



Universität Potsdam

Diether Hopf

Anreize zu erfolgreichem Lehren in der
Schule : Forschungsstand und
methodologische Überlegungen ;
Langfassung

first published in:
www.dds-home.de/v2010/aktuelle_hefte.htm
(Online-Ressource nicht mehr verfügbar)

Postprint published at the Institutional Repository of the Potsdam University:
In: Postprints der Universität Potsdam Humanwissenschaftliche Reihe ; 231
<http://opus.kobv.de/ubp/volltexte/2011/5382/>
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-53829>

Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 231

Hopf, Diether: Erfolgreiches Lehren in der Schule.
Anmerkungen zum Stand der Forschung.
In: DDS – Die Deutsche Schule 102 (3/2010), S. 268-279.

Diether Hopf

Anreize zu erfolgreichem Lehren in der Schule

Forschungsstand und methodologische Überlegungen

– Langfassung –¹

Inhaltsverzeichnis

1. Einleitung	2
2. Forschung über Unterricht und Lehrerverhalten	6
2.1 Forschungsmethodisch liberale Überblicke	6
2.2 Forschungsmethodischer Exkurs	8
2.3 Forschungsmethodisch anspruchsvolle Überblicke	12
3. Forschung über Anreizsysteme für Lehrer/innen	17
4. Forschung über Lehrertraining und Lehrerbildung	20
5. Schlussbemerkungen	28
6. Literatur	36

¹ Stand: 16.05.2010.

1. Einleitung

Wenn man guten Unterricht und erfolgreiches Lehren verstärken und verbreiten möchte, muss man verlässliche Kenntnisse besitzen

1. über die Eigenschaften und Formen guten Unterrichts und erfolgreichen Lehrens, welche sich auf die Lernergebnisse der Schüler/innen auswirken, und
2. über die Möglichkeiten, erfolgreiches Lehrerhandeln und -verhalten zu vermitteln und zu verstärken.

Von Interesse ist zugleich die Betrachtung der Forschungen über den Gegenpol, nämlich über ineffektive Lehrer/innen und kontraproduktiven Unterricht und die Möglichkeiten negativer Verstärkung des entsprechenden Verhaltens.

Verlässliche Kenntnisse hierüber bilden auch die Voraussetzung für Schulinspektionen und Beurteilungen von Lehrer/inne/n, insbesondere wenn diese Konsequenzen positiver oder negativer Art nach sich ziehen. Dennoch beruhen in Deutschland z.B. so genannte Schul-Inspektionen² oder Beurteilungen bei der Einstellung oder Beförderung von Lehrer/inne/n in der Regel nicht oder nur zu einem kleinen Teil auf belastbaren Forschungsbefunden, sondern sie werden auf der Grundlage von Erfahrungen und *Common Sense* angefertigt, wobei nicht selten eine große Selbstsicherheit über das Wissen von Ursache und Wirkung zu bemerken ist oder die Lernfortschritte der Schüler/innen, um derentwillen Schule ja veranstaltet wird, aus dem Blickfeld geraten. Der zuletzt genannte Eindruck drängt sich ebenfalls bei der Betrachtung der Diskussion um die Lehrerbildung auf, wo ständig neue Modelle entwickelt werden, in denen oftmals die Lehrerzufriedenheit und die Verhinderung von Burn-out wichtiger zu sein scheinen als die Förderung und die Lernfortschritte der Schüler/innen, oder wo unterstellt wird, diese seien selbstverständlich mit der Lehrerzufriedenheit gegeben.

Zum 1. Punkt:

Es existiert eine Fülle von literarischen Zeugnissen und Erfahrungsberichten, in denen über guten Unterricht berichtet wird oder über eindrucksvolle Lehrer/innen, an die sich die Berichterstatter/innen oder Autor/inn/en gern erinnern und durch die sie sich erkennbar gefördert gefühlt haben. So gibt es, um ein Beispiel zu nennen, eine seit Jahrzehnten laufende Serie in dem britischen Wochenblatt „The Times Educational Supplement“ mit dem Titel „My Best Teacher“. Hier berichten Erwachsene aus allen möglichen Berufsfeldern rückblickend über ihre Schulerfahrungen.

Bei der – nicht nur amüsanten, sondern auch interessanten – Inspektion dieser Berichte aus unterschiedlichen Jahrzehnten stellt man schnell fest, dass

² Schul-TÜVs gibt es in mehreren Bundesländern. Erfreulicherweise sind nicht nur die Ergebnisse, sondern auch viele der dabei verwendeten Instrumente zugänglich, so dass eine Diskussion darüber möglich wird. Ich komme später auf die Schul-TÜVs zurück.

es außerordentlich schwierig wäre, auf ihrer Grundlage eine brauchbare Definition dessen zu geben, was einen guten Lehrer oder eine gute Lehrerin ausmacht; zumindest aus der Sicht der betroffenen, ehemaligen Schüler/innen und beurteilt nach vielen Jahren außerschulischer Lebenserfahrung. Ein Blick auf einige der Äußerungen³ zeigt dies deutlich: Zum einen gibt es eine riesige Variation von Merkmalen, die den ehemaligen Schüler/innen im Gedächtnis geblieben sind, und man könnte sich auf der Grundlage dieser Berichte gewiss nicht darauf einigen, was ein guter Lehrer oder eine gute Lehrerin ist. Zweitens zeigt sich eine ausgeprägte Wechselwirkung zwischen Eigenschaften, Verhaltensweisen und Werten der ehemaligen Schüler/innen und denen der Lehrer/innen – ein Thema, das übrigens in der bisher vorliegenden Forschungsliteratur zu diesem Thema nur sporadisch angesprochen wird. Und drittens wäre es schwierig, sich über die Wirkungen der beschriebenen Lehrer/innen und ihres Unterrichts zu einigen: sind die ihnen nachgesagten und nachgerühmten Effekte konsensfähig und generalisierbar? Sind sie überhaupt wissenschaftlich überprüfbar?

Die Folge ist, dass man sich auf der Grundlage einer solchen Datenbasis nur schwer vorstellen kann, Anreizsysteme und Anerkennungsformen zu entwickeln, die der Vielfalt des geschilderten positiven Verhaltens der Lehrer/innen gerecht würden. Selbst bei einer einzelnen Lehrkraft ist im übrigen mit einer erheblichen Varianz erfolgreichen Verhaltens zu rechnen; man denke nur an Frank McCourts eindrucksvolles Roman-Dokument „Teacher Man“.

Erwähnungen und Beschreibungen schlechter und traumatisch erinnerten Lehrer/innen sind in dem Corpus des *Times Educational Supplement* naturgemäß selten, es gibt sie immerhin: Lehrer/innen beispielsweise, die ihre Schüler/innen verängstigt oder Talente zugeschüttet haben.⁴ Sicher wäre es überschaubarer, einen Überblick über negative und zu sanktionierende Lehrerverhaltensweisen als über den „guten Lehrer“ oder die „gute Lehrerin“ zu gewinnen, so wie Verhaltensanleitungen ja meistens nicht in Form von Geboten, sondern von Verboten anzutreffen sind.⁵ Eine Durchsicht literarischer und autobiographischer Zeugnisse würde wahrscheinlich ergeben, dass weniger Texte über negatives Lehrerverhalten und schlechten Unterricht als

³ Beispiele:

„ ... was teaching the subject, not the pupil, and had a passion for his subject ... he couldnt stand anyone chatting if work was going on ... there was a lot of discipline ... and strictness ... within the structure there was room for experimenting“ (24.3.2000).

„ ... he taught me humility and a lot about how to relate to people“ (3.3.2000).

„ ... wonderful teacher because he didn't do a thing for my education“ (10.12.2004).

„ ... he wasn't cruel, but he could be acerbic, sarcastic or mocking. I am sure some people didn't like his manner, but it always used to make me smile“ (15.9.2000).

⁴ Vgl. z.B.: „School for me was unrelenting torture ... recurring nightmare ... School did teach me what a terminal moraine is ... I also know the French word for turkey – but not a lot more. What I learned most was how to cope with an institution I loathed ...“ (11.7.1997). Oder: „I don't remember any of my school teachers with affection. What I remember are the threats, the punishments, the slaps around the face, which is sad really“ (3.10.1997).

⁵ Z.B. sind von den biblischen zehn „Geboten“ acht Verbote.

über die positive Seite existieren. Auf die Bedeutung wechselnder gesellschaftlicher Kontexte für die Perspektiven der Schulerinnerungen ist Konrad Jarausch (2003) auf interessante Weise eingegangen.

Neben den literarischen oder quasi-literarischen Quellen gibt es, und damit befasst sich der vorliegende Aufsatz, nach Tausenden zählende wissenschaftliche Untersuchungen zum Thema, die sich über viele Jahrzehnte international angesammelt haben. Ob sich aus diesem Corpus generalisierbare Gewissheiten ableiten oder eher nur Anregungen für weiterführende Forschung gewinnen lassen, wird uns später beschäftigen. Guter Unterricht ist in diesen Studien auf unterschiedliche Weise definiert, manchmal begnügt man sich nicht mit einer vorab gesetzten Norm der Unterrichtsgestaltung, sondern spricht erst dann von „gutem Unterricht“, wenn sich bei den Schüler/inne/n erwünschte Veränderungen ergeben haben. Arbeiten über negativ sich auswirkendes Lehrerverhalten sind in der Forschung kaum⁶ zu finden.

Zum 2. Punkt:

Forschungen über die Möglichkeiten, erfolgreiches Lehrerhandeln und -verhalten zu verstärken, gehen einen Schritt weiter, insofern sie die Kenntnis, was guten Unterricht und erfolgreiches Lehren ausmacht, zur Voraussetzung haben. Daher ist es nicht verwunderlich, dass sich hierzu weniger Studien entdecken lassen, jedenfalls wenn man sich auf diejenigen Arbeiten beschränkt, bei denen Unterricht und Lehrerhandeln mit messbarer Förderung der Schüler/innen verbunden sind. Ähnliches gilt für Studien über Lehreraus- oder -fortbildung, die zu einer erhöhten Wirksamkeit des Lehrens führen; im deutschsprachigen Raum herrscht ein akuter Mangel an solchen Studien.

Fehlende Kenntnisse zu diesen Fragen haben zwar bisher niemanden davon abgehalten, beispielsweise Zensuren in Lehramts- oder Assessordprüfungen zu erteilen (sogar mit Zensurenbruchteilen), so als wüsste man genau, wer eine gute Lehrkraft sein wird, und als verfüge man über reliable und valide Messinstrumente. Oder es wurden und werden immer aufs Neue Lehrerbildungs-Modelle und Reformvorschläge entwickelt, ohne dass die Frage beantwortet – oder auch nur empirisch angegangen – würde, ob das Neue besser sei als das Alte.⁷ Immerhin hat man sich bei den in den letzten Jahren in mehreren Bundesländern entworfenen und praktizierten „Schul-TÜVs“, soweit ich sehen kann, bisher zurückgehalten, spürbare Sanktionen zu verhän-

⁶ Ein Beispiel sind die leistungsmindernden Effekte von *Open Education*, die Kulik/Kulik (1989) in ihrer Metaanalyse identifiziert haben; vgl. auch Bennett (1979).

⁷ Man denke nur an das medienwirksam präsentierte sog. „Potsdamer Modell der Lehrerbildung“ und seine nüchterne Kritik durch die Lemmermöhle-Kommission (vgl. http://www.uni-potsdam.de/pressmitt/2008/Gutachten_Lehrerbildung.pdf; Zugriffsdatum: 18.6.2008). Umgekehrt könnte niemand belegen, dass z.B. ein unmittelbar nach dem 2. Weltkrieg in einem beliebigen Bundesland verwendetes Modell der Lehrerbildung untauglich war und heutzutage nicht ebenso gut verwendet werden könnte.

gen bzw. konkrete positive Verstärkungen für als erfolgreich erachtetes Handeln zu setzen.⁸

Im Vorgriff auf einen späteren Abschnitt sei hier eingefügt, dass es nicht zwingend erforderlich ist, vorab zu klären, welches konkrete Lehrerverhalten sich wie auf die Schüler/innen auswirkt, sondern auch methodologisch andere Herangehensweisen möglich (und erfolgreich) sein können, wenn die Frage beantwortet werden soll, durch welche Anreize sich erfolgreiches Lehren verstärken lässt. So kann man z.B. den Lehrer/inne/n Anreize anbieten, die sie unter der Bedingung erhalten, dass ihre Schüler/innen bessere Lernerfolge erzielen, und es den Lehrer/inne/n selbst überlassen, wie sie ihr Expertenwissen zur Förderung der Lernfortschritte ihrer Schüler/innen umsetzen. Ein solches „*black-box*“-Modell ist in der Tat mit Erfolg eingesetzt worden, worüber an späterer Stelle zu berichten sein wird.

Erfolgreiches Lehren ist im vorliegenden Aufsatz dadurch charakterisiert, dass es bei den Schüler/inne/n zu erwünschten Lernfortschritten führt: zu höheren Schulleistungen beispielsweise oder zu besserem sozialem Verhalten oder zu höherer Motivation. Nur solche Lehrer/innen werden als erfolgreich definiert, deren professionelle Tätigkeit positive Auswirkungen auf die Schüler/innen hat. Lehrerverhalten, das zwar Herangehensweisen entspricht, die in der allgemeinen pädagogischen Diskussion für erstrebenswert gehalten werden, dessen Auswirkungen auf die Schüler/innen aber nicht bekannt sind, bleiben außerhalb der Betrachtung. Die meisten der hier beigezogenen Untersuchungen beziehen sich im übrigen auf Schüler/innen der Sekundarstufe des Schulwesens.

Es ist immer wieder darauf hingewiesen worden – zuletzt von Seidel und Shavelson (2007) –, von wie großer Bedeutung die Auswahl eines geeigne-

⁸ Beispielsweise verwendet der im Juni 2008 veröffentlichte Bericht über die Qualität der Berliner Schulen zurückhaltende Berichts- und Urteilsformen. Als Sanktion ist dort lediglich die Ankundigung eines erneuten Schul-TÜVs nach nur zwei Jahren vorgesehen (vgl. Senatsverwaltung 2008, S. 14); im Übrigen werden unterstützende Maßnahmen angeboten. An mehreren Stellen wird dabei allerdings, wie es weithin üblich ist, nicht vorhandenes Wissen unterstellt, etwa wenn von den „Stärken des Unterrichts“ oder dem „Entwicklungsbedarf des Unterrichts“ (vgl. S. 24 u.a.) die Rede ist, so als wüsste man z.B., dass innere Differenzierung besonders förderlich für die Lernfortschritte der Schüler/innen sei. Ein bei den Schulinspektionen schwer lösbares Problem besteht darin, dass es in Deutschland nur sehr wenige Schülerdaten gibt, zu denen man Schulverhältnisse, Lehrerverhalten, Unterrichtsformen etc. in Beziehung setzen könnte. In dem jüngsten Berliner Bericht sind als Indikatoren auf der Schülerenebene dementsprechend nur wenige angegeben, etwa Verspätungen (vgl. S. 19), Klassenwiederholungen und Fehlquoten (vgl. S. 11), Schullaufbahn- und Schulleistungsdaten, die schulweit kommuniziert werden (vgl. S. 30). Gleichwohl werden Indikatoren, deren Zusammenhang mit den Lernfortschritten der Schüler/innen völlig unklar sind, bewertet, etwa wenn von „erfreulicher“ PC-Ausstattung die Rede ist (vgl. S. 22) oder wenn der vorgefundene Mangel an innerer Differenzierung beklagt wird (vgl. S. 24). Der Mangel an Schülerdaten wird besonders fühlbar, wenn sich kein Zusammenhang aufzeigen lässt zu den Lernergebnissen der Schüler/innen an den (wenigen) Schulen, die dort als defizitär bezeichnet werden. Vgl. auch Senatsverwaltung 2006a und 2006b. – Eine neue Übersicht über externe Evaluationen in den 16 Bundesländern findet sich bei Ditton (2008).

ten theoretischen Modells für die Aufarbeitung des Forschungsstandes sind. In dem vorliegenden Aufsatz nehme ich die Debatte über solche Modelle nicht erneut auf, sondern verweise auf die Vielzahl bereits existierender Modelle, von denen man bei der Strukturierung des Feldes Anregungen erhält.⁹

2. Forschung über Unterricht und Lehrerverhalten

Bei der Beantwortung der Frage, über welche Kenntnisse zum Zusammenhang von Lehren und Lernen im schulischen Kontext wir verfügen, verfolge ich zwei verschiedene Stränge, die sich methodologisch voneinander unterscheiden: Einerseits werden Studien bzw. Überblicke, Metaanalysen oder Zusammenfassungen von Studien diskutiert, in denen nicht nach methodischen Kriterien unterschieden wird, bei denen also korrelative Untersuchungen, Surveys oder Fallstudien gemeinsam mit z.B. experimentellen Studien oder Längsschnitten aufgelistet werden (Abschnitt 2.1). Andererseits werden Zusammenfassungen von Untersuchungen betrachtet, die sich auf solche Studien beschränken, welche Ursache-Wirkungs-Zusammenhänge aufzudecken erlauben oder nahe legen (experimentelle, quasi-experimentelle, longitudinale Untersuchungen) (Abschnitt 2.3). Zwischen diesen beiden Abschnitten steht ein forschungsmethodischer Exkurs (Abschnitt 2.2).

2.1 Forschungsmethodisch liberale Überblicke

Bereits in den 60er-Jahren des vorigen Jahrhunderts sind Überblicke über den Stand der damals schon reichhaltigen Unterrichtsforschung vorgelegt worden, in denen zunächst ein recht positives Bild unseres Wissens über Unterricht bzw. Lehrerhandeln gezeichnet wurde (vgl. z.B. Biddle/Elena 1964). Allerdings bezog eine Vielzahl der Studien nicht die Lernergebnisse der Schüler/innen mit ein, sondern beschränkte sich auf die Betrachtung und Klassifizierung des Lehrerverhaltens.

In den 70er-Jahren kann man dann eine eher skeptische Einschätzung beobachten: Schule und Unterricht hätten nur eine geringe Auswirkung auf die Lernfortschritte oder die Verringerung der Ungleichheiten zwischen den Schüler/inne/n. Beispielsweise kommt Jencks (1972) nach Durchsicht der ihm vorliegenden (überwiegend korrelativen) Forschung zu Schlussfolgerungen wie: „Wenn wir die Qualität der Grundschulen gleich machen könnten, würde die kognitive Ungleichheit [zwischen den Schüler/inne/n; D.H.] um 3 Prozent oder weniger verringert.“ Und: „Wenn wir die Qualität der Sekundarschulen gleich machen könnten, würde die kognitive Ungleichheit

⁹ Zur Diskussion theoretischer Modelle sei hier neben der Metaanalyse von Seidel und Shavelson auf die Arbeiten u.a. von Haertel (1983), Helmke/Weinert (1997), Baumert/Kunter (2006) sowie auf die den PISA-Analysen zugrunde liegenden Konzepte hingewiesen, vgl. z.B. Baumert u.a. 2003, S. 18.

um 1 Prozent oder weniger verringert“. Dagegen: „Wenn wir Jedermanns totale Umwelt gleich machen könnten, würde die Testergebnis-Ungleichheit um 25-40 Prozent fallen.“ Good, Biddle und Brophy (1975, S. 3) fragen: „Machen Schulen oder Lehrer eigentlich einen Unterschied aus? Es gibt keine sichere Antwort“. ¹⁰ Zur gleichen Zeit gibt es aber auch weiterhin optimistische Einschätzungen der Unterrichtsforschung, beispielsweise in den Teams um Gage und Berliner¹¹, beides einflussreiche Forscher, die unermüdlich eine positive Einschätzung unseres Wissensstandes und der Belastbarkeit unseres Wissens formuliert haben.

Im Unterschied zu den skeptischen Einschätzungen der 70er-Jahre geht die Schulforschung 30 Jahre später auf der Basis des inzwischen erreichten Forschungsstandes davon aus, dass Schule und Unterricht von erheblicher Bedeutung für Lernen und Entwicklung sind. Darauf haben bereits international vergleichende Surveys hingewiesen, bei denen sich zeigt, dass Entwicklungsländer mit kurzen Schulzeiten und ungenügend befolgter Schulpflicht – z.B. bei jenen 40% der Länder, in denen weniger als 40% der Altersgruppe Sekundarschulen besuchen¹² – in den Schulleistungen weit hinter Ländern zurück liegen, bei denen alle Schüler/innen mindestens acht oder neun Jahre die Schule besuchen, dass also die Höhe der Lernergebnisse allein schon mit dem zeitlichen Umfang des Unterrichtsangebots kovariert. Dies knüpft an die frühen Untersuchungen von Wiley (1974) an, der innerhalb den USA den Zusammenhang zwischen der schieren Unterrichtszeit in Stunden pro Jahr und den Lernergebnissen belegt hat.

Die neueste Übersicht über den Stand der Unterrichtsforschung in der Tradition von Biddle und Elena (1964) findet sich bei Anderson (2004), der seinen für die UNESCO verfassten Forschungsbericht von 1991 aktualisiert hat und den Stand unseres Wissens über die Wirksamkeit von Lehrer/inne/n aus seiner Sicht vorstellt. Anderson zeichnet ein positives Bild unserer Kenntnislage und formuliert zahlreiche konkrete Hinweise, die unbedingt befolgt werden müssten, wenn man gute von schlechten Lehrer/inne/n unterscheiden oder effektive Lehrerbildung betreiben wolle. Folgerichtig gibt er auch am Ende der meisten Kapitel ausführliche und direkte Empfehlungen, wie man die Effektivität des Unterrichts und der Lehrer/innen erhöhen kann, und zwar mit Blick sowohl auf die konkrete Schul- und Unterrichtspraxis wie auf die Schuladministration und Schulpolitik.

¹⁰ Man muss hier aber u.a. zwischen Lernfortschritten per se und der Einebnung von kognitiver Ungleichheit der Schüler/innen unterscheiden (vgl. Weinert 1996, S. 76): „... dass Schulen zwar das Lernen im Sinne des Erwerbs von Wissen und Können bei allen Schülern massiv beeinflussen, dass sie aber dadurch die individuellen Unterschiede der intellektuellen Fähigkeiten zwischen ihnen nicht bedeutsam verändern“.

¹¹ Vgl. z.B. deren Pädagogische Psychologie von 1975, deutsch 1979.

¹² Vgl. Anderson 2004, S. 19.

2.2 Forschungsmethodischer Exkurs

Es stellt sich die Frage, wie belastbar die in den erwähnten Texten referierten Forschungsergebnisse sind und ob man aus ihnen Empfehlungen für die Praxis ableiten darf.

Bei der Verwendung der erwähnten Übersichten und Metaanalysen aus den unterschiedlichen Zeitperioden steht man fast immer vor dem Dilemma, dass die darin verarbeiteten Originalarbeiten nicht nach den verwendeten Forschungsmethoden getrennt werden, sondern dass experimentelle Untersuchungen, Surveys, Einzelfallstudien, Längsschnitte etc. vermischt aufgeführt und meist auch gleich gewichtet werden. Eine Differenzierung nach Forschungsmethode wäre aber erforderlich, um die Tragfähigkeit und Generalisierbarkeit der Ergebnisse abschätzen zu können. Belastbare und in angebbare Weise generalisierbare Befunde erhält man vor allem aus experimentellen Untersuchungen; korrelative Studien dagegen beschreiben Zusammenhänge, lassen aber keine Schlüsse über Ursache und Wirkung zu. Dies gilt auch für die methodisch oft raffiniert verfahrenen Datenanalysen, die bei Schneider u.a. (2007) diskutiert und gern von der Bildungsökonomie verwendet werden. Sicherlich haben auch andere als (quasi-)experimentelle oder longitudinale Studien ihre wichtige Funktion im Fortgang der Forschung; Erkenntnisse über Ursache und Wirkung sind aus ihnen aber nicht zu gewinnen.

Anhand von einigen wenigen, einflussreichen Texten, die den Stand der Unterrichtsforschung widerspiegeln, soll die angesprochene Problematik der forschungsmethodischen Indifferenz der Autor/inn/en bei der Auswahl der Arbeiten, auf die sie sich beziehen, verdeutlicht werden.

a) In der oben genannten Übersicht über den Forschungsstand zum Thema Teacher Effectiveness (Anderson 2004) wird an keiner Stelle die Frage aufgeworfen, wie man mit unterschiedlich qualifizierten Arbeiten und deren Ergebnissen umgehen sollte; überhaupt gibt es in dem ganzen Text nur spärliche Hinweise auf forschungsmethodische Fragen.¹³ Dabei ist das Problem mit Händen zu greifen: Die überwiegende Zahl der von Anderson in seinen Überblick einbezogenen Untersuchungen ist schon auf den ersten Blick erkennbar nicht dazu geeignet, Lehrerhandeln als *Ursache* für Schülerleistung zu identifizieren, da sie forschungsmethodisch nicht so angelegt sind, dass aus ihnen kausale Schlussfolgerungen gezogen werden können. Dennoch wird Kausalität unterstellt: „A positive classroom climate is necessary in order to bring out the best in students“, heißt es z.B. (S. 50); oder: „... nine factors that define an effective classroom climate. These factors are clarity, fairness, interest, order, participation, physical environment, safety, standards and support“ (S. 53); oder: „The presentations made by teachers should be clear and understandable“ (S. 93). Die als Belege von ihm zitierten, meist korrelativen Arbeiten sind in ihren Ergebnissen zwar oft plausibel

¹³ Vgl. z.B. S. 56.

und ähneln den Empfehlungen von Seminarleiter/inne/n in der Referendar-
ausbildung, in die in der Regel viel Praxiserfahrung eingegangen ist, basie-
ren aber nicht auf generalisierbarer und Kausalität prüfender Forschung.

Entsprechendes gilt für die von Anderson gegebenen unzähligen Empfeh-
lungen und Vorschriften zur Verbesserung der Lehrer-Effektivität sowie für
seine aufsummierten Übersichtstabellen, in denen erwünschtes Lehrerver-
halten zusammengefasst wird.¹⁴ Sie lesen sich durchaus plausibel, sind aber
nicht durch Studien abgestützt, die es erlauben, solche generalisierenden
und scheinbar auf sicheren Erkenntnissen fußenden Aussagen zu machen. In
der Tat ist sogar fundierte methodische Kritik an einer Reihe von Studien
geübt worden, auf die sich Anderson an zentraler Stelle beruft.¹⁵ Insgesamt
wird man den Praxisempfehlungen Andersons und seinen optimistischen
Urteilen über die Forschungslage mit größter Skepsis begegnen müssen.

b) Gage und Berliner (1979) stellen in ihrer bekannten, praxisbezogenen
Pädagogischen Psychologie zwar die methodischen Grundlagen der For-
schung ausführlich dar,¹⁶ unterscheiden im Verlauf des Buches dann aber
nicht mehr nach forschungsmethodischer Qualität der referierten Studien, so
dass einerseits kunstgerechte psychologische Experimente (z.B. zum operan-
ten Konditionieren), andererseits korrelative und oft recht anspruchsvolle
Studien berichtet und ohne Unterschied auf schulische Lehr- und Lernsitua-
tionen angewandt und verallgemeinert werden. Bei näherem Zusehen lässt
sich feststellen, dass es sich bei den einbezogenen Arbeiten sogar überwie-
gend um korrelative Untersuchungen handelt, aus denen sich nicht ableiten
lässt, ob das jeweils untersuchte Lehrer- oder Schülerverhalten in kausaler
Beziehung zu dem berichteten Effekt steht. Unberücksichtigt bleibt darüber
hinaus in der Regel die Generalisierbarkeit der Befunde, die ein wie auch
immer geartetes Stichprobenverfahren voraussetzen würde.

Zu ähnlichen Ergebnissen wie bei der Durchsicht der Bücher von Anderson
oder Gage und Berliner kommt man bei der Analyse auch zahlreicher ander-
er, den jeweiligen Forschungsstand abdeckender Texte zur Unterrichtsfors-
chung und zum Lehrerhandeln. Selbst Weinert (1996) macht uns zwar mit
den methodischen Qualitätsunterschieden bekannt, findet aber in ambivalen-
ter Weise für beide Positionen – einerseits die milde, alle Untersuchungser-
gebnisse großzügig verwertende Position, andererseits die methodisch

¹⁴ Vgl. z.B. S. 54f.

¹⁵ So wird z.B. bei Anderson (2004, S. 20) als einer der besonders wichtigen Belege für
die Effektivität von Lehrer/inne/n die Studie von Sanders und Rivers (1996) zitiert. Die-
se ist aber wegen methodischer Fehler überzeugend widerlegt (vgl. Kupermintz 2002);
letzterer wird von Anderson nicht erwähnt, obwohl er die Literatur bis 2003 berücksich-
tigt hat. Im Übrigen werden bei Anderson nur eine verschwindend geringe Anzahl von
Studien einer methodischen Kritik unterzogen, die meisten werden ungeprüft über-
nommen. Es entbehrt nicht einer gewissen Tragik, dass die enorme Mühe, die in der
Aufarbeitung der Literatur durch Anderson steckt, wegen der fehlenden Sensibilität ge-
genüber den verwendeten Forschungsmethoden größtenteils als vergeblich zu bezeich-
nen ist.

¹⁶ Band 1, Kap. 2.

strenge, evidenz-basierte – Worte der Akzeptanz und bezeichnet sie als Kontroversen, hinter denen sich „Differenzen eines aristotelischen oder galileischen Verständnisses der wissenschaftlichen Psychologie“ (S. 7) verbergen könnten. Es ist dann nicht verwunderlich, dass er an späterer Stelle (vgl. S. 39f.) zwar acht Instruktionsprinzipien auflistet, diesen jedoch sofort die damit verbundenen Probleme und Kenntnislücken an die Seite stellt und fortfährt: „Es ist offensichtlich, dass die postulierten und in der einschlägigen Literatur vielfach propagierten Instruktionsprinzipien mehr Probleme aufwerfen als dass sie spezifische Handlungsanweisungen liefern. Diese Situation wird sich auf der Basis der gegenwärtig bevorzugten Forschungsstrategien auch künftig kaum verändern. Instruktion bleibt vermutlich die wissenschaftlich zwar fundierte, aber nur durch gesunden Menschenverstand, praktische Vernunft und plausible Erfahrungsgeneralisierung nutzbare Anwendung von Prinzipien, die eigentlich keine Prinzipien sind“ (S. 41).

So ambivalent Weinert auch argumentiert, ist doch zu erkennen, dass er dem seinerzeit vorhandenen Wissen über Unterricht und Lehrerhandeln skeptisch gegenübersteht. Dies zeigt sich bereits in seiner Arbeit von 1989, wo er von der riesigen (fünfstelligen) Zahl vorhandener korrelativer Untersuchungsergebnisse zu dieser Thematik spricht, denen man aber kaum brauchbare Informationen für die Verbesserung des Lehrens und Lernens in Schule und Unterricht entnehmen könne.¹⁷ Anwendbare und generalisierbare Untersuchungsergebnisse lassen sich eben nicht aus korrelativen Studien gewinnen, sondern sie bedürfen experimentell gewonnener Belege.

Freilich muss man auch zugeben, dass es außerordentlich schwierig ist, in den hochkomplexen Situationen und Handlungskontexten, welche für Schule und Unterricht typisch sind, diejenigen Parameter zu identifizieren, die sich auf die Lernergebnisse auswirken. Allein schon die seit Jahrzehnten diskutierte und beforschte Frage der Schulorganisation – homogene oder heterogene Gruppierungen, innerhalb und zwischen den Schulen – nähert sich erst in jüngster Zeit aufgrund der kompetenten Auswertung internationaler Studien einer Antwort. Dennoch darf das Faktum, dass nur unter großen Schwierigkeiten belastbare Befunde zu erzielen sind, nicht als Anlass akzeptiert werden, wenig tragfähigen Untersuchungen unzulässige Beweiskraft zuzusprechen, um darauf aufbauend Lehrerleistung zu beurteilen oder Lehrerbildung zu modellieren.

Auch bei den großen, nach den besten verfügbaren methodischen Qualitätsstandards durchgeführten internationalen Querschnittsuntersuchungen (z.B. IEA, TIMSS, PISA) sind die spezifischen Einflüsse des Unterrichts und des Lehrerhandelns in der Regel nicht von denen anderer Einflussfaktoren in einer Weise abgrenzbar, dass man aus ihnen verlässliche Kenntnisse über die Wirksamkeit des Lehrerhandelns gewinnen kann, da es die methodische

¹⁷ Hierzu gehören nach meiner Beurteilung auch viele der Studien, „die“ – man kann es nicht eleganter formulieren – „auf der Grundlage kleiner Konvenienzstichproben durchgeführt wurden und oftmals methodischen Einwänden offen stehen“ (Baumert/Kunter 2006, S. 503).

Anlage dieser Untersuchungen generell verbietet, *rerum cognoscere causas* (Redslob), wie auch die Autor/inn/en der Studien selbst immer wieder explizit betonen.¹⁸ Sehr wohl aber sind die Ergebnisse der hoch differenzierten Analysen dieser Daten eine ausgezeichnete Grundlage, um gezielte Experimente anzusetzen.

Um in dieser Situation weiterzukommen, sind zunächst zwei Schritte notwendig:

- Zum einen müssen die – buchstäblich – nach Tausenden zählenden Untersuchungen der Unterrichtsforschung *nach methodischer Tragfähigkeit sortiert, gewichtet und bewertet* werden, bevor sie den Empfehlungen für die Praxis zugrunde gelegt werden. Bestandsaufnahmen vorliegender Forschung, die in der beschriebenen Weise verfahren, sind bereits anzutreffen, allerdings nur vereinzelt und unvollständig (vgl. Kap. 2.3 und 5).
- Zum anderen gilt es, bei künftiger Forschung solche forschungsmethodischen Designs und Auswertungsverfahren zu bevorzugen, die Ursache-Wirkungs-Zusammenhänge aufzudecken und Generalisierungen¹⁹ vorzunehmen erlauben, damit der Weg zu einer evidenz-basierten (und nicht nur „eminenz“-basierten) Praxis offen steht.²⁰ Forschungsethische Überlegungen stehen dem keineswegs entgegen, wie sich längst bei denjenigen Studien gezeigt hat, die den so genannten „Gold Standard“ der Forschung, das Experiment, in der Unterrichtsforschung einhalten. Denn vor einem Versuch weiß man ja nicht, ob man z.B. bei einer Interventionsstudie mit den neuen oder den alten, in den Kontrollgruppen verwendeten Verfahren (z.B. Unterrichtsmethoden) den Schüler/inne/n Nachteile oder Vorteile verschafft.

Gewiss sollen die methodisch anders angelegten Studien damit nicht entwertet werden, zumal sich ein Teil der erziehungswissenschaftlichen Fragestellungen nur schwer mit einem experimentellen Design bearbeiten lässt. Solche Studien können dazu dienen, eine plausible Auswahl der mit Priorität zu untersuchenden Variablen zu treffen. Insbesondere können qualitative

¹⁸ Vgl. z.B. Prenzel u.a. 2005, S. 35: „Der PISA-Datensatz lässt eine angemessene Prüfung von Hypothesen über kausal relevante Bedingungen nicht zu“. Die Anbindung von PISA-Befunden an die Daten einer – Kausalitätsaussagen erlaubenden – Längsschnittstudie (BIJU) ist ein ganz neuer Ansatz und gehört zu den bemerkenswerten, aber bislang nur sehr selten angewandten forschungsmethodischen Fortschritten (vgl. Baumert u.a. 2006). Ähnlich bedeutsam ist die PISA-i-Plus-Studie, bei der die 2003 getesteten Jugendlichen ein Jahr später erneut untersucht wurden, soweit sie noch erreichbar waren (vgl. Prenzel u.a. 2006, besonders S. 30). Vgl. auch das Projekt COACTIV des Max-Planck-Instituts für Bildungsforschung. – Ein Musterbeispiel methodisch verantwortungsvoller Interpretation von Forschungsbefunden findet sich z.B. bei NICHD (2006).

¹⁹ Deren *Essentials* sind Kontrolle der unabhängigen Variablen und Randomisierung, vgl. z.B. *Encyclopedia of Educational Research* 1969, S. 1136.

²⁰ Vgl. auch Slavin 2002.

Studien höchst sinnvoll sein zur Gewinnung Erfolg versprechender und plausibler Forschungshypothesen.²¹

2.3 Forschungsmethodisch anspruchsvolle Überblicke

a) Einen wichtigen Schritt in die notwendige Richtung stellt die Publikation von Kulik und Kulik (1989) dar, in der die Autoren zum einen grundsätzliche Überlegungen anstellen zur Methode der Zusammenfassung und Nutzung von Forschungsbefunden in Form von Metaanalysen, wobei sie auf zahlreiche vorausgehende Literaturübersichten und Metaanalysen eingehen. Zum anderen beschreiben und bewerten sie den Stand des Wissens in ausgewählten inhaltlichen Bereichen der Schul- und Unterrichtsforschung. Diese Publikation – sie wird befremdlicherweise u.a. in der Arbeit von Anderson (2004) nicht erwähnt – fordert dazu heraus, fortgesetzt und auf den neuesten Stand gebracht zu werden; was aber erst vor kurzem von Seidel und Shavelson geleistet worden ist,²² wobei auch diese die Arbeit von Kulik und Kulik (1989) nicht erwähnen, obwohl dort ein Fundus relevanter Informationen enthalten ist und teilweise ähnliche Ergebnisse berichtet werden.

In die Metaanalysen Kuliks werden vor allem solche Studien einbezogen, die experimentell oder quasi-experimentell angelegt sind und deren Ergebnisse standardisiert wurden und statistisch aufeinander bezogen werden können. So lässt sich die so genannte Effektstärke, also quasi der Wirkungsgrad der untersuchten Variablen im Vergleich mit Kontrollgruppen feststellen.

Kulik und Kulik gliedern ihre Übersicht über den Forschungsstand nach Themen, listen die jeweils verfügbaren empirischen Studien bzw. bereits vorhandenen Metaanalysen z.B. über Unterrichtsformen, neue Curricula, Lehrerbildung, Klassen- und Schulorganisation oder Chancengleichheit auf und bewerten sie nach den Regeln einer professionellen Metaanalyse. Es finden sich eine ganze Reihe von bemerkenswerten Befunden, die auf relevante Variablen verweisen, die auch hierzulande in die Bewertung und Verbesserung von Lehrerhandeln eingehen können, sobald sichergestellt ist, dass die aus dem angelsächsischen Raum stammenden Arbeiten auch im Kontext des bundesrepublikanischen Schulwesens Gültigkeit beanspruchen können.

Ein Ergebnis von allgemeiner Bedeutung ist, dass die meisten der gesicherten Befunde aus (quasi-)experimentellen Studien nur mäßig starke Effekte aufweisen. Der stärkste beobachtete Effekt ergab sich bei Studien, die auf die Lernfortschritte von besonders begabten Schüler/inne/n in eigens für sie eingerichteten Kursen bezogen waren; er betrug ca. 0,90 Standardabweichung.

²¹ Vgl. auch Baumert/Kunter 2006, S. 490, sowie die Wertung qualitativer Studien bei Baumert u.a. 2010.

²² Wenn man einmal von Arbeiten zu einzelnen Aspekten absieht.

chungen, um welche die Schüler/innen in den Spezialkursen über den Leistungen gleich Begabter in normalen Klassen lagen. Die Mehrzahl der gefundenen Effektstärken liegt dagegen zwischen 0,10 und 0,55 Standardabweichungen. Dass innovative Treatments in den meisten Fällen nur relativ geringe Veränderungen bewirken, stellt eine wichtige Erkenntnis für die Anwendung der Befunde sowie für künftige Forschung dar. Andererseits darf nicht übersehen werden, dass auch geringe Effekte von großer politischer und sozialer Bedeutung sein können, etwa im Hinblick auf Chancengleichheit.

Im engeren Sinne auf Schule und Unterricht bezogen lehren uns die von Kulik und Kulik zusammengefassten und bewerteten Untersuchungsergebnisse und Metaanalysen u.a. folgendes:²³

- Tutorenprogramme, computerbasierter Unterricht, bestimmte Formen des *Mastery Learning* und andere unterrichtsbezogene Methoden führen zu höheren Leistungen der Schüler/innen und zu besserer Motivation; andere didaktische und methodische Neuerungen haben der Überprüfung nicht standgehalten und teilweise sogar negative Effekte gezeigt.
- Eine klare Definition der anstehenden Lernaufgaben, wie immer sie sich konkretisiert – hier sind zahlreiche Optionen in unterschiedlichen Kontexten und Fächern in einer großen Zahl von Studien experimentell untersucht worden –, wirken sich ebenso wie bestimmte Formen des Feedback und der Überprüfung der Lernergebnisse positiv aus.
- Verschiedene Formen der Unterrichtsorganisation beeinflussen die Schülerleistungen nur schwach, abgesehen vom getrennten Unterricht für besonders Begabte. Klassengröße hat zwar nicht auf die Schulleistungen, wohl aber auf die Einstellungen der Schüler/innen Einfluss, soweit sich die Variation der Klassengröße in normalen Grenzen hält. Auch die Wirkungen, die heterogene Klassen auf das Selbstbild begabter und weniger begabter Schüler/innen haben, sind verlässlich untersucht und haben differenzierte Ergebnisse erbracht.
- Bilinguale Programme für Schüler/innen mit Migrationshintergrund ließen sich nicht als wirksam identifizieren, da Kulik und Kulik sich nicht auf akzeptable Studien und Metaanalysen stützen konnten und demzufolge dringlichen Forschungsbedarf anmelden. Hier gibt es seither, mehr als zu vielen anderen Themen, deutliche Erkenntnisfortschritte, welche die besondere Bedeutung des Erlernens der Verkehrssprache für die Leistungsentwicklung der Schüler/innen mit Migrationshintergrund nachweisen.²⁴

Die Arbeiten von Kulik und Kulik haben in methodisch reflektierter Weise die vorliegenden Bestände unseres Wissens über erfolgreiches Lehren und Lernen in der Schule gesichtet. Eine Schwäche ihrer Arbeit besteht in der Bereitschaft, auch solche Studien in die Metaanalysen einzubeziehen, deren Anlage und Auswertung strengen methodischen Standards nicht standhält, so dass man nicht unmittelbar diejenigen Arbeiten identifizieren kann, die

²³ Ich kann hier nur in sehr vergrößerter Form auf die Befunde eingehen.

²⁴ Vgl. z.B. Hopf 2005.

eine verlässliche Grundlage für Forschung, Entwicklung und praktisches Handeln darstellen. Insofern ist es verständlich, dass sie den anspruchsvolleren Ansatz von Slavin, die so genannte *best-evidence synthesis*, verwerfen (vgl. S. 255). Slavins Herangehensweise hat sich aber in den Folgejahren mit guten Gründen durchgesetzt und dazu geführt, dass in den USA inzwischen bevorzugt solche Untersuchungen durchgeführt und finanziert werden, die aufgrund ihrer methodischen Anlage zu belastbaren Ergebnissen führen und evidenzbasiertes Lehrerhandeln ermöglichen.

b) In mehrerer Hinsicht weiter entwickelt als die Arbeit von Kulik und Kulik stellt sich die neueste Metaanalyse über *Teaching Effectiveness Research* dar, die im Dezember 2007 von Seidel und Shavelson publiziert worden ist. Sie weist den besonderen Vorteil auf, dass im Unterschied zu den sonst ganz überwiegend auf den angelsächsischen Raum beschränkten Übersichten die Forschung aus insgesamt 25 Ländern einschließlich des deutschsprachigen Raums einbezogen ist (vgl. S. 488). Wie bei Kulik und Kulik werden wichtige vorausgehende Metaanalysen ausführlich berücksichtigt, insbesondere die beiden jeweils zehn Jahre vorhergehenden von Scheerens and Bosker (1997) sowie von Fraser, Walberg, Welch und Hattie (1987).

Die Mehrzahl (zwei Drittel) der von Seidel und Shavelson aufgearbeiteten Studien gehören methodisch nicht zu denjenigen Arbeiten, mit deren Hilfe man über Ursachen und Wirkungen Näheres erfahren kann, sondern bestehen in korrelativen Surveys. Aber es werden auch alle verfügbaren experimentellen und quasi-experimentellen Studien aufgearbeitet und getrennt von den anderen dargestellt. Erstaunlicherweise wird allerdings die methodisch bedingte Erklärungsschwäche von Korrelationsstudien nicht eigens benannt, sondern lediglich mit Sympathie auf die US-amerikanische Debatte über den „gold standard“ der empirischen Bildungsforschung – experimentelle Vorgehensweisen – verwiesen, so als stünden nicht Welten zwischen der Belastbarkeit und Tragfähigkeit von Befunden aus Studien der einen oder der anderen Art – vermutlich eine Art kollegialer Höflichkeit der beiden Autoren (vgl. z.B. S. 484).

Es ist hier nicht der Ort, auf die zahlreichen interessanten Ergebnisse der Arbeit von Seidel und Shavelson einzugehen, einige ihrer für die vorliegende Diskussion wichtigen Punkte seien gleichwohl aufgeführt:

- Die berücksichtigten Effekte auf Seiten der Schüler/innen erstrecken sich nicht nur auf Schulleistungen, sondern auch auf motivational - affektive Aspekte sowie Prozessvariablen des Lernens (vgl. S. 470). Diese Ausdifferenzierung der Lernergebnisse erwies sich als fruchtbar, da durchaus differentielle Effekte zu beobachten sind, die bei summarischem Vorgehen unentdeckt bleiben.
- Unterschiedliche theoretische Modelle wurden auf ihre Tauglichkeit zur Einordnung der Daten geprüft; die Klassifizierung gemäß neueren kognitiven Modellen des Lehrens und Lernens bewährt sich (vgl. S. 461, 472).

- Abgesehen von wenigen Ausnahmen waren die Effektstärken der Ergebnisse der analysierten Untersuchungen niedrig bis moderat, wobei freilich auch niedrige Effekte für die Praxis relevant sein können (S. 471)²⁵.
- Über alle Untersuchungen hinweg zeigt die Variable Lernzeit²⁶ (*time on task, opportunity to learn, homework*; vgl. S. 467f.) bemerkenswerte Effektstärken, übertroffen nur von schulfachspezifischen Prozessvariablen (mathematisches Problemlösen, naturwissenschaftliches Forschen, Strategien des Lesens und Schreibens; vgl. S. 469, 473).
- Unterschiedliche Unterrichtsstrategien zeigen Effekte auf unterschiedliche abhängige Variablen, die einen fördern mehr motivationale, die anderen mehr kognitive Lernergebnisse (vgl. S. 473).
- Wenn man die Effekte nach Schulfach unterteilt, zeigen wiederum die schulfachspezifischen Prozessvariablen (neben der Lernzeit) die stärksten Wirkungen, und zwar in Mathematik, Lesen, Naturwissenschaften. Ebenso erweisen sie sich als wirksam sowohl in der Primar- als auch in der Sekundarstufe (vgl. S. 473).
- Wenn die untersuchten Studien methodologisch unterschieden werden in korrelative und experimentelle, ergeben sich für die korrelativen insgesamt auffällig niedrige Effektstärken, was die Autoren vor allem mit den dort meist verwendeten, stärker distalen Indikatoren erklären, die vom konkreten Lerngeschehen weiter entfernt sind als die bei experimentellen Studien verwendeten. Genau genommen dürfte man im Übrigen hier gar nicht von „Effekt“-Stärken reden, da korrelative Untersuchungen keine Wirkungszusammenhänge im kausalen Sinn aufdecken können.
- (Quasi-)experimentelle Studien zeigen vielfach nicht nur größere Effektstärken, sondern es erweisen sich auch neue Variablen als wirksam, vor allem *Organization of learning (classroom management, discipline, control)*. Allerdings ließen sich auch eine Reihe von Variablen identifizieren, die zwar in korrelativen Surveys, nicht jedoch in experimentellen Studien erfasst wurden, so dass hier ein deutlicher Nachholbedarf in der Forschung besteht.
- Die in unserem Zusammenhang besonders wichtigen Befunde aus den experimentellen oder quasi-experimentellen Studien lauten:
 - Auf den Lernprozess als abhängige Variable (z.B. Stärke der Beteiligung, der Lernmotivation, Verwendung intensiver Lernstrategien; vgl. S. 470) wirken sich aus (nach Größe geordnet) schulfachspezifische Tätigkeiten, Lernzeit, soziale Interaktion (kooperatives Lernen, Schüler-Diskussion, aktive Schülerarbeit), Unterrichtsorganisation.

²⁵ So schon Kulik/Kulik 1989, S. 332.

²⁶ So schon Wiley 1974. Der Vergleich zwischen Ländern mit zeitlich unterschiedlichen Unterrichtsangeboten und Beschulungsquoten beispielsweise zeigt die erheblichen Auswirkungen dieser Variable auf die Schulleistungen. Vgl. auch u.a. die Ergebnisse der internationalen TIMS- und PISA- Studien (z.B. Beaton 1996; Baumert u.a. 2001). Baumert u.a. (2001, S. 416ff.) weisen darauf hin, dass die zur Verfügung gestellte Lernzeit für sich allein kein befriedigender Prädiktor von Schulleistung ist. Die Variable *time on task* ist im übrigen von Soziolog/inn/en und Pädagog/inn/en erstaunlicher Weise oft nicht verstanden worden, was zu den merkwürdigsten Schlussfolgerungen geführt hat; vgl. z.B. Söhn 2005.

- Für *motivational-affektive* Lernergebnisse sind am bedeutsamsten schulfachspezifische Tätigkeiten sowie Unterstützung, Feedback, Selbststeuerung.
- *Kognitive* Lernergebnisse werden am nachhaltigsten gefördert durch schulfachspezifische Tätigkeiten, Lernorganisation sowie Zielsetzung bzw. Strukturierung (vgl. S. 479/80).
- Als insgesamt einflussreichste Variable erwiesen sich die schulfachspezifischen Tätigkeiten. Wenn es gelingt, Schüler/innen in fachspezifische Lernaktivitäten zu verwickeln, kann man darauf setzen, in allen Altersgruppen, Fächern und Aspekten der Lernergebnisse fördernd zu wirken (vgl. S. 483).
- Methodologisch interessant ist die Gegenüberstellung von experimentellen und quasi-experimentellen Studien bei Seidel und Shavelson, die das Vertrauen in die Ergebnisse der quasi-experimentellen Arbeiten, die ohne Randomisierung auskommen, stärken (vgl. S. 480).²⁷ Aus methodologischen Gründen ist dieser Befund nicht generalisierbar, er gilt nur für die konkret verglichenen Untersuchungen.

c) Ausdrücklich hingewiesen sei hier auf die ausführliche Arbeit von Baumert und Kunter (2006), die aber nicht im Einzelnen dargestellt werden kann. Sie ist sowohl theoretisch anspruchsvoll und ergiebig als auch ordnet sie eine große Zahl von Befunden, meist nach ihrer methodisch bedingten Belastbarkeit charakterisiert, in das entworfene theoretische Gerüst ein. Die Fülle der dort ausgebreiteten Informationen hätte ein metaanalytisches Vorgehen wünschenswert gemacht; der aufgespannte breite Rahmen, in dem das Thema „professionelle Kompetenz von Lehrkräften“ behandelt wird, hätte allerdings zu einem für die Zwecke des Artikels unangemessenen Umfang geführt. Die Leser/innen seien aber ausdrücklich auf diese Arbeit, die auch eine ganze Reihe von belastbaren und unmittelbar praxisrelevanten Forschungsergebnissen berichtet, verwiesen.

d) Ferner gibt es eine Reihe von Untersuchungen, die zu neu sind, um bereits Gegenstand einer Metaanalyse geworden zu sein, die aber relevante und methodisch robuste Ergebnisse entweder schon erzielt haben oder in Aussicht stellen. Beispiele hierfür finden sich etwa in der Studie PISA-I-Plus, einer längsschnittlichen Ergänzung der PISA-Erhebung 2003 (vgl. Prenzel u.a. 2006). So verweist etwa Senkbeil (2006, vgl. S. 303) auf die signifikant höheren Förderleistungen sog. aktiver Lehrkräfte (mit hohen Werten in Kooperation, Evaluationspraxis, Engagement) im Unterschied zu disziplinerorientierten hin. Wengleich sich diese Befunde auf eine Schülergruppe beziehen, die beim Übergang von der 9. zur 10. Jahrgangsstufe den überwiegenden Anteil der schwächeren Schüler/innen verloren hatte, sind die Ergebnisse von hohem Interesse. Derselbe Datensatz belegt auch, dass

- eine durch die Lehrer/innen verwirklichte effektive Klassenführung, in der die „Zeit hauptsächlich für lernrelevante Aktivitäten genutzt wird“,
- lernbezogene konstruktive Unterstützung,

²⁷ Zu diesem Befund kamen bereits Kulik/Kulik (vgl. 1989, S. 270), deren verdienstvolle Metaanalysen von Seidel/Shavelson nicht erwähnt werden.

- die Quantität der Lerngelegenheiten sowie
 - deren Qualität im Sinne eines kognitiven Anregungspotentials
- Mathematikleistungen nachweislich fördern (vgl. Kunter u.a. 2006, S. 189 u.a.). – Über weitere Ergebnisse des Projektes COACTIV („Kognitiv-aktivierender Unterricht“) des Max-Planck-Instituts für Bildungsforschung wird unten in Abschnitt 4 berichtet, weil es sich gut mit Fragen der Lehrerbildung verbinden lässt. Sie stellen einen Durchbruch in der bisherigen Forschung dar.

e) Schließlich sei verwiesen auf die neu erschienene, eindrucksvolle Studie von Slavin und Lake (2008), welche die in den USA erschienenen Untersuchungen über die Wirkungen von Projekten der Grundschulmathematik nach methodisch strengen Gesichtspunkten vollständig durchgesehen haben. Slavin und Lake trennen nach Studien zu curricularen Änderungen, zum Einsatz von computerunterstütztem Unterricht sowie zu veränderten Unterrichtsstrategien der Lehrer/innen. Vor allem einige der untersuchten veränderten Formen des Lehrerverhaltens²⁸ erwiesen sich dabei als besonders förderlich für den Lernzuwachs der Schüler/innen.

3. Forschung über Anreizsysteme für Lehrer/innen

Es zeigt sich also, dass es eine – begrenzte – Anzahl von Forschungsbefunden gibt, welche die ursächliche Wirkung von umschriebenen Formen des Unterrichts und des Lehrerverhaltens auf die Lernergebnisse der Schüler/innen belegen, und dass eine Reihe entsprechender Arbeiten der Auswertung und Durchführung harren. Dabei handelt es sich ganz überwiegend um Treatments, die zu einem nachweisbaren Lernzuwachs führen und nur sehr selten um solche mit negativem Ausgang.

Wie lässt es sich aber erreichen, dass die durch (quasi-)experimentelle Untersuchungen identifizierten, erwünschten Verhaltensweisen verstärkt angewendet werden, und gibt es Untersuchungen, die zeigen, dass dies mit Hilfe von Anreizen erreicht werden kann und sich auf die Lernergebnisse der Schüler/innen auswirkt? Und gibt es Untersuchungen darüber, wie man kontraproduktive Treatments zurückfahren kann?

Es ist schwierig, Studien zu finden, in denen auf methodisch überzeugende Weise dieser Schritt getan worden wäre, dass also konkrete, nachgewiesenermaßen wirksame Formen von Unterricht und Lehrerhandeln gezielt verstärkt und anschließend die Auswirkungen des so modifizierten Lehrerhandelns auf die Schüler/innen untersucht worden wären. Nach dem Eindruck zu urteilen, den man aus der umfangreichen Literatur und den zahlreichen

²⁸ S. 466 ff: Effektiv sind: *cooperative learning / individualized programs / direct instruction / mastery learning / professional development focused on math content / professional development focused on classroom management and motivation / supplemental programs*. Vgl. neuerdings auch die analoge Studie von Slavin u.a. (2009) für die Sekundarstufen.

Dokumenten im Internet erhält, nähern sich aber inzwischen eine Reihe von Arbeiten diesem Ziel, so dass in absehbarer Zeit Einsichten in solche Wirkungszusammenhänge zu erwarten sind. Allein schon aus den erheblichen Forschungsanstrengungen im Zusammenhang mit dem US-amerikanischen Projekt *No Child Left Behind* (NCLB), das seit 2002 realisiert wird, sind wichtige Ergebnisse zu erwarten, da dort denjenigen Schulen, welche die angestrebten Standards nicht erreicht haben, geraten wird, insbesondere solche Unterrichtsformen und Lehrerverhaltensweisen verstärkt einzusetzen, die sich als wirksam erwiesen und in belastbaren Untersuchungsergebnissen niedergeschlagen hatten.²⁹ Dabei laufen die nicht-erfolgreichen Schulen Gefahr, erheblichen Sanktionen bis hin zu Schulschließungen unterworfen zu sein, wenn sie die Lernergebnisse ihrer Schüler/innen nicht in der gewünschten Weise verbessern.

Allerdings finden sich auch ganz anders verführende Untersuchungen, in denen Anreize gesetzt werden für messbare Erfolge von Unterricht oder Lehrerhandeln, ohne dass vorgegeben würde, was die Lehrer/innen tun oder lassen sollen. Unterricht und Lehrerhandeln werden hier quasi als *black box* behandelt und es wird nur überprüft, ob die Anreize letztlich zu besseren Lernergebnissen der Schüler/innen führen, auf welche Weise auch immer dies geschieht. Man kennt lediglich die Ausgangslage und den Endzustand der Schüler/innen und geht davon aus, dass die Lehrer/innen, wenn man ihnen Anreize für bessere Leistungen ihrer Schüler/innen zusagt, in der Regel wissen, wie sie diese erzielen können, und vertraut in ihre professionelle Kompetenz und ihr Expertenwissen. Als Beispiel kann hier eine der Untersuchungen von Lavy stehen.³⁰

Lavy (2003) hat in einer Längsschnittstudie an einer großen Zahl israelischer Schulen den Lehrer/inne/n zusätzliche finanzielle Einkünfte³¹ angebo-

²⁹ Vgl. z.B. U.S.Department of Education, Institute of Education Sciences (2003). – Das Projekt *No Child Left Behind* (NCLB) hat eine große Zahl von Forschungsarbeiten in Gang gesetzt, die durchzusehen hier nicht geleistet werden kann. Vgl. dazu Zeichner (2006) sowie zahlreiche Texte des U.S. Department of Education; Hess/Petrilli (2007); Fuller u.a. 2007, mit ihrer fundierten Kritik an dem Projekt und den Zweifeln an seiner Effektivität. Vgl. aber auch Slavin/Lake 2008, S. 428. Einen interessanten Überblick über die Forschungsergebnisse des NCLB vermitteln Mintrop u.a. (2009). Der Aufsatz widmet sich verdienstvollerweise vor allem dem Sanktions-Aspekt der in dem Projekt NCLB enthaltenen Maßnahmen – ein Fokus, den man bei den meisten anderen Arbeiten vermisst. – Zu methodischen Fragen des NCLB vgl. die wichtigen Texte von Amrein-Beardsley (2008) sowie Kim/Sunderman (2005). – Auch in England wird mit Sanktionen bis hin zu Schulschließungen gearbeitet, die denen von *No Child Left Behind* nicht unähnlich sind.

³⁰ Mit weiterführender Literatur zu *performance-related incentives* und Beispielen solcher Programme in den USA, England, Neuseeland etc., wo es aber an evidenz-basierter Forschung mangle; vgl. auch Lavy (2007) mit weiteren Berichten und einer Diskussion möglicher ungewollter und unerwünschter Konsequenzen wie der Konzentration der Lehrertätigkeit auf das Erreichen des belohnten Schülerverhaltens. – Einen guten Überblick über weitere Studien geben Prentice u.a. (2007).

³¹ Es handelte sich hierbei um Beträge bis zu einem Viertel eines Jahresgehaltes, nicht um so niedrige Werte wie 3%, die in der deutschen Diskussion kursieren. Vgl. aber die Studie von Muralidharan/Sundaraman (2008).

ten, wenn sie die Leistungen ihrer Schüler/innen, gemessen u.a. an der Quote der Abschlussexamina, verbesserten. Es blieb den Lehrer/inne/n überlassen, wie sie ihr Ziel erreichten, und die nach Rangordnung gestaffelten Honorare wurden denjenigen Lehrer/inne/n übergeben, welche die besten Schülerwerte erzielt hatten. Verglichen wurden die Endergebnisse nicht nur mit den Schülerleistungen, die vor Einführung der Anreize an den Experimentalschulen zu beobachten gewesen waren, sondern auch mit denen an vergleichbaren anderen Schulen, an denen es keine Anreize gab.

In der Tat haben sich aufgrund der den Lehrer/inne/n gebotenen Zusatzeinkünfte die Leistungen der Schüler/innen erheblich gesteigert. Lavy versuchte *ex post*, die Gründe für die Verbesserungen zu identifizieren und fand heraus, dass die wirksamen Lehrer/innen vor allem ihren Schüler/inne/n mehr Zeit zur Verfügung gestellt hatten (auch nach der Schule), stärker auf ihre individuellen Bedürfnisse eingegangen waren und ihre eigenen Unterrichtsstrategien modifiziert hatten. Das verwendete Anreizsystem war besonders wirksam bei lernlangsamen Schüler/inne/n.

Nach einem ähnlichen Prinzip ist auch die eindrucksvolle Experimentalstudie von Muralidharan und Sundararaman (2008) konzipiert, an der 500 indische Primarschulen beteiligt waren. Die Zusatzeinkünfte der Lehrer/innen in den Experimentalgruppen beliefen sich auf ca. 3% des jährlichen Schulbudgets. Lese- und Rechenleistungen wurden zwei Jahre lang gemessen (2005 bis 2007). Auch hier ergaben sich, ohne dass den Lehrer/inne/n vorgegeben worden war, was sie zu tun hätten, im Lauf der beforschten Schuljahre signifikante Verbesserungen der Lernergebnisse der Schüler/innen. Neben den zahlreichen interessanten Aspekten der Studie verdient zum einen hervorgehoben zu werden, dass es sich hier um eines der seltenen echten Experimentalstudien handelt. Außerdem ist bemerkenswert, dass auch in dieser Studie die von den Lehrer/inne/n selbst gewählten Verhaltensänderungen im Umgang mit den Schüler/inne/n vor allem im Bereitstellen von mehr Zeit für die Betreuung bestanden (*conducting extra classes after school, additional homework and classwork, providing practice tests*; vgl. S. 3 und S. 26, Tabelle 9). *Time on task* bewährt sich einmal mehr als zentrale Variable für die Lernfortschritte der Schüler/innen.³²

Es gibt aber nicht nur *Forschungsprojekte* wie die genannten, in denen Anreizsysteme als *black box* erprobt werden, sondern es finden sich international auch quantitativ umfangreiche, von den Schulbehörden bereits implementierte Formen, in denen Erfahrungen mit der Wirksamkeit von Anreizsystemen³³ gesammelt werden, ohne dass vorab als effektiv erwiesene For-

³² Vgl. Carroll (1963, 1989) sowie für die Ergiebigkeit dieses Modells z.B. Hopf (2005). – Weitere Literatur zum Thema *performance pay* bei Lehrern vgl. Muralidharan u.a. (2008, S. 4, Anm. 5).

³³ Bei solchen Anreizformen sind im übrigen nicht solche gemeint, wie wir sie in Deutschland seit langem kennen, wo z.B. Gymnasiallehrer/innen verschiedene Beförderungsstufen durchlaufen können (Oberstudienrat bzw. –rätin, Studiendirektor/in etc.); denn dabei bezieht man sich nicht explizit, und schon gar nicht aufgrund belastbarer Forschungsbefunde, auf die Lernfortschritte der Schüler/innen. Sondern es geht um Ansät-

men von Unterricht oder Lehrerverhalten vorgegeben wären. Beispiele hierfür gibt es im Umfeld des Makinson-Report (2000) für England.³⁴

4. Forschung über Lehrertraining und Lehrerbildung

Anreizsysteme können als Sofortmaßnahmen konzipiert werden, die in bestehende Schulen eingeführt werden und ihre Wirkung ohne lange Verzögerung entfalten können. Im Unterschied hierzu haben Veränderungen, die in die Lehrerbildung eingebracht werden, eine deutlich längere Vorlaufzeit, ehe sie den Schüler/innen zu Gute kommen können. Gleichwohl wäre es sinnvoll, Kenntnisse über evidenz-basierte Formen erfolgreichen Unterrichts oder, allgemeiner, des produktiven Lehrerhandelns möglichst rasch in die Lehrerbildung zu integrieren, damit künftige Lehrergenerationen von vornherein über das aktuell mögliche Expertenwissen verfügen. Wesentlich beschleunigen ließen sich die Wirkungen, wenn mit Hilfe von Lehrerfortbildung bereits praktizierende Lehrer/innen ihr Verhalten modifizierten. Anreizsysteme könnten an die Umsetzung solcher Treatments im Unterricht gebunden werden, wenn diese positive Auswirkungen auf die Schüler/innen haben.

Die bisherige Forschung gibt wenig Auskunft über differenzielle Effekte unterschiedlicher Formen der Lehrerbildung. So gern man wüsste, ob beispielsweise das Potsdamer Modell der Lehrerbildung gegenüber dem an der FU Berlin praktizierten sich positiver auf die Lernergebnisse der Schüler/innen auswirkt (wie implizit unterstellt wird), gibt es keinen solchen Vergleich, ja nicht einmal den Ansatz, auch nur eines dieser Modelle oder seiner Elemente auf seine Wirkungen zu überprüfen. Es gibt jedenfalls im Unterschied zu der großen Fülle von Untersuchungen über die Effektivität von Unterricht und Lehrerverhalten nur sehr wenige Studien zur schülerbezogenen Effektivität von Lehrertraining.³⁵ Unter ihnen befinden sich im Üb-

ze, bei denen entweder für nachgewiesene Lernerfolge der Schüler/innen positive Anreize gesetzt werden (z.B. durch zusätzliche finanzielle Angebote an die Lehrer/innen und/oder Schulen) oder bei Ausbleiben der angestrebten Lernfortschritte Sanktionen anstehen (z.B. Schulschließungen).

³⁴ Im Internet findet sich eine umfangreiche und interessante Literatur hierzu; vgl. URL: <http://www.ome.uk.com/downloads/Performance%20pay%20in%20the%20Public%20Sector.%20A%20review%20of%20the%20issues%20and%20evidence.pdf> (Zugriffdatum: 21.04.2008), wo britische, israelische, US-amerikanische Studien berichtet werden und es z.B. heißt: „There is strong evidence that teachers do respond to financial incentives. Several studies suggest that this response does not universally affect all students: most of the improvement appears to come from previously weak students performing better under such schemes“. Weitere Beispiele finden sich bei Lavy (2007), bei Prentice u.a. (2007) oder bei Marsden u.a. (2006). – Die so genannten *Charter Schools* in mehreren US-amerikanischen Bundesstaaten lassen sich ebenfalls hier subsumieren; vgl. z.B. C. Hoxby (2004).

³⁵ So schon Anderson (2004, S. 115): „... far more is known about teacher effectiveness than is known about increasing teacher effectiveness“. Auch Oser/Oelkers (2001) führen in dieser Frage nicht weiter; vgl. hierzu auch Gehrman (2007).

rigen auch einige kompetent gemachte Studien über Lehrerfortbildung und ihre Auswirkungen auf die teilnehmenden Lehrer/innen, beispielsweise aus der Forschungstradition des so genannten Microteaching, wo man u.a. untersucht hat, ob das Verhaltenstraining den erwarteten Erfolg bei den trainierten Lehrer/innen hatte und ob sich die Verhaltensänderungen über die Zeit erhalten.³⁶ Der notwendige nächste Schritt aber, nämlich die Überprüfung der Effekte des veränderten Verhaltens der Lehrer/innen auf die Lernerfolge der Schüler/innen, wird dort selten getan, so dass wir nicht erfahren, ob es auf der Schülerseite die erwünschten Entwicklungen gibt.

Allerdings ist auch zu diesem Thema die Metaanalyse von Kulik und Kulik (vgl. 1989, S. 303ff.) bislang übersehen worden. Sie zitiert auch ältere Metaanalysen und berichtet über die Auswirkungen von *preservice* und *inservice* Programmen auf die Lehrer/innen sowie auf die Leistungen der Schüler/innen. Dabei wird sowohl über die Ergebnisse korrelativer als auch experimenteller Arbeiten berichtet, wobei die Autoren allerdings, wie oben bereits bemerkt, die Untersuchungen nicht nach den verwendeten Forschungsmethoden trennen. Überprüft wurden Studien in unterschiedlichen Schulfächern, in denen als Treatments z.B. Formen forschenden Lehrens, durch Lehrer/innen vorgenommene Unterrichtsanalysen oder die Verwendung anspruchsvoller Fragen im Unterricht Verwendung fanden. Das Ergebnis der Übersicht ist eindeutig: „The preservice programs examined have not had an important effect on teacher knowledge, attitudes, or behavior. The inservice programs have had clear effects on these same outcome measures. In addition, their effect on student learning has been clearly demonstrated“.³⁷

In einigen neueren Studien werden Lehrer/innen mit unterschiedlichen Abschlussexamina miteinander verglichen. So haben Akiba u.a. (2007) auf der Grundlage der Daten aus der *Third International Mathematics and Science Study* (TIMSS) von 2003 geprüft, wie die unterschiedlichen Lehrer-Examina in 46 Ländern mit Schülerleistungen zusammenhängen: Es gibt in der Tat solche Beziehungen in der erwarteten Richtung; die Studie ist freilich korrelativ. Akiba u.a. kommen zu dem Ergebnis: „Our data confirmed the importance of ensuring a qualified teaching workforce to produce higher national achievement“ (S. 379), wobei die Qualität der Lehrer/innen getrennt nach *Teacher Certification*, *Math major*, *Math education major*, *Teaching experience* sowie der daraus resultierenden *Overall teacher quality* bestimmt worden war. Interessant ist auch der Literaturbericht von Akiba u.a., aus dem die Bedeutung des Fachwissens und des pädagogischen Wissens der Lehrer/innen für die Schülerleistungen hervorgeht (vgl. S. 371). Fachwissen der Lehrer/innen erweist sich in den Sekundarschulen als besonders relevant; dies gilt auch für Studien mit einem anderen als korrelativen Forschungsdesign.

³⁶ Vgl. z.B. Gage/Berliner 1979, S. 707ff.

³⁷ Von Interesse ist hier auch das umfangreiche Werk von Cochran-Smith/Zeichner (2005), das sich im wesentlichen auf die Phase des *Preservice Training* der Lehrer/innen bezieht, sowie die konstruktive Kritik von Reichwein Zientek u.a. (2008).

Zu einer ähnlich positiven Einschätzung kommt Darling-Hammond (2002), die neben der Fachkompetenz und pädagogischen Kompetenz auch die Bedeutung der verbalen Intelligenz der Lehrer/innen für den Unterrichtserfolg hervorhebt. Dieselbe Autorin zeigt (2007) auch an weiteren (korrelativen) Datensätzen, dass ein enger Zusammenhang besteht zwischen den Leistungen der Schüler/innen in verschiedenen Fächern und der Qualität der Lehrer/innen (z.B. *license in field*). Diesen Befund bewertet die Autorin als besonders wichtig angesichts der Tatsache, dass unterprivilegierte Schüler/innen in den USA eine besonders hohe Quote von Lehrer/inne/n haben, die Fächer unterrichten, für die sie keine Fakultas besitzen.

Auch Fuchs und Woessmann (2004) berichten Ähnliches anhand der internationalen PISA-Daten: „Students in schools whose teachers had on average a higher level of education performed statistically significantly better than otherwise. This is true for degrees in pedagogy, in the respective subject and for specific teacher certificates. In math and reading, the relationship is substantially stronger for a masters degree in the subject than a masters degree in pedagogy. In math and science, a full teacher certification by the appropriate authority shows the strongest relationship to student performance, at 15 AP“ (S. 24). Vor allem die Ausbildung der Lehrer/innen in den Schulfächern, aber auch die pädagogische Vorbildung stehen also mit den Schülerleistungen in Beziehung, je nach Fach verschieden stark. Leider liegen diesen Aussagen nur Surveydaten zugrunde.

Die Bedeutung der Lehrerbildung wird ebenfalls unterstrichen in der Metaanalyse von Rice (2003), wo es zusammenfassend heißt, dass „teachers who have earned advanced degrees have a positive impact on high school mathematics and science achievement when degrees earned are in those subjects“.

Rockoff (2004) hat die Ergebnisse einer Panel-Studie über die Wirkungen von Lehrer/inne/n auf Schülerleistung ausgewertet und kommt zu der Aussage: „The empirical evidence above suggests that raising teacher quality may be a key instrument in improving student outcomes. However, ... policies that reward teachers based on credentials may be less effective than policies that reward teachers based on performance“ (S. 251). Auch hier also gibt es deutliche Hinweise auf die Bedeutung sowohl der Lehrerbildung als auch eines gezielten Anreizsystems für die Leistungen der Schüler/innen.

In immer mehr Studien lässt sich, wie bereits dargestellt, erkennen, dass es ergiebig ist, einzelne Elemente der Lehrerbildung zu isolieren und in wechselnden Konstellationen zu betrachten.³⁸ So gibt es gute Gründe zu vermuten, dass einerseits die fachlichen und andererseits die fachdidaktisch-pädagogischen Kenntnisse und Fähigkeiten der Lehrer/innen von Bedeutung

³⁸ Vgl. oben. Zahlreiche Belege hierfür finden sich auch in dem Aufsatz von Baumert und Kunter (2006), in dem u.a. auch die im deutschsprachigen Raum wichtigsten Ansätze (Terhart, Oser etc.) ausführlich diskutiert werden (vgl. S. 478ff.).

sind. Wichtige diesbezügliche Erkenntnisse verdanken wir neuerdings den Ergebnissen der COACTIV-Studie³⁹ sowie der MT21-Untersuchung⁴⁰; beide Projekte verweisen auf Vorgängerliteratur, in der von vergleichbaren Differenzierungen die Rede ist. – Da sich die MT21-Studie auf Studierende und Referendar/inn/e/n, also auf die Lehrerbildungsperiode bezieht, diskutiere ich diese und die COACTIV-Untersuchung wegen ihrer zahlreichen Berührungspunkte in diesem Kapitel.

Krauss u.a. (2008a) haben zunächst an einer repräsentativen Stichprobe von Mathematiklehrer/inne/n der Sekundarstufe I aus der PISA-I-Plus-Studie⁴¹ gezeigt, dass sich fachliche (*CK, content knowledge*) und fachdidaktisch-pädagogische (*PCK, pedagogical content knowledge*) Kenntnisse mit Hilfe der von ihnen konstruierten Instrumente empirisch unterscheiden lassen.⁴² Gymnasiallehrer/innen sind in beiden Dimensionen den Nicht-Gymnasiallehrer/inne/n deutlich überlegen. Ferner hat sich herausgestellt, dass sich bei Gymnasiallehrer/inne/n – im Unterschied zu Nicht-Gymnasiallehrer/inne/n – fachliche und fachdidaktisch-pädagogische Kenntnisse zu einer einzigen Kategorie zusammenschließen. Den Fachkenntnissen kommt eine Schlüsselrolle zu, da sie eine Voraussetzung für die Herausbildung guter fachdidaktisch-pädagogischer Kenntnisse darstellen: „Given that GY teacher candidates receive additional in-depth mathematics training at university, but no additional training in teaching mathematics (relative to NGY teacher candidates), their substantially higher PCK scores are remarkable, and may – although very tentatively – be interpreted as a first indication that CK supports the development of PCK.“⁴³ Diese (repräsentative) Studie ist insofern von besonderem Wert, als die Daten aus deutschen Schulen stammen und die Lehrer-Kompetenzen mit eigens konstruier-

³⁹ Vgl. vor allem Krauss u.a. 2008a (eingereicht 2006) und 2008b; Baumert u.a. 2010.

⁴⁰ Blömeke u.a. 2008.

⁴¹ Vgl. Prenzel u.a. 2006.

⁴² Zur begrifflichen Unterscheidung vgl. auch Shulman 1987 oder Borko u.a. 1996.

⁴³ Vgl. S. 724. – In Übereinstimmung damit steht der an einer für Westdeutschland repräsentativen Gruppe von Mathematiklehrer/inne/n der 7. Klasse des Gymnasiums schon vor Jahrzehnten gewonnene Befund, dass fachlich besonders gut ausgebildete Lehrer/innen, die vergleichsweise höhere Examina abgelegt haben, sich von einer engen Orientierung ihres Unterrichts am Lehrbuch distanzieren, vielmehr den Unterricht nach eigenen Vorstellungen planen und gestalten; im Gegensatz zu denjenigen Gymnasiallehrer/inne/n, die Mathematik ohne Fakultas unterrichten und sich eng an die Lehrbücher als methodischen Leitfaden klammern. Der Unterricht ist daher umso weniger lehrbuchbezogen, je mehr die Lehrer/innen auf ihre Fachkompetenz vertrauen können. Auch hier scheint demnach höheres Fachwissen den Weg zur Entwicklung eines differenzierteren fachdidaktisch-pädagogischen Wissens zu eröffnen (vgl. Hopf 1980, S. 37f.). Leider war es bei diesen aus der Schulleistungsstudie des Max-Planck-Instituts für Bildungsforschung stammenden Befunden seinerzeit aus institutsinternen Gründen nicht möglich, die vorhandenen Leistungsdaten der Schüler/innen jener Lehrer/innen in die Analyse einzubeziehen. Andernfalls hätten damals schon Erkenntnisse über den Zusammenhang zwischen Lehrerverhalten (operationalisiert in unterrichtsnahen Fragebögen) und Leistungs*zuwachs* der Schüler/innen (operationalisiert in eigens konstruierten curriculumvaliden Tests) präsentiert werden können.

ten Tests⁴⁴ erfasst wurden; allerdings sind sie in der genannten Publikation noch nicht mit Leistungsdaten der Schüler/innen verknüpft.

Die Studie von Blömeke u.a. (2008) bezieht sich ebenfalls auf die deutsche Szene und auf das Fach Mathematik. Auch diesen Autoren gelingt es, unabhängig von der vorgenannten Untersuchung, mit Hilfe speziell für diese Zwecke konstruierter Tests in einer sorgfältigen empirischen Analyse und theoriebezogenen Interpretation zunächst Fachwissen und fachdidaktisch-pädagogisches Wissen empirisch zu trennen. Sie fächern dann darüber hinaus insbesondere die Dimension des fachdidaktisch-pädagogisches Wissens weiter auf. Insgesamt kommen sie zu zahlreichen ähnlichen Befunden wie die COACTIV-Studie.⁴⁵ So trennen die Autoren, um die „Wirksamkeit der Lehrerbildung differenziert analysieren zu können“ (Kap. 1), ebenfalls zwischen Lehrer/inne/n mit gymnasialer Oberstufenfakultas und den anderen Lehrer/inne/n. Und auch bei ihnen zeigt sich die erhebliche Überlegenheit der Gymnasial- und Gesamtschullehrer/innen gegenüber den anderen sowohl was das mathematische Fachwissen (vgl. z.B. Tab. 3.1) als auch das mathematikdidaktisch-pädagogische Wissen betrifft (vgl. Tab. 3.2). Man kann diese Befunde, trotz der Abweichungen im Detail – die bei Blömeke ausführlich diskutiert werden (vgl. z.B. Kap. 2.3.2) –, als Validierung der COACTIV-Befunde werten (und umgekehrt), insbesondere weil in den beiden Studien unterschiedliche Tests verwendet und unterschiedliche Stichproben untersucht wurden. Bei der Interpretation der Ergebnisse werden von Blömeke u.a. auch sorgsam die Treatments verglichen und es wird im Detail expliziert, wie verschieden die Studien- und Ausbildungsanforderungen der beiden Lehrergruppen sind, so dass sich allein daraus schon ein Gutteil der gefundenen Kompetenzunterschiede erklären ließe.⁴⁶

In der MT21-Studie wird neben dem fachwissenschaftlichen und dem spezifisch fachdidaktischen noch das fächerübergreifende erziehungswissenschaftliche Wissen definiert, welches ebenfalls durch eigens konstruierte Tests erfasst wurde (vgl. Kap. 6). Die Aufgaben kommen aus den Bereichen Allgemeine Didaktik, Pädagogische Psychologie und Bildungssoziologie und konkretisieren sich in Fragen zur Unterrichtsplanung, Lernzielkontrolle sowie zum Umgang mit sozialer Ungleichheit. Diese Inhalte haben ihren Platz sowohl in der ersten wie in der zweiten Phase der Lehrerbildung (vgl.

⁴⁴ Vgl. zu weiteren Messinstrumenten Frey 2006.

⁴⁵ Ein nicht unwichtiger Unterschied zwischen den beiden Studien besteht allerdings darin, dass bei Blömeke u.a. Studierende und Referendar/inn/e/n untersucht worden sind, während es sich bei COACTIV um praktizierende, im Mittel knapp 50jährige Lehrer/innen handelt, deren Schüler/innen in PISA 2003 einbezogen waren. Außerdem stellt bei ihrem Forschungsdesign die professionelle Kompetenz der Lehrer/innen die von der Lehrerbildung abhängige Variable dar, während sie in der COACTIV-Studie als unabhängige Variable fungiert, deren Auswirkungen auf die Schülerleistungen geprüft wird.

⁴⁶ Vgl. z.B. S. 115 ff. Danach haben Gymnasiallehrer/innen in Mathematik und ihrer Didaktik 60 bis 80 Semesterwochenstunden studiert, gefolgt von 24 Monaten Referendariat; für die anderen Lehrer/innen liegen die Zahlen bei nur 40 bis 64 Semesterwochenstunden und bei 18 bis 24 Monaten Dauer der zweiten Phase.

S. 197); inhaltliche Schwerpunktsetzungen sowie Zeitdauer der Ausbildungsphasen beeinflussen die Lerngelegenheiten.

Im Ergebnis zeigten die Lehrer/innen der verschiedenen Schultypen in etwa gleich hohe Kompetenzen in Bezug auf Unterrichtsplanung und soziale Ungleichheit; was den Bereich der Lernzielkontrollen betrifft, waren die Gesamtschul-/Hauptschul-/Realschullehrer/innen der Gruppe der Lehrer/innen mit Oberstufenfakultas überlegen, was seinen Grund vermutlich in der stärker praxisbezogenen Ausbildung der überlegenen Gruppe hat (vgl. Kap. 7). Die zeitliche Erstreckung der Ausbildung der Lehrer/innen in den erziehungswissenschaftlichen Themen ist quantitativ bedeutsam (vgl. S. 201).

Festzuhalten bleibt, dass sowohl in der MT21-Studie wie in der COACTIV-Untersuchung der Nachweis gelingt, dass die unterschiedlichen Formen der Lehrerbildung (gymnasial versus nicht-gymnasial) unterschiedliche Ergebnisse hervorbringen; meist sogar dann, wenn man die Eingangsvoraussetzungen der Proband/inn/en statistisch kontrolliert. Diese beiden Studien liefern somit die ersten belastbaren Belege für die (differentielle) Wirksamkeit der Lehrerbildung auf die Lehrer/innen in Deutschland.

Den entscheidenden Schritt weiter geht schließlich die jüngste Publikation aus der COACTIV-Studie (Baumert u.a. 2010), in der die Lehrerdaten mit den Schülerdaten in Zusammenhang gebracht werden.⁴⁷ Zunächst werden die notwendigen Definitionen gegeben, wobei CK als dasjenige Fachwissen bestimmt wird, welches sich auf „a profound mathematical understanding of the mathematics taught at school“ bezieht (vgl. S. 142). PCK wird genauer bestimmt als „body of instruction- and student-related mathematical knowledge and skill – the knowledge that makes mathematics accessible to students“, bestehend aus drei Teilen: „knowledge of mathematical tasks as instructional tools, knowledge of students’ thinking and assessment of understanding, and knowledge of multiple representations and explanations of mathematical problems“. Nach der Beschreibung der Stichprobe und Instrumente sowie des Datenanalyseprozesses folgt der Beleg, dass die identifizierten Lehrerkompetenzen nachweislich Auswirkungen auf die Schülerleistungen im Fach Mathematik besitzen.

Die zahlreichen interessanten Details der Untersuchung und ihrer Befunde können hier nicht wiedergegeben werden.⁴⁸ Wichtig ist, dass wir aufgrund

⁴⁷ Die recht große Zahl von Untersuchungen zu Lehrer/inne/n und zur Lehrerbildung darf nicht darüber hinwegtäuschen, dass fast alle von ihnen das Kerngeschäft des Lehrers bzw. der Lehrerin, die Förderung der Entwicklung der Schüler/innen, nicht zum Gegenstand haben, sondern sich auf die Lehrpersonen und die Institutionen beschränken (vgl. z.B. den Überblick bei Lüders u.a. 2007). Untersuchungen, die dies Kriterium nicht einbeziehen, lassen sich im Grunde nur als Vorarbeiten betrachten.

⁴⁸ Nicht unwichtig ist es z.B., die um etwa 20% reduzierte Schülerstichprobe im Auge zu haben, welche die Generalisierbarkeit der Ergebnisse leicht einschränkt (es fehlen vor allem die schwächeren Schüler/innen wie die Hauptschüler/innen oder Sitzbleiber/innen, aber auch ein überdurchschnittlicher Anteil der Migrantenkinder). Darauf wird S. 147f. hingewiesen, der Frage aber nicht weiter nachgegangen.

der Qualität des Forschungsdesigns und der Datenanalysen⁴⁹ nun erstmals im deutschsprachigen Raum über belastbare Befunde zu erfolgreichem Lehren in der Schule in Gestalt von Auswirkungen der Lehrerkompetenzen auf die Lernergebnisse der Schüler/innen verfügen. Die Länge der Praxiserfahrung der Lehrer/innen ist dabei ohne Bedeutung (vgl. S. 158); hieraus kann man schließen, dass sowohl CK als auch PCK im Wesentlichen während der Ausbildungszeit der Lehrer/innen erworben wurden.⁵⁰

Von Interesse ist im Zusammenhang mit der Thematik des vorliegenden Aufsatzes der Bericht über die *differentiellen* Wirkungen des CK und PCK auf die Schülerleistungen, die sich in der Tat identifizieren ließen. Dabei konnte nur für das PCK – *teachers' domain-specific instructional knowledge* – ein *direkter* Effekt auf die Schülerleistungen nachgewiesen werden (vgl. S. 162). Dies bedeutet freilich nicht, dass das CK damit uninteressant würde. Im Gegenteil geht nicht nur aus anderen, in dem genannten Artikel detailliert besprochenen qualitativen Studien hervor, „that the repertoire of teaching strategies and the pool of alternative mathematical representations and explanations available to teachers in the classroom are largely dependent on the breadth and depth of their conceptual understanding of the subject“, so dass „an insufficient understanding of mathematical content limits teachers' capacity to explain and represent that content to students in a sense-making way, a deficit that cannot be offset by pedagogical skills“ (S. 138). Sondern die Autoren folgern auch aus den Analysen ihrer eigenen Daten, dass das CK eine *conditio sine qua non* für das PCK ist („PCK is inconceivable without CK“; S. 145). Zwar weist das CK trotz seiner hohen Korrelation mit dem PCK einen schwächeren Zusammenhang mit den Schülerleistungen auf, doch gibt es gleichsam die Grenzen für die Entwicklung des PCK vor.⁵¹ „Hohes Fachwissen ‚schützt‘ gewissermaßen vor einem niedrigen fachdidaktischen Wissensniveau“ (Krauss u.a. 2008b, S. 244).

⁴⁹ Methodisch handelt es sich bei der COACTIV-Studie um ein Quasi-Experiment, bei dem der im deutschen Lehrerbildungssystem vorfindbare Unterschied zwischen gymnasialer und nicht-gymnasialer Lehrerbildung als Gruppierungsmerkmal zugrunde gelegt werden konnte. Darüber hinaus ist die abhängige Variable Schülerleistung als *Leistungszuwachs* bestimmt. Vom methodischen Ansatz her lässt sich im gegebenen Forschungskontext ein besseres Forschungsdesign kaum vorstellen. Darüber hinaus hat sich bei systematischen Vergleichen von Befunden aus Experimenten gegenüber Quasi-Experimenten immer wieder herausgestellt, dass die Übereinstimmung beträchtlich ist, vgl. z.B. Kulik/Kulik (1989, S. 270) oder Seidel/Shavelson (2007, S. 480). Freilich gilt eine solche Übereinstimmung nicht automatisch, sondern muss im Einzelfall geprüft werden. – Einen besonderen Hinweis verdient auch der Ansatz der COACTIV-Studie, bei welchem die beiden vorfindbaren Typen der Lehrerbildung als „unique natural experiment“ wissenschaftlich fruchtbar gemacht und miteinander verglichen wurden. Beispiele für die Ergiebigkeit dieses Ansatzes finden sich auch in anderen Arbeiten, vgl. z.B. Baumert zur Koedukation (1992) oder Hopf u.a. (1999) zum Lehrer-geschlecht.

⁵⁰ Die Lehrer/innen waren bei der Untersuchung im Durchschnitt 47,2 Jahre alt (vgl. Krauss 2008b, S. 718), stellen also eine im Vergleich zu Blömeke u.a. 2008 deutlich ältere Stichprobe dar (Mathematikstudierende und Referendar/inn/e/n). Sie haben ihre Hochschul- und Referendarausbildung in den 80er-Jahren absolviert, in denen aber die Ausbildungsordnungen nicht einschneidend anders gewesen sind als 20 Jahre später.

⁵¹ Wichtig wäre die genauere Analyse der aus dem Rahmen fallenden, kleinen Gruppe von Lehrer/inn/e/n, die trotz niedrigem CK hohe PCK-Werte aufweisen (vgl. Krauss u.a.

Es gibt also inzwischen belastbare Forschungsbefunde, auf deren Grundlage man die Lehrerbildung überdenken sollte. Sie weisen darauf hin, dass die Fachkenntnisse der Lehrer/innen im unterrichteten Schulfach hoch bedeutsam sind; auf ihnen wiederum scheinen die fachdidaktisch-pädagogischen Kenntnisse aufzubauen, für die dann eine direkte Wirkung auf die Schülerleistungen nachweisbar ist. Man wird bei diesen Befunden nicht nur an die Ergebnisse von Seidel und Shavelson erinnert, welche die herausragende Rolle der schulfachspezifischen Aktivitäten herausgearbeitet haben, sondern auch an Baumert und Kunters Formulierung (2006, S. 496), „dass das tatsächlich im Unterricht verfügbare fachdidaktische Handlungsrepertoire von Lehrkräften weitgehend von der Breite und Tiefe ihres konzeptuellen Fachverständnisses abhängt. ... Fachwissen ist die Grundlage, auf der fachdidaktische Beweglichkeit entstehen kann“. Erinnert sei auch an den Befund von Wayne und Youngs (2006, S. 86): „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.“⁵²

Die genannten Befunde beruhen auf einer soliden Datenbasis und sind darüber hinaus plausibel. Sie erinnern an den Disput unter Philosophen und Rednern in der Antike, in welchem einerseits der Standpunkt des Sokrates viele Befürworter hatte, dass jedermann auf dem Gebiet, das er beherrsche, auch ein hinreichend guter Redner sei; andererseits wurde der *stultitia loquax*, der geschwätzigen Dummheit, dem Reden ohne Sachkenntnis also, die *indiserta prudentia*, das unberedete Expertentum, gegenübergestellt. Auch bei dieser Figur bestand aber Einigkeit über die *conditio sine qua non*: die Sachkenntnis. „Denn“, so Cicero, „es kann ja keine Meisterschaft im Reden geben, wenn der Redner den Gegenstand der Rede nicht beherrscht“.⁵³ *Rem tene, viae sequentur*, so könnte man also in Abwandlung der berühmten Formulierung des Cato major sagen.

Es liegt nahe, die genannten Kompetenzen gezielt zu fördern, d.h. also Anreize zu setzen, sie sich gründlich anzueignen und kontinuierlich zu verbessern; schulstufenspezifische⁵⁴ Argumente sind dabei zu berücksichtigen.

2008b, S. 244). In unserem Zusammenhang könnten insbesondere Hinweise auf die Ausbildung dieser Lehrer/innen aufklärend sein.

⁵² Vgl. auch Wayne u.a. 2008: „... it is generally accepted that intensive, sustained, job-embedded professional development focused on the content of the subject that teachers teach is more likely to improve teacher knowledge, classroom instruction, and student achievement.“ Siehe dort auch den Verweis auf Kennedy (1998). Vgl. auch Lipowsky (2006, S. 50ff.), der allerdings die referierten Studien nicht nach den verwendeten Forschungsmethoden gewichtet.

⁵³ *Dicendi enim virtus, nisi ei, qui dicet, ea, quae dicet, percepta sunt, exstare non potest.* (Cicero, *De Oratore* I, 48). Priorität gebührt dabei der Sachkenntnis: *Itaque praestantior est indiserta prudentia quam stultitia loquax.*

⁵⁴ Die meisten Studien beziehen sich auf die Sekundarstufe, nicht die Grundstufe des Schulwesens (vgl. auch Akiba u.a. 2007; Baumert/Kunter 2006). Für die Grundstufe liegen aber ebenfalls bemerkenswerte Forschungsbefunde vor, etwa aus dem britischen Längsschnittprojekt EPPE; vgl. z.B. Sammons u.a. 2008. Dieses Projekt geht allerdings

5. Schlussbemerkungen

Der Stand der *Forschung über Unterricht und Lehrerverhalten* (Kap. 2) lässt sich zusammenfassend so charakterisieren, dass wir zwar über eine hinreichende Zahl von Überblicken in verschiedenen Zeitperioden verfügen, aber aufgrund der meist fehlenden Trennung der Forschungsbefunde nach ihrer methodischen Qualität im Dunklen tapfen hinsichtlich ihrer Tragfähigkeit und Generalisierbarkeit. In letzter Zeit wird aber in den Arbeiten von Kulik und Kulik und ihrer Nachfolger/innen, insbesondere von Seidel und Shavelson, das bestehende methodische Problem weitgehend gelöst. Seidel und Shavelson betonen freilich, dass wir, wenn wir die methodischen Fragen gebührend ernst nehmen, erst über einen geringen Bestand an belastbaren Befunden verfügen, weil die ganz überwiegende Anzahl der Studien weder generalisierbar ist noch Ursache-Wirkungs-Zusammenhänge aufdeckt; diese stellen gleichwohl schon heute einen wertvollen Bestand dar, dessen sich künftige Forschung zumindest zur Bildung von zu prüfenden Hypothesen sinnvoll bedienen kann; für die Bewertung von Lehrer/innen und Unterricht und die Setzung von Anreizen oder Sanktionen sind sie jedoch nicht tragfähig genug.

Betrachtet man die seit dem Abschluss der Metaanalyse von Seidel und Shavelson erschienenen Arbeiten sowie die dort nicht einbezogenen früheren einschlägigen Studien, so ist das Resümee der beiden Autoren nach meiner Einschätzung zu pessimistisch. Insbesondere stellt die erwähnte COACTIV-Studie neuerdings belastbare Befunde auch zu solchen Eigenschaften und Verhaltensweisen der Lehrer/innen bereit, deren fördernde Auswirkungen auf die Schüler/innen hat gezeigt werden können. Aufgrund ihrer ökologischen Validität können diese Befunde leichter als die zuvor bekannten praxisnah gewendet werden. Ich komme auf diese Studie zurück.

Um die Forschungen über *Anreizsysteme für Lehrer/innen* (Kap. 3) steht es weniger gut. Das liegt vor allem daran, dass die oben beschriebenen Forschungslücken zum Unterricht und Lehrerhandeln die Möglichkeiten zur Untersuchung von Anreizen begrenzen; denn man muss ja zunächst wissen, was sich zu verstärken lohnt oder durch Sanktionen reduziert werden sollte. Immerhin treten auch hier neue Erkenntnisse aus belastbaren Untersuchungen zu Tage (z.B. im Kontext des Programms *No Child Left Behind*). Dar-

nicht der Frage der Bedeutung der schulfachlichen Kenntnisse der Lehrer/innen für die Schüler/innen nach. Die Studie ist aus vielen Gründen von Interesse, nicht zuletzt wegen der Beobachtung, dass es Übereinstimmung zwischen den in England üblichen offiziellen Schulinspektionen und den Ergebnissen dieser Längsschnittuntersuchungen gibt. In Spannung zu diesen Ergebnissen stehen nämlich die Befunde von Rosenthal (2004), der negative Auswirkungen der britischen Schulinspektionen auf die Schulen belegt hat (vgl. auch Ditton 2008, S. 45). Tiedemann u.a. (2007) haben für deutsche Grundschulen keinen Zusammenhang zwischen Fachkompetenz und Berufserfahrung der Lehrer/innen und den Fachleistungen der Schüler/innen (Mathematik, Lesen, Rechtschreibung) und ihrer Motivation entdecken können. Bei Slavin u.a. (2008) ergeben sich dagegen in ihrer umfassenden Metaanalyse klare Hinweise für lehrerinduzierten erfolgreichen Mathematikunterricht in der Grundstufe. Vgl. auch Hill u.a. (2005).

über hinaus ist es gelungen, im Rahmen eines ganz anderen Forschungsdesigns, das unabhängig ist von den Ergebnissen der traditionellen *teacher effectiveness* Forschung, durch (z.B. finanzielle) Anreize latente Kompetenzen von Lehrer/inne/n zu aktivieren und dadurch bessere Lernerfolge der Schüler/innen hervorzurufen; ohne ein vorgegebenes Treatment (vgl. z.B. Lavy; Muralidharan/Sundaraman).

Auch die *Lehrerbildungs-Forschung* (Kap. 4) liefert erst wenige belastbare Befunde, auf die man zurückgreifen könnte, wenn man Anreize setzen oder Sanktionen ausüben wollte, um die Lernergebnisse der Schüler/innen zu verbessern. Fachkompetenz, pädagogisch-didaktische Kompetenz, verbaler IQ, Faktoren dieser Art sind als wirksam identifiziert worden und bieten sich zumindest teilweise als Basis für eine Veränderung der Lehrerbildung an, die sich dann wiederum auf die Lernfolge der Schüler/innen auswirken würde. Vor allem kann man hier auf die oben genannten Studien zurückgreifen, die neuerdings zur Verfügung stehen (COACTIV und MT21). Da diese sich auf die Lehrerkompetenzen richten, welche im Wesentlichen während der Ausbildung in der ersten und zweiten Phase erworben wurden, werden sie hier der Lehrerbildung zugeordnet. Das Professionswissen von Lehrer/inne/n wird hier differenziert behandelt:

Was das Fachwissen betrifft – in den meisten belastbaren Studien geht es um das *mathematische* Fachwissen –, bedarf es einer genauen Bestimmung, was damit gemeint ist. Blömeke (vgl. 2008, Kap. 2 und 3) und die Publikationen der COACTIV-Studie (vgl. z.B. Baumert u.a. 2010, vor allem aber Krauss 2008b) machen diesbezüglich einleuchtende Vorschläge. So ist es durchaus überlegenswert, die fachmathematische Ausbildung der Lehrer/innen stärker als bisher auf die Inhalte der Schulmathematik zu konzentrieren (vertieftes Hintergrundwissen zur Schulmathematik), sofern diese mit dem auch sonst üblichen wissenschaftlichen Anspruch gelehrt werden; zumindest in kontrollierten Versuchen würde das Sinn machen. Man muss allerdings auch festhalten, dass der nachgewiesenen Überlegenheit des Fachwissens der Gymnasiallehrer/innen gegenüber den Nicht-Gymnasiallehrer/inne/n die übliche universitäre Mathematikausbildung zugrunde liegt; diese ist insoweit offensichtlich erfolgreich, auch ohne Beschränkung auf die Inhalte der Schulmathematik, und eine Umsteuerung könnte durchaus kontraproduktiv sein. Das Fachwissen spielt in der COACTIV-Studie – und nur in dieser werden deutsche Lehrerdaten mit Schülerleistungen in Zusammenhang gebracht – jedenfalls eine hoch bedeutsame Rolle⁵⁵ und kann als Voraussetzung für die Herausbildung methodisch-didaktischer Kompetenzen betrachtet werden.

⁵⁵ Vgl. z.B. Baumert/Kunter 2006, S. 496: „Fachwissen scheint eine notwendige, aber nicht hinreichende Bedingung für qualitativollen Unterricht und Lernfortschritte der Schülerinnen und Schüler zu sein“. – Baumert u.a. (2010, S. 136): „consensus ... that a strong knowledge of the subject taught is a core component of teacher competence“ mit Hinweis auf andere Literatur. – Baumert u.a. (2010, S. 163): „... learning support seems to be dependent not only on a caring ethos but also on domain-specific knowledge. The qualitative findings indicate that deficits in CK are to the detriment of PCK, limiting the scope for its development. Our findings suggest that it is not possible to off-

Hinsichtlich des fachdidaktischen und pädagogischen Wissens lohnt es, die Bestandteile des PCK bei Baumert u.a. (2010) zu betrachten, die sich als wirksam erwiesen haben (z.B. *effective classroom management*, *cognitive activation*, *individual learning support*), sowie die weitere empirische Ausdifferenzierung bei Blömeke u.a. (2008) zu berücksichtigen. Dabei geht es nicht nur um Fachdidaktik im engeren Sinne und um Lehrerverhalten im täglichen Umgang mit den Schüler/innen im Unterricht, sondern es werden auch Elemente der Lehrerbildung wie allgemeine Didaktik, Bildungssoziologie etc. zu diskutieren sein. Letztere haben derzeit ein erhebliches Gewicht in der Lehrerbildung, über die Auswirkungen solcher Wissensbestände auf Schule und Schüler/innen wissen wir allerdings nichts.

Studien, die diese Faktoren systematisch in der Lehrerbildung variieren, um ihre Wirkung auf die Schüler/innen besser zu verstehen, gibt es meines Wissens bisher nicht, sondern es werden im besten Fall vorhandene „natürliche Experimente“ ausgewertet, welche sich bei einer uneinheitlichen Lehrerbildung in einem nationalen Bildungssystem vorfinden.

Zum gegenwärtigen Zeitpunkt lässt sich also feststellen, dass es hierzulande weder evidenzbasiertes Lehren noch eine evidenzbasierte Lehreraus- und -fortbildung gibt, was angesichts des begrenzten gesicherten Wissens und seiner üblicherweise zögerlichen Umsetzung nicht verwunderlich ist. Es erschwert die Erforschung dieser Fragen, dass wir im deutschsprachigen Raum bislang nicht über genügend zuverlässige Messinstrumente für die uns interessierenden abhängigen Variablen, z.B. die Schülerleistungen⁵⁶, verfügen, so dass es stets sehr aufwendig ist, Effekte zu identifizieren.

In dieser Situation gibt es (mindestens) vier Handlungsmöglichkeiten:

set this relationship by increasing the specific focus on PCK in teacher training.“ – „This study shows that both CK and PCK are largely dependent on the type of training program attended, with program-specific differences in CK of more than one standard deviation. The qualitative findings indicate that deficits in CK are to the detriment of PCK, limiting the scope for its development. Our findings suggest that it is not possible to offset this relationship by increasing the specific focus on PCK in teacher training.“ (Baumert u.a. 2010, S. 167). – Baumert u.a. (2010, S. 167): „... tentative conclusions ... teacher training programs. It seems that teacher training programs that compromise on subject matter training ... have detrimental effects on PCK, and consequently negative effects on instructional quality and student progress. Differences in CK that emerge during preservice training persist across the entire teaching career.“ – Auf die notorische Unterschätzung der Bedeutung des Fachwissens in der Lehrerforschung verweisen Baumert/Kunter (vgl. 2006, S. 490).

⁵⁶ In Deutschland finden, im Unterschied zu den USA, im schulischen Alltag so gut wie keine standardisierten und validierten Schulleistungstests Verwendung. Daher ist es sowohl für die Forschung wie für die Praxis besonders schwierig, die Auswirkungen von Schule und Unterricht an Schüler-Daten festzumachen. Dies ist ein wichtiges Thema, das hier aber nicht genauer besprochen werden kann. Einen Einblick in die Vielfalt und die Bedeutung, aber auch in die hier auftauchenden Probleme geben z.B. Solórzano (2008), Kim/Sunderman (2005). Zu den neuerdings in Deutschland in Entwicklung befindlichen Qualitätsstandards vgl. z.B. Köller (2008).

1. Insbesondere durch die zusammenfassenden und frühere Arbeiten einbeziehenden Metaanalysen von Kulik und Kulik (1989) und Seidel und Shavelson (2007) sowie durch die weiterführenden Arbeiten z.B. von Baumert und Kunter (2006) verfügen wir gegenwärtig über einen theoretisch angeleiteten Überblick über die Forschung zur *Teacher Effectiveness*, der zahlreiche wichtige Ergebnisse enthält. Allerdings weisen alle mir bekannten Übersichten beklagenswerte Lücken bezüglich der Aufarbeitung des zum Zeitpunkt ihrer Abfassung erreichten Forschungsstandes auf, und es wäre auch beim heutigen Stand außerordentlich nützlich, wenn sich eine Gruppe von Wissenschaftler/inne/n daran machen würde, den Stand der Forschung endlich einmal vollständig und akribisch darzustellen.⁵⁷ Dazu ist eine erhebliche Anzahl kompetenter Personen erforderlich; ein Einzelner oder eine kleine Gruppe von Forscher/inne/n kann die Fülle der vorliegenden Arbeiten⁵⁸ nicht bewältigen, zumal die – unbedingt notwendigen – me-

⁵⁷ Um nur einige Beispiele für die genannte Unvollständigkeit zu geben: Kulik und Kulik (1989) erwähnen nicht die einschlägige Arbeit von Fraser, Walberg u.a. (1987), obwohl sie Literatur bis 1988 verarbeiten und die Arbeit von Fraser u.a. sogar in derselben Fachzeitschrift erschienen war. Auch tauchen Arbeiten von Gage u.a. bei ihnen nicht auf. Seidel und Shavelson (2007) erwähnen die höchst relevanten Arbeiten von Kulik und Kulik nicht, obwohl sie sich bei manchen ihrer Befunde mit Gewinn auf ganz ähnliche Ergebnisse hätten berufen können (Beispiel: die Ähnlichkeit der Resultate bei experimentellen und quasi-experimentellen Studien; oder die mögliche Bedeutung auch von Studien mit niedriger Effektstärke für die Praxis). Von den beklagenswerten Lücken bei Anderson (2004) war bereits die Rede. Baumert und Kunter (2006) nennen weder Kulik und Kulik (1989) noch Anderson (2004) noch Fraser u.a. (1987) noch Scheerens und Bosker (1997); letztere beiden sind dagegen zentral für Seidel und Shavelson (2007). Auch Helmke u.a. (1997) finden mit ihren Überblicken über Metaanalysen und ihren eigenen Aufarbeitungen der Literatur wenig Beachtung. Wang u.a. (1993) werden zwar mehrfach gelobt, aber offenbar seltener gelesen und zur Grundlage darauf aufbauender Literaturstudien gewählt. Wayne und Youngs (2006) scheinen Lavy nicht zu kennen. Cramer, Horn und Schweitzers Literaturbericht zur Lehrerbildungsforschung irritiert beispielsweise dadurch, dass Blömeke u.a. (2008) zwar genannt, nicht aber verarbeitet werden, obwohl sie für diese Autoren unumgehbare Forschungsergebnisse vorgelegt hatten.

⁵⁸ Zusätzlich zu den üblichen Recherchen sollte bei der Erarbeitung des Forschungsstandes auch der reiche Informationsbestand von Interesse sein, der unter der Überschrift „What Works in Education“ in den USA gesammelt, aufbereitet und mit Empfehlungen bzw. Warnungen publiziert wird; vgl. hierzu auch U.S. Department of Education (2003) mit Hinweis auf weitere Quellen im Internet, wo man entsprechende evidenzbasierte Studien finden kann. Zu beachten sind außerdem www.ed.gov/nclb sowie die bei Slavin gegebenen Hinweise auf weitere einschlägige Publikationen (Slavin 2008). – Unbedingt berücksichtigt werden muss auch die Weiterentwicklung des US-amerikanischen Projekts *No Child Left Behind* sowohl in der Praxis als auch in der reichhaltigen begleitenden Forschung. Das Projekt ist zusätzlich dadurch interessant, dass nicht die individuelle Lehrkraft das Ziel möglicher Sanktionen ist, sondern die Schule als Ganzes. Dies fördert offenbar gegenseitige Unterstützung im Lehrerkollegium, um die Lernziele der Schüler/innen besser zu erreichen und damit im Team die erstrebte Anerkennung zu erhalten (die darin besteht, nicht als defizitäre Schule klassifiziert zu werden, sondern professionell das Notwendige in befriedigendem Maße geleistet zu haben). Dass bei einem solchen Konzept auch hoher sozialer Druck auf die Lehrer/innen entstehen und zu unerwünschten Folgen führen kann, ist keineswegs auszuschließen; die entstehenden Berichte über das Projekt sollten auch hierauf geprüft werden. Außerdem muss offensichtlich dafür Sorge getragen werden, dass unterschiedlich belastete Schulen differenziell betrachtet und behandelt werden (vgl. die interessante und konstruktive Diskussion zu

thodischen Differenzierungen vorgenommen und Fragen der Validität, Generalisierbarkeit und Übertragbarkeit beachtet werden müssen.⁵⁹ Beispiel für eine vorbildliche Arbeit der benötigten Art ist die vollständige Durchsicht der Studien zur Grundschul- bzw. Sekundarschulmathematik in den USA unter Berücksichtigung der erforderlichen methodischen Standards durch Slavín u.a. (2008 bzw. 2009).

2. Es besteht Aussicht auf relativ schnelle und zentrale Erkenntnisgewinne, wenn Forschungen wie beispielsweise die im Rahmen des DFG-Projektes COACTIV intensiv weiterverfolgt werden. Dies kann auch geschehen, bevor der Stand der Forschung abschließend erarbeitet ist, da in solchen Projekten neuartige Ansätze verfolgt werden, für welche die bisherige Forschung nur partielle Relevanz besitzt. Wenn möglichst viele Untersuchungen im deutschsprachigen Bereich erfolgen, kann zudem die stets heikle Frage der Übertragbarkeit von Forschungsergebnissen besser beantwortet werden. Von hoher Bedeutung ist auch die Ausdehnung dieserart Forschung auf andere Schulfächer sowie auf weitere Variablen als die bisher erforschten, welche ja nur einen Teil der Varianz der Schülerleistungen erklären (vgl. auch Baumert u.a. 2010).

Was man in der bisherigen Forschungsliteratur ebenfalls vermisst, ist der Versuch, als wirksam erwiesene Variablen miteinander zu verknüpfen und kombinierte Auswirkungen zu untersuchen.⁶⁰ Denn in dem hoch komplexen Geflecht der in Schule und Unterricht wirkenden Variablen genügt es nicht, eine einzelne Variable zu isolieren und mit einer anderen in Beziehung zu setzen, sondern man muss der existierenden Komplexität durch angemessene Forschungsdesigns und statistische Verfahren Rechnung tragen. Dieser Forderung wurde in zahlreichen der bisherigen Studien der Lehr-Lern-Forschung nicht Genüge getan, überstieg auch für lange Zeit die Leistungsfähigkeit des verfügbaren Forschungsinstrumentariums. Neben anderen ha-

dieser Frage bei Kim/Sunderman (2005)). Kritik an dem Projekt wegen möglicher Verletzung der Chancengleichheit findet sich ebenfalls in dem interessanten Aufsatz von Kim u.a. (2005); enttäuschend dagegen die Arbeit von Bellmann (2006).

⁵⁹ Hier sollte man von vornherein streng verfahren, damit sich nicht Entwicklungen wiederholen, wie man sie z.B. beim Umgang mit den Ergebnissen der Bilingual-Forschung beobachten konnte. Dort folgte auf eine frühe Metaanalyse von Rossell und Baker (1996) eine scharfe methodisch begründete Kritik von Greene, der von den ursprünglich 72 als tragfähig erklärten Studien nur noch 11 akzeptiert; von diesen haben Limbird und Stanat (2006) weitere begründet disqualifiziert. Sinnvoll ist es daher, von vornherein eine strenge Auswahl zu treffen und um der Sache willen die Kollegialität nicht zu weit zu treiben. Andernfalls ergeben sich unnötige Verzögerungen bei der Beantwortung drängender Fragen.

⁶⁰ Vgl. hierzu auch Rice (2003). – Selbst Studien, die das Geschlecht der Lehrkraft berücksichtigen, besitzen Seltenheitswert: Informationen darüber, ob Lehrerinnen und Lehrer unterschiedliche Wirkungen auf Schülerinnen und Schüler ausüben oder ob für sie verschiedene Formen von Bestätigungen und Anreizen angemessen sind, fehlen weitgehend (vgl. hierzu z.B. Hopf/Hatzichristou 1999). Lavy (2008) konnte immerhin zeigen, dass die Steigerung der Schülerleistungen, die durch finanzielle Anreize der Lehrer/innen in Gang gesetzt wurde, nicht nach Geschlecht unterschiedlich ausfiel; lediglich die Überzeugung, etwas bei den Schüler/inne/n bewirken zu können, war bei den weiblichen Lehrerinnen schwächer ausgeprägt.

ben Helmke und Weinert (1997) die Bedeutung der Verflochtenheit der wirkenden Kräfte herausgearbeitet und Wege für eine angemessene Vorgehensweise aufgezeigt.

Bei der Suche nach potentiell einflussreichen Variablen für neue Untersuchungen lassen sich gut die zahlreich vorhandenen korrelativen Studien nutzen, die plausibel machen, was Lehrer/innen sinnvollerweise tun bzw. unterlassen sollten, um die Lernerfolge ihrer Schüler/innen zu verbessern bzw. nicht zu gefährden. Hierzu gehören auch die neueren, kompetent gemachten Surveys (TIMSS, PISA etc.), die ebenfalls wertvolle Anregungen liefern, auch hinsichtlich schulstruktureller und sozialpsychologischer Faktoren.⁶¹

Es wäre zu hoffen, dass die Prognose von Seidel und Shavelson, erst in weiteren zwñhn Jahren werde man wohl genügend experimentelle Untersuchungen antreffen, um eine neue Metaanalyse zu füllen (vgl. S. 464), nicht eintrifft.

3. So sehr sich die methodologische Grundeinstellung von Sozialwissenschaftler/inne/n auch dagegen sträuben mag, sollten doch auch Studien wie die oben beschriebenen, in welchen den Lehrer/inne/n keine Treatments empfohlen werden, sondern ihnen freie Hand gelassen wird, in Angriff genommen werden, weil sie schnelle und vielfach auch unmittelbar praxisrelevante Ergebnisse erbringen können (*black-box-Paradigma*). Sicher wird man dabei z.B. über Lavy hinausgehen müssen, indem man die veränderten Verhaltensweisen der erfolgreichen Lehrer/innen nicht nur ex post und oberflächlich per Telefoninterview abfragt (vgl. Lavy 2008), sondern sich möglichst präzise vergewissert, welche ihrer Kompetenzen Lehrer/innen aktiviert haben, um die Schüler/innen zu fördern.⁶² Die Vorgehensweise besteht dabei, vereinfacht ausgedrückt, darin, Anreize, Anerkennungen, Belohnungen für definierte Lernerfolge der Schüler/innen festzulegen, ohne den Lehrer/inne/n Vorschriften für ihr Verhalten zu machen, es also ihrer Professionalität und Erfahrung zu überlassen, auf welche Weise sie die Lernergebnisse ihrer Schüler/innen verbessern. Die dabei gewonnenen Erkenntnisse können dann in einem nächsten Schritt auch als Forschungshypothesen genutzt werden.

⁶¹ Beispielsweise haben Baumert, Stanat und Watermann (2006) nachgewiesen, dass Schulen unter höchst unterschiedlichen Bedingungen arbeiten; bei Anreizen oder Sanktionen müssten daher differentielle Kriterien angewendet werden. Vgl. auch die britische Diskussion um die so genannten *Educational Priority Areas*, z.B. Halsey (1972). Von zentraler Bedeutung ist hier auch der Befund der COACTIV-Studie, dass ähnlich wie in den USA die Lehrer/innen in den nicht-akademischen Schulen über ein geringeres Fachwissen verfügen und dass darüber hinaus das PCK stärkere Wirkungen auf Schüler/innen mit niedrigerem Sozialstatus und niedriger Leistung haben (vgl. Baumert u.a. 2010, S. 147 und S. 158). „Weaker students from lower SES families and immigrant families tend to be taught by teachers who are less competent in terms of CK and PCK“ (S. 168).

⁶² Ansätze zu genaueren Erhebungen finden sich bei Muralidharan/Sundaraman (2008).

Ein weiteres Argument für die Option, auch in nationalen Studien auf dies methodische Paradigma zu setzen, leitet sich schließlich ab aus der immer wieder neu zu beantwortenden Frage, ob man in anderen Schulsystemen gewonnene Befunde – auch wenn sie belastbar und dort generalisierbar sind – auf die hiesigen Verhältnisse übertragen kann. Auch Prentice u.a. (vgl. 2007, S. 38) erinnern bei ihrer Übersicht über Studien zur *Performance Pay* an die Problematik der externen Validität.⁶³

4. Sobald umfassendere gesicherte Befunde vorliegen, lassen sich diese in die Lehrerbildung und die Lehrerweiterbildung einbringen; wobei insbesondere letztere nach Auskunft der Metaanalyse von Kulik und Kulik zu positiven Auswirkungen auf die Schüler/innen führt.⁶⁴

In der Tat können aber bereits heute auf der Grundlage der Erkenntnisse aus den oben zitierten Studien Elemente einer evidenzbasierten Lehrerbildung tentativ formuliert werden:

- In der ersten, universitären Phase der Lehrerbildung wäre danach der Aufbau einer möglichst soliden und differenzierten Fachkenntnis in den Schulfächern ins Zentrum zu rücken. Denn erst gute Fachkenntnisse ermöglichen ein flexibles und variantenreiches didaktisches Verhalten. Was dies (fach-)inhaltlich bedeutet, ist an anderer Stelle andiskutiert worden. – Der Erwerb des Fachwissens ist unter den gegebenen Umständen sicherlich in der ersten Phase der Lehrerbildung zu verorten, wie die zitierten Studien ja auch zu Recht davon ausgehen, dass das Fachwissen während der Hochschuljahre gewonnen wird. Die Länge der Praxiserfahrung in der Schule ist hierfür irrelevant.

- Die Entwicklung des pädagogisch-didaktischen Wissens ließe sich schwerpunktmäßig in die zweite Phase der Lehrerbildung (Referendariat) einbinden, wo auch unmittelbare Möglichkeiten zur Einübung gegeben sind. Hauptbestandteil des zu erwerbenden Repertoires müssten dann die Fachdidaktik im engeren Sinne sowie die allgemeindidaktischen Verhaltenspotentiale wie *classroom mangement*, *cognitive activation*, *individual learning support* etc. sein. Microteaching, welches sich als Lernform bei Lehrer/inne/n nachweislich bewährt hat, wird dabei sicher eine Rolle spielen, insbesondere wenn es sich einmal als wirkungsvoll auch bis hin zu den Lernergebnissen der Schüler/innen erwiesen haben sollte. Von großer Bedeutung ist gerade in diesem Bereich des pädagogisch-didaktischen Wissens der Befund, dass die Dauer der Unterrichtserfahrung für seine Ausprägung irrelevant ist, man also nicht davon ausgehen kann, dass Lehrer/innen mit wachsendem Dienstalter automatisch besser werden.

⁶³ „Care must be taken when generalising the results of individual studies across other countries ...“.

⁶⁴ Zu den nachgewiesenen Effekten der Lehrerweiterbildung vgl. auch Wayne u.a. (2008) mit Verweis auf den sorgfältigen Literaturbericht von Yoon u.a. (2007).

- Der Stellenwert der allgemeinen Erziehungswissenschaft ist bislang unbestimmt, was die Auswirkungen auf die Schüler/innen betrifft. Sicher gehören Themen wie beispielsweise: soziale Ungleichheit; kindliche Entwicklung; historische Formen von Erziehung und Unterricht; Heterogenität; Grundlagen der Diagnostik, zum Grundlagenwissen eines Pädagogen und einer Pädagogin. Was solches allgemeinpädagogische Wissen jedoch für die Förderung der Schüler/innen bedeutet, wissen wir nicht.⁶⁵ Solange es keine Untersuchungen über die Auswirkungen solcher Wissensbestände auf die Schüler/innen gibt, wird man durch philosophische Reflexion ihren Stellenwert beschreiben müssen. Dabei könnte sich herausstellen, dass sie am besten an den Hochschulen in einem Studium Generale für alle künftigen Pädagog/inn/en aufgehoben sind.
- Neben der Lehrerbildung in Universität und Referendariat könnte eine gleichzeitige Lehrerweiterbildung schon praktizierenden Lehrer/inne/n das neu gewonnene, evidenzbasierte Wissen zugänglich machen, so dass es ohne langen Vorlauf Eingang in Schule und Unterricht findet.

Anreize unterschiedlicher Art hätten in allen genannten Zusammenhängen und Phasen die Funktion, die Kompetenzentwicklung der Lehrer/innen zu unterstützen. Letzten Endes dienen sie dazu, die Entwicklung der Lernergebnisse der Schüler/innen zu fördern.

⁶⁵ Teile der von Blömeke u.a. (2008) als fächerübergreifendes erziehungswissenschaftliches Wissen beschriebenen Bestände gehören nach der Definition von COACTIV dort zum pädagogisch-didaktischen Wissen.

6. Literatur

- Akiba, M./LeTendre, G.K./Scribner, J.P. (2007): Teacher Quality, Opportunity Gap, and National Achievement in 46 Countries. In: *Educational Researcher* 36, H. 7, S. 369-387.
- Anderson, L.W. (2004): *Increasing Teacher Effectiveness*. Paris: UNESCO, International Institute for Educational Planning.
- Baumert, J. (1992): Koedukation oder Geschlechtertrennung. In: *Zeitschrift für Pädagogik* 38, H. 1, S. 83-110.
- Baumert, J./Bos, W./Lehmann, R. (2000): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Band 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen: Leske + Budrich.
- Baumert, J./Artelt, C./Klieme, E. u.a. (Hrsg.) (2003): *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- Baumert, J./Stanat, P./Watermann, R. (2006): Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In: Baumert, J./Stanat, P./Watermann, R. (Hrsg.): *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000*. Wiesbaden: VS, Kap. 4, S. 95-188.
- Baumert, J./Kunter, M. (2006): Stichwort: Professionelle Kompetenz von Lehrkräften. In: *Zeitschrift für Erziehungswissenschaft* 9, H. 4, S. 469-520.
- Baumert, J./Kunter, M./Blum, W./Brunner, M./Voss, T./Jordan, A./Klusmann, U./Krauss, S./Neubrand, M./Tsai, Y.-M. (2010): Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. In: *American Educational Research Journal* 47, S. 133-180.
- Bellmann, J. (2006): Bildungsforschung und Bildungspolitik im Zeitalter „Neuer Steuerung“. In: *Zeitschrift für Pädagogik* 52, H. 4, S. 487-504.
- Bennett, N. (1976): *Teaching Styles and Pupil Progress*. Cambridge/Mass.: Harvard University Press 1976 (Deutsch: *Unterrichtsstil und Schülerleistung*. Stuttgart: Klett 1979).
- Biddle, B.J./Ellena, W.J. (1964): *Contemporary Research on Teacher Effectiveness*. New York: Holt, Rinehart & Winston.
- Blömeke, S./Kaiser, G./Lehmann, R. (Hrsg.) (2008): *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Münster: Waxmann.
- Borko, H./Putnam, R.T. (1996): Learning to Teach. In: Berliner, D.C./Calfee, R.C. (Hrsg.): *Handbook of Educational Psychology*. New York: Macmillan, S. 673-708.
- Carroll, J.B. (1963): A Model of School Learning. In: *Teachers College Record* 64, H. 8, S. 723-733.
- Carroll, J.B. (1989): The Carroll Model. A 25-Year Retrospective and Prospective View. In: *Educational Researcher* 18, H. 1, S. 26-31.
- Cochran-Smith, M./Zeichner, K.M. (Hrsg.) (2005): *Studying Teacher Education. The Report of the AERA Panel on Research and Teacher Education*. Washington.

- Cramer, C./Horn, K.-P./Schweitzer, F. (2009): Zur Bedeutsamkeit von Ausbildungskomponenten des Lehramtsstudiums im Urteil von Erstsemestern. Erste Ergebnisse der Studie „Entwicklung Lehramtsstudierender im Kontext institutioneller Rahmenbedingungen“ (ELKiR). In: Zeitschrift für Pädagogik 55, H. 5, S. 761-780.
- Darling-Hammond, L./Youngs, P. (2002): Defining “Highly Qualified Teachers”: What Does “Scientifically-Based Research” Actually Tell Us? In: Educational Researcher 31, H. 9, S. 13-25.
- Darling-Hammond, L. (2007): The Flat Earth and Education: How America’s Commitment to Equity Will Determine Our Future. Third Annual Brown Lecture in Education Research. In: Educational Researcher 36, H. 6, S. 318-334.
- Ditton, H. (2008): Qualitätssicherung in Schulen. In: Klieme, E./Tippelt, R. (Hrsg.): Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz. In: Zeitschrift für Pädagogik, 53. Beiheft, Weinheim: Beltz, S. 36-58.
- Düker, H./Tausch, R. (1957): Über die Wirkung der Veranschaulichung von Unterrichtsstoffen auf das Behalten. In: Zeitschrift für experimentelle und angewandte Psychologie 4, S. 384-400.
- Ebel, R.L. (Hrsg.) (1969): Encyclopedia of Educational Research. American Educational Research Association. 4th Edition. Toronto: The Macmillan Company.
- Fraser, B.J./Walberg, H.J./Welch, W.W./Hattie, J.A. (1987): Synthesis of Educational Productivity Research. In: International Journal of Educational Research 11, S. 145-252.
- Frey, A. (2006): Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften – eine erste Standortbestimmung zu bereits publizierten Instrumenten. In: Zeitschrift für Pädagogik, 51. Beiheft, S. 30-46.
- Fuchs, T./Woessmann, L. (2004): What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data. München: CESifo Working Paper No. 1235.
- Gage, N.L./Berliner, D.C. (1979): Pädagogische Psychologie. Band 1: Grundlagen, Konzepte, Ergebnisse. Band 2: Lehrmethoden, Bewertung des Lernerfolge. 2. Aufl. München: Urban & Schwarzenberg.
- Gehrmann, A. (2007): Kompetenzentwicklung im Lehramtsstudium. In: Lüders, M./Wissinger, J. (Hrsg.): Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation. Münster: Waxmann.
- Gläser-Zikuda, M./Seifried, J. (Hrsg.) (2008): Lehrerexpertise. Analyse und Bedeutung unterrichtlichen Handelns. Münster: Waxmann.
- Good, T.L./Biddle, B.J./Brophy, J.E. (1975): Teachers Make a Difference. New York: Holt, Rinehart & Winston.
- Haertel, G.D./Walberg, H.J./Weinstein, T. (1983): Psychological Models of Educational Performance: A Theoretical Synthesis of Constructs. In: Review of Educational Research 53, H. 1, S. 75-91.
- Halsey, A.H. (Hrsg.) (1972): Educational Priority. Vol. 1: E.P.A. Problems and Policies. Report of a Research Project. Department of Education and Science. London: Her Majesty’s Stationery Office.
- Helmke, A./Weinert, F.E. (1997): Bedingungsfaktoren schulischer Leistung. In: Weinert, F.E. (Hrsg.): Psychologie des Unterrichts und der Schule, Band 3. Göttingen: Hogrefe, S. 71-176.
- Hess, F.M./Petrilli, M.J. (2007): No Child Left Behind Primer. New York.
- Hill, H.C./Rowan, B./Ball, D. (2005): Effects of Teachers’ Mathematical Knowledge for Teaching on Student Achievement. In: American Educational Research Journal 42, H. 2, S. 371-406.

- Hopf, D. (1980): *Mathematikunterricht. Eine empirische Untersuchung zur Didaktik und Unterrichtsmethode in der 7. Klasse des Gymnasiums*. Stuttgart: Klett-Cotta.
- Hopf, D./Hatzichristou, C. (1999): *Teacher Gender-Related Influences in Greek Schools*. In: *British Journal of Educational Psychology* 68, H. 1, S. 1-18.
- Hopf, D. (2005): *Zweisprachigkeit und Schulleistung bei Migrantenkindern*. In: *Zeitschrift für Pädagogik* 51, H. 2, S. 236-251.
- Hopf, D. (2007): *Mehrsprachigkeit und Schulerfolg bei Schülern mit Migrationshintergrund*. In: *Sprache – Stimme – Gehör* 31, H. 4, S. 163-169.
- Jarausch, K.H. (2003): *Professor Unrat*. In: François, E./Schulze von Beck, H. (Hrsg.): *Deutsche Erinnerungsorte, Band II*. München: Beck, S. 315-331.
- Jencks, C.S./Smith, M./Acland, H./Bane, M.J./Cohen, D./Gintis, H./Heyns, B./Michelson, S. (1972): *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books.
- Jürges, H./Schneider, K. (2007): *Leistungsgerechtere Bezahlung von Lehrern: Sinnvolle Option oder sinnlose Ökonomisierung?* In: *Recht der Jugend und des Bildungswesens* 55, H. 1, S. 83-97.
- Jürges, H./Schneider, K. (2008): *Ressourcen und Anreize im Bildungswesen. Aufgaben und Handlungsmöglichkeiten des Staates aus Sicht der Bildungsökonomik*. In: *Zeitschrift für Erziehungswissenschaft* 11, H. 2, S. 234-252.
- Karweit, N. (1989): *Time and Learning: A Review*. In: Slavin, R.E. (Hrsg.): *School and Classroom Organization*. Hillsdale: Lawrence Erlbaum Associates, S. 69-95.
- Kennedy, M. (1998): *Form and Substance of Inservice Teacher Education*. Research Monograph No. 13. Madison: University of Wisconsin-Madison, National Institute for Science Education.
- Kim, J.S./Sunderman, G.L. (2005): *Measuring Academic Proficiency Under No Child Left Behind Act: Implications for Educational Equity*. In: *Educational Researcher* 34, H. 8, S. 3-13.
- Kölller, O. (2008): *Bildungsstandards in einem Gesamtsystem der Qualitätssicherung im allgemein bildenden Schulsystem Deutschlands*. In: Klieme, E./Tippelt, R. (Hrsg.): *Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz*. *Zeitschrift für Pädagogik*, 53. Beiheft, Weinheim: Beltz, S. 59-75.
- Krauss, S./Brunner, M./Kunter, M./Baumert, J./Blum, W./Neubrand, M./Jordan, A. (2008a): *Pedagogical Content Knowledge and Content Knowledge of Secondary Mathematics Teachers*. In: *Journal of Educational Psychology* 100, H. 3, S. 716-725.
- Krauss, S./Neubrand, M./Blum, W./Baumert, J./Brunner, M./Kunter, M./Jordan, A. (2008b): *Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie*. In: *Journal für Mathematik-Didaktik* 29, H. 3/4, S. 223-258.
- Kulik, J.A./Kulik, C.-L.C. (1989): *Meta-Analysis in Education*. In: *International Journal of Educational Research* 13, H. 3, S. 221-340.
- Kunter, M./Dubberke, T./Baumert, J./Blum, W./Brunner, M./Jordan, A./Klusmann, U./Krauss, S./Löwen, K./Neubrand, M./Tsai, Y.-M. (2006): *Mathematikunterricht in den PISA-Klassen 2004: Rahmenbedingungen, Formen und Lehr-Lernprozesse*. In: *PISA-Konsortium Deutschland (Hrsg.): PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann, Kap. 7, S. 161-194.
- Kupermintz, H. (2002): *Teacher Effects as a Measure of Teacher Effectiveness: Construct Validity Considerations in TVAAS (Tennessee Value-Added Assessment System)*. Los Angeles: University of California.

- Lavy, V. (2003): *Paying for Performance. The Effect of Individual Financial Incentives on Teachers' Productivity and Students' Scholastic Outcomes*. Jerusalem: Hebrew University.
- Lavy, V. (2007): *Using Performance-Based Pay to Improve the Quality of Teachers*. In: *The Future of Children* 17, H. 1, S. 87-109.
- Limbird, C./Stanat, P. (2006): *Sprachförderung bei Schülerinnen und Schülern mit Migrationshintergrund: Ansätze und ihre Wirksamkeit*. In: Baumert, J./Stanat, P./Watermann, R. (Hrsg.): *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000*. Wiesbaden: VS, Kap. 7, S. 257-307.
- Lipowsky, F. (2006): *Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler*. In: *Zeitschrift für Pädagogik*, 51. Beiheft, Weinheim: Beltz, S. 47-70.
- Lüders, M./Wissinger, J. (Hrsg.) (2007): *Forschung zur Lehrerbildung: Kompetenzentwicklung und Programmevaluation*. Münster: Waxmann.
- Marsden, D./Belfield, R. (2006): *Pay for Performance Where Output is Hard to Measure: the Case of Performance Pay for School Teachers*. In: *Advances in Industrial and Labor Relations*. London: JAI Press, S. 1-34.
- Masuch, S. (2010, in Vorbereitung): *Geld, Zeit oder Blumen? Eine empirische Erkundung der Ansichten von Schulleiter/inne/n und Lehrer/inne/n über Anreizsysteme im Lehrerberuf*.
- McCourt, F. (2005): *Teacher Man. A Memoir*. New York: Scribner.
- Mintrop, H./Sunderman, G.L. (2009): *Predictable Failure of Federal Sanctions-Driven Accountability for School Improvement – And Why We May Retain It Anyway*. In: *Educational Researcher* 38, H. 5, S. 353-364.
- Morkel, A. (2008): *Das Zerwürfnis zwischen Zunge und Verstand. Anmerkungen zu einer These von Cicero*. In: *Forschung & Lehre*, H. 12, S. 838-841.
- Muralidharan, K./Sundaraman, V. (2008): *Teacher Performance Pay: Experimental Evidence from India*. Paper presented at CESifo Conference Economic Incentives: Do They Work in Education? Insights and Findings from Behavioral Research. München.
- NICHD Early Child Care Research Network (2006): *Child-Care Effect Sizes for the NICHD Study of Early Child Care and Youth Development*. In: *American Psychologist* 61, H. 2, S. 99-116.
- Oser, F. (1997a): *Standards in der Lehrerbildung. Teil 1: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen*. In: *Beiträge zur Lehrerbildung* 15, H. 1, S. 26-37.
- Oser, F. (1997b): *Standards in der Lehrerbildung. Teil 2: Wie werden Standards in der schweizerischen Lehrerbildung erworben? Erste empirische Ergebnisse*. In: *Beiträge zur Lehrerbildung* 15, H. 2, S. 210-228.
- Oser, F./Oelkers, J. (2001): *Die Wirksamkeit der Lehrerbildungssysteme. Von der Allrounderbildung zur Ausbildung professioneller Standards*. Zürich: Rüegger.
- Oser, F./Renold, U. (2005): *Kompetenzen von Lehrpersonen – über das Auffinden von Standards und ihre Messung*. In: Gogolin, I. u.a. (Hrsg.): *Standards und Standardisierungen in der Erziehungswissenschaft. Zeitschrift für Erziehungswissenschaft, Beiheft 4*, S. 119-140.
- Prentice, G./Burgess, S./Propper, C. (2007): *Performance Pay in the Public Sector: A Review of the Issues and Evidence*. URL: <http://www.ome.uk.com/downloads; Zugriffsdatum: 21.04.2008>.

- Prenzel, M./Baumert, J./Blum, W./Lehmann, R./Leutner, D./Neubrand, M./Pekrun, R./Rost, J./Schiefele, U. (Hrsg.) (2005): PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche? Münster: Waxmann.
- Prenzel, M./Baumert, J./Blum, W./Lehmann, R./Leutner, D./Neubrand, M./Pekrun, R./Rost, J./Schiefele, U. (Hrsg.) (2006): PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres. Münster: Waxmann.
- Rauin, U./Maier, U. (2007): Subjektive Einschätzungen des Kompetenzerwerbs in der Lehramtsausbildung. In: Lüders, M./Wissinger, J. (Hrsg.): Forschung zu Lehrerbildung. Münster: Waxmann, S. 103-133.
- Reichwein Zientek, L./Capraro, M.M./Capraro, R.M. (2008): Reporting Practices in Quantitative Teacher Education Research: One Look at the Evidence Cited in the AERA Panel Report. In: Educational Researcher 37, H. 4, S. 208-216.
- Rice, J.K. (2003): Teacher Quality. Understanding the Effectiveness of Teacher Attributes. Economic Policy Institute (EPI).
- Rockoff, L. (2004): The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. In: American Economic Review 94, H. 2, S. 247-252.
- Rosenthal, L. (2004): Do School Inspections Improve School Quality? Ofsted Inspections and School Examination Results in the UK. In: Economics of Education Review 23, H. 2, S. 143-151.
- Sammons, P./Taggart, B./Siraj-Blatchford, I./Sylva, K./Melhuish, E./Barreau, S. (2006): The Effective Pre-School and Primary Education 3-11 Project (EPPE 3-11): Summary Report – Variations in Teacher and Pupil Behaviours in Year 5 Classes. London: DfES/Institute of Education, University of London.
- Sammons, P./Sylva, K./Melhuish, E./Siraj-Blatchford, I./Taggart, B./Barreau, S./Grabbe, Y. (2008): The Influence of School and Teaching Quality on Children's Progress in Primary School. Effective Pre-School and Primary Education 3-11 Project (EPPE 3-11). Report No. DCSF-RR028. Nottingham: DfES Publications.
- Scheerens, J./Bosker, R.J. (1997): The Foundations of Educational Effectiveness. Oxford: Pergamon.
- Schneider, B./Carnoy, M./Kilpatrick, J./Schmidt, W.H./Shavelson, R.J. (2007): Estimating Causal Effects. Washington: AERA.
- Seidel, T./Shavelson, R.J. (2007): Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. In: Review of Educational Research 77, H. 4, S. 454-499.
- Senatsverwaltung für Bildung, Jugend und Sport (2006a): Materialien zur Schulinspektion. Ausgabe für Grundschulen. 1. Fassung. Berlin. 76 S.
- Senatsverwaltung für Bildung, Jugend und Sport (2006b): Materialien zur Schulinspektion. Ausgabe für weiterführende allgemein bildende Schulen, berufliche Schulen, Schulen mit sonderpädagogischem Förderschwerpunkt. 1. Fassung. Berlin. 78 S.
- Senatsverwaltung für Bildung, Wissenschaft und Forschung (2008): Bildung in Berlin. Schulinspektionen im Schuljahr 2006/2007. Berlin.
- Shulman, L.S. (1987): Knowledge and Teaching: Foundations of the New Reform. In: Harvard Educational Review 57, S. 1-22.
- Slavin, Robert E. (2002): Evidence-based Education Policies: Transforming Educational Practice and Research. In: Educational Researcher 31, H. 7, S. 15-21.
- Slavin, Robert E. (2008): Perspectives on Evidence-Based Research in Education. What works? Issues in Synthesizing Educational Program Evaluations. In: Educational Researcher 37, H. 1, S. 5-14.

- Slavin, R.E./Lake, C. (2008): Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis. In: *Review of Educational Research* 78, H. 3, S. 427-515.
- Slavin, R.E./Lake, C./Groff, C. (2009): Effective Programs in Middle and High School Mathematics: A Best-Evidence Synthesis. In: *Review of Educational Research* 79, H. 2, S. 839-911.
- Söhn, J. (2005): Zweisprachiger Schulunterricht für Migrantenkinder. Ergebnisse der Evaluationsforschung zu seinen Auswirkungen auf Zweitspracherwerb und Schulerfolg. Berlin: Wissenschaftszentrum für Sozialforschung (WZB), Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI).
- Solórzano, R.W. (2008): High Stakes Testing. Issues, Implications, and Remedies for English Language Learners. In: *Review of Educational Research* 48, H. 2, S. 260-339.
- Terhart, E. (Hrsg.) (2000): Perspektiven der Lehrerbildung in Deutschland. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Kommission. Weinheim/Basel: Beltz.
- Terhart, E. (2007): Erfassung und Beurteilung der beruflichen Kompetenzen von Lehrkräften. In: Lüders, M./Wissinger, J. (Hrsg.): *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation*. Münster: Waxmann, S. 37-62.
- Tiedemann, J./Billmann-Mahecha, E. (2007): Macht das Fachstudium einen Unterschied? Zur Rolle der Lehrerexpertise für Lernerfolg und Motivation in der Grundschule. In: *Zeitschrift für Pädagogik* 53, H. 1, S. 58-73.
- U.S. Department, Institute of Education Sciences (IES) (2002ff.): *What Works. Clearinghouse (WWC)*.
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (2003): *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*. Washington D.C.
- Wang, M.C./Haertel, G.D./Walberg, H.J. (1993): Toward a Knowledge Base for School Learning. In: *Review of Educational Research* 63, H. 3, S. 249-294.
- Wayne, A.J./Youngs, P. (2006): Die Art der Ausbildung von Lehrern und die Lerngewinne ihrer Schüler. Eine Übersicht über aktuelle empirische Forschung. In: *Zeitschrift für Pädagogik*, 51. Beiheft, S. 71-98 (Original: Wayne, A.J./Youngs, P. [2003]: *Teacher Characteristics and Student Achievement Gains: A Review*. In: *Review of Educational Research* 73, H. 1, S. 89-122).
- Wayne, A.J./Yoon, K.S./Zhu, P./Cronen, S./Garet, M.S. (2008): Experimenting With Teacher Professional Development: Motives and Methods. In: *Educational Researcher* 37, H. 8, S. 469-479.
- Weinert, F.E. (1989): Psychologische Orientierungen in der Pädagogik. In: Röhrs, H./Scheuerl, H. (Hrsg.): *Richtungsstreit in der Erziehungswissenschaft und pädagogische Verständigung*. Frankfurt a.M./Bern: Lang, S. 203-214.
- Weinert, F.E. (1996): Lerntheorien und Instruktionsmodelle. In: Weinert, F.E. (Hrsg.): *Psychologie des Lernens und der Instruktion*. Enzyklopädie der Psychologie, Serie I. Pädagogische Psychologie, Band 2. Göttingen: Hogrefe, S. 1-48.
- Weinert, F.E. (Hrsg.) (2001): *Leistungsmessungen in Schulen*. Weinheim/Basel: Beltz.
- Wiley, D.E./Harnischfeger, A. (1974): *Explosion of a Myth: Quantity of Schooling and Exposure to Instruction, Major Educational Vehicles*. *Studies of Educational Processes*, No. 8. Chicago.

- Wilson, S.M./Youngs, P. (2005): Research on Accountability Processes in Teacher Education. In: Cochran-Smith, M./Zeichner, K.M. (Hrsg.): Studying Teacher Education. Washington D.C.: Erlbaum, S. 591-643.
- Winkler, C. (2010, in Vorbereitung): Carrot and Stick: Zum Einfluss von Unterstützungs- und Anreizsystemen auf die Berufskultur von Lehrern.
- Yoon, K.S./Duncan, T./Lee, S.W.-Y./Scarloss, B./Shapley, K. (2007): Reviewing the Evidence how Teacher Professional Development Affects Student Achievement. (Issues/Answers Report, REL 2007-No. 033). Washington D.C.: U.S. Department of Education.
- Zeichner, K. (2006): Konzepte von Lehrerexpertise und Lehrerausbildung in den Vereinigten Staaten. In: Zeitschrift für Pädagogik, 51. Beiheft, Weinheim: Beltz, S. 97-113.

Diether Hopf, Prof. Dr., Universitätsprofessor für Schulpädagogik i.R., Universität Potsdam.

Anschrift: Lützelsteiner Weg 45, 14195 Berlin
E-Mail: dhopf@uni-potsdam.de