

[Seitenzahlen bitte dem Originalbeitrag zufolge zitieren.]

1.3 Empirische Befunde zur Wirksamkeit der Lehrerbildung

Sigrid Blömeke

Im folgenden Überblick werden zunächst theoretische und methodologische Probleme der Erfassung von ‚Wirksamkeit‘ der Lehrerbildung herausgearbeitet. Von besonderer Bedeutung ist dabei die Frage des Kriteriums: Kognitionen von Lehrerinnen und Lehrern, ihr Handeln oder Schülerleistungen? Die anschließende Darlegung der empirischen Befunde (zu unterschiedlichen Ansätzen von Evaluationen der Lehrerbildung vgl. Kap. 2.3 des Handbuchs) macht einerseits die Bedeutung der beliefs deutlich, die bei Studierenden bereits vor Beginn der Lehrerausbildung vorliegen. Andererseits wird offenkundig, dass derzeit weder entschieden werden kann, welcher Faktor dafür ausschlaggebend ist, dass jemand zu einer ‚guten‘ Lehrperson wird, noch wie Lehrerbildung konkret zu gestalten ist. In der Interpretation dieser Befunde lassen sich für die USA zwei gegensätzliche Strömungen ausmachen: ‚Deregulierer‘, die sich für eine Abschaffung staatlich vorgegebener Ausbildungsregelungen einsetzen, und ‚Professionalisierer‘, die sich für eine Anhebung dieser Standards aussprechen. Welche Folgen eine verstärkte Überprüfung der Wirksamkeit von Lehrerbildung haben könnte, wird ebenfalls auf empirischer Basis nachgegangen. Abschließend werden Anforderungen künftiger Lehrerbildungsforschung formuliert, in deren Kern internationale Leistungsvergleiche und die Verknüpfung von empirischer Bildungsforschung mit Bildungstheorie und -geschichte gesehen werden.

Die Lehrerbildung war hinsichtlich ihrer Gestaltung und Leistungsfähigkeit während des gesamten 20. Jahrhunderts starker Kritik ausgesetzt. Allerdings führten erst zwei ‚externe‘ Entwicklungen dazu, dass verstärkt eine konkrete empirische Überprüfung ihrer Wirksamkeit gefordert wird: die wiederholte Feststellung defizitärer Leistungen des deutschen Schulsystems und eine zunehmende Autonomie der Universitäten, ihre Studiengänge zu gestalten, die notwendigerweise mit einer intensiveren Rechenschaftslegung einher geht. Eine einfache Übertragung der vorhandenen Modelle und Methoden empirischer Forschung auf die Erfassung von ‚Wirkungen‘ der Lehrerbildung ist allerdings theoretisch und methodologisch problematisch. Nur selten sind daher zu dieser Thematik größere empirische Studien durchgeführt worden.

Im folgenden Beitrag wird der Versuch unternommen, einen systematischen Überblick über die vorliegenden empirischen Befunde und wichtige Diskussionslinien in diesem Zusammenhang zu geben. Zu diesem Zweck werden einleitend die Probleme empirischer Forschung zur Wirksamkeit der Lehrerbildung dargelegt (1.3.1). Anschließend erfolgt in zwei Schritten eine Zusammenfassung der zentralen Ergebnisse: zunächst in Bezug auf die Bedeutung der Lernvoraussetzungen von zukünftigen Lehrerinnen und Lehrern zu Beginn ihrer Ausbildung (1.3.2), dann in Bezug auf deren Wirksamkeit (1.3.3). Angesichts der defizitären Situation empirischer Lehrerbildungsforschung im deutschsprachigen Raum und der vergleichsweise umfangreichen Erfahrung damit in den USA, die zudem traditionell als wichtiger Bezugspunkt bildungsbezogener Diskussionen gilt, erfolgt jeweils eine Fokussierung auf US-amerikanische Studien. Nach einem Zwischenfazit (1.3.4) wird eine übergeordnete Fragestellung aufgegriffen und auf empirischer Basis behandelt: Welche Folgen würde eine verstärkte Überprüfung der Wirksamkeit von Lehrerbildung vermutlich haben? (1.3.5) Abschließend werden die eingangs formulierten Probleme empirischer Lehrerbildungsforschung aufgegriffen und es werden – auf der Basis der dargelegten Befunde – Anforderungen zukünftiger Forschung formuliert (1.3.6).

Mit dieser Gliederung wird versucht, eine strikt empirische Perspektive auf die Lehrerbildung einzunehmen. Im Vordergrund stehen die *Ergebnisse* von systematischen Untersuchungen. Unterschiedliche *Modelle* von Evaluationen der Lehrerbildung bzw. unterschiedliche *Konzeptionen von Qualitätsmaßstäben* („Standards“) werden nur am Rande thematisiert. Dieses Thema ist Gegenstand eines eigenen Beitrags (vgl. Kap. 2.3 des Handbuchs).

1.3.1 Probleme empirischer Forschung zur Wirksamkeit der Lehrerbildung

Empirische Forschung zur Wirksamkeit der Lehrerbildung steht vor der zentralen Herausforderung zu klären, welche *Variablen* als relevant angesehen werden können, wenn es um die Evaluation von ‚Wirkungen‘ der Lehrerbildung geht. Herrmann (2003, 625): „Institutionen und Personen, Faktoren und Prozesse, Ziele, Erwartungen und Interessen, Realbedingungen und Beteiligungsvoraussetzungen bei Bildungsprozessen bilden fast unauflösbare Ursachen-, Bedingungs- und Wirkungszusammenhänge.“ Insofern ist die Variablenwahl keineswegs unumstritten und hängt u.a. vom zugrunde liegenden theoretischen Modell zum Zusammenhang von Lehrerbildung, Lehrerhandeln und Schülerleistungen ab. So stellt sich *zum ersten* die Frage, von welchen Faktoren der Lehrerbildung eine Wirksamkeit angenommen wird: Sind es die Inhalte der Lehrangebote, ihre didaktisch-methodische Präsentation, der Grad der Strukturierung der Ausbildung, die Abfolge ihrer Phasen, ihr Ort, ihre Länge usw.? *Zum zweiten* ist es grundlegend, wodurch Lehrerhandeln beeinflusst wird und welche Rolle die Lehrerbildung in diesem Zusammenhang spielt. *Zum dritten* ist die Auswahl des Kriteriums für ‚Wirksamkeit‘ zentral. Bereits auf den ersten Blick lassen sich diesbezüglich drei unterschiedliche Herangehensweisen ausmachen: eine Erfassung der fach- und berufsbezogenen Kognitionen der angehenden Lehrpersonen, eine Beobachtung ihres Handelns im Unterricht und eine Erfassung der Leistungen ihrer Schülerinnen und Schüler. Nimmt man die professionstheoretische

Diskussion ernst (vgl. z.B. Blömeke 2002), muss aber mindestens eine weitere Ebene hinzugezogen werden: die des Ethos' der Lehrerinnen und Lehrer.

Nach Klärung dieser drei Fragen stellt die Wahl der jeweils angemessenen Untersuchungsmethoden eine weitere Herausforderung dar. Eine Durchsicht der Literatur zeigt, dass es weder eine einheitliche Terminologie zur Bezeichnung der erhobenen Fragestellungen noch eine einheitliche Methodologie gibt (vgl. Kane, Sandretto & Heath 2002, 180ff.). Darüber hinaus werden unterschiedliche Erhebungsmethoden und innerhalb dieser unterschiedliche Items verwendet, so dass die vergleichsweise geringe Konsistenz in den Ergebnissen der vorliegenden Studien zur Wirksamkeit der Lehrerbildung möglicherweise auf diese Uneinheitlichkeit zurückzuführen ist.

Für die Bundesrepublik Deutschland lässt sich ein generelles Defizit an lehrerbildungsbezogener empirischer Forschung feststellen. Zahlreichen Veröffentlichungen mit normativ-konzeptioneller Orientierung stehen nur punktuelle empirische Studien gegenüber, die zudem in der Regel lediglich auf Befragungen und Dokumentenanalysen beruhen (vgl. zum Überblick Merzyn 2002). Schlee (1992, 559) formuliert daher als Fazit in Bezug auf die Forschung bis 1990, „dass zu den konzeptionellen Kernfragen und Brennpunkten der Lehrerbildungsreform keine systematische, theoriebezogene und experimentelle Forschung durchgeführt wurde“. Zum selben Ergebnis kommen Döbrich u.a. (2003, 3; vgl. auch Fried 1998, Oelkers 2002b, Schaefers 2002) in Bezug auf die Gegenwart: „Verfügbar ist eine große Fülle von thematisch und regional begrenzten Einzeluntersuchungen, die aber oft kein Urteil darüber erlauben, inwiefern die gewonnenen Erkenntnisse generalisierbar sind.“ Larcher und Oelkers gehen in ihrem Beitrag für das vorliegende Handbuch sogar so weit zu sagen, „wenn es eine Krise in der Lehrerbildung gibt, dann ist es wesentlich eine Krise der fehlenden Daten“ (s. Kap. 1.6).

Eine Untersuchung der Wirksamkeit von Lehrerbildung ist allerdings nur sinnvoll, wenn man davon ausgeht, dass a) Lehrerhandeln Schülerleistungen beeinflusst und dass b) das Lehrerhandeln von der Qualität der Ausbildung – wie auch immer gemessen – abhängt. Im Hinblick auf den ersten Punkt gibt es sowohl im deutschsprachigen (vgl. z.B. Weinert & Helmke 1997, Helmke & Jäger 2002, Ditton 2002, Baumert, Watermann & Schümer 2003) als auch im anglo-amerikanischen Raum zahlreiche Studien (vgl. z.B. als klassische, noch immer bedeutsame Studie Coleman u.a. 1966 sowie Murnane & Phillips 1981; Coleman, Hoffer & Kilgore 1982; Monk & King 1994; Rowan, Chiang & Miller 1997; Wright, Horn & Sanders 1997), die auch weitgehend übereinstimmend interpretiert werden. Zum zweiten Ausgangspunkt gibt es dagegen durchaus widersprüchliche Interpretationen der vorliegenden Befunde. Sie hängen einerseits davon ab, welche methodischen Zugänge als angemessen angesehen werden, und sie hängen andererseits damit zusammen, welche bildungspolitische Position vertreten wird.

Für die US-amerikanische Forschung lässt sich hier ein interessanter Zusammenhang feststellen. Seit einigen Jahren wird hier eine intensive Auseinandersetzung über die Bedeutung von gesetzlichen Regelungen der formalen Anforderungen an die Ausbildung von Lehrerinnen und Lehrern geführt. Auf der einen Seite stehen diejenigen, die sich für eine weitgehende Freigabe aller Vorschriften aussprechen. Letztlich solle der Markt das Problem regeln, indem die besten Lehrerinnen und Lehrer – ob mit oder ohne Ausbildung – über Angebot und Nachfrage gefunden würden. Diese Gruppe der so genannten ‚Deregulierer‘ – vorrangig vertreten durch die Abell Foundation –, stützt sich auf eine der umfassendsten Reviews englischsprachiger (v.a. US-amerikanischer) empirischer Literatur, die es derzeit gibt. Unter der Voraussetzung, dass nur Ergebnisse berücksichtigt werden, die a) in *large scale*-Studien entstanden sind und Drittvariablen wie soziale Herkunft etc. kontrollieren, die b) die Wirksamkeit der Lehrerbildung anhand von Schülerleistungen messen und die c) in peer reviewten-Organen mit Blindbegutachtung – mindestens aber überhaupt mit einem Review-Verfahren – erschienen sind, bleiben nur wenige Studien zur Wirksamkeit der Lehrerbildung übrig (mit zum Teil widersprüchlichen Aussagen). Auf der anderen Seite steht die Gruppe derjenigen, die sich für

eine Reform der Lehrerbildung einsetzt in dem Sinne, dass für alle Staaten verbindlich eine Mindestnorm an zu belegenden Inhalten (v.a. in der Kombination von Fachwissenschaft und Erziehungswissenschaft sowie angemessener schulpraktischer Ausbildungsanteile) sowie zu erbringenden Abschlüssen (Master) gefordert wird. Prominenteste Vertreterin der so genannten ‚Professionalisierer‘ ist Darling-Hammond. Von ihr stammen ebenfalls umfassende Reviews empirischer Forschungsarbeiten – mit dem Unterschied, dass sie (mindestens eingeschränkt) auch Ergebnisse von Doktorarbeiten, qualitativen Studien und Beiträge in Zeitschriften ohne *peer review*-Verfahren akzeptiert.

Ohne Zweifel müssen die Methoden, mit denen in den USA in der Vergangenheit Wirkungen der Lehrerbildung untersucht wurden, kritisch diskutiert werden (vgl. Wideen, Mayer-Smith & Moon 1998, 161ff.; Chatterij 2002, 367): So handelt es sich in der Regel nur um (qualitative) Fallstudien. Ihre Ergebnisse können daher weniger als empirisch geprüfte Aussagen genommen werden denn als Hypothesen, die noch einer Überprüfung in large scale-Studien bedürfen. Bei Studien mit größeren Fallzahlen handelt es sich oft um Re-Analysen von *surveys*, die andere Fragestellungen verfolgten, so dass eine präzise und konsistente Bearbeitung von lehrerbildungsbezogenen Fragen schwierig ist. (Feld-)experimentelle Studien und Langzeitstudien finden sich so gut wie nie. Hinzu kommt, dass sich die Erhebungen häufig auf *Befragungen* (von Schülern, Lehrpersonen oder Schulfunktionären) beschränken, gelegentlich werden noch *Beobachtungen* (durch Schulfunktionäre oder Wissenschaftler) einbezogen. Schülerleistungen werden in der Regel nicht erhoben – und wenn, dann werden sie (v.a. in älteren Untersuchungen) so gut wie nie auf der Basis standardisierter Tests festgestellt. Noch immer ist darüber hinaus eine erhebliche Forschungslücke festzustellen, was die Bedeutsamkeit der fachwissenschaftlichen Ausbildung angeht; die überwiegende Zahl der Studien fokussiert auf *beliefs* oder pädagogisches Wissen. Und schließlich ist ein eklatantes Defizit an gegenseitigen Bezugnahmen festzustellen. Die Mehrheit der Forscher macht keinen Versuch, die eigenen Arbeiten mit denen anderer zu verbinden.

Um diese Mängel systematisch zu erfassen, kann auf Kennedys (1999) Differenzierung der empirischen Studien zum Lernprozess von Schülerinnen und Schülern Bezug genommen und diese auf die Lehrerbildungsforschung übertragen werden: Auf dem höchsten Level werden standardisierte Beobachtungen und Tests in realen Lehr-Lernsituationen vorgenommen. Ein Level darunter erfolgt eine Erhebung mit Hilfe hypothetischer Lehr-Lernsituationen, von *stimulated recalls* bzw. schriftlichen Unterlagen der Lehrenden. Auf dem dritten Level begnügt man sich mit nicht-situationsbezogenen Befragungen zum eigenen Lehrhandeln. Die vierte und am wenigsten valide Annäherung ist schließlich die der Befragung von Lehrenden nach Wirkungen z.B. der Lehrerbildung auf ihr Handeln. Tatto (2000, 7) hält in ihrem Review empirischer Studien zur Lehrerbildung fest: „Studies of teacher development have more commonly used third and fourth level approximations. [...] Less frequently have researchers used second level approximations [...] and rarely have they used first level approximations.“

Diese Probleme empirischer Lehrerbildungsforschung gilt es zu bedenken, wenn im Folgenden ihre Ergebnisse dargelegt werden.

1.3.2 Empirische Befunde zur Bedeutung der Lernvoraussetzungen

Die Wirksamkeit der Lehrerbildung kann nur schwer angemessen beurteilt werden, wenn nicht zuvor geklärt wird, mit welchen Lernvoraussetzungen zukünftige Lehrerinnen und Lehrer ihre Ausbildung aufnehmen. Diese lassen sich hinsichtlich zweier Dimensionen unterscheiden: zum einen im Hinblick auf sozio-kulturelle Merkmale und zum anderen im Hinblick auf Vorannahmen, Einstellungen bzw. Erwartungen (mangels konsistenter Begrifflichkeit in der Forschung im Folgenden mit dem in anglo-amerikanischen Studien verbreiteten Begriff ‚*beliefs*‘ bezeichnet) sowie Berufswahlmotive.

Während zu den sozio-kulturellen Merkmalen der deutschen Lehramtsstudierenden – mit Ausnahme historischer Untersuchungen (vgl. z.B. Kaelble 1978, Bölling 1983) – kaum aktuelle und detaillierte repräsentative Informationen vorliegen, lässt sich in Bezug auf Nordamerika festhalten: „The typical candidate for teacher education in the United States and Canada is a White, Anglo-Saxon, lower- or middle-class female who has grown up in a suburban or rural area. She is monolingual in English, has traveled very little beyond a 100-mile radius of her home, and has attended a local college or university close to her home.“ (Wideen, Mayer-Smith & Moon 1998, 141).

Zu den Berufswahlmotiven lassen sich im deutschsprachigen Raum zahlreiche Studien finden, die im Kern auf die beiden folgenden Aussagen reduziert werden können: Lehramtsstudierende nehmen ihr Studium in der Absicht auf, später als Lehrer bzw. Lehrerin zu arbeiten (vgl. Terhart u.a. 1994, Ulich 2000, Drechsel 2001), und sie haben dafür in erster Linie pädagogische oder fachliche Motive (vgl. als Überblick Schaefer 2002, 67f.). Die Validität dieser Erkenntnisse kann allerdings trotz mehrfacher Replikation insofern in Frage gestellt werden, als es sich jeweils nur um *Befragungen* der Betroffenen handelt – und dies auch nur im Nachhinein.

Im Hinblick auf die *beliefs* lassen sich folgende zentrale Ergebnisse der empirischen Forschung festhalten (vgl. Koch 1972; Zeichner 1986; Clark 1988; Kagan 1992; Pajares 1992; Block & Hazelip 1995; Richardson 1996; Bullough 1997; Wideen, Mayer-Smith & Moon 1998; Richardson & Placier 2001; Kane, Sandretto & Heath 2002):

- Lehramtsstudierende treten mit Vorstellungen über schul- und unterrichtsrelevante Aspekte in die Lehrerbildung ein.
- Diese *beliefs* sind erfahrungsbasiert. Ihr Zustandekommen lässt sich sozialisatorisch (Lortie 1975: Schule als ‚apprenticeship of observation‘), evolutionär oder psychoanalytisch einordnen.
- *Beliefs* sind weitgehend veränderungsresistent, so dass es im Laufe der Lehrerbildung eher selten zu grundlegenden Veränderungen kommt.
- Ihre Funktionsweise in der Ausbildung kann als Filter beschrieben werden. Es werden überwiegend nur solche Informationen aufgenommen, die sich in das vorhandene System an Überzeugungen einpassen lassen.
- Erfolg versprechend scheinen Lehrprozesse zu sein, die aktiv an die vorhandenen *beliefs* anknüpfen und diese so schrittweise verändern.

In inhaltlicher Hinsicht ist in Bezug auf die *beliefs* als bedeutsames Ergebnis festzuhalten, dass die Lehramtsstudierenden sehr optimistisch sind, was ihre zukünftigen Fähigkeiten zu unterrichten angeht (vgl. Kagan 1992, 140). Sie meinen zu wissen, wie Unterricht auszusehen hat (eine Lehrperson überliefert Wissen frontal vor der Klasse an die Schülerinnen und Schüler, die dieses rezeptiv aufnehmen), und sie meinen, nur noch ein bestimmtes Methodenrepertoire erlernen müssen, um starten zu können. Als besonders wichtig sehen es die Studierenden an, ihre Klassen ‚im Griff‘ zu haben. Gleichzeitig – und in einem gewissen Spannungsverhältnis hierzu – ist ihr Fokus auf soziale Faktoren des Unterrichtens ausgerichtet, indem sie ein verständnisvolles und sorgendes Verhältnis zu ihren Schülerinnen und Schülern für bedeutsam halten. Damit ist die Annahme verbunden, dass die Persönlichkeit einer Lehrkraft von höherer Bedeutung sei als ihre kognitiven Fähigkeiten oder ihr pädagogisches bzw. fachliches Wissen (vgl. Wideen, Mayer-Smith & Moon 1998, 142f.).

Um den konkreten Verlauf des Einflusses der Lehrerbildung auf anfängliche Einstellungen, Interessen und Werthaltungen der Lehramtsstudierenden deutlich zu machen, kann auf die bereits 30 Jahre alte, immer noch maßgebliche, da einzige Längsschnittuntersuchung im deutschsprachigen Raum von Koch (1972) Bezug genommen werden. Eingesetzt wurden

Skalen zu schul- und unterrichtsbezogenen Einstellungen und Überzeugungen: Bedeutsamkeit von ‚Anlage versus Umwelt‘ und ‚Druck versus Zug‘ als Bedingungen von Lernen sowie von ‚Allgemeinbildung versus Spezialisierung‘ als Ziel von Lernen, Verständnis des eigenen Berufs als ‚Berufung versus Job‘ bzw. als ‚Pädagoge versus Fachwissenschaftler‘, und schließlich ‚negative Reformeinstellung versus Veränderungsbereitschaft‘. Zwei Kernergebnisse dieser Studie lauten: (1) Lehramtsstudierende unterscheiden sich in ihren Einstellungen weder im Vergleich zu anderen Abiturienten noch im Vergleich zu Studienanfängern in anderen Studiengängen. (2) Im Prozess der Lehrerausbildung lassen sich durchaus Veränderungen gegenüber den ursprünglichen zu Beginn des Studiums vorhandenen Einstellungen, Interessen und Werthaltungen feststellen. Diese werden aber mit dem Eintritt in die Schulpraxis während des Referendariats wieder rückgängig gemacht (‚Konstanzer Wanne‘). Von der Tendenz her gilt, dass vor dem Studium eher konservative Einstellungen vorhanden sind, während des Studiums liberalere Auffassungen stärkere Zustimmung finden (stärkere Zuschreibung von Umweltfaktoren und Unterstützung als Erfolgsfaktoren für Lernen sowie größere Veränderungsbereitschaft), bevor im Referendariat wieder zu konservativeren Positionen zurückgekehrt wird. Eine Follow up-Untersuchung zeigt, dass sich die festgestellte Tendenz zu konservativen pädagogischen Einstellungen, Interessen und Werthaltungen in den ersten Jahren des Berufslebens weiter fortsetzt (vgl. Dann, Müller-Fohrbrod & Cloetta 1981).

Dass angehende Lehrerinnen und Lehrer in der Ausbildung erworbene innovative Einstellungen – vor allem die Bereitschaft, gegenüber Schülern auf übermäßigen Zwang zu verzichten – wieder aufgeben, stellten 20 Jahre später auch Brouwer und ten Brinke (1995a, b) in einer Längsschnittuntersuchung fest, die sie in den Niederlanden zur Erhebung der Wirkungen eines neuen Lehrerbildungsprogramms durchgeführt haben. In Form von Befragungen suchten die Autoren auch die Ursache dieses Wandels zu klären. Sie kommen zu dem Ergebnis, dass sich die neu in eine Schule kommenden Lehrerinnen und Lehrer gezwungen sehen, gemäß der dort üblichen Handlungsrouninen und gegen ihre eigenen Einstellungen zu handeln (vgl. Brouwer/ ten Brinke 1995a, 4).

Vor diesem Hintergrund wird verständlich, wenn Lortie (1975) – wie Koch (1972) – die Lehrerausbildung als „low impact enterprise“ und Richardson (1996) sie als „weak intervention“ (v.a. im Vergleich zur eigenen Schulerfahrung und der späteren beruflichen Sozialisation) bezeichnet (speziell in Bezug auf die Ausbildung von Mathematiklehrerinnen und -lehrern vgl. auch Ball, Lubienski & Mewborn 2001, 437). Was immer Lehrpersonen im Laufe ihrer Ausbildung gelernt haben erscheint „washed out by school experience“ (Zeichner & Tabachnik 1981).

1.3.3 Ergebnisse zur Wirkung von Lehrerbildungsprogrammen

Im Folgenden wird auf zentrale Ergebnisse so genannter ‚Programmevaluationen‘ eingegangen, in denen die Lehrerbildung als Ganze oder zumindest substanzielle Teile untersucht werden. Gegenstand der Untersuchung ist hier nicht die individuelle Lehrperson, ihre Lernvoraussetzungen und ihr Kompetenzgewinn im Prozess der Lehrerbildung, sondern in *large scale*-Studien und Meta-Analysen werden mit Hilfe multipler Regression Merkmale des Berufsstandes in Beziehung zu Schülerleistungen oder Lehrerhandeln gesetzt. Um die Effekte der Lehrerbildung abschätzen zu können, werden bekannte Einflüsse (z.B. finanzielle Ausstattung der Schulen, Dauer der beruflichen Erfahrung) konstant gehalten (*product function model*, vgl. Floden 2001, 7). Während sich entsprechende Studien in der Bundesrepublik Deutschland nicht finden lassen, stellt diese Art von Studien in den USA das Kernverständnis von ‚Evaluation der Lehrerbildung‘ dar. Ihre zentralen Erkenntnisse können für die deutsche Diskussion wichtige Hinweise liefern, auch wenn sich die hiesige Situation aufgrund der historischen Entwicklung des Bildungssystems und der bildungspolitischen Strukturen sowie auf-

grund der sozialen Stellung von Lehrpersonen und ihrer individuellen Mentalität deutlich unterscheidet.

Die gegenwärtig umstrittenste Frage bei der Untersuchung von Wirkungen der Lehrerbildung ist die nach der Bedeutsamkeit der Ausbildung im Verhältnis zu bereits vorhandenen Persönlichkeitsmerkmalen von Lehrpersonen. Diese Frage kann auch umformuliert werden in: Wird man als guter Lehrer ‚geboren‘ oder können die zugrunde liegenden Kenntnisse, Fertigkeiten und Fähigkeiten vollständig bzw. teilweise erlernt werden? Für einen Überblick zu den zentralen Erkenntnissen ist an erster Stelle auf die großen Reviews der Abell Foundation (2001a, b) und von Darling-Hammond (2000) einzugehen, die sich z.T. auf dieselbe Datenlage beziehen und die unterschiedliche Interpretationsansätze repräsentieren.

- *Kernergebnisse empirischer Forschung aus Sicht der ‚Professionalisierer‘*

Darling-Hammonds Review besteht aus zwei Teilen, zum einen wertet sie die seit Ende des Zweiten Weltkriegs erschienene englischsprachige Literatur im Hinblick auf den Einfluss von Lehrerbildung und Lehrerhandeln auf Schülerleistungen unter verschiedenen systematischen Gesichtspunkten aus. Anschließend nimmt sie eine ergänzende Analyse neuerer Daten aus drei US-weiten Surveys sowie aus Fallstudien zu einzelnen Staaten vor. Dabei macht Darling-Hammond (2000, 24) darauf aufmerksam, dass eine besondere Problematik in der Auswertung dieser großen Zahl an Studien im unterschiedlichen Grad der Aggregation der Daten (individuelle, institutionelle bzw. staatliche Ebene) liegt: „The direction of the differences cannot be predicted with certainty. While, on one hand, the possibility of greater variability or noise exists in disaggregated analyses, it is possible that omitted variables may bias the coefficients of school input variables upward when the data are aggregated to the district or state level.“ Im Hinblick auf die Wirksamkeit der Lehrerbildung sind unter Berücksichtigung dieser Einschränkung folgende Ergebnisse systematisiert hervorzuheben (zu den detaillierten bibliographischen Angaben s. Darling-Hammond 2000):

- Abschluss eines lehrerbildenden Studiengangs (*certification*; ohne nähere Spezifizierung, in der Regel handelt es sich um einen Bachelor mit einem Major in einer Fachwissenschaft oder der Erziehungswissenschaft): Lehrpersonen mit einer entsprechenden Ausbildung erreichen den meisten Studien zufolge durchweg höhere Schülerleistungen bzw. positivere Einschätzungen ihres Handelns durch Fachpersonal (Schulleiter, Schulaufsicht, Wissenschaftler etc.) oder Schüler als Lehrpersonen ohne eine solche bzw. nur mit einer Notausbildung (LuPone 1961; Hice 1970; McNeil 1974; Roupp u.a. 1979; Erikson & Barr 1985; Evertson, Hawley & Zlotnik 1985; Knoblock 1986; Mitchell 1987; Hansen 1988; Armour-Thomas u.a. 1989; Lenk 1989; Bents & Bents 1990; Feiman-Nemser & Parker 1990; Gomez & Grobe 1990; Texas Education Agency 1993; NCES 1994; Sanders, Skonie-Hardin & Phelps 1994; Jelmberg 1995; Campbell u.a. 1996; Greenwald, Hedges & Laine 1996; Council for School Performance 1997; Goldhaber & Brewer 1999; Fetler 1999; Fuller 1999). Die Art des Majors – ob in einer Fachwissenschaft oder in Erziehungswissenschaft – mache keinen Unterschied (Murnane 1985). Darüber hinaus finden sich in zahlreichen Studien Zusammenhänge in der Form, dass voll ausgebildete Lehrpersonen weniger Schwierigkeiten zu Beginn ihres Berufslebens haben, besser mit der Unterrichtsplanung zurechtkommen, zufriedener mit ihrer Ausbildung sind und später seltener aus dem Schuldienst ausscheiden. Einzelnen Studien zufolge erreichen Lehrpersonen mit weniger als einer vollen Lehrerausbildung (also z.B. mit Notausbildungen) zwar keine geringeren Schülerleistungen (Barnes, Salmon & Wale 1989; Goebel, Romacher & Sanchez 1989; Goldhaber & Brewer 1999). Die methodische Sorgfalt dieser Studien wird von Darling-Hammond (2000, 8) aber z.T. in Zweifel gezogen.
- Vorhandensein eines für die spätere Unterrichtstätigkeit relevanten Masterabschlusses (die Mehrzahl der Lehrpersonen hat einen solchen in Erziehungswissenschaft erwor-

ben, die übrigen besitzen einen in einer Fachwissenschaft, in der sie unterrichten): Lehrpersonen mit einem Master erreichen höhere Schülerleistungen als Lehrpersonen, die nur einen Bachelor besitzen (Knoblock 1986; Ferguson 1991; NCES 1994; Sanders, Skonie-Hardin & Phelps 1994; Ferguson & Ladd 1996; Council for School Performance 1997).

- Qualität der fachwissenschaftlichen Ausbildung (erhoben in Form von Tests, Kurszählungen oder akademischen Abschlüssen): Ein konsistenter Zusammenhang zwischen dem Fachwissen und Schülerleistungen bzw. positiven Einschätzungen des Lehrerhandelns durch Fachpersonal (Schulleiter, Schulaufsicht, Wissenschaftler etc.) bzw. Schüler ist nicht zu erkennen, da in den vorliegenden Studien entweder nicht-signifikante oder widersprüchliche (also mal positive und mal negative Zusammenhänge) festgestellt werden (Perkes 1967, 1968; Begle & Geeslin 1972; Quirk, Witten & Weinberg 1973; Summers & Wolfe 1975; Ayers & Qualls 1979; Andrews, Blackmon & Mackey 1980; Byrne 1983; Hawk, Coble & Swanson 1985; Haney, Madaus & Kreitzer 1986; Ashton & Crocker 1987; Guyton & Farokhi 1987; Ferguson & Womack 1993; Monk & King 1994). Nur eine Studie (Carroll 1975) und ein Review (Druva & Anderson 1983) behaupten einen signifikant positiven Zusammenhang. Darling-Hammond zieht für die Interpretation dieser widersprüchlichen Ergebnisse Monks (1994) Langzeitstudie heran, in der von einem ‚Deckeneffekt‘ ausgegangen wird: Um hohe Schülerleistungen zu erreichen, ist ein gewisses Maß an Fachwissen notwendig, über einer gewissen Schwelle ist aber kein zusätzlicher positiver Effekt mehr festzustellen. Diesem Interpretationsansatz wird auch in der mathematikspezifischen Lehrerbildungsforschung gefolgt, die im Unterschied zu der in anderen Unterrichtsfächern vergleichsweise häufig zu finden ist (vgl. Ball & Bass 2000; Ball, Lubienski & Mewborn 2001). In diesen Studien wird auch versucht, die Ursache für den ‚Deckeneffekt‘ zu erklären: Die aus dem höheren fachwissenschaftlichen Niveau folgende Strukturierung der Inhalte mache eine Elementarisierung schwieriger. Die Vertiefung fachwissenschaftlicher Inhalte sei zudem in der Regel mit einer stärkeren Orientierung an konventionellen Lehrmethoden verbunden, was schülerorientiertem Lehren entgegenstehe. Darüber hinaus weist Darling-Hammond (2000, 4) darauf hin, dass die widersprüchlichen Ergebnisse ggf. auch auf unterschiedliche Erhebungsmethoden zurückgeführt werden können.
- Qualität der fachdidaktischen Ausbildung: Dieser Bestandteil der Lehrerbildung spielt in den USA eine deutlich geringere Rolle als in Deutschland. Zudem wird er – je nach Ansiedlung innerhalb der Ausbildung – mit verschiedenen Begriffen belegt (*pedagogical content knowledge, subject methodology, subject education*), was es schwierig macht, darauf ausgerichtete Studien zu identifizieren. Zieht man aus Darling-Hammonds Review diejenigen Studien heraus, die einen eindeutig fachdidaktischen Gegenstand haben, stellen sich die Zusammenhänge zu Schülerleistungen bzw. Einschätzungen des Lehrerhandelns durch Fachpersonal (Schulleiter, Schulaufsicht, Wissenschaftler etc.) bzw. Schüler als signifikant positiv dar (Perkes 1967, 1968; Begle 1979; Brown, Smith & Stein 1995, Wiley & Yoon 1995, Cohen & Hill 1997).
- Qualität der erziehungswissenschaftlichen Ausbildung: Im Vergleich zur fachwissenschaftlichen Ausbildung haben die meisten Studien „a somewhat stronger and more consistently positive influence of education coursework“ auf Schülerleistungen bzw. positive Einschätzungen des Lehrerhandelns durch Fachpersonal (Schulleiter, Schulaufsicht, Wissenschaftler etc.) bzw. Schüler festgestellt (Druva & Anderson 1983; Denton & Lacina 1984; Evertson, Hawley & Zlotnik 1985; Ashton & Crocker 1987; Guyton & Farokhi 1987; Armour-Thomas u.a. 1989; Ferguson & Womack 1993; Monk 1994). Gestützt auf Byrne (1983) interpretiert Darling-Hammond (2000, 5) die Ergebnisse allerdings insofern komplexer, als vermutlich erst durch das Zusammen-

spiel von umfangreicher erziehungswissenschaftlicher Ausbildung und Fachwissen Effekte verzeichnet werden können: „The degree of pedagogical skill may interact with subject matter knowledge to bolster or reduce teacher performance.“ Dies ist insofern plausibel, als für die umfassenden US-weiten Surveys in Mathematik und Lesen der vierten und achten Klassen des ‚National Assessment of Educational Progress‘ (NAEP) in den Jahren 1990, 1992, 1994 und 1996 (Campbell u.a. 1996) nur Lehrpersonen mit einem Abschluss in einem lehrerbildenden Studiengang berücksichtigt wurden, die einen Major im unterrichteten Fach hatten, was bedeutet, dass diese Form der Operationalisierung „is a proxy for both strong disciplinary knowledge (a major in the field taught) and substantial knowledge of education (full certification)“ (Darling-Hammond 2000, 29).

- Vorhandensein genereller Fähigkeiten der Lehrpersonen: Es bestehen keine oder nur nicht-signifikante Zusammenhänge zwischen allgemeinen kognitiven Fähigkeiten von Lehrpersonen (Indikator: Intelligenzquotient) und Schülerleistungen bzw. positive Einschätzungen des Lehrerhandelns durch Fachpersonal (Schulleiter, Schulaufsicht, Wissenschaftler etc.) bzw. Schüler (Hellfritsch 1945; LaDuke 1945; Rosteker 1945; Skinner 1947; Schalock 1979; Soar, Medley & Coker 1983). Dagegen sind die verbalen Fähigkeiten einer Lehrperson mit Schülerleistungen signifikant assoziiert (Coleman u.a. 1966; Bowles & Levin 1968; Hanushek 1971; Summers & Wolfe 1975; Murnane 1985).

Aus den analysierten Studien zieht Darling-Hammond (2000, 1) die zentrale Schlussfolgerung, „that measures of teacher preparation and certification are by far the strongest correlates of student achievement in reading and mathematics, both before and after controlling for student poverty and language status“. Auf dieser Basis spricht sie sich dann für ein stärkeres staatliches Engagement aus, was Regelungen hinsichtlich Mindestanforderungen an die Ausbildung von Lehrpersonen und die Unterstützung ihrer professionellen Entwicklung angeht. In den USA erlauben immerhin noch 30 Staaten die Anstellung von Lehrpersonen, die keinen oder nur einen unpassenden Studienabschluss haben.

- *Kernergebnisse empirischer Forschung aus Sicht der ‚Deregulierer‘*

Zu einer entgegengesetzten Schlussfolgerung kommt die Abell Foundation. In ihrem Hauptbericht lautet die Kernaussage, dass sich keine hinreichenden Belege dafür finden lassen, dass Personen mit abgeschlossener Lehrerbildung tatsächlich höhere Schülerleistungen erzielen als Personen mit anderen Abschlüssen: „Teacher certification is neither an efficient nor an effective means by which to ensure a competent teaching force.“ (Abell Foundation 2001a, iii) Wenn überhaupt ein stabiler Prädiktor für gute Lehrpersonen gefunden werden könne, dann nur außerhalb der Lehrerausbildung, bei Persönlichkeitsmerkmalen nämlich. Die Folgerungen aus diesen Erkenntnissen lauten (vgl. ebd., vii f.): Abschaffung aller Regularien im Hinblick auf spezielle lehrerausbildende Qualifikationen von einzustellenden Lehrpersonen; stattdessen lediglich Forderung eines Bachelors (ohne nähere Spezifizierung der Fachrichtung, wie es gegenwärtig in den USA in Bezug auf Privatschulen gilt), kombiniert mit einem Einstellungstest, der verbale Fähigkeiten sowie – bei Elementarschullehrern – die Fähigkeit, Lesen zu unterrichten, bzw. – bei Sekundarschullehrern – das Fachwissen misst. Als Unterstützung für ihre Forderungen zieht die Stiftung heran, dass im Privatschulwesen der USA, das im Vergleich zum öffentlichen Schulwesen als qualitativ besser angesehen wird, der Anteil an Lehrpersonen ohne spezielle Lehrerausbildung höher ist als im öffentlichen Schulwesen (vgl. Walsh 2001, o.P. [6]). Lehrerausbildende Bemühungen sollen weniger auf erziehungswissenschaftliche Qualifikationen denn auf die Förderung allgemeiner mündlicher und schriftlicher Sprach- sowie von Lesefertigkeiten gerichtet sein; insgesamt solle stärker die Berufseingangsphase (zu Lasten der Lehrerausbildung) unterstützt werden (vgl. Abell Foundation 2001a).

Diese Interpretation resultiert in hohem Maße daraus, dass empirische Erkenntnisse keine Berücksichtigung finden, wenn sie den geforderten Standards (s.o.) nicht entsprechen. Die Autoren schließen in diesem Sinne selbst Aussagen aus blindbegutachteten (!) Studien aus, wenn in diesen beispielsweise Doktorarbeiten, die der Abell Foundation generell als nicht akzeptabel erscheinen, oder qualitative Fallstudien zitiert wurden bzw. wenn die Qualität des Lehrerhandelns nicht auf der Basis von Tests von Schülerleistungen erhoben worden ist, sondern beispielsweise auf Schülerbefragungen oder Schulleitereinschätzungen zurück ging.

- *Re-Analyse der vorgelegten Reviews*

Letztlich lässt sich aber die Frage stellen, inwieweit eine solche methodische Einschränkung berechtigt ist, die die eingangs angesprochenen theoretischen und methodologischen Herausforderungen empirischer Lehrerbildungsforschung weitgehend unberücksichtigt lässt und die ggf. den Kompetenzen von Lehrpersonen nicht differenziert genug gerecht wird. Darüber hinaus kann auch die Konsequenz kritisiert werden: ‚keine Studie = keine Lehrerausbildung notwendig‘. In Ergänzung der Analysen der Abell Foundation soll daher in derselben Systematik, die der Auswertung des Darling-Hammond-Reviews zugrunde gelegt wurde, eine Re-Analyse erfolgen. In einem Anhang stellt die Abell Foundation die vorliegenden empirischen Studien umfassend vor, indem Hauptaussagen und Datenqualität eingeschätzt werden. Akzeptiert man einen mittleren Review-Standard (zwar keine Blindbegutachtung, aber Vorhandensein eines anderen Gutachter- oder redaktionsinternen Reviewverfahrens) und löst man sich von der ausschließlichen Orientierung an Schülerleistungen (das bedeutet im Kern: Berücksichtigung auch anderer Operationalisierungen von ‚Wirksamkeit‘ der Lehrerbildung), können die Befunde der verbleibenden Studien wie folgt zusammengefasst werden (zu den detaillierten bibliographischen Angaben s. Abell Foundation 2001b):

- Abschluss eines lehrerbildenden Studiengangs (ohne nähere Spezifizierung, in der Regel handelt es sich dann um einen Bachelor in einer Fachwissenschaft oder Erziehungswissenschaft): Obwohl von dem Vorhandensein eines entsprechenden Abschlusses auf der Plausibilitätsebene vermutlich am ehesten positive Wirkungen auf Schülerleistungen bzw. positive Einschätzungen des Lehrerhandelns durch Fachpersonal (Schulleiter, Schulaufsicht, Wissenschaftler etc.) bzw. Schüler erwartet werden kann, sind die Ergebnisse der empirischen Studien widersprüchlich. Während in einigen Studien kein signifikanter Zusammenhang festgestellt wird (Popham 1971; Link & Ratledge 1979; Everston, Hawley & Zlotnik 1985; Hanushek 1989, 1996), behaupten andere, dass Lehrpersonen mit abgeschlossener Lehrerbildung höhere Schülerleistungen bzw. bessere Einschätzungen durch Schulleiter/ Schüler erreichen als Lehrpersonen ohne einen solchen Abschluss (Ellis 1961; LuPone 1961; Greenberg 1983; Ashton 1996; Fetler 1999; Darling-Hammond, Barry & Thoreson 2001).
- Vorhandensein eines für die spätere Unterrichtstätigkeit relevanten Masterabschlusses (die Mehrzahl der Lehrpersonen erwirbt einen solchen in Erziehungswissenschaft, die übrigen einen in einer Fachwissenschaft): In einer Reihe von Studien wird kein signifikanter positiver Zusammenhang zu Schülerleistungen festgestellt (Summers & Wolfe 1975, 1977; Murnane & Phillips 1981; Kiesling 1984; Hanushek 1986, 1992; Harnisch 1987; Ferguson 1991; Greenwald, Hedges & Laine 1994; Monk 1994; Monk & King 1994; Ferguson & Ladd 1996; Hanushek, Kain & Rivkin 1998; Grissmer u.a. 2000). Zwei Studien stellen eine Ausnahme dar (Greenwald, Hedges & Laine 1996; Hedges & Greenwald 1996).
- Qualität der fachwissenschaftlichen Ausbildung (erhoben in Form von Tests, Kurszählungen oder akademischen Abschlüssen): Die Ergebnisse hierzu sind insofern widersprüchlich, als in einigen Studien kein signifikanter Zusammenhang zwischen der Höhe des Fachwissens und Schülerleistungen bzw. Einschätzungen durch Schulleiter/ Schüler (Everston, Hawley & Zlotnik 1985; Ashton/ Crocker 1987; Guyton & Farokhi

1987; Goldhaber & Brewer 1997; Friedman 2000), in den meisten allerdings doch ein signifikanter Zusammenhang festgestellt wird (Perkes 1967; Summers & Wolfe 1975, 1977; Eisenberg 1977; Druva & Anderson 1983; Hawk, Coble & Swanson 1985; Chaney 1995; Goldhaber & Brewer 1996, 2000; Rowan, Chiang & Miller 1997; Hawkins, Stancavage & Dossey 1998), wenn auch z.T. auf einen Deckeneffekt hingewiesen wird (Monk 1994).

- Qualität der fachdidaktischen Ausbildung: Die als hierauf ausgerichtet identifizierbaren Studien deuten auf einen positiven Zusammenhang zu Schülerleistungen hin (Perkes 1967; Monk 1994).
- Qualität der erziehungswissenschaftlichen Ausbildung: Eine solche Ausbildung hat den meisten Studien zufolge positive Auswirkungen auf Schülerleistungen bzw. Einschätzungen durch Schulleiter/ Schüler (Hall 1964; Druva & Anderson 1983; Everston, Hawley & Zlotnik 1985; Ashton/ Crocker 1987; Grossman 1990), wenn auch andere Ergebnisse vorliegen (Summers & Wolfe 1975, 1977).
- sozio-kulturelle Herkunft der Lehrpersonen: Die Ergebnisse stellen sich widersprüchlich dar. So stellen einige Studien fest, dass Schichtzugehörigkeit oder ethnische Herkunft für Schülerleistungen bedeutsam sind (Murnane 1975; Alexander/ Entwisle/ Thompson 1987); andere Studien sehen keinen Zusammenhang zwischen Herkunft und Schülerleistungen (Summers & Wolfe 1975, 1977; Farkas u.a. 1990; Ferguson 1991; Ehrenberg, Goldhaber & Brewer 1994).
- generelle Fähigkeiten der Lehrpersonen: Die verbalen Fähigkeiten der Lehrpersonen sind das einzige Merkmal, das durchgängig in allen empirischen Studien einen positiven Zusammenhang zu Schülerleistungen aufweist (Coleman u.a. 1966; Bowles/ Levin 1968; Hanushek 1971, 1992; Levin 1976; McLaughlin & Marsh 1978; Bruno/ Doscher 1981; Murnane & Phillips 1981; Murnane 1983; Strauss & Sawyer 1986; Ferguson 1991; Ehrenberg & Brewer 1994, 1995; Greenwald, Hedges & Laine 1996; Hedges & Greenwald 1996; Kain & Singleton 1996). Hohe generelle kognitive Fähigkeiten wirken sich ebenfalls positiv auf Schülerleistungen aus (Hanushek 1971; Ferguson & Ladd 1996; Ferguson 1998).

Verschiebt man also das Kriterium für die Akzeptanz wissenschaftlicher Beiträge, ist die Übereinstimmung zwischen den zugrunde gelegten Reviews durchaus hoch. Diese Übereinstimmung wird in ihrem Kern weiter gefestigt, wenn man eine dritte Literaturübersicht hinzuzieht, die parallel im Auftrag der US-Regierung entstanden ist und die weniger eindeutig einer der beiden angeführten Positionen zuzuordnen ist (vgl. Wilson, Floden & Ferrini-Mundy 2001). Dieses Review sucht dieselben systematischen Fragen zu beantworten, wie sie oben herausgearbeitet wurden. Etwa 300 empirische Studien werden auf ihre Tauglichkeit geprüft, von denen letztlich 57 als wissenschaftlichen Kriterien entsprechend akzeptiert werden. Aus der Zusammenfassung der Kernergebnisse sowie einem detaillierten Anhang, in dem das methodische Vorgehen und die wichtigsten Ergebnisse aller 57 Studien dargestellt werden, ergeben sich vergleichbare Aussagen, wie sie oben in Bezug auf Darling-Hammond und die Abell Foundation herausgearbeitet wurden. Auf eine erneute Einzeldarstellung soll daher an dieser Stelle verzichtet werden.

- *Empirische Forschung zu punktuellen Fragestellungen*

Im Unterschied zu den bisher angesprochenen Programmevaluationen untersuchen zahlreiche Studien einzelne kleinere Reformansätze (*short-term interventions*), wie sie sich z.B. in neu etablierten Pflichtkursen äußern. In diesen Studien können überwiegend nur begrenzte Effekte festgestellt werden. Generell lässt sich hierzu festhalten, dass solche Reformen in der Regel im Rahmen der traditionellen Lehrerbildung stattfinden, womit sich die Wirkungen von Inter-

vention und generellem Ausbildungsprogramm, deren Bestandteil die Reform ist, nicht trennen lassen (vgl. Wideen, Mayer-Smith & Moon 1998, 149). Auf den ersten Blick erfolgreicher scheinen Reformen zu sein, die über längere Zeit (mindestens ein Jahr) laufen und in allen Bestandteilen des Ausbildungsprogramms auf ein spezifisches Ziel fokussiert sind (vgl. ebd., S. 151f.). Aber auch hier muss festgehalten werden, dass die empirischen Belege sehr gering sind, da die Wirkungen in der Regel anhand von Selbstaussagen der teilnehmenden Studierenden erhoben werden und nicht anhand von Verhaltensbeobachtungen.

Die vielleicht umfassendste Forschung liegt zur Frage des Verhaltenstrainings bei Lehrerinnen und Lehrern vor. In Kursen – häufig weniger Bestandteil der Erstausbildung als der Lehrerfortbildung – wird versucht, Teilaspekte des Lehrerhandelns unterrichtsorientiert im Hinblick auf neue Konzepte zu verändern (‘Microteaching’). Waren mit diesen Trainings anfangs hohe Erwartungen verknüpft, zeigen Evaluationen eher begrenzte Wirkungen (vgl. Zeichner 1986). Schlee (1992, 560): „Je sorgfältiger die Untersuchungen geplant wurden und je größer das Bemühen um den Nachweis einer externen Validität war, desto nüchterner wurde das Verhältnis von Aufwand und dauerhaftem Effekt beurteilt.“ Schlees Hauptkritik richtet sich allerdings eher auf die mangelnde theoretische Fundierung der bisher vorhandenen Trainingsansätze und die fehlende Verzahnung der Trainings mit der Lehrerbildung. Eine systematische Integration von Verhaltenstrainings in die Lehrerbildung – obwohl ebenso wenig auf ihre Wirksamkeit überprüft – gelang im Übrigen nie. Ohne detaillierte Aussagen machen zu können, lässt sich anhand zahlreicher Studien als Problem in Bezug auf die handlungsbezogene Ausbildung aber festhalten, dass Lehramtsstudierende in ihrer Ausbildung in der Regel mit vollständig anderen Lehrkonzepten vertraut gemacht werden sollen, als sie als Schülerinnen und Schüler selbst erlebt haben, als sie als *student teacher* in Praxisphasen parallel erleben und als sie an der Hochschule unterrichtet werden. Ginsburg und Clift (1990) stellen hierzu die Hypothese auf, dass möglicherweise der darin verborgene ‚heimliche Lehrplan‘ das inhaltlich intendierte Curriculum konterkariert (vgl. auch Zeichner 1986). Und Schlee (1992, 560) kritisiert, dass „in der Lehrerbildung unüberprüfte Lehr-Lernformen als Prozessvariable die Regel sind, während solche, deren Wirksamkeit zwar nicht generell, aber doch hinreichend und mit ermutigender Tendenz empirisch nachgewiesen werden konnte, die Ausnahme bilden“.

Zur Frage der konkreten Wirksamkeit von Praxiserfahrungen in der Lehrerausbildung liegen zwar umfangreiche Studien vor, die sich allerdings nur schwer auf einen Nenner bringen lassen, da sie ihre Fragestellungen oftmals unterschiedlich akzentuieren, diese mit unterschiedlichen Methoden erheben (in der Regel qualitative Fallstudien) oder unterschiedliche theoretische Rahmenkonzepte verwenden (vgl. Zeichner 1986). Konsistent scheint die folgende Erkenntnis zu sein, die Wideen, Mayer-Smith und Moon (1998, S. 154) festhalten und die auch mit den oben dargestellten Erkenntnissen zu *prior beliefs* kompatibel ist: „The student teachers were initially concerned with the affective side of learning, wanting to establish warm, positive classroom atmospheres. When they moved from teaching small groups to whole classes, their views began to change to focus on classroom management.“ Der Handlungsdruck, dem sie ausgesetzt sind, lässt sie zudem an der ‚Verwendbarkeit‘ von Theorie zweifeln – für sie geht es eher um ‚Überleben‘, was ihnen offensichtlich mit traditionellem Verhalten leichter zu bewerkstelligen erscheint als mit dem in der bisherigen Lehrerausbildung erworbenen Wissen. Im Hinblick auf die Betreuung, die die Studierenden im Prozess des *student teaching* erfahren, muss offensichtlich zwischen akademischer und praktischer Betreuung differenziert werden. Während Erstere nach Kagan (1992, 148) „oversimplify the reality of student teaching and ignore the many social and pedagogical variables that can affect a novice’s instructional decisions“, orientieren sich die Studierenden gern an den betreuenden Lehrpersonen als Vorbildern.

1.3.4 Zwischenfazit: Wie wirksam ist die Lehrerbildung?

Fokussiert man die Frage der Wirksamkeit von Lehrerbildung darauf, ob durch sie eine hohe Qualität von Lehrerhandeln erreicht wird oder ob sie nicht Resultat von Persönlichkeitseigenschaften („Talent“) bzw. gar erst, wie von der Expertiseforschung behauptet (vgl. z.B. Berliner 1994), von langjähriger Berufspraxis („Übung“) ist, muss man feststellen, dass eine solche Frage (wenn sie auch im Zentrum vieler Auseinandersetzungen steht) auf der Basis der vorhandenen empirischen Befunde nicht beantwortet werden kann. Fraglich ist, ob das in dieser Form jemals möglich sein wird. Berliner (2001, 465) weist darauf hin, dass es sich schon bei Talent um „an extremely complicated interaction of many human characteristics“ handelt: „These might include sociability, persuasiveness, trustworthiness, nurturant style, ability to provide logical and coherent stories and explanations, ability to do more than one thing at a time, physical stamina, the chance to ‚play teacher‘ with a younger sibling or playmate, and so forth.“ Zudem wird mit einer solchen Fokussierung möglicherweise die Bedeutung des Kontextes von Lehrerhandeln im Sinne von Arbeitsplatzbedingungen, Schulentwicklung und Schulklima, gesellschaftlichen Erwartungen und Bildungspolitik systematisch unterschätzt, die sich in Selbst- und Fremdwahrnehmungen von Lehrpersonen niederschlagen und ihr Handeln beeinflussen (vgl. ebd., 466).

Darüber hinaus ist von normativen Unterschieden auszugehen, was als Kriterium hoher Lehrqualität akzeptiert wird. Standardisierte Erhebungen von Schülerleistungen stellen vermutlich auch in Bezug auf die Wirksamkeit der Lehrerbildung den weit reichendsten Konsens dar, da Schülerleistungen als „ultimate goal of teacher education“ (Cochran-Smith 2001, 530) unbestritten sein dürften. Allerdings endet die Einigkeit vermutlich bei genau dieser – oberflächlichen – Beschreibung. Cochran-Smith (ebd.): „If we attempt to describe the relationship between teacher learning and professional practice, attempt to explain what we mean by teacher learning and student learning, attempt to elaborate the theoretical bases and consequences of the kind of student learning we are trying to account for [...], we uncover differences, some of which represent deep philosophical and political divides.“ Darüber hinaus würde das Rekurrenieren auf Schülerleistungen bedeuten, nur einen indirekten Weg der Erfassung von Lehrqualität – über das Handeln von Schülern, das u.a. von ihrer sozio-kulturellen Herkunft, peer-Effekten etc. abhängt – zu beschreiten. Im Vergleich zur Forschung in anderen Domänen ist das ein bedeutsamer und nicht unproblematischer qualitativer Unterschied.

Schließlich ist darauf hinzuweisen, dass sich auf der Basis der vorliegenden Befunde nur schwer erkennen lässt, wie Lehrerausbildung strukturell und inhaltlich optimalerweise auszu-sehen hat. Wie viel Fachwissenschaft, Fachdidaktik und Erziehungswissenschaft werden denn für welche Schulform genau benötigt? Ist ein stark reguliertes Kerncurriculum sinnvoll oder eher eine freie Wahl aus einem breiten Angebot? Wie viel und an welcher Stelle ist Praxis vonnöten? An welcher Institution sollte die Lehrerausbildung stattfinden, wie lange sollte sie dauern, in wie vielen Phasen, für alle Lehrpersonen gemeinsam oder nach Stufen, Fächern oder Sonstigem getrennt? Diese Fragen lassen sich bisher – auf empirischer Basis – nicht beantworten. Allerdings gilt das sowohl für Verfechter der bestehenden Formen als auch für ihre Gegner, die ebenso wenig empirische Belege für die Sinnhaftigkeit anderer Modelle vorweisen können. Am stärksten zu erkennen ist, dass unabhängig von Persönlichkeitsfaktoren „the only characteristic that seem to importantly relate to complex pupil learning is the content and not the structure of the teacher development programs (within reasonable limits)“, dass aber – und diese Feststellung gilt für Deutschland ebenso wie für die USA – „paradoxically, it is program structure what seems to drive teacher education reform in the U.S. and abroad“ (Tatto 2000, 12).

1.3.5 Empirische Befunde zu Problemen von Standards und Evaluationen

Im Rahmen einer Zusammenfassung empirischer Befunde zur Wirksamkeit der Lehrerbildung, bei der im Kern immer wieder die Lückenhaftigkeit entsprechender Forschung in

Deutschland festgestellt werden muss, interessiert auf einer Meta-Ebene auch die Fragestellung, welche Veränderungen/ Folgen/ Probleme verstärkte Überprüfungen der Wirksamkeit haben würden (= die Einführung von Standards und Evaluationen, wie sie derzeit für die Lehrerbildung – vgl. Terhart 2002 und als konkretes Beispiel PLAZ 2003 – sowie für die Schule – vgl. für Deutschland Tenorth 2001, Buhlman, Wolff & Klieme 2003, KMK 2003a, b, c; vgl. für die Schweiz Oser & Oelkers 2001 – diskutiert werden). Während die Chancen eines derartigen Vorhabens auf der Hand liegen – in Bezug auf Standards: Herausarbeitung von Kompetenzniveaus und deren Konkretisierung anhand von Beispielaufgaben; in Bezug auf ihre Evaluation: systematische Rückmeldungen zur Qualität der eigenen Arbeit im Vergleich zu anderen, so dass nach Ansatzpunkten für (weitere) Niveausteigerungen gesucht werden kann – gilt es, rechtzeitig auch mögliche Probleme in den Blick zu nehmen. Dies kann konsequenterweise erneut nur am Beispiel der USA stattfinden. *Accountability* – die regelmäßige Überprüfung der Funktionsfähigkeit des Bildungssystems auf der Ebene der Einzelinstitution und -personen – ist seit mehreren Jahrzehnten das zentrale Schlagwort der US-amerikanischen Bildungspolitik; in ‚Standards‘ werden dabei zwei ansonsten nicht vereinbare Ansprüche zusammengeführt: Dezentralisierung in Form weitgehender Autonomie der Schulen und der Lehrerbildungseinrichtungen bei gleichzeitiger Sicherung zentraler Qualitätsansprüche (vgl. Böttcher 2003, 154). Die Lehrerbildung wird dabei als treibende Kraft für avisierte Schulreformen gesehen. Criblez (1998, 53f.): „Das Anliegen der Schulreform ist wahrscheinlich im ganzen englischsprachigen Raum eines der Hauptmotive für die Reform der Lehrerbildung. Die Lehrerbildung wird so zum Mittel der Schulreform.“ Um ein vollständiges Bild zu erhalten, muss daher neben der Ebene der Lehrerbildung häufiger auch auf die Schulebene Bezug genommen werden.

- *Entwicklungsschritte in der Etablierung von Standards und Evaluationen*

Für die USA hatte der 1983 erschienene Bericht der *National Commission on Excellence in Education* ‚A Nation at Risk‘ eine vergleichbare Wirkung wie die Ergebnisse der TIMS- und PISA-Studien in Deutschland, indem detailliert auf die defizitären Lernergebnisse der amerikanischen Schülerinnen und Schüler aufmerksam gemacht wurde (vgl. NCEE 1983). Unter Verweis auf die nationalen ökonomischen Folgen, die ein niedriger Bildungsstand im globalen Wettbewerb haben würde, wurde eine Ablösung der bisher geltenden Minimalstandards zugunsten von „more rigorous“ Standards von „excellence in education“ – und die Etablierung eines *high-stakes testing* – gefordert (ebd.). In schneller Folge kam es zu einer Reihe von staatlich initiierten Reformen, die darauf gerichtet waren, die Schulqualität – definierbar in den drei Kategorien kognitive Leistungen der Schülerinnen und Schüler, Chancengleichheit und öffentliche Kosten – zu erhöhen. Im Mittelpunkt stand die Entwicklung von Standards ‚What students [Schülerinnen und Schüler; S.B.] should know and be able to do‘. Bis heute haben mindestens 47 Staaten solche Standards entwickelt (vgl. Darling-Hammond 2003); fast alle Staaten verbanden ihre Überprüfung mit Tests (vgl. Amrein & Berliner 2002a, 4).

Vergleichbar der Diskussion in Deutschland, die hier fünfzehn bis zwanzig Jahre später einsetzte, blieb die Rolle der Lehrpersonen zunächst weitgehend unbeachtet, obwohl bereits bekannt war, dass die Lehrerausbildung einer der Studiengänge war, die am wenigsten von leistungsstarken Schülerinnen und Schülern gewählt wurden (vgl. NCEE 1983). Diese Randständigkeit lehrerbezogener Evaluationen änderte sich mit dem Engagement weiterer *think tanks* in diesem Zusammenhang. 1986 gründeten die Präsidenten der Universitäten mit Lehrerbildung die ‚Holmes Group‘ und veröffentlichten zunächst den Bericht ‚Tomorrow’s Teachers‘, in dem – vergleichbar zu den schülerbezogenen Standards – Standards für die Lehrerbildung und ihre Evaluation gefordert wurden (vgl. The Holmes Partnership 1986), und dann den Bericht ‚Tomorrow’s Schools of Education‘, in dem insbesondere auf eine Prüfung der fachwissenschaftlichen Qualifikation Wert gelegt wird, die als einer der Schwachpunkte der amerikanischen Lehrerausbildung gilt (vgl. The Holmes Partnership 1995). Parallel forderte die Carnegie Foundation in ‚A Nation Prepared. Teachers for the 21st Century‘ ebenfalls Standards

für Lehrerinnen und Lehrer: ‚What teachers should know and be able to do‘. Zudem wurde der Aufbau einer bundeseinheitlichen Institution, eines *National Board for Professional Teaching Standards*, vorgeschlagen, die eine auf Standards basierende Zertifizierung der im Beruf befindlichen Lehrerinnen und Lehrer vornehmen sollte. Vergleichbar zur Schulreform kam es in der Folge auch in Bezug auf die Lehrerbildung zu Neuerungen, u.a. wurden die Einschreibungsvoraussetzungen für ein Lehramtsstudium deutlich hoch gesetzt (z.B. Erfordernis einer Mindestpunktzahl im High School-Abschluss und Eingangstests).

1987 wurde das von der Carnegie Foundation vorgeschlagene *National Board for Professional Teaching* gegründet, das seither auf der Basis von Portfolios und Tests Lehrpersonen zertifiziert. Für mehr als 30 Unterrichtsfächer, differenziert in vier Altersgruppen (*pre-school*, *elementary school*, *lower secondary school* und *high school*), wurden Standards entwickelt, die die fundamentalen Anforderungen an Lehrpersonen fachlich differenziert spezifizieren. Diese lauten: „a broad grounding in the liberal arts and sciences; knowledge of the subjects to be taught, of the skills to be developed, and of the curricular arrangements and materials that organize and embody that content; knowledge of general and subject-specific methods for teaching and for evaluating student learning; knowledge of students and human development; skills in effectively teaching students from racially, ethnically and socioeconomically diverse backgrounds; and the skills, capacities and dispositions to employ such knowledge wisely in the interest of students“ (NBPTS 2002, 2). Die Standards sind professionstheoretisch fundiert, indem versucht wird, den Paradoxien des Lehrberufs gerecht zu werden: Theoriewissen und Situationsbezug werden ebenso ausgewogen berücksichtigt wie die sich aus so unterschiedlichen Aufgaben wie Unterricht, Erziehung und Schulentwicklung ergebende Anforderungen (vgl. ebd.). Bis Ende der 90er Jahre erkannten 34 Staaten die NBPTS-Standards als formalen Maßstab für die Qualität von Lehrpersonen an, 26 Staaten fördern durch Bonussysteme die Zertifizierung ihrer Lehrpersonen (Zuschuss zu den Gebühren, bessere Aufstiegschancen etc.), acht Staaten gehen sogar soweit, eine Gehaltserhöhung damit zu verbinden (vgl. Darling-Hammond 2001, 768).

Eine erneute Zäsur stellt das Jahr 1998 dar, in dem der Bericht ‚A Nation Still at Risk‘ der Thomas H. Fordham Foundation erschien. In diesem wird herausgehoben, dass nicht mehr die Gefahr von ökonomischem oder technischem Rückstand als Motor für Bildungsanstrengungen angeführt werden brauche, da sich die nationale Situation in dieser Hinsicht gewandelt habe. Allerdings wurde auf die noch immer bestehenden inneramerikanischen Verwerfungen hingewiesen: die Ungleichheit der Bildungschancen abhängig von sozialer oder ethnischer Herkunft. Aus der Perspektive der Bürgerrechte wurde der Anspruch jeden einzelnen Kindes auf einen Mindeststandard an Bildung hervorgehoben, und zwar als *realisierter* Anspruch. Diese Argumentation hat in den USA eine starke Tradition, die sich zuletzt im kürzlich ergangenen Urteil des Obersten Gerichtshofes zur bevorzugten Zulassung von Studierenden aus ethnischen Minderheiten widerspiegelt. In diesem wird ein staatliches Interesse an ethnischer Diversität unter den Studierenden formuliert, „in order to cultivate a set of leaders with legitimacy in the eyes of the citizenry. [...] Effective participation by members of all racial and ethnic groups in the civil life of our nation is essential if the dream of one nation, indivisible, is to be realized“ (zit. nach Greenhouse 2003). Die dahinter stehende Grundidee ist, dass unabhängig von ihrer sozialen oder ethnischen Herkunft alle Menschen gleich geboren werden. Wenn es dennoch zu systematischen Ungleichverteilungen bei Schulabschlüssen u.ä. kommt, müssen diese sozial hergestellt sein und sind daher auch sozial (durch *affirmative action* z.B.) wieder auszugleichen. Den Lehrpersonen und der Lehrerbildung wurde erneut eine zentrale Rolle zugeschrieben, indem im Unterricht eine individuelle Förderung aller Schülerinnen und Schüler in der Form stattfinden solle, dass keine systematischen Ungleichheiten mehr auftreten.

Im Jahr 2001 hatte die Bush-Regierung diese Ansprüche gesetzlich im ‚No Child Left Behind Act‘ gefasst. Reformen in der Lehrerbildung kommt danach ein hoher Stellenwert zu. Spätes-

tens am Ende des Schuljahrs 2005/06 sollen alle Lehrerinnen und Lehrer „highly qualified“ (U. S. Department 2002, vii) sein. Zwei Merkmale wurden als Definition hierfür festgelegt: „full state certification“ und „solid content knowledge“. Seit 2002 müssen daher alle neuen Lehrpersonen an *elementary schools* Tests in Mathematik, Lesen und Schreiben ablegen, und zwar im Hinblick auf ihr fachliches Wissen und Können. Lehrpersonen an *middle* bzw. *high schools* müssen sich einem Wissenstest in Bezug auf die zu unterrichtenden Fächer unterziehen oder mindestens einen BA-Major hierin besitzen. Das Prinzip der individuellen Zertifizierung wird auf der institutionellen Ebene durch ein weiteres Zertifizierungsverfahren ergänzt. Die 1.300 lehrerausbildenden Einrichtungen der USA können sich bereits seit den 50er Jahren akkreditieren lassen. Allein 600 Institutionen haben dies bei dem National Council for the Accreditation of Teacher Education (vgl. NCATE 2003) vorgenommen, die als Akkreditierungsagentur in dem Ruf steht, besonders strikte Standards aufgestellt zu haben (vgl. Abell Foundation 2001a, 34). Das NCATE, in der 33 nationale Organisationen Mitglieder sind, ist von der US-Regierung als Akkreditierungsagentur anerkannt; die von ihr akkreditierten Institutionen bilden rd. zwei Drittel aller Lehramtsstudierenden aus (vgl. Darling-Hammond 2001, 753). Gleichzeitig etabliert der ‚No Child Left Behind Act‘ ein umfassendes Testsystem auch auf Schülerebene. In Ergänzung zu den bereits bestehenden Tests sollen nun auch die Klassen 3 bis 8 und alle High School-Klassen jährlichen Tests unterzogen werden. Ziel ist, die einzelnen Schulen verantwortlich machen zu können, wenn sie vergleichsweise hohe Drop out-Raten haben oder systematisch bei bestimmten ethnischen Populationen geringe Schülerleistungen erreichen.

- *Probleme von Standards und Evaluationen*

So überzeugend die Ansprüche sind, die mit der Entwicklung von Standards und dem No Child Left Behind Act aufgestellt wurden, so kritisch sind allerdings einige Aspekte im Detail zu betrachten. Darling-Hammond (2003, 1) stellt fest: „In a number of states, the notions of standards and ‚accountability‘ have become synonymus with mandates for students testing.“ Und weitergehend ist festzustellen, dass Testergebnisse (obwohl für diagnostische Zwecke nicht geeignet, da zu einem Themengebiet zu wenig Aufgaben enthalten sind) zunehmend entscheidend wurden für die weitere individuelle Schulkarriere oder für die finanzielle Zuweisung an Schulen (so genanntes *high-stakes testing*; vgl. Amrein & Berliner 2002a, b). Die dahinter stehende Idee ist, dass auf diese Weise gute Schulen (z.B. über zusätzliche finanzielle Mittel oder vordere Plätze in öffentlichen Rankings) belohnt und schlechte Schulen (z.B. drohende Schließungen) motiviert würden, sich zu verbessern. Stipendien bzw. Wiederholungen von Klassen sollen dieselben Funktionen auf der Ebene der Schülerinnen und Schüler erfüllen. Für diese ist die wohl gravierendste Konsequenz, dass die Vergabe des High School-Diploms (in der Regel nach Klasse 10 oder 11) mittlerweile in 18 Staaten in Ergänzung zum bestehenden System ‚studienbegleitender Prüfungen‘ an das Bestehen der Leistungstests gebunden wurde (*do-or-die exams*; vgl. Dillon 2003b). Die Quote derjenigen, die hieran scheitern, liegt bei bis zu 12 Prozent (vgl. Amrein & Berliner 2002a, 18).

Bereits wenige Jahre nach ihrer Einführung wurde die Kritik an dieser ‚High stakes‘-Testpolitik immer stärker, weil die intendierten Ziele der Leistungssteigerung offensichtlich nicht erreicht werden. So zeigt eine umfassende Analyse, dass nach Einführung des *high stakes testing* die Schulabschlussraten in mehr Staaten gefallen als gestiegen und dass umgekehrt die Dropoutraten in mehr Staaten gestiegen als gefallen sind (vgl. Amrein & Berliner 2002b, 33). Zudem müssten die Schülerinnen und Schüler – wenn das Ziel der Leistungssteigerung erreicht würde – in anderen üblichen Tests (v.a. in den Zulassungstests für die Colleges, die von 60 bis 70 Prozent eines jeden High school-Jahrgangs absolviert werden) bessere Leistungen als vor Einführung der High stakes-Tests erzielen. Genau dies ist aber offensichtlich (abgesehen von kurzfristigen Effekten im Jahr nach Einführung des Testsystems) auf lange Sicht gesehen nur sehr selten der Fall (vgl. Amrein und Berliner 2002a, 27). Eine andere Konsequenz aus diesem System steht dem ursprünglichen Anliegen der Verbesserung des Bildungs-

zugangs für alle geradezu entgegen: „In states where ‚high stakes testing‘ is the primary policy reform, disproportionate numbers of minority, low-income, and special needs students have failed tests for promotion and graduation.“ (Darling-Hammond 2003, 1).

Neben dem Nichterreichen ursprünglicher Ziele sind gleichzeitig zahlreiche unerwünschte Nebenwirkungen des *high-stakes testing* festzustellen. Ein besonders drastisches Beispiel könnte der texanische Schulbezirk Houston sein, wo der gegenwärtige Erziehungsminister der Bush-Regierung, Paige, zuvor für die Bildungspolitik verantwortlich war. Das von ihm hier erzeugte ‚Wunder‘ von Texas hat laut New York Times erheblich dazu beigetragen, dass Bush Präsident werden konnte (vgl. Schemo 2003a, b). Houston war – wie Texas überhaupt – eine der Regionen mit der geringsten Zahl an Schulabschlüssen, besonders schwachen TIMSS-Ergebnissen und besonders großen ethnischen sowie sozialen Diskrepanzen. Paige testete in seiner Zeit als Superintendent die Maßnahmen, die später mit dem ‚No Child Left Behind Act‘ bundesweite Gültigkeit bekamen – mit dem geradezu sensationellen Ergebnis, dass es ihm gelang, die Dropout-Rate fast gegen Null zu senken. Drei Jahre später stellte sich anlässlich einer Überprüfung dieser Ergebnisse an 16 ausgewählten *middle* und *high schools* allerdings heraus, dass im Schuljahr 2000/01 mehr als die Hälfte der 5.500 Jugendlichen, die diese Schule verlassen hatten, als *dropout* hätte gemeldet werden müssen (vgl. Schemo 2003a). Die Schulleiter hatten schlicht falsch gemeldet, dass die entsprechenden Schülerinnen und Schüler an andere Schulen gewechselt seien etc. (vgl. Schemo 2003b).

Eine weitere unerwünschte Nebenwirkung ist die teilweise mangelnde Testfairness, die sich in sehr hohen und jährlich stark schwankenden Durchfallquoten ausdrückt. Obwohl für die Konstruktion der Tests ein hoher Aufwand betrieben wird (mehrere Begutachtungsdurchgänge unter unterschiedlichen Perspektiven, z.B. methodische Korrektheit, Verständlichkeit, inhaltliche Angemessenheit, sowie ein anschließender Feldtest; vgl. Dillon 2003a), gelingt es offensichtlich nicht, regelmäßig Tests zu entwickeln, die es einer großen Mehrheit der Schülerinnen und Schülern ermöglicht, auf der Basis ihres Unterrichts zu bestehen. Unmittelbar mit den hohen Durchfallquoten verbunden ist eine andere unerwünschte Nebenwirkung: die Verlagerung unterrichtlicher Schwerpunktsetzungen. So sind im Staat New York mit Ausnahme von Schulen, die besonders niedrige Durchfallquoten haben, für alle anderen Schulen die Standards in Englisch und Mathematik hoch gesetzt worden, um mehr Schülerinnen und Schülern das Bestehen der *high stakes*-Tests zu ermöglichen. Die neuen Standards lassen sich in der Regel nur durch zusätzlichen Unterricht – also eine Verlängerung des Schultages, der die Gewerkschaften zustimmen müssen – oder durch Streichung anderer Kurse erreichen. Und genau dies geschieht derzeit (vgl. Herszenhorn 2003). Davon betroffen sind v.a. die achten Klassen, an deren Ende die Tests in Englisch und Mathematik stehen.

Amrein und Berliner (2002b, 35ff.) berichten darüber hinaus von sprunghaft angestiegenen Wiederholungsraten in der Klasse 9 und von ebenso sprunghaft angestiegenen Prozentsätzen an *special needs* und *limited english proficient* Schülerinnen und Schülern, die von der Testung ausgenommen werden können. Geringere bzw. höhere Testergebnisse erklären sich also durch höhere bzw. geringere Teilnehmerraten (Amrein & Berliner 2002a, 33: „illusion from exclusion“). Im Hinblick auf die Unterrichtsinhalte scheinen sich zudem Lehren und Lernen in Schulen mit unterdurchschnittlichen Testergebnissen zu einem *teaching to the test* zu verschieben (Amrein & Berliner 2002b, 40). So werden die Tests gleichzeitig immer weniger valide im Hinblick auf ihr ursprüngliches Anliegen (= Indikator, ob im Unterricht die gewünschten Standards erreicht werden). Kein Einzelfall scheinen auch unzulässige Hilfen zu sein. Amrein und Berliner (ebd., 46) folgern: „As stakes attached to tests become more severe, the likelihood that school personnel will cheat on tests increases.“

Diese unerwünschten Nebenwirkungen sind in erster Linie auf die unmittelbare Verknüpfung von Testergebnissen und Konsequenzen (z.B. individuell über Knicke in der Schullaufbahn oder systemisch für eine einzelne Schule über finanzielle bzw. personelle Einbußen) zurückzuführen. In den USA zeigt sich damit die Gültigkeit des Heisenbergschen Unschärfepinzips,

das schon häufiger auf soziale Prozesse angewendet wurde, auch für *high stakes testing*: „The more important that any quantitative social indicator becomes in social decision-making, the more likely it will be to distort and corrupt the social process it is intended to monitor.“ (ebd., 5) Diese Erfahrungen sprechen gegen die in der Bundesrepublik jüngst erhobene Forderung, notwendige Evaluationen der Lehrerbildung müssten gleichzeitig „folgenreich“ (Terhart 1999, 148) für Universitäten und Studienseminare sein. Was sich allerdings nicht belegen lässt, sind Befürchtungen, bereits die *Einführung* von Tests – beispielsweise zu Rückmelde- und Vergleichszwecken ohne unmittelbare Konsequenzen – würde unerwünschte Nebenwirkungen (*teaching to the test* z.B.) mit sich bringen, ja solche würden sogar die Regel sein (vgl. Rolff 2003).

1.3.6 Aufgaben für die Zukunft

Vor dem Hintergrund der dargelegten empirischen Befunde zur Wirksamkeit der Lehrerbildung und der Erfahrungen mit Standards und Evaluationen in den USA lassen sich die anstehenden Aufgaben von empirischer Forschung zur Lehrerbildung in drei Schlagworten zusammenfassen: 1. international einheitliche Sammlung von Strukturdaten, 2. standardbasierte, international vergleichende Untersuchung der Wirksamkeit von Lehrerbildung sowie 3. die Verknüpfung von empirischer Bildungsforschung, Bildungstheorie und Bildungsgeschichte.

- *International einheitliche Sammlung von Strukturdaten*

Der ersten Aufgabe stellt sich derzeit die OECD, indem sie ihre Bildungsberichte erheblich ausweitet (vgl. OECD 2002a) und indem sie das Programm ‚Attracting, Developing and Retaining Effective Teachers‘ aufgelegt hat (vgl. OECD 2002b). Als eine Konsequenz aus dem Entschluss der OECD-Bildungsminister, den Blick in Zukunft stärker auf Faktoren zu lenken, die Schülerleistungen beeinflussen, als diese nur über Tests zu erheben, wurde eine international einheitliche – auf die Faktoren gesellschaftliche Rahmenbedingungen, institutionelle Strukturen und Arbeitsbedingungen vor Ort abhebende – Analyse des Lehrberufs beschlossen (vgl. ebd.). Ein erster deutscher Bericht liegt bereits vor (vgl. KMK 2003d), ebenso eine vertiefende Stellungnahme von Bildungsexperten (vgl. Döbrich u.a. 2003). Bei dem KMK-Bericht handelt es sich um eine eher formal-administrative, in erster Linie auf die Sammlung von Strukturdaten und Organisations- bzw. Entscheidungsabläufe konzentrierte Darstellung, die insofern ihren Wert hat, als die entsprechenden Angaben bisher nur verstreut dokumentiert waren. Döbrich u.a. haben ihre Stellungnahme dagegen stärker bewertend angelegt und sie legen – auf der Basis ihrer Expertise in Fragen der Lehrerbildung – Traditionen, Diskussionslinien, Defizite und Reformansätze offen. Im Wesentlichen werden acht „Defizite der gegenwärtigen Form der Ausbildung“ (ebd., 24) identifiziert: Desinteresse der Fachwissenschaftler an der Lehrerbildung (vgl. hierzu im Einzelnen Kap. 5.1 und 5.2 des Handbuchs), Mangel an fachdidaktischer Forschung (Kap. 5.3), Fehlen eines Kerncurriculums v.a. in den Erziehungswissenschaften (Kap. 5.4), Praxisferne der Lehrerausbildung (Kap. 5.5), Desintegrationsprozesse in der Ersten Phase (Kap. 3.1), keine systematische Qualifizierung der Ausbilder in der Zweiten (Kap. 3.2) und Dritten Phase (Kap. 3.3), mangelnde Abstimmung zwischen den drei Phasen (Kap. 3.4) und schließlich Fehlen von Standards in der Lehrerbildung (Kap. 2.3).

- *International vergleichende Untersuchung der Wirksamkeit von Lehrerbildung*

So wesentlich die Zusammenstellung solcher Informationen auch ist, um mehr über das Bildungssystem zu erfahren, zu dem bisher nicht hinreichend Daten vorhanden sind, sollte die empirische Forschung dabei allerdings nicht stehen bleiben. Zu fragen ist, was ,effektive

Lehrpersonen' sind und wie ihre Qualifikationen mit der Lehrerbildung zusammenhängen, wobei in Bezug auf Letztere wiederum zwischen – inhaltlich und strukturell – avisiertem (Lehr-)Programm und tatsächlich implementiertem Lernprozess (auch als *opportunity to learn* zu bezeichnen) zu unterscheiden ist. Eine theoriegeleitete Entwicklung von international akzeptierten Standards (bzw. ggf. eine Operationalisierung vorliegender Modelle) und ihre systematische Prüfung in international einheitlich angelegten Längsschnittstudien, die Lernvoraussetzungen, Lehr- und Lernprozesse sowie Wirkungen erfassen, sind die beiden zentralen Herausforderungen an die empirische Lehrerbildungsforschung der kommenden Jahre. Tatto (2000, 24f.) stellt hierzu zwei, nicht leicht miteinander zu vereinbarende Aspekte in den Mittelpunkt: „There is a dearth of sound cross-cultural research on the influence of teacher preparation on teacher quality and pupil learning“, wobei angesichts der hohen kulturellen Bindung von Bildungsprozessen gleichzeitig kontextuelle Differenzen Beachtung finden müssen. Das ist „aufwändig und methodisch anspruchsvoll, weil es quantitative mit qualitativen Daten verbindet, mehrere Ebenen und Kontexte der Lehrerausbildung miteinander in Beziehung setzt und überdies in seinem Zentrum ein *Theorieproblem* bearbeitet, das zur Sprengung der konventionellen Evaluationsansätze von *teacher education programs* führen kann“ (von Prondczynski 2001, 130).

Eine entsprechende internationale Vergleichsstudie wird derzeit von der International Association for the Evaluation of Educational Achievement (IEA) vorbereitet (vgl. Tatto u.a. 2003; zur Konzeption der deutschen Beteiligung vgl. Blömeke 2003). In dieser soll erstmals über Befragungen hinausgegangen werden (worauf sich beispielsweise Brouwer und ten Brinke (1995a, b) in ihrer nationalen Längsschnittstudie für die Niederlande beschränken), indem Lehrerkognitionen getestet werden und Lehrerhandeln systematisch beobachtet wird. Inwieweit tatsächlich auch Schülerleistungen erhoben werden, ist vor dem Hintergrund ungeklärter Wirkungszusammenhänge (s.o.) noch offen.

Eine langfristig auch in der Bundesrepublik Deutschland bedeutsame Frage ist in diesem Zusammenhang allerdings, inwieweit Lehrpersonen überhaupt über den angenommenen Weg der Lehrerbildung in den Schuldienst kommen und wie lange sie in ihrem Beruf verbleiben. Für die Bundesrepublik mit ihrer strikten staatlichen Kontrolle der Eingangsqualifikationen stellt sich die Frage so genannter ‚Quereinsteiger‘ zwar erst seit wenigen Jahren (bzw. seit rd. 40 Jahren das erste Mal wieder). In den USA handelt es sich aber um vielfach beklagte Probleme. Hier verlässt schätzungsweise die Hälfte (!) aller Lehrpersonen innerhalb der ersten sechs Jahre den Beruf wieder (vgl. LaTurner 2002, 654). Zudem sehen die staatlichen Regelungen des Berufseinstiegs sehr unterschiedlich aus. Auf der Basis einer Langzeitstudie über den beruflichen Verbleib der BA-Absolventen der Jahre 1992/93 arbeitet LaTurner für die USA vier Wege von Lehrpersonen heraus, die in den Klassen 6 bis 12 Mathematik oder Naturwissenschaften unterrichten: 1. Abschluss eines lehrerbildenden Studiengangs mit erziehungswissenschaftlicher und mindestens 18 Semesterwochenstunden (SWS) fachwissenschaftlicher Ausbildung; 2. zwar Erfüllung der 18 SWS fachwissenschaftlichen, aber nicht der erziehungswissenschaftlichen Voraussetzungen; 3. Erfüllung der erziehungswissenschaftlichen, aber nicht der fachwissenschaftlichen Voraussetzungen; 4. Erfüllung keine der Voraussetzungen (vgl. ebd., 656). Weniger als die Hälfte der Stichprobe (46,4%) fällt in Gruppe 1, während jeweils rund 15% den Gruppen 2 und 3 und sogar jede vierte Mathematik- bzw. Naturwissenschafts-Lehrperson der Gruppe 4 zugerechnet werden muss (vgl. ebd., 658). Unterschwellig gibt es allerdings auch in der Bundesrepublik gegenwärtig das Problem, dass nicht in allen Fächern und Stufen formal ausgebildete Lehrpersonen unterrichten (vgl. Döbrich u.a. 2003), und zwar v.a. in Grund-, Haupt- und Sonderschulen, wo aus pädagogischen Gründen das Klassenlehrerprinzip vorherrscht, sowie z.T. in einzelnen Fächern der Sekundarstufe I, für die aufgrund von Bewerbermangel bzw. ausbleibenden Neueinstellungen nicht hinreichend fachlich ausgebildete Lehrpersonen zur Verfügung stehen, so dass der Unterricht ‚fachfremd‘

erfolgt. Darüber hinaus verfügen nur wenige Lehrpersonen über spezifische Qualifikationen, Schülerinnen und Schüler zu unterrichten, deren Muttersprache nicht Deutsch ist.

- *Verknüpfung von empirischer Bildungsforschung, Bildungstheorie und -geschichte*

Um angemessen eine empirische Prüfung der Wirksamkeit von Lehrerbildung durchführen zu können, ist schließlich eine Verknüpfung mit bildungstheoretischen und bildungshistorischen Diskursen unabdingbar, da sonst das Problem entsteht, ggf. komplexe Prozesse unangemessen zu reduzieren. So sieht beispielsweise von Prondczynski (2001, 94) schon im Heranziehen des Kriteriums der *teaching excellence* eine Reduktion der Lehrerbildung „auf eine kurzsichtige Praxis“. Angesichts der normativen Ausrichtung von Bildungsprozessen und ihrer nur indirekten Messmöglichkeiten ist in der Tat eines der gravierendsten Probleme empirischer Forschung, dass ‚Bildung‘ – je nach Begriffsverständnis – gar nicht, mindestens aber nur schwer messbar ist (wenn sie nicht auf ‚Leistung‘ – in der Regel zudem noch auf Kognitionen eingeschränkt – allein reduziert werden soll). In jedem Fall wird man den Anspruch an Forschung zur ‚Wirksamkeit‘ von Lehrerbildung einschränken müssen; hierbei handelt es sich eher um *Grundlagenforschung*. *Politikberatung* dagegen, wie es der Begriff der ‚Evaluation‘ in einem zentralen Verständnis nahe legt, erscheint – unabhängig von dem generellen Problem der Vermischung von Wertfreiheit im Begründungszusammenhang und legitimatorischen Interessen – nur schwer möglich. Cochran-Smith (2001, 532): „No matter what conclusions are ultimately drawn about the evidence, they will depend to a great extent on complex and prior decisions about what counts and how it counts as evidence, decisions that are based on values, goals, and priorities. In that sense, the debate will never be settled solely through empirical means“.

Die komplexe bildungstheoretische Diskussion angemessen zu berücksichtigen, stellt in vielerlei Hinsicht eine Chance dar. Seit rd. 250 Jahren wird in Deutschland eine intensive Diskussion zur Lehrerbildung geführt, in der nicht nur Ziele, sondern auch Wege zu ihrer Erreichung entfaltet werden. Da eine systematische Prüfung unterschiedlicher Modelle jedoch immer ausgeblieben ist, spricht von Prondczynski (2001, 98) nicht zu Unrecht von „einem Teufelskreis von Reform, Kritik, erneuter Reform und Kritik“, den es zu durchbrechen gilt. Zu empirischen Erkenntnissen zu kommen, kann nun in technokratischer Weise geschehen, das muss aber nicht zwangsläufig der Fall sein. So wie es ein Unterschied ist, ob Schülerleistungen wissens- oder kompetenzorientiert erfasst werden bzw. ob man Faktoren zum Unterrichtsklima berücksichtigt (weil man sie aus sozialen Gründen für bedeutsam hält) oder nicht (weil sie ggf. wenig zu Schülerleistungen beitragen), ist es ein Unterschied, ob man im Hinblick auf die Lehrerbildung Persönlichkeitsfaktoren mit aufnimmt oder sich auf die Erhebung des Fachwissens konzentriert und ob man die ‚Wirksamkeit‘ der Lehrerbildung an Lehrerkognitionen, Lehrerhandeln und/ oder Schülerleistungen bindet. Dahinter stehen unterschiedliche Modelle von und Ansprüche an Schule und Lehrerbildung. Tatto (2000, 2): „Empirical research on teacher quality has been strongly influenced by larger ideas about what knowledge is valued, what it means for pupils to learn that knowledge and what it means to teach it.“ Diese ‚larger ideas‘ lassen sich durchaus benennen – und damit in Variablen umsetzen.

Internationale Vergleiche öffnen darüber hinaus den Blick dafür, dass Systeme auch anders organisiert werden können, als es national üblich ist. Allerdings muss gleichzeitig auf die Zusammenführung von empirischer Forschung und Bildungsgeschichte hingewiesen werden. So ist gerade das Bildungssystem eines Landes untrennbar mit historischen Entwicklungen und sozio-kulturellen Einflüssen verbunden. Hat man die Geschichte im Blick, ist auch zu erkennen, dass sich viele Probleme – und häufig auch die entsprechenden Lösungsansätze – wiederholen, ohne dass nach Ursachen für das vormalige Scheitern gesucht wurde. Auch dies lässt sich am Beispiel der USA zeigen (vgl. Oelkers 2002a; vgl. auch Brügelmann 2003).

Schon in den 50er Jahren wurde in Bezug auf das Schulsystem eine Mängelliste aufgemacht, die eine analoge Beschreibung der gegenwärtig festgestellten Probleme sein könnte: defizitäre Lehrerbildung, Dropouts, ungleiche Bildungschancen für ethnische Minderheiten, hohe Kosten. Als Konsequenz wurde z.B. in Bezug auf die Lehrerbildung die Einführung von Standards gefordert, die überprüft werden sollten. Obwohl diese und andere ‚Reformen‘ durchgeführt wurden, wiederholen sich heute Kritik und Forderungen. Oelkers (2002a, 3): „Warum fünfzig Jahre später immer noch keine adäquate Problemlösung vorliegt, bleibt unklar und wird zumeist gar nicht erst diskutiert, weil die historische Dimension fehlt. Die Frage nach der Wiederholung des Neuen stellt sich erst gar nicht, so dass bestimmte Ideen immer neu ins Spiel gebracht werden können, ohne sich auf Erfahrungen beziehen zu müssen.“

Literatur

- [The] Abell Foundation [2001a]: Teacher Certification Reconsidered. Stumbling for Quality. Baltimore: Abell <http://www.abell.org/pubsitems/ed_cert_1101.pdf > [30.09.2003]
- [The] Abell Foundation [2001b]: Teacher Certification Reconsidered. Stumbling for Quality – Appendix. Review of Research Teacher Certification and Effective Teaching. Baltimore: Abell <http://www.abell.org/pubsitems/ed_cert_appendix_1101.pdf> [30.09.2003]
- Amrein, A. L./ Berliner, D. C. [2002b]: An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing. In: Education Policy Studies Laboratory. Tempe: Arizona State University <<http://www.asu.edu/educ/eps/EPRU/documents/EPSTL-0211-125-EPRU.pdf> > [30.09.2003]
- Amrein, A. L./ Berliner, D. C. [2002a]: High-Stakes Testing, Uncertainty, and Student Learning. In: Education Policy Analysis Archives 10 (18) <<http://epaa.asu.edu/epaa/v10n18>> [03.07.2003]
- Ball, D. L./ Bass, H. [2000]: Interweaving content and pedagogy in teaching and learning to teach. Knowing and Using Mathematics. In: Boaler, J. (Hrsg.): Multiple Perspectives on the Teaching and Learning of Mathematics. Westport: Ablex, S. 83-104 <<http://www-personal.umich.edu/~dball/BallBassInterweavingContent.pdf> > [30.09.2003]
- Ball, D. L./ Lubienski, S. T./ Mewborn, D. S. [2001]: Research on Teaching Mathematics. The Unsolved Problem of Teachers' Mathematical Knowledge. In: Richardson, V. (Hrsg.): Handbook of Research on Teaching. Washington: American Educational Research Association (4. Aufl.), S. 433-456
- Baumert, J./ Watermann, R./ Schümer, G. [2003]: Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. Ein institutionelles und individuelles Mediationsmodell. In: Zeitschrift für Erziehungswissenschaft (Opladen) 6 (1), S. 46-72
- Berliner [1994], D. C.: Teacher Expertise. In: Husen, T./ Postlethwait, T. N. (Hrsg.): The International Encyclopedia of Education. Bd. 10. London: Pergamon, S. 6020-6026
- Berliner [2001], D. C.: Learning about and learning from expert teachers. In: International Journal of Educational Research 35, 463-482
- Block, J. H./ Hazelip, K. [1995]: Teachers' beliefs and belief systems. In: Anderson, L. W. (Hrsg.): International encyclopaedia of teaching and teacher education. Oxford: Elsevier Science Ltd. (2. Aufl.), S. 25-28
- Blömeke [2002], S.: Universität und Lehrerausbildung. Bad Heilbrunn/Obb.: Klinkhardt
- Blömeke [2003], S.: Teacher preparation system in Germany. Paper presented at the IEA TEDS Expert Panel Meeting (16-20 June 2003, Amsterdam)
- Bölling [1983], R.: Sozialgeschichte der deutschen Lehrer. Ein Überblick von 1800 bis zur Gegenwart. Göttingen: Vandenhoeck und Ruprecht
- Böttcher [2003], W.: Bildung, Standards, Kerncurricula. Ein Versuch, einige Missverständnisse auszuräumen. In: Die Deutsche Schule (Weinheim) 95 (2), 152-165
- Brouwer, N./ ten Brinke, St. [1995a]: Der Einfluss integrativer Lehrerausbildung auf die Unterrichtskompetenz (I). In: Empirische Pädagogik (Landau) 9 (1), S. 3-31
- Brouwer, N./ ten Brinke, St. [1995b]: Der Einfluss integrativer Lehrerausbildung auf die Unterrichtskompetenz (II). In: Empirische Pädagogik (Landau) 9 (3), S. 289-330
- Brügelmann [2003], H.: Das kurze Gedächtnis der großen Reformer. Anmerkungen zum Beitrag von Wolfgang Böttcher in diesem Heft. In: Die Deutsche Schule (Weinheim) 95 (2), S. 168-171
- Buhlman, E./ Wolff, K./ Klieme, E. [2003]: Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Berlin <http://www.bmbf.de/pub/zur_entwicklung_nationaler_bildungsstandards.pdf> [30.09.2003]
- Bullough [1997], R.V.: Becoming a teacher. Self and the social location of teacher education. In: Biddle, B. J./ Good, T. L./ Goodson, I. F. (Hrsg.): International handbook of teachers and teaching. Dordrecht: Kluwer Academic Publishers, S. 87-148

- Chatterji [2002], M.: Models and Methods for Examining Standards-Based Reforms and Accountability Initiatives. Have the Tools of Inquiry Answered Pressing Questions on Improving Schools? In: *Review of Educational Research* 72 (3), S. 345-386
- Clark [1988], C.: Asking the right questions about teacher preparation. Contributions of research on teacher thinking. In: *Educational Researcher* (Washington), 17 (2), S. 5-12
- Cochran-Smith [2001], M.: The outcomes question in teacher education. In: *Teaching and teacher education* 17, S. 527-546
- Coleman u.a. 1966 = Coleman, J. S./ Campbell, E. Q./ Hobson, C. J./ McPartland, J. M./ Mood, A./ Weinfeld, F. D./ York, R.L.: Equality of educational opportunity. Washington: Government Printing Office
- Coleman, J. S./ Hoffer, T./ Kilgore, S. [1982]: Cognitive outcomes in public and private schools. In: *Sociology of Education* (Washington) 55 (2-3), S. 65-76
- Criblez, L./ Wild-Näf, M. [1998]: Lehrerbildungsforschung in der Schweiz. Forschungsdesiderata zur Stellung der Erziehungswissenschaften in der Lehrerbildung. In: *Zeitschrift für Pädagogik* (Weinheim) 44 (1), S. 21-39
- Criblez [1998], L.: Die Reform der Lehrerbildung in England und Amerika. In: *Zeitschrift für Pädagogik* (Weinheim) 44 (1), S. 41-60
- Dann, H.-D./ Müller-Fohrbrod, G./ Cloetta, B. [1981]: Sozialisation junger Lehrer im Beruf. „Praxischock“ drei Jahre später. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 13 (3), S. 251-262
- Darling-Hammond [2000], L.: Teacher Quality and Student Achievement. A Review of State Policy Evidence. In: *Education Policy Analysis Archives* 8 (1) <<http://epaa.asu.edu/epaa/v8n1/>> [27.05.2003]
- Darling-Hammond [2001], L.: Standard Setting in Teaching. Changes in Licensing, Certification, and Assessment. In: Richardson, V. (Hrsg.): *Handbook of Research on Teaching*. Washington: American Educational Research Association (4. Aufl.), S. 751-776
- Darling-Hammond [2003], L.: Standards and Assessments. Where We Are and What We Need. <<http://www.tcrecord.org/printcontent.asp?contentid=11109>> [10.04.2003]
- Dillon [2003a], S.: Before the Answer, the Question Must Be Correct. In: *The New York Times* (New York) v. 16.07.2003
- Dillon [2003b], S.: Citing Flaw, New York State Voids Math Scores. In: *The New York Times* (New York) v. 24.06.2003
- Ditton [2002], H.: Unterrichtsqualität – Konzeption, methodische Überlegungen und Perspektiven. In: *Unterrichtswissenschaft* (Weinheim) 30 (3), S. 197-212
- Döbrich u.a. 2003 = Döbrich, P./ Klemm, K./ Knauss, G./ Lange, H. [o.J. (2003)]: Ausbildung, Einstellung und Förderung von Lehrerinnen und Lehrern (OECD-Lehrerstudie). Ergänzende Hinweise zu dem Nationalen Hintergrundbericht (CBR) für die Bundesrepublik Deutschland, Bericht für die Kultusministerkonferenz
- Drechsel [2001], B.: Subjektive Lernbegriffe und Interesse am Thema Lernen bei angehenden Lehrpersonen. Münster u.a.: Waxmann (= Internationale Hochschulschriften; 363)
- Fenstermacher [1993], G. D.: Where are we going? Who will lead us there? Washington: American Association of Colleges of Teacher Education
- Floden [2001], R. E.: Research on Effects of Teaching. A Continuing Model for Research on Teaching. In: Richardson, V. (Hrsg.): *Handbook of Research on Teaching*. Washington: American Educational Research Association (4. Aufl.), S. 3-16
- Fried [1998], L.: Zwischen Wissenschaftsorientierung und Orientierung an der Berufspraxis – Bilanz der Lehrerbildungsforschung. In: *Empirische Pädagogik* (Landau) 12 (1), S. 49-90
- Greenhouse [2003], L.: Justices Back Affirmative Action by 5 to 4. In: *The New York Times* (New York) v. 24.06.2003
- Ginsburg, M. B./ Clift, R. T. [1990]: The hidden curriculum of preservice teacher education. In: Houston, W. R./ Haberman, M./ Sikula, J. (Hrsg.): *Handbook of Research on Teacher Education*. New York: Macmillan, S. 450-465
- Helmke, A./ Jäger, R. S. (Hrsg.) [2002]: Das Projekt MARKUS. Mathematik-Gesamterhebung Rheinland-Pfalz. Kompetenzen, Unterrichtsmerkmale, Schulkontext. Landau: VEP (= VEP-Aktuell; 2)
- Herrmann [2003], U.: „Bildungsstandards“ – Erwartungen und Bedingungen, Grenzen und Chancen. In: *Zeitschrift für Pädagogik* (Weinheim) 49 (5), S. 625-639
- Herszenhorn [2003], D. M.: Basic Skills Forcing Cuts in Art Classes. In: *The New York Times* (New York) v. 23.07.2003
- [The] Holmes Partnership [1986]: Tomorrow's Teachers. A Report of the Holmes Group. <<http://www3.baylor.edu/soe/scholmes/tt.html>> [16.09.2003]
- [The] Holmes Partnership [1995]: Tomorrow's Schools of Education. A Report of the Holmes Group. <<http://www3.baylor.edu/soe/scholmes/tsoe.html>> [16.09.2003]
- Kaelble [1978], H.: Probleme der Modernisierung in Deutschland. Sozialhistorische Studien zum 19. und 20. Jahrhundert. Opladen: Westdeutscher Verlag

- Kagan [1992], D. M.: Professional Growth Among Preservice and Beginning Teachers. In: Review of Educational Research 62 (2), S. 129-169
- Kane, R./ Sandretto, S./ Heath, Ch. [2002]: Telling Half the Story. A Critical Review of Research on the Teaching Beliefs and Practices of University Academics. In: Review of Educational Research 72 (2), S. 177-228
- Kennedy [1999], M.: Approximations to indicators of student outcomes. In: Educational Evaluation and Policy Analysis 21 (4), S. 345-363
- KMK 2003a = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [2003]: Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss. Entwurf. Bonn: KMK
- KMK 2003b = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [2003]: Bildungsstandards für die erste Fremdsprache (Englisch/ Französisch) für den Mittleren Schulabschluss. Entwurf. Bonn: KMK
- KMK 2003c = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [2003]: Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Entwurf. Bonn: KMK
- KMK 2003d = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [2003]: Ausbildung, Einstellung und Förderung von Lehrerinnen und Lehrern. Nationaler Hintergrundbericht (CBR) für die Bundesrepublik Deutschland (Ms.)
- Koch [1972], J.-J.: Lehrer-Studium und Beruf. Einstellungswandel in den beiden Phasen der Ausbildung. Ulm: Süddeutsche Verlagsgesellschaft
- LaTurner [2002], R. J.: Teachers' Academic Preparation and Commitment to Teach Math and Science. In: Teaching and Teacher Education 18, S. 653-663
- Lortie [1975], D.: School teacher. A sociological study. Chicago: University of Chicago Press
- Luhmann, N./ Schorr, K.-E. [1976]: Ausbildung für Professionen – Überlegungen zum Curriculum für Lehrerbildung. In: Haller, H.-D./ Lenzen, D. (Hrsg.): Lehrjahre in der Bildungsreform. Resignation oder Rekonstruktion? Stuttgart: Ernst Klett (= Jahrbücher für Erziehungswissenschaft), S. 247-277
- Merzyn [2002], G.: Stimmen zur Lehrerbildung. Überblick über die Diskussion zur Gymnasiallehrerbildung, basierend vor allem auf Stellungnahmen von Wissenschafts- und Bildungsgremien sowie auf Erfahrungen von Referendaren und Lehrern. Hohengehren: Schneider
- Monk, D. H./ King, J. A. [1994]: Multilevel teacher resource effects in pupil performance in secondary mathematics and science. The case of teacher subject matter preparation. In: Ehrenberg, R. G. (Hrsg.): Choices and consequences. Contemporary policy issues in education. Ithaca: ILR Press, S. 29-58
- Munby, H./ Russell, T./ Martin, A. K. [2001]: Teachers' Knowledge and How It Develops. In: Richardson, V. (Hrsg.): Handbook of Research on Teaching. Washington: American Educational Research Association (4. Aufl.), S. 877-904
- Murnane, R. J./ Phillips, B. R. [1981]: What do effective teachers of inner-city children have in common? In: Social Science Research, 10, S. 83-100
- NCATE = National Council for Accreditation of Teacher Education <<http://www.ncate.org>> [30.09.2003]
- NBPTS 2002 = National Board for Professional Teaching Standards [2002]: What Teachers Should Know and Be Able to Do. Arlington/ Southfield: NBPTS <http://www.nbpts.org/standards/know_do/know_do.html> [30.09.2003]
- NCEE 1983 = The National Commission on Excellence in Education: A nation at risk. The imperative for educational reform. <<http://www.ed.gov/pubs/NatAtRisk/title.html>> [16.09.2003]
- OECD 2002a = Organisation für wirtschaftliche Zusammenarbeit und Entwicklung [2002a]: Bildung auf einen Blick. OECD-Indikatoren. Paris: OECD
- OECD 2002b = Organisation for Economic Co-Operation and Development/ Directorate for Education, Employment, Labour and Social Affairs/ Education and Training Division [2002]: Attracting, Developing and Retaining Effective Teachers. Design and Implementation Plan for the Activity. Paris: OECD
- Oelkers [2002a], J.: Die Entwicklung der Lehrerbildung als internationales Problem? Vortrag im Rahmen der Ringvorlesung der Universität Erfurt „Herausforderung der Bildungsgesellschaft“ am 25. Juni 2002 in der Michaeliskirche in Erfurt. <<http://www.paed.unizh.ch/ap/ErfurtLAB.rtf>> [30.09.2003]
- Oelkers [2002b], J.: Befunde und Fragen zur Wirksamkeit der Lehrerbildung. In: Daschner, P./ Drews, U./ Durdel, A. (Hrsg.): Kursbuch Referendariat. Weinheim/ Basel: Beltz 4. Aufl., S. 152-174
- Oser, F./ Oelkers, J. (Hrsg.) [2001]: Die Wirksamkeit der Lehrerbildungssysteme. Von der Allrounderbildung zur Ausbildung professioneller Standards. Chur/ Zürich: Rüegger (= Nationales Forschungsprogramm; 33)
- Pajares [1992], F.: Teachers' beliefs and educational research. Cleaning up a messy construct. In: Review of Educational Research (Washington) 62 (3), S. 307-332
- PLAZ 2003 = Paderborner Lehrerbildungszentrum [2003]: Positionspapier des PLAZ zur Diskussion um Standards, Qualität und Leitbild der Lehrerbildung an der Universität Paderborn. Paderborn: PLAZ (Stand: 11.02.2003)

- Prondczynski [2001], A. v.: Evaluation der Lehrerausbildung in den USA. Geschichte, Methoden, Befunde. In: Keiner, E. (Hrsg.): Evaluation (in) der Erziehungswissenschaft. Weinheim/ Basel: Beltz (= Beiträge zur Theorie und Geschichte der Erziehungswissenschaft; 23), S. 91-140
- Richardson [1996], V.: The role of attitudes and beliefs in learning to teach. In: Sikula, J./ Buttery, T./ Guyton, E. (Hrsg.): Handbook of Research on Teacher Education. New York: Macmillan (2. Aufl.), S. 102-119
- Richardson, V./ Placier, P.: Teacher Change. In: Richardson, V. (Hrsg.): Handbook of Research on Teaching. Washington: American Educational Research Association (4. Aufl.), S. 905-947
- Rolff [2003], H.-G.: Bildungsstandards sind attraktiv – und problematisch. In: Frankfurter Rundschau (Frankfurt) 60 (12.03.2003)
- Rowan, B./ Chiang, F. S./ Miller, R. L. [1997]: Using research on employees' performance to study the effect of teachers on student performance. In: Sociology of Education (Washington) 70 (4). S. 256-284
- Schaefers [2002], Ch.: Forschung zur Lehrerausbildung in Deutschland – eine bilanzierende Übersicht der neueren empirischen Studien. In: Schweizerische Zeitschrift für Bildungswissenschaften 24 (1), S. 65-88
- Schema [2003a], D. J.: Education Secretary Defends School System He Once Led. In: The New York Times (New York) v. 26.07.2003
- Schema [2003b], D. J.: Questions on Data Cloud Luster of Houston Schools. In: The New York Times (New York) v. 11.07.2003
- Schlee [1992], J.: Empirische Forschung zur Lehrerbildung. In: Ingenkamp, Karlheinz/ Jäger, Reinhold S./ Petillon, Hanns/ Wolf, Bernhard (Hrsg.): Empirische Pädagogik 1970-1990. Eine Bestandsaufnahme der Forschung in der Bundesrepublik Deutschland. Weinheim: Deutscher Studien Verlag, S. 558-565
- Tatto [2000], M. T.: Teacher Quality and Development. Empirical Indicators and Methodological Issues in the Comparative Literature. <<http://www7.nationalacademies.org/bicse/Tatto.pdf>> [12.06.2003]
- Tatto u.a. 2003 = Tatto, M./ Schwille, J./ Ingvarson, L./ Beavis, A. [2003]: TEDS – Teacher Education Study. Teacher Education and Development Study – Progress Report to the IEA General Assembly. o.O.: IEA
- Tenorth [2001], H.-E.: Kerncurriculum Oberstufe. Mathematik, Deutsch Englisch. Weinheim/ Basel: Beltz
- Terhart u.a. 1994 = Terhart, E./ Czerwenka, K./ Ehrich, K./ Jordan, F./ Schmidt, H.J. [1994]: Berufsbiographien von Lehrern und Lehrerinnen. Frankfurt/M.: Lang
- Terhart [1999], E.: Strukturprobleme der Lehrerinnen- und Lehrerbildung. Konfligierende Modernisierungen. In: Beiträge zur Lehrerbildung (Großhöchstetten) 17 (2), S. 141-150
- Terhart [2002], E.: Standards für die Lehrerbildung. Eine Expertise für die Kultusministerkonferenz. Münster: Westfälische Wilhelms-Universität/ Institut für Schulpädagogik und Allgemeine Didaktik
- Ulich [2000], K.: Lehrer/innen-Ausbildung im Urteil der Betroffenen. Ergebnisse und Folgerungen. In: Die Deutsche Schule (Weinheim) 88 (1), S. 81-97
- U.S. Department of Education/ Office of Postsecondary Education/ Office of Policy Planning and Innovation: Meeting the Highly Qualified Teachers Challenge. The Secretary's Annual Report on Teacher Quality. Washington, D.C.: ED Pubs 2002 <<http://ed.gov/offices/ope/news/teacherprep/index.html>> [...]
- Walsh [2001], K.: Teacher Certification Reconsidered. Stumbling for Quality – A Rejoinder. Baltimore: Abell Foundation <http://www.abell.org/pubsitems/ed_cert_rejoinder_1101.pdf> [30.09.2003]
- Weinert, F. E./ Helmke, A. [1997]: Entwicklung im Grundschulalter. Weinheim: Psychologie Verlags Union
- Wideen, M./ Mayer-Smith, J./ Moon, B. [1998]: A Critical Analysis of the Research on Learning to Teach. Making the Case for an Ecological Perspective on Inquiry. In: Review of Educational Research 68 (2), S. 130-178
- Wilson, S. M./ Floden, R. E./ Ferrini-Mundy, J. [2001]: Teacher Preparation Research. Current Knowledge, Gaps, and Recommendations. Washington: Center for the Study of Teaching and Policy <<http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf>> [30.09.2003]
- Wright, S. P./ Horn, S. P./ Sanders, W. L. [1997]: Teacher and classroom context effects on student achievement. Implications for teacher evaluation. In: Journal of Personnel Evaluation in Education (Boston) 11, S. 57-67
- Zeichner [1986], K. M.: Lehrersozialisation und Lehrerausbildung. Forschungsstand und Perspektiven. In: Bildung und Erziehung (Köln) 39 (3), S. 263-278
- Zeichner, K./ Tabachnik, B. R. [1981]: Are the Effects of University Teacher Education Washed Out by School Experience? In: Journal of Teacher Education (Washington), 32, S. 7-11