helen, and Paul W. Richardson. 2007. Motivational Factors Influencing Teaching as a Career Choice: Development and Validation of the FIT-Choice Scale. *The Journal of Experimental Education* 75: 167-202.
- Watt, Helen M. G., Paul W. Richardson, Uta Klusmann, Mareike Kunter, Beate Beyer, Ulrich Trautwein and Jürgen Baumert. 2012. Motivations for Choosing Teaching as a Career: An International Comparison Using the FIT-Choice Scale. *Teaching and Teacher Education* 28: 791-805.
- Weishaupt, Horst, and Radoslaw Huth. 2012. Systematisierung der Lehrerforschung und Verbesserung ihrer Datenbasis. *Bildungsforschung* Band 36. Berlin: BMBF.
- Wong, Angel K. Y., Sylvia Y. F. Tang and May M. H. Cheng. 2014. Teaching Motivations in Hong Kong: Who Will Choose Teaching as a Fallback Career in a Stringent Job Market? *Teaching and Teacher Education* 41: 81-91.
- Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*. 4th Edition, Mason, Ohio: South-Western Cengage Learning.
- Young, Betty J. 1995. Career Plans and Work Perceptions of Preservice Teachers. *Teaching and Teacher Education* 11: 281-292.

Kapitel 4 Unmasking the myth of the same-sex teacher advantage

Erschienen als: Neugebauer, Martin, Marcel Helbig und Andreas Landmann. 2011. Unmasking the myth of the same-sex teacher advantage. *European Sociological Review* 27: 669-689. Abdruck mit Erlaubnis von Oxford University Press.

Abstract: Trend statistics reveal a striking reversal of a gender gap that has once favoured males: Girls have surpassed boys in many aspects of the educational system. At the same time, the share of female teachers has grown in almost all countries of the western world. There is an on-going, contentious debate on whether the gender of the teacher can account, in part, for the growing educational disadvantage of males. In this study, we use large-scale data from IGLU-E, an expansion of PIRLS in Germany, to estimate whether there is a causal effect of having a same-sex teacher on student outcomes. We estimate effects for typical ‘female’ subjects and typical ‘male’ subjects as well as for different student outcomes (objective test scores and more subjective teacher’s grades). We find virtually no evidence of a benefit from having a same-sex teacher, neither for boys nor for girls. These findings suggest that the popular call for more male teachers in primary school is not the key to tackle the growing disadvantage of boys.

4.1 Introduction

In past decades, school systems in nearly all countries of the world have seen a gender-specific change in educational achievement. Even though girls’ performance in math and science continues to lag behind that of boys in most countries, girls have been catching up with boys regarding both their math and science skills (Baker and Jones 1993; Cole 1997; Hedges and Nowell 1995; IES 2009; Willingham and Cole 1997). What is more, girls have higher reading skills than boys and significantly higher writing skills (Cole 1997; German PISA Consortium 2007; Hedges and Nowell 1995; IES 2009; Wagemaker et al. 1996; Willingham and Cole 1997). In nearly all countries of the world, women’s participation in tertiary education, when compared to that of men, is at much higher levels than it was 40 years ago (UNESCO 2009). In some European countries, the share of women among first-year university students has increased to more than 60 percent of all first-year students (OECD 2009: own calculations). This development has led some researchers to talk about a “boy crisis” (Dammach 2007; Pollack 2006) or even a “war against boys” (Sommers 2000). As a consequence, girls’ and women’s most recent levels of educational success in relation to boys and men are interpreted as male educational failure. Research on this rare case of an inversion of a pattern of social stratification

(Buchmann and DiPrete 2006; Quenzel and Hurrelmann 2010) is still in its fledgling stages, however. This is because a change of perspective with regard to gender-specific educational achievement only occurred in the late 1990s, and the causes of this new phenomenon have only been studied in recent years.

At the same time that girls' opportunities for educational success have been on the increase compared to those of boys, the share of women among the teaching population has been rising in virtually all countries of the Western hemisphere (Eurostat 2009). This so-called "feminization of schooling" has been repeatedly linked to boys' poor educational opportunities when compared to girls (BBC-News 2000; Dee 2007; Sexton 1969). It is conjectured that girls and boys tend to perform better with teachers of their own gender and/or receive better grades than with teachers of the opposite gender. It is assumed that girls have benefited academically from the increased share of female teachers among the teaching population, whereas boys have been negatively affected by the simultaneous decrease in male teachers at school. At the aggregate level, this hypothesis seems to be plausible. Figure 1 shows the correlation between the share of female teachers in primary and secondary schools in relation to all teachers and the share of female students enrolled in tertiary education in relation to all students in tertiary education in 135 countries. The two indicators are highly correlated - The higher the share of female teachers in a given country, the higher the share of female students who are enrolled in tertiary education.¹

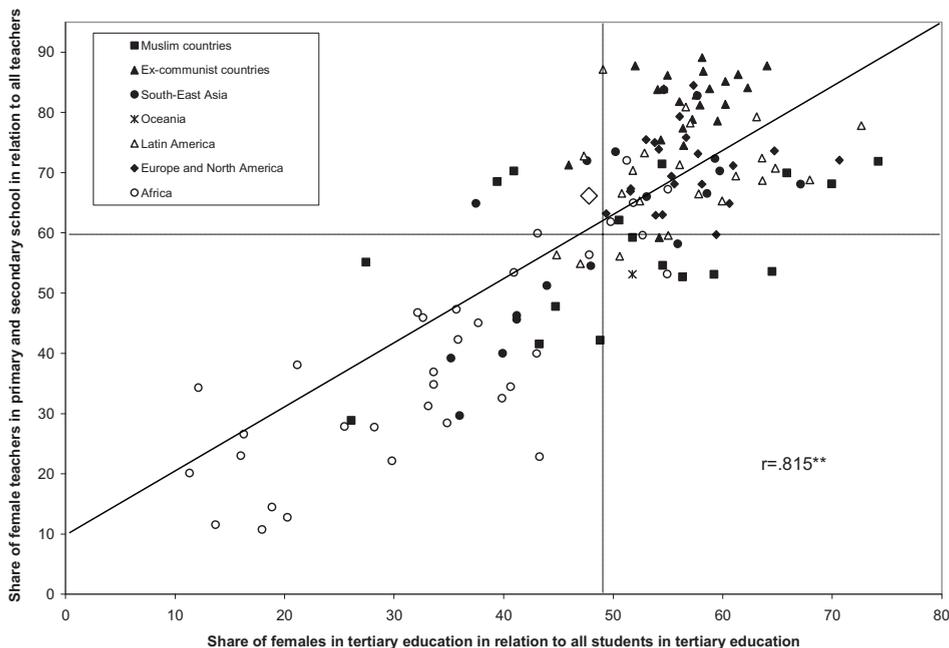
So far however, robust empirical evidence to confirm – at the individual level – the strong association which may be observed at the country level has been scarce. This lack of empirical support notwithstanding, efforts have been made to increase the share of male teachers, most importantly in Anglo-American countries (Driessen 2007). This is considered an effective instrument for improving the performance of boys. In the United Kingdom, the Training and Development Agency for Schools started an initiative to encourage more men to become teachers (TDA 2009), as did the Federal Government of Australia (ABC-News 2004).

In this paper, we seek to examine whether teacher gender in fact has an impact on the academic achievement of male and female students, using the German supplemental sample from the 2001 PIRLS study (Progress in International Reading Literacy).²

¹ Of course, this correlation can be spurious. Cf. 'Summary and Conclusion' for a discussion on this.

² The PIRLS study was conducted in 2001. 35 countries participated in the study, which assessed the reading literacy of fourth-graders. The 2001 German supplemental sample collected additional data on students' math and science skills as well as on their grades in German, Mathematics, and Science ('Sachkunde').

Figure 1: Correlation between the share of female teachers in primary and secondary schools in relation to all teachers and the share of females into tertiary education in relation to all students in tertiary education in 135 countries (2007).



Notes: The figure includes all countries for which data was available. Countries with a current tertiary student population of fewer than 3000 students were not included. The horizontal and vertical dashed lines indicate the respective median value for all countries. The diagonal line represents the regression line. The bigger white square close to the median values represents Germany.

Source: Own figure based on UNESCO (2010).

The evidence presented here aims to make several contributions to existing research. First, this study focuses on Germany, for which teacher gender effects have not yet been estimated directly (i.e. with individual-level data). Second, in addition to “gender-blind” performance scores, the educational outcomes in this study include teacher grades. Comparing grades and performance outcomes indicates whether teachers’ gender in fact influences students’ actual performance development or rather relates to their assessment of students, net of performance.³ Third, we estimate effects across several subjects (German, Mathematics, and Science). Concentrating on these three domains is a reasonable approach because reading may generally be

³ In addition, grades are a useful component because of their relevance for students’ transition into one of Germany’s distinct secondary school types, which in large parts determines a student’s future educational and occupational career.

called a “girls’ domain”, whereas math is considered to be more of a “boys’ domain”.⁴ Fourth, the rich data set allows us to estimate effects not only for the student population as a whole, but also for various subgroups, such as students from migrant or low socio-economic status families. Fifth, and perhaps most importantly, this study adopts an identification strategy that exploits the distinct features of the German primary education system, thus allowing us to (arguably) come very close to the “true” teacher gender effect: In Germany, primary school students typically stay with the same teacher for two or more years. More than 50 percent of primary school students in our sample even stayed with the same teacher for four years (that is, their entire school career up to this point). In other words, the majority of students in our data set were taught by only one teacher per subject. In addition, many students at primary school are taught by the same teacher in all major subjects. As a result, we may attribute students’ learning process to that one teacher. To our knowledge, all previous studies encountered the problem that information on the gender or other characteristics of previous teachers was usually unavailable. In fact, in most instances, it was not even clear how long the teacher under examination had been teaching a student. This problem may partially account for the low degree of effects of teachers’ gender on students’ gender reported in previous studies. A second important feature of the German primary school system is the fact that school choice, for both teachers and students, is very limited. At primary school level, children are assigned to a school within their school district and may not freely choose an institution. Private schools are practically non-existent at this level. Furthermore, the procedure through which teachers apply to schools used to be very indirect. Up until the mid-1990s, new teachers were usually appointed by a central, state-level agency and then assigned to a school anywhere within the federal state in which they were applying (Treptow and Rothland 2005).⁵ As a result, the chance of being taught by a male versus female teacher should not depend on either the student’s or the school’s characteristics. Thus, student-teacher-gender pairing comes very close to a natural experiment.

⁴ Science (Sachunterricht) is a subject concerned with social, historical, scientific and geographical topics. The Sachunterricht curriculum suggests that this is neither a clear “boys’ domain” nor a clear “girls’ domain” but rather “gender neutral”.

⁵ This so-called Listeneinstellungsverfahren was changed in recent years in favor of a procedure that grants more agency and autonomy to schools (schulscharfes Verfahren). These reforms do not pertain to our 2001 data set, however, since the teachers in question were appointed between the 1960s and the 1990s.

4.2 Theoretical considerations and empirical evidence

The feminization of schooling discourse is by no means a new phenomenon in Western societies. In the United States, where the teaching profession became a predominantly female profession as early as in the late 19th century, initial efforts to recruit more men into teaching were made between the two world wars (Blount 2000; Martino 2008; Rury 1989). The main argument for doing so was the call for male role models, driven by an anxiety about the “feminization of boys”, who would be less able to develop their masculinity without appropriate role models. This line of argument also characterized discussions in the 1960s (Sexton 1969). In past as well as more recent contributions on the “feminization of schooling,” the main focus of study is on boys. It is predominantly boys, according to this point of view, whose development is negatively affected by the absence of male teachers and whose academic performance, as a result, is weaker than that of girls. This interpretation tends to forget, however, that the relative loss of boys may also be explained by the rising educational success of girls. Gender-specific changes in educational success, therefore, may just as well be the result of a positive impact of female teachers on girls. This possibility is often neglected, however.

Today, there are several lines of argument that consider the feminization of schooling to have a negative impact on the educational performance of boys. The first line of argument points to a *lack of male role models* in school (Budde 2008; Driessen 2007; Holmund and Sund 2008). From this point of view, the main complaint concerns the increasing absence of men in all stages of the education process, which results in a situation in which boys are unsure about their identity, lack clear patterns of gender role orientation (Budde 2008, p. 492), and, as a result, are unable to form positive ideas of masculinity (Bacher et al. 2008). This is closely related to the “same-sex hypothesis”, which is based on the idea that girls imitate women and that boys need men in order to eventually become men themselves (Powell and Downey 1997). Psychology-based learning theory assumes that children tend to observe and identify with same-sex agents of socialization (Hannover 2008; Kohlberg 1966). At the same time, gender identification needs to be viewed in a bipolar way. Teachers, according to this theory, are better able to identify with a child of their own gender, possibly because they feel more competent in responding to their problems, having experienced them themselves (Powell and Downey 1997). However, the link between gender identity and educational performance is inherently inconclusive (Blossfeld et al. 2009). And yet, scholars, educators, and policy-makers keep using it for explaining gender-specific educational achievement.

Another possibility for explaining the gender achievement gap is a phenomenon referred to as *stereotype threat* (cf. Steele 1997). According to this theory, students tend to perform less well when they fear they are being evaluated through the lens of a negative stereotype about the group they belong to. Teacher gender may enforce such negative stereotyping. If the math teacher is male, for example, girls may underperform simply because they believe math is a male domain. The negative stereotype is deactivated, however, when the teacher herself is female. Conversely, the same negative stereotype may be activated when boys take reading and writing classes with a female teacher.

Furthermore, there is some evidence that the *context in which a task is presented*, which may refer to the world of boys or the world of girls, leads to differences in the performance of boys and girls (Walther et al. 2008). Math problems, for example, may be presented in a female context (using girls' names, pets, or cakes) or in a male context (using boys' names, buildings, trains, or income). If female teachers generally tended to choose female contexts and male teachers tended to choose male contexts to frame curricular content, they would thereby favor students of their own gender.

If all or any of the aforementioned arguments holds true, (H1:) boys' academic skills should be higher when they were taught by a male teacher, and girls' academic skills should be higher when they were taught by a female teacher.

Empirically, Dee (2007) was able to show that boys and girls have higher skills when they were taught by a same-sex teacher. With school-level aggregate data, Helbig (2010) did find positive effects regarding the reading skills of girls at schools with a higher share of female teachers, but he could not confirm the same-sex hypothesis with regard to boys' reading skills and the math skills of both genders. Ammermüller and Dolton's (2006) findings for England suggest that boys at the age of 13 do benefit somewhat from male math teachers, whereas girls benefit from female English teachers. They did not find significant effects in the US, however. In addition, they did not find any positive same-sex teacher effects for students around the age of 9. Likewise, several other authors were unable to come up with convincing empirical evidence for the claim that students benefit from being taught by a same-sex teacher with regard to their cognitive skills (Driessen 2007; Ehrenberg et al. 1995; Sokal et al. 2007).

Another line of argument relates to the gender-specific evaluation of students, net of their actual performance. It is argued that female teachers may expect and reward certain types of behavior that girls have acquired in the process of socialization, whereas boys have not. While there is the possibility that female teachers actively discriminate against boys, the disadvantages that boys may experience from being taught by female teachers are more likely to be

unintended consequences of the actions of female teachers, who may *interpret and judge boys' and girls' behavior in different ways* (Diefenbach and Klein 2002). Types of behaviors that tend to disrupt school activities and negatively affect students' academic performance are more frequently found to originate from boys than from girls (Eagly and Chryvala 1986). Possibly, female teachers are more "upset" with such types of behaviors than male teachers because their standards of judgment are derived from their own gender-specific socialization. In other words, one might say there is a certain mismatch between the "habitus" of female teachers and the "habitus" of male students (Brandes 2002). According to this interpretation, feminization puts boys at a disadvantage especially when it comes to assessing their performance (Stamm 2008; Holmund and Sund 2008).

Gender influences on grades (net of performance) may also be caused by students' behavior instead of teacher evaluation. According to the same-sex hypothesis, students taught by same-sex teachers may be more 'well-behaved' (e.g. participation in school lessons, disruption during class, doing homework) than students taught by other-sex teachers. As a consequence, the fact that students tend to get better grades from same-sex teachers would result more from a better behavior on the part of the student than from a conscious or subconscious grading bias in the teacher.

Based on these theoretical implications, (H2:) the grades boys earn from male teachers should be better than those they earn from female teachers, whereas girls should get better grades from female teachers than from male teachers. Previous research offers mixed results in this regard, too. With respect to grading, Ehrenberg et al. (1995), Hopf and Hatzichristou (1999), and Helbig (2010) were able to show that boys receive inferior grades from female teachers. The correlations that were identified must be considered quite weak, however.⁶ What is more, a same-sex teacher advantage could not be confirmed by other studies, neither with regard to boys and girls (Carter 1952; Driessen 2007; Holmund and Sund 2008) nor with regard to girls (Helbig 2010).

In the following, we will assess whether girls or boys benefit from being taught by a same-sex teacher with respect to their academic performance (H1) and with respect to their grades, net of performance (H2).

⁶ Helbig (2010) analysed the effect of the ratio of male teachers at the school level on boys' and girls' grades. He found that an increase of the male teacher ratio in a school by 10 percent would raise the grades of boys in mathematics by less than 0.1 grade points. There were no effects of the ratio of male teachers on mathematics grades for girls or on German grades for both, boys and girls.

4.3 Data and variables

35 countries participated in the Progress in International Reading Literacy Study (PIRLS). This study was conducted by the International Association for the Evaluation of Educational Achievement (IEA) in 2001 to assess the reading literacy of nine- and ten-year-old students (Mullis et al. 2003; Mullis et al. 2002). Germany embedded this study within its national IGLU-E study, which focused on the reading and writing ability of fourth graders, as well as on their performance in math and science.⁷ The PIRLS background questionnaires that were administered to students, parents, teachers, and schools were also expanded in IGLU-E. These modifications required a second test day to gather the information without overburdening the students (Schwippert 2007). 12 out of 16 states in the Federal Republic of Germany participated in the extended PIRLS (IGLU-E). Data was collected in a two-stage stratified sampling design. First, participating schools were chosen. Within each school, a sample of classes from the targeted grade (fourth grade) was drawn. Within each class, in principle, all students were sampled. In practice, however, the number of sampled students was sometimes smaller than the actual class size because of student non-participation. After applying weights to compensate for non-proportionality and to ensure representativeness, the sample consisted of 5858 students from 166 schools (308 classes).⁸ The participation rate was satisfactorily high with 84% (details: Bos et al. 2004; Bos et al. 2003).⁹ In the data set, extensive information on student performance, as well as on home and school environment is available through student, parent, teacher, and school questionnaires.

Dependent Variables. To test whether boys and girls actually know more when being taught (for at least 2 years) by a same-sex teacher (H1), we utilize test scores in reading literacy, mathematical literacy, and scientific literacy. Tests were constructed and evaluated by experts to ensure broad coverage of third and fourth grade curricula in the respective subjects (details: Bos et al. 2003). In order to keep the test time to a reasonable length while at the same time

⁷ Access to the IGLU-E datasets has been kindly provided by the Research Data Centre at the Institute for Educational Progress (Institut zur Qualitätsentwicklung im Bildungswesen, IQB).

⁸ Note that the German sample for the international comparison consists of 8997 students in 211 schools from all 16 states. However, the states of Brandenburg, Mecklenburg-West Pomerania, Lower Saxony, and Saxony-Anhalt did not participate in the extension and hence do not provide information on Mathematics and Science, amongst other background information. We see no reason, why results might be biased by the exclusion of these 4 states, as they are very similar to the included 12 states in terms of school system characteristics, student performance and demographic composition including the feminisation of the teacher workforce.

⁹ The fact that the participation rate is somewhat lower than the international average is probably caused by the necessity of obtaining written parental consent.

ensuring a comprehensive coverage of the curricula, not all items were administered to all students. Instead, the test material was distributed among different booklet versions according to a multi-matrix design. Subsequently, within the framework of Item-Response-Theory, Maximum-Likelihood Estimates (MLE) were estimated which we employ as best estimates of the student performance (IPN 2005). To test whether boys and girls are assessed differently (or behave differently) when being taught by a same-sex teacher (H2), we utilize teacher grades in German, Mathematics, and Science as dependent variables, controlling for the respective test scores. Grades run from 6 (insufficient/fail) to 1 (very good). Information about individual student grades was retrieved from the teacher questionnaires.

Independent Variables. The independent variables in our study are gender dummies for the teacher and the student. We are able to identify 2434 male students (2389 female students) with female German teachers and 553 male students (482 female students) with male German teachers. The share of female teachers in our sample ranges from 79% to 85%, which is why the number of teacher-student pairings varies somewhat across subjects (cf. Table 2). At the primary school level, the same teacher often teaches several subjects per class. Accordingly, in many cases one and the same teacher had to fill out several separate subject teacher questionnaires. In order not to be redundant, some teachers provided the standard socio-demographic information, including information on gender, in only one of the questionnaires. When socio-demographic information was missing, we imputed gender and other socio-demographic items when we could make sure we were dealing with the identical teacher.¹⁰ In addition, we constructed a variable indicating whether a teacher was teaching all three subjects.

The remaining variables in this study represent controls for student, teacher, class, and school observables (cf. Table 1). To ensure that students' characteristics do not differ systematically between male and female teachers, we include at the student level parental socioeconomic status¹¹, students' migration background, number of children books at home, and students' age.¹² At the teacher/class level, we control for class size, share of male students in the class, and share of students with German language problems. For each teacher, we include age, number of years of teaching experience, a dummy indicating whether the teacher works part-time or full-time, and the amount of further training they engaged in while already working

¹⁰ We used a range of socio-demographic, teaching, and school related variables to identify identical teachers across the different teacher questionnaires.

¹¹ Socio-economic status is measured by the Highest International Socio-economic Index of Occupational Status (HISEI), which corresponds to the highest occupational index score of the student's father or mother (Ganzeboom et al. 1992).

¹² We also included a quadratic age term to capture non-linearities. The same procedure has been applied to teachers' age.

as a teacher.¹³ External evidence (GFSO 2009: own calculations) supports our data suggesting that female teachers are, on average, younger and less experienced and have a higher tendency to work part-time. Furthermore, female teachers tend to invest more in their further training (own auxiliary calculations based on IGLU-Data). We want to make sure that potential teacher-gender effects are not confounded by these factors. Next, to control for potential bias caused by the possibility that male and female teachers work at schools with different characteristics, we include school level controls. All school level information was provided through the school principal. We capture school resources by a dummy variable indicating whether or not a school library is available and by a factor score “shortage of teaching aids” based on a row of questions asking whether the lack of books, teaching materials, computers, audiovisual aids etc. constrains the school’s capacity for providing adequate instruction. We also include a factor score “special needs” which indicates whether a school provides assistance for low-achieving students and students with language deficiencies, or whether it offers additional programs for high-achieving students. In addition, we include the percentage of students from economically disadvantaged backgrounds. The location of the school is controlled in terms of urbanity (urban, suburban, rural) and through dummies for East and West German states (‘Länder’), between which the share of female teachers differs (GFSO 2009).

We utilize multiple imputation to address missing data, which replaces missing values with several plausible values based on information observed in the sample (Allison 2001; Rubin 1987). Most missing values were found for parental socio-economic status, for which 25 percent of the sample had unknown values. Even in the face of a large amount of missing values, Rubin (1987, p. 114) states that only a few imputations are required to obtain estimates with a relatively high efficiency. We generate five datasets through imputation by chained equations (ICE), which has recently been implemented in Stata (Royston 2004). All analyses are based on the imputed datasets, for which the results are combined to account for variation within and between imputed data sets.¹⁴

¹³ German teachers were asked how often they read material about instruction or didactics. Mathematics and science teachers were asked how many times within the last two years they participated in further education seminars in their subjects.

¹⁴ It has been shown that this procedure has advantages over listwise deletion in terms of obtaining unbiased estimators. For a recent overview see Graham (2009).

Table 1: Descriptive statistics (averaged over multiply imputed data sets, weights apply)

| | Girls (n=2871) | | | Boys (n=2987) | | |
|---|----------------|-------|-------------|---------------|-------|-------------|
| | Mean | SD | Min/Max | Mean | SD | Min/Max |
| <i>Performanceⁱ</i> | | | | | | |
| German | 511.9 | 97.9 | -83.4/896.8 | 498.4 | 95.9 | -78.0/904.8 |
| Mathematics | 494.6 | 98.3 | 48.1/916.6 | 511.5 | 98.5 | 42.1/916.6 |
| Science | 492.9 | 99.7 | -26.0/844.5 | 510.6 | 98.9 | 124.0/844.6 |
| <i>Grades</i> | | | | | | |
| German | 2.59 | 0.93 | 1/6 | 2.94 | 0.89 | 1/6 |
| Mathematics | 2.80 | 0.99 | 1/6 | 2.69 | 0.96 | 1/6 |
| Science | 2.48 | 0.90 | 1/6 | 2.53 | 0.87 | 1/6 |
| <i>fraction male teachers</i> | | | | | | |
| German | 0.16 | 0.37 | 0/1 | 0.17 | 0.38 | 0/1 |
| Mathematics | 0.22 | 0.41 | 0/1 | 0.22 | 0.41 | 0/1 |
| Science | 0.15 | 0.36 | 0/1 | 0.15 | 0.36 | 0/1 |
| <i>Student level controls</i> | | | | | | |
| Migration background | 0.25 | | 0/1 | 0.25 | | 0/1 |
| parental hisei | 49.59 | 14.99 | 16/90 | 49.78 | 15.19 | 16/90 |
| age ⁱⁱ | 10.48 | 0.47 | 8.92/13.08 | 10.58 | 0.52 | 8.5/13.25 |
| books (0-10) ⁱⁱⁱ | 5.77 | | | 8.29 | | |
| (11-25) | 17.43 | | | 20.09 | | |
| (26-50) | 33.00 | | | 32.31 | | |
| (51-100) | 27.21 | | | 23.80 | | |
| (more than 100) | 16.59 | | | 15.51 | | |
| <i>Class level controls</i> | | | | | | |
| Size | 22.86 | 3.92 | 9/32 | 22.58 | 4.02 | 9/32 |
| fraction male students | 0.48 | 0.11 | 0.15/0.88 | 0.53 | 0.11 | 0.15/1 |
| fraction language problems | 0.06 | 0.12 | 0/1 | 0.06 | 0.13 | 0/1 |
| German | | | | | | |
| teacher age ⁱⁱ | 48.25 | 9.71 | 26/63 | 48.62 | 9.66 | 26/63 |
| teacher experience | 23.26 | 11.59 | 1/42 | 23.80 | 11.59 | 1/42 |
| teacher further training ⁱⁱⁱ | 2.45 | 0.60 | 1/4 | 2.47 | 0.60 | 1/4 |
| fraction part-time | 0.38 | 0.49 | 0/1 | 0.38 | 0.49 | 0/1 |
| Mathematics | | | | | | |
| teacher age | 47.79 | 9.66 | 26/64 | 47.92 | 9.60 | 26/64 |
| teacher experience | 22.74 | 11.31 | 0/42 | 22.91 | 11.35 | 0/42 |
| teacher further train. (>5) | 6.14 | | | 4.74 | | |
| (3-4) | 14.71 | | | 15.12 | | |
| (1-2) | 39.48 | | | 39.04 | | |
| (0) | 39.67 | | | 41.10 | | |
| fraction part-time | 0.30 | 0.46 | 0/1 | 0.28 | 0.45 | 0/1 |
| Science | | | | | | |
| teacher age | 47.42 | 10.23 | 25/63 | 47.67 | 10.14 | 25/63 |
| teacher experience | 22.40 | 12.04 | 1/42 | 22.82 | 11.99 | 1/42 |
| teacher further train. (>5) | 3.73 | | | 3.22 | | |
| (3-4) | 8.89 | | | 10.96 | | |
| (1-2) | 31.44 | | | 28.89 | | |
| (0) | 55.94 | | | 56.93 | | |

| fraction part-time | 0.34 | 0.47 | 0/1 | 0.33 | 0.47 | 0/1 |
|---|-------|------|--------|-------|------|--------|
| <i>School level controls</i> | | | | | | |
| school library available | 0.51 | 0.50 | 0/1 | 0.50 | 0.50 | 0/1 |
| special needs | 0.46 | 0.29 | 0/1 | 0.46 | 0.30 | 0/1 |
| shortage of teaching aids | 2.02 | 0.63 | 0.29/4 | 2.03 | 0.66 | 0.29/4 |
| East Germany | 0.09 | 0.28 | 0/1 | 0.09 | 0.29 | 0/1 |
| % students econ. disadvant. backgrounds | 49.07 | | | 49.80 | | |
| (0-10) | | | | | | |
| (11-25) | 31.36 | | | 29.84 | | |
| (26-50) | 10.79 | | | 12.61 | | |
| (>50) | 8.78 | | | 7.75 | | |
| school location (urban) | 32.27 | | | 30.46 | | |
| (suburban) | 23.75 | | | 23.05 | | |
| (rural) | 43.98 | | | 46.49 | | |

ⁱ Performance scores vary up to very high or sometimes negative values. This is due to the maximum-likelihood-estimates, which are estimated based on a rotated test design within the framework of item-response-theory. This should not affect regression outcomes compared to a score normalized to some tighter bounds.

ⁱⁱ we also include a quadratic age term in the models to capture non-linearity.

ⁱⁱⁱ we display percentages for all categorical variables.

ⁱⁱⁱⁱ cf. footnote 13.

4.4 Results

We start out by documenting girls’ and boys’ performance scores and grades with same-sex and other-sex teachers across different subjects (Table 2). The share of female teachers differs across subjects (German: 83%, Mathematics: 78%, Science: 86%). Accordingly, the number of student-teacher-gender pairings differs across subjects as well.

The same-sex hypothesis states that students benefit from a same-sex teacher. Looking at girls, the raw data lends only weak support for this hypothesis. Girls with female teachers score between 8 and 16 points higher on performance tests than girls with male teachers, which equates roughly the winning margin of a ¼ school year (cf. Bos et al. 2004, p. 56). However, the effects are not statistically significant. They also receive better (i.e. numerically lower) grades; however, the effect is only significant in Mathematics. Looking at boys, the hypothesis cannot prevail. Neither performance scores nor grades differ significantly between boys with male teachers and boys with female teachers. In addition, the (insignificant) direction of the teacher-gender effect is reversed, i.e. boys receive inferior grades from male teachers than from female teachers. One explanation could be, that male teachers grade rather strictly compared to their female teacher colleagues, independent of the student’s gender. We tested this explicitly in a pooled dataset of girls and boys and found (insignificant) negative effects of male teachers

across all subjects and outcomes (table not shown).¹⁵ The direction of the effects suggests that there is no advantage of having a same-sex teacher for boys. If anything, male teachers have a consistent negative impact on both, boys and girls. However, such a negative male teacher effect (independent of student gender) is too weak to pass the test of significance.

Table 2: Students' performance and grades by student and teacher gender (averaged over multiply imputed data sets, weights apply)

| | Girls | | | Boys | | |
|--------------------------|------------------|-------------------|---------|------------------|-------------------|-------|
| | Same-sex teacher | Other-sex teacher | Diff. | Same-sex teacher | Other-sex teacher | Diff. |
| Performance Score | | | | | | |
| German | 513 | 505 | 8 | 498 | 498 | 0 |
| Mathematics | 496 | 488 | 8 | 515 | 510 | 5 |
| Science | 495 | 479 | 16 | 510 | 510 | 0 |
| Grade | | | | | | |
| German | 2.58 | 2.69 | -0.11 | 2.99 | 2.92 | 0.07 |
| Mathematics | 2.77 | 2.90 | -0.13** | 2.72 | 2.68 | 0.04 |
| Science | 2.48 | 2.53 | -0.05 | 2.62 | 2.52 | 0.10 |
| Sample size ⁱ | | | | | | |
| German | 2389 | 482 | | 553 | 2434 | |
| Mathematics | 2252 | 619 | | 650 | 2337 | |
| Science | 2486 | 385 | | 438 | 2549 | |

*** p<0.01, ** p<0.05, * p<0.1 (two-tailed T-Test, standard errors adjusted for class-level clustering).

ⁱ The sample sizes differ slightly across imputations because some of the teacher's gender has been imputed. We display the case numbers of the first imputation in the tables. All analyses, however, are based on all 5 imputations according to Rubin's (1987) rule.

In a next step, we run OLS-regressions of performance scores on teachers' gender separately for girls and boys (cf. specifications 1a and 1b, Table 3). We start with the empty model which naturally reproduces the descriptive figures from Table 2. The adjusted R-squared indicates that the teachers' gender explains virtually no variance in performance scores, neither for girls nor for boys. The same findings emerge when looking at grades (cf. specifications 5a and 5b, Table 4). Thus, across the three main subjects in primary school, and across objective performance measures (H1) as well as teachers' grades (H2), the 'same-sex teacher'-hypothesis has to be rejected.

¹⁵ Analysis can be obtained from the authors upon request. Another explanation could be that students do not learn as much from male teachers. We test this in the remainder of the study.

However, these results could be quite misleading due to selection effects or because the hypotheses only hold in subgroups of the population. Selection effects could exist when students – or factors influencing students – differ systematically between male and female teachers, creating a spurious relationship between teacher gender and the outcome of interest. In the following section, we model a range of checks to verify whether our initial findings are robust. First, we try to tackle potential selection effects that usually pose a threat to the validity of empirical findings based on non-experimental data. Second, we try to estimate whether the initial absence of a teacher-gender effect is caused by a non-linear effect across different subgroups. For example, the gender of the teacher may have a significant effect only if students are exposed to the same teacher in all subjects.

4.5 Robustness checks

4.5.1 Does selection mask a same-sex teacher effect?

First, suppose, for example, that male teachers are more likely to be assigned to boys with a propensity for lower achievement. In this case, the results in Table 2 could falsely suggest the absence of teacher-gender effects, because boys with a same-sex teacher would start out with a higher propensity for lower achievement, but would still benefit from their teachers. A comparison between test scores or grades of boys with same-sex teachers and boys with other-sex teachers would not reveal such teacher-gender effects.¹⁶ Without longitudinal data, there is no direct way to address this classical concern of selection. However, we can control empirically whether male and female teachers are assigned to students with certain ascriptive traits that are usually confounded with high or low achievement, such as parents' socio-economic status, cultural capital in the home, migration background, or students' age.¹⁷ If we hold such student traits constant and still find no evidence for a teacher-gender effect, we can most likely rule out the alternative explanation of biased student-teacher assignment.¹⁸ We apply this strategy in specifications 2a (girls) and 2b (boys) for performance (Table 3) across several subjects. For the sake

¹⁶ The same argument may be applied to girls and female teachers, of course.

¹⁷ In specifications not reported here, we also included intelligence test scores, which are supposed to be time-invariant. However, the time-invariance of intelligence assessment is controversial in psychological research and could be influenced by the teacher's gender as well. We therefore do not include intelligence test scores in the models shown. The inclusion of intelligence test scores does not change results regarding the teacher-gender effect.

¹⁸ Selection on unobservable characteristics would in principle still be possible, but given the wide range of covariates available to us, we consider this rather unlikely.

of clarity, we display only the main effect (that is the teacher gender effect).¹⁹ As expected, the inclusion of student level controls increases the proportion of explained variance in the models drastically.²⁰ The teacher-gender effects, however, do not change by and large, and remain insignificant.

For our models with grades as dependent variables (testing H2), we first control for test scores (specifications 6a and 6b, Table 4) to see whether the only significant teacher gender effect in Mathematics for grade for girls (.13, cf. Table 2) can be explained by actual performance differences or by teacher's evaluation. As can be seen from the table, the teacher gender effect partially disappears, thus the slight difference in grading reflects the fact that girls with male teachers perform slightly less well than girls with female teachers, and are, net of performance differences, perceived more negatively. However, this effect is not significant. Next, we include student level controls (specifications 7a / 7b), just as in the performance model testing H1. There is very little change in coefficients across specifications; male teacher effects are small in magnitude, and none become significant.

Second, the bivariate findings in Table 2 could be biased by the presence of teacher traits (e.g. age) or classroom traits (e.g. class size) that are associated with a teacher's gender. We know from auxiliary analysis that female teachers are, on average, younger and therefore less experienced than their male counterparts. They also have a higher tendency to work part-time and are more likely to engage in further education while already filling a teaching position.²¹ In specifications 3a/8a (girls) and 3b/8b (boys) we therefore control for observable teacher traits such as experience, amount of further education, employment status (part-time vs. full-time), and age. We also include a quadratic age term to capture non-linearity. In addition, we include the share of male students in class because we found that the share of male students in classes taught by a male teacher is slightly higher (53%) compared to classes taught by female teachers (48%). Again, results indicate that the teacher gender effect is not affected by these controls, neither in models where test scores are used as dependent variables (Table 3), nor in models where grades are used as dependent variables (Table 4). There is one exception: the effect of a male teacher for the reading performance of boys increases from -6.27 to -11.76 and

¹⁹ The full set of covariates can be seen in Table A1 (appendix), exemplarily for reading test score as an outcome.

²⁰ Interestingly, the adjusted R-squared increases most in models with reading literacy. This is because our control variables 'books at home' and migration background (which is highly correlated with language proficiency) are better predictors of reading literacy than of the other performance domains.

²¹ They do not differ according to their formal training, as all teachers have to obtain a teaching degree in order to be eligible for teaching.

becomes marginally significant. Note however, that the effect is negative and thus is inconsistent with a same-sex teacher benefit (i.e. boys should perform better with a male teacher). Given the weak level of significance one should be careful in interpreting this result. In any case there is no sign of a positive same-sex effect.

There might still be unobservable teacher traits (e.g. a teacher's socio-economic background) or class characteristics (e.g. classroom facilities) which confound our findings. Pooling the data on boys and girls, it is possible to control for all unobservable class and teacher traits by including class/teacher fixed effects (FE). Regarding the fixed effect regressions, we assume for the following structure:

$$y_a = x_a \beta_{g_a}^i + x_c \beta_{g_a}^c + g_a \gamma + g_t \delta_{g_a} + \varepsilon_c + \varepsilon_a \quad (1)$$

y_a is the outcome of interest, x_a are covariates on the individual, and x_c on the class (including teacher) level. Furthermore g_a denotes student's gender, g_t teacher's gender (let $g_t = 1$ for male teachers) and ε_a is an idiosyncratic (unobserved) error term that is assumed to be uncorrelated with all covariates. ε_c denotes (unobserved) class fixed effects - potentially correlated with the observed covariates. We identify our parameters by within class differences between students a and b:

$$\begin{aligned} \Delta y_{a,b} = y_a - y_b &= \begin{pmatrix} x_i \\ x_c \\ g_i \\ g_t \end{pmatrix}' \begin{pmatrix} \beta_{g_a}^i \\ \beta_{g_a}^c \\ \gamma \\ \delta_{g_a} \end{pmatrix} + \varepsilon_c + \varepsilon_a - \begin{pmatrix} x_b \\ x_c \\ g_b \\ g_t \end{pmatrix}' \begin{pmatrix} \beta_{g_b}^i \\ \beta_{g_b}^c \\ \gamma \\ \delta_{g_b} \end{pmatrix} - \varepsilon_c - \varepsilon_b \\ &= \begin{pmatrix} x_c \\ g_t \end{pmatrix}' \begin{pmatrix} \beta_{g_a}^c - \beta_{g_b}^c \\ \delta_{g_a} - \delta_{g_b} \end{pmatrix} + (g_a - g_b) \gamma + x_a \beta_{g_a}^i - x_b \beta_{g_b}^i + (\varepsilon_a - \varepsilon_b) \end{aligned} \quad (2)$$

From this last equation, we see that the class fixed effects cancel out. Additionally, coefficients of variables not varying within class are no longer identified. As a result, the effect of teacher gender itself cannot be identified any more, as it is part of the class FE. However, an interaction effect, i.e. the difference between the male teacher effect on boys and girls (or the sum of the same-sex teacher effect), is still identified under weaker assumptions than before.²² Now, if

²² The sum of the same-sex teacher effects for boys and girls is equivalent to the difference $\delta_{g_a} - \delta_{g_b}$ from equation (2). This result can be obtained by just defining g_t relative to g_a instead of as always one for male teachers. Also

there actually were a positive same-sex effect, the difference between a male teacher effect on boys and girls should be large and positive.

FE-Models for all subjects are presented in the last specification of Tables 3 and 4. There are 308 unique teachers/classes per subject associated with 5858 first-differenced observations. Results support what we observed so far. All effects have the same sign and magnitude as indicated by the differences between specifications (3b)-(3a) and (8b)-(8a), as summarized in Table A2 (Appendix). Interactions from the FE regressions are less than one standard deviation from what is indicated by standard OLS. Moreover, they are far from being significant except for Science where they pass the test at the 10% level.²³

Finally, a bias could occur if school characteristics such as school resources or the urbanity of a school location is associated with a teacher's gender. For example, it is conceivable that male teachers are more likely to find employment at less favourable schools. If, in addition, boys happen to be especially prone to suffer from unfavourable school conditions, the male teacher effect for boys would be biased downwards and a same-sex teacher effect could be masked. It needs to be emphasized that we do not expect such biases because, as mentioned above, the application procedure of teachers to schools used to be very indirect. Auxiliary analysis confirms the expectation that the likelihood of finding a male teacher assigned to a school with certain observable school characteristics (such as different kinds of resources, school location, and student composition) does not differ systematically by observable school characteristics (cf. Table A3 in Appendix). However, the likelihood of finding a male teacher does differ across the German states. This is because the share of female teachers is higher in former East Germany (GFSO 2009), which is a relict of the higher female labour market participation in the German Democratic Republic. The inclusion of a dummy to account for East/West differences, however, does not change our findings.

note that in our specification, for each subject, we have one teacher per class. In this respect, teacher and class fixed effects are identical.

²³ In the case of test scores, the loss of degrees of freedom is not sufficiently compensated by higher explanatory power of the model, thus leading to a lower adjusted R-squared.

Table 3: Estimated effect of a male teacher on test scores for different subjects by student gender (OLS & FE Regressions)

| | <i>Girls OLS</i> | | | <i>Boys OLS</i> | | | <i>Pooled Boys&Girls FEⁱ</i> | |
|-----------------------------|-------------------|----------------------------|-----------------|-----------------|----------------------------|-------------------|---|------------------|
| | (1a) | (2a) | (3a) | (1b) | (2b) | (3b) | (4) | |
| Male teacher | -7.65 (8.03) | Reading -9.42 (6.30) | -5.69 (7.31) | -0.36 (7.65) | Reading -6.27 (5.73) | -11.76* (6.84) | Reading | -2.87 (6.55) |
| Student level controls | - | x | x | - | x | x | x | x |
| Class/teach. level controls | - | - | x | - | - | x | x | x |
| Class/teach. fixed effects | - | - | - | - | - | - | - | x |
| Adj. R ² | 0.001 | 0.219 | 0.228 | 0.000 | 0.198 | 0.204 | Math | 0.167 |
| Male teacher | -8.38 (6.70) | -7.01 (4.85) | -4.27 (5.48) | 5.49 (6.44) | 3.73 (5.25) | 2.46 (6.01) | Math | 8.69 (6.64) |
| Student level controls | - | x | x | - | x | x | x | x |
| Class/teach. level controls | - | - | x | - | - | x | x | x |
| Class/teach. fixed effects | - | - | - | - | - | - | - | x |
| Adj. R ² | 0.001 | 0.131 | 0.140 | 0.000 | 0.124 | 0.127 | Science | 0.099 |
| Male teacher | -16.28 (10.64) | -12.03 (7.72) | -8.11 (8.81) | -0.49 (7.98) | 1.38 (6.05) | -0.66 (6.95) | Science | 15.52* (9.33) |
| Student level controls | - | x | x | - | x | x | x | x |
| Class/teach. level controls | - | - | x | - | - | x | x | x |
| Class/teach. fixed effects | - | - | - | - | - | - | - | x |
| Adj. R ² | 0.004 | 0.164 | 0.170 | 0.000 | 0.148 | 0.149 | Science | 0.124 |
| Observations | 2871 | 2871 | 2871 | 2987 | 2987 | 2987 | 5858 | 308 |
| Number of classes | 308 | 308 | 308 | 308 | 308 | 308 | | |

*** p<0.01, ** p<0.05, * p<0.1. Standard errors, adjusted for class-level clustering, are reported in parentheses. Models are averaged over multiply imputed data sets, weights apply. FE Regressions allow for differing coefficients of the control variables by student's gender, to reach the same flexibility as the separate gender models. The full set of covariates can be seen in table A1 (appendix), exemplarily for reading test scores as an outcome.

ⁱ The interaction effects denote the difference between the male teacher effect on boys and girls (or the sum of the same-sex teacher effect).

Table 4: Estimated effect of a male teacher on teachers' grades for different subjects by student gender (OLS & FE Regressions)

| | <i>Girls</i> | | | | <i>Boys</i> | | | | <i>Pooled Boys&Girls FE</i> |
|-----------------------------|-----------------------------|-----------------|-----------------|-----------------|----------------|----------------|----------------|----------------|---------------------------------|
| | (5a) | (6a) | (7a) | (8a) | (5b) | (6b) | (7b) | (8b) | (9) |
| Male teacher | German | | | | German | | | | German |
| | 0.11 (0.08) | 0.06 (0.08) | 0.08 (0.08) | 0.02 (0.08) | 0.07 (0.07) | 0.07 (0.07) | 0.11 (0.07) | 0.03 (0.08) | 0.01 (0.06) |
| | Reading lit. test score | - | X | X | X | X | X | X | X |
| | Student level controls | - | - | X | X | - | - | X | X |
| | Class/teach. level controls | - | - | - | X | - | - | - | X |
| | Class/teach. fixed effects | - | - | - | - | - | - | - | X |
| | Adj. R ² | 0.002 | 0.340 | 0.398 | 0.406 | 0.001 | 0.295 | 0.358 | 0.372 |
| | Math | | | | Math | | | | Math |
| | 0.13** (0.06) | 0.08 (0.05) | 0.08 (0.05) | 0.06 (0.06) | 0.04 (0.06) | 0.07 (0.06) | 0.09 (0.06) | 0.04 (0.07) | -0.03 (0.07) |
| | Math lit. test score | - | X | X | X | X | X | X | X |
| Student level controls | - | - | X | X | - | - | X | X | |
| Class/teach. level controls | - | - | - | X | - | - | - | X | |
| Class/teach. fixed effects | - | - | - | - | - | - | - | X | |
| Adj. R ² | 0.003 | 0.307 | 0.365 | 0.374 | 0.000 | 0.282 | 0.317 | 0.325 | |
| Male teacher | Science | | | | Science | | | | Science |
| | 0.05 (0.14) | -0.02 (0.12) | -0.02 (0.10) | -0.03 (0.09) | 0.10 (0.08) | 0.09 (0.08) | 0.10 (0.08) | 0.10 (0.09) | 0.15* (0.08) |
| | Science lit. test score | - | X | X | - | X | X | X | X |
| | Student level controls | - | - | X | X | - | - | X | X |
| | Class/teach. level controls | - | - | - | X | - | - | - | X |
| | Class/teach. fixed effects | - | - | - | - | - | - | - | X |
| | Adj. R ² | 0.002 | 0.204 | 0.302 | 0.322 | 0.004 | 0.181 | 0.265 | 0.285 |
| | Observations | 2871 | 2871 | 2871 | 2871 | 2987 | 2987 | 2987 | 2987 |
| | Number of classes | 308 | 308 | 308 | 308 | 308 | 308 | 308 | 308 |

*** p<0.01, ** p<0.05, * p<0.1. Standard errors, adjusted for class-level clustering, are reported in parentheses. Models are averaged over multiply imputed data sets, weights apply. FE Regressions allow for differing coefficients of the control variables by student's gender, to reach the same flexibility as the separate gender models.

ⁱ The interaction effects denote the difference between the male teacher effect on boys and girls (or the sum of the same-sex teacher effect).

4.5.2 Do subgroup differences mask a same-sex teacher effect?

So far, the only finding which is striking is the extent to which the match between teacher gender and student gender is unimportant. However, before negating a phenomenon many scholars and politicians believe in, we run an additional set of tests before conceding the case for a no same-sex teacher effect. An effect of the teachers' gender may become relevant, if the 'treatment' (that is the exposure to the same teacher) is stronger. In Germany, each primary school class usually has one teacher, who teaches most major subjects in this class and who stays with the same class for two or more years. The idea is to have a stable psychological parent for the students, who knows the students very well and is thus able to respond more closely to their personal difficulties. The downside is that potential difficulties, such as negative teacher-gender effects, are more prevalent and pertain to all subjects. We ran the full set of regressions on the subgroup of students who were taught by the same teacher in all subjects under investigation, i.e. German, Mathematics, and Science (49 percent of the sample). Table 5 displays our findings. Again, we must conclude that a teacher-gender effect is not detectable. There is only one marginally significant effect of having a male teacher for boys in mathematics when looking at grades. The direction of the effect, however, speaks against the same-sex teacher hypothesis.

Next, we ran the same set of regressions on students who were most likely taught by the same teacher for four years (i.e. we excluded students from schools for which the principle stated that students are usually taught by the same teacher for only 3 years or less). For this subgroup (56 percent of the sample), a negative effect of having a male teacher becomes evident when looking at German test scores (cf. Table 5). But this effect is around -21 points for both boys and girls, which speaks once more against a same-sex teacher advantage (with our coding one would expect a positive effect for boys).

In accordance with theoretical considerations in the same-sex literature, we hypothesised in H2 that a mismatch between the habitus of boys and the habitus of female teachers might be responsible for the academic underachievement of boys. It might be argued that this mismatch should be stronger for boys with a particularly pronounced male habitus. Since boys from migrant and low-SES backgrounds are commonly assumed to have a particularly pronounced male habitus (Connell 1987; Connell 1995; King 2005), these subgroups should be especially harmed by female teachers. We tested this hypothesis by running regressions on the subsample of students from low-SES backgrounds, i.e. the 25th quantile of the socio-economic status distribution as well as on students with a migration background (cf. Table 5). Once again,

results suggest a firm decline of our hypothesis. Across all outcomes (i.e. performance scores and grades in German, Mathematics, and Science), male teacher effects are insignificant for boys and for girls in both subgroups.

Table 5: Estimated effect of a male teacher on students’ test scores and grades for different subgroups (OLS Regressions)

| | Girls | | | | Boys | | | |
|---------------------|------------------------|----------------------|-------------------|----------------------|------------------------|----------------------|------------------|----------------------|
| | same teacher all subj. | same teacher 4 years | migrants | .25 low SES quantile | same teacher all subj. | same teacher 4 years | migrants | .25 low SES quantile |
| <i>Test scores</i> | | | | | | | | |
| German | | | | | German | | | |
| Male teacher | -7.80 (7.54) | -20.24* (10.98) | -11.94 (12.86) | 1.51 (17.05) | -8.36 (8.14) | -21.44** (9.66) | -5.54 (16.07) | -8.12 (13.69) |
| Adj. R ² | 0.244 | 0.207 | 0.179 | 0.174 | 0.216 | 0.201 | 0.190 | 0.116 |
| Math | | | | | Math | | | |
| Male teacher | -10.72 (7.28) | 10.06 (7.91) | 12.19 (12.56) | 14.24 (9.49) | 1.88 (7.69) | -0.74 (8.41) | -3.67 (9.97) | -2.22 (11.25) |
| Adj. R ² | 0.154 | 0.134 | 0.143 | 0.129 | 0.121 | 0.125 | 0.108 | 0.083 |
| Science | | | | | Science | | | |
| Male teacher | -11.65 (10.13) | -9.72 (13.54) | -12.20 (16.27) | -7.85 (16.99) | 1.40 (7.52) | -3.88 (10.62) | -4.91 (14.09) | -4.70 (12.61) |
| Adj. R ² | 0.171 | 0.174 | 0.164 | 0.163 | 0.165 | 0.144 | 0.138 | 0.096 |
| <i>Grades</i> | | | | | | | | |
| German | | | | | German | | | |
| Male teacher | 0.08 (0.09) | -0.16 (0.12) | 0.17 (0.14) | 0.12 (0.13) | 0.11 (0.10) | -0.07 (0.13) | 0.10 (0.14) | 0.06 (0.12) |
| Adj. R ² | 0.424 | 0.411 | 0.356 | 0.365 | 0.359 | 0.386 | 0.398 | 0.352 |
| Math | | | | | Math | | | |
| Male teacher | 0.06 (0.08) | 0.04 (0.07) | 0.15 (0.12) | 0.15 (0.13) | 0.14* (0.08) | -0.09 (0.09) | 0.02 (0.10) | 0.05 (0.11) |
| Adj. R ² | 0.373 | 0.366 | 0.315 | 0.306 | 0.335 | 0.340 | 0.306 | 0.334 |
| Science | | | | | Science | | | |
| Male teacher | 0.09 (0.09) | -0.25** (0.12) | 0.09 (0.16) | -0.03 (0.15) | 0.16 (0.10) | -0.04 (0.13) | 0.09 (0.19) | -0.02 (0.14) |
| Adj. R ² | 0.279 | 0.357 | 0.267 | 0.283 | 0.272 | 0.300 | 0.293 | 0.298 |
| Observations | 1414 | 1611 | 754 | 691 | 1464 | 1666 | 805 | 748 |

*** p<0.01, ** p<0.05, * p<0.1. Standard errors, adjusted for class-level clustering, are reported in parentheses. Models are averaged over multiply imputed data sets, weights apply. All models include student level and class/teacher level controls.

4.6 Summary and conclusion

In past decades, school systems in nearly all countries of the world have seen a rather unanticipated gender-specific change in educational achievement, with females gaining advantages over males. At the same time, the share of female teachers has grown in almost all countries of the western world. In this context, we examine the empirical evidence regarding one question increasingly addressed in educational research and politics alike: Can the growing “feminization” of the teacher profession explain the emerging disadvantage for boys in educational attainment? The claim that a lack of male teachers is responsible for the ‘boy crisis’ in education is highly popular internationally, both in some parts of academia and among the broader public. This is astonishing, given that there is hardly any solid empirical evidence to support the same-sex hypothesis.

After evaluating the evidence on this research question, we isolate a teacher’s gender effect by exploiting two features of the primary German educational system: a long ‘exposure time’ of students to the same teacher (at least two years) and a quasi-random assignment of students to teachers. We use large-scale data from IGLU-E/PIRLS and focus on different scholastic outcomes (‘gender-blind’ test scores and more subjective teacher’s grades) across different subjects (German, Mathematics, and Science). We find that boys do not benefit from male teachers and girls do not - at least not significantly - benefit from female teachers, neither with regard to their academic performance as measured by test scores, nor with regard to their grades. This non-existence of a teacher-gender effect is evident across the three major subjects in German primary school, namely German (a typical ‘female’ domain), Math (a typical ‘male’ domain), and Science (which has to be considered rather ‘gender neutral’ in the German primary school curriculum (cf. footnote 6)). In summary, we can say that the “feminization” of (primary) schooling has neither led to a weaker development of the academic skills of boys, nor has it improved the academic skills of girls. Grading has also been found to be independent of teacher gender. Children from migrant and low-SES families do not benefit academically from being taught by a same-sex teacher; neither do children who are taught by the same teacher in all subjects. In other words, the call for more male teachers to address boys’ educational and psycho-social needs in school is not based on sound empirical evidence, at least not with regard to primary schools. Furthermore, our results suggest that the unqualified call for more male teachers may have unintended consequences. Both boys and girls who were taught by a male German teacher for 4 years (56 percent of the sample) had significantly less well developed reading skills than students who were taught by a female German teacher for four years. More

research is needed on why this finding might be and in how far it can be generalized. At this point, we can only speculate why male German teachers who have taught a class for 4 years should have a negative effect on all students. It is conceivable that female teachers may have higher reading skills themselves, which they are better able to convey to their students. To our knowledge, studies on the gender-specific reading skills of German teachers do not exist, however. Nevertheless, calling for more male teachers involves the danger of unintentionally harming the reading skills of both boys and girls by having them taught by male teachers.

What is most striking about our findings is the extent to which the match between teacher gender and student gender is irrelevant.²⁴ In this context, it is important to carefully consider the degree to which our findings may be generalized. First, there might be teacher gender effects at a later student age, i.e. during adolescence. This may help to explain why Dee (2007) found a correlation between the academic achievement of US secondary school students and same-sex teachers. Second, it is possible that boys who grow up without a male role model at home do benefit from having male teachers. This hypothesis could not be tested with the data available.

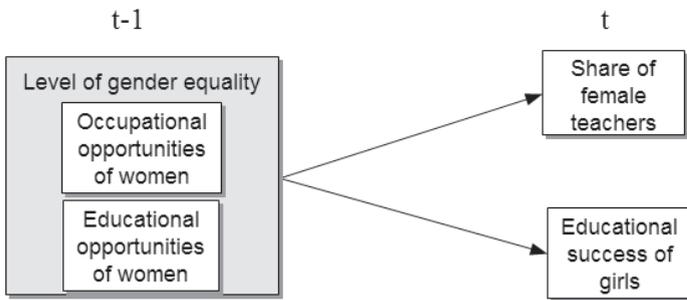
One additional aspect of this paper seems to require further explanation, however. Why is it that the correlation between the share of female teachers and gender-specific educational achievement, as clearly shown to exist at the country level in Figure 1, cannot be supported with empirical evidence at the individual level? We would like to argue that this puzzle results from a classical ecological fallacy. To find out about this, we need to ask how cross-country differences regarding the percentage of female teachers came about in the first place, and why the total share of female teachers has risen over past decades. Two aspects of the teaching profession become evident when comparing it across countries: Firstly, completing higher education is a prerequisite for gaining teacher licensure in most countries. Secondly, teaching is a female sex-segregated profession in literally all advanced societies, while it is male dominated in developing countries (most visibly in African countries, cf. figure 1). In the latter countries, women do not obtain higher education and are consequently not able to work as teachers – thus the share of female teachers is low. In addition, recent research has shown, that once the level of female tertiary enrolment begins to rise in a given country, having a higher percentage of women in female sex-segregated academic professions is an almost automatic consequence, as long as female preferences for traditionally female occupations do not change

²⁴ This is not to say that teacher effects do not exist. In fact, many studies show that a teacher explains large shares of the variance in student outcomes (Hanushek and Rivkin 2006). However, the important factor is probably the quality of teaching rather than the gender of the person delivering it (Marsh et al. 2008).

(Charles and Bradley 2009). The argument is that in advanced industrial societies, it is more ‘culturally legitimate’ to pursue self-realization. As such, men and women are free to express their ‘gendered selves’, when choosing a field of study and an occupation. As a consequence, the gender-segregation across academic fields (such as education) and subsequently across occupations (such as teaching) intensifies. This means that the current share of female teachers at school may be deduced from women’s educational opportunities in past decades and their opportunities of participating in the labour market today - two indicators of general gender equality (figure 2).

But why are girls and women today academically more successful in countries, in which the gender equality has been higher in the past? First of all, it is conceivable that socio-cultural and institutional factors that have helped women in past decades to become successful in the academic labour market are beneficial for girls and women today. For example, a declining discrimination in the labour market (possibly because ‘female’ qualities are appreciated more in societies with a rising service sector and a decreasing industrial sector) may give rise to girls’ and women’s aspirations. As girls and women think about their expectations to succeed in education, the prevailing labour market opportunities for women in their country influences whether they expect to complete an academic degree and obtain an academic job position (cf. McDaniel 2010).

Figure 2: Relationship of gender equality with the share of female teachers and the educational success of girls



Source: Own illustration

This argument is closely related to the expectation, that in countries where women have experienced relatively higher gender equality in the past, girls have more role models who demonstrate: females can be successful in the labour market if they are successful in the education system (Mickelson 1989). This may very likely encourage females to appreciate

educational outcomes (Buchmann and DiPrete 2006) and utilize their academic potential. Evidence of the fact that female employment levels and general gender equity have an impact on gender-specific educational achievement could be provided by several authors (Baker and Jones 1993; Else-Quest et al. 2010; Guiso et al. 2008; McDaniel 2010; Riegle-Crumb 2005). The so called “gender-stratification hypothesis” posits, that gender differences in some educational outputs, like mathematic competencies, are closely related to cultural variations in opportunity structures for girls and women.

In sum, not the gender of the teacher is relevant for the increasing educational success of girls. Rather, equal gender opportunities in a given country encourage girls to realize their academic potential, while at the same time making it more likely for women to become teachers.

Acknowledgements

We wish to thank Walter Müller, Zerrin Salikutluk, Steffen Schindler, and an anonymous reviewer for helpful comments on an earlier version of this article.

4.7 Appendix

Table A1: Effect of a male teacher on girls and boys reading test scores (OLS Regressions), all control variables shown

| | Girls – reading test score | | | Boys – reading test score | | |
|-------------------------------------|----------------------------|-------------------------|-------------------------|---------------------------|------------------------|------------------------|
| | (1a) | (2a) | (3a) | (1b) | (2b) | (3b) |
| Male teacher | -7.65 (8.03) | -9.42 (6.30) | -5.69 (7.31) | -0.36 (7.65) | -6.27 (5.73) | -11.76* (6.84) |
| <i>Student level controls</i> | | | | | | |
| age | | 426.50*** (101.87) | 427.80*** (105.10) | | 347.00*** (102.77) | 321.50*** (97.45) |
| age ² | | -20.67*** (4.78) | -20.71*** (4.94) | | -17.00*** (4.82) | -15.75*** (4.54) |
| books (0-10) | | -37.30*** (11.40) | -35.59*** (11.16) | | -38.78*** (10.00) | -37.28*** (10.03) |
| books (11-25) | | -19.99*** (6.64) | -19.38*** (6.51) | | -16.85*** (5.33) | -15.72*** (5.37) |
| books (26-50) | | <i>ref.</i> | <i>ref.</i> | | <i>ref.</i> | <i>ref.</i> |
| books (51-100) | | 20.59*** (4.79) | 21.17*** (4.72) | | 22.00*** (5.26) | 22.83*** (5.28) |
| books (>100) | | 44.43*** (5.73) | 44.75*** (5.65) | | 41.22*** (5.82) | 41.80*** (5.84) |
| parental hisei | | 1.25*** (0.15) | 1.25*** (0.15) | | 0.97*** (0.17) | 0.94*** (0.17) |
| migration backgr. | | -36.83*** (6.04) | -33.75*** (5.77) | | -23.22*** (5.45) | -22.94*** (5.50) |
| <i>Class/teacher level controls</i> | | | | | | |
| % male stud. | | | 37.36* (21.81) | | | 19.12 (20.72) |
| % language problems | | | -37.59 (25.62) | | | -54.09* (30.98) |
| class size | | | 0.25 (0.69) | | | 0.14 (0.66) |
| teacher further training | | | -4.55 (4.37) | | | -6.90 (4.87) |
| teach. age | | | -1.43 (2.23) | | | 0.86 (2.50) |
| teach. age ² | | | -0.00 (0.03) | | | -0.01 (0.03) |
| teach. experience | | | 1.04* (0.60) | | | 0.02 (0.48) |
| teach. parttime | | | 7.11 (5.35) | | | -5.09 (5.34) |
| Constant | 513.13*** (3.63) | -1741.86*** (543.67) | -1716.28*** (567.79) | 498.44*** (3.36) | -1311.82** (549.75) | -1189.94** (535.08) |
| Observations | 2871 | 2871 | 2871 | 2987 | 2987 | 2987 |
| No. of classes | 308 | 308 | 308 | 308 | 308 | 308 |
| Adj. R ² | 0.001 | 0.219 | 0.228 | 0.000 | 0.198 | 0.204 |

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, adjusted for class-level clustering, are reported in parentheses. Models are averaged over multiply imputed data sets, weights apply.

Table A2: Differences between specifications for boys and girls compared to interaction effects

| | Test scores | | | | Grades | | | |
|---------|-----------------|-------------------|-------------|------------------|-----------------|----------------|-------------|-----------------|
| | (3a) | (3b) | (3b) - (3a) | (4) | (8a) | (8b) | (8b) - (8a) | (9) |
| Reading | -5.69 (7.31) | -11.76* (6.84) | -6.07 | -2.87 (6.55) | 0.02 (0.08) | 0.03 (0.08) | 0.01 | 0.01 (0.06) |
| Math | -4.27 (5.48) | 2.46 (6.01) | 6.73 | 8.69 (6.64) | 0.06 (0.06) | 0.04 (0.07) | -0.02 | -0.03 (0.07) |
| Science | -8.11 (8.81) | -0.66 (6.95) | 7.45 | 15.52* (9.33) | -0.03 (0.09) | 0.10 (0.09) | 0.12 | 0.15* (0.08) |

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Figures of specifications are drawn from table 3 (test scores) and table 4 (grades). 'Differences between' by own calculations.

Table A3: Likelihood of schools to have a male teacher (Logistic Regressions)

| | male German teacher OR (S.E.) | male Math teacher OR (S.E.) | male Science teacher OR (S.E.) |
|----------------------------------|-------------------------------------|-----------------------------------|--------------------------------------|
| library | 0.58 (0.19) | 0.90 (0.28) | 0.65 (0.27) |
| special needs | 1.53 (0.91) | 0.91 (0.51) | 1.10 (0.76) |
| shortage of teaching aids | 0.71 (0.18) | 0.67* (0.19) | 0.79 (0.22) |
| fraction disadv. students 0-10% | <i>ref.</i> | <i>ref.</i> | <i>ref.</i> |
| fraction disadv. students 11-25% | 0.70 (0.28) | 1.07 (0.37) | 0.66 (0.29) |
| fraction disadv. students 26-50% | 0.99 (0.50) | 0.76 (0.46) | 1.01 (0.65) |
| fraction disadv. students >50% | 0.75 (0.46) | 1.10 (0.67) | 0.77 (0.63) |
| rural | <i>ref.</i> | <i>ref.</i> | <i>ref.</i> |
| suburban | 0.84 (0.34) | 0.91 (0.38) | 0.87 (0.40) |
| urban | 0.74 (0.30) | 0.92 (0.37) | 1.04 (0.58) |
| Observations | 308 | 308 | 308 |
| Pseudo R-squared | 0.025 | 0.013 | 0.012 |

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are reported in parentheses. Models are averaged over multiply imputed datasets.

4.8 Literature

- ABC-News. 2004. Scholarships to Target Declining Male Teacher Numbers. <http://www.abc.net.au/news/stories/2004/05/03/1100090.htm> (accessed: 08.07.2010).
- Allison, Paul D. 2001. *Missing Data*. Thousand Oaks, CA: Sage.
- Ammermüller, Andreas, and Peter Dolton. 2006. Pupil-Teacher Gender Interaction Effects on Scholastic Outcomes in England and the USA. *ZEW Discussion Paper* 06-060: 1-32.
- Bacher, Johann, Martina Beham and Norbert Lachmayr. 2008. *Geschlechterunterschiede in der Bildungswahl*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baker, David P., and Deborah Perkins Jones. 1993. Creating Gender Equality: Cross-National Gender Stratification and Mathematical Performance. *Sociology of Education* 66: 91-103.
- BBC News. 2000. *Male Teachers for 'Role Models'*. http://news.bbc.co.uk/2/hi/uk_news/education/893313.stm (accessed: 08.07.2010).
- Blossfeld, Hans Peter, Wilfried Bos, Bettina Hannover, Dieter Lenzen, Detlef Müller-Böling, Manfred Prenzel and Ludger Wößmann. 2009. *Geschlechtsdifferenzen im Bildungssystem. Jahresgutachten 2009 Aktionsrat Bildung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Blount, Jackie. 2000. Spinsters, Bachelors and other Gender Transgressors in School Employment, 1850-1990. *Review of Educational Research* 70: 83-101.
- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin and Gerd Walther. 2004. *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walther, Renate Valtin and Gerd Walther. 2003. *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Brandes, Holger 2002. *Der männliche Habitus*. Opladen: Leske+Budrich.
- Buchmann, Claudia, and Thomas A. DiPrete. 2006. The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement. *American Sociological Review* 71: 515-541.
- Budde, Jürgen 2008. *Bildungs(miss)erfolge von Jungen und Berufswahlverhalten bei Jungen / männlichen Jugendlichen*: BMBF, Referat Chancengerechtigkeit in Bildung und Forschung.
- Carter, Robert S. 1952. How invalid are marks assigned by teachers. *Journal of Educational Psychology* 43: 218-228.
- Charles, Maria, and Karen Bradley. 2009. Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries. *American Journal Of Sociology* 114: 924-976.
- Cole, Nancy S. 1997. *The ETS Gender Study: How Females and Males Perform in Educational Settings*. Princeton, NJ: Educational Testing Service.
- Connell, Robert W. 1987. *Gender and Power. Society, the Person, and Sexual Politics*. Stanford: Stanford University Press.
- Connell, Robert W. 1995. *Masculinities*. Berkeley: University of California Press.
- Dammasch, Frank 2007. *Jungen in der Krise. Das schwache Geschlecht? Psychoanalytische Überlegungen*. Frankfurt a. M.: Brandes & Apsel

- Dee, Thomas S. 2007. Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources* 42: 528-554.
- Diefenbach, Heike, and Michael Klein. 2002. "Bringing Boys Back In". Soziale Ungleichheit zwischen den Geschlechtern im Bildungssystem zuungunsten von Jungen am Beispiel der Sekundarschulabschlüsse. *Zeitschrift für Pädagogik* 48: 938-958.
- Driessen, Geert. 2007. The Feminization of Primary Education: Effects of Teachers' Sex on Pupil Achievement, Attitudes and Behaviour. *International Review of Education* 53: 183-203.
- Eagly, Alice H., and Carole Chryala. 1986. Sex Differences in Conformity: Status and Gender Role Interpretations. *Psychology of Women Quarterly* 10: 203-220.
- Ehrenberg, Ronald G., Daniel G. Goldhaber and Dominic J. Brewer. 1995. Do Teachers' Race, Gender and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review* 48: 547-561.
- Else-Quest, Nicole M., Janet S. Hyde and Marcia C. Linn. 2010. Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. *Psychological Bulletin* 136: 103-127.
- Eurostat. 2009. Eurostat Database. <http://epp.eurostat.ec.europa.eu/portal/page/portal/education/data/database> (accessed: 08.07.2010).
- Ganzeboom, Harry B. G., Paul M. De Graaf, Donald J. Treiman and Jan De Leeuw. 1992. A Standard International Socio-Economic Index of Occupational Status. *Social Science Research* 21: 1-56.
- German PISA Consortium. 2007. *PISA. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- GFSO. 2009. *Fachserie 11.1 - Allgemeinbildende Schulen Schuljahr 2008/2009*. German Federal Statistic Office. Wiesbaden.
- Graham, John W. 2009. Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology* 60: 549-576.
- Guiso, Luigi, Ferdinando Monte, Paola Sapienza and Luigi Zingales. 2008. Culture, Gender, and Math. *Science* 320: 1164-1165.
- Hannover, Bettina. 2008. Vom biologischen zum psychologischen Geschlecht: Die Entwicklung von Geschlechtsunterschieden. In *Lehrbuch der pädagogischen Psychologie*, Ed. Alexander Renkl, 339-388. Bern: Huber.
- Hanushek, E. A., and S. G. Rivkin. 2006. Teacher Quality. In *Handbook of Economics of Education*. 2nd Edition, Eds. Eric A. Hanushek, Finis Welch, 1053-1078. Amsterdam: Elsevier.
- Hedges, Larry V., and Amy Nowell. 1995. Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals. *Science* 269: 41-45.
- Helbig, Marcel. 2010. Sind Lehrerinnen für den geringeren Schulerfolg von Jungen verantwortlich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62: 93-111.
- Holmlund, Helena, and Krister Sund. 2008. Is the Gender Gap in School Performance affected by the Sex of the Teacher? *Labour Economics* 15: 37-53.
- Hopf, Diether, and Chryse Hatzichristou. 1999. Teacher Gender-Related Influences in Greek Schools. *British Journal of Educational Psychology* 69: 1-18.

- IES. 2009. NAEP Data Explorer. <http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx> (accessed: 08.07.2010).
- IPN. 2005. Research area 5 - Educational monitoring and research methodology. Research Report 2003/2004. Kiel: IPN – Leibniz Institute for Science and Mathematics Education.
- King, Vera. 2005. Bildungskarrieren und Männlichkeitsentwürfe bei Adoleszenten aus Migrantenfamilien. In *Männliche Adoleszenz. Sozialisation und Bildungsprozesse zwischen Kindheit und Erwachsensein*, Eds. Vera King, Karin Flaake, 57-76. Frankfurt a. M.: Campus.
- Kohlberg, Lawrence. 1966. A Cognitive-Developmental Analysis of Children's Sex Role Concepts and Attitudes. In *The development of sex differences*, Ed. Eleanor E. MacCoby, 82-172. Stanford: Stanford University Press.
- Marsh, Herbert W., Andrew J. Martin and Jacqueline H. S. Cheng. 2008. A Multilevel Perspective on Gender in Classroom Motivation and Climate: Potential Benefits of Male Teachers for Boys? *Journal of Educational Psychology* 100: 78-95.
- Martino, Wayne J. 2008. Male Teachers as Role Models: Addressing Issues of Masculinity, Pedagogy and the Re-Masculinisation of Schooling. *Curriculum Inquiry* 38: 189-223.
- McDaniel, Anne. 2010. Cross-National Gender Gaps in Educational Expectations: The Influence of National-Level Ideology and Educational Systems. *Comparative Education Review* 54: 27-50.
- Mickelson, Roslyn Arlin. 1989. Why Does Jane Read and Write so Well? The Anomaly of Women's Achievement. *Sociology of Education* 62: 47-63.
- Mullis, Ina V.S., Michael O. Martin, Ann M. Kennedy and Cheryl L. Flaherty (Eds.). 2002. *PIRLS 2001 Encyclopedia*. Boston: PIRLS International Study Center.
- Mullis, Ina V.S., Michael O. Martin, Eugenio J. Gonzalez and Ann M. Kennedy. 2003. *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools*. Chestnut Hill, MA: Boston College.
- OECD. 2009. *OECD Stat Extracts*. <http://stats.oecd.org/> (accessed: 08.07.2010).
- Pollack, William S. 2006. The "War" for Boys: Hearing "Real Boys" Voices, Healing their Pain. *Professional Psychology - Research and Practice* 37: 190-195.
- Powell, Brian, and Douglas B. Downey. 1997. Living in Single-Parent Households: An Investigation of the Same-Sex Hypothesis. *American Sociological Review* 62: 521-539.
- Quenzel, Gudrun, and Klaus Hurrelmann. 2010. Geschlecht und Schulerfolg: Ein soziales Stratifikationsmuster kehrt sich um. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62: 61-91.
- Riegle-Crumb, Catherine. 2005. The Cross-National Context of the Gender Gap in Math and Science. In *The Social Organization of Schooling*, Eds. Larry V. Hedges, Barbara Schneider, 227-243. New York: Russell Sage Foundation.
- Royston, Patrick. 2004. Multiple Imputation of missing Values. *Stata Journal* 4: 227-241.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Rury, John L. 1989. Who Became Teachers? The Social Characteristics of Teachers in American History. In *American Teachers: Histories of a Profession at Work*, Ed. Donald Warren, 9-48. New York: Macmillan.

- Schwippert, Knut. 2007. The Impact of PIRLS in Germany. In *Progress in Reading Literacy: The Impact of PIRLS 2001 in 13 Countries*, Ed. Knut Schwippert, 93-108. Münster: Waxmann.
- Sexton, Patricia C. 1969. *The Feminized Male: Classrooms, White Collars, & the Decline of Manliness*. New York: Random House.
- Sokal, Laura, Herb Katz, Les Chaszewski and Cecilia Wojcik. 2007. Good-bye, Mr. Chips: Male Teacher Shortages and Boys' Reading Achievement. *Sex Roles* 56: 651-659.
- Sommers, Christina H. 2000. *The War Against Boys: How Misguided Feminism Is Harming Our Young Men*. New York: Touchstone.
- Stamm, Margrit. 2008. Underachievement von Jungen: Perspektiven eines internationalen Diskurses. *Zeitschrift für Erziehungswissenschaft* 11: 106-124.
- Steele, Claude M. 1997. A Threat in the Air: How Stereotypes Shape the Intellectual Identities and Performance of Women and African-Americans. *American Psychologist* 52: 613-629.
- TDA. 2009. Wanted: Male Primary School Teachers. Training and Development Agency for Schools (TDA). www.tda.gov.uk/about/mediarelations/2009/200909.aspx (accessed: 08.07.2010).
- Treptow, Eva, and Martin Rothland. 2005. „Jeder Schule ihre Lehrer“? - Empirische Befunde zum Auswahlverfahren bei der Besetzung von Lehrerstellen in Nordrhein-Westfalen. *Zeitschrift für Erziehungswissenschaft* 8: 305-320.
- UNESCO. 2009. *Global Education Digest 2009: Comparing Education Statistics Across the World*. Montreal: UNESCO Institute for Statistic.
- UNESCO. 2010. Unesco Institute for Statistic - Key Statistical Tables for Education. http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx?IF_ActivePath=P,50&IF_Language=eng (accessed: 08.07.2010).
- Wagemaker, Hans, Karin Taube, Ingrid Munck, Georgia Kontogiannopoulou-Polydorides and Michael Martin. 1996. *Are Girls Better Readers? Gender Differences in Reading Literacy in 32 Countries*. Amsterdam: IEA.
- Walther, Gerd, Knut Schwippert, Eva-Maria Lankes and Tobias C. Stubbe. 2008. Können Mädchen doch rechnen? Vertiefende Analysen zu Geschlechtsdifferenzen im Bereich Mathematik auf Basis der Internationalen Grundschul-Lese-Untersuchung IGLU. *Zeitschrift für Erziehungswissenschaft* 11: 30-46.
- Willingham, Warren W., and Nancy S. Cole. 1997. *Gender and Fair Assessment*. Mahwah, NJ: Lawrence Erlbaum.

Kapitel 5 Werden Jungen von Lehrerinnen bei den Übergangsempfehlungen für das Gymnasium benachteiligt? Eine Analyse auf Basis der IGLU-Daten.

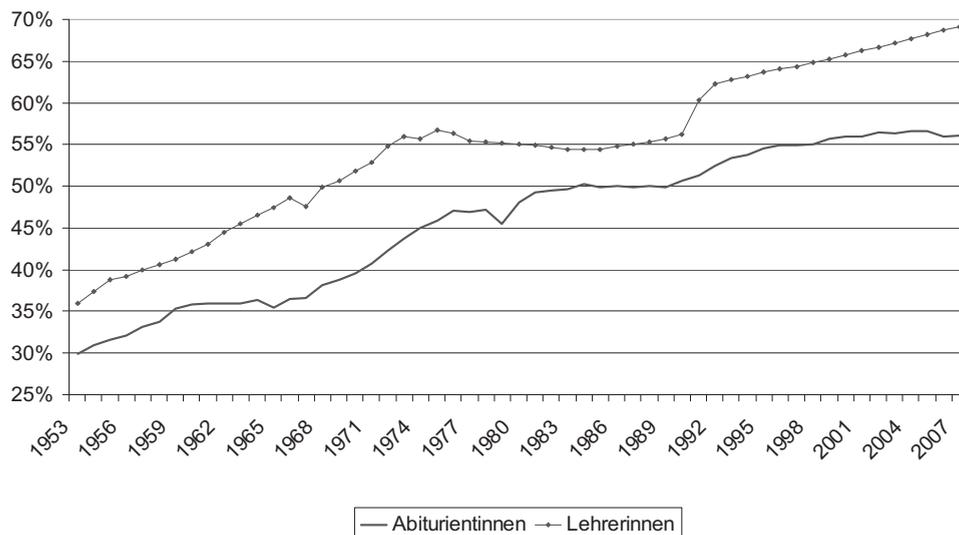
Erschienen als: Neugebauer, Martin. 2011. Werden Jungen von Lehrerinnen bei den Übergangsempfehlungen für das Gymnasium benachteiligt? Eine Analyse auf Basis der IGLU-Daten. In *Geschlechtsspezifische Bildungsungleichheiten*, Hrsg. Andreas Hadjar, 235-260. Wiesbaden: VS Verlag für Sozialwissenschaften. Abdruck mit Erlaubnis von Springer Science+Business Media.

5.1 Einleitung

In vielen schulischen Aspekten sind Mädchen heute erfolgreicher als ihre männlichen Altersgenossen. Sie bleiben seltener sitzen und seltener ohne Ausbildungsabschluss, erzielen dafür höhere Lesekompetenzen und bessere Schulnoten. Auch beim Übergang auf das Gymnasium sind Mädchen erfolgreicher als Jungen und erlangen in der Folge die höheren Bildungszertifikate (Blossfeld et al. 2009). Das war nicht immer so: Trendstatistiken zeigen, dass es in den Schulsystemen fast aller Länder der Welt in den letzten Dekaden zu einem geschlechtsspezifischen Wandel von Bildungserfolg gekommen ist (UNESCO 2009). Dieser Trend ist auch für Deutschland erkennbar. Für die Zeit ab 1953 ist in Abbildung 1 anhand der Abiturabschlüsse illustriert, wie die Mädchen die Jungen überholt haben. Im Jahr 1953 waren nur etwa 30 Prozent der Abiturienten weiblich. Dieser Anteil stieg relativ kontinuierlich und erreichte Anfang der 1980er Jahre die 50 Prozent Marke. Bis etwa 1990 erlangten ebenso viele Mädchen wie Jungen das Abitur. Seit den neunziger Jahren schließen mehr Mädchen als Jungen das Gymnasium erfolgreich ab; im Jahr 2007 betrug der Anteil der Abiturientinnen 56 Prozent. Angesichts dieser Zahlen wurden in den letzten Jahren kritische Stimmen laut, die eine „Krise der Jungen“ (Damasch 2007) oder sogar einen „war against boys“ (Sommers 2000) beklagen. Die verbesserte Situation der Frauen im Bildungssystem wird als Bildungsmisserfolg des männlichen Geschlechts interpretiert. In den wissenschaftlichen und gesellschaftspolitischen Debatten über die Ursachen dieser „Jungenkrise“ wird immer wieder auf die Feminisierung des Lehrerberufs verwiesen (Arnot et al. 1999; Dee 2007; Diefenbach und Klein 2002; Driessen 2007; Hannan 2001; Horstkemper 1999; Preuss-Lausitz 2005; Rose und Schmauch 2005; Sexton 1969). Die Benachteiligung der Jungen, so die These, sei kaum verwunderlich, wenn im Lehrberuf fast ausschließlich Frauen anzutreffen sind. „Das schlechtere Abschneiden von Jungen liegt unter

anderem daran, dass Kindergärten und Schulen weiblich dominiert sind“ (Bundesfamilienministerin Kristina Schröder in einem Interview in „Die Zeit“ vom 22.04.2010). Die Feminisierung, so die Argumentation, bringe den Jungen vor allem bei der Bewertung von Schulleistungen Nachteile. Weil Lehrerinnen den Jungen weniger Verständnis entgegenbrächten und geringere Erwartungen an sie hätten, würden sie Jungen auch seltener für das Gymnasium empfehlen (vgl. Stamm 2008). Auf der Aggregatebene scheint die Feminisierungsthese plausibel zu sein. Gleichzeitig zu dem gestiegenen Schulerfolg der Mädchen gegenüber den Jungen ist in fast allen Ländern der westlichen Welt auch der Frauenanteil in der Lehrerschaft angestiegen (Eurostat 2009). Zur Illustration ist für Deutschland in Abbildung 1 neben den Abiturientinnenanteilen der nahezu parallel verlaufende Anstieg der Lehrerinnenanteile abgetragen, der von 1953 bis 2007 von 36 Prozent auf 69 Prozent gewachsen ist. Anfang der 1980er Jahre verflachen beide Kurven, ab etwa 1990 steigen sie wieder steiler an.

Abbildung 1: Anteil der Abiturientinnen und Lehrerinnen in allgemeinbildenden Schulen (1953-2007)



Anmerkung: Der Lehrerinnenanteil bezieht sich auf den Frauenanteil unter allen hauptberuflichen Lehrkräften (Voll- und Teilzeitbeschäftigt) in allgemeinbildenden Schulen. Dies umfasst Grund- und Hauptschulen (bis 1973 als Volksschulen bezeichnet, Hauptschulen bis 2006 inkl. schulartunabhängiger Orientierungsstufen), Real-/Mittelschulen, Gymnasien, Gesamtschulen und Schularten mit mehreren Bildungsgängen (SMBG) und Abendschulen. Bis 1990 Früheres Bundesgebiet; 1991 Deutschland ohne Brandenburg und Sachsen-Anhalt; ab 1992 Gesamtdeutschland.

Quelle: Statistische Jahrbücher, verschiedene Jahrgänge, eigene Berechnungen.

Ob nun aber in der „Feminisierung der Schule“ die (kausale) Ursache des niedrigen Schulerfolgs der Jungen und des hohen Schulerfolgs der Mädchen liegt, ist bislang kaum untersucht. Dennoch versucht man vor allem in den angelsächsischen Ländern, den Anteil männlicher Lehrkräfte zu erhöhen (Driessen 2007). In Großbritannien (TDA 2009) als auch in Australien (ABC-News 2004) wurden von staatlicher Seite Initiativen gestartet mit dem Ziel, mehr Männer für den Lehrerberuf zu gewinnen. Auch in Deutschland gibt es in jüngster Zeit Bestrebungen, den Lehrerberuf für Männer attraktiver zu machen (BMFSFJ 2010).

In diesem Beitrag soll die weit verbreitete Meinung, dass das Geschlecht des Lehrers einen Einfluss auf die Beurteilung von geschlechtsspezifischen Schulleistungen hat, empirisch überprüft werden. Der Fokus liegt dabei auf der Grundschulempfehlung (GSE), in der die Lehrpersonen empfehlen oder – je nach Bundesland – festlegen, welche Art der weiterführenden Schule ein Kind nach der Grundschule besuchen soll.¹ Die GSE ist im Zusammenhang mit der Fragestellung von besonderer Bedeutung. Zum einen sind die Vorgaben zur Vergabe der GSE in den meisten Bundesländern so formuliert, dass sich für die (überwiegend weiblichen) Lehrkräfte der nötige Interpretationsspielraum für geschlechtsspezifische Benachteiligungen bietet. Zum anderen spielt die GSE beim Übergang auf die weiterführende Schule und damit für die weitere Bildungs- und Erwerbskarriere eine überragende Rolle (Ditton und Krüskens 2006; Stubbe und Bos 2008; Wiese 1982), da sich die Eltern zu großen Teilen nach der GSE richten (Bos et al. 2004a; Bos et al. 2003; Ditton 1992; Lehmann et al. 1997). Die Darstellung des Lehrerinnenanteils an allen allgemeinbildenden Schulen (Abbildung 1) verschleiert die Tatsache, dass der Lehrerinnenanteil an Grundschulen besonders hoch ist (2007: 88 Prozent).² Wenn nun die These stimmt, dass Jungen durch die Feminisierung der Lehrerschaft benachteiligt sind, dann sollten entsprechende Effekte in denjenigen Klassen besonders sichtbar werden, in denen die Grundschulempfehlung von überwiegend weiblichen Lehrkräften getroffen wird.

Der empirischen Überprüfung geschlechtsspezifischer Benachteiligung bei der GSE ist zunächst die Frage vorgelagert, inwiefern man überhaupt von einer ‚Krise der Jungen‘ sprechen kann (Abschnitt 5.2.1). Anschließend werden die theoretischen Grundlagen der Feminisierungsthese dargelegt – es wird also erörtert, warum weibliche Lehrkräfte Schüler gegenüber

¹ Die Empfehlung der Grundschule wird je nach Bundesland anders bezeichnet. Um Wortverwirrungen zu vermeiden, werden im Folgenden einheitlich die Begriffe „GSE“ oder „Empfehlung“ verwendet. Zum – je nach Bundesland – variierenden Einfluss der Grundschulempfehlung auf die Bildungsentscheidung vgl. Neugebauer (2010).

² Besonders niedrig ist der Lehrerinnenanteil hingegen an Gymnasien (2007: 54 Prozent) und Hauptschulen (2007: 58 Prozent). Da die amtliche Statistik Grund- und Hauptschullehrer erst ab 1989 getrennt ausweist, wird für die lange Zeitreihe (Abbildung 1) der Frauenanteil an allen allgemeinbildenden Schulen dargestellt.

Schülerinnen benachteiligen sollten (Abschnitt 5.2.2) und warum diese Mechanismen besonders beim Übergang auf das Gymnasium relevant sein sollten (Abschnitt 5.2.3). Aus dem im Anschluss dargestellten Forschungsstand zur geschlechtsspezifischen Benachteiligung in den Beurteilungen der Lehrpersonen (Abschnitt 5.3) ergeben sich unmittelbare Anknüpfungspunkte für die eigene empirische Studie. In Abschnitt 5.4 werden die zu Grunde liegenden IGLU Daten und die Operationalisierungen der Variablen beschrieben, bevor in Abschnitt 5.5 untersucht wird, ob sich eine Benachteiligung in den Empfehlungen nachweisen lässt. Dabei zeigt sich, dass männliche Klassenkonferenzen verglichen mit weiblichen Klassenkonferenzen seltener Gymnasialempfehlungen aussprechen – und zwar für Mädchen und Jungen gleichermaßen. Eine Benachteiligung der Jungen durch weiblich besetzte Klassenkonferenzen findet nicht statt. Lehrer, so scheint es, urteilen insgesamt strenger als Lehrerinnen. Darüber hinaus finden sich Hinweise, dass Jungen bei den Gymnasialempfehlungen gar nicht benachteiligt werden, wenn man die Faktoren berücksichtigt, die laut den Vorgaben der Kultusministerkonferenz in die Empfehlung einfließen sollen. Jungen, die die gleiche Lernbereitschaft an den Tag legen wie Mädchen, werden sogar signifikant häufiger für das Gymnasium empfohlen. Abschließend werden die zentralen Ergebnisse zusammengefasst und diskutiert (Abschnitt 5.6).

5.2 Die „Krise“ der Jungen, die Feminisierung des Lehrerberufs und die Bedeutung der Grundschulempfehlung

Die Feminisierungsthese besteht aus zwei Teilen. Erstens wird konstatiert, dass die Jungen in ihrem Schulerfolg gegenüber den Mädchen zurückfallen. Zweitens wird dann die Feminisierung des Lehrerberufs, insbesondere in der Primarstufe, als Ursache für dieses Zurückfallen angesehen.

5.2.1 Gibt es eine „Krise der Jungen“?

Ob man tatsächlich davon sprechen kann, dass Jungen hinter den Mädchen zurückfallen, hängt davon ab, wie Schulerfolg definiert und welche Altersgruppe betrachtet wird. Betrachtet man Bildungsabschlüsse als Indikator für Schulerfolg, zeigt sich dieser relative Nachteil der Jungen im Zeitverlauf sehr deutlich (vgl. Abbildung 1). Zieht man als Indikator jedoch Kompetenzen heran, relativiert sich dieses Bild. Auf Basis einer Meta-Analyse von Leistungstests in Schreiben, Mathematik und Naturwissenschaften stellen Nowell und Hedges (1998) fest, dass Geschlechtsunterschiede in Leistungstests über die letzten 30 Jahre relativ stabil geblieben sind.

Auch mit Blick auf die Schulnoten zeigt sich, dass Mädchen bereits in den 1950er und 1960er Jahren bessere Noten erzielten als ihre männlichen Klassenkameraden (Rodax und Hurrelmann 1986). Eine differenzierte Betrachtung der einzelnen Unterrichtsbereiche zeigt, dass Mädchen besser lesen, während Jungen besser rechnen können (Baker und Jones 1993; Bos et al. 2007; Bos et al. 2003; Buchmann et al. 2008; Marks 2007; Nowell und Hedges 1998; PISA-Konsortium 2000, 2005). Bei den Schulnoten zeigt sich tendenziell, dass Mädchen bei gleichen Leistungen besser benotet werden: Auswertungen der IGLU Daten von 2001 zeigen, dass Jungen in der vierten Jahrgangsstufe in den Fächern Deutsch und Sachunterricht bei vergleichbaren Kompetenzen schlechter benotet werden als Mädchen. In Mathematik sind nach Kontrolle der Kompetenzen die Noten der Jungen besser als die der Mädchen, dieser Unterschied ist jedoch nicht signifikant (Valtin et al. 2005). Jungen scheinen ihre Kompetenzen nicht in gleichem Maße in Noten umsetzen zu können wie Mädchen. Ob dies auf die Feminisierung der Schule zurückgeführt werden kann, ist jedoch fraglich. Ebenso könnten geringere Selbstdisziplin (Duckworth und Seligman 2006), geringere Motivation oder höhere Arbeitsvermeidung (Steinmayr und Spinath 2008) der Jungen für die Differenzen verantwortlich sein. Zudem zeigt sich sowohl bei den Kompetenzen als auch bei den Noten, dass die Geschlechterdifferenzen auf höheren Schulstufen ansteigen (Buchmann et al. 2008; Willingham und Cole 1997). Das spricht insofern gegen die Feminisierungsthese, als das auf höheren Schulstufen die Geschlechterkomposition der Lehrerschaft ausgewogener ist.

Es gibt sie also, die Geschlechterdifferenzen. Sie gehen jedoch keineswegs durchgängig zu Lasten der Jungen und sind tendenziell größer, je älter die Schüler sind – und je ausgeglichener die Geschlechterzusammensetzung des Kollegiums ist. Darüber hinaus haben die Mädchen die Jungen im Zeitverlauf nicht überholt, wenn man die Kompetenzen oder Noten heranzieht. Ein differenzierterer Blick auf die „Krise der Jungen“ lässt die populäre These von der Feminisierung unplausibler erscheinen, als dies in Abbildung 1 suggeriert wird. Vielmehr scheint es, dass Mädchen anders als früher ihre besseren Noten in entsprechende Bildungszertifikate umsetzen. Dieser Aspekt wird allerdings weitgehend vernachlässigt (Buchmann et al. 2008; Mickelson 1989). Die in der Öffentlichkeit geführte Debatte von der „Krise der Jungen“ läuft Gefahr, die Wirklichkeit zu vereinfachen (Budde 2009; Stamm 2008).

5.2.2 Die Feminisierung der Lehrerschaft als Ursache?

Nichtsdestoweniger wird die Feminisierung der Schule häufig herangezogen, um das schlechtere Abschneiden der Jungen zu erklären. Die Diskussion um eine nachteilige Wirkung von

Lehrerinnen auf Jungen ist allerdings kein neues Phänomen. In den USA, in denen der Lehrberuf bereits Ende des 19. Jahrhunderts zum Frauenberuf wurde (Martino 2008; Rury 1989), gab es bereits zu Beginn des 20. Jahrhunderts Kampagnen um mehr Männer für den Lehrberuf zu gewinnen. Auch zwischen den beiden Weltkriegen wurden Anstrengungen unternommen, um den Männeranteil im Lehrberuf zu erhöhen (Blount 2000; Martino 2008). Das Hauptargument hierfür war die Sorge, dass Jungen ohne gleichgeschlechtliche Rollenmodelle „verweichlichen“ würden und in ihrer männlichen Entwicklung behindert wären (Sexton 1969). Heute dagegen wird der Ruf nach mehr männlichen Lehrkräften eher mit dem relativ niedrigen Schulerfolg der Jungen begründet. Im Hinblick auf die Forderung nach mehr männlichen Lehrkräften hat also eine gewisse Verschiebung in der Begründung stattgefunden. Unverändert in der zyklisch wiederkehrenden Debatte um mehr männliche Lehrkräfte ist aber die Fokussierung auf die Jungen. Vor allem Jungen würden durch das Fehlen von Männern im Lehrberuf negativ in ihrer Entwicklung beeinflusst und erbrächten schlechtere Schulleistungen. Dabei wird jedoch vergessen, dass die Entwicklung des geschlechtsspezifischen Bildungserfolgs ebenso durch den steigenden Bildungserfolg der Mädchen zu erklären ist.

Die Argumentationslinien, die einen negativen Einfluss der Feminisierung des Lehrberufs auf den Schulerfolg der Jungen postulieren, lassen sich in zwei Gruppen einteilen.³ Die erste Gruppe von Argumenten nimmt an, dass weibliche Lehrkräfte einen negativen Einfluss auf die schulischen Leistungen von Jungen haben. Der Mangel an männlichen Rollenvorbildern, so die Argumentation, wirke sich negativ auf die Lernbereitschaft und in der Folge auf die Kompetenzentwicklung von Jungen aus. Eine zweite Gruppe von Argumenten geht davon aus, dass Jungen bei gleicher schulischer Leistung von weiblichen Lehrkräften schlechter bewertet werden (Arnot et al. 1999; Diefenbach und Klein 2002; Hannan 2001; vgl. Neugebauer et al. 2011 für eine detaillierte Darstellung). Verhaltensweisen, die den schulischen Alltag stören und vermutlich auch die schulischen Leistungen beeinträchtigen, sind bei Jungen häufiger als bei Mädchen anzutreffen (Eagly und Chvala 1986). Möglicherweise werden Lehrerinnen durch solches Verhalten stärker „irritiert“ als männliche Lehrer, wenn sie als Maßstab die eigene geschlechtsspezifische Sozialisation heranziehen. „Lehrerinnen prägen die Schulkultur, möglicherweise erwarten und prämiieren sie solche Verhaltensweisen, die Mädchen im Rahmen ihrer Sozialisation einüben, Jungen aber nicht (in demselben Maß).“ (Diefenbach und Klein

³ Darüber hinaus gibt es eine weitere Argumentationslinie, die die „Feminisierung der Schulkultur“ als Ursache für das Schulversagen von Jungen ansieht. Jungen haben schlechtere Schulleistungen, weil die Schule, unabhängig von dem Geschlecht der Lehrkräfte, in ihrer Wissensvermittlung „weiblich“ ausgerichtet sei (Bacher et al. 2008). Diese Argumente sind weitgehend unabhängig vom Geschlecht des Lehrkörpers und werden daher in diesem Beitrag nicht näher ausgeführt oder geprüft.

2002, S. 950). Weil Lehrerinnen dem Verhalten von Jungen weniger Verständnis entgegenbringen und geringere Erwartungen an sie haben, würden sie Jungen auch seltener für Schulen mit höherem Anspruchsniveau empfehlen. Aber auch das tatsächliche Verhalten der Kinder kann bei Lehrerinnen anders sein als bei Lehrern und damit eine mögliche Benachteiligung der Jungen erklären (Helbig 2010; Neugebauer et al. 2011). So ist vorstellbar, dass Schülerinnen und Schüler sich bei gleichgeschlechtlichen Lehrkräften leistungsbereiter zeigen als bei gegengeschlechtlichen Lehrkräften, so dass eine bessere Benotung bei gleichgeschlechtlichen Lehrkräften die Folge höherer Leistungsbereitschaft wäre und nicht Folge bewusster oder unbewusster Bewertung.

5.2.3 Die Rolle der Grundschulempfehlung

Die dargestellten Mechanismen wären besonders in der Grundschule bedeutsam, denn hier werden die Grundlagen für jedes zukünftige Lernen gelegt. Die Entscheidung am Ende der Grundschule über die weiterführende Schulform hat einen nachhaltigen Einfluss auf die weitere Bildungs- und Berufskarriere (z. B. Müller und Haun 1994). In der Grundschule ist der Anteil der Lehrerinnen verglichen mit anderen Schulformen am größten (2007: 88 Prozent). Die überwiegend weiblichen Lehrkräfte prägen durch die Vergabe der GSE die weiterführende Schulwahl maßgeblich mit (Bos et al. 2004b; Ditton und Krüskens 2006; Wiese 1982), indem sie in Form der GSE eine „Bildungsentscheidung“ für das jeweilige Kind treffen müssen (Neugebauer 2010). In den Beschlüssen der Kultusministerkonferenz sind die Selektionskriterien bei der Vergabe der GSE sehr weit interpretierbar. Demnach sind neben „Kenntnissen und Fertigkeiten“ auch „Eignung, Neigung und Wille des Kindes zu geistiger Arbeit“ bei der Empfehlung zu berücksichtigen. Weiter heißt es, dass die Empfehlung „nicht nur die Leistungen in Bezug auf die fachlichen Ziele der Lehrpläne, sondern auch die für den Schulerfolg wichtigen allgemeinen Fähigkeiten“ berücksichtigen soll (KMK 2010, S. 4 ff.). Insgesamt soll also erkennbar sein, dass ein Kind die Anforderungen der jeweiligen Schulform erfüllen kann. Diese Vorgaben eröffnen den Lehrkräften einen „pädagogisch sinnvollen Entscheidungsspielraum“ (Maier 2007, S. 272) bei der Feststellung, welches Anforderungsniveau die weitere Lernentwicklung des Kindes vermutlich am besten fördern wird (Bellenberg et al. 2004, S. 48; Cortina und Trommer 2003).

Wenn nun die Feminisierungsthese zutrifft, sollten Jungen bei ansonsten gleichen Voraussetzungen seltener auf das Gymnasium überwiesen werden, wenn die für die GSE

verantwortliche Klassenkonferenz weiblich dominiert ist. Mädchen wiederum sollten bei gleichen Kompetenzen häufiger auf das Gymnasium überwiesen werden, wenn die verantwortliche Klassenkonferenz weiblich dominiert ist.

5.3 Forschungsstand zur Feminisierungsthese

Die empirische Evidenz zum Einfluss des Geschlechts der Lehrperson auf den geschlechtsspezifischen Schulerfolg von Kindern ist gemischt, wobei die Mehrzahl der Studien keine signifikanten Effekte findet. Dee (2007) berichtet für die USA, dass Jungen und Mädchen in der 8. Klasse sowohl höhere Testleistungen erzielen als auch vorteilhafter bewertet werden, wenn sie von einer gleichgeschlechtlichen Lehrkraft unterrichtet werden. Die Ergebnisse von Ammermüller und Dolton (2006) deuten darauf hin, dass in England Jungen im Alter von 13 Jahren bei Lehrern etwas höhere Matheleistungen erzielen als bei Lehrerinnen. Für Mädchen finden sie signifikante Vorteile in Englischtests, wenn diese von Lehrerinnen unterrichtet werden. Für die USA finden sie im Gegensatz zu Dee (2007) keine signifikanten Effekte des Geschlechts der Lehrperson. Für Kinder im Alter von 9 Jahren ergeben sich sowohl in England als auch den USA keine Effekte. Auch andere Autoren können den Zusammenhang zwischen dem Geschlecht der Lehrkraft und den Kompetenzen oder Schulnoten von Jungen und Mädchen nicht oder nur sehr schwach nachweisen (für die Niederlande: Driessen 2007; für die USA: Ehrenberg et al. 1995; für die Schweiz: Hadjar und Lupatsch 2011; für Schweden: Holmlund und Sund 2008; für Kanada: Sokal et al. 2007).

Für Deutschland gibt es bisher erst zwei entsprechende Studien: Helbig (2010) untersuchte mit der Berliner ELEMENT-Studie, ob der Anteil männlicher Lehrkräfte an einer Grundschule Einfluss auf die Kompetenzen, Noten und Grundschulempfehlungen für Jungen und Mädchen am Ende der Grundschulzeit hat. Da in ELEMENT selbst keine Informationen über das Geschlecht der Lehrkräfte enthalten sind, wurde den untersuchten Grundschulen mittels Onlinerecherchen der Anteil männlicher Lehrkräfte zugespielt. Die Studie liefert Hinweise, dass Jungen weder in ihrem Leseverständnis noch in ihren Mathematikkompetenzen von mehr männlichen Lehrkräften profitieren. Mädchen dagegen weisen bessere Leseleistungen auf, wenn sie Schulen mit einem höheren Lehrerinnenanteil besuchen. Auf die Mathematikkompetenzen hatte der Lehrerinnenanteil keinen Einfluss. Bei der Notenvergabe zeigt sich, dass Jungen in Schulen mit einem höheren Anteil männlicher Lehrpersonen bessere Mathematiknoten bekommen, in Deutsch jedoch nicht. Der gefundene Zusammenhang ist jedoch schwach.

Bei den Mädchen war kein Einfluss des Anteils männlicher Lehrer auf die Notengebung erkennbar. Die Empfehlungen für die weiterführenden Schulen hängen bei beiden Geschlechtern von den Noten ab, nicht jedoch vom Anteil männlicher Lehrer an einer Schule. Die Studie untersucht allerdings nicht den direkten Zusammenhang zwischen dem Geschlecht der Lehrkraft und dem Bildungserfolg der Schüler, sondern konzentriert sich auf den Einfluss des Anteils männlicher Lehrpersonen an Berliner Grundschulen. Unklar bleibt damit also, ob Jungen tatsächlich auch von einem Mann unterrichtet wurden und wie sich dies auf die Leistungen auswirkt. In diese Lücke stößt eine zweite Studie (Neugebauer et al. 2011), die erstmals in Deutschland das Geschlecht der unterrichtenden Lehrkraft direkt in Beziehung zur Kompetenzentwicklung und den Noten von Mädchen und Jungen setzt. Diese Untersuchung zeigt für Gesamtdeutschland unter Verwendung der IGLU Daten von 2001, dass weder Jungen noch Mädchen bei Kompetenzentwicklung oder Noten in Mathematik, Deutsch oder Sachkunde von einer Lehrperson des gleichen Geschlechts profitieren. Nominell erreichen alle Kinder bessere Noten bei weiblichen Lehrkräften, keine dieser Differenzen ist aber signifikant. Darüber hinaus stellen die Autoren fest, dass die Leseleistungen von Mädchen und Jungen signifikant schlechter sind, wenn sie in Deutsch vier Jahre lang von einem Mann unterrichtet wurden. Zwar lassen sich die Ursachen hinter diesem Befund nicht präzise benennen. Es wird jedoch deutlich, dass der pauschale Ruf nach mehr männlichen Lehrkräften unbeabsichtigte Folgen nach sich ziehen kann, die für die Kompetenzentwicklung bei Mädchen und Jungen sogar nachteilig sein können.

Noch nicht abschließend geklärt ist die Frage, ob ein direkter Zusammenhang zwischen dem Geschlecht der an der Grundschulempfehlung beteiligten Lehrkräfte und der Wahrscheinlichkeit einer Gymnasialempfehlung für Mädchen und Jungen besteht. Auf höheren Aggregatebenen gibt es dazu 3 Studien. Wie Helbig (2010) in seiner bereits erwähnten Untersuchung zeigen konnte, findet sich auf Schulebene für Jungen (nicht jedoch für Mädchen) nur ein indirekter, über die Schulnoten vermittelter Einfluss des Anteils männlicher Lehrer auf die Empfehlung. Ein solcher Zusammenhang konnte auf Kreisebene für den Übergang auf die Allgemeine höhere Schule (AHS) in Österreich ebenfalls nicht festgestellt werden (Bacher et al. 2008). Zu einem gegenläufigen Befund kommt eine Studie von Diefenbach und Klein (2002). Die Autoren stellen fest, dass der Anteil weiblicher Grundschullehrkräfte in einem Bundesland mit den geschlechtsspezifischen Hauptschulabschluss- und Abiturquoten zusammenhängt und schließen daraus, dass Jungen von weiblichen Lehrkräften benachteiligt werden. Allerdings überzeugen diese Studien hinsichtlich ihres methodischen Vorgehens nur bedingt, da sie durch

die Analyse von Aggregatdaten auf Prozesse auf der Individualebene schließen. Die nachfolgenden Analysen schließen unmittelbar an die bisherigen Studien an und testen, ob und inwiefern Jungen bei der Gymnasialempfehlung durch weibliche Lehrkräfte benachteiligt werden.

5.4 Daten und Operationalisierungen

Die empirischen Analysen dieses Beitrags wurden mit dem deutschen Datensatz aus der internationalen Progress in Reading Literacy Study von 2001 (PIRLS) durchgeführt (Mullis et al. 2003), die in Deutschland unter dem Akronym IGLU (Internationale Grundschul-Lese-Untersuchung) bekannt ist. In IGLU wurden die Lesekompetenzen von Schülern am Ende der vierten Jahrgangsstufe getestet. In 12 von 16 Bundesländern wurde die Untersuchung um Kompetenzmessungen in den Kernfächern Mathematik und Sachkunde erweitert.⁴ Neben den Schulleistungstests wurden in IGLU umfangreiche Hintergrundinformationen durch Befragung der Schüler/-innen, sowie deren Eltern, Lehrpersonen und der Schulleitung erhoben. Die Stichprobenziehung verlief in einem zweistufigen Verfahren: Zunächst wurden proportional zur Einwohnerzahl zufällig Schulen gezogen, aus denen dann zufällig Klassen in Stufe 4 gezogen wurden. Innerhalb der gezogenen Klassen wurden im Prinzip alle Kinder sowie deren Eltern befragt. Die Teilnehmerate war mit rund 84 Prozent zufrieden stellend (Details: Bos et al. 2004a; Bos et al. 2003).⁵ 5.943 Schülerinnen und Schüler an 168 Schulen umfasste die den Zufallskriterien genügende Stichprobe für die nationale Erweiterung IGLU-E.

Die abhängige Variable ist die von der Klassenkonferenz ausgesprochene Grundschulempfehlung (GSE) am Ende der vierten Klassenstufe.⁶ Einige Bundesländer mussten von den Analysen ausgeschlossen werden, weil für sie keine GSE vorlag: In Berlin und Brandenburg findet der Übergang auf eine weiterführende Schule erst nach einer 6-jährigen Grundschulzeit statt. In Niedersachsen, Sachsen-Anhalt und Bremen besuchten im Jahr 2001 alle Schüler in der fünften und sechsten Klasse eine gemeinsame (schulartunabhängige) Orientierungsstufe (gemeinsame Förderstufe). Nach Ausschluss der genannten Bundesländer verbleiben 5022 Schüler aus 255 Klassen an 137 Schulen im Datensatz. Die Vielfalt der Schulsysteme hat zur

⁴ Die Länder Brandenburg, Mecklenburg-Vorpommern, Niedersachsen und Sachsen-Anhalt haben an der Erweiterung um Mathematik und Sachkunde (IGLU-E) nicht teilgenommen.

⁵ Grund für den im internationalen Vergleich etwas niedrigeren Wert ist vermutlich die in Deutschland notwendige schriftliche Einverständniserklärung der Eltern.

⁶ Die Angaben stammen direkt von den verantwortlichen Lehrkräften.

Folge, dass systematische Vergleiche über die Bundesländer nur für das Gymnasium möglich sind, welches als einzige Schulart überall existiert.⁷ Aus diesem Grund wird die abhängige Variable für die späteren Analysen dichotomisiert in eine Empfehlung für oder gegen das Gymnasium. 35.2 Prozent der Kinder erhielten eine Gymnasialempfehlung und 64.8 Prozent eine Real- oder Hauptschulempfehlung.

Um abzuschätzen, ob Jungen und Mädchen von gleichgeschlechtlichen Lehrkräften bei der Vergabe der GSE bevorzugt oder benachteiligt werden, muss das Geschlecht der an der GSE beteiligten Lehrkräfte ermittelt werden. Die GSE wird in aller Regel von einer Klassen- oder Versetzungskonferenz entschieden. Die Zusammensetzung der Konferenz ist in den Bundesländern im Detail unterschiedlich geregelt. Länderübergreifend gilt jedoch, dass diejenigen Lehrpersonen, die eine Klasse unterrichten, gemeinsam eine Entscheidung fällen. Zudem ist der Schulleiter oder die Schulleiterin stimmberechtigtes Mitglied der Konferenz.⁸ Für die vorliegende Fragestellung muss nun identifiziert werden, ob die Entscheidungsträger überwiegend weiblich oder männlich sind. In IGLU-E stehen Informationen zum Geschlecht der Deutsch-, Mathematik- und Sachkundelehrkräfte sowie der Schulleiter zur Verfügung. Diese können mit den einzelnen Schülern verknüpft werden. In einem ersten Schritt wurde eine Variable konstruiert, die anzeigt, ob die Lehrkräfte aller drei Hauptfächer ausschließlich weiblich, ausschließlich männlich, oder gemischtgeschlechtlich sind. Dabei handelt es sich in vielen Fällen um dieselbe Lehrkraft, die alle drei Fächer unterrichtet (49.8 Prozent der Schüler/in-Lehrpersonen-Dyaden). In einem zweiten Schritt wurde das Geschlecht der Schulleitenden zugespielt. Insgesamt waren bei 33.8 Prozent der Schüler alle Hauptfachlehrpersonen sowie die Schulleiterin weiblich (= weibliche Klassenkonferenz). Bei 59.4 Prozent der Schüler war die Klassenkonferenz gemischtgeschlechtlich besetzt (=gemischte Klassenkonferenz) und bei 6.8 Prozent gehörten ausschließlich Männer der Klassenkonferenz an.⁹

⁷ So gibt es die Hauptschule als eigenständige Schulart nur in 10 von 16 Ländern, die Realschule sogar nur in 8, was in erster Linie darauf zurückzuführen ist, dass einige Länder diese Schularten (und damit auch die Empfehlungen) zusammengefasst haben. Durch die Dichotomisierung können wie im Falle Hamburg auch Länder in der Stichprobe verbleiben, in denen nicht zwischen Hauptschul- und Realschulempfehlung, aber zwischen Gymnasialempfehlung und weniger als Gymnasialempfehlung unterschieden werden kann.

⁸ vgl. etwa für BW: §10 Abs.1, §12 Abs. 2, §13 Abs. 1 Konferenzordnung des Kultusministeriums; NRW: §50 Abs. 2 SchulG; RP: §16 Abs. 2, §46 Abs. 2 Schulordnung für öff. Grundschulen, Verordnungspunkt 3.2 der Konferenzordnung; TH: §37 Abs. 3 ThürSchulG; HE: §135 Abs.2 HSchG; SN: §21 Abs.1 SOGS, §5 Abs.2 und §9 Abs.1 LKonf; HH: §62 Abs.1 und 2 HmbgSG; SH: §65 Abs.4 SchulG; SL: §12 Abs.1 und 2 SchumG.

⁹ Dass der Anteil der weiblichen Konferenzen nicht höher liegt, ist der Tatsache geschuldet, dass die meisten Schulen (56 Prozent der Schulen im Analysesample) einen männlichen Schulleiter haben. Traditionelle Geschlechterhierarchien herrschen also selbst im „feminisierten“ Lehrerberuf vor.

Für ein adäquates Prüfungsmodell geschlechtsspezifischer Benachteiligung müssen alle für die GSE relevanten Einflussfaktoren kontrolliert werden. Die Regelungen zur Grundschulempfehlung sind Ländersache. Allerdings hat die Kultusministerkonferenz einige allgemeine Vorgaben definiert, nach denen sich die Bundesländer richten (KMK 2010, S. 4 ff.).¹⁰ Nach diesen Vorgaben sollen zum einen "Kenntnisse und Fertigkeiten" der Kinder ausschlaggebend für die Empfehlung sein. Diese wurden mit Hilfe von standardisierten Leistungstests in den Bereichen Lesen, Mathematik und Sachkunde operationalisiert. Die Tests weisen laut Expertenmeinung eine hohe curriculare Lehrplanvalidität auf (vgl. Lankes et al. 2003). Zum anderen wurden auch die Noten in den Fächern Deutsch, Mathematik und Sachkunde berücksichtigt, in denen sich die Leistungseinschätzung der Kinder durch die Lehrkräfte widerspiegelt. Darüber hinaus sollen nach den Vorgaben der KMK aber auch „Eignung, Neigung und Wille des Kindes zu geistiger Arbeit insgesamt“ in der GSE berücksichtigt werden. Die allgemeine Eignung zu geistiger Arbeit wird über die Punktzahl erfasst, die ein Kind in einem allgemeinen kognitiven Fähigkeitstest (Intelligenztest) erreicht hat. Neigung und Wille zu geistiger Arbeit wird über die Lern- und Leistungsbereitschaft der Kinder erfasst. Die Eltern beurteilten dazu (auf einer 4er-Skala) sieben Aussagen zum Lern- und Leistungsverhalten ihres Kindes, bspw. "Mein Kind erledigt Hausaufgaben ordentlich und genau" oder "Mein Kind ist konzentriert und ausdauernd", aus denen eine Skala gebildet wurde (Eigenwert des ersten Faktors 3.43, Cronbachs Alpha .82). Als weitere Kontrollvariablen wurden die soziale Herkunft¹¹ und der Migrationshintergrund des Kindes berücksichtigt, die üblicherweise mit dem Schulerfolg zusammenhängen. In der Literatur werden häufiger Stadt/Land Disparitäten sowie Effekte der durchschnittlichen Klassenleistung („big fish little pond“-Effekt) oder der Klassengröße auf die Wahrscheinlichkeit, eine Gymnasialempfehlung zu erhalten, berichtet. Mit den IGLU Daten ist es möglich, diese Einflussfaktoren zu kontrollieren. Alle im Folgenden berichteten multivariaten Analysen wurden zur Kontrolle unter Berücksichtigung dieser Faktoren gerechnet. Die Ergebnisse waren in Bezug auf die Fragestellung die gleichen.

Mit Blick auf die Item-Nonresponse hat sich die Einsicht durchgesetzt, dass die Schätzung fehlender Werte durch multiple Imputationen gegenüber dem paar- oder fallweisen Ausschluss vorzuziehen ist (Rubin 1987; Allison 2001). Um die Unsicherheit aufgrund fehlender Werte zu berücksichtigen, wurden mit Hilfe des Stata Programms ICE (Royston 2007) für

¹⁰ Die Vorgaben der Kultusministerkonferenz (KMK) finden sich entsprechend in den Landesschulgesetzen zur Übergangsempfehlung wieder.

¹¹ abgebildet über den sozioökonomischen Status der Eltern (HISEI) (vgl. Ganzeboom et al. 1992).

jeden fehlenden Wert 5 plausible Werte erzeugt. Die Analysen wurden nach dem von Rubin (1987) vorgeschlagenen Verfahren kombiniert.

5.5 Ergebnisse

Zunächst werden in Tabelle 1 die Verteilungen der zentralen Variablen getrennt für Jungen und Mädchen dargestellt. Knapp 38 Prozent der Mädchen erhalten eine Gymnasialempfehlung, bei den Jungen sind es mit rund 33 Prozent signifikant weniger. Mädchen haben zudem bessere Kompetenzen und Noten in Deutsch. In Mathematik hingegen sind die Jungen kompetenter und erlangen in der Folge auch bessere Noten. In Sachkunde ist das Bild gemischt: hier weisen die Jungen höhere Kompetenzen auf, allerdings erhalten die Mädchen die etwas besseren Noten. Dies deckt sich mit früheren Befunden (vgl. Abschnitt 5.2.1). Bei dem allgemeinen kognitiven Fähigkeitstest schneiden Mädchen etwas besser ab als Jungen. Darüber hinaus legen sie eine höhere Lern- und Leistungsbereitschaft an den Tag, d.h. sie sind im Mittel konzentrierter, erledigen regelmäßiger und ordentlicher ihre Hausaufgaben, zeigen mehr Fleiß – kurzum: sie „tun mehr“ für die Schule.

Tabelle 2: Mittelwerte zentraler Variablen nach Geschlecht der Schüler

| | Mädchen | Jungen |
|---------------------------------|------------------------|------------------------|
| Gymnasialempfehlung | 37.63 % ^{***} | 32.92 % ^{***} |
| Lesekompetenz | 514.66 ^{***} | 499.24 ^{***} |
| Deutschnote | 2.60 ^{***} | 2.97 ^{***} |
| Mathematikkompetenz | 496.91 ^{***} | 512.58 ^{***} |
| Mathematiknote | 2.80 ^{***} | 2.70 ^{***} |
| Sachkundekompetenz | 495.67 ^{***} | 510.99 ^{***} |
| Sachkundenote | 2.49 ^{**} | 2.57 ^{**} |
| kognitiver Fähigkeitstest | 49.02 ^{***} | 47.91 ^{***} |
| Lern- und Leistungsbereitschaft | 3.04 ^{***} | 2.79 ^{***} |
| Klassenkonferenz | | |
| -männlich | 6.78 % | 6.85 % |
| -gemischt | 60.42 % | 58.38 % |
| -weiblich | 32.79 % | 34.77 % |
| Migrationshintergrund | 23.59 % | 24.44 % |
| Höchster ISEI der Familie | 49.81 | 49.92 |
| N | 2456 | 2566 |

Anmerkung: Signifikante Unterschiede zwischen Jungen und Mädchen ^{***} p<0.01, ^{**} p<0.05, ^{*} p<0.1 (zweiseitiger T-Test)

Quelle: IGLU-E 2001, eigene Berechnungen

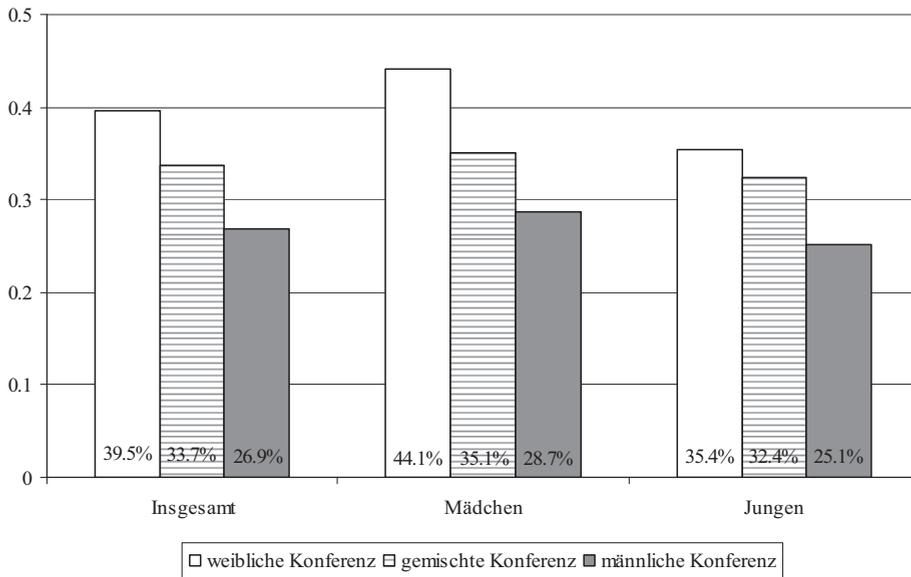
Im Hinblick auf askriptive Merkmale wie Migrationshintergrund oder soziale Herkunft unterscheiden sich Mädchen und Jungen erwartungsgemäß nicht. Auch die Zuteilung zu männlichen oder weiblich dominierten Lehrerkollegien variiert nicht systematisch mit dem Geschlecht der Kinder. Die Wahrscheinlichkeit beispielsweise eine weibliche Klassenkonferenz zu haben ist für Mädchen und Jungen gleich. Am häufigsten haben alle Kinder eine gemischtgeschlechtlich besetzte Konferenz.

5.5.1 Werden Jungen bei der Gymnasialempfehlung von weiblichen Lehrkräften benachteiligt?

Nach der Feminisierungsthese sollte die Erhöhung der Männerquote an Grundschulen zu einer geringeren Benachteiligung der Jungen bei den Gymnasialempfehlungen führen. Demnach sollten männlichen Klassenkonferenzen Jungen häufiger auf das Gymnasium überweisen als weibliche Klassenkonferenzen. Umgekehrt sollten Mädchen häufiger auf das Gymnasium überwiesen werden, wenn die Klassenkonferenz weiblich dominiert ist. In Abbildung 2 ist der Anteil der Gymnasialempfehlungen von Mädchen und Jungen in Abhängigkeit von der geschlechtlichen Zusammensetzung der Klassenkonferenz dargestellt. Es zeigt sich, dass alle Kinder, Jungen wie Mädchen, eine geringere Wahrscheinlichkeit haben eine Gymnasialempfehlung zu erhalten, wenn die zuständige Klassenkonferenz ausschließlich männlich zusammengesetzt ist. Zwar scheinen insbesondere Mädchen von weiblichen Konferenzen zu profitieren. Ein differenzieller Effekt des Geschlechts der Lehrperson wie ihn die Feminisierungsthese annimmt, also eine positive Wirkung von männlichen Konferenzen bei Jungen und eine negative Wirkung bei Mädchen, lässt sich aber nicht erkennen. Eher liegt der Schluss nahe, dass eine Erhöhung des Männeranteils für Jungen wie für Mädchen die Wahrscheinlichkeit reduzieren würde, eine Gymnasialempfehlung zu erhalten.

Diese Schlussfolgerung wäre jedoch voreilig, denn der berichtete bivariate Zusammenhang könnte durch unbeobachtete Drittvariablen entstehen. So erhalten beispielsweise Kinder auf dem Land seltener eine Gymnasialempfehlung. Ebenso werden an Schulen mit geringerem Leistungsniveau seltener Schüler für das Gymnasium empfohlen. Wenn nun männliche Lehrkräfte eine höhere Wahrscheinlichkeit haben, an ebensolchen Schulen zu unterrichten, wären vielmehr die Schulcharakteristika als die männlichen Lehrer verantwortlich für den beobachteten Effekt.

Abbildung 2: Anteil der Gymnasialempfehlungen für Mädchen und Jungen in Abhängigkeit von der geschlechtlichen Zusammensetzung der Klassenkonferenz



Anmerkung: Insgesamt sind alle Differenzen auf dem 5% Niveau signifikant. Betrachtet man nur die Mädchen, ist die Differenz zwischen männlicher und weiblicher Konferenz auf dem 1% Niveau signifikant ($t=4.08$), die Differenz zwischen männlicher und gemischter Konferenz auf dem 6% Niveau ($t= 1.92$) und die Differenz zwischen weiblicher und gemischter Konferenz auf dem 5% Niveau ($t= 2.83$). Bei den Jungen ist nur die Differenz zwischen männlicher und weiblicher Konferenz auf dem 10% Niveau signifikant ($t=1.88$).

Quelle: IGLU-E 2001, eigene Berechnungen

Um diesen Einwand zu prüfen, wurde in multinomialen logistischen Regressionen geprüft, ob die geschlechtliche Zusammensetzung der Klassenkonferenz systematisch mit der Urbanität des Schulstandortes, mit dem Anteil von Kindern aus armen Verhältnissen an der Schule, mit der Ausstattung der Schule, oder mit dem Bundesland variiert (tabellarisch nicht ausgewiesen). Dabei erwiesen sich einzig die bekannten Bundeslandunterschiede als signifikante Prädiktoren: Im Osten gibt es weniger männliche Lehrkräfte (vgl. Diefenbach und Klein 2002). Da aber nach den IGLU Daten die Wahrscheinlichkeit einer Gymnasialempfehlung weder für Jungen noch für Mädchen im Osten größer ist als im Westen, kann der Ost-Effekt den berichteten bivariaten Zusammenhang nicht wegerklären. Zur weiteren Überprüfung der Befunde wurden logistische Regressionen zur Vorhersage der Gymnasialempfehlung jeweils für Mädchen und Jungen berechnet (vgl. Tabelle 2). Dabei wurde versucht, möglichst genau diejenigen Aspekte empirisch abzubilden, aus denen sich die Empfehlung laut den Vorgaben der Kultusministerkonferenz

ergeben soll. Zusätzlich flossen die soziale Herkunft sowie der Migrationshintergrund der Kinder in das Modell ein, die sich in vielen empirischen Studien als relevante Einflussfaktoren erwiesen haben. Es wurden für beide Geschlechter zunächst Modelle ohne die Schulnoten berechnet und anschließend Modelle, in denen die Schulnoten als zusätzliche Prädiktoren berücksichtigt wurden. Dargestellt sind vollstandardisierte Regressionskoeffizienten, die einerseits einen Vergleich der Effektstärken innerhalb eines Modells zulassen und andererseits einen Vergleich der Koeffizienten über genestete Modelle erlauben (Mood 2010; Winship und Mare 1984).¹²

Tabelle 2: Modellschätzungen zur Vorhersage der Gymnasialempfehlung am Ende der 4. Jahrgangsstufe (binäre logistische Regressionen)

| | Mädchen | | Jungen | |
|--|-------------|-------------|-------------|-------------|
| | M1 | M2 | J1 | J2 |
| männliche Konferenz | -0.0226** | -0.0099 | -0.0317*** | -0.0206** |
| gemischte Konferenz | -0.0300* | -0.0227* | -0.0125 | -0.0063 |
| weibliche Konferenz | <i>ref.</i> | <i>ref.</i> | <i>ref.</i> | <i>ref.</i> |
| Lesekompetenz | 0.1527*** | 0.0491*** | 0.1449*** | 0.0494*** |
| Mathematikkompetenz | 0.0860*** | 0.0196* | 0.0474*** | 0.0047 |
| Sachkundekompetenz | 0.0184 | 0.0027 | 0.0173 | 0.0128 |
| allg. kognitive Fähigkeiten | 0.0593*** | 0.0333*** | 0.0557*** | 0.0116 |
| Lernbereitschaft | 0.1704*** | 0.0431*** | 0.2134*** | 0.0522*** |
| Deutschnote | | -0.1338*** | | -0.1486*** |
| Mathematiknote | | -0.1314*** | | -0.1464*** |
| Sachkundenote | | -0.0832*** | | -0.0689*** |
| HISEI | 0.0878*** | 0.0489*** | 0.1064*** | 0.0571*** |
| Migrationshintergrund | -0.0122 | 0.0262*** | -0.0224 | 0.0175* |
| N | 2456 | 2456 | 2566 | 2566 |
| Wald Test aller Koeff. (Chi ²) | 359.37*** | 456.04*** | 425.17*** | 467.80*** |
| Wald Test (Chi ²) | 5.42* | 3.46 | 8.90** | 8.57** |
| $\beta_{m. \text{Konf}} = \beta_{g. \text{Konf}} = \beta_{w. \text{Konf}} = 0$ | | | | |
| Pseudo R ² | 0.45 | 0.64 | 0.44 | 0.65 |
| BIC | 232417.697 | 143724.675 | 235837.681 | 139696.186 |

Anmerkung: Voll-standardisierte β -Koeffizienten, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, robuste Schätzung der Standardfehler mit Berücksichtigung der Cluster (Schulen), gewichtete Daten.

Quelle: IGLU-E 2001, eigene Berechnungen

Die Angaben zur Modellpassung (McFadden R², BIC, Wald) zeigen, dass sich die Gymnasialempfehlung mit diesen Variablen recht gut vorhersagen lässt. Die multivariaten Analysen, in denen die von der Kultusministerkonferenz vorgegebenen „Kenntnisse und Fertigkeiten“ (=

¹² Bei der vollen Standardisierung werden alle X und die (latente) Y* Variable auf einen Mittelwert von 0 und eine Standardabweichung von 1 standardisiert. Die Standardisierungen wurden mit dem Stata ado *listcoeff* (Long und Freese 2006) berechnet.

Kompetenzen), „Eignung“ (= allgemeine kognitive Fähigkeiten) sowie „Neigung und Wille des Kindes zu geistiger Arbeit“ (= Lernbereitschaft und -verhalten) berücksichtigt werden, bestätigen insgesamt die bisherigen Erkenntnisse: Gemischte Konferenzen – der häufigste Fall (~60%) – vergeben an Mädchen nur tendenziell seltener Gymnasialempfehlungen als weibliche Konferenzen; bei Jungen ist diese Differenz nicht statistisch bedeutsam. Erneut zeigt sich aber auch, dass männliche Klassenkonferenzen alle Kinder signifikant seltener auf das Gymnasium überweisen als weibliche Klassenkonferenzen. Jungen werden von weiblichen Lehrkräften also nicht benachteiligt – im Gegenteil: weibliche Konferenzen erhöhen für Mädchen und Jungen die Chancen das Gymnasium zu besuchen.

5.5.2 Bewerten Lehrer strenger als Lehrerinnen?

Es stellt sich die Frage, ob die Schülerinnen und Schüler von männlichen Lehrern seltener auf das Gymnasium überwiesen werden, weil diese strenger bewerten, oder weil die Kinder bei ihnen weniger lernen und daher gewissermaßen „folgerichtig“ seltener eine Gymnasialempfehlung erhalten. Wie Tabelle 2 zeigt, bleiben die negativen Effekte männlicher Konferenzen unter Kontrolle der Kompetenzen bestehen, was auf eine strengere Bewertung durch männliche Lehrkräfte hindeutet. Diese Einschätzung bestätigt sich durch die Neugruppierung der Daten in Tabelle 3.

Tabelle 3: Leistungsindikatoren getrennt nach der geschlechtlichen Zusammensetzung der Klassenkonferenz

| | weibliche Konferenz | gemischte Konferenz | männliche Konferenz |
|---------------------------------|------------------------|------------------------|------------------------|
| Lesekompetenz | 511.70 _a | 504.02 _a | 506.92 _a |
| Mathematikkompetenz | 510.70 _a | 501.77 _a | 503.82 _a |
| Sachkundekompetenz | 507.27 _a | 501.36 _a | 503.47 _a |
| allg. kognitive Fähigkeiten | 48.80 _a | 48.18 _a | 49.09 _a |
| Gymnasialempfehlung | 39.5 % _a | 33.7 % _b | 26.9 % _c |
| Deutschnote | 2.72 _a | 2.81 _a | 2.94 _b |
| Mathematiknote | 2.67 _a | 2.78 _b | 2.87 _b |
| Sachkundenote | 2.45 _a | 2.55 _b | 2.72 _c |
| Lernverhalten und -bereitschaft | 2.93 _a | 2.92 _a | 2.82 _b |

Anmerkung: Mittelwerte mit unterschiedlichen Subskripten unterscheiden sich signifikant auf dem 5%-Niveau (2-seitiger T-Test), robuste Schätzung der Standardfehler mit Berücksichtigung der Cluster (Schulen), gewichtete Daten.

Quelle: IGLU-E 2001, eigene Berechnungen

Dargestellt sind die verschiedenen, für die GSE relevanten Leistungsindikatoren nach der Geschlechterzusammensetzung der Klassenkonferenz. Dabei zeigt sich, dass sowohl die Kompetenzen als auch die allgemeinen kognitiven Fähigkeiten in allen Fächern unabhängig vom Geschlecht der Lehrperson sind. Schlechtere Leistungen der Kinder können also nicht die selteneren Gymnasialempfehlungen bei männlichen Klassenkonferenzen erklären. Im zweiten Teil der Tabelle sind Bewertungsdimensionen abgetragen. Neben den bereits bekannten Unterschieden auf den Gymnasialempfehlungen ist erkennbar, dass es auch schon bei den Schulnoten signifikante Differenzen zwischen männlichen und weiblichen Klassenkonferenzen gibt: so werden Kinder in allen drei Fächern unabhängig von ihren Kompetenzen bei Lehrern schlechter bewertet als bei Lehrerinnen.¹³

Die GSE ist nicht die einzige Bewertung durch die Lehrpersonen. In erster Linie beurteilen Lehrkräfte die Leistung der Kinder in Form von Schulnoten, die dann wiederum in hohem Maße in die GSE einfließen. Ein Blick auf die Modellschätzungen M2 und J2 in Tabelle 2 zeigt, dass es hierbei scheinbar einen Geschlechtsunterschied gibt: Während Jungen auch unter Kontrolle der Schulnoten bei männlichen Konferenzen seltener eine Gymnasialempfehlung erhalten als bei weiblichen Konferenzen, verschwindet der Effekt bei den Mädchen. Bei ihnen ist die strengere Bewertung durch männliche Konferenzen also bereits komplett über die Noten vermittelt.

5.5.3 Wie wichtig ist die Zusammensetzung der Klassenkonferenz?

Insgesamt betrachtet hat die geschlechtliche Zusammensetzung der Konferenz verglichen mit anderen Einflussfaktoren keinen starken Effekt auf die Empfehlung. Ein Vergleich der vollstandardisierten Koeffizienten (Tabelle 2) untereinander ermöglicht Aussagen zur relativen Stärke der Prädiktoren. Dabei erweisen sich im Einklang mit früheren Studien bei simultaner Schätzung die Schulnoten – insbesondere in Deutsch – als die stärksten Prädiktoren (vgl. Bos et al. 2004b; Stubbe und Bos 2008).¹⁴ Die Lese- und Mathematikkompetenzen haben zwar einen starken Einfluss auf die Empfehlung, da sie aber zu einem großen Teil in die Noten einfließen und dadurch „absorbiert“ werden, haben sie bei einer simultanen Schätzung (unter Kontrolle der Noten) keinen großen eigenen Erklärungsanteil mehr. Die Sachkundekompetenz liefert in einer simultanen Schätzung keinen eigenständigen Erklärungsbeitrag zur Vorhersage

¹³ Interessanterweise zeigt sich darüber hinaus, dass die (von den Eltern wahrgenommene) Lernbereitschaft von Kindern in Klassen mit männlichen Konferenzen etwas schlechter ist als bei den übrigen Kindern.

¹⁴ Technisch bedeutet der Effekt der Deutschnote: Eine um eine Standardabweichung bessere Deutschnote erhöht die logarithmierte Chance auf eine Gymnasialempfehlung durchschnittlich um 0.13 Standardabweichungen.

der Empfehlung, anders als die allgemeinen kognitiven Fähigkeiten, deren relative Vorhersagestärke aber nicht besonders gewichtig ist. Einen erheblichen Einfluss auf die Empfehlung hat das Lernverhalten. Ähnlich wie die Lesekompetenz fließt das Lernverhalten in erster Linie indirekt, vermittelt über die Schulnoten, in die Empfehlung ein. Darüber hinaus bleibt aber auch unter Berücksichtigung der Noten ein nachweisbarer Effekt. Von den sozialen Merkmalen erweist sich die soziale Herkunft auch unter Berücksichtigung leistungsbezogener (primärer) Effekte (Boudon 1974) als wichtiger Prädiktor, während der Migrationshintergrund bei simultaner Berücksichtigung der übrigen Faktoren einen leichten positiven Effekt hat.

5.5.4 Warum werden Jungen seltener für das Gymnasium empfohlen?

Die Feminisierung der Schule ist also nicht dafür verantwortlich, dass Jungen seltener für das Gymnasium empfohlen werden – eher im Gegenteil. Woran liegt es dann, dass nur 33 Prozent der Jungen aber knapp 38 Prozent der Mädchen eine Gymnasialempfehlung bekommen? In Tabelle 4 sind (gepoolte) Schätzungen zur Vorhersage der Gymnasialempfehlung dargestellt, in denen das Geschlecht der Lernenden als Prädiktor berücksichtigt wird. In Modell 1 zeigt sich zunächst das bekannte Muster (vgl. z. B. Lehmann et al. 1997): Jungen werden seltener für das Gymnasium empfohlen als Mädchen. Verglichen mit der Benachteiligung der bildungsfernen Schichten (HISEI) sind die Geschlechterunterschiede aber klein. Das gilt auch, wenn zusätzlich die Kompetenzen und kognitiven Fähigkeiten berücksichtigt werden (Modell 2). Wird nun zusätzlich die Lernbereitschaft berücksichtigt, dreht sich der Effekt um. Anders ausgedrückt: Ein Junge der bei vergleichbaren Kompetenzen und Fähigkeiten eine gleichgroße Lernbereitschaft an den Tag legt wie ein Mädchen, wird häufiger – und nicht seltener – für das Gymnasium empfohlen. Die zentrale Bedeutung der Lernbereitschaft wird deutlich, wenn man die Koeffizienten vergleicht. In einem Modell ohne Noten (Modell 3) ist sie der stärkste Prädiktor, gefolgt von der Lesekompetenz und der sozialen Herkunft. Unter zusätzlicher Berücksichtigung der Noten (Modell 4) zeigt sich erneut, dass die Lernbereitschaft wie auch die Kompetenzen einen starken Zusammenhang mit den Noten aufweisen, die bei einer simultanen Schätzung alle anderen Einflussfaktoren überragen. Auf diesen Befund wird zusammen mit den übrigen Erkenntnissen dieses Beitrags in der abschließenden Diskussion eingegangen.

Tabelle 4: Modellschätzungen zur Vorhersage der Gymnasialempfehlung am Ende der 4. Jahrgangsstufe (Binäre logistische Regressionen)

| | M1 | M2 | M3 | M4 |
|---------------------------------|-------------|-------------|-------------|-------------|
| Junge | -0.0506*** | -0.0219** | 0.0200** | 0.0148** |
| männliche Konferenz | -0.0604*** | -0.0377*** | -0.0277*** | -0.0145** |
| gemischte Konferenz | -0.0512*** | -0.0201 | -0.0210 | -0.0145 |
| weibliche Konferenz | <i>ref.</i> | <i>ref.</i> | <i>ref.</i> | <i>ref.</i> |
| HISEI | 0.3362*** | 0.1124*** | 0.0974*** | 0.0533*** |
| Migrationshintergrund | -0.1042*** | -0.0040 | -0.0173* | 0.0220*** |
| Lesekompetenz | | 0.2141*** | 0.1498*** | 0.0487*** |
| Mathematikkompetenz | | 0.1168*** | 0.0661*** | 0.0112 |
| Sachkundekompetenz | | 0.0182 | 0.0186 | 0.0086 |
| allg. kognitive Fähigkeiten | | 0.0792*** | 0.0568*** | 0.0219*** |
| Lernverhalten und -bereitschaft | | | 0.1959*** | 0.0482*** |
| Deutschnote | | | | -0.1450*** |
| Mathematiknote | | | | -0.1387*** |
| Sachkundenote | | | | -0.0763*** |
| N | 5022 | 5022 | 5022 | 5022 |
| Wald Test aller Koeff. (Chi2) | 522.87 | 764.74 | 687.19 | 903.43 |
| Pseudo R ² | 0.10 | 0.34 | 0.44 | 0.64 |
| BIC | 771801.826 | 554494.252 | 466791.573 | 282188.572 |

Anmerkung: Voll-standardisierte β -Koeffizienten, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, robuste Schätzung der Standardfehler mit Berücksichtigung der Cluster (Schulen), gewichtete Daten

Quelle: IGLU-E 2001, eigene Berechnungen

5.6 Zusammenfassung und Diskussion

Sowohl in der fachwissenschaftlichen als auch in der bildungspolitischen Diskussion um die „Krise der Jungen“ erfährt die Feminisierung der Lehrerschaft zunehmende Aufmerksamkeit. Der Feminisierungsthese zufolge ist es dem gestiegenen Lehrerinnenanteil – vor allem in der Primarstufe – geschuldet, dass die Jungen strenger bewertet werden und in der Folge seltener auf das Gymnasium wechseln. Die vorliegende Studie findet keine Bestätigung für eine solche These. Jungen werden nicht von weiblichen Lehrkräften benachteiligt, wenn es um eine Gymnasialempfehlung geht. Dieses Ergebnis fügt sich in die Befundlage anderer aktueller Untersuchungen ein, nach denen Jungen auch im Hinblick auf die Kompetenzen und Noten nicht unter der Feminisierung leiden (Helbig 2010; Neugebauer et al. 2011). Die vorliegende Studie zeigt weiter, dass männliche Klassenkonferenzen verglichen mit weiblichen Klassenkonferenzen seltener Gymnasialempfehlungen aussprechen – und zwar für Mädchen und Jungen gleichermaßen. Lehrer bewerten bei gleicher Kompetenzlage der Kinder offensichtlich

etwas strenger als Lehrerinnen, oder umgekehrt ausgedrückt: Lehrerinnen bewerten permissiver. Woran liegt es dann, dass Jungen seltener für das Gymnasium empfohlen werden? In einem Modell, welches die Vorgaben der Kultusministerkonferenz möglichst passgenau abbildet, erweisen sich „Neigung und Wille des Kindes zu geistiger Arbeit“ – erfasst über die Lernbereitschaft – als entscheidende Faktoren zur Erklärung der Geschlechterdifferenz. Jungen haben eine geringere Lernbereitschaft als Mädchen.¹⁵ Ein Junge der bei vergleichbaren Kompetenzen und Fähigkeiten eine gleichgroße Lernbereitschaft an den Tag legt wie ein Mädchen, wird sogar etwas häufiger – und nicht seltener – für das Gymnasium empfohlen. Die Lehrkräfte benachteiligen die Jungen also nicht. Vielmehr folgen sie der Kultusministerkonferenz und berücksichtigen entsprechend den Vorgaben Faktoren, die von Mädchen im Schnitt besser erfüllt werden. Wenn überhaupt von einer Benachteiligung der Jungen gesprochen werden kann, dann werden diese durch die Vorgaben benachteiligt, die dem Verhalten der Mädchen entgegenkommen. Das hat aber nichts mit dem Geschlecht der Lehrkräfte zu tun. Ein Vergleich der Einflussfaktoren auf die Gymnasialempfehlung zeigt, dass das Geschlecht der Lehrkräfte insgesamt vernachlässigbar ist. Deutlich wichtiger ist neben der Lernbereitschaft vor allem die Lesekompetenz; beides beeinflusst die Noten und in einem weiteren Schritt die Empfehlungen in starkem Maße. Von den askriptiven Merkmalen erweist sich bei simultaner Schätzung lediglich die soziale Herkunft als bedeutsamer Einflussfaktor auf die Empfehlungsvergabe. Verglichen damit spielt das Geschlecht der Kinder nur eine kleine Rolle.

Ein Aspekt dieses Artikels scheint weiterhin erklärungsbedürftig. Warum ist der in Abbildung 1 gezeigte Zusammenhang zwischen dem Lehrerinnenanteil und dem geschlechtsspezifischen Bildungserfolg auf Individualebene nicht nachweisbar? Es kann angenommen werden, dass es sich dabei um einen klassischen ökologischen Fehlschluss handelt. Sowohl der gestiegene Anteil der Abiturientinnen als auch der gestiegene Anteil der Lehrerinnen ist vermutlich der Tatsache geschuldet, dass das weibliche Geschlecht seine vorhandenen Bildungspotenziale über die Jahrzehnte in zunehmendem Maße in entsprechende Bildungs- und Berufsabschlüsse umsetzt. Die Fokussierung auf die Jungen hat dazu geführt, dass die relativen Bildungserfolge des weiblichen Geschlechts fälschlicherweise als Bildungsmisserfolge des männlichen Geschlechts interpretiert werden. Ein stärkerer Fokus auf die Mädchen zeigt, dass diese schon immer lernwilliger waren und dadurch bessere Noten bekamen. Allerdings haben sie diese schulischen Vorteile lange nicht in höhere Bildungsabschlüsse umgesetzt. Schließlich war Frauen früher eine andere gesellschaftliche Rolle zugeordnet. Während der Mann für die

¹⁵ Darüber hinaus gibt es weitere Faktoren wie etwa das (mediale) Freizeitverhalten (Möble et al. 2006), aber schon dieser einzelne Faktor reicht aus um die Benachteiligung „wegzuerklären“.

bezahlte Erwerbsarbeit außer Haus verantwortlich war, sollte sich die Frau um die unbezahlte Erziehungs- und Haushaltsarbeit kümmern. Höhere Bildungs- oder gar Berufsabschlüsse gehörten nicht zur klassischen Rolle der Frau. Diese gesellschaftlichen Hürden sind heute jedoch zu großen Teilen ausgeräumt, so dass es sich für Mädchen zunehmend lohnt, ihr schulisches Potenzial in höhere Bildungsabschlüsse umzusetzen (Buchmann und DiPrete 2006). Damit zusammen hängt auch der gestiegene Lehrerinnenanteil. Ein genauerer Blick auf Abbildung 1 zeigt, dass auch schon Anfang der 1950er Jahre der Lehrerberuf für Frauen in besonderem Maße attraktiv war. Der Anteil der Lehrerinnen lag zwar bis 1968 unter 50 Prozent. Das lag aber daran, dass es insgesamt weniger Abiturientinnen gab – eine Grundvoraussetzung für ein Lehramtsstudium. In allen Jahren liegt jedoch der Anteil der Lehrerinnen über dem Anteil der Abiturientinnen. Mit anderen Worten: Schon immer haben sich überproportional viele Frauen die die Zugangsberechtigung zum Studium hatten, für den Lehrerberuf entschieden (vgl. auch Flyer und Rosen 1997). Dass in den letzten Jahrzehnten der Lehrerinnenanteil gestiegen ist, ist demnach ebenfalls der Tatsache geschuldet, dass Frauen in zunehmendem Maße entsprechende Abschlüsse erreichen. Mit dieser Erklärung werden auch der größere weibliche Bildungserfolg von Frauen und der höhere Lehrerinnenanteil in Ostdeutschland (vgl. Diefenbach und Klein 2002) nachvollziehbar. In der DDR gehörte die Gleichstellung der Frau zu den offiziellen Zielen der sozialistischen Gesellschaftspolitik (Geißler 2002, S. 365 ff.). Frauen sollten höhere Bildungsabschlüsse erzielen und gleichberechtigt auf dem Arbeitsmarkt partizipieren. Entsprechend gab es mehr Frauen mit der Möglichkeit zu studieren und in der Folge mehr Lehrerinnen. Genauso waren auch die von den Lehrerinnen unterrichteten Mädchen ambitionierter, ihre Bildungsreserven auszuschöpfen. Folglich findet sich auf Aggregatebene ein Zusammenhang zwischen dem Lehrerinnenanteil und dem Bildungserfolg – gemessen in Abschlüssen – von Mädchen. Dieser hat aber nichts mit einer Benachteiligung der Jungen durch mehr Lehrerinnen zu tun. Es ist also nicht zielführend, zur Verbesserung männlicher Bildungserfolge den Anteil der Männer unter den Lehrkräften zu erhöhen.

5.7 Literatur

- ABC-News. 2004. Scholarships to Target Declining Male Teacher Numbers. <http://www.abc.net.au/news/stories/2004/05/03/1100090.htm> (Zugegriffen: 08.07.2010).
- Ammermüller, Andreas, und Peter Dolton. 2006. Pupil-Teacher Gender Interaction Effects on Scholastic Outcomes in England and the USA. *ZEW Discussion Paper* 06-060: 1-32.
- Arnot, Madeleine, Miriam E. David und Gaby Weiner. 1999. *Closing the Gender Gap: Postwar Education and Social Change*. Cambridge, UK: Polity.
- Bacher, Johann, Martina Beham und Norbert Lachmayr. 2008. *Geschlechterunterschiede bei der Bildungswahl*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baker, David P., und Deborah Perkins Jones. 1993. Creating Gender Equality: Cross-National Gender Stratification and Mathematical Performance. *Sociology of Education* 66: 91-103.
- Bellenberg, Gabriele, Gertrud Hovestadt und Klaus Klemm. 2004. Selektivität und Durchlässigkeit im allgemein bildenden Schulsystem. Rechtliche Regelungen und Daten unter besonderer Berücksichtigung der Gleichwertigkeit von Abschlüssen: Arbeitsgruppe Bildungsforschung/Bildungsplanung, Universität Duisburg-Essen.
- Blossfeld, Hans Peter, Wilfried Bos, Bettina Hannover, Dieter Lenzen, Detlef Müller-Böling, Manfred Prenzel und Ludger Wößmann. 2009. *Geschlechtsdifferenzen im Bildungssystem. Jahresgutachten 2009 Aktionsrat Bildung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Blount, Jackie. 2000. Spinsters, Bachelors and Other Gender Transgressors in School Employment, 1850-1990. *Review of Educational Research* 70: 83-101.
- BMFSFJ. 2010. Pressemitteilung Bundesministerin Schröder: "Sowohl Mädchen als auch Jungen sollen die Chance haben, ihre Talente zu entdecken" vom 22.04.2010.
- Bos, Wilfried, Sabine Hornberg, Karl-Heinz Arnold, Gabriele Faust, Lilian Fried, Eva-Maria Lankes, Knut Schwippert und Renate Valtin (Hrsg.). 2007. *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther. 2004a. *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Bos, Wilfried, Andreas Voss, Eva-Maria Lankes, Knut Schwippert, Oliver Thiel und Renate Valtin. 2004b. Schullaufbahnpfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe. In *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*, Hrsg. Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther, 191-228. Münster: Waxmann.
- Bos, Wilfried, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walther, Renate Valtin und Gerd Walther. 2003. *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Boudon, Raymond. 1974. *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*. New York: Wiley.

- Buchmann, Claudia, und Thomas A. DiPrete. 2006. The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement. *American Sociological Review* 71: 515-541.
- Buchmann, Claudia, Thomas A. DiPrete und Anne McDaniel. 2008. Gender Inequalities in Education. *Annual Review of Sociology* 34: 319-337.
- Budde, Jürgen. 2009. Bildungs(miss)erfolge von Jungen in der Schule?! In *Jugendhilfe und Schule. Handbuch für eine gelingende Kooperation*. 2. Auflage, Hrsg. Angelika Henschel, Rolf Krüger, Christof Schmitt und Waldemar Stange, 394-408. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Cortina, Kai S., und Luitgard Trommer. 2003. Bildungswege und Bildungsbiographien in der Sekundarstufe I. In *Das Bildungswesen in der Bundesrepublik Deutschland*, Hrsg. Kai S. Cortina, Jürgen Baumert, Achim Leschinsky, Karl Ulrich Mayer und Luitgard Trommer, 342-391. Reinbek: Rowohlt.
- Dammasch, Frank. 2007. *Jungen in der Krise. Das schwache Geschlecht? Psychoanalytische Überlegungen*. Frankfurt am Main: Brandes & Apsel
- Dee, Thomas S. 2007. Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources* 42: 528-554.
- Diefenbach, Heike, und Michael Klein. 2002. "Bringing Boys Back In". Soziale Ungleichheit zwischen den Geschlechtern im Bildungssystem zuungunsten von Jungen am Beispiel der Sekundarschulabschlüsse. *Zeitschrift für Pädagogik* 48: 938-958.
- Ditton, Hartmut. 1992. *Ungleichheit und Mobilität durch Bildung. Theorie und empirische Untersuchung über sozialräumliche Aspekte von Bildungsentscheidungen*. Weinheim: Juventa.
- Ditton, Hartmut, und Jan Krüskens. 2006. Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft* 9: 348-372.
- Driessen, Geert. 2007. The Feminization of Primary Education: Effects of Teachers' Sex on Pupil Achievement, Attitudes and Behaviour. *Review of Education* 53: 183-203.
- Duckworth, Angela Lee, und Martin E. P. Seligman. 2006. Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores. *Journal of Educational Psychology* 98: 198-208.
- Eagly, Alice H., und Carole Chryala. 1986. Sex Differences in Conformity: Status and Gender-Role Interpretations. *Psychology of Woman Quarterly* 10: 203-220.
- Ehrenberg, Ronald G., Daniel G. Goldhaber und Dominic J. Brewer. 1995. Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review* 48: 547-561.
- Eurostat. 2009. Eurostat Database. <http://epp.eurostat.ec.europa.eu/portal/page/portal/education/data/database> (Zugegriffen: 08.07.2010).
- Flyer, Fredrick, und Sherwin Rosen. 1997. The New Economics of Teachers and Education. *Journal of Labor Economics* 15: 104-139.
- Ganzeboom, Harry B. G., Paul M. De Graaf, Donald J. Treiman und Jan De Leeuw. 1992. A Standard International Socio-Economic Index of Occupational Status. *Social Science Research* 21: 1-56.
- Geißler, Rainer. 2002. *Die Sozialstruktur Deutschlands*. Wiesbaden: Westdeutscher Verlag.

- Hadjar, Andreas, und Judith Lupatsch. 2011. Geschlechterunterschiede im Schulerfolg: Spielt die Lehrperson eine Rolle? *Zeitschrift für Soziologie der Erziehung und Sozialisation* 31: 79-94.
- Hannan, Geoff. 2001. *Improving Boys' Performance*. Oxford: Heinemann Educational.
- Helbig, Marcel. 2010. Sind Lehrerinnen für den geringeren Schulerfolg von Jungen verantwortlich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62: 93-111.
- Holmlund, Helena, und Krister Sund. 2008. Is the Gender Gap in School Performance affected by the Sex of the Teacher? *Labour Economics* 15: 37-53.
- Horstkemper, Marianne. 1999. Ausgewählte Aspekte der Lehrer-Schüler-Interaktion In *Koedukation. Erbe und Chancen*, Hrsg. Marianne Horstkemper, 139-158. Weinheim: Juventa.
- KMK. 2010. Übergang von der Grundschule in Schulen des Sekundarbereichs I und Förderung, Beobachtung und Orientierung in den Jahrgangsstufen 5 und 6 (sog. Orientierungsstufe). Informationsschrift des Sekretariats der Kultusministerkonferenz. www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_06_16-Uebergang-Grundschule-Orientierungsstufe.pdf (Zugegriffen: 12.08.2010).
- Lankes, Eva-Maria, Wilfried Bos, Joachim Mohr, Nike Plaßmeier, Knut Schwippert, Heiko Sibberns und Andreas Voss. 2003. Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU) und ihrer nationalen Erweiterung um Mathematik und Naturwissenschaften (IGLU-E). In *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*, Hrsg. Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walther und Renate Valtin, 7-28. Münster: Waxmann.
- Lehmann, Rainer H., Rainer Peek und Rüdiger Gänsfuß. 1997. Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern, die im Schuljahr 1996/97 eine fünfte Klasse an Hamburger Schulen besuchten. Hamburg: Behörde für Schule, Jugend und Berufsbildung.
- Long, J. Scott, und Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables using Stata*. College Station, TX: Stata Press.
- Maier, Uwe. 2007. Systematische Lehrereffekte bei Übergangsquoten auf weiterführende Schulen. *Zeitschrift für Erziehungswissenschaft* 10: 271-284.
- Marks, Gary N. 2007. Accounting for the Gender Gap in Reading and Mathematics: Evidence from 31 Countries. *Oxford Review of Education* 34: 1-21.
- Martino, Wayne J. 2008. Male Teachers as Role Models: Addressing Issues of Masculinity, Pedagogy and the Re-Masculinisation of Schooling. *Curriculum Inquiry* 38: 189-223.
- Mickelson, Roslyn A. 1989. Why Does Jane Read and Write so Well? The Anomaly of Women's Achievement. *Sociology of Education* 62: 47-63.
- Mood, Carina. 2010. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26: 67-82.
- Mößle, Thomas, Matthias Kleinmann, Florian Rehbein und Christian Pfeiffer. 2006. Medienutzung, Schulerfolg, Jugendgewalt und die Krise der Jungen. *Zeitschrift für Jugendkriminalrecht und Jugendhilfe* 3: 295-309.
- Müller, Walter, und Dieter Haun. 1994. Bildungsungleichheit im Sozialen Wandel. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 46: 1-42.

- Mullis, Ina V.S., Michael O. Martin, Eugenio J. Gonzalez und Ann M. Kennedy. 2003. *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools*. Chestnut Hill, MA: Boston College.
- Neugebauer, Martin. 2010. Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte. *Zeitschrift für Soziologie* 39: 202-214.
- Neugebauer, Martin, Marcel Helbig und Andreas Landmann. 2011. Unmasking the Myth of the Same-Sex Teacher Advantage. *European Sociological Review* 27: 669-689.
- Nowell, Amy, und Larry V. Hedges. 1998. Trends in Gender Differences in Academic Achievement from 1960 to 1994: An Analysis of Differences in Mean, Variance, and Extreme Scores. *Sex Roles* 39: 21-43.
- PISA-Konsortium, Deutsches (Hrsg.). 2000. *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske+Budrich.
- PISA-Konsortium, Deutsches (Hrsg.). 2005. *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* Münster: Waxmann.
- Preuss-Lausitz, Ulf. 2005. Anforderungen an eine jugendfreundliche Schule. *Die Deutsche Schule* 97: 222-235.
- Rodax, Klaus, und Klaus Hurrelmann. 1986. Die Bildungsbeteiligung für Mädchen und Frauen - ein Indikator für wachsende Chancengleichheit? *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*: 138-146.
- Rose, Lotte, und Ulrike Schmauch. 2005. *Jungen - die neuen Verlierer?* Königsstein: Helmer.
- Rury, John L. 1989. Who Became Teachers? The Social Characteristics of Teachers in American History. In *American Teachers: Histories of a Profession at Work*, Hrsg. Donald Warren, 9-48. New York: Macmillan.
- Sexton, Patricia C. 1969. *The Feminized Male: Classrooms, White Collars, & the Decline of Manliness*. New York: Random House.
- Sokal, Laura, Herb Katz, Les Chaszewski und Cecilia Wojcik. 2007. Good-bye, Mr. Chips: Male Teacher Shortages and Boys' Reading Achievement. *Sex Roles* 56: 651-659.
- Sommers, Christina H. 2000. *The War Against Boys: How Misguided Feminism Is Harming Our Young Men*. New York: Touchstone.
- Stamm, Margrit. 2008. Underachievement von Jungen: Perspektiven eines internationalen Diskurses. *Zeitschrift für Erziehungswissenschaft* 11: 106-124.
- Steinmayr, Ricarda, und Birgit Spinath. 2008. Sex Differences in School Achievement: What are the Roles of Personality and Achievement Motivation? *European Journal of Personality* 22: 185-209.
- Stubbe, Tobias C., und Wilfried Bos. 2008. Schullaufbahneempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik* 22: 49-63.
- TDA. 2009. Wanted: Male Primary School Teachers. Training and Development Agency for Schools (TDA). www.tda.gov.uk/about/mediarelations/2009/200909.aspx (Zugegriffen: 08.07.2010).
- UNESCO. 2009. *Global Education Digest 2009: Comparing Education Statistics Across the World*. Montreal: UNESCO Institute for Statistic.

- Valtin, Renate, Christine Wagner und Knut Schwippert. 2005. Schülerinnen und Schüler am Ende der vierten Klasse - schulische Leistungen, lernbezogene Einstellungen und außerschulische Lernbedingungen. In *IGLU - Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien*, Hrsg. Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther, 187-238. Münster: Waxman.
- Wiese, Wilhelm. 1982. Elternstatus, Lehrerempfehlung und Schullaufbahn: Eine empirische Analyse des Einflusses des Grundschullehrers auf die Bildungslaufbahn des Schülers. *Zeitschrift für Soziologie* 11: 49-63.
- Willingham, Warren W., und Nancy S. Cole. 1997. *Gender and Fair Assessment*. Mahwah, NJ: Lawrence Erlbaum.
- Winship, Christopher, und Robert D. Mare. 1984. Regression Models with Ordinal Variables. *American Sociological Review* 49: 512-525.

Kapitel 6 Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte.

Erschienen als: Neugebauer, Martin. 2010. Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte. *Zeitschrift für Soziologie* 39, 202-214. Abdruck mit Erlaubnis von Lucius & Lucius.¹

Abstract: In Anlehnung an die klassische Arbeit von Boudon (1974) gibt es zwei Ursachenkomplexe für soziale Bildungsungleichheit: primäre Effekte, die sich im Zusammenhang von sozialer Herkunft und Schulleistungen ausdrücken, und sekundäre Effekte, die schichtspezifische Übergangsnäigungen widerspiegeln, selbst bei gleichen schulischen Leistungen. Die Abschätzung der relativen Bedeutung von primären und sekundären Effekten ist das Ziel einer wachsenden Anzahl internationaler Beiträge. Die vorliegende Untersuchung möchte hierzu einen Beitrag leisten, indem sie erstens die relative Bedeutung der Effekte am Übergang auf das Gymnasium in Deutschland abschätzt und zweitens untersucht, inwieweit die bundeslandspezifische Ausgestaltung der Grundschulempfehlung die relative Bedeutung von primären zu sekundären Effekten beeinflusst. Mit bundesweiten Längsschnittdaten des DJI Kinderpanels der Jahre 2002 bis 2005 lassen sich primäre und sekundäre Effekte durch eine neue kontrafaktische Dekompositionsanalyse zerlegen. Es wird gezeigt, dass sekundäre Effekte 59 Prozent der Bildungsungleichheit am Übergang auf das Gymnasium ausmachen. Besonders bedeutsam sind sie bei Kindern im mittleren Notenbereich. Weiterhin finden sich Hinweise, dass der relative Erklärungsbeitrag von sekundären Effekten größer ist, wenn Eltern frei entscheiden können und geringer, wenn die elterliche Wahlfreiheit durch eine verbindliche Grundschulempfehlung eingeschränkt wird. Theoretische und bildungspolitische Konsequenzen werden diskutiert.

6.1 Einleitung

Kinder aus niedrigeren sozialen Schichten gehen weitaus seltener auf das Gymnasium als Kinder aus höheren sozialen Schichten. In Anlehnung an die klassische Arbeit von Boudon (1974, S. 29 ff.) wird bei den Ursachen dieser Ungleichheit häufig zwischen primären und sekundären Effekten unterschieden.² Primäre Effekte ergeben sich, weil Kinder aus niedrigeren Sozial-schichten infolge geringerer Anregungen, niedrigerer kultureller und materieller Ausstattung

¹ Dieser Beitrag basiert zu großen Teilen auf meiner 2008 eingereichten Diplomarbeit an der Universität Mannheim.

² Interessanterweise findet sich die Einteilung in primäre und sekundäre Effekte bereits bei Girard und Bastide (1963), allerdings hat Boudon (1974) das Konzept in der Forschungsgemeinschaft populär gemacht.

im Elternhaus und möglicherweise auch aufgrund genetischer Voraussetzungen schlechtere schulische Leistungen erzielen. Deshalb scheitern sie eher an den Selektionshürden des Bildungssystems, insbesondere beim Übergang auf das Gymnasium. Sekundäre Effekte hingegen beruhen auf Unterschieden im Bildungsverhalten: Selbst bei gleichen Leistungen variieren Bildungsentscheidungen nach sozialer Herkunft.³ Es gibt inzwischen verschiedene empirische Überprüfungen und Anwendungen des von Boudon (1974) vorgestellten und von neueren weiterwartungstheoretischen Ansätzen erweiterten und formalisierten Modells (für einen Überblick vgl. Goldthorpe 2007, S. 73 ff.). Bislang jedoch wissen wir wenig über die relative Bedeutung der beiden Effekttypen. Diese Frage ist theoretisch und bildungspolitisch relevant, weil die beiden Effekte durch unterschiedliche Ursachenkomplexe und Mechanismen zustande kommen (Müller-Benedict 2007). Nur wenn man die relative Bedeutung der beiden Ursachenkomplexe kennt, lassen sich politische Maßnahmen mit dem Ziel der Ungleichheitsreduktion informiert priorisieren. Auch die Beurteilung und Wahrnehmung von Leistungsgerechtigkeit im Bildungssystem ist mit der relativen Bedeutung der Effekte verbunden, denn vor allem Ungleichheiten bei Übergängen zu weiterführender Bildung, die sich als unabhängig von gezeigten Schulleistungen erweisen, werden kritisch beurteilt (vgl. Baumert et al. 2003). Wie groß ist also dieser leistungsunabhängige (sekundäre) Anteil?

In jüngster Zeit widmet sich eine wachsende Anzahl internationaler Aufsätze der Quantifizierung primärer und sekundärer Effekte an Bildungsübergängen in verschiedenen Ländern (Contini et al. 2008; Erikson und Rudolphi 2009; Jackson et al. 2007; Kloosterman et al. 2009; Schindler und Reimer 2010). Diese Aufsätze verwenden eine neuartige Methode der Effektkomposition. Im vorliegenden Beitrag möchte ich diesen Forschungsstrang ergänzen, indem ich *mit der gleichen Methode Ergebnisse zum relativen Gewicht primärer und sekundärer Herkunftseffekte bei dem in Deutschland entscheidenden Übergang auf das Gymnasium liefere*. Bildungsungleichheit hängt von den institutionellen Gegebenheiten des Bildungssystems ab (z. B. Pfeffer 2008). Als zweiten Schwerpunkt greift der Beitrag daher ein zentrales institutionelles Merkmal heraus – die Grundschulempfehlung – und untersucht, inwiefern Unterschiede in der Empfehlungspraxis die relative Bedeutung von primären und sekundären Effekten beeinflussen. In den letzten Jahren hat sich ein Trend abgezeichnet, die Verantwortung der ersten Bildungsentscheidung zunehmend auf die Lehrer zu übertragen (Arnold et al. 2007, S. 292). Beeinträchtigt diese Verschiebung die Chancengleichheit? Zur Beantwortung dieser Frage wird

³ Diese Entscheidungsmuster sind geprägt durch das schichtspezifisch variiende Motiv des intergenerationellen Stauerhalts, unterschiedliche Erfolgserwartungen und schichtabhängigen Kosten-Nutzen-Relationen von Bildungsentscheidungen (vgl. Breen und Goldthorpe 1997; Erikson und Jonsson 1996; Esser 1999, S. 265 ff.).

untersucht, *inwieweit die nach Bundesländern variierende Einflussnahme der Lehrer das relative Gewicht primärer und sekundärer Effekte beeinflusst.*

6.2 Forschungsstand: Relative Bedeutung primärer und sekundärer Effekte

Die relative Bedeutung primärer und sekundärer Effekte variiert deutlich nach Ländern: Eine vergleichsweise geringe Bedeutung sekundärer Effekte beim Übergang zu weiterführender Bildung findet sich in Großbritannien (Jackson et al. 2007) und Schweden (Erikson und Rudolphi 2009), wo sie nur etwa ein Drittel der Bildungsungleichheit ausmachen. Etwas stärkere Anteile sekundärer Effekte (40-48%) finden Kloosterman et al. (2009) für die Niederlande. Noch gewichtiger scheinen sekundäre Effekte in Italien zu sein, wo sie nach Befunden von Contini et al. (2008) rund 60 bis 71 Prozent der Ungleichheit am Übergang in die Sekundarstufe I ausmachen.

Für Deutschland kommt Müller-Benedict (2007) auf Basis von Querschnittsdaten über 15-Jährige zu dem Schluss, dass primäre und sekundäre Effekte beim Übergang auf das Gymnasium in etwa von gleicher Bedeutung sind, während Stocké (2007) mit Längsschnittdaten aus Rheinland-Pfalz je nach Operationalisierung einen sekundären Anteil von 30-82% errechnet. Beide Studien verwenden andere Methoden als die sich international durchsetzende Effektdekompositionsmethode nach Erikson et al. (2005). Eine Anwendung der Dekompositionsmethode auf Deutschland erscheint erstens aufgrund ihrer konzeptionellen Klarheit sinnvoll; zweitens sollte sie es ermöglichen, international vergleichbare Ergebnisse zu erlangen.

6.3 Freier Elternwille versus verbindliche Grundschulempfehlung – substantielle Länderunterschiede

In Deutschland treffen Familien Bildungsentscheidungen nicht alleine. Lehrer beeinflussen den schulischen Werdegang durch ihre Grundschulempfehlung (GSE)⁴ maßgeblich mit (z. B. Becker 2000; Bos et al. 2004; Ditton und Krüsken 2006; Lehmann et al. 1997; Wiese 1982). Mit

⁴ Die Empfehlung der Grundschule wird je nach Bundesland anders bezeichnet. Um Wortverwirrungen zu vermeiden werden im Folgenden einheitlich die Begriffe „GSE“ oder „Empfehlung“ verwendet.

anderen Worten: Auch Lehrer treffen in Form der GSE eine „Bildungsentscheidung“ für das jeweilige Kind. Generell sind die Übergangsregelungen von Bemühungen geprägt, einen Konsens zwischen der Lehrermeinung und den Vorstellungen der Eltern herzustellen (Cortina und Trommer 2003). Im Falle einer Diskrepanz findet sich allerdings ein grundlegender Unterschied zwischen zwei Gruppen von Bundesländern hinsichtlich der „Entscheidungshoheit“: In der ersten Ländergruppe (10 von 16 Bundesländern) ist die GSE eine „Empfehlung“ im eigentlichen Wortsinn und also nicht verbindlich – sie stellt für die Eltern lediglich eine Einschätzung der Lehrer hinsichtlich der Angemessenheit einer bestimmten weiterführenden Schulart dar. In der zweiten Ländergruppe (Baden-Württemberg, Bayern, Brandenburg, Sachsen, Saarland und Thüringen) sind die Eltern an die GSE gebunden.⁵ Im Falle einer Diskrepanz ist in diesen Ländern ein Beratungsgespräch mit den Eltern vorgesehen, verbunden mit der Möglichkeit eines zusätzlichen Eignungstests. Die Lehrer haben gleichwohl das „letzte Wort“: Wenn bestimmte Leistungskriterien nicht erfüllt werden, wird der Zugang zum Gymnasium verwehrt.⁶

Inwiefern beeinflusst die „Entscheidungshoheit“ die relative Bedeutung von primären und sekundären Effekten? Vieles spricht dafür, dass Lehrer ihre Entscheidung insgesamt in deutlich geringerem Maße von der sozioökonomischen Situation des Elternhauses abhängig machen als die Eltern (Ditton et al. 2005). Aus werterwartungstheoretischer Sicht ist dies gut nachvollziehbar: Da die typischerweise für die Elternentscheidung ins Feld geführten Erklärungen (wie das Motiv des familiären Stuserhalts und die sozialschichtabhängige Kosten-Nutzen-Relationen) für die Lehrer keine Rolle spielen sollten, sollten subjektive Erfolgserwartung die Entscheidung der Lehrer dominieren. Es erhalten dann diejenigen Schüler eine Gymnasialempfehlung, deren Schulleistungen so gut sind, dass sie mit hoher Wahrscheinlichkeit das Gymnasium erfolgreich abschließen werden. Dieser „primäre“ Prozess entspricht dem „normativen Selbstverständnis des Lehrerberufs“ (Wiese 1982, S. 52). Doch auch Lehrer treffen ihre Entscheidungen nicht unabhängig von leistungsfremden Herkunftseinflüssen der Kinder, wie einschlägige Studien zeigen (Lehman et al. 1997; Bos et al. 2004). Wird keine bewusste Diskriminierung durch die Lehrer unterstellt, lassen sich diese schichtspezifischen GSEs ebenfalls über die Erfolgserwartung fassen. Zum Beispiel ist es plausibel, dass Lehrer prognoserelevante motivationale Merkmale berücksichtigen, die mit der sozialen Herkunft eines Kindes korrelieren (Baumert et al. 2003). Weiterhin antizipieren Lehrer möglicherweise

⁵ Die Darstellung entspricht dem Stand 2004/05, als die hier untersuchte Schülerkohorte auf die weiterführende Schule gewechselt ist. Eine Übersicht der Übertrittsbestimmungen auf das Gymnasium nach Bundesländern findet sich im Anhang.

⁶ In diesen Ländern muss i.d.R. zudem ein festes Notenkriterium erreicht werden.

eine bessere Unterstützungsleistung der Elternhäuser bei Kindern aus höheren Schichten. Auch ist es theoretisch nachvollziehbar, dass Eltern mit hohen Bildungsaspirationen oder Statusinteressen häufiger Kontakt zu Lehrern suchen, um diese in ihrer Empfehlung zu beeinflussen. Dieses Engagement kann von den Lehrern als günstige Voraussetzung für das erfolgreiche Bestehen einer anspruchsvolleren Schulart gewertet werden. Die Lehrerentscheidung sollte verglichen mit der Elternentscheidung stärker an den Leistungen des Kindes orientiert sein (Ditton et al. 2005), doch sekundäre Effekte können auch bei ihr eine gewichtige Rollen spielen.

Die Grundschullehrer beeinflussen durch die Erteilung der GSE die Schulwahl also maßgeblich mit. In Ländern mit verbindlicher GSE sollte der Spielraum für die Durchsetzung schichtverzerrter Elternpräferenzen zugunsten einer durch die Lehrer induzierten, meritokratischen Schulwahl eingeschränkt sein.

6.4 Daten und Operationalisierungen

6.4.1 Datenbasis

Die Datenbasis der vorliegenden Arbeit bildet die Schulkinder-Kohorte des auf drei Wellen angelegten DJI-Kinderpanels (Alt und Quellenberg 2005; Infas 2006).⁷ Es wurden neben den Kindern jeweils auch die Eltern befragt. Die Befragung für die erste Welle erfolgte im Herbst 2002, in der 1042 Interviews realisiert wurden (Ausschöpfungsquote 50.6%). Zu diesem Zeitpunkt befanden sich 97% der Kinder in der dritten oder zweiten Grundschulklasse. Im Abstand von jeweils ca. anderthalb Jahren wurden dann die zweite (Frühjahr 2004) und dritte Welle (Herbst 2005) realisiert. Wie in allen Längsschnittstudien kam es auch im DJI-Kinderpanel zu Panelmortalität.⁸ Ausfallanalysen haben gezeigt, dass der Ausfall mit Blick auf die interessierenden Variablen – schulische Leistungen und Bildungshintergrund – nicht selektiv ist und die Schätzer nicht verzerren sollte. Mit Blick auf die Item-Nonresponse hat sich die Einsicht durchgesetzt, dass die Schätzung fehlender Werte durch multiple Imputationen gegenüber dem paar- oder fallweisen Ausschluss vorzuziehen ist (Rubin 1987; Allison 2001). Um die Unsicherheit aufgrund fehlender Werte zu berücksichtigen, wurden mit Hilfe des Stata Programms ICE

⁷ Die Grundgesamtheit der ersten Welle besteht aus in Deutschland lebenden Kindern im Alter von 8 bis 9 Jahren, die zwischen Oktober 1993 und September 1994 geboren wurden. Dazu wurden auf Basis des Einwohnermeldeamtsregisters Adressen aus 100 repräsentativ ausgewählten Gemeinden gezogen.

⁸ Von 1042 vollständigen Interviews der ersten Welle konnten in der zweiten Welle 722 (69.3%) und in der dritten Welle noch 618 (59.3%) realisiert werden.

(Royston 2007) für jeden fehlenden Wert 10 plausible Werte erzeugt.⁹ Dem Vorschlag von Hippels (2007) folgend wurden nach der Imputation Beobachtungen gelöscht, die ursprünglich einen fehlenden Wert bei der abhängigen Variable „gewählte weiterführende Schulart“ hatten. Dadurch verbleiben 606 Beobachtungen im Sample. Die Analysen wurden nach dem von Rubin (1987) vorgeschlagenen Verfahren kombiniert.¹⁰

6.4.2 Operationalisierungen

Bildungsentscheidung: Die abhängige Variable ist die gewählte weiterführende Schulart. Die Vielfalt der Schulsysteme hat zur Folge, dass systematische Vergleiche über die Bundesländer nur für das Gymnasium möglich sind, das als einzige Schulart überall existiert.¹¹ Aus diesem Grund wird die abhängige Variable für die späteren Analysen dichotomisiert in eine Entscheidung für oder gegen das Gymnasium.¹² Von 606 Familien haben sich 283 (47%) eindeutig für und 255 (42%) eindeutig gegen das Gymnasium entschieden. In 68 Familien (11%) ist zum Zeitpunkt der dritten Welle noch keine eindeutige Entscheidung gefallen, weil die Kinder ein Schuljahr wiederholen, oder weil sie eine schulartunabhängige Orientierungsstufe, eine 6-jährige Grundschule, oder eine weiterführende Schule besuchen, auf der sowohl ein Abitur als auch ein anderer Abschluss erzielt werden kann. Diese Kinder werden für die folgenden Analysen ausgeschlossen, so dass ein Analysesample von 538 Fällen verbleibt.

Soziale Herkunft: Die soziale Herkunft des Kindes wird über den höchsten allgemeinen Bildungsabschluss der Eltern operationalisiert. Die geringe Fallzahl im Analysesample macht es erforderlich, soziale Herkunft zu dichotomisieren. Dazu wird der Bildungsstand der Eltern zusammengefasst zu Familien in denen mindestens ein Elternteil über die (Fach-) Hochschulreife verfügt und solchen, in denen beide über einen geringeren Abschluss als die (Fach-) Hochschulreife verfügen. Eine solche Dichotomisierung ist vor dem Hintergrund differenzierter familiärer Bildungsmilieus sicherlich eine starke Vereinfachung der Realität, erweist sich

⁹ Mit rund 33 Prozent weist der Bildungsabschluss des Vaters den mit Abstand höchsten Anteil fehlender Werte auf.

¹⁰ Es wurden zusätzlich Analysen ohne multipel imputierte Daten durchgeführt. Die Ergebnisse sind im Wesentlichen gleich, die Standardfehler sind kleiner bei Verwendung multipel imputierter Daten.

¹¹ So gibt es die Hauptschule als eigenständige Schulart nur in 10 von 16 Ländern, die Realschule sogar nur in 8, was in erster Linie darauf zurückzuführen ist, dass einige Länder diese Schularten zusammengefasst haben.

¹² Durch dieses Vorgehen können zudem solche Kinder einer eindeutigen Entscheidungsalternative zugeordnet werden, die eine Schule mit mehreren Bildungsgängen (SMBG) besuchen, auf denen kein Abitur möglich ist. Auch Kinder, die eine kooperative Gesamtschule besuchen, lassen sich in aller Regel einem eindeutigen Schulzweig und damit einer Entscheidungsalternative zuordnen.

jedoch im Hinblick auf die Bildungschancen der Kinder als guter Prädiktor, der mehr Varianz erklärt als eine Operationalisierung nach Herkunftsklassen.

Primäre Effekte: Die grundlegende Entscheidung bei der Operationalisierung primärer Effekte ist die Frage, ob man darunter Unterschiede im Leistungspotenzial oder Unterschiede im Schulerfolg versteht. Ich beziehe mich hier auf Erikson und Rudolphi (2009, S. 11) und betrachte den Schulerfolg als den theoretisch entscheidenden Faktor, denn dieser ist in der Entscheidungssituation sichtbar und entscheidungsrelevant (vgl. auch Jackson et al. 2007). Die Messung von Schulerfolg geschieht über den Durchschnitt aus den Noten der Hauptfächer Deutsch und Mathematik, die im Entscheidungsprozess von hoher Bedeutung sind und den Familien signalisiert, wie erfolgreich das Kind eine bestimmte Schulart besuchen kann. Herangezogen werden die Noten aus dem Halbjahreszeugnis der 4. Klassenstufe, also die zum Entscheidungszeitpunkt vorliegenden Noten.

Insgesamt ist dies sicherlich keine ideale Datenbasis für die hier berichteten Analysen – ein größeres Panel wäre wünschenswert. Auch die Operationalisierung primärer Effekte über die Durchschnittsnote lässt sich kritisieren, weil sie Elemente enthält, die eher als sekundäre Effekte gelten können (etwa leistungsfremde Herkunftseinflüsse in der Notengebung). Zur Eingrenzung solcher Elemente wären Daten aus standardisierten Intelligenz- oder Leistungstests erforderlich, die diese Datenbasis nicht bietet. In der Interpretation ist daher zu berücksichtigen, dass durch die Verwendung der Durchschnittsnote der Anteil sekundärer Effekte tendenziell unterschätzt wird. Für die Verwendung der DJI-Daten spricht, dass der Datensatz Bildungverläufe aus allen 16 Bundesländern abbildet und daher über institutionelle Variation im Hinblick auf die Verbindlichkeit der GSE verfügt. Weiterhin erfassen die Daten den Bildungsweg im Längsschnitt – und ermöglichen somit eine Separierung von individuellen Leistungs- und herkunftsbedingten Entscheidungsmerkmalen. Bisher gibt es in Deutschland meines Wissens keinen anderen Datensatz, der diese notwendigen Kriterien für die Untersuchung der Forschungsfrage besser erfüllen würde.

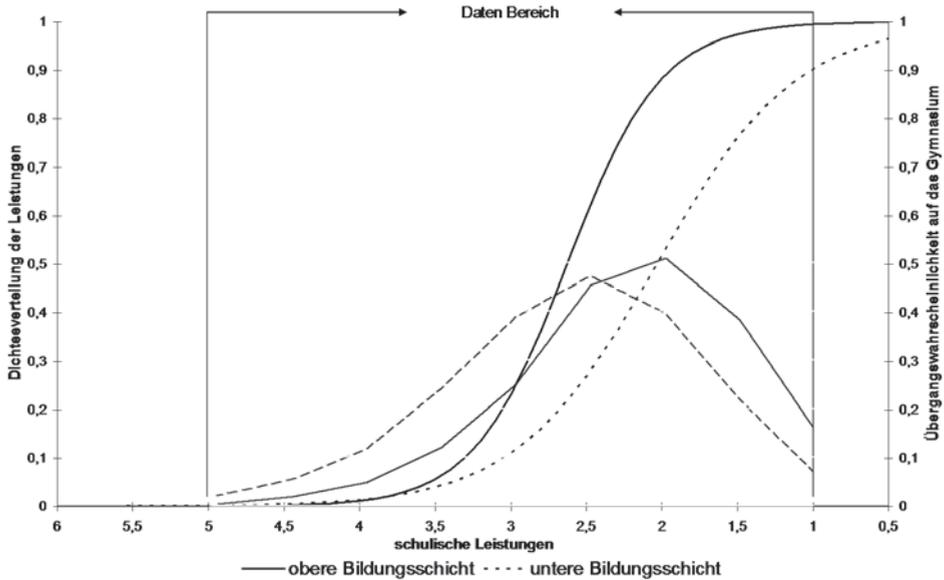
6.5 Ergebnisse

6.5.1 Soziale Herkunft und schulische Leistungen beim Übergang auf das Gymnasium

Für Gesamtdeutschland zeigt Abbildung 1 den Zusammenhang zwischen sekundären Effekten und über schulische Leistungen vermittelten primären Effekten. Dargestellt sind vorhergesagte Übergangswahrscheinlichkeiten (auf Basis logistischer Regressionen) auf das Gymnasium für

beide Bildungsschichten, und zwar an jedem Punkt der Leistungsverteilung. Im Folgenden wird diese konditionale, von Leistung und Schichtzugehörigkeit abhängige Übergangswahrscheinlichkeit als Übergangsneigung bezeichnet. Zusätzlich sind die Verteilungen der schulischen Leistungen beider Schichten abgetragen (Kerndichteschätzer mit Bandbreite 0.4).

Abbildung 1: Wahrscheinlichkeit des Übergangs auf das Gymnasium in Abhängigkeit von der schulischen Leistung und Kerndichteschätzer der Leistungsverteilung, nach Schichtzugehörigkeit



Aus der Abbildung werden unmittelbar zwei Tatsachen ersichtlich: Erstens unterscheiden sich die Leistungsverteilungen der zwei Bildungsschichten in der vermuteten Art und Weise; Kinder aus bildungsnahen Elternhäusern haben bessere Schulnoten als bildungsferne Kinder. Diese unterschiedlich gelagerten Leistungsverteilungen spiegeln die primären Effekte wider, die offensichtlich am Übergang von der Primarstufe auf das Gymnasium wirken. Zweitens haben die bildungsnahen Kinder an jedem Punkt der Leistungsverteilung höhere Übergangsneigungen auf das Gymnasium.

Betrachten wir den Verlauf genauer: Während die Differenzen der Übergangsneigungen im unteren und oberen Leistungsbereich gering sind, öffnen sich die Abstände zur Mitte der Leistungsverteilung hin. Vor allem im mittleren Notenbereich fallen bildungsnah und bildungsferne Familien also unterschiedliche Bildungsentscheidungen, während die Differenzen

im Entscheidungsverhalten an den Rändern der Notenverteilung gering sind. Aus werterwartungstheoretischer Sicht ist dies folgerichtig: Ein Kind mit einer Durchschnittsnote von 1.0 werden Eltern, ganz gleich welcher Bildungsschicht sie angehören, kaum auf die Real- oder Hauptschule schicken. Umgekehrt wird ein Kind, das knapp an einer Nichtversetzung „vorbeigeschrammt“ ist, nicht auf das Gymnasium wechseln – selbst wenn die Eltern aus der oberen Bildungsschicht kommen und daher, sei es aufgrund von hohen Bildungsaspirationen, rationalen Abwägungsprozessen oder beidem, besondere Anreize dazu verspüren. Im mittleren Notenbereich hingegen, in dem die Erfolgswahrscheinlichkeiten weniger eindeutig ausfallen, ist der Wirkungsspielraum für sekundäre Effekte am größten.¹³

6.5.2 Effektdekomposition

Sind nun aber primäre oder sekundäre Effekte wichtiger für die Entstehung von Bildungsungleichheit beim Übergang auf das Gymnasium? Zur Beantwortung dieser Frage wird der Gesamteffekt der Bildungsungleichheit zwischen den Schichten, ausgedrückt durch die Odds Ratio, nach der Methode von Erikson et al. (2005) zerlegt in einen primären und einen sekundären Anteil.¹⁴ Dieser Methode liegt eine kontrafaktische Idee zugrunde: Die Notenverteilung einer Bildungsschicht kann mit der Übergangsnegung einer anderen Bildungsschicht kombiniert werden. Dadurch können erstens Übergangsraten von kontrafaktischen Gruppen berechnet und durch den Vergleich mit den Übergangsraten tatsächlicher Gruppen primäre und sekundäre Effekte isoliert werden. Zweitens kann über mathematische Umformungen der relative Erklärungsanteil der Effekte berechnet werden. Aus Platzgründen wird zu den Details des Verfahrens auf die Darstellung bei Jackson et al. (2007) verwiesen.

Tatsächliche und kontrafaktische Übergangsraten in das Gymnasium sind in Tabelle 1 abgetragen. Das erste Subskript repräsentiert jeweils die Leistungsverteilung, das zweite Subskript die Übergangsnegung; „o“ steht für obere Bildungsschicht, „u“ für untere Bildungsschicht. Die Bedeutung sekundärer Effekte wird deutlich, wenn man zwei Gruppen

¹³ Es kann vermutet werden, dass die größeren Abstände im mittleren Notenbereich durch die logistische Funktion zustandekommen, die allein durch ihre funktionale Form engere Abstände an den Rändern produziert. Um diesen Einwand zu überprüfen, wurden Lowess Regressionen für die tatsächlichen Übergangsraten der beiden Schichten geplottet. Dieses nicht-parametrische Vorgehen bestätigt, dass es sich bei der beobachteten Vergrößerung der Abstände nicht um ein Artefakt aufgrund der logistischen Linkfunktion, sondern um ein inhaltliches Ergebnis handelt (siehe Anhang).

¹⁴ Dabei wird eine Spezifikation von Buis (2010) verwendet, in der die unrealistische Annahme normalverteilter Schulleistungen fallengelassen wird zugunsten der tatsächlichen empirischen Verteilung. Die Analysen wurden mit dem STATA ado *ldecomp* (ebenfalls Buis 2010) gerechnet.

vergleicht (eine kontrafaktische und eine tatsächliche), die sich nur im Hinblick auf die bergangsnegung unterscheiden. Zum Beispiel ist die tatsächliche Übergangsnegung der oberen Bildungsschicht (P_{oo}) 73 Prozent. Hätten diese Kinder bei gleicher Leistung die Übergangsnegung der unteren Bildungsschicht (P_{ou}), würden lediglich 48 Prozent auf das Gymnasium wechseln. Durch den Vergleich der beiden Gruppen P_{oo} und P_{ou} lässt sich der sekundäre Effekt „isolieren“, da sich die Gruppen nur in ihrer Übergangsnegung unterscheiden, während die Leistungsverteilung identisch ist. In ähnlicher Weise lässt sich der Einfluss von primären Effekten „isolieren“, wenn etwa P_{oo} (73%) mit P_{uo} (56%) verglichen wird. Hier beträgt der Unterschied in den Übergangsnegungen 17 Prozentpunkte und ist ausschließlich auf Leistungsunterschiede zurückzuführen.¹⁵

Tabelle 1: Tatsächliche und kontrafaktische Übergangsnegungen auf das Gymnasium (gemittelt über multiple imputierte Datensätze)

| Tatsächliche Kombinationen | Übergangsnegungen auf das Gymnasium (in %) |
|-------------------------------|--|
| P_{oo} | 73 |
| P_{uu} | 31 |
| Kontrafaktische Kombinationen | |
| P_{ou} | 48 |
| P_{uo} | 56 |

Im nächsten Schritt werden nun die tatsächlichen und kontrafaktischen Übergangsnegungen nach der Methode von Erikson et al. (2005) kombiniert und transformiert. Der sekundäre Effekt lässt sich dann als prozentualer Anteil am Gesamteffekt der Bildungsungleichheit ausdrücken (Tabelle 2). Sekundäre Effekte verantworten 59 Prozent der Bildungsungleichheit am Übergang auf das Gymnasium, primäre Effekte 41 Prozent.

Tabelle 2: Relative Anteile primärer und sekundärer Effekte (gemittelt über multiple imputierte Datensätze)

| | Gesamteffekt | Anteil primärer Effekte [#] | Anteil sekundärer Effekte [#] |
|------------|----------------|--------------------------------------|--|
| Odds-Ratio | 5.78 (SE=1.21) | 41% | 59% |

[#]Durchschnitt aus Methoden a und b (vgl. z. B. Jackson et al. 2007), Berechnung sekundärer Anteile auf Basis ungerundeter Werte

¹⁵ Die Effekte lassen sich ebenfalls isolieren, wenn nicht die Gruppe P_{oo} als Vergleichsbasis herangezogen wird (Methode a), sondern die Gruppe P_{uu} (Methode b). Entsprechend gibt es bei der Quantifizierung der Effekte zwei Methoden, deren Ergebnisse minimal voneinander abweichen. Analog zu Jackson et al (2007) wird hier der Durchschnitt aus beiden Methoden berichtet.

6.5.3 Lehrer oder Eltern: Der Einfluss der Entscheidungshoheit auf die relative Bedeutung der sekundären Effekte

Wenn die Untersuchungsstichprobe geteilt wird in Schüler aus Ländern, in denen die Entscheidungshoheit bei den Eltern liegt (61% des Analysesamples) und in Schüler aus Ländern, in denen die Lehrer in letzter Instanz über die Art der weiterführenden Schule entscheiden (39% des Analysesamples), können diese beiden Ländergruppen im Hinblick auf die Bedeutung sekundärer Effekte beim Übergang auf das Gymnasium verglichen werden. Bei entsprechender Anwendung der Effektdekomposition findet sich für die Länder ohne verbindliche GSE ein Anteil von sekundären Effekten an der gesamten Bildungsungleichheit von 61 Prozent (Tabelle 3). Bei Kindern, deren Lehrer über die Erteilung einer weitgehend verbindlichen GSE einen stärkeren Einfluss auf die Wahl der weiterführenden Schule ausüben, fällt der Anteil der sekundären Effekte mit 54 Prozent etwas geringer aus.

Tabelle 3: Übergangsraten auf das Gymnasium und relative Anteile sekundärer Effekte nach Ländergruppen (gemittelt über multiple imputierte Datensätze)

| | unverbindliche GSE (10 Länder) | | verbindliche GSE (6 Länder) | |
|--|-----------------------------------|--------------------|--------------------------------|--------------------|
| | Übergangsraten (in %) | | Übergangsraten (in %) | |
| | tatsächlich | kontrafaktisch | tatsächlich | kontrafaktisch |
| obere Bildungsschicht | P _{oo} 76 | P _{ou} 52 | P _{oo} 66 | P _{ou} 44 |
| untere Bildungsschicht | P _{uu} 36 | P _{uo} 60 | P _{uu} 25 | P _{uo} 48 |
| Odds-Ratio | 5.61 (SE=1.55) | | 5.91 (SE=1.94) | |
| Anteil sekundärer Effekte [#] : | 61 % | | 54 % | |

[#] Durchschnitt aus Methoden a und b, Berechnung sekundärer Anteile auf Basis ungerundeter Werte

Dieses Ergebnis deutet auf ein theoretisch erwartetes interessantes Phänomen hin: Mehr elterliche Entscheidungsfreiheit führt anscheinend zu sozial selektiveren Bildungsentscheidungen. Oder umgekehrt: Der Entscheidungsprozess ist in Ländern mit verbindlicher GSE meritokratischer. Allerdings sei an dieser Stelle darauf hingewiesen, dass es bislang keine gesicherte Möglichkeit gibt, Aussagen über Konfidenzintervalle für die relativen Effektteile zu treffen.¹⁶ Insofern bleibt offen, ob sich die relative Bedeutung von sekundären Effekten zwischen den beiden Ländergruppen zufallskritisch unterscheidet.

¹⁶ Ein erster Versuch, Konfidenzintervalle über die Methode des Bootstrapping zu konstruieren, ist bislang nicht validiert worden.

Eine weitere Erkenntnis lässt sich aus Tabelle 3 ableiten: Die absolute Ungleichheit (ausgedrückt in der Odds Ratio) ist in Ländern mit verbindlicher Empfehlung nicht kleiner – der relative Erklärungsbeitrag von sekundären Effekten sagt also noch nichts über die absolute Ungleichheit aus.

6.6 Zusammenfassung und Diskussion

Die Abschätzung der relativen Bedeutung von primären und sekundären Effekten ist das Ziel einer wachsenden Anzahl internationaler Forschungsbeiträge. Das Ziel des vorliegenden Beitrags war *erstens* die Ergänzung dieses Forschungsstrangs durch die Quantifizierung der Effekte für den Übergang auf das Gymnasium in Deutschland. Mittels kontrafaktischer Dekompositionsanalysen konnte gezeigt werden, dass sekundäre Effekte rund 60 Prozent der Bildungsungleichheit an diesem Übergang verantworten. *Zweitens* konnte gezeigt werden, dass sekundäre Effekte vor allem bei „mittelmäßigen“ Schülern bedeutsam sind. Während die Bildungsentscheidung bei sehr guten oder sehr schlechten Schüler relativ unabhängig von ihrer sozialen Herkunft getroffen wird, schlagen direkte Herkunftseffekte im mittleren Notenbereich voll durch. *Drittens* wurde überprüft, inwiefern die relative Bedeutung primärer und sekundärer Effekte mit der institutionell variierenden Entscheidungshoheit – Eltern versus Lehrer – variiert. Diese institutionelle Variation wurde bislang empirisch kaum untersucht. Sollten Eltern frei entscheiden dürfen oder die Lehrer das „letzte Wort“ haben? Es wurde angenommen, dass in Ländern mit verbindlicher GSE der Spielraum für die Durchsetzung schichtverzerrter Elternpräferenzen eingeschränkt ist zugunsten einer durch die Lehrer induzierten, meritokratischeren Schulwahl. Die Ergebnisse deuten auf eine Bestätigung der Hypothese hin, allerdings ist die Reduktion sekundärer Effekte nur moderat und konnte in der vorliegenden Untersuchung nicht zufallskritisch abgesichert werden.

Lässt sich der Ländergruppenunterschied nun eindeutig auf die unterschiedliche Empfehlungsregelung zurückführen? Wie bei jeder komparativen Forschung ist die Isolierung einzelner institutioneller Effekte schwierig, weil andere kontextuelle Faktoren ebenfalls zwischen den Systemen variieren können. Faktoren, die den postulierten Ländergruppenunterschied wegerklären könnten, müssten aber *erstens* systematisch mit den Ländergruppen korrelieren und *zweitens* das relative Gewicht von primären und sekundären Effekten selektiv beeinflussen. Zu denken ist hier an den Entscheidungszeitpunkt (6-jährige Grundschulzeit), von dem angenommen werden darf, dass eine spätere Entscheidung sekundäre Disparitäten reduziert. Dieser Faktor wird jedoch indirekt mitkontrolliert, da Kinder, die eine

6-jährige Grundschule oder eine schulartunabhängige Orientierungsstufe besuchen, von der Analyse ausgeschlossen sind. Weiterhin könnte man argumentieren, dass die West-Länder mit hohem Migrantenanteil häufiger keine verbindliche GSE haben als die Ost-Länder. Migrationshintergrund ist mit niedriger sozialer Herkunft konfundiert, gleichzeitig treffen Migranten unter Kontrolle der Schulleistungen ambitioniertere Bildungsentscheidungen. Diese Konfundierungen könnten im Mittel zu geringeren sekundären Effekten in Ländern ohne verbindliche GSE führen. Insofern unterschätzt der Ländergruppenvergleich ohne Kontrolle des Migrationshintergrundes die wahre Differenz zwischen Ländern mit und ohne GSE. Die potenzielle Verzerrung wirkt beim Test der Hypothese also konservativ. Darüber hinaus gibt es möglicherweise weitere Faktoren, die zwischen den Ländergruppen variieren und die Schlussfolgerungen beeinträchtigen könnten, etwa lokale Arbeitsmarktbedingungen,¹⁷ das Verhältnis von angebotenen und nachgefragten Gymnasialplätzen oder die wahrgenommene Durchlässigkeit des Sekundarschulsystems (vgl. dazu Ditton 2004). Auch wenn auf den ersten Blick keine klaren theoretischen Argumente zu finden sind, in welche Richtung diese Faktoren den Ländergruppenunterschied im relativen Verhältnis von primären und sekundären Effekten beeinflussen sollten, wäre eine Kontrolle prinzipiell wünschenswert, bei der vorhandenen Datenlage in Deutschland derzeit aber kaum möglich. Grundsätzlich ist ein innerdeutscher Vergleich für das Abschätzen und Isolieren institutioneller Effekte aber vorteilhaft, da aufgrund der relativen Ähnlichkeit der Systeme Verzerrungen durch unkontrollierte Drittvariablen weniger wahrscheinlich sind als bei international vergleichenden Studien.

Die Analysen fördern einen weiteren interessanten Befund zu Tage: Die absolute Ungleichheit ist in Ländern mit verbindlicher GSE *nicht* geringer. Wie lässt sich dieser Befund erklären? Ich sehe mindestens zwei Ansatzpunkte für potenzielle Erklärungen: Die drohende verbindliche Nicht-Empfehlung für das Gymnasium führt möglicherweise dazu, dass bildungsnahe Eltern ungleich häufiger in die schulischen Leistungen ihrer Kinder investieren, um die Notengrenze für eine Gymnasialempfehlung zu unterschreiten. Dadurch würden gewissermaßen sekundäre Effekte in primäre Effekte „konvertiert“, das absolute Niveau bliebe aber gleich. Zudem üben bildungsnahe Eltern möglicherweise Druck auf Lehrer aus, wenn aufgrund der Schulleistungen eine Gymnasialempfehlung fraglich ist (Pohlmann 2008). Dieser Druck kann zum einen dazu führen, dass Lehrer eine drohende Auseinandersetzung im Voraus umgehen, indem sie bildungsnahen Kindern bessere Noten geben und in der Folge eine Gymnasialempfehlung aussprechen – auch hier wäre die Ungleichheit nur scheinbar über

¹⁷ In einer strukturschwachen Region mag ein niedriger Schulabschluss wertvoller sein für die Realisierung der Berufsvorstellungen als in einer strukturstarken Region.

Leistungsunterschiede vermittelt. Ebenso ist es vorstellbar, dass Lehrer ihren Spielraum nutzen und bei gleichen Noten die GSE als Reaktion auf den elterlichen Druck gleichsam schichtverzerrt aussprechen – in diesem Fall würde die Ungleichheit direkt über sekundäre Effekte vermittelt, die verglichen mit der unverbindlichen Ländergruppe kaum reduziert ausfielen, ebenso wenig wie das absolute Ungleichheitsniveau.

Die Erklärungen deuten auf eine grundlegende theoretische Schwierigkeit hin: Primäre und sekundäre Effekte operieren nicht unabhängig voneinander. Vorgelagerte „antizipatorische“ Entscheidungen (vgl. auch Jackson et al. 2007, S. 212) beeinflussen die gemessenen schulischen Leistungen, dadurch sind bei der hier getroffenen Unterscheidung entscheidungsbasierte, sekundäre Anteile zum Teil in der primären Komponente enthalten. Eine verbindliche GSE beeinflusst solche vorgelagerten Entscheidungen womöglich zusätzlich. Insofern stellen die im empirischen Teil der Arbeit berichteten Anteile sekundärer Effekte die „untere Grenze“ der tatsächlich operierenden sekundären Effekte dar. Teilweise könnte das Problem eingegrenzt werden, wenn primäre Effekte ergänzend über Schulleistungstests abgebildet würden. Aber selbst dann erfasst die Messung der primären Effekte alle familiären „Leistungsförderentscheidungen“ bis zum Zeitpunkt der weiterführenden Schulwahl mit. Eine konsequente Herauspatisierung der entscheidungsbasierten Komponenten aus dem primären Effekt wäre allenfalls möglich, wenn mit Längsschnittdaten alle bewussten Fördermaßnahmen von der Geburt eines Kindes bis zum Entscheidungszeitpunkt abgebildet und dadurch kontrolliert werden könnten. Es ist zu hoffen, dass in einigen Jahren mit den Daten des Nationalen Bildungspanels eine differenziertere Identifikation der Mechanismen möglich wird.

Zu guter Letzt: Was bedeuten die Ergebnisse dieser Studie für die bildungspolitische Praxis? Erstens ist deutlich geworden, dass die in der öffentlichen Diskussion häufig geforderten Maßnahmen zur Förderung von benachteiligten Gruppen durch Ganztagsbetreuung oder frühkindliche Förderung nur einen Teil der Ungleichheit (und zwar den allem Anschein nach kleineren Teil) reduzieren können. Solche Maßnahmen könnten den primären Effekt reduzieren, doch sekundäre Effekte würden dadurch nur indirekt tangiert (Müller-Benedict 2007). Vor dem Hintergrund der vorliegenden Ergebnisse wäre es möglicherweise bildungspolitisch „effizienter“, sekundäre Disparitäten zu reduzieren. Der im internationalen Vergleich hohe Anteil sekundärer Effekte ist ein weiterer Hinweis darauf, dass eine frühe Selektion soziale Ungleichheit fördert – will man dennoch an der Selektion festhalten, kann letztlich nicht mit Sicherheit beantwortet werden, ob man Eltern oder Lehrern die Entscheidungshoheit zusprechen sollte. Die Ergebnisse deuten auf geringere sekundäre Effekte hin, wenn Lehrer das „letzte Wort“ bei

der weiterführenden Schulwahl haben. Zunächst wäre es aber notwendig, den Entscheidungsprozess und das Zusammenspiel zwischen Eltern- und Lehrerpräferenzen für beide Ländergruppen explizit auszuformulieren und mit adäquaten Daten empirisch zu überprüfen.

Danksagung

Besonderer Dank für hilfreiche Kommentare und Diskussionen gilt Katherin Barg, Robert Erikson, Walter Müller, Volker Stocké, Steffen Schindler und Felix Weiss, sowie den Herausgebern und zwei anonymen Gutachtern der Zeitschrift für Soziologie.

6.7 Appendix

Tabelle A1: Deskriptive Angaben zu den empirischen Daten (gemittelt über multipel imputierte Datensätze)

| | % | n | Quelle |
|---|-----------|--------------------|---------------------------|
| Bildungsentscheidung | | | Welle 3 (Mütter-FB) |
| Gymnasium | 52.6 | 283 | |
| weniger als Gymnasium | 47.4 | 255 | |
| Höchster allg. Bildungsabschluss der Eltern | | | W 1 (Mütter-FB, Väter-FB) |
| (Fach)Hochschulreife | 51.4 | 277 | |
| weniger als (Fach)Hochschulreife | 48.6 | 261 | |
| Länder mit verbindlicher Grundschulempfehlung | 38.8 | 209 | |
| | \bar{x} | $\sigma_{\bar{x}}$ | |
| Durchschnittsnote aus Mathematik und Deutsch | 2.36 | .03 | W 2 (Kinder-FB) |

%: Anteilswerte; n: Zahl der gültigen Fälle; \bar{x} : arith. Mittel; $\sigma_{\bar{x}}$: Standardfehler

Abbildung A1: Lowess Regressionen für die Übergangsraten der oberen und unteren Bildungsschicht

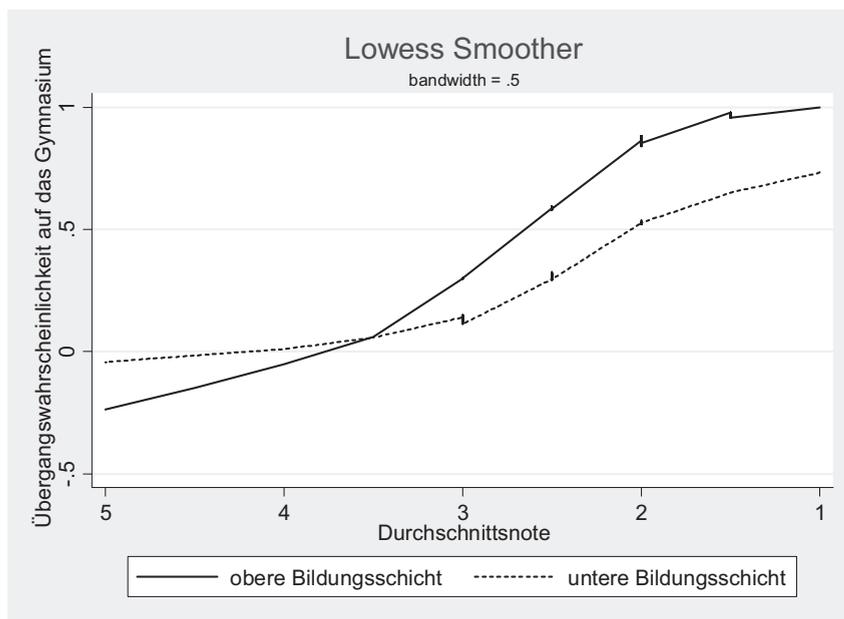


Tabelle A2: Übertrittsbestimmungen auf das Gymnasium nach Bundesländern, vereinfachte Darstellung

| Land | Kriterium für Gymnasialempfehlung; Bindekraft Empfehlung (++) - (0) | Entscheidungshoheit | Möglichkeiten, den Haupt-Entscheidungssträger zu beeinflussen |
|------|--|---------------------|--|
| | | | Eltern können Lehrerentscheidung beeinflussen: |
| BW | Ø Mathe u. Deutsch ≤ 2.5 ; bisheriges Lern- und Arbeitsverhalten (++) | Lehrer | Beratungsverfahren; Aufnahmeprüfung (muss mit ≤ 2.5 bestanden werden). |
| BY | Ø Mathe, Deutsch u. HSK ≤ 2.33 , wobei Mathe u. Deutsch ≤ 2.0 ; pädagogisches Wortgutachten (++) | Lehrer | Bei einer 3 in Mathe oder Deutsch ist Übertritt ins Gymnasium nach Beratungsgespräch möglich; Probeunterricht ^a . |
| SN | Ø Mathe u. Deutsch ≤ 2.5 (++) | Lehrer | Schriftliche Aufnahmeprüfung (muss mit ≤ 2.5 bestanden werden). |
| TH | Ø Mathe, Deutsch u. HSK ≤ 2.0 ; Kompetenzentwicklung (++) | Lehrer | Möglichkeit Antrag für die Aussprechung einer Gymnasialempfehlung zu stellen; 3-tägiger Probeunterricht. |
| BB | Ø Mathe, Deutsch u. erste Fremdspr. ≤ 2.33 ^b ; GSGutachten (++) | Lehrer | 2-tägige Eignungsprüfung. |
| SL | In Deutsch u. Mathe mind. eine 2 und eine 3. Bei abw. Noten hat Klassenkonf. Ermessensspielraum (++) | Lehrer | Schriftliche Prüfung in Deutsch u. Mathe. |
| | | | Lehrer können Elternentscheidung beeinflussen: |
| BE | Ø Mathe, Deutsch, erste Fremdspr. u. Naturwiss. ≤ 2.2 ; Einschätzung der Lernkompetenz (+) | Eltern | Sofern sich im ersten Halbjahr die Nichteignung herausstellt, muss Kind vom Gymnasium genommen werden. |
| SH | Empfehlung auf Basis der Lernentwicklung, aktuellen Leistungen (+) | Eltern | Abweichung nur nach vorherigem Beratungsgespräch möglich; Abweichung über zwei Schularten hinweg nicht möglich. |
| HE | Lernentwicklung, Leistungsstand, Arbeitshaltung (+) | Eltern | Sofern sich im ersten Halbjahr die Nichteignung herausstellt, muss Kind vom Gymnasium genommen werden (Querversetzung). |
| ST | Noten, Lernentwicklung und -verhalten (0) | Eltern | -- |
| HB | Lernergebnisse, Lernentwicklung, Persönlichkeitsentwicklung ^c (0) | Eltern | -- |
| HH | Lernstand und -entwicklung, Selbstständigkeit des Lernens (0) | Eltern | -- |
| MV | Schulleistungen, Arbeits- und Sozialverhalten (0) | Eltern | -- |
| NI | Arbeits- und Sozialverhalten, Leistungsstand, Lernentwicklung (0) | Eltern | -- |
| NW | Leistungsstand, Elterngespräch ^d (0) | Eltern | -- |
| RP | Leistungen, Lernverhalten und -entwicklung (0) | Eltern | -- |

Abkürzungen: Ø - Durchschnittsnote; HSK - Heimat- und Sachkunde.

^a Ca. 4% aller Schüler am Schulübertritt nehmen am Probeunterricht teil, davon bestehen rund 50% und können im Anschluss auf das Gym. wechseln (Amtstatistik des Bayrischen Kultusministeriums, Schuljahr 2004/05).

^b Maßgebend ist das Halbjahreszeugnis der 6. Klasse, da die Grundschule i.d.R. 6-jährig ist.

^c In seltenen Fällen gibt es sechsjährige Grundschulen (5 von insg. 84 Grundschulen in der Hansestadt Bremen), in denen die Empfehlung bindend ist.

^d Bei Neuregelungen 2006 wurde das freie Elternrecht eingeschränkt (betrifft die beobachtete Kohorte nicht).

6.8 Literatur

- Allison, Paul D. 2001. *Missing Data*. Thousand Oaks, CA: Sage.
- Alt, Christian, und Holger Quellenberg. 2005. Daten, Design und Konstrukte. Grundlagen des DJI-Kinderpanels. In *Kinderleben - Aufwachsen zwischen Familie, Freunden und Institutionen*, Hrsg. Christian Alt, 277-303. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Arnold, Karl-Heinz, Wilfried Bos, Peggy Richert und Tobias C. Stubbe. 2007. Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In *IGLU 2006 - Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Hrsg. Wilfried Bos, Sabine Hornberg, Karl-Heinz Arnold, Gabriele Faust, Lilian Fried, Eva Maria Lankes, Knut Schwippert und Renate Valtin, 271-298. Münster: Waxmann.
- Baumert, Jürgen, Rainer Watermann und Gundel Schümer. 2003. Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. *Zeitschrift für Erziehungswissenschaft* 6: 46-71.
- Becker, Rolf. 2000. Klassenlage und Bildungsentscheidungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 52: 450-474.
- Bos, Wilfried, Andreas Voss, Eva-Maria Lankes, Knut Schwippert, Oliver Thiel und Renate Valtin. 2004. Schullaufbahneempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe. In *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*, Hrsg. Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin und Gerd Walther, 191-228. Münster: Waxmann.
- Boudon, Raymond. 1974. *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*. New York: Wiley.
- Breen, Richard, und John H. Goldthorpe. 1997. Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society* 9: 275-305.
- Buis, Maarten L. 2010. Direct and Indirect Effects in a Logit Model. *Stata Journal* 10: 11-29.
- Contini, Dalit, Andrea Scagni und Annette Riehl. 2008. Primary and Secondary Effects in Educational Attainment in Italy. *Equalsoc Working Paper* 2008 / 4.
- Cortina, Kai S., und Luitgard Trommer. 2003. Bildungswege und Bildungsbiographien in der Sekundarstufe I. In *Das Bildungswesen in der Bundesrepublik Deutschland*, Hrsg. Kai S. Cortina, Jürgen Baumert, Achim Leschinsky, Karl Ulrich Mayer und Luitgard Trommer, 342-391. Reinbek: Rowohlt.
- Ditton, Hartmut. 2004. Der Beitrag von Schule und Lehrern zur Reproduktion von Bildungsungleichheit. In *Bildung als Privileg? Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit*, Hrsg. Rolf Becker, Wolfgang Lauterbach, 251- 279. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ditton, Hartmut, und Jan Krüskens. 2006. Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft* 9: 348-372.
- Ditton, Hartmut, Jan Krüskens und Magdalena Schauenberg. 2005. Bildungsungleichheit - der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft* 8: 285-304.

- Erikson, Robert, und Jan O. Jonsson. 1996. Explaining Class Inequality in Education: The Swedish Test Case. In *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, Hrsg. Robert Erikson, Jan O. Jonsson, 1-63. Boulder: Westview Press.
- Erikson, Robert, und Frida Rudolph. 2010. Change in Social Selection to Upper Secondary School - Primary and Secondary Effects in Sweden. *European Sociological Review* 26: 291-305.
- Erikson, Robert, John H. Goldthorpe, Michelle Jackson, Meir Yaish und David R. Cox. 2005. On Class Differentials in Educational Attainment. *Proceedings of the National Academy of Sciences* 102: 9730-9733.
- Esser, Hartmut. 1999. *Soziologie: Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt a. M.: Campus.
- Girard, Alain, und Henri Bastide. 1963. La stratification sociale et la démocratisation de l'enseignement. *Population* 18: 435-472.
- Goldthorpe, John H. 2007. *On Sociology. Volume 2*. 2nd Edition. Stanford: Stanford University Press.
- Infas, Institut für angewandte Sozialwissenschaft. 2006. Methodenbericht zur 3. Welle des DJI Kinderpanels "Wie wachsen Kinder auf?".
- Jackson, Michelle, Robert Erikson, John H. Goldthorpe und Meir Yaish. 2007. Primary and Secondary Effects in Class Differentials in Educational Attainment: The Transition to A-Level Courses in England and Wales. *Acta Sociologica* 50: 211-229.
- Kloosterman, Rianne, Stijn Ruiter, Paul M. de Graaf und Gerbert Kraaykamp. 2009. Parental Education, Children's Performance and the Transition to Higher Secondary Education: Trends in Primary and Secondary Effects Over Five Dutch School Cohorts (1965-99). *British Journal of Sociology* 60: 377-398.
- Lehmann, Rainer H., Rainer Peek und Rüdiger Gänsfuß. 1997. Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern, die im Schuljahr 1996/97 eine fünfte Klasse an Hamburger Schulen besuchten. Hamburg: Behörde für Schule, Jugend und Berufsbildung.
- Müller-Benedict, Volker. 2007. Wodurch kann die soziale Ungleichheit des Schulerfolgs am stärksten verringert werden? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59: 615-639.
- Pfeffer, Fabian T. 2008. Persistent Inequality in Educational Attainment and its Institutional Context. *European Sociological Review* 24: 543-565.
- Pohlmann, Sanna. 2008. Der Übergang von der Primar- in die Sekundarstufe aus der Sicht bayerischer und hessischer Lehrer. *Diskurs Kindheits- und Jugendforschung* 3: 123-140.
- Royston, Patrick. 2004. Multiple Imputation of missing Values. *Stata Journal* 4: 227-241.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Schindler, Steffen, und David Reimer. 2010. Primäre und sekundäre Effekte der sozialen Herkunft beim Übergang in die Hochschulbildung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62: 623-653.
- Stocké, Volker. 2007. Strength, Sources, and Temporal Development of Primary Effects of Families' Social Status on Secondary School Choice. *Working Paper Sonderforschungsbereich 504*, Universität Mannheim.

Von Hippel, Paul T. 2007. Regression with missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data. *Sociological Methodology* 37: 83-117.

Wiese, Wilhelm. 1982. Elternstatus, Lehrerempfehlung und Schullaufbahn: Eine empirische Analyse des Einflusses des Grundschullehrers auf die Bildungslaufbahn des Schülers. *Zeitschrift für Soziologie* 11: 49-63.

Kapitel 7 Ausblick

Lehrkräfte sind die zentralen Akteure im Bildungssystem. Ihr Handeln im Unterricht ebenso wie ihre Entscheidungen bei schulischen Angelegenheiten entscheiden mit über die Lernerfolge und Bildungskarrieren der Schülergenerationen. Vor diesem Hintergrund war das Anliegen dieser Arbeit, einige Aspekte zur Rolle von Lehrkräften auf Basis empirischer Studien zu analysieren, die immer wieder im Fokus wissenschaftlicher wie öffentlicher Debatten stehen.

Kapitel 2 und 3 beschäftigten sich mit Fragen der (Selbst)selektion in das Lehramt. Ausgehend von der Annahme, dass der Kompetenzerwerb zukünftiger Lehrer auch von den kognitiven und motivationalen Merkmalen der Aspiranten abhängt, untersuchten die Arbeiten verschiedene individuelle Voraussetzungen von Lehramtsanwärtern im Vergleich zu anderen Studierenden bzw. Absolventen. In Kapitel 2 lag der Schwerpunkt darauf, eine Vielzahl an leistungsbezogenen, motivationalen und soziodemographischen Merkmalen möglichst differenziert nach Art des Lehramtsstudiums anhand einer relativ aktuellen Kohorte zu analysieren. In Kapitel 3 lag der Fokus darauf, einige wenige Eingangsmerkmale verschiedener Lehramtsanwärterkohorten über einen Zeitraum von 30 Jahren zu vergleichen und in Bezug zu setzen zu gleichsam schwankenden Arbeitsmarktbedingungen, von denen angenommen werden kann, dass sie die Selbstselektion in die Profession beeinflussen. Die Analysen in den beiden Kapiteln liefern Hintergrundwissen zu der Frage, wie die Rekrutierung zukünftiger Lehramtsanwärter gestaltet werden könnte. Die Beiträge im zweiten Teil der Dissertation befassten sich mit der Rolle von Lehrkräften bei der Genese ungleicher Bildungschancen. In den Kapiteln 4 und 5 ging es um geschlechtliche Bildungsungleichheiten. Die Arbeiten untersuchten, inwiefern die Feminisierung des Lehrerberufs für die Bildungsmisserfolge der Jungen verglichen mit den Mädchen (mit)verantwortlich sein kann. Die Studien liefern Antworten auf die Frage, inwiefern die Rekrutierung von mehr männlichen Lehrkräften zu einem Abbau der schulischen Nachteile von Jungen beitragen kann. Kapitel 6 thematisierte soziale Bildungsungleichheiten an der Schnittstelle zur Sekundarstufe. In dem Artikel wurde unter anderem der Frage nachgegangen, ob eine von den Grundschullehrkräften ausgesprochene verbindliche Schullaufbahnpfehlung das Potenzial hat, soziale Ungleichheiten am Übergang zum Gymnasium zu reduzieren.

Die wesentlichen Befunde, Einschränkungen und Schlussfolgerungen wurden in den einzelnen Kapiteln benannt. Sie sollen hier nicht wiederholt werden. Stattdessen möchte ich zum Abschluss auf zukünftige Forschungsfragen aufmerksam machen.

7.1 Selektion in das Lehramt

Hinsichtlich der Rekrutierung in das Lehramt drängen sich vor allem Fragen nach der Vorhersagekraft von Eingangsmerkmalen auf. Die Wirkungskette (vgl. Kapitel 1, Abbildung 1) zwischen individuellen Voraussetzungen von Lehramtsanwärtern, der Qualität der Lehrerbildung, dem Kompetenzerwerb der Lehrer und schließlich den Lernerfolgen der Schüler ist lang und komplex (vgl. auch Terhart 2008). Die Identifizierung relevanter Eingangsmerkmale ist deshalb alles andere als trivial. Dadurch lässt sich erklären, warum die bisherige Kenntnislage zur prädiktiven Wirkung von Eingangsmerkmalen bescheiden ist. Das bedeutet jedoch nicht, dass solche Merkmale vernachlässigt werden sollten. Bestimmte kognitive und motivationale Voraussetzungen erleichtern den Erwerb komplexer professioneller Kompetenzen. Solche Kompetenzen wiederum sind Voraussetzungen für erfolgreiches Lehrerhandeln. Die beschriebene Wirkungskette muss Schritt für Schritt geprüft werden. Zukünftige Forschung sollte klären, inwiefern bestimmte Eingangsvoraussetzungen mit der Nutzung von Lerngelegenheiten zusammenhängen. In einem weiteren Schritt muss dann analysiert werden, inwiefern Eingangsvoraussetzungen und Nutzung von Lerngelegenheiten im Zusammenspiel den Erwerb der verschiedenen Kompetenzdimensionen beeinflussen. Bislang wurde die theoretische Annahme weitgehend vernachlässigt, dass Eingangsmerkmale und die Qualität der Lehrerbildung multiplikativ (und nicht additiv) miteinander verbunden sind. Das bedeutet: Aus den intelligentesten und motiviertesten Anwärtern werden kaum geeignete Lehrkräfte, wenn die Ausbildung nicht die entsprechenden Kompetenzen vermittelt. Und umgekehrt: Selbst die beste Lehrerbildung wird kaum fähige Lehrer ausbilden, wenn sich nur ungeeignete Anwärter in die Ausbildung selektieren. Die Herausforderung ist es, trotz methodischer Schwierigkeiten die relevanten Eingangsmerkmale auf Basis längsschnittlicher Studien zu identifizieren. Dabei kann an die bisherigen Befunde der Kompetenzforschung angeknüpft werden, die zeigen, dass kognitive aber auch motivationale Merkmale den Erwerb von Professionswissen fördern (vgl. Kapitel 1). Um die Kompetenzen zukünftiger Lehrergenerationen zu steigern, wäre es eine vertane Chance, die Bedeutung der Eingangsmerkmale zu ignorieren und ausschließlich auf die Ausbildungsqualität zu fokussieren (vgl. auch Mayr 2010; Spinath 2012).

Auch das letzte Glied in der Wirkungskette (Kompetenzen → Schülerleistungen) muss weiter untersucht werden. Die COACTIV-Studie (Kunter et al. 2011) ist durch ihre längsschnittliche Anlage und die direkte Testung von Lehrerkompetenzfacetten sowie ihren Auswirkungen auf Schülerleistungen ein vorbildliches Beispiel dafür, in welche Richtung weitere Forschungsbemühungen zielen sollten. Wichtig wäre es, COACTIV zu replizieren,

insbesondere im Hinblick auf andere Fächer als Mathematik (beispielsweise Deutsch, Sachkunde, Fremdsprachen) und mit Blick auf andere Schulstufen (insbesondere müsste der Primarbereich in den Blick genommen werden).

Eine notwendige, direkte Erweiterung der in Kapitel 2 präsentierten Studie wäre es, die Lehramtsanwärter differenziert nach Geschlecht und angestrebter Fächerkombination zu analysieren. Die in den Kapiteln 4 und 5 präsentierten Befunde deuten an, dass männliche Grundschullehrkräfte tendenziell weniger Lesekompetenzen vermitteln als ihre weiblichen Kollegen. Inwiefern sich ein solches Muster bereits in ungünstigeren Eingangsvoraussetzungen von männlichen Anwärtern für das Grundschullehramt niederschlägt, sollte untersucht werden. Bei der komplexen Studienentscheidung geht es für Studienberechtigte neben der Entscheidung für eine Hochschul- und Abschlussart vor allem auch um die Wahl eines inhaltlichen Studienfaches. Da das Fächerspektrum in Lehramtsstudiengängen eingeschränkt ist, sind Fachwahl und angestrebter Studienabschluss konfundiert. Damit ist nicht abschließend geklärt, ob die beschriebenen Differenzen zwischen Lehramtsstudierenden und anderen Studierenden auf unterschiedliche Studienfachprofile zwischen den Alternativen zurückzuführen sind, oder ob es sich tatsächlich um genuin charakteristische Merkmale von Lehramtsstudierenden handelt. Um diese Konfundierung aufzulösen, sollten den Lehramtsstudierenden in zukünftigen Analysen fachhomogene Vergleichsgruppen gegenübergestellt werden.

7.2 Lehrkräfte und Bildungsungleichheiten

Wie die Systematik in Kapitel 1 (Abschnitt 1.5) aufgezeigt hat, gleicht das Themenfeld ‚Lehrkräfte und Bildungsungleichheiten‘ einer Landkarte voller weißer Flecken. Anknüpfend an den Einfluss von Lehrkräften auf Leistungsdisparitäten (primäre Effekte) erscheinen folgende Forschungsfragen besonders dringlich:

- Inwiefern verfügen Lehrkräfte über ein unterschiedliches Ausmaß an Kompetenz zum Umgang mit Heterogenität? Inwiefern hängt diese Kompetenzfacette mit dem Ausmaß an sozialen, ethnischen und geschlechtlichen Ungleichheiten in einer Klasse zusammen? Bisherige Studien haben untersucht, inwiefern Lehrerkompetenz mit Lernzuwächsen in Verbindung steht. Eine wichtige Erweiterung wäre es, zu prüfen, inwiefern die Kompetenz im Umgang mit soziodemographischer und leistungsbezogener Heterogenität im Klassenzimmer mit der Leistungsstreuung innerhalb einer Klasse zusammenhängt. Das Thema

Leistungsheterogenität gewinnt derzeit zusätzlich an Aktualität angesichts der Herausforderungen von inklusiver Bildung (UNESCO 2013).

- Vertreter aus Politik, Praxis und Verwaltung fordern seit längerem die Erhöhung des Anteils von Lehrkräften und Pädagogen mit Migrationshintergrund; sowohl im Bundesweiten Integrationsprogramm (BAMF 2010) als auch im Nationalen Integrationsplan (Die Bundesregierung 2007) sind dazu konkrete Empfehlungen formuliert. Verbunden mit diesen Empfehlungen ist die Hoffnung, einerseits die Zusammenarbeit der Lehrkräfte mit zugewanderten Eltern zu verbessern; andererseits wird angenommen, dass solche Lehrkräfte gegenüber Kindern mit Migrationshintergrund als „wichtige Rollenvorbilder“ fungieren, was sich positiv auf deren schulische Leistungen auswirken sollte. Inwiefern solche Hoffnungen berechtigt sind, ist allerdings eine empirisch offene Frage. Inwiefern profitieren Kinder mit Migrationshintergrund von Lehrkräften mit Migrationshintergrund?
- Bislang werden die Schulkinder mit den größten Nachteilen von den am wenigsten kompetenten Lehrkräften unterrichtet (Baumert et al. 2010), was sicherlich zu den differenziellen und ungleichheitsfördernden „Lern- und Entwicklungsmilieus“ beiträgt (Baumert et al. 2006). Wie kann es gelingen, die kompetentesten Lehrkräfte auf den Schulformen unterrichten zu lassen, auf denen die benachteiligten Schüler überrepräsentiert sind? Eine Möglichkeit besteht sicherlich in der Angleichung oder Zusammenlegung der Ausbildung von Haupt-/Realschullehrern und Gymnasiallehrern. Ich sehe keinen unmittelbaren Grund, warum die Ausbildung auf den nicht-gymnasialen Schulstufen kürzer und akademisch weniger anspruchsvoll gestaltet werden soll. Es erscheint geboten, die Ausbildung für diese Lehrämter aufzuwerten, sowohl was die Länge als auch den akademischen Anspruch angeht. Ein zweiter Ansatzpunkt wäre die Schaffung einer Anreizstruktur, um besonders kompetente Lehrkräfte für nicht gymnasiale Lehrerpositionen zu gewinnen. Zu denken ist hier beispielsweise an finanzielle Anreize, etwa Zulagen für die Anstellung an Problemschulen, oder nicht monetäre Anreize, wie mehr Autonomie bei der Gestaltung des Unterrichts oder einem erleichterten Übergang in ein Sabbatjahr.

Die Einflussnahme von Lehrkräften auf Bildungsentscheidungen (sekundäre Effekte) ist in vielerlei Hinsicht ungeklärt. Forschungsbedarf besteht insbesondere bei den folgenden Themen:

- Das Gewicht der Lehrerentscheidung bei anderen schulischen Entscheidungen als der Grundschulempfehlung ist bislang kaum systematisch untersucht worden. Zukünftige Forschungsvorhaben sollten klären, inwiefern soziodemographische Merkmale der Schüler unabhängig von leistungsbezogenen Eigenschaften eine Rolle spielen bei Entscheidungen

zu Klassenwiederholungen, Querversetzungen, Rückstellungen und vorzeitigen Einschulungen.

- Baden-Württemberg hat seit 2012 die verbindliche Grundschulempfehlung abgeschafft. Kurz vor der Fertigstellung dieser Dissertation zeigen die neuesten Zahlen des Kultusministeriums Baden-Württemberg, dass die Klassenwiederholungen in den 5. Klassenstufen zugenommen haben (Südkurier 2013; SWR 2013). Rund 1.300 (3,9 Prozent) der 34.000 Realschüler müssen 2013 die fünfte Klasse wiederholen. Das sind fünf Mal mehr als im vergangenen Schuljahr. Von den 38.000 Fünftklässlern im Gymnasium müssen rund 500 (1,3 Prozent) die Klasse wiederholen. Im Schuljahr 2011/2012, als die Grundschulempfehlung noch verbindlich war, betrug die Wiederholer-Quote an Gymnasien 0,5 Prozent. Inwiefern diese gestiegenen Quoten mit der Abschaffung der Verbindlichkeit in Zusammenhang steht, sollte analysiert werden. Seit dem Wegfall der verbindlichen Grundschulempfehlung dürfen Eltern ihre Kinder an der Schule ihrer Wahl anmelden. Bei der Realschule hatten 23,3 Prozent der Schüler eine Empfehlung für Haupt- und Werkrealschule, bei den Gymnasien hatten 10,2 Prozent eine Empfehlung für die Realschule und 0,8 Prozent eine Empfehlung für die Haupt- und Werkrealschule. Angesichts der vielen Familien, die entgegen einer anderslautenden Empfehlung den Weg auf eine ambitioniertere Schulform gewählt haben, bleibt zu prüfen, ob vor allem Kinder mit anderslautender Empfehlung an der Versetzungshürde gescheitert sind. Können Lehrkräfte das schulische Potenzial besser einschätzen als die Eltern? Sollte diese Vermutung zutreffen, müsste auch untersucht werden, inwiefern Kinder unter ambitionierten Bildungsentscheidungen ihrer Eltern leiden, wenn sie mit dem Erleben schulischen Scheiterns konfrontiert sind.

7.3 Literatur

- BAMF. 2010. *Bundesweites Integrationsprogramm. Angebote der Integrationsförderung in Deutschland – Empfehlungen zu ihrer Weiterentwicklung*. Nürnberg: Bundesamt für Migration und Flüchtlinge.
- Baumert, Jürgen, Petra Stanat und Rainer Watermann. 2006. Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000*, Hrsg. Jürgen Baumert, Petra Stanat und Rainer Watermann, 95-188. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baumert, Jürgen, Mareike Kunter, Werner Blum, Martin Brunner, Thamar Voss, Alexander Jordan, Uta Klusmann, Stefan Krauss, Michael Neubrand und Yi-Miau Tsai. 2010. Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal* 47: 133-180.
- Die Bundesregierung. 2007. *Der Nationale Integrationsplan. Neue Wege – Neue Chancen*. Berlin: Presse- und Informationsamt der Bundesregierung.
- Kunter, Mareike, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand (Hrsg.). 2011. *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Mayr, Johannes. 2010. Selektieren und/oder qualifizieren? Empirische Befunde zu guten Lehrpersonen. In *Wirkt Lehrerbildung? Antworten aus der empirischen Forschung*, Hrsg. Jürgen Abel, Gabriele Faust, 73-89. Münster: Waxmann.
- Spinath, Birgit. 2012. Beiträge der Pädagogischen Psychologie zur Professionalisierung von Lehrerinnen und Lehrern: Diskussion zum Themenschwerpunkt. *Zeitschrift für Pädagogische Psychologie* 26: 307-312.
- Südkurier. 2013. Mehr Wiederholer in fünften Klassen. <http://www.suedkurier.de/nachrichten/baden-wuerttemberg/themensk/Mehr-Wiederholer-in-fuenften-Klassen;art417921,6184339> (Zugegriffen: 26.07.2013).
- SWR. 2013. Sitzenbleiber in Baden-Württemberg: Immer mehr Schüler machen Ehrenrunde. <http://www.swr.de/landesschau-aktuell/bw/-/id=1622/nid=1622/did=11795884/p3sw79/> (Zugegriffen: 26.07.2013).
- Terhart, Ewald. 2008. Die Lehrerbildung. In *Das Bildungswesen in der Bundesrepublik Deutschland*, Hrsg. Kai S. Cortina, Jürgen Baumert, Achim Leschinsky, Karl Ullrich Mayer und Luitgard Trommer, 745-772. Reinbek bei Hamburg: Rowohlt.
- UNESCO. 2013. Inklusive Bildung. http://www.unesco.de/inklusive_bildung.html (Zugegriffen: 26.07.2013)

