

Wayne, Andrew J.; Youngs, Peter

## **Die Art der Ausbildung von Lehrern und die Lerngewinne ihrer Schüler. Eine Übersicht über aktuelle empirische Forschung**

*Allemann-Ghionda, Cristina [Hrsg.]; Terhart, Ewald [Hrsg.]: Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern. Weinheim u.a. : Beltz 2006, S. 71-96. - (Zeitschrift für Pädagogik, Beiheft; 51)*



Quellenangabe/ Reference:

Wayne, Andrew J.; Youngs, Peter: Die Art der Ausbildung von Lehrern und die Lerngewinne ihrer Schüler. Eine Übersicht über aktuelle empirische Forschung - In: Allemann-Ghionda, Cristina [Hrsg.]; Terhart, Ewald [Hrsg.]: Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern. Weinheim u.a. : Beltz 2006, S. 71-96 - URN: urn:nbn:de:0111-opus-73718 - DOI: 10.25656/01:7371

<https://nbn-resolving.org/urn:nbn:de:0111-opus-73718>

<https://doi.org/10.25656/01:7371>

in Kooperation mit / in cooperation with:

# **BELTZ** **JUVENTA**

<http://www.juventa.de>

### **Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### **Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Zeitschrift für Pädagogik · 51. Beiheft

# **Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf**

Herausgegeben von  
Cristina Allemann-Ghionda und Ewald Terhart

Beltz Verlag · Weinheim und Basel

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2006 Beltz Verlag · Weinheim und Basel  
Herstellung: Klaus Kaltenberg  
Gesamtherstellung: Druckhaus »Thomas Müntzer«, Bad Langensalza  
Printed in Germany

ISSN 0514-2717  
Bestell-Nr. 41152

# Inhaltsverzeichnis

*Cristina Allemann-Ghionda/Ewald Terhart*

Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf. Zur Einleitung in das Beiheft .....	7
--	---

## 1. Grundlagen

*Renate Girmes*

Lehrprofessionalität in einer demokratischen Gesellschaft. Über Kompetenzen und Standards in einer erziehungswissenschaftlich fundierten Lehrerbildung .....	14
---	----

*Andreas Frey*

Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften – eine erste Standortbestimmung zu bereits publizierten Instrumenten .....	30
--	----

*Frank Lipowsky*

Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler .....	47
--	----

*Andrew J. Wayne/Peter Youngs*

Die Art der Ausbildung von Lehrern und die Lerngewinne ihrer Schüler. Eine Übersicht über aktuelle empirische Forschung .....	71
--	----

*Ken Zeichner*

Konzepte von Lehrerexpertise und Lehrerausbildung in den Vereinigten Staaten ....	97
---	----

## 2. Forschungsbeiträge

### 2.1 Universitäre Phase

*Manfred Lüders/Sabine Eisenacher/Steffen Pleßmann*

Der Umgang mit Studienzeit. Eine empirische Untersuchung bei Studierenden der Lehramter und im Diplom-Studiengang Erziehungswissenschaft .....	116
---	-----

*Tina Hascher*

Veränderungen im Praktikum – Veränderungen durch das Praktikum:  
Eine empirische Untersuchung zur Wirkung von schulpraktischen Studien  
in der Lehrerbildung ..... 130

*Johannes Mayr*

Theorie + Übung + Praxis = Kompetenz? Empirisch begründete Rückfragen  
zu den „Standards in der Lehrerbildung“ ..... 149

*Rainer Lersch*

Lehrerbildung im Urteil der Auszubildenden. Eine empirische Studie zu beiden  
Phasen der Lehrerausbildung ..... 164

## 2.2 Vorbereitungsdienst

*Christian Reintjes*

Wie beurteilen die Ausbilder der zweiten Phase die pädagogischen  
Kenntnisse der Absolventen der ersten Phase? Eine empirische Studie mit  
Hauptseminarleitern in Nordrhein-Westfalen ..... 182

*Peter Menck/Michaela Schulte*

Lehrer werden: erste Ergebnisse einer Untersuchung zum Referendariat ..... 199

*Hermann Josef Abs*

Zur Bildung diagnostischer Kompetenz in der zweiten Phase der Lehrerbildung ..... 217

## 2.3 Beruf

*Doris Edelmann*

Pädagogische Professionalität im transnationalen sozialen Raum.  
Eine Studie über Sichtweisen und Erfahrungen von Primarlehrpersonen in  
Bezug auf die kulturelle Heterogenität ihrer Schulklasse ..... 235

*Cristina Allemann-Ghionda/Georg Auernheimer/Helga Grabbe/Angelika Krämer*

Beobachtung und Beurteilung in soziokulturell und sprachlich  
heterogenen Klassen: die Kompetenzen der Lehrpersonen ..... 250

*Wolfgang Fichten/Hilbert Meyer*

Kompetenzentwicklung durch Lehrerforschung – Möglichkeiten und Grenzen ..... 267

Andrew J. Wayne/Peter Youngs

## Die Art der Ausbildung von Lehrern und die Lerngewinne ihrer Schüler

*Eine Übersicht über aktuelle empirische Forschung<sup>1</sup>*

Auf ihrer Suche nach Möglichkeiten zur Verbesserung von Schule und Unterricht hat sich sowohl die Bildungsforschung als auch die Bildungsadministration früher wie heute auf den Lehrer konzentriert. Lehrer sind eine der zentralen Ressourcen des Schulsystems. Ihre Gehälter machen den allergrößten Anteil an den Kosten des Schulsystems aus. Und Intuition wie auch empirische Forschung legen nahe, dass die Schulleistung von Schülern substanziell von den Lehrern abhängt, von denen sie unterrichtet werden.<sup>2</sup> Vor kurzem hat die Frage nach der notwendigen Zahl von Lehrkräften eine sehr breite Beachtung in der Öffentlichkeit gefunden, und zwar im Zusammenhang mit Sorgen hinsichtlich der Überalterung der Lehrerschaft und der Nachfrage nach neuen Lehrern. Seit dem epoche machenden Bericht der *National Commission for Teaching and America's Future* (1996) haben zahlreiche Forscher, Bildungsadministratoren und Kommissionen diese Frage sehr gründlich behandelt. Erst vor kurzem hat eine Expertengruppe des *National Research Council* einen Bericht über die Frage der Qualität von Lehrern und der Beurteilung ihrer Qualität vorgelegt (Mitchell u.a. 2001), und der Bildungsminister der USA, Rod Paige, hat einen sehr dezidierten Bericht über die Qualität der Lehrerschaft vorgelegt (U.S. Department of Education 2002).

Auf der Suche nach Ergebnissen, die für die Bildungsadministration von Bedeutung sein könnten, sind von Seiten der Bildungsforscher empirische Untersuchungen durchgeführt und zusammenfassende Forschungsübersichten verfasst worden, die sich mit verschiedenen Aspekten der Lehrerproblematik befassen. Manche Studien konzentrieren sich allerdings nur auf die Frage der *Quantität*, indem etwa errechnet wird, wie viele Lehrer Jahr für Jahr den Beruf verlassen und wie viele Lehrer in Zukunft benötigt wer-

1 Orig.: Teacher Characteristics and Student Achievement Gains: A Review. In: Review of Educational Research 73 (2003) 1, 89-122; Übersetzung (gekürzt) von E. Terhart. Übersetzung und Abdruck mit Erlaubnis der AERA.

Einige der hier präsentierten Ergebnisse sind Bestandteil des Einleitungskapitels der Dissertation von Wayne (2000). Obwohl wir (die Autoren) die volle Verantwortung für mögliche Irrtümer in diesem Artikel übernehmen, möchten wir uns bei Dale Ballou, Linda Darling-Hammond, William Galston, Daniel Goldhaber, Willis Hawley, Mark Lopez und zwei anonymen Gutachtern für ihre hilfreichen Kommentare zu früheren Fassungen des Aufsatzes bedanken.

2 Die empirische Forschung, die diese Annahme stützt, hat sich mit der Leistungsentwicklung von Schülern ohne Berücksichtigung spezieller Eigenschaften von Lehrern beschäftigt, sondern nach der Variation der Leistungszuwächse von Schülern in Abhängigkeit von verschiedenen Lehrern gefragt (vgl. Goldhaber/Brewer 1997b; Jordan/Mendro/Weerasinghe 1997; Hanushek 1971, 1992; Murnane 1975; Murnane/Phillips 1981; Rivkin, Hanushek/Kain 2001; Sanders/Rivers 1996 und Wright/Horn/Sanders 1997).

den (z.B. Hussar 1999). Die folgende Forschungsübersicht zielt demgegenüber auf die *Qualität* von Lehrern. Es liegt eine ganze Reihe von empirischen Studien vor, in denen die Charakteristika effektiver Lehrer untersucht worden sind. Bei ihren Versuchen, sich auf diese Studien zu stützen, sind Forscher und Analytiker jedoch zu sehr unterschiedlichen Interpretationen gekommen – vielleicht auch aufgrund der Schwierigkeiten der systematischen Auswertung von thematisch einschlägigen Forschungsberichten.<sup>3</sup> Wir werden im Folgenden eine systematische Methodik anzuwenden, um die Ergebnisse der Studien zu den Eigenschaften von Lehrern sowohl der Bildungsforschung als auch der Bildungsadministration zugänglich zu machen.

Forschungsergebnisse zur Relation zwischen bestimmten Eigenschaften von Lehrern (*teacher characteristics*) und den Lernleistungen von Schülern sind für die Diskussion um den Lehrerberuf und seine Gestaltung von großer einschlägiger Bedeutung. Die einzelnen Staaten der USA benennen sehr genaue Voraussetzungen für die Ausübung des Lehrerberufs in Gestalt von Abschlüssen, Ausbildungsordnungen und notwendigen Examensergebnissen. Andere Bereiche, für die Erkenntnisse über den Zusammenhang von Lehrereigenschaften und Schülerleistungen von Bedeutung sind, sind etwa die Identifikation von Entwicklungen innerhalb der Qualität der Lehrerschaft oder die der Benennung von Problemen, die vordringlich gelöst werden müssen. So kann es z.B. sein, dass Schüler aus niedrigen Einkommensschichten einen geringeren Anteil von Lehrern mit bestimmten Eigenschaften haben als Schüler aus Bevölkerungsschichten mit höheren Einkommen (vgl. z.B. Ingersoll 1996; Wayne 2002).

Nun ist es in diesem Kontext sehr wichtig zu berücksichtigen, dass in den USA die Konzepte des „effektiven Unterrichts“ sich im Laufe der Zeit sehr deutlich gewandelt haben. Für den allergrößten Teil des 20. Jahrhunderts galt die Grundregel: Diejenigen Personen, die ein entsprechendes staatlich definiertes Ausbildungsprogramm erfolgreich abgeschlossen haben, sind geeignet für den Lehrerberuf. Durch Ausbildungs-, Prüfungs- und Studienordnungen für die Fachstudien, für die pädagogischen Studien und für Schulpraktika haben die staatlichen Stellen ihren Einfluss auf die Lehrerbildung ausgeübt. Es wurde jedoch staatlicherseits nicht in Richtung auf spezifische, während der Ausbildung zu vermittelnden didaktischen Handlungsformen (*instructional practice*) hin gearbeitet. Dies änderte sich in den 1980er-Jahren, als mehrere Staaten damit begannen, bei den neu eingestellten Lehrkräften in der Berufseingangsphase leistungsbezogene *assessments* einzuführen. Viele dieser Tests zur Feststellung der beruflichen Kompetenz gründeten auf dem Prozess-Produkt-Paradigma innerhalb der empirischen Unterrichtsforschung und konzentrierten sich auf ein uniformes Muster des Lehrerverhaltens unabhängig von dem jeweiligen Fach, der Jahrgangsstufe oder dem Leistungsniveau.

Im Verlauf der letzten 10 Jahre haben viele Staaten die Standards vom *Interstate New Teacher Assessment and Support Consortium* (INTASC) und dem *National Council for*

3 In mehreren neueren Diskussionsbeiträgen werden die empirischen Evidenzen unterschiedlich interpretiert (vgl. (Ballou/Podgursky 1999, 2000; Darling-Hammond 2000, 2001; Darling-Hammond/Berry/Thorensen 2001; Goldhaber/Brewer 2001; Walsh 2001).

*the Accreditation of Teacher Education* (NCATE) bei ihren Entscheidungen über die Lizenzierung und Akkreditierung übernommen. Diese Standards basieren auf Unterrichtskonzeptionen, welche sich signifikant von denjenigen didaktischen Vorstellungen unterscheiden, die bislang Bestandteil der staatlichen Lehrerbildungspolitik waren. Die neuen Konzeptionen betonen insbesondere den kontextspezifischen Charakter des Unterrichts sowie die Notwendigkeit, dass Lehrer ein integriertes Wissen über Fachinhalte, Schülerlernen und die Kontexte ihres Unterrichts haben sowie dazu in der Lage sein müssen, Schüler zu aktivem Lernen und zur Reflektion auf ihr Handeln zu veranlassen (Youngs/Odden/Porter 2003).

Unsere Übersicht über die einschlägige empirische Forschung ist folgendermaßen strukturiert: Als erstes beschreiben wir unsere Methode der Erstellung der Übersicht, d.h. wir stellen dar, aufgrund welcher Kriterien Einzeluntersuchungen in unsere Übersicht aufgenommen werden, wie wir die Einzelstudien interpretieren, und wie wir sie schließlich synthetisieren (1). Im zweiten Teil wenden wir diese Methode an, um Forschungsergebnisse zu der Frage darzustellen, ob Lehrer mit bestimmten Eigenschaften einen größeren Einfluss auf die Lernergebnisse (genauer: die Lernzuwächse) ihrer Schüler (gemessen mit standardisierten Tests) haben als solche Lehrer, die diese Charakteristika nicht aufweisen. Es geht dabei um vier Bereiche: die Einschätzung der Qualität ihrer Ausbildungsstätte, ihre Examensergebnisse, ihre inhaltlichen Studienerfahrungen und die Art ihrer formalen Qualifikation oder Zulassung zum Lehrerberuf (2). Im abschließenden Teil werden Konsequenzen für Bildungsforschung und Bildungsverwaltung erörtert (3).

## 1. Methode

### 1.1 Der Rahmen der folgenden Forschungsübersicht

Nur 21 Studien erfüllten die Kriterien, die zur Aufnahme in den ausgewerteten Pool berechtigten. Am Beginn der Suche nach einschlägigen Studien haben wir Datenbanken und bereits vorliegende Forschungsübersichten ausgewertet. Dabei stützten wir uns auf die folgenden Metaanalysen und Übersichten: Darling-Hammond (1999a), Greenwald/Hedges/Laine (1996), Hanushek (1997) und Wilson/Floden/Ferrini-Mundy (2001). Zusätzlich wurden elektronische Datenspeicher für die Jahre 1975 – 2002 befragt: ERIC, PsychLit, und EconLit. Um in unsere Auswertung einbezogen werden zu können, mussten Studien den folgenden vier Kriterien genügen:

*Erstens* sollte bereits deutlich geworden sein, dass es uns um den Zusammenhang von Lehrermerkmalen und den durch standardisierte Tests gemessenen Lernleistungen bzw. Lernzuwächsen von Schülern dieser Lehrer geht. Sicher gibt es auch wichtige und untersuchungswerte Beziehungen zwischen Lehrereigenschaften und anderen Wirkungen aufseiten der Schüler, wie z.B. Abschlussquoten, Übergänge zu postsekundären Bildungswegen, Erwerb von Wissen und Fähigkeiten, die nicht ohne weiteres durch standardisierte Tests gemessen werden können. Gleichwohl: Zum Zeitpunkt der Zusammenstellung unserer Datenbasis lag keine Studie vor, die sich auf solche breiteren, nicht

durch standardisierte Tests gemessene Wirkungen aufseiten der Schüler beziehen und zugleich die weiteren drei Kriterien erfüllen. *Zweitens* haben wir uns auf Studien beschränkt, die mit Schülern innerhalb der USA durchgeführt wurden, da wir Informationen für die US-amerikanische Situation bereitstellen wollen. Unseres Wissens wurde aufgrund dieses Kriteriums nur eine Studie ausgeschlossen (Hanushek/Gomes-Neto/Harbinson 1996). Die folgenden beiden Kriterien zielen darauf ab, solche Ergebnisse in den Mittelpunkt zu stellen, die wirklich zwingend und nicht nur irgendwie überzeugend sind. So haben wir – *drittens* – nur Studien einbezogen, die nicht nur eine Wirkungserfassung bei den Schülern durchführten, sondern auch deren vorhergehenden Leistungsstand erfassten. Dieses Kriterium hat zum Ausschluss zahlreicher Studien geführt. Schließlich wurden – *viertens* – nur Studien aufgenommen, in denen auch der sozioökonomische Status der Schüler berücksichtigt wurde. Dieses Kriterium hat zum Ausschluss nur sehr weniger Studien geführt.<sup>4</sup>

Sowohl den vorherigen Leistungsstand als auch den sozioökonomischen Status zu erfassen gibt den Ergebnissen einer Studie einen höheren Überzeugungsgrad, denn die Frage „Lernen Schüler mehr bei Lehrern mit dieser (oder jener) Eigenschaft?“ richtet sich immer auch auf ein kausales Verhältnis. Es gibt viele Studien, die die Leistungen von Schülern am Ende eines Schuljahres in ein Verhältnis setzen zu den Qualifikationen der Lehrer dieser Schüler. Um aber die Leistungsunterschiede von Schülern auf die Eigenschaften ihrer Lehrer zurückführen zu können, muss man alle anderen Erklärungsfaktoren neutralisieren. Randomisierte Experimente tragen zum Ausschluss anderer Erklärungsfaktoren bei, aber solche Experimente sind im Rahmen der Forschung zur Effektivität von Lehrern kaum angewandt worden. Nicht-randomisierte (quasi-)experimentelle Designs haben sich jedoch als hinreichend aussagekräftig erwiesen. Um unter nicht-randomisierten Bedingungen andere intervenierende Faktoren zu kontrollieren, liegt diesen Designs eine bestimmte Theorie der Determinanten von Schulleistung zugrunde. Diese Konzepte werden ständig weiterentwickelt, aber in der einschlägigen Forschung ist man davon überzeugt, dass sowohl vorausgehende Leistungsstände als auch der sozioökonomische Status statistisch kontrolliert werden müssen, um näher an den interessierenden Zusammenhang – die Effekte von Lehrern auf die Lernleistung, genauer: die Lernzuwächse ihrer Schüler – heranzukommen (vgl. allgemein Ehrenberg u.a. 2001; Ferguson 1996; Rowan 2002). Solche Modelle der Determinanten von Schulleistung werden „*value added*“ genannt, weil sie von der Annahme ausgehen, dass die Lehrer zum Lernfortschritt in der Periode zwischen Pre- und Posttest beitragen. – Die Kriterien, die bei der Einbeziehung von thematisch einschlägigen Studien berücksichtigt wurden, lauten also zusammengefasst:

1. Die Studie muss auf Daten basieren, die sich sowohl auf Eigenschaften von Lehrern wie auf die durch standardisierte Tests gemessene Leistung ihrer Schüler beziehen.
2. Die Daten sind in den USA gesammelt worden.

4 Leser, die mehr über die hier ausgeschlossenen Studien wissen möchten, können sich in den Anhängen von Greenwald/Hedges/Laine (1996) und Mitchell u.a. (2001) informieren.

3. Im Forschungsdesign wurden auch die Ausgangsleistungen der Schüler erfasst.
4. Im Design der Studie wurde der sozioökonomische Status der Schüler berücksichtigt.

## 1.2 Interpretation der Einzelstudien

*Anm. des Übersetzers: Im Folgenden geben die Autoren ausführliche Hinweise über die Art der Interpretation von Einzelstudien sowie zur Deutung ihrer Resultate. Es handelt sich um eine relativ technisch gehaltene Argumentation, die insgesamt den skrupulösen Umgang der Autoren mit den statistischen Grundlagen der Einzelstudien, auf die sich stützen, dokumentiert. Aufgrund der komplexen Verhältnisse zwischen Lehrermerkmalen und Schülerleistungen ist die Tendenz erkennbar, sich tatsächlich nur auf Verlässliches zu stützen und Deutungen und Schlussfolgerungen mit allergrößter Vorsicht vorzunehmen. Die Autoren fassen diese Überlegungen folgendermaßen zusammen:*

Man kann also sagen, dass solche quasi-experimentellen Studien zum Zusammenhang von Lehrereigenschaften und Schülerleistungen zwar unsere Aufmerksamkeit verdienen, ihre Interpretation aber gleichwohl mit großer Vorsicht erfolgen sollte. Erstens haben Forscher nie untersucht, ob ein bestimmtes Lehrermerkmal *keine* Auswirkung auf die Schülerleistung hat. Dementsprechend interpretieren wir unbestimmte bzw. inkonsistente Ergebnisse (*indeterminate findings*) genau so – als unbestimmt. Zweitens haben wir Studien, in denen solche Beziehungen entdeckt wurden, aufgrund bestimmter Merkmale ihres Designs gewichtet: Studien, die einen hohen Schutzwall gegenüber Schein-Resultaten errichtet hatten, wurden stärker gewichtet. Drittens schließlich muss man davon ausgehen, dass die in den Studien berichteten Effektstärken alle mit einem *bias* belastet sind, sodass wir nicht wissen können, ob der tatsächliche Effekt stärker oder schwächer ist. Dementsprechend werden wir die erhaltenen Effektstärken nicht diskutieren. [...]

## 2. Ergebnisse

Im Folgenden sollen die Ergebnisse unserer Auswertung präsentiert werden. Für vier Charakteristika von Lehrern werden wir a) alle relevanten Studien und Ergebnisse beschreiben, b) zusammenfassende Interpretationen vorstellen und c) Implikationen für Bildungsverwaltung und Bildungsforschung erörtern. Die im Einzelnen betrachteten Charakteristika von Lehrern sind

1. die Beurteilung der Qualität ihrer Ausbildungsstätte (teachers' college ratings)
2. ihre Testergebnisse bei Berufseingangstests (test scores)
3. ihr Ausbildungscurriculum und der erworbene Grad (course taking and degrees)
4. ihre formelle Zulassung als Lehrer (certification status).

Ergänzt wird dies durch eine offene Restklasse von Eigenschaften, die keiner der vier genannten Eigenschaften zuzurechnen sind.

## 2.1 Die Qualität der Ausbildungsinstitution (undergraduate level)

### 2.1.1 Relevante Studien und Ergebnisse

Es gibt nur drei Untersuchungen, die der Frage nachgegangen sind, ob Schüler mehr von denjenigen Lehrern lernen, deren Ausbildungsstätte allgemein als von hoher Qualität beurteilt wird, als von denjenigen, deren Ausbildungsstätte schlechter bewertet wird. Die erste Untersuchung (Summers/Wolfe 1975a, 1977) basiert auf Studien in Philadelphia im Schuljahr 1970/71 in den Jahrgangsstufen 6, 8 und 12. Die Akten in den Schulen gaben Auskunft über die Leistungen in diesem speziellen Schuljahr sowie in früheren Schuljahren, sodass die Leistungsentwicklung jedes Schülers über eine längere Zeit verfolgt werden konnte. Die Analysen umfassten die Lernentwicklung von der 3. zur 6. Jahrgangsstufe, von der 6. zur 8. und von der 9. zur 12. Jahrgangsstufe. Während die meisten Studien bis dahin bei den schulischen Voraussetzungen den Durchschnitt auf Schul-Ebene zugrunde gelegt hatten, verwendeten Summers/Wolfe Daten auf der Ebene des einzelnen Schülers. Die Unterlagen in den Schulen enthielten Informationen über die gegenwärtigen und früheren Lehrer der Schüler, ebenso die Testergebnisse dieser Lehrer bei der *National Teacher Examination* (NTE), weiterhin Informationen zu den Dienstjahren der Lehrer und zur Beurteilung der Qualität der Ausbildungsinstitution dieser Lehrer entsprechend dem Gourman-Bericht.<sup>5</sup>

Die Analysen der 6. Jahrgangsstufe (627 Schüler) basierten auf Schülerseite auf den zusammengefassten Ergebnissen des *Iowa Test of Basic Skills* (ITBS). Es zeigte sich, dass die Qualität der Ausbildungsinstitution der in dieser Jahrgangsstufe arbeitenden Lehrer eine Bedeutung für den Lernzuwachs der Schüler hatte. Als die Forscher jedoch versuchten, die Gourman-Einteilung und alle anderen schulischen Bedingungen auf Schul-Ebene zu aggregieren und Durchschnitte zu bilden, ergaben sich keine Beziehungen – ein überraschendes Ergebnis, wenn man die übliche Richtung des *bias* bei Aggregationen in Rechnung stellt. Die Analysen zur 8. Jahrgangsstufe verwendeten ebenfalls ITBS-Ergebnisse zu Lernzuwachsen und bezogen diese auf Informationen zu den Englisch-, Mathematik- und Sozialkunde-Lernern dieser Schüler. Basierend auf einer Stichprobe von 553 Schülern fanden die Autoren heraus, dass sich nur im Fach Sozialkunde eine stabile Beziehung zwischen den Lernzuwachsen der Schüler und den Gourman-Ratings der Ausbildungsinstitution der Sozialkundelehrer dieser Schüler ergab.

Die Analysen in der 12. Jahrgangsstufe (716 Schüler) verwendeten auf Schülerseite ein anderes Instrument. Konzentrierte man sich auf Testwerte für Lesekompetenz, wurden nur die Qualitätsdaten der Lehrer im muttersprachlichen Unterricht (Englisch) dieser Schüler herangezogen. Anders als bei der 6. und 8. Jahrgangsstufe wurden hier keine Zusammenhänge zwischen Lehrerqualität und Schülerlernen gefunden.

Die zweite Studie in diesem Bereich wurde von Murnane/Phillips (1981) durchgeführt. Den Unterlagen zu einem bestimmten, in den frühen 1970er-Jahren durchgeführ-

5 Der Gourman-Report beinhaltet eine Bewertungsliste von Ausbildungsstätten im Hochschulbereich (Ranking von Undergraduate Colleges und Universitäten. [www.collegeconfidential.com/college\\_rankings/gourman\\_report/htm](http://www.collegeconfidential.com/college_rankings/gourman_report/htm)).

ten wohlfahrtsstaatlich ausgerichteten Reformexperiment in Gary (Indiana) konnten leistungsbezogene Daten von einigen hundert schwarzen Grundschulern entnommen werden, von denen die meisten aus Familien mit niedrigem Einkommen stammten. Als nun die per ITBS (*Iowa Test of Basic Skills*) gemessenen Lernzuwächsen der Schüler mit den Eigenschaften ihrer Lehrer, genauer: mit der Qualität der Ausbildungsstätte dieser Lehrer in Beziehung gesetzt wurde, konnte kein statistischer Zusammenhang hergestellt werden.

Die dritte und letzte Studie in diesem Bereich wurde von Ehrenberg/Brewer (1994) durchgeführt. Es wurde der Datensatz aus *High School and Beyond* (HS&B) verwendet, der auf einer Stichprobe von Zehntklässlern aus dem Jahre 1980 sowie deren erneuter Erfassung im Jahre 1982 basiert. Das Maß für die Schülerleistung waren die zusammengefassten Testwerte zu mathematischen Fähigkeiten, Lesefähigkeit und Wortschatz. Der HS&B-Datensatz selbst enthielt keine Lehrerdaten, aber eine Nachfolgestudie mit je 25 Lehrern an 320 Schulen ermöglichte es Ehrenberg und Brewer, der Frage nachzugehen, ob die einzelnen Schüler mehr lernten, wenn die Qualität des Lehrkörpers an ihrer Schule über dem Durchschnitt lag. Das Maß für die Qualität des Lehrkörpers einer Schule beinhaltete u.a. den Anteil der Lehrer mit einem Master-Grad oder mit mehr als 10 Jahren Berufserfahrung. Für jede Schule ermittelten die Forscher die durchschnittliche Bewertung der Ausbildungsinstitutionen der Lehrer dieser Schule, und zwar auf der Basis von Barrons sechsstufiger Bewertungsskala. Bei den statistischen Analysen wurde auch die Hautfarbe sowie die ethnische Zugehörigkeit berücksichtigt. Es zeigte sich, dass Lehrer, die von höher bewerteten Ausbildungsinstitutionen kommen, bei weißen und schwarzen Schülern höhere Lernzuwächse erreichen. Bei Hispanics zeigten sich keine Zusammenhänge.

### 2.1.2 Zusammenfassende Interpretation

Obwohl es bei einer Interpretation auf Systemebene (s.u.) anders aussehen mag, so lautet die angemessene Interpretation der Studien folgendermaßen: Es existiert ein Zusammenhang zwischen der Qualität der Ausbildungsstätte von Lehrern und den Lernzuwächsen ihrer Schüler. Die Forschung hat nicht immer diese Zusammenhänge entdeckt, aber dort, wo Beziehungen gefunden wurden, waren sie positiv. Die beiden Studien, die Zusammenhänge feststellen konnten, haben aber eine unterschiedliche Aussagekraft. Die von Ehrenberg und Brewer verwendete Kontrolle des sozioökonomischen Status' der Schüler war auf der Ebene der Schüler angesetzt und sehr differenziert (elterlicher Bildungsabschluss, Familieneinkommen, Familienstruktur). Im Gegensatz dazu war es eine Stärke in Sommers und Wolfes Studie, dass alle Analysen auf der Ebene der Einzelschüler durchgeführt wurden und eine große Zahl von schulischen Faktoren berücksichtigt wurde, wie etwa der tatsächliche Schulbesuch, Maßzahlen für die Gleichaltrigengruppe, Schulstruktur und die auf Disziplin bezogene Schumatmosphäre sowie weitere Bedingungen. In ihrer Analyse der Lernzuwächse wurde nicht nur die Zeit zwischen 6. und 8. sowie zwischen 9. und 12. Jahrgangsstufe, sondern auch Daten aus früheren Jahrgangsstufen berücksichtigt, um zusätzliche Evidenzen zu erhalten.

### 2.1.3 Implikationen

Obwohl die Ergebnisse weitere Untersuchungen zu den Auswirkungen der Qualität von Ausbildungsinstitutionen notwendig machen und dazu ermutigen, unterscheiden sich die Implikationen für die Schuladministration darin, welches Qualitätseinschätzungssystem man verwendet. Mit Blick auf Barrons Selektionsraten legen die Ergebnisse nahe, dass an den Zugängen zum Lehrerberuf eine sorgfältigere Auswahl der Interessenten stattfinden sollte. Die traditionelle Beurteilungs- und Auswahlprozedur der staatlichen Schulverwaltung war die Examensnote (*licensure test*), nicht die Einschätzung der Selektivität der Ausbildungsinstitution. Im nächsten Abschnitt werden die Examensprozeduren näher erläutert.

Mit Blick auf die Grouman-Ratings wird deutlich, dass diese Einschätzung tatsächlich auf die breit definierte institutionelle Qualität der Ausbildungsinstitution eingeht. Wenn also die institutionelle Qualität eine Rolle spielen soll, könnte der Schulverwaltung daran gelegen sein, dass die einzustellenden Lehrkräfte mit Abschlüssen aus hoch eingeschätzten Institutionen kommen. Eine Akkreditierung durch eine dritte, neutrale Institution ist der typische Weg, um institutionelle Qualität einschätzen zu können. Forschung, die die Schulverwaltung beraten möchte, wird deshalb die relative Effektivität von Lehrern aus Lehrerbildungsinstitutionen mit unterschiedlichem Akkreditierungsstatus und also: mit unterschiedlicher Ausbildungsqualität untersuchen müssen.

## 2.2 Ergebnisse bei Berufseingangstests<sup>6</sup>

### 2.2.1 Relevante Studien

Die Bedeutung der Testergebnisse von Lehrern (bei Tests zu ihren sprachlichen Fähigkeiten und anderen Tests) für die Lernzuwächse ihrer Schüler ist in sieben anspruchsvollen empirischen Studien untersucht worden. Um eine bessere Übersichtlichkeit zu erreichen, werden diese Studien im Folgenden in drei Gruppen zusammengefasst: a) Studien, die auf den Examensergebnissen der Lehrer basierten; 2) Studien, in denen die sprachlichen Fähigkeiten von Lehrern in Beziehung gesetzt wurden zu den Lernzuwächsen ihrer Schüler, und c) neuere Studien, in denen – auf Lehrerseite – andere Testergebnisse zugrunde gelegt wurden.

*Examensergebnisse von Lehrern:* In zwei Studien ist untersucht worden, ob Schüler von Lehrern mit besseren Examensresultaten größere Lernzuwächse aufweisen als solche, die von Lehrern mit niedrigeren Examensresultaten unterrichtet werden. In diesen beiden Studien wurden Daten aus den *National Teacher Examinations (NTE)* und von den *Texas Examinations of Current Administrators and Teachers (TECAT)* zugrunde gelegt.

6 Frühere Arbeiten zu diesem Abschnitt wurden von Andrew J. Wayne während seiner Arbeit für die National Partnership for Excellence in Teaching fertiggestellt. Diese Initiative wurde vom U.S. Department of Education, Office of Educational Research and Development, Contract RD97124001, gefördert.

Vor den 1980er-Jahren wurden kaum Eignungstests für Lehrer als Grundlage für die Lizenzierung verwendet. Seit dieser Zeit ist die Zahl der Staaten, die Tests zu den sprachlichen Fähigkeiten, zum Wissen in den Unterrichtsinhalten sowie zu berufsbezogenem Wissen als Grundlage für die Lizenzierung durchführen, sehr stark gestiegen: Es sind mittlerweile mehr als 40. Seit 2002 verwenden 41 Staaten Tests zu unterschiedlichen Bereichen, um den Zugang zur Lehrerausbildung zu regulieren und über erste Stufen der Lizenzierung zu entscheiden.

Der NTE wurde vom *Educational Testing Service* in den 1940er-Jahren entwickelt, und in den 1970ern wurde er von einigen Staaten verwendet, um Lehramtsbefähigungen zu vergeben. (vgl. Haney/Madaus/Kreitzer 1987). In dem Maße, wie das Testen von Lehrern zunahm, wurde der NTE zum am häufigsten eingesetzten Instrument. Spezialisiert ausgebildete Lehrer unterzogen sich bereichsspezifischen NTE-Prüfungen, in denen sowohl das Fachwissen als auch die Unterrichtsmethoden in einem der mehr als 30 speziellen Bereiche überprüft wurde. Andere Lehrer unterzogen sich dem üblichen NTE-Verfahren, welches sich auf vier Bereiche richtete: a) Allgemeine Pädagogik sowie psychologische und gesellschaftliche Grundlagen der Erziehung, b) schriftliche Ausdruckskompetenz in der Muttersprache, c) Sozial- und kulturwissenschaftliche Kenntnisse; (d) mathematisch-naturwissenschaftliche Kenntnisse. Der NTE blieb das dominierende Instrument zur Examensprüfung, bis der *Educational Testing Service* ihn in den 1990er-Jahren durch PRAXIS ersetzte.<sup>7</sup>

Die einzige geeignete Studie zu Schülerleistungen, die auch die NTE-Ergebnisse der Lehrkräfte beinhaltet, ist von Summers/Wolfe (1975a, 1977) im Jahre 1971 bei Schülern in Philadelphia durchgeführt worden, wie weiter oben bereits beschrieben.<sup>8</sup> Von den insgesamt 627 Schülern der Jahrgangsstufe 6 lernten diejenigen *weniger dazu (!)*, die von Lehrern unterrichtet wurden, die ein *besseres* Ergebnis beim allgemeinen NTE erzielt hatten. Bei den Schülern der Jahrgangsstufe 8 und 12 ergaben sich keine stabilen Beziehungen zwischen den Lernzuwächsen der Schüler und den Ergebnissen der fachbezogenen NTE.<sup>9</sup>

7 Von mehreren Forschern ist auf den diskriminierenden Bias von Lehrertests gegenüber Kandidaten aus Minoritäten hingewiesen worden, insbesondere in den Südstaaten (Goertz/Pitcher 1985; Graham 1987; Smith/Miller/Joy 1988; Ludlow 2001). Die Verwendung von Tests, die von National Evaluation Systems entwickelt worden waren, wurde in den 1980er Jahren in Alabama verboten, da sie aus testimmanenten Gründen Afroamerikaner diskriminierten (Ludlow 2001).

8 Auch Strauss/Sawyer (1986) haben die Beziehung zwischen den Examensergebnissen von Lehrern und den Schülerleistungen der Schüler dieser Lehrer untersucht. Diese Studie mussten wir jedoch ausschließen, da in ihr kein Pretest eingesetzt wurde.

9 Ein Ergebnis einer dritten Veröffentlichung von Summers/Wolfe (1975b) könnte in diesem Zusammenhang für Verwirrung sorgen. In dieser Veröffentlichung werden zwei positive Ergebnisse vermerkt, die in den anderen Publikationen dieser beiden Autoren nicht anzutreffen sind: Schüler der 8. Jahrgangsstufe lernten mehr von Sozialkundelehrern, die ein hohes Examensresultat bei NTE in Social Studies hatten, und überdurchschnittliche Schüler der Jahrgangsstufe 12 lernten mehr von Englischlehrern, die beim NTE in Englisch besser abgeschnitten hatten. Gleichwohl sind wir der Auffassung, dass wir diese Ergebnisse als „unbestimmt“ zusammenfassen können. Dieses Urteil deckt sich mit den Aussagen von Summers/Wolfe in einer stärker wissenschaftlich ausge-

Die einzige andere Studie, die Examensergebnisse von Lehrern verwendet hat, stützt sich auf den TECAT (*Texas Examination of Current Administrators and Teachers*). Dieser Test bewertet Lese- und Schreibfähigkeiten, wobei hierin inhaltlich sowohl Fachwissen als auch pädagogisch-professionelles Wissen eingeschlossen ist. Ferguson (1991, 1998) hat die TECAT-Ergebnisse für alle texanischen Lehrer aus dem Jahre 1986 verwendet (vgl. Kain/Singleton 1996). Er hat die auf Lesefähigkeit bezogenen Testergebnisse der Lehrer zugrunde gelegt, wobei diese Testergebnisse auch das Ausmaß des professionellen Wissens zu bestimmen erlaubten.<sup>10</sup> Ferguson errechnete das durchschnittliche TECAT-Ergebnis für jeden Schulbezirk in Texas, aus dem er Daten erhalten konnte.<sup>11</sup> Die Schülerdaten wurden ebenfalls auf der Schulbezirksebene aggregiert. Die Testergebnisse (Lesekompetenz und Mathematik) der Schüler basierten auf den Ergebnissen des 1986 durchgeführten *Texas Educational Assessment of Minimum Skills* (TEAMS), einem Multiple-Choice-Test, der mit Schülern der 1., 3., 5., 7., 9. und 11. Jahrgangsstufe durchgeführt wurde.

Auf der Basis einer Analyse der Daten aus 900 Schulbezirken errechnete Ferguson – für jeden Bezirk gesondert – die Differenz zwischen den durchschnittlichen Testergebnissen der 1., 3., 5. und 7. Jahrgangsstufen aus dem Jahre 1986 und den 5., 7., 9. und 11. Jahrgangsstufen aus dem Jahre 1990, also eine gleich bleibende Kohorte von Schülern, wobei angenommen wurde, dass man Ortswechsel von Schülern ignorieren kann. Er fand heraus, dass in solchen Schulbezirken, in denen die Lehrer bessere TECAT-Ergebnisse hatten, die Schüler größere Lernfortschritte aufwiesen, insbesondere zwischen der 3. und 7. Jahrgangsstufe. Später hat Ferguson diese Ergebnisse noch verstärkt (1998), indem er zeigen konnte, dass in den einzelnen Schulbezirken die Lernzuwächse der Grundschüler sich von den Lernzuwächsen der Sekundarschüler in Abhängigkeit von den TECAT-Ergebnissen der Grundschul- bzw. Sekundarschullehrer unterscheiden.

*Sprachliche Fähigkeiten:* In drei Studien wurden die Testergebnisse zu den sprachlichen Fähigkeiten von Lehrern verwendet, allerdings waren dies schon ältere Daten. In der ersten Studie wurden Daten aus dem Coleman-Report (*Equality of Educational Opportunity*; Coleman u.a. 1966) verwendet. Die von Ehrenberg/Brewer durchgeführte Re-Analyse der Daten – unter Hinzufügung von Messergebnissen zu den sprachlichen Fähigkeiten von Lehrern – konnte zeigen, dass diese Messergebnisse einiges an Varianz zwischen Schulen erklären konnten. Die verbleibenden zwei Studien, die sich auf die

richteten Publikation (1975a), in der den erwähnten Zusammenhängen auf allen drei Jahrgangsstufen nachgegangen worden ist.

10 Um mehr über den abgedeckten Umfang und die erhobenen Fähigkeiten des TEACAT zu erfahren, haben wir Kontakt zu den offiziellen Stellen der Texas Education Agency und dem Texas State Board for Educator Certification aufgenommen. Unserer Meinung nach evaluiert der Test nicht nur die Lese- und Schreibfähigkeiten der Lehrer, sondern erhebt in diesem Zusammenhang ganz offensichtlich auch berufsbezogenes Wissen.

11 Ferguson hat die Daten aus etwa 900 (von insgesamt mehr als 1000) Schulbezirken in Texas analysiert. Nur wenige Bezirke sind aufgrund fehlender Daten ausgeschlossen worden. Ebenso sind Dallas und Houston ausgeschlossen worden, da das Gewichtungsschema diesen beiden Bezirken quasi automatisch eine übergroße Dominanz gegenüber allen anderen Bezirken gegeben und das Gesamtergebnis verfälscht hätte.

sprachlichen Fähigkeiten von Lehrern bezogen, basierten auf Daten aus dem oben bereits erwähnten *Gary Income Maintenance Experiment* aus den 1970er-Jahren. Dessen Datensatz beinhaltete neben den Schülerdaten auch unterschiedliche Angaben zu den Lehrern, u.a. auch zu Ergebnissen eines „Quick Word Test“ genannten Instruments.

In ihrer Analyse konzentrierten sich Murnane/Phillips (1981) auf die Testergebnisse der Schüler im ITBS-Wortschatztest. Nachdem verschiedene Lehrereigenschaften statistisch kontrolliert wurden, inklusive Rasse, Geschlecht, Dauer der Berufserfahrung, Vorhandensein eines Master-Grades und Einschätzung der Ausbildungsstätte, ergaben sich keine Zusammenhänge zwischen den Ergebnissen des Wortschatztests der Lehrer und den Lernzuwächsen der Schüler. Als jedoch die Schülerdaten nach Jahrgangsstufen aufgeteilt wurden, zeigte sich, dass die Schüler der 6. Jahrgangsstufe weniger lernten, wenn sie von Lehrern unterrichtet worden waren, die *hohe* Werte im Wortschatztest aufwiesen. Murnane/Phillips wiesen diese Ergebnisse jedoch als unzuverlässig zurück, „denn eine große Zahl von Lehrern hatte zusätzliche Hilfen bei der Erledigung des Tests verwendet“ (ibid., S. 97f.).

Hanushek (1992) reanalyisierte dieselben Daten, wobei seine abhängige Variable der Lernzuwachs der Schüler sowohl in ITBS-Lesetest als auch im ITBS-Wortschatztest war. Er untersuchte Lernzuwächse von der 2. bis zur 6. Jahrgangsstufe innerhalb eines Schuljahres (z.B. Leistungszuwächse vom Ende der 2. bis zum Ende der 3. Jahrgangsstufe). Hanushek berichtet, dass die Testergebnisse der Lehrer im Wortschatztest die Ergebnisse ihrer Schüler beim Lesetest beeinflussten, nicht aber die Ergebnisse beim Wortschatztest. Im Unterschied zu Murnane/Phillips hat Hanushek nicht die Einschätzung der Ausbildungsinstitutionen statistisch kontrolliert.

*Neuere Studien:* Kürzlich sind zwei weitere Studien erschienen. Die erste Studie nutzte die lehrerbezogenen Daten aus einem einzigen Item eines Multiple-Choice-Mathematik-Tests aus. Rowan/Chiang/Miller (1997) analysierten Leistungsdaten aus dem für die USA repräsentativen *National Educational Longitudinal Study* aus dem Jahre 1998 (NELS 1988). Die Schüler wurden einem Mathematik-Test in der 8. und in der 10. Jahrgangsstufe unterzogen. Die Erfassung ihrer Lehrer in der 10. Jahrgangsstufe umfasste u.a. ein einziges mathematisches Testitem auf High-School-Niveau. Die Forscher fanden heraus, dass die Schüler derjenigen Lehrer, die dieses Item korrekt beantwortet hatten, größere Lernzuwächse in Mathematik zwischen der 8. und der 10. Jahrgangsstufe verzeichnen konnten – auch wenn man statistisch kontrollierte (also gewissermaßen als Faktor außer Kraft setzte), ob die Lehrer einen Abschluss in Mathematik hatten oder nicht. Diese Studie ist jedoch nicht so aussagekräftig wie sie wäre, wenn die Erfassung der mathematischen Fähigkeiten der Lehrer sich nicht nur auf ein Item, sondern auf mehrere Items oder auf einen kompletten Mathematik-Test beziehen würde.

Eine zweite neuere Studie basiert auf Personalakten aus Alabama, von denen einige auch die ACT-Ergebnisse der (späteren) Lehrer aus den College-Eingangstests beinhalteten. Ferguson/Ladd (1996) verwendeten zusammengesetzte Testwerte aus dem *American College Testing* (ACT), und zwar kombinierten sie Testkomponenten zu den muttersprachlichen Fähigkeiten (Englisch), Mathematik, Sozialkunde und Naturwissenschaften (American College Testing 1998). Ihre erste Analyse zeigte, dass die Lernzuwächse

der Schüler im Lesen zwischen der 3. und der 4. Jahrgangsstufe positiv mit den durchschnittlichen ACT-Ergebnissen der Lehrer an den Schulen dieser Schüler zusammenhängt. Für die Lernzuwächse im Mathematikunterricht ergaben sich keine klaren Resultate. Ferguson/Ladd analysierten ebenso die Durchschnitte von 127 Schuldistrikten in Alabama. Sie erzeugten auf statistischem Wege Lernzuwächse auf der Basis der Ergebnisse der 3. und 4. Jahrgangsstufe und der 8. und 9. Jahrgangsstufe für das Schuljahr 1990/91. In jedem Distrikt hing der durchschnittliche Unterschied zwischen den Mathematik-Ergebnissen jüngerer und älterer Schüler positiv mit den durchschnittlichen ACT-Ergebnissen der Lehrer dieses Bezirks zusammen. Bei der Analyse auf der Ebene der Einzelschule wie bei der Analyse auf Ebene des Distrikts wurden die Faktoren ‚Dauer der Berufserfahrung der Lehrer‘ und ‚Haben die Lehrer einen Master-Grad: ja/nein?‘ statistisch kontrolliert, also neutralisiert.

### 2.2.2 *Zusammenfassende Interpretation*

Fasst man zusammen, so zeigt sich, dass die sieben Studien zum Zusammenhang von Testleistungen der Lehrer und Lernzuwächsen ihrer Schüler letztendlich uneinheitliche, divergente Ergebnisse aufweisen: Unter den inhaltlich klar bestimmten Ergebnissen gab es fünf positive und zwei negative. Für diese Divergenz gibt es unterschiedliche Erklärungen.

Dass diese sieben klar artikulierten Ergebnisse alle zufällig aufgetreten sind, und dass es also in der Tat keinen Zusammenhang zwischen der Testleistung der Lehrer und den Lernzuwächsen ihrer Schüler gibt, ist sehr unwahrscheinlich. Aber auch wenn man nun annimmt, dass die beiden negativen Ergebnisse zufällig sind, so ist dies ebenfalls nicht recht plausibel, denn diese beiden negativen Ergebnisse resultieren aus Studien mit hoher Beweiskraft: Beide Studien basierten auf individuellen Schülerdaten, und insbesondere Summers/Wolfe haben eine sehr hohe Zahl an sonstigen schulischen Faktoren statistisch kontrolliert.

Eine andere Möglichkeit zur Relativierung der Studien, die kein klares Ergebnis zeigten, besteht in dem Hinweis auf die Inhalte der Tests für Schüler und Lehrer, die verwendet bzw. zugrunde gelegt wurden. So fand Hanushek (1992) heraus, dass die Testleistung von Lehrern bei Wort-Tests die Lernzuwächse ihrer Schüler, gemessen durch ITBS-Tests, erhöhte, nicht aber die Lernzuwächse beim ITBS-Wortschatztest. Dies passt zu dem Ergebnis von Murnane/Phillips (1981), die ebenfalls keinen Zusammenhang zwischen den Testergebnissen von Lehrern beim Wort-Test und den Lernzuwächsen der Schüler dieser Lehrer gefunden haben.

Zukünftige Studien werden vielleicht aufdecken, dass die Beziehung zwischen den Testergebnissen von Lehrern und den Lernzuwächsen ihrer Schüler auch davon abhängen, dass die bei beiden Gruppen eingesetzten Testinstrumente aufeinander abgestimmt sind. Auf Grundlage der vorliegenden Studien kann man dies jedoch noch nicht mit Sicherheit sagen.

Letztendlich stützt sich die plausibelste Erklärung der uneinheitlichen Ergebnisse auf eine genaue Inspektion der statistischen Kontrollen, die in den sieben Studien prak-

tiziert worden sind. Die beiden Studien, die negative Ergebnisse erbrachten, haben insbesondere die Einschätzung der Qualität der Ausbildungsinstitutionen kontrolliert. Im Gegensatz dazu hat keine der Studien, die positive Ergebnisse zwischen den Testergebnissen der Lehrer und den Lernzuwächsen ihrer Schüler aufzeigen konnten, die Einschätzung der Qualität der Ausbildungsinstitutionen dieser Lehrer statistisch kontrolliert. Die negativen Ergebnisse spiegeln den Effekt von Lehrertestergebnissen wider, nachdem die unterschiedliche Einschätzung der Qualität ihrer Ausbildungsinstitution statistisch neutralisiert worden ist. Möglicherweise ist die Qualität der Ausbildungsinstitution ein ähnlich wichtiger Faktor wie die Testleistung der Lehrer. Somit könnten die negativen Ergebnisse die positiven sogar noch stützen/untermauern: Schüler lernen mehr von Lehrern mit höheren Testwerten. Diese Testwerte sind aussagekräftig – so lange man nicht die Qualität der Ausbildungsinstitutionen dieser Lehrer mit berücksichtigt.

### 2.2.3 Implikationen

Obwohl das Ergebnis, dem zufolge Lehrkräfte mit besseren Testwerten größere Lernzuwächse bei Ihren Schülern erzeugen, sicherlich dazu motiviert, sehr gründliche und anspruchsvolle Prüfungen von angehenden Lehrern durchzuführen, muss man doch berücksichtigen, dass die Schulverwaltung auf unterschiedliche Typen von Lehrbefähigungen (*licensure tests*) zurückgreifen kann. Zwischen 1990 und 2000 haben 34 Staaten bei zukünftigen Lehrern grundlegende Sprach- und Rechentests angewandt. Zusätzlich haben 30 Staaten Tests im Bereich des Wissens zu den Unterrichtsfächern eingesetzt, und 25 Staaten verlangten von zukünftigen Lehrern auch das Bestehen eines Tests zu pädagogischem Wissen (Youngs/Odden/Porter 2001). Die Synthese aus den vorliegenden Studien sagt nichts darüber aus, welche Wissensart getestet werden sollte.

Um der Schulverwaltung brauchbare Informationen zu geben, sollte es mehr Forschung zum Zusammenhang zwischen dem Lernzuwachs der Schüler und den Examensergebnissen der Lehrer dieser Schüler geben, die sich auf solche Tests für angehende Lehrer stützen, welche gegenwärtig tatsächlich verwendet werden. (Die Tests in den Studien, über die wir berichten konnten, sind allesamt nicht mehr in Gebrauch). Gegenwärtig setzen viele Staaten den Test PRAXIS I zu grundlegenden Fähigkeiten ein, den PRAXIS II zum Wissen in den Unterrichtsfächern und/oder den Test PRAXIS II zu pädagogischem Wissen ein (alle vom *Educational Testing Service* in den 1990 Jahren entwickelt). Andere Staaten haben eigene Testverfahren entwickelt oder von anderen privatwirtschaftlichen Testinstituten entwickeln lassen.

Schließlich sollten Forscher auch die kürzlich neu entwickelten Formen der Beurteilung der beruflichen Leistung von Lehrern berücksichtigen, die in Connecticut, North Carolina, Ohio und anderen Staaten entwickelt und eingesetzt worden sind. Diese *assessment*-Verfahren wurden erarbeitet, um die herkömmlichen Papier- und Bleistift-Tests zu verbessern. Studien, die unter Zugrundelegung solcher Testverfahren für Lehrer eine Verbindung zu den Lernzuwächsen der Schüler dieser Lehrer herstellen, sind sicherlich informativ für alle, die über neue Beurteilungsverfahren für Lehrer nachdenken

(z.B. Pecheone/Rogers/Moirs 2001). Diese Empfehlungen decken sich mit denjenigen Schlussfolgerungen, die eine Expertengruppe des *National Research Council* zur Frage der Zugangsprüfungen zum Lehrerberuf und der Qualität der Lehrerverarbeit entwickelt hat. Im Bericht dieser Gruppe wird ein „multidisziplinäres, multimethodisches Forschungsprogramm“ gefordert, das „Zusammenhänge zwischen Examensnoten, anfänglicher beruflicher Kompetenz und den Leistungen der Schüler untersucht“ (*National Research Council Committee on Assessment and Teacher Quality* 2001, S. 29). Wir hoffen, dass unsere Forschungsübersicht wertvolle Informationen für ein solches Forschungsprogramm bereitstellt.

### 2.3 Abschlussart und Studienprogramm

#### 2.3.1 Relevante Studien

Bis vor kurzem lagen noch keine Daten dazu vor, die es der Forschung erlaubt hätten zu ermitteln, ob Schüler mehr von solchen Lehrern lernen, die über bestimmte Abschlüsse verfügen bzw. die ein bestimmtes Studienprogramm durchlaufen haben. Die zur Verfügung stehenden Datensätze beinhalteten Informationen zu den Abschlüssen der Lehrer (z.B. Bachelor, Master), und die Ergebnisse waren uneinheitlich. Die meisten Studien zeigten in dieser Frage uneinheitliche und unbestimmte Ergebnisse (Harnisch 1987; Hanushek 1992; Link/Ratledge 1979; Murnane 1975; Murnane/Phillips 1981; Rivkin/Hanushek/Kain 2001; Summers/Wolfe 1975a, 1977), und die vier Studien, die zu aussagekräftigen Ergebnissen führten, waren teils positiv (Ferguson/Ladd 1996), teils negativ (Eberts/Stone 1984; Ehrenberg/Brewer 1994; Kiesling 1984). Eine nähere Überprüfung dieser vier Studien erlaubte keine ‚Versöhnung‘ der konträren Ergebnisse. Obwohl drei von ihnen einen negativen Einfluss nahe legen, sind die positiven Ergebnisse von Ferguson/Ladd (1996) doch überzeugend. Insgesamt stabilisierte sich jedoch das uneinheitliche Bild.

Die neueren Fortschritte bei der Datensammlung zu Abschlüssen und Studienprogrammen führten zu Ergebnissen, die deutlich werden ließen, dass die früheren, gemischt-unbestimmten Resultate zu den Auswirkungen der Abschlussart z.T. darauf zurückzuführen waren, dass diese früheren Studien überhaupt nicht mit erfasst hatten, ob der Inhaltsbereich oder das Fach, in dem ein Lehrer einen akademischen Grad erworben hatte, auch derjenige bzw. dasjenige war, in dem der Lehrer unterrichtete. Man hatte nur die Abschlussart, nicht aber die Fächer erfasst. Dieses Defizit wurde deutlich, als interessierte Forscher die Daten aus NELS:88 verwendeten; diese Studie zu Schülerleistungen ist weiter oben bereits erwähnt worden. NELS:88 ist eine national repräsentative Erfassung von über 24.000 Schülern der 8. Jahrgangsstufe, die im Frühjahr 1988 durchgeführt worden ist. Eine Teilstichprobe dieser Schüler wurde in der 10. Jahrgangsstufe (1990) und der 12. Jahrgangsstufe (1992) erneut getestet. Jedes Mal wurden ein oder mehrere fachbezogene Tests durchgeführt, und zwar in Mathematik, Naturwissenschaften, Englisch (schriftlich) und Geschichte. Aus diesem Grunde erlauben die NELS.88-Daten längsschnittliche Analysen zum Lernfortschritt der Schüler von der 8. bis zur

12. Jahrgangstufe in einzelnen Fächern. Die NELS:88-Daten enthielten auch Informationen zu relevanten Schüler-, Lehrer- und Schul-Variablen.

Drei Analysen basieren auf den detaillierten lehrerbezogenen Daten in NELS:88 über Abschlussarten. Die Analyse von Goldhaber/Brewer (1997a) macht den zentralen Punkt deutlich: Wenn die Autoren lediglich fragten, ob Schüler der 10. Jahrgangsstufe in Mathematik von solchen Lehrern, die einen Master-Grad haben, besser lernen als die Schüler von Lehrern, die diesen Grad nicht aufweisen, so zeigten sich keine Unterschiede. Erst wenn man berücksichtigt, *in welchem Fach* diese Lehrer einen Master-Grad inne hatten, zeigten sich signifikante Ergebnisse. Die erwähnten Schüler hatten dann größere Lernzuwächse in Mathematik, wenn sie von Lehrern unterrichtet worden waren, die einen Master-Grad in Mathematik hatten, verglichen mit denjenigen Schülern, deren Lehrer entweder einen niedrigeren akademischen Abschluss oder aber einen Master-Abschluss in irgendeinem anderen Fach (als Mathematik) hatten. Das Gleiche konnte man auf der Bachelor-Ebene feststellen: Schüler lernten mehr von Lehrern mit einem Bachelor-Grad im Fach Mathematik als von Lehrern mit einem Bachelor-Grad in einem anderen Fach. Die Auswirkungen dieser Faktoren – Innehaben eines fachlich parallelen Bachelors zum Lernzuwachs der Schüler in diesem Fach – war sehr stabil und unabhängig von anderen Lehreigenschaften; statistisch kontrolliert wurde auf der Seite der Lehrer dabei die Note in Mathematik sowie die Dauer der Berufserfahrung. Die anderen beiden Studien auf der Basis von NELS:88 (Goldhaber/Brewer 2000; Rowan/Chiang/Miller 1997) bestätigten im Wesentlichen diese Ergebnisse.

Falls es also entscheidend ist, dass man einen fachlich adäquaten Master haben muss, um in diesem Fach als Lehrer erfolgreich zu sein, so müsste man annehmen, dass in irgendeiner Weise auch die fachbezogenen Studiererfahrungen ein Prädiktor für den späteren beruflichen Erfolg sein müssten. Zwei Studien sind dieser Vermutung nachgegangen. Eberts/Stone (1984) erfassten die Zahl der Mathematik-Kurse auf College-Ebene der Lehrer. Es zeigte sich kein Zusammenhang zu den Lernzuwächsen in Mathematik der Viertklässler. Eine zweite Studie, die ebenfalls die Studiererfahrung berücksichtigte, fand jedoch Zusammenhänge. Monk/King (1994) verwendeten Leistungsdaten aus dem *Longitudinal Survey of American Youth* (LSAY), eine Studie, die auf einer Zufallsauswahl von 2831 Schülern staatlicher Schulen der 10. Jahrgangsstufe vom Herbst 1987 bis zur Abschlussklasse der Highschool. Die Schüler absolvierten in jedem Herbst Tests in Mathematik und Naturwissenschaften von 1987 bis 1989, die auf dem NEAP basierten. Im Rahmen von LSAY wurden auch die Mathematik- und Naturwissenschaften-Lehrer der zur Stichprobe gehörenden Schüler erfasst; eine gesonderte Umfrage erfasste die Berufserfahrung der Lehrer und ihre Studiererfahrungen. Um diese Studiererfahrungen richtig zu bewerten, wurden alle Undergraduate- und Graduate-Kurse der Lehrer in drei Kategorien eingeteilt: Mathematik, Life Sciences, physikalische Wissenschaften. Obwohl es viele unbestimmte Relationen gab, fanden Monk/King doch auch einige positive Zusammenhänge zwischen Lernzuwächsen der Schüler in Mathematik und den Studiererfahrungen der Lehrer in Mathematik heraus: Bei Kontrolle der Berufserfahrung der Lehrer zeigte sich, dass Zehntklässler in Mathematik mehr lernen, wenn sie von Lehrern unterrichtet wurden, die mehr Mathematik-Kurse absolviert hatten als andere.

Die Ergebnisse der ebenfalls auf LSAY-Daten basierenden Studie von Monk (1994) werden zwar häufig in diesem Kontext zitiert, wir haben sie jedoch aus unserer Übersicht ausgeschlossen, da es keine Kontrolle des sozioökonomischen Status gegeben hat. Hätten wir sie aufnehmen können, hätte sie die These unterstützt, dass die Intensität des Fachstudiums in Mathematik mit den Lernzuwächsen der Schüler dieser Lehrer in Mathematik zusammenhängt. Für Mathematik scheint es also einen klaren Zusammenhang zwischen der Intensität des Fachstudiums und dem Lernerfolg der späteren Schüler zu geben.

Wie sieht es in andern Fächern aus? Hier liegen nur inkonsistente und unbestimmte Ergebnisse vor. Goldhaber/Brewer (1997a) haben Analysen im Bereich der Fächer Naturwissenschaften, Englisch und Geschichte durchgeführt. Im Bereich der Naturwissenschaften wurden positive Zusammenhänge entdeckt, nicht aber in den anderen Fächern. In einer späteren Studie (Goldhaber/Brewer 2000) wurden zwar auch Zusammenhänge entdeckt, die jedoch nicht signifikant waren. Monk/King (1994) sind auch Zusammenhängen in Naturwissenschaften nachgegangen und erhielten nur ein bestimmtes Ergebnis: In den ersten Jahren der High School lernen die Schüler dann *weniger* dazu, wenn sie von Lehrern unterrichtet worden waren, die mehr Physik-Kurse absolviert hatten als von solchen, die weniger absolviert hatten.

### 2.3.2 Zusammenfassende Interpretation

Bei einer zusammenfassenden Interpretation muss man die Ergebnisse fachbezogen interpretieren. In den Fächern Geschichte und Englisch haben sich keine klaren Ergebnisse gezeigt; insofern kann unsere Übersicht keine Schlussfolgerungen für diese Fächer ziehen.

In den Naturwissenschaften haben wir zwei klare Ergebnisse – sie weisen jedoch in die entgegengesetzte Richtung: Naturwissenschaften beinhalten jedoch sowohl die physikalischen Wissenschaften wie die *Life Sciences*, sodass die konträren Ergebnisse vielleicht auch darauf zurückzuführen sind, dass hier nicht differenziert genug vorgegangen wurde. Wenn man also in Zukunft dem Zusammenhang von Lehrerkompetenzen und den Lerngewinnen der Schüler dieser Lehrer nachgehen will, so wird man eine feinkörnigere Perspektive anlegen müssen.

Im Fach Mathematik jedoch waren alle klaren Ergebnisse eindeutig positiv, sodass man begründet annehmen kann, dass Schüler dann von ihren Lehrern mehr lernen, wenn diese Lehrer möglichst viele Kurse in Mathematik belegt hatten und in diesem Fach auch die höheren Abschlussarten erworben hatten. Allerdings beziehen sich alle diese Ergebnisse auf High School-Schüler. Die unbestimmten Ergebnisse von Eberts/Stone (1984) bezogen sich auf Grundschüler.

### 2.3.3 Implikationen

Beinahe ausnahmslos bezahlen die Schuldistrikte in den USA die Lehrer nach der Höhe ihres Abschlusses, d.h. je höher der Abschluss, desto höher das Gehalt. Die Studien, die

wir ausgewertet haben, bestreiten nicht, dass dies möglicherweise eine sinnvolle Strategie ist. Sie machen allerdings deutlich, dass im Fach Mathematik der Zusammenhang zwischen studiertem Fach und dem Lehrerfolg in diesem Fach wichtig ist. Ein anderer wichtiger Punkt für die Bildungsverwaltung ist, dass sie genau festlegen kann, welche Studienerfahrungen notwendig sind, um eine bestimmte Arbeit als Lehrer übernehmen zu dürfen. Viele Staaten haben solche Anforderungsprofile definiert. Unsere Forschungsübersicht legt nahe, dass anspruchsvolle Anforderungsprofile in Verbindung mit steigenden Gehältern wahrscheinlich einen positiven Effekt auf die Schülerleistungen haben – insbesondere in Mathematik. Ob es diesen *payoff* gibt, hängt von den Kosten ab, die angehende Lehrer haben, um die Anforderungen zu erfüllen sowie natürlich von der Stärke der Beziehung, über deren Größe man aus methodischen Gründen jedoch nichts aussagen kann.

Abschließend muss auf den interessanten Punkt aufmerksam gemacht werden, dass die von uns analysierten Studien eine unklare Antwort auf die Frage geben, ob Schüler mehr von solchen Lehrern lernen, die fach- und fachdidaktische Studien absolviert und entsprechende Abschlüsse haben (also z.B. Mathematikdidaktik – und nicht nur Mathematik). Keine der im Zusammenhang mit NELS:88 durchgeführten Lehrerstudien haben danach gefragt. Hatte ein Lehrer auch Mathematikdidaktik studiert, wurde dies unter „Pädagogik“ subsumiert, aber es ist möglich, dass viele dieser Lehrer lieber das Fach angegeben hätten. Monk/King (1994) berichten nichts darüber, wie sie mit diesem Problem umgegangen sind.<sup>12</sup> Eberts/Stone (1984) verwendeten die Kategorie „auf Mathematikunterricht bezogene Kurse“ und schlossen vermutlich darin sowohl Fachstudien als auch fachdidaktische Studien ein. Zusammenfassend kann man also sagen, dass bislang die Forschung noch nicht hinreichend differenziert genug vorgegangen ist. Die Konsequenz ist klar: Bei der Untersuchung der Wirkung von Studienerfahrungen der Lehrer wird man zwischen den reinen Fachstudien und den fachdidaktischen Studien unterscheiden müssen.

## 2.4 Art der Zulassung (certification status)

### 2.4.1 Relevante Studien und Ergebnisse

Wie schon bei den Abschlüssen, so zeigen sich die Effekte der formellen Zulassung zum Lehrerberuf (*certification*) nur dann, wenn man bei der Zulassung immer auch das Fach mit berücksichtigt, für das die Zulassung erfolgt; dies hat sich im Fach Mathematik ergeben. Nur zwei Studien, die die Kriterien dieser Forschungsübersicht erfüllten, haben den Effekt der Zulassung untersucht; weitere Studien, die sich diesem Thema gewidmet haben, weisen keinen längsschnittlichen Charakter auf oder haben den sozioökonomischen Status der Schüler nicht erfasst.

<sup>12</sup> Die Studie von Monk/King (1994), die aufgrund der nicht durchgeführten Kontrolle des Faktors sozioökonomischer Status ausgeschlossen werden musste, hat jedoch explizit zwischen fachdidaktischen Studien und Fachstudien unterschieden.

In der oben bereits beschriebenen Studie von Goldhaber/Brewer (1997a) wurden zwei Indikatoren für die Zulassung überprüft. Im ersten Fall wurde lediglich unspezifiziert gefragt, ob die Lehrer überhaupt eine regelgerechte Zulassung als Lehrer haben oder nicht. Es ergab sich nur ein klarer Zusammenhang: Schüler hatten im muttersprachlichen Unterricht (Englisch) dann *geringere* Lernfortschritte, wenn sie von Lehrern mit (irgendeiner) Lehrer-Zulassung unterrichtet wurden. Im zweiten Fall wurden zusätzliche Informationen über das spezifische Fach mit berücksichtigt, in dem Lehrer ihre Zulassung hatten. Die Ergebnisse für den muttersprachlichen Unterricht wurden unbestimmt, wodurch nahe gelegt wird, dass das im ersten Fall erhaltene negative Ergebnis durch Lehrer verursacht worden ist, die über keine Zulassung für das Erteilen des muttersprachlichen Unterrichts hatten. Die Ergebnisse im Fach Geschichte waren ebenfalls unklar. Für das Fach Mathematik wurde jedoch berichtet, dass Schüler höhere Lernfortschritte erzielten, wenn sie von Lehrern unterrichtet worden waren, deren Lehramtsberechtigung im Fach Mathematik lag, als solche Schüler, die auf keine auf das Fach Mathematik bezogene Lehramtsbefähigung verweisen konnten – wobei die letztgenannte Gruppe sowohl solche Lehrer umfasste, die überhaupt keine Lehramtsberechtigung hatten als auch solche, die zwar eine Lehramtsberechtigung haben, aber nicht im Fach Mathematik. Im Bereich der naturwissenschaftlichen Fächer zeigten sich ähnliche positive Beziehungen.

Bemerkenswerterweise wurde in dieser Studie nichts über die Unterschiede in der Art der Zulassung zum Lehrerberuf ausgesagt. Die meisten Staaten sprechen unterschiedlich geartete Lehramtsberechtigungen für ein bestimmtes Fach aus, und sowohl bei Bildungsforschern als in der Bildungsadministration gibt es Unsicherheiten über die unterschiedliche Effektivität der Lehrer mit einer nicht-regelkonformen Lehramtsbefähigung (dt. in etwa: „Seiteneinsteiger“). Die Befürworter von solchen „alternativen Zulassungsformen“ behaupten, dass eine gewisse Reduktion der Zugangsstandards solche Personen anzieht, die über höhere Qualitäten verfügen, und zwar deshalb, weil die Eingangsbarrieren niedriger sind.

In einer zweiten Studie, bei der fachbezogene Lehramtsberechtigungen zugrunde gelegt wurden, haben Goldhaber/Brewer explizit nach unterschiedlichen Arten von Zulassungen gefragt. Wie schon erwähnt, haben sich Goldhaber/Brewer (2000) auf die Lernfortschritte der von NELS:88 erfassten Schüler (10. und 12. Jahrgangsstufe) gestützt. Die Befragung der Lehrer der Schüler der 12. Jahrgangsstufe umfasste auch eine Frage nach der Art der Lehramtszulassung im Fach Mathematik und in naturwissenschaftlichen Fächern. Diese Angaben wurden in Beziehung gesetzt zu den Lernfortschritten der Schüler dieser Lehrer in Mathematik und in den Naturwissenschaften. Das klare Ergebnis dieser Untersuchung war, dass die Schüler von Lehrern, die eine regelgerechte Standard-Qualifizierung für ihren Beruf innehatten, größere Lernfortschritte aufwiesen als solche Schüler, deren Lehrer entweder keine Zulassung für Mathematik hatten oder aber eine von einer privaten Institution erteilte Zulassung für Mathematik aufweisen konnten.

Diese Studie hat dann zu weiteren Analysen und Diskussionen geführt (vgl. Darling-Hammond u.a. 2001; Goldhaber/Brewer 2001). Goldhaber/Brewer (2000) verglichen

die Lernzuwächse von Schülern in Mathematik, deren Lehrer die Standard-Qualifikation für Mathematik-Lehrer aufwiesen, mit den Lernzuwächsen derjenigen Schüler, deren Lehrer davon abweichende Zulassungen aufwiesen: „zeitlich befristete, provisorische, oder Not-Qualifikationen (bevor eine volle Zulassung erteilt wird, muss Weiterbildung erfolgen)“<sup>13</sup>. Bei Kontrolle anderer Variablen ihres Modells konnten die Autoren keine Unterschiede feststellen und interpretierten dieses unbestimmte Ergebnis dahingehend, dass die Lehrer trotz unterschiedlicher Zulassungen gleich effektiv seien. Die sehr strengen Standards, die wir unserer Übersicht zugrunde legen, verbietet es jedoch, irgendwelche Schlussfolgerungen aus unbestimmten Ergebnissen zu ziehen.<sup>13</sup>

Schließlich wurde eine weitere, für diesen Zusammenhang relevante Studie von Raymond/Fletcher/Luque (2001) durchgeführt. Sie basierte auf Daten des Schuldistrikts Houston und wollte ermitteln, ob Schüler dann mehr lernen, wenn ihre Lehrer sich dem TFA unterzogen hatten (*Teach for America*). Faktisch haben sie jedoch nicht die TFA-Lehrer mit solchen Lehrern verglichen, die eine Standard-Berechtigung in ihrem Unterrichtsfach aufwiesen. Stattdessen haben sie TFA-Lehrer mit *allen anderen* Berufsanfängern im Lehrerberuf verglichen, wobei andere Faktoren kontrolliert wurden. Die Gruppe der Berufsanfänger umfasste eine große Bandbreite von unterschiedlich qualifizierten Personen. Insofern informiert diese Studie lediglich den Schuldistrikt Houston darüber, ob sie TFA-Lehrer akzeptieren sollen oder nicht; sie erlaubt aber keine weitergehenden Verallgemeinerungen.

#### 2.4.2 Zusammenfassende Interpretation

Die beiden in diesem Abschnitt zum Status der Zulassung ausgewerteten Studien sind methodisch sehr aussagekräftig angelegt. Beide gehen von der Individualebene aus, und beide verwenden eine beachtliche Zahl von sozioökonomischen Statusfaktoren. Die Kontrolle anderer Schul- und Lehrerfaktoren war nicht so ausführlich wie in anderen Studien, aber gleichwohl vorhanden. Weil nicht deutlich wurde, wie der Status der Zulassung erfasst wurde, und die zweite Studie lediglich Aussagen über die Auswirkungen von Standard-Zulassungen erlaubte, schließen wir insgesamt hieraus, dass Schüler im Mathematikunterricht dann mehr lernen, wenn ihre Lehrer eine Standard-Zulassung für den Mathematikunterricht innehaben (dies im Vergleich mit Lehrern, deren Mathematik-Examen von privaten Ausbildungsinstitutionen stammt oder die eben keine Examen in Mathematik haben).

13 Aus Raumgründen können wir nicht die gesamte Diskussion zwischen Goldhaber/Brewer (2000) und Darling-Hammond u.a. (2001) nachzeichnen. Es ist aber wichtig zur Kenntnis zu nehmen, dass die Kontrahenten sich darüber einig sind, dass in der o.g. Studie die Lehrer mit Nicht-Standard-Zulassungen im Durchschnitt in anderer Hinsicht etwa die gleich hohen Qualifikationen aufwiesen wie Lehrer mit Standard-Zulassungen. Für uns stellt sich die Frage nach den Kriterien, anhand derer man diesbezüglich zwischen Lehrern unterscheidet; weitere Forschung hierzu wird notwendig sein.

### 2.4.3 Implikationen

Die Ergebnisse zu den Auswirkungen der formalen Zulassung entsprechen denjenigen zu den Auswirkungen von Abschlüssen und Studienerfahrungen. Kurz gefasst: fachspezifische Faktoren wirken sich aus. Dass die Lehrer mit regulären Mathematik-Examina und Zulassungen diejenigen Lehrer übertreffen, die diese nicht aufweisen, weist darauf hin, dass die Unterrichtsstandards für den Mathematikunterricht, die viele Staaten formuliert haben, alles in allem Sinn machen.

Wenn man hieraus Konsequenzen für die Bildungsadministration ziehen will, so muss man immer im Auge behalten, dass jeder Staat seine spezielle Stellenanforderungen definiert, die erfüllt sein müssen, damit man eine Standard-Zulassung als Mathematiklehrer bekommt. Das Ergebnis sagt nichts über die je spezifischen Anforderungen aus, die dabei der Sache nach wichtig sind. Die Frage, ob eine Vereinheitlichung dieser Anforderungen die Qualität verändern würde, kann angesichts der gegebenen Forschungslage nicht beantwortet werden.

Um klarere Ergebnisse hinsichtlich der Zulassung zu erhalten, müssten Designs entwickelt werden, die die je besonderen, in verschiedenen Staaten geltenden Anforderungen an je spezielle Zulassungen in den Staaten miteinander in Beziehung setzt. Der von der U.S. Bundesregierung finanzierte *Schools and Staffing Survey* hat zwar Muster der Zertifizierung beschrieben, hat aber keine Schülerleistungen erfasst. Immerhin ist Einiges über die konkrete Art der Lehrerausbildung deutlich geworden.<sup>14</sup> Um den Zusammenhang zwischen der Art der Zulassung von Lehrern und den Schülerleistungen dieser Lehrer zu erfassen, könnte man sich zunächst auf eine kleinere Gruppe von Staaten konzentrieren, die über ähnliche Ausbildungs- und Zulassungsstrukturen verfügen. Weil es das Ziel der Reform der Zulassungsverfahren ist, die durchschnittliche Qualität derjenigen Lehrer zu verbessern, die sich fest an den Lehrerberuf binden, müsste man auch die Zulassungen derjenigen Lehrer untersuchen, die den Lehrerberuf verlassen. Diese Empfehlungen decken sich mit denjenigen ähnlicher Forschungsübersichten.<sup>15</sup>

## 2.5 Andere Charakteristika

Die auf die Art der Ausbildung zurückgehenden Eigenschaften von Lehrern, die im Vorangegangenen thematisiert worden sind (Einschätzung der Ausbildungsstätte, Testleistungen, Abschlüsse und Studienerfahrungen, Art der Zulassung), sind von der Forschung soweit untersucht worden, dass bestimmte Schlussfolgerungen für die Bildungsverwaltung daraus gegeben werden können. Hinsichtlich vieler anderer Eigenschaften von Lehrern existiert entweder keine Forschung oder aber die Ergebnisse sind uneindeutig. In diesem letzten Abschnitt werden solche Eigenschaften bzw. Forschungen kurz erörtert.

Die wohl hervorstechendste Lehrereigenschaft, über die im Vorangegangenen noch nichts gesagt wurde, ist die Dauer der Berufserfahrung. Von den 21 Studien zu Schüler-

14 Nähere Informationen zu diesem Survey finden sich unter [www.nces.ed.gov/surveys/sass](http://www.nces.ed.gov/surveys/sass).

15 Vgl. dazu Goldhaber/Brewer (2000) und Darling-Hammond u.a. (2001).

leistungen beinhalten 19 auch Informationen darüber, über wie viele Jahre Berufserfahrung die Lehrer verfügten. Die meisten der eindeutigen Ergebnisse waren positiv (d.h. größere Berufserfahrung von Lehrern führt zu größeren Lernfortschritten der Schüler dieser Lehrer). Aufgrund der zuerst von Murnane/Phillips (1981, S. 94 – 97) genannten Gründe haben wir jedoch entschieden, dass die auf die Dauer der Berufserfahrung basierenden Ergebnisse kaum zu interpretieren sind. Erstens schließt das Merkmal ‚Berufserfahrung‘ naturgemäß die Wirkung des Faktors, ob ein Lehrer in einer Zeit des Lehrermangels oder der Lehrerarbeitslosigkeit eingestellt worden ist, mit ein. Sinnvolle Interpretationen setzen aber voraus, dass man genau kontrolliert, ob die Einstellung eines Lehrers während der einen oder der anderen Phase gefallen ist. Zweitens schließt Berufserfahrung Unterschiede hinsichtlich der Motivation für die Arbeit mit ein, die je nach biografischer Situation eines Lehrers unterschiedlich aussehen kann. So wirkt sich etwa sehr stark aus, wenn Lehrkräfte selbst Eltern werden und insofern viel Energie auf die eigenen Kinder verwendet werden muss. Und wenn drittens eine unterschiedliche Effektivität zwischen denen besteht, die den Beruf verlassen, und denen die ihn weiterführen, würden Berufserfahrung diese Unterschiede ebenfalls verdecken – und dabei sind diese Unterschiede selbst noch dynamisch, d.h. sie wandeln sich zusammen mit generellen kulturellen Trends sowie bedingt durch die jeweiligen Bedingungen am Lehrerarbeitsmarkt. Indem also der Faktor ‚Dauer der Berufserfahrung‘ gewissermaßen zahllose andere Effekte mitbeinhaltet, kann man ihn als eine wertvolle Kontroll-Variable einsetzen. Aus dem gleichen Grund sind aber die Zusammenhänge zwischen der Dauer der Berufserfahrung von Lehrern und der Lernleistung der Schüler dieser Lehrer sehr schwer zu interpretieren.

Eine andere Eigenschaft von Lehrern, auf die wir in unserer Forschungsübersicht nicht eingegangen sind, ist die Rasse. Dieser Faktor wurde ausgeschlossen, weil wir uns auf klare, eindeutige Ergebnisse konzentrieren wollten. Die Forschung zu den Auswirkungen der Hautfarbe eines Lehrers – und insbesondere zu der Frage, ob Schüler dann mehr lernen, wenn sie von Lehrern ihrer eigenen Rasse oder Ethnie unterrichtet werden – sind sehr gemischt (vgl. Ferguson 1998). Falls man überhaupt jemals eindeutige Ergebnisse erhalten wird, sollte man vielleicht nur nach sehr sorgfältigen Überlegungen hieraus überhaupt Konsequenzen ziehen.

Unsere Forschungsübersicht ist ebenfalls nicht der Frage nachgegangen, wie es sich auswirkt, wenn Lehrer über erziehungswissenschaftliche Abschlüsse verfügen. Die Untersuchungen unterscheiden nicht explizit zwischen fachbezogenen Abschlüssen, fachdidaktischen Abschlüssen und erziehungswissenschaftlichen Abschlüssen. Gäbe es solche Studien, ließe sich eine empfindliche Forschungslücke schließen.

### **3. Schlussfolgerungen**

Es fällt nicht schwer, die Ergebnisse zusammenzufassen. Forschung hat gezeigt, dass Schüler von Lehrern mit bestimmten Charakteristika mehr lernen als von Lehrern, die diese Charakteristika nicht aufweisen. Lehrer, die aus positiv beurteilten Ausbildungs-

stitutionen kommen, erzeugen höhere Lerngewinne bei ihren Schülern als Lehrer, die aus schlechter bewerteten Ausbildungsinstitutionen kommen. Es sollte detaillierter untersucht werden, wie groß die relative Bedeutung dieses Faktors sowie auch des bei Lehrern gemessenen Wissens und der Fähigkeiten ist. Wenn wir auf die Abschlüsse, die Studienerfahrungen und die Art der Zulassung schauen, so haben wir es mit inkonsistenten Ergebnissen zu tun – mit Ausnahme des Faches Mathematik. In diesem Fach lernen die Schüler der High School mehr von Lehrern, die Mathematik studiert haben, die einen Abschluss in Mathematik vorweisen können, und die eben auch für dieses Fach eine Zulassung besitzen.

Obwohl dies für die Bildungsverwaltung sicherlich wichtig ist, bleiben doch viele Fragen unbeantwortet. Es könnte z.B. sein, dass die Forschungsergebnisse anders aussehen, wenn man spätere Bildungsabschlüsse oder das spätere Einkommen von Schülern mit berücksichtigt. Für viele vermutlich wichtige Eigenschaften von Lehrern – Dauer der Berufserfahrung, Rasse bzw. Ethnie oder Kompetenz auf der Ebene der Unterrichtsmethoden – existieren jedoch keine Studien oder aber nur solche, die inkonsistente Ergebnisse erbracht haben. Weiterhin ist es wichtig zu berücksichtigen, dass sich die Zusammensetzung der Lehrerschaft über längere Zeitstrecken hinweg verändert. In den letzten drei Jahrzehnten hat es hier sicherlich beträchtliche Veränderungen gegeben.<sup>16</sup> Und schließlich: Die oben als sicher ausgewiesenen Zusammenhänge liegen zwar vor – aber es ist nichts über ihre Effektstärke bekannt. Was wäre, wenn deren Effektstärke extrem klein wäre? Vieles in diesem Feld bleibt also immer noch der Intuition und der Einschätzung von Bildungsadministratoren vorbehalten.

Angaben über Effektstärken sind in der Tat zentral für die gegenwärtige Diskussion um den Lehrerberuf und seine Zukunft. Ein wichtiger Denkansatz innerhalb der bildungspolitischen Diskussion um die Qualität von Lehrern behauptet, dass es nicht sinnvoll ist, die Eigenschaften von Lehrern bei der Planung zukünftiger Strukturen zu berücksichtigen, weil damit der allergrößte Anteil an der Variation von Lehrerqualität überhaupt nicht erfasst wird. Anders formuliert: Lehrer unterscheiden sich zwar deutlich in ihrer Effektivität – aber Lehrer mit oder ohne bestimmte Eigenschaften oder Qualitäten unterscheiden sich nur wenig. Diesem Denk- und Handlungsansatz zufolge sollte man die Schulleitungen stärken und ihre Freiheit bei der Einstellung von geeigneten Lehrern ausweiten anstatt von Seiten der Schulverwaltung aus bestimmte Qualifikationen zur Voraussetzung zu machen.

Die einzige Studie, die sich explizit mit der Frage der Effektstärke beschäftigt hat, ist die von Rivkin/Hanushek/Kain (2001). Auf der Basis von Studien zu Schülerleistungen,

16 Seit den 1970er-Jahren haben sowohl Frauen als auch Angehörige rassistischer Minderheiten sicherlich bessere berufliche Möglichkeiten außerhalb des Lehrerberufs bekommen. In den 1970er-Jahren gab es relativ wenige neu zu besetzende Lehrerstellen, was auf demographische Faktoren auf der Schüler- wie Lehrerseite zurückzuführen ist. Dies hat natürlich auch die Haltung von Collegestudenten zum Lehrerberuf bzw. zum Lehramtsstudium beeinflusst. Vgl. dazu generell Warren (1989) und Sedlak/Schlossmann (1986). Zusätzlich gibt es eine kontinuierliche Debatte über die akademischen Qualifikationen (intellektuellen Leistungen) zukünftiger Lehrer (vgl. Gitomer/Lathaam/Ziomek 1999; Henke/Chen/Geis 2000).

die auf mehr als zwei Messzeitpunkten beruhen, haben sich die Autoren darum bemüht, den Einfluss solcher Schüler und Schulen statistisch auszuschließen, die sich im Verlaufe der Zeit nicht verändert haben. Diese Methode ermöglicht eine bessere Isolation der Effektstärke der Lehrervariable, sie sollte jedoch so lange als entwicklungsbedürftig betrachtet werden, bis andere Forscher diese Technik haben überprüfen können. Ein anderer vielversprechender Ansatz für eine bessere Erfassung der Effektstärke besteht in der Strategie der zufälligen Zuordnung. Eine neue Studie zu TFA geht so vor, dass Schüler nach Zufall TFA-Lehrern oder Nicht-TFA-Lehrern zugeordnet werden (Decker/Mayer 2002). Solche Designs sind nicht problemlos (vgl. Ehrenberg u.a. 2001), aber sie sind transparenter als quasi-experimentelle Anordnungen, und ihre Ergebnisse zur Effektstärke könnten die Diskussion über den Lehrerberuf voranbringen.

### Abkürzungsverzeichnis

ACT	Ein weit verbreiteter Zulassungstest zum College
HS&B	High School and Beyond
INTASC	Interstate New Teacher Support and Assessment Consortium
ITBS	Iowa Test of Basic Skills
LSAY	Longitudinal Survey of American Youth
NCATE	National Council for Accreditation of Teacher Education
NEAP	National Educational Advancement Programs
NELS	National Educational Longitudinal Survey
NTE	National Teacher Examination
TEAMS	Texas Educational Assessment of Minimum Skills
TECAT	Texas Examination of Current Administrators and Teachers
TFA	Teach for America

### Literatur

- Ballou, D./Podgursky, M. (1999): Reforming teacher preparation and licensing: What is the evidence? In: Teachers College Record 102, S. 5–27.
- Ballou, D./Podgursky, M. (2000): Reforming teacher preparation and licensing: Continuing the debate. In: Teachers College Record 102, H. 1, S. 5–27.
- Coleman, J.S. (1966): Equality of educational opportunity. Washington, DC: Office of Education, U.S. Department of Health, Education, and Welfare.
- Darling-Hammond, L. (1999a): Teacher quality and student achievement: A review of state policy evidence. Seattle: Center for the Study of Teaching and Policy, University of Washington.
- Darling-Hammond, L. (2000): Reforming teacher preparation and licensing: Debating the evidence. In: Teachers College Record 102, S. 28–56.
- Darling-Hammond, L. (2001): The research and rhetoric on teacher certification: A response to „Teacher certification reconsidered“. New York: National Commission on Teaching and America's Future.
- Darling-Hammond, L./Berry, B./Thoreson, A. (2001): Does teacher certification matter? Evaluating the evidence. In: Educational Evaluation and Policy Analysis 23, H. 1, S. 57–77.
- Decker, P./Mayer, D. (2002): Studying alternative certification using random assignment. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.

- Eberts, R.W./Stone, J.A. (1984): Unions and public schools: The effect of collective bargaining on American education. Lexington, MA: Heath.
- Ehrenberg, R.G./Brewer, D.J. (1994): Do school and teacher characteristics matter? Evidence from high school and beyond. In: *Economics of Education Review* 13, S. 1–17.
- Ehrenberg, R.G./Brewer, D.J./Gamoran, A./Wilms, J.D. (2001): Class size and student achievement. In: *Psychological science in the Public Interest* 2, H. 1, S. 1–30.
- Ferguson, R.F. (1991): Paying for public education: New evidence on how and why money matters. In: *Harvard Journal on Legislation* 28, S. 465–498.
- Ferguson, R.F. (1998): Can schools narrow the Black-White test score gap? In: Jencks, C./Phillips, M. (Hrsg.): *The Black-White test score gap*. Washington, DC: Brookings Institution, S. 318–374.
- Ferguson, R.F./Ladd, H.F. (1996): How and why money matters: An analysis of Alabama schools. In: Ladd, H.F. (Hrsg.): *Holding schools accountable: Performance-based reform in education*. Washington, DC: Brookings Institution, S. 265–298.
- Gitomer, D.H./Latham, A.S./Ziomek, R. (1999): *The academic quality of prospective teachers: The impact of admission and licensure testing*. Princeton, NJ: Educational Testing Service.
- Goertz, M.E./Pitcher, B. (1985): *The impact of NTE use by states on teacher selection*. Princeton, NJ: Educational Testing Service.
- Goldhaber, D.D./Brewer, D.J. (1997a): Evaluating the effect of teacher degree level on educational performance. In: Fowler, W.J. (Hrsg.): *Developments in school finance*. Washington, DC: National Center for Education Statistics, U.S. Department of Education, S. 197–210.
- Goldhaber, D.D./Brewer, D.J. (1997b): Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. In: *Journal of Human Resources* 32, S. 505–523.
- Goldhaber, D.D./Brewer, D.J. (2000): Does teacher certification matter? High school certification status and student achievement. In: *Educational Evaluation and Policy Analysis* 22, S. 129–146.
- Goldhaber, D.D./Brewer, D.J. (2001): Evaluating the evidence on teacher certification: A rejoinder. In: *Educational Evaluation and Policy Analysis* 23, S. 79–86.
- Graham, P.A. (1987): Black teachers: A drastically scarce resource. In: *Phi Delta Kappan* 68, S. 598–605.
- Greenwald, R./Hedges, L.V./Laine, R.D. (1996): The effect of school resources on student achievement. In: *Review of Educational Research* 66, S. 361–396.
- Haney, W./Madaus, G./Kreitzer, A. (1987): *Charms talismanic: Testing teachers for the improvement for American education*. In: *Review of Research in Education*, Vol. 14., Washington, DC: American Educational Research Association.
- Hanushek, E.A. (1971): Teacher characteristics and gains in student achievement: Estimation using micro-data. In: *American Economic Review* 61, S. 280–288.
- Hanushek, E.A. (1992): The trade-off between child quantity and quality. In: *Journal of Political Economy* 100, S. 85–117.
- Hanushek, E.A. (1997): Assessing the effects of school resources on student performance: An update. In: *Educational Evaluation and Policy Analysis* 19, S. 141–164.
- Hanushek, E.A./Gomes-Neto, J.B./Harbison, R.W. (1996): Efficiency-enhancing investments in school quality. In: Birdsall, N./Sabot, R.H. (Hrsg.): *Opportunity foregone: Education in Brazil*. Washington, DC: Inter-American Development Bank/Johns Hopkins University Press, S. 385–424.
- Harnisch, D.L. (1987): Characteristics associated with effective public high schools. In: *Journal of Educational Research* 80, S.233–241.
- Henke, R.R./Chen, X./Geis, S. (2000): *Progress through the teacher pipeline: 1992–93 college graduates and elementary/secondary school teaching as of 1997*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Hussar, W.J. (1999): *Predicting the need for newly hired teachers in the United States to 2008–09*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Ingersoll, R.M. (1996): *Out-of-field teaching and educational equity*. Washington, DC: U.S. Department of Education.

- Jordan, H. R./Mendro, R./Weerasinghe, D. (1997): Teacher effects on longitudinal student achievement. Paper presented at the annual meeting of the Center for Research on Educational Accountability and Teacher Education, Indianapolis, IN.
- Kain, J.F./Singleton, K. (1996, May-June): Equality of educational opportunity revisited. In: *New England Economic Review*, S. 87–111.
- Kiesling, H.J. (1984): Assignment practices and the relationship of instructional time to the reading performance of elementary school children. In: *Economics of Education Review* 3, S. 341–350.
- Link, C.R./Ratledge, E.C. (1979): Student perceptions, I.Q. and achievement. In: *Journal of Human Resources* 14, S. 98–111.
- Ludlow, L. (2001): Teacher test accountability: From Alabama to Massachusetts. In: *Education Policy Analysis Archives* 9(6).
- Mitchell, K.J./Robinson, D.Z./Plake, B.S./Knowles, K.T. (2001): *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.
- Monk, D.H. (1994): Subject area preparation of secondary math and science teachers and student achievement. In: *Economics of Education Review* 13, S. 125–145.
- Monk, D.H./King, J. (1994): Multilevel teacher resource effects on pupil performance in secondary mathematics and science: The case of teacher subject-matter preparation. In: Ehrenberg, R. (Hrsg.): *Contemporary policy issues: Choices and consequences in education*. Ithaca, NY: ILR, S. 29–58.
- Murnane, R.J. (1975): *The impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger.
- Murnane, R.J./Phillips, B.R. (1981): What do effective teachers of inner-city children have in common? In: *Social Science Research* 10, S. 83–100.
- Pecheone, R./Rogers, J./Moirs, K. (2001): A beginning teacher survey study: A policy-practice perspective. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Raymond, M./Fletcher, S.H./Luque, J. (2001): *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford, CA: Center for Research on Education Outcomes.
- Rivkin, S.G./Hanushek, E.A./Kain, J.F. (2001): *Teacher, schools, and academic achievement*. Amherst, MA: Amherst College.
- Rowan, B. (2002): What large scale survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. Ann Arbor: University of Michigan.
- Rowan, B./Chiang, F.-S./Miller, R.J. (1997): Using research on employees' performance to study the effects of teachers on students' achievement. In: *Sociology of Education* 70, S. 256–284.
- Sanders, W.L./Rivers, J.C. (1996): *Cumulative and residual effects of teachers on future academic achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.
- Sedlak, M./Schlossman, S. (1986): *Who will teach? Historical perspectives on the changing appeal of teaching as a profession*. Santa Monica, CA: Rand.
- Smith, G.P./Miller, M.C./Joy, J. (1988): A case study of the impact of performance-based testing on the supply of minority teachers. In: *Journal of Teacher Education* 39, H. 4, S. 45–53.
- Strauss, R.P./Sawyer, E.A. (1986): Some new evidence on teacher and student competencies. In: *Economics of Education Review* 5, S. 41–48.
- Summers, A.A./Wolfe, B.L. (1975a): *Equality of educational opportunity quantified: A production function approach*. Philadelphia: Federal Reserve Bank of Philadelphia, Department of Research.
- Summers, A.A./Wolfe, B.L. (1975b): *Which school resources help learning? Efficiency and equity in Philadelphia public schools*. Philadelphia: Federal Reserve Bank of Philadelphia.
- Summers, A.A./Wolfe, B.L. (1977): Do schools make a difference? In: *American Economic Review* 67, S. 639–652.

- U.S. Department of Education (2002): Meeting the highly qualified teachers challenge. Washington, DC.
- Walsh, K. (2001): Teacher certification reconsidered: Stumbling for quality. Baltimore: Abell Foundation.
- Warren, D. (1989): American teachers: Histories of a profession at work. New York: Macmillan.
- Wayne, A. (2002): Teacher inequality: New evidence on disparities in teachers' academic skills. In: Education Policy Analysis Archives 10(30).
- Wayne, A.J. (2000a): Federal policies to improve teacher quality for low-income students. Unveröffentlichte Doktorarbeit, University of Maryland, College Park.
- Wayne, A.J. (2000b): Teacher supply and demand: Surprises from primary research. In: Education Policy Analysis Archives 8(47).
- Wilson, S.M./Floden, R.E./Ferrini-Mundy, J. (2001): Teacher preparation research: Current knowledge, gaps, and recommendations. Seattle, WA: Center for the Study of Teaching and Policy.
- Wright, S.P./Horn, S.P./Sanders, W.L. (1997): Teacher and classroom context effects on student achievement: Implications for teacher evaluation. In: Journal of Personnel Evaluation in Education 11, S. 57–67.
- Youngs, P./Odden, A./Porter, A.C. (2001): State leadership in teacher licensure. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Youngs, P./Odden, A./Porter, A.C. (2003): State policy related to teacher licensure. In: Educational policy 17, S. 217–236.

*Anschrift der Autoren:*

Andrew Wayne, SRI International, 1100 Wilson Blvd., Suite 2800, Arlington, VA 22209-3915,  
E-Mail: [andrew.wayne@sri.com](mailto:andrew.wayne@sri.com).

Peter Young, Research Associate, Stanford University, School of Education, CERAS Building, Room 109-O, Stanford CA 94305, E-Mail: [pyoungs@stanford.edu](mailto:pyoungs@stanford.edu).