

Sigrid Blömeke
Lehrerbildung – Lehrerhandeln – Schülerleistungen
Perspektiven nationaler und internationaler empirischer Bildungsforschung
– Antrittsvorlesung an der Humboldt-Universität zu Berlin am 10.12.2003¹ –

Lehrerbildung war und ist immer wieder starker Kritik ausgesetzt. Sie sei praxisfern, in ihrer ersten Phase beliebig, in ihrer zweiten Phase mehr eine Meisterlehre als eine theoriegeleitete Ausbildung, sie dauere zu lange, sei nicht koordiniert und kenne keine geteilten Standards. Empirisch geprüft wurde ihre Wirksamkeit allerdings nie. Erst eine ‚externe‘ Entwicklung führte dazu, dass die Bereitschaft hierzu momentan wächst: Die Ergebnisse der internationalen Schulleistungsvergleiche legten Defizite offen, die den Blick u.a. auf die Lehrerbildung als deren mögliche Ursache richten. Durch eine empirische Prüfung ihrer Qualität könnte entweder eine Ursache der schulischen Defizite ausgemacht oder die Lehrerbildung als Ursache weitgehend ausgeschlossen werden.

So nahe liegend die Forderung nach einer Wirksamkeitsprüfung der Lehrerbildung ist, so schwierig ist deren konkrete Durchführung. Eine Übertragung der in der empirischen Bildungsforschung vorhandenen Modelle und Verfahren ist theoretisch problematisch und methodisch unzureichend. Pilotstudien existieren nur in einem sehr eingegengten Verständnis von Wirksamkeitsprüfung. Ich möchte Ihnen im Folgenden Chancen, aber auch Probleme und Grenzen nationaler und internationaler empirischer Bildungsforschung in diesem Zusammenhang vorstellen.

Ich beginne im ersten Teil meines Vortrags mit Grundsatzfragen. Anschließend gehe ich im zweiten Teil auf den derzeitigen Forschungsstand ein, bevor ich im dritten Teil Aufgaben für die Zukunft herausarbeite. Abschließend reflektiere ich diese in ihren Chancen und Grenzen.

1 Grundsatzfragen empirischer Forschung zur Lehrerbildung

Die empirische Forschung steht in Bezug auf die Lehrerbildung vor der zentralen Herausforderung zu klären, was überhaupt untersucht werden soll. Ulrich Herrmann führte hierzu kürzlich aus: „Institutionen und Personen, Faktoren und Prozesse, Ziele, Erwartungen und Interessen, Realbedingungen und Beteiligungsvoraussetzungen bilden bei Bildungsprozessen fast unauflösbare Ursachen-, Bedingungs- und Wirkungszusammenhänge.“ Dementsprechend umstritten ist die Variablenwahl. Sie hängt besonders von der zugrunde liegenden Modellierung des Zusammenhangs von Lehrerbildung, Lehrerhandeln und Schülerleistungen ab.

Auf der Seite der Einflussfaktoren, also bei den Variablen, von denen man Wirkungen annimmt, stellen sich in Bezug auf diesen Zusammenhang im Kern zwei Fragen.

Die erste Frage umfasst, wodurch das unterrichtliche Handeln von Lehrerinnen und Lehrern beeinflusst wird: durch ihre kognitiven Strukturen, durch Persönlichkeitsmerkmale, durch Übung oder z.B. durch schulische Rahmenbedingungen? Wer v.a. Persönlichkeitsmerkmale für ausschlaggebend hält, wird ‚Talent‘ für wichtiger halten als eine gründliche Ausbildung; wer v.a. Übung für ausschlaggebend hält, wird Praxiserfahrung eine hohe Bedeutung zuschreiben.

Wie immer man den Zusammenhang von Lehrerhandeln und Lehrerbildung auch modelliert: Sobald man von einer Bedeutung der Lehrerbildung ausgeht, stellt sich die zweite Kernfrage,

¹ Es gilt das gesprochene Wort. Die bibliographischen Angaben der zugrunde liegende Literatur können bei der Verfasserin angefordert werden.

welchen Einflussfaktoren konkret Wirksamkeit zugeschrieben wird: nur den Ausbildungsinhalten oder auch der didaktisch-methodischen Ausrichtung der Ausbildung, dem Grad an Verbindlichkeit, ggf. auch der Abfolge der Ausbildungsphasen? Dieser Fragenkatalog ließe sich fast beliebig erweitern.

Vergleichbar mit der Seite der Einflussfaktoren stellt sich auf der Seite der Wirkungen von Lehrerbildung eine ähnliche Frage: Welches Kriterium kann als angemessen angesehen werden, um ‚Wirksamkeit‘ zu bestimmen? Bereits auf den ersten Blick lassen sich diesbezüglich drei Herangehensweisen ausmachen: zum ersten eine Erfassung der fach- und berufsbezogenen Kognitionen der angehenden Lehrpersonen (dahinter steht die Frage: Was wissen sie?); zum zweiten eine Beobachtung ihres Handelns im Unterricht (also: Was können sie?) und zum dritten eine Erfassung der Leistungen ihrer Schülerinnen und Schüler (ist doch letztlich Ziel von Schule und Lehrerbildung das Lernen und die Entwicklung von Kindern und Jugendlichen). Nimmt man die professionstheoretische Diskussion ernst, muss mindestens noch eine vierte Herangehensweise in Betracht gezogen werden: die des beruflichen Ethos‘ der Lehrerinnen und Lehrer.

Leider ist die empirische Forschung, die zur Beantwortung dieser Fragen hilfreich sein könnte, äußerst defizitär. Zwar kann man konstatieren, dass im Zuge von PISA 2003 und im DFG-Schwerpunktprogramm „Bildungsqualität von Schule“ das unterrichtliche Handeln der Lehrpersonen immer stärker in den Blick gerät. Doch zur Lehrerbildung muss weiterhin ein Forschungsdefizit festgestellt werden. Wir verfügen über zahllose Veröffentlichungen mit normativ-konzeptioneller Orientierung, da will ich meine eigenen Arbeiten gar nicht ausschließen. Diesen stehen aber nur punktuelle empirische Studien gegenüber. Schlee hat 1992 formuliert, „dass zu den konzeptionellen Kernfragen und Brennpunkten der Lehrerbildungsreform keine systematische, theoriebezogene und experimentelle Forschung durchgeführt wurde“. Zehn Jahre später kommen die Reviewer weiterhin zum selben Ergebnis. Larcher und Oelkers gehen sogar so weit zu sagen, „wenn es eine Krise in der Lehrerbildung gibt, dann ist es wesentlich eine Krise der fehlenden Daten“.

Mit der Gewinnung von Daten ist die Frage der Wirksamkeit von Lehrerbildung allerdings noch nicht beantwortet. Dies machen die widersprüchlichen Interpretationen der wenigen vorliegenden empirischen Befunde deutlich. Welche Interpretationsperspektive eingenommen wird, hängt von der bildungspolitischen Position und methodischen Kriterien ab. Für die anglo-amerikanische Forschung lässt sich hier eine interessante Polarisierung feststellen. Seit einigen Jahren wird in den USA eine erbitterte Auseinandersetzung darüber geführt, welche Bedeutung gesetzlichen Regelungen zukommen soll, mit denen Mindestanforderungen an die Ausbildung von Lehrerinnen und Lehrern festgelegt werden. Auf der einen Seite stehen die so genannten ‚Deregulierer‘, die sich für einen weitgehenden Verzicht auf Vorschriften aussprechen. Auf der anderen Seite stehen die so genannten ‚Professionalisierer‘, die eine striktere Regelung der Lehrerbildung empfehlen.

Ich komme zunächst zu den ‚Deregulierern‘, vorrangig vertreten durch die Abell Foundation. Diese vertritt die Ansicht, dass letztlich der Markt das Problem regeln solle, indem die Auslese der besten Lehrerinnen und Lehrer – ob mit oder ohne Ausbildung – über Angebot und Nachfrage gesteuert werde. Die Abell Foundation stützt sich auf eines der umfassendsten Reviews englischsprachiger Literatur zur Lehrerbildung, das es derzeit gibt. Seine Erstellung zeichnet sich durch ein besonders strenges methodisches Vorgehen aus. Aufgenommen wurden nur Arbeiten, die drei Voraussetzungen entsprechen:

Erstens: Es muss sich um so genannte large scale-Studien handeln, also um Studien mit großen Fallzahlen. Damit finden beispielsweise die Ergebnisse qualitativer Studien keine Berücksichtigung. Die zweite Voraussetzung ist, dass die Wirksamkeit der Lehrerbildung anhand von Schülerleistungen gemessen wird. Damit fallen solche Studien heraus, die sich z.B. auf Befragungen von Schulleitern stützen. Um in das Review der Abell Foundation aufgenommen zu werden, müssen die Ergebnisse einer Studie drittens in Zeitschriften mit einem externen Gutachterverfahren erschienen sein. Mit Blick auf diese drei Kriterien schließt die Abell Foundation selbst Aussagen aus blindbegutachteten Studien aus, wenn in diesen Doktorarbeiten oder qualitative Fallstudien zitiert werden.

Prominenteste Vertreterin der ‚Professionalisierer‘ ist Darling-Hammond. Von ihr stammen ebenfalls umfassende Reviews der empirischen Forschungsarbeiten zur Lehrerbildung – mit dem Unterschied, dass sie auch Ergebnisse von Doktorarbeiten akzeptiert. Darüber hinaus akzeptiert Darling-Hammond Beiträge in Zeitschriften mit einem redaktionsinternen Gutachterverfahren. Haben qualitative Studien diese Hürde genommen, fließen deren Ergebnisse in ihre Reviews ein.

Ich möchte an dieser Stelle anmerken, dass die Methoden, mit denen Wirkungen der Lehrerbildung bisher untersucht wurden, generell kritisch diskutiert werden müssen: Häufig handelt es sich nur um Re-Analysen von Erhebungen, die andere Fragestellungen verfolgten. Eine konsistente Bearbeitung von lehrbildungsbezogenen Fragen ist daher schwierig. Experimentelle Arbeiten und Langzeitstudien finden sich so gut wie gar nicht. Noch immer ist darüber hinaus eine erhebliche Forschungslücke festzustellen, was die fachliche Ausbildung angeht; die überwiegende Zahl der Studien fokussiert auf Einstellungen z.B. zur Rolle der Lehrperson im Unterricht. Und schließlich ist ein eklatantes Defizit an gegenseitigen Bezugnahmen zu konstatieren. Die Mehrheit der Forscher macht keinen Versuch, die eigenen Studien mit denen anderer zu verbinden.

Wenn ich im folgenden zweiten Teil meines Vortrags auf zentrale Befunde v.a. der US-Forschung zur Wirksamkeit der Lehrerbildung eingehe, müssen diese Qualitätsmängel bedacht werden. Ich werde meine Darstellung auf so genannte ‚Programmevaluationen‘ fokussieren. In diesen wird die Lehrerbildung als Ganze untersucht oder es werden substanzielle Teile von ihr erfasst. Gegenstand ist also nicht die individuelle Lehrperson und ihr Kompetenzgewinn im Prozess der Lehrerbildung. Vielmehr werden in Meta-Analysen formale Merkmale der Lehrerschaft – z.B. Studienabschlüsse – in Beziehung zu Lehrerhandeln oder Schülerleistungen gesetzt. Um die Effekte der Lehrerbildung abschätzen zu können, werden alle übrigen Merkmale (z.B. die Dauer der beruflichen Erfahrung) konstant gehalten. In den USA stellt diese Art von Studien das Kernverständnis einer Evaluation der Lehrerbildung dar. Für Europa bzw. Deutschland liegen leider keine entsprechenden Studien vor.

2 Kerneergebnisse der empirischen Forschung zur Lehrerbildung

Zunächst zur Frage, was sich generell in Bezug auf den Abschluss eines lehrerbildenden Studiengangs im Vergleich zu Notausbildungen bzw. gänzlich fehlender Lehrerbildung feststellen lässt. Unter das Merkmal ‚lehrerbildender Studiengang‘ fallen dabei sowohl Studien in Erziehungswissenschaft als auch in der Fachwissenschaft, in der unterrichtet wird. Obwohl von diesem Merkmal vermutlich am ehesten signifikante Wirkungen auf Schülerleistungen bzw. Lehrerhandeln erwartet werden könnten, sind die Ergebnisse der Studien aus den letzten 50 Jahre nicht durchgängig positiv. Auf der Ebene eines Bachelor-Abschlusses lässt sich zwar überwiegend folgern, dass Lehrpersonen mit einem lehrerbildenden Abschluss höhere Schülerleistungen bzw. positivere Einschätzungen ihres Handelns erreichen als Lehrpersonen ohne einen solchen

Abschluss. Einer Minderheit an Studien zufolge ist ein lehrerbildender Abschluss allerdings bedeutungslos. Die Art des Bachelors – ob in einer Fachwissenschaft oder in Erziehungswissenschaft – ist im Übrigen weder für die eine noch für die andere Richtung bedeutsam.

Dieses widersprüchliche Ergebnis spiegelt sich auf der Ebene der Masterabschlüsse wider. Den meisten Studien zufolge erreichen Lehrpersonen mit einem Master höhere Schülerleistungen oder ihr Handeln wird besser beurteilt als das von Lehrpersonen, die nur einen Bachelor besitzen. Es liegen aber auch Studien vor, die einen solchen Unterschied nicht feststellen können.

Soweit zur Lehrerbildung als Ganze. In Bezug auf die Wirksamkeit einzelner ihrer Komponenten ist zunächst die Frage nach der fachwissenschaftlichen Ausbildung interessant. Diese wird in Form von Tests des Lehrwissens, durch Zählungen belegter Fachkurse oder in Form von akademischen Abschlüssen operationalisiert. Auch hier kann kein durchgängig positiver Zusammenhang zwischen Fachwissen und Schülerleistungen bzw. Lehrerhandeln gefunden werden. Zwar spricht erneut eine Mehrheit der Studien für die Wirksamkeit einer umfassenden fachwissenschaftlichen Ausbildung, gleichzeitig liegen allerdings Studien vor, die nicht-signifikante oder sogar negative Zusammenhänge feststellen.

Um diese Befunde angemessen interpretieren zu können, kann die Langzeitstudie von Monk herangezogen werden. In dieser wird von einem ‚Deckeneffekt‘ ausgegangen: Um hohe Schülerleistungen zu erreichen, ist ein Mindestmaß an Fachwissen notwendig. Über eine gewisse Schwelle hinaus ist jedoch kein zusätzlicher positiver Effekt mehr festzustellen. Die mathematikspezifische Lehrerbildungsforschung analysiert dieses Phänomen noch ein wenig tiefer: Die aus dem höheren fachwissenschaftlichen Niveau folgende Art der Strukturierung von Inhalten macht zum einen eine Elementarisierung schwieriger. Darüber hinaus ist die Vertiefung fachwissenschaftlicher Inhalte in der Regel mit einer stärkeren Orientierung an konventionellen Lehrmethoden verbunden, was schülerorientiertem Lehren entgegensteht.

Die Frage nach der Wirksamkeit lässt sich auch in Bezug auf die fachdidaktische Ausbildung stellen. Dieser Bestandteil der Lehrerbildung spielt allerdings im gesamten anglo-amerikanischen Raum eine deutlich geringere Rolle als in Deutschland. Den wenigen fachdidaktisch ausgerichteten Studien lässt sich – im Unterschied zu den bisher angesprochenen Merkmalen – ein konsistent positiver Zusammenhang mit Schülerleistungen bzw. Lehrerhandeln entnehmen. Die Bedeutung dieser Erkenntnis gleich wieder einschränkend ist allerdings darauf hinzuweisen, dass das Verständnis von ‚Fachdidaktik‘ in den USA vergleichsweise eng ist. Vor dem Hintergrund eines pragmatischen Konzepts von *literacy* werden vor allem fachspezifische Unterrichtsmethoden gelehrt.

In Bezug auf die erziehungswissenschaftliche Ausbildung sind die Ergebnisse erneut widersprüchlich. Einzelne Studien sprechen gegen ihre Wirksamkeit, in der Mehrheit sind aber positive Einflüsse auf Schülerleistungen bzw. Lehrerhandeln festgestellt worden. Zwei Aspekte sind in diesem Zusammenhang herauszustellen: So sind die positiven Effekte vermutlich nur bei einem Zusammenspiel von umfangreicher erziehungswissenschaftlicher und fachlicher Ausbildung zu verzeichnen. Darüber hinaus ist das Verständnis von Erziehungswissenschaft – wie im Fall der Fachdidaktik – in der anglo-amerikanischen Lehrerbildung deutlich enger als in Mitteleuropa. Es gibt in der Regel nur Angebote, die unmittelbar auf den Lernprozess der Schülerinnen und Schüler bzw. auf das methodische Handeln der Lehrpersonen bezogen sind. Selbst Bereiche wie Geschichte und Theorie der Schule oder Allgemeine Didaktik, v.a. aber fast sämtliche Angebote,

die bei uns traditionell unter die Allgemeine Pädagogik fallen, sind für die Lehrerbildung in den USA so gut wie unbedeutend.

Neben der Wirkung dieser Komponenten von Lehrerbildung wurde das Vorhandensein berufsunspezifischer Fähigkeiten der Lehrpersonen untersucht. In Bezug auf ihr intellektuelles Niveau müssen widersprüchliche Zusammenhänge zu Schülerleistungen bzw. Lehrerhandeln festgehalten werden. Die verbalen Fähigkeiten einer Lehrperson sind dagegen offensichtlich signifikant positiv mit diesen assoziiert.

Darling-Hammond zieht angesichts der Tendenz in der Mehrheit der Studien die Schlussfolgerung [Zitat in deutscher Übersetzung]: „Umfang und Abschluss einer Lehrerbildung sind bisher die Maße mit den höchsten Korrelationen zu Schülerleistungen.“ Sie spricht sich daher für ein stärkeres staatliches Engagement aus, was Regelungen von Mindestanforderungen an die Lehrerbildung angeht. Inhaltlich geht es ihr in erster Linie um eine Kombination von Fachwissenschaft, Erziehungswissenschaft und schulpraktischen Ausbildungsanteilen. Der Abschluss soll ein Master sein (bisher ist in den USA ein Bachelor die Regel).

Im Review der Abell Foundation wird ein Großteil der Studien, die positive Wirkungen der Lehrerbildung belegen, nicht akzeptiert, da sie die zuvor festgelegten Kriterien verfehlen. Die Stiftung kommt daher zu einer Schlussfolgerung, die der von Darling-Hammond genau entgegen gesetzt ist. In ihrem Hauptbericht lautet die Kernaussage, dass sich keine hinreichenden Belege dafür finden ließen, Personen mit abgeschlossener Lehrerbildung erzielten höhere Schülerleistungen als Personen ohne Ausbildung. Wenn überhaupt ein stabiler Prädiktor gefunden werden könne, dann nur außerhalb der Lehrerbildung, bei berufsunspezifischen Fähigkeiten. Die Folgerungen der Abell Foundation aus diesen Erkenntnissen lauten: Abschaffung aller Regularien im Hinblick auf lehrerbildungsbezogene Qualifikationen von Lehrpersonen; stattdessen lediglich die Forderung eines Bachelors (und zwar ohne nähere Spezifizierung der Fachrichtung, wie es in den USA in Bezug auf Privatschulen gilt).

Diese Formalie soll mit einem Einstellungstest kombiniert werden, der verbale Fähigkeiten misst sowie – bei Elementarschullehrern – ihre Fähigkeit, Lesen zu unterrichten, bzw. – bei Sekundarschullehrern – ihr Fachwissen. Wenn es überhaupt lehrerbildende Bemühungen geben müsse, so die Abell Foundation, dann sollten diese weniger auf erziehungswissenschaftliche Qualifikationen zukünftiger Lehrerinnen und Lehrer als auf die Förderung ihrer allgemeinen Sprach- und Lesefertigkeiten gerichtet sein. Insgesamt soll – zu Lasten der Grundausbildung – stärker die Berufseingangsphase unterstützt werden.

Ich habe eingangs bereits darauf verwiesen, dass diese Interpretation in hohem Maße daraus resultiert, dass empirische Befunde keine Berücksichtigung finden, wenn sie nicht die vorab definierten Kriterien erfüllen. Letztlich lässt sich aber die Frage stellen, inwieweit eine solche Selektion berechtigt ist. In meinen Augen lässt sie die theoretischen und methodischen Probleme empirischer Forschung zur Lehrerbildung weitgehend unberücksichtigt. Ich gehe darauf gleich noch näher ein.

Zuvor möchte ich ein erstes Zwischenfazit ziehen. Die oben aufgeworfene Ausgangsfrage, ob eine hohe Qualität von Lehrerhandeln durch Lehrerbildung erreicht wird oder ob die Qualität eher Resultat von Persönlichkeitseigenschaften („Talent“) bzw. von langjähriger Berufspraxis („Übung“) ist – diese Frage kann auf der Basis der vorhandenen Befunde nicht eindeutig beantwortet werden. Aus den Studien lässt sich auch nicht ableiten, wie Lehrerbildung strukturell und inhaltlich optimalerweise auszusehen hat. Wie viel Fachwissenschaft, Fachdidaktik und

Erziehungswissenschaft werden denn mindestens benötigt? Sind Unterschiede nach Schulformen zu machen? Und wie sollen die jeweiligen Inhalte konkret aussehen? Eine Operationalisierung allein über Abschlüsse und Kurszählungen geht möglicherweise am Kern der Problematik von Lehrerbildung vorbei. Dass sich diese Fragen bisher auf empirischer Basis nicht beantworten lassen, gilt allerdings sowohl für Verfechter der bestehenden Lehrerbildung als auch für ihre Gegner.

Darüber hinaus werden in den dargestellten Reviews normative Unterschiede zur Frage deutlich, was als Kriterium hoher Qualität von Lehrerbildung akzeptiert wird. Das Kriterium der Schülerleistungen stellt vermutlich einen weitreichenden Konsens dar, weil es wohl in aller Augen als wesentliches Ziel von Erziehung und Bildung akzeptiert wird. Allerdings endet die Einigkeit vermutlich bei genau dieser oberflächlichen Beschreibung. Damit meine ich nicht nur die Definition des Begriffs ‚Schülerleistung‘, die kontrovers ausfallen kann. Schon das Rekurren auf dieses Kriterium muss problematisiert werden. Schülerleistungen zum alleinigen Maßstab für die Qualität von Lehrerbildung zu nehmen, wie es die Abell Foundation tut, unterstellt eine Linearität pädagogischer Prozesse, die nicht vorhanden ist. Ein Ingenieur kann an der Beschaffenheit seines Werkes gemessen werden. Pädagogen haben es aber mit nicht-trivialen, autonomen Subjekten zu tun. Sie sind mit dem Paradox konfrontiert, etwas von außen zu wollen, was nur von den Subjekten selbst hervorgebracht werden kann. Schülerleistungen können daher immer nur ein Maßstab unter mehreren sein.

Ich schließe damit die Darlegung des aktuellen Forschungsstands ab und gehe im dritten Teil meines Vortrags auf zukünftige Aufgaben ein. Ich möchte darlegen, wie vor dem Hintergrund des bisher Ausgeführten angemessene Forschung in Bezug auf die Lehrerbildung aussehen könnte – und zwar insbesondere eine empirische Forschung, die mehr sein will als eine auf das Zählen von Formalia ausgerichtete Programmevaluation. Geprägt von einer intensiven fächerübergreifenden Zusammenarbeit mit Kolleginnen und Kollegen aus dem In- und Ausland lassen sich aus meiner Sicht die anstehenden Aufgaben in folgenden programmatischen Stichworten zusammenfassen: international vergleichende Untersuchung der Lehrerbildung auf der Basis einer Verknüpfung von empirischer Bildungsforschung, Bildungstheorie und Bildungsgeschichte. Ich lege zunächst Aktivitäten dar, die in meiner Abteilung derzeit anlaufen, bevor ich darauf eingehe, welche Chancen und Grenzen ich sehe.

3 Aufgaben empirischer Forschung zur Lehrerbildung

Den Ruf an die Humboldt Universität habe ich vor allem deswegen angenommen, weil ich hier gute Chancen sehe, die in der empirischen Lehrerbildungsforschung bestehenden Forschungsdefizite anzugehen. Während meiner Vorarbeiten hierzu traf es sich daher glücklich, dass mein neuer Kollege Rainer Lehmann einen empirisch erfahrenen Experten für Lehrerbildungsfragen suchte, der sich an der Vorbereitung einer internationalen Vergleichsstudie beteiligt. Trägerin dieser Studie wird die IEA sein, die *International Association for the Evaluation of Educational Achievement*, die auch die TIMS-Studien durchführt. Was ich Ihnen im Folgenden vorstelle, ist z.T. bereits in diese seit nunmehr sechs Monaten bestehende länderübergreifende Zusammenarbeit von Wissenschaftlern aller fünf Kontinente eingeflossen. Die Studie selbst wird in den Jahren 2005 bis 2010 in voraussichtlich mehr als 30 Ländern durchgeführt werden.

Bei der Konzeptionierung einer solchen Studie steht man immer im Zwiespalt zwischen Notwendigkeiten aus theoretischer Sicht und Machbarkeit. Ein hilfreicher Weg, um diesen Spagat zu bewältigen, ist die Modellbildung – also zu versuchen, wichtige Variablen systematisch

abzubilden und ihre Relationen darzustellen. Dabei ist die erste und wichtigste Entscheidung, in welcher Systematik vorgegangen wird. Die empirische Bildungsforschung kann auf zwei Traditionen zurückblicken, in denen bedeutsame Zusammenhänge herausgearbeitet wurden: die Wechselwirkung zwischen dem Lehr-Lernprozess und den Lernergebnissen als Produkt sowie die Wechselwirkung zwischen Lernvoraussetzungen und Lehr-Lernprozess. Dem daraus entwickelten Standard, die drei Elemente Voraussetzungen, Prozess und Produkt aufzugreifen, wird in der geplanten Lehrerbildungsstudie gefolgt.

Zunächst werden die Lernvoraussetzungen derjenigen genauer untersucht, die mit ihrer Ausbildung zum Lehrer bzw. zur Lehrerin beginnen. Der Prozess der Lehrerbildung ist die zweite Differenzierung. Dabei mache ich besonders darauf aufmerksam, dass dieser Prozess nicht mit dem Tag endet, an dem die Grundausbildung abgeschlossen wird, sondern erst nach fünf Jahren Berufserfahrung. Dahinter steht die Absicht, konkurrierenden Professionstheorien gerecht zu werden. In einigen Ländern lassen sich Traditionslinien feststellen, denen zufolge ein Lehrer seinen Beruf erst in dem Moment wirklich lernt, in dem er anfängt zu unterrichten. Die dritte Differenzierung ist dann das Produkt der Lehrerbildung, ihre Wirkung auf die zukünftigen Lehrerinnen und Lehrer.

Bisher war nur von einem und dazu noch vergleichsweise traditionellen Zugang zur Untersuchung der Lehrerbildung die Rede. Auf der Ebene des Individuums verlaufsorientiert vorzugehen ist typisch für die meisten empirischen Erhebungen. Im Hinblick auf die Untersuchung der Lehrerbildung würde man auf diese Weise aber nur einen Bruchteil der Realität abbilden. Historische Entwicklungen, kulturelle Traditionen und ‚Philosophien‘ der Institutionen in der Lehrerbildung blieben außen vor. Für die geplante internationale Vergleichsstudie haben wir uns daher dafür entschieden, bei der Systematisierung eine zweite Achse zu Hilfe zu nehmen und eine Mehrebenenanalyse durchzuführen. Neben dem Verlauf der Lehrerbildung sollen auch die unterschiedlichen gesellschaftlichen Ebenen, die an ihrer Gestaltung beteiligt sind, in den Blick kommen, insbesondere institutionelle und systemische Einflüsse. Die daraus entstehende Systematik können Sie sich wie eine 3 x 3-Matrix bzw. wie eine 9-Felder-Tafel vorstellen. Im anglo-amerikanischen Raum werden bereits seit einigen Jahren Erfahrungen mit solchen Mehrebenenanalysen gesammelt. In Deutschland sind solche Studien derzeit noch selten.

[Power Point Präsentation] Sie sehen hier den derzeitigen Stand des Modells der IEA-Lehrerbildungsstudie. Ich möchte Ihnen zunächst ihre Systematik verdeutlichen. In den Spalten finden Sie den Verlauf der Lehrerbildung wieder – beginnend mit den Voraussetzungen, die je nach Ebene als *input* bzw. als intendiertes Curriculum bezeichnet werden können. Der Prozess der Lehrerbildung kann als implementiertes Curriculum angesehen werden, während es sich bei dem Produkt um das erreichte Curriculum handelt. Die unterschiedlichen gesellschaftlichen Ebenen sind horizontal abgebildet. In der oberen Zeile finden sich jeweils die systemischen Vorgaben. In der mittleren Zeile sind die institutionellen Merkmale aufgenommen. Und unten ist der traditionelle individuelle Zugang erfasst. Die Modellbildung ist für jede einzelne Variable mit einer Suche nach angemessenen Indikatoren, also messbaren Eigenschaften, und der Entwicklung von Erhebungsinstrumenten verbunden.

Betrachtet man zunächst die Voraussetzungen, die Einfluss nehmen auf die Wirksamkeit der Lehrerbildung, sind auf systemischer Ebene z.B. die folgenden Einflussvariablen zu berücksichtigen: die nationalen und regionalen Lehrerbildungstraditionen sowie die konkrete Steuerung des Lehrerbildungssystems durch Politik und Verwaltung. In Deutschland würden hier Elemente wie die starke Rolle des Staates und der Föderalismus von Bedeutung sein. Sie nehmen Einfluss darauf, wie Lehrerbildung institutionalisiert ist und unter welchen Bedingungen eine

Ausübung des Lehrerberufs erfolgt. Sie nehmen auch Einfluss darauf, aus welchem Personenkreis sich die zukünftigen Lehrerinnen und Lehrer v.a. rekrutieren.

Auf der institutionellen Ebene sind z.B. die Festlegung der formalen Zugangsvoraussetzungen zur Lehrerbildung zu nennen, weiterhin Trends auf dem Arbeitsmarkt, der Status von Lehrpersonen und ihr gesellschaftliches Ansehen. Bei der Einordnung dieser Fragen ist im Übrigen schon jetzt eine interessante Konfliktlinie zu erkennen: Wer stellt die Vergleichsgruppe dar, vor der Status, Ansehen und Berufschancen von Lehrerinnen und Lehrern beurteilt werden, alle Arbeitnehmerinnen und Arbeitnehmer oder nur die Gruppe der Akademiker? Je nach Herkunftsland der Diskutanten – z.B. ob Industrieland oder Entwicklungsland – wird diese Frage mit überzeugenden Gründen unterschiedlich beantwortet.

Auf der individuellen Ebene der Studienanfänger ist die Erhebung eines Bündels an Lernvoraussetzungen notwendig: zunächst die Kognitionen zu berufsrelevanten Gegenständen und Situationen. Weiterhin affektiv-motivationale Komponenten wie z.B. Einstellungen zu Kindern und Jugendlichen oder zum fachlichen Gegenstand; aber auch berufsunspezifische Merkmale wie das intellektuelle Niveau und verbale Fähigkeiten.

Während diese Untersuchungsschritte auch länderübergreifend noch vergleichsweise einfach zu realisieren sind, nimmt die Komplexität zu, sobald man den Prozess der Lehrerbildung betrachtet. Als systemische Einflussgrößen können beispielhaft die Folgenden festgehalten werden: die Institutionalisierung der Lehrerbildung mit der deutschen Besonderheit, dass in zwei Phasen – an der Universität und im Studienseminar – lehramtsspezifisch ausgebildet wird. Weitere systemische Vorgaben sind z.B. Rahmenrichtwerte für Ausbildungsinhalte und Praxisanteile, Regelungen der Examina sowie Einstellungsverfahren in den Schuldienst.

In Bezug auf den Prozess der Lehrerbildung kommt aber sehr viel stärker die einzelne Institution in den Blick sowie die individuelle Wahrnehmung des Lehrangebots durch die zukünftigen Lehrpersonen. Was machen die einzelnen Universitäten und Studienseminare aus den staatlichen Vorgaben? Welche Inhalte werden konkret gelehrt? Wie sehen die Praxiserfahrungen der Auszubildenden aus? Das sind keine trivialen Fragen, da die Unterschiede in den Darbietungen groß sein können. Ich möchte ein Beispiel nennen, über das wir hier an der Humboldt-Universität im Rahmen der Modularisierung diskutiert haben: Sollen jeweils die Lehrveranstaltungen innerhalb der Fächer, innerhalb der Fachdidaktiken und innerhalb der Erziehungswissenschaft Module bilden oder trauen wir uns eine querstrukturelle Modularisierung zu, indem eine fachwissenschaftliche Lehrveranstaltung mit einer darauf ausgerichteten fachdidaktischen und einer erziehungswissenschaftlichen Lehrveranstaltung gekoppelt wird? Trotz eines quantitativ gleichen Lehrangebots würden die beiden Richtungen mit einem qualitativen Unterschied einhergehen. Um Wirkungen von Lehrerbildung angemessen beurteilen zu können, ist es wichtig, diese qualitativen Unterschiede korrekt zu erfassen.

In Bezug auf den Prozess der Lehrerbildung gilt es zudem, eine gänzlich andere Perspektive in den Blick zu nehmen, die in einzelnen Professionstheorien als das ‚versteckte‘ Curriculum bezeichnet wird. Es gibt Hinweise, dass es möglicherweise nicht so sehr ausschlaggebend ist, welche Inhalte dargeboten werden, sondern in welcher Form dies geschieht. Ob z.B. vorwiegend strukturiert anhand der Systematik der zugrunde liegenden akademischen Disziplin oder fallbasiert, ob z.B. vorwiegend in Form von Präsentationen oder über aktive Aneignung durch die Studierenden. Um diese wichtige Frage klären zu können, müssen entsprechende Indikatoren entwickelt werden.

Damit möchte ich zum letzten Schritt der anstehenden Forschungsaufgaben übergehen, zur Erfassung der Wirksamkeit der Lehrerbildung, der Produktperspektive also.

Auf der Ebene der individuellen Lehrperson kann wie auf der Seite der Voraussetzungen zwischen Kognitionen, Einstellungen und Handeln differenziert werden. Das nimmt unterschiedliche professionstheoretische Überlegungen zur Differenz von Wissen und Handeln sowie die Bedeutsamkeit des professionellen Ethos auf. Die Problematik liegt hier darin, wie man diese Komponenten erheben will. Eine sorgfältige Untersuchung würde einen umfangreichen Test der zukünftigen Lehrpersonen, ihre systematische Beobachtung und eine intensive Befragung notwendig machen, und zwar zu unterschiedlichen Zeitpunkten: nach Abschluss der Lehrerbildung – in Deutschland ggf. nach Phasen differenziert – und nach fünf Jahren im Beruf.

Die institutionelle Ebene umfasst Wirkungen der Lehrerbildung auf die gesamte Berufsgruppe der Lehrerinnen und Lehrer. Als Beispiele für mögliche Kriterien möchte ich hier die Rate der Lehrpersonen nennen, die den Beruf verlässt bzw. die beruflich veranlasste gesundheitliche Probleme davon tragen.

Forschungsergebnisse sollen als Beispiel für systemische Wirkungen der Lehrerbildung angeführt werden. Ich habe mit Absicht diesen Indikatoren gewählt, um deutlich zu machen, dass die Lehrerbildung nicht nur Ausbildungsaufgaben hat, sondern als akademisches Gebiet auch die Verpflichtung hat, zur wissenschaftlichen Erkenntnisgewinnung beizutragen. Die IEA-Studie selbst stellt beispielsweise ein solches Produkt dar.

Mein Ziel war, Ihnen mit diesen Ausführungen einen Einblick in die zukünftigen Aufgaben der empirischen Bildungsforschung im Bereich der Lehrerbildung zu geben. Es ging mir v.a. darum, die Komplexität des Vorhabens deutlich zu machen. Aus dieser ergibt sich eine völlig andere Untersuchungsanlage im Vergleich zu dem, was ich unter dem Stichwort ‚Programmevaluation‘ vorgestellt habe. Notwendigen bildungstheoretischen Ansprüchen kann man so aber stärker gerecht werden.

Im letzten Teil meines Vortrags möchte ich nun vor diesem Hintergrund die Chancen, aber auch die verbleibenden Grenzen des dargelegten Vorhabens reflektieren.

4 Chancen und Grenzen empirischer Forschung zur Lehrerbildung

Zunächst gehe ich auf Chancen ein, und zwar zunächst in bildungstheoretischer, dann in methodologischer und schließlich in entwicklungsorientierter Hinsicht.

Seit rd. 250 Jahren wird in Deutschland eine intensive bildungstheoretische Diskussion über die Lehrerbildung geführt. Dabei wird nicht nur über legitime Ziele diskutiert, es werden auch Wege zur Erreichung dieser Ziele entfaltet. Worin liegt aber eigentlich die Überzeugung begründet, dass die jeweils entwickelten Modelle tragen? Von Prondczynski spricht zu Recht von einem „Teufelskreis von Reform, Kritik, erneuter Reform und Kritik“. Diesen Teufelskreis wenigstens ansatzweise zu durchbrechen, indem zumindest die von Oelkers so bezeichnete ‚Krise der Daten‘ beendet wird, ist in meinen Augen eine der Hauptchancen der von mir skizzierten IEA-Vergleichsstudie.

Ein solches Vorhaben zu unterstützen, könnte dabei gerade aus bildungstheoretischer Perspektive interessant sein. Denn zu empirischen Erkenntnissen über die Lehrerbildung kann man nicht nur in technokratischer Weise kommen. Auf der Schulebene ist es ein Unterschied, ob Schülerleistungen

wissens- oder kompetenzorientiert erfasst werden. Es ist ein Unterschied, ob man Faktoren zum Unterrichtsklima und zur moralischen Entwicklung berücksichtigt (weil man sie aus sozialen Gründen für bedeutsam hält) oder ob man sie unberücksichtigt lässt (weil sie ggf. wenig zu kognitiven Schülerleistungen beitragen). Entsprechend ist es auf der Ebene der Lehrerbildung ein Unterschied, ob man Persönlichkeitsmerkmale der zukünftigen Lehrpersonen mit aufnimmt oder sich nur auf die Erhebung ihres Fachwissens konzentriert, ob man den Lehr-Lernprozess in der Ausbildung einbezieht oder sich auf das Produkt konzentriert, ob man dieses Produkt über die Kognitionen der Lehrpersonen und ihr Handeln oder über die Leistungen ihrer Schülerinnen und Schüler definiert. Dahinter stehen unterschiedliche Ansprüche an Schule und Lehrerbildung. Diese Ansprüche lassen sich durchaus benennen – und damit in Variablen umsetzen. Auf diese Weise bildungstheoretisch fundiert, kann die Studie dann auch einen Beitrag zur erziehungswissenschaftlichen Theoriebildung leisten.

In diesem Zusammenhang stellt schon allein die für empirische Studien notwendige Modellbildung eine bildungstheoretische Chance dar. Erst ein Modell mit seiner systematischen Unterscheidung von gesellschaftlichen Ebenen und Stadien von Bildungsprozessen macht abstrakte Bildungsfragen diskutierbar. So kann man dann z.B. den rein funktionalen Lesebegriff der PISA-Studien kritisieren. Aus einer fachdidaktischen Tradition, die sich nicht ausschließlich an der Pädagogischen Psychologie orientiert, kann so der Beitrag der Unterrichtsfächer zur Allgemeinbildung kritisch und anders reflektiert werden, als dies in den PISA-Studien der Fall ist. Vergleichbare Diskussionen erhoffe ich mir auch im Zusammenhang der anstehenden IEA-Lehrerbildungsstudie.

Das Problem der bisher eher wenig empirisch orientierten Erziehungswissenschaft und Fachdidaktiken ist aus meiner Sicht nämlich, dass ihre Reflexionen zwar bedeutsam sind, dass in der Vergangenheit aber häufig Operationalisierungen ausblieben, die es ermöglichen, einen plausiblen Lösungsweg tatsächlich daraufhin zu überprüfen, ob er das leistet, was seine Verfechter anstreben. Dietrich Benner formulierte in diesem Sinne kürzlich zu Recht aus der Perspektive der Bildungstheorie: „Zwar lassen sich Bildungstheorie und Empirie nicht in ein Kontinuum aufheben. Die Spannungen zwischen bildungstheoretischen und empirischen Fragen wechselseitig fruchtbar zu machen, ist dagegen ein Unternehmen, das für beide Seiten lohnend sein könnte.“

Ergänzend zu den angeführten bildungstheoretischen Chancen sehe ich einen methodologischen Mehrwert. Einerseits kann durch die Studie eine Weiterentwicklung der vergleichsweise neuen Methode der Mehrebenenanalyse erfolgen. In Bezug auf die Lehrerbildung handelt es sich um den ersten Versuch, diese Herangehensweise umzusetzen. Am Ende zu Indikatoren für Qualität in der Lehrerbildung zu kommen, die theoretisch und empirisch fundiert sind, halte ich für eine reizvolle Herausforderung. Eine weitere wichtige Leistung der Studie sehe ich darin, Instrumente zu entwickeln, die auch für Evaluationen genutzt werden könnten. Mit dem vorgesehenen Systemmonitoring ist nicht zu leisten, die Ziele einer bestimmten Universität oder eines bestimmten Studienseminars aufzunehmen. Gleichzeitig sind diese bei ihren Evaluationen jedoch häufig überfordert, was eine angemessene Instrumententwicklung angeht. Durch die IEA-Studie werden aber Institutionen, die sich einer intensiveren Evaluation stellen wollen, Instrumente bereitgestellt, die sie adaptieren können.

Neben diesen wissenschaftsinternen theoretischen und methodologischen Chancen sehe ich eine andere wichtige Perspektive der Studie, dass aus ihr fundierte Ansatzpunkte für eine Weiterentwicklung der Lehrerbildung gewonnen werden können – insbesondere aufgrund ihrer internationalen Orientierung. Dass in Deutschland überhaupt wieder über das Bildungssystem diskutiert wird, dass andere Institutionalisierungsformen als die bestehenden in den Sinn kommen,

ist eine Folge internationaler Vergleiche. Internationalität öffnet den Blick dafür, dass Systemstrukturen und Lehr-Lernprozesse auch ganz anders organisiert werden können, als es vor Ort der Fall ist. Im globalen Maßstab kann man die unterschiedlichen Institutionalisierungen der Lehrerbildung mit einer quasi-experimentellen Situation vergleichen. Die Variation zwischen den Ländern ermöglicht einen Vergleich ihrer relativen Wirksamkeit. Man ist also noch nicht einmal gezwungen, normativ ein Kriterium zu setzen, das erreicht werden muss, um von einer ‚guten‘ Lehrerbildung zu sprechen. In einer ersten Annäherung würde es auch genügen festzustellen: Verglichen mit xy, wird bei ab mehr oder weniger erreicht.

Soweit zu den Chancen der anstehenden Studie in theoretischer, methodologischer und entwicklungsorientierter Hinsicht. Ich möchte aber nicht versäumen, auch auf einige Grenzen aufmerksam zu machen.

Von Prondczynski kritisiert, dass das Kriterium der Qualität von Lehrerhandeln Lehrerbildung „auf eine kurzsichtige Praxis“ reduziert. Abgesehen davon, dass hier mit ‚Qualität‘ vermutlich ein sehr eingeschränktes Begriffsverständnis assoziiert wird, spricht von Prondczynski dennoch ein Grundsatzproblem an. Angesichts der normativen Ausrichtung von Bildungsprozessen und ihrer nur indirekten Messmöglichkeiten ist in der Tat eines der gravierendsten Probleme empirischer Forschung, dass selbst bei noch so komplexer Untersuchungsanlage ‚Bildung‘ – je nach Begriffsverständnis – gar nicht, mindestens aber nur schwer messbar ist. Das gilt v.a., wenn Bildung nicht allein auf kognitive Leistungen reduziert werden soll. Aber selbst bei einer differenzierten Berücksichtigung affektiv-motivationaler Komponenten und berufunspezifischer Persönlichkeitsmerkmale wird man in jedem Fall den Anspruch an empirische Forschung zur Lehrerbildung einschränken müssen. Der Technologisierbarkeit von Bildungsprozessen sind enge Grenzen gesetzt.

Bescheidenheit ist vermutlich auch angebracht, was die praktischen Folgen empirischer Studien angeht. Schon in den 50er Jahren wurde in den USA in Bezug auf das Schulsystem eine Mängelliste aufgemacht, die eine analoge Beschreibung der gegenwärtig festgestellten Probleme sein könnte. 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 sagt daher: „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.“

Vor diesem Hintergrund stellt sich die Frage, wem mit solchen Studien wie der von mir vorgestellten in praktischer Hinsicht ‚gedient‘ wird. Ich möchte in diesem Fall dem Habermas’schen Verständnis von empirischer Forschung als ‚Orientierungsforschung‘ im Interesse von Information und Aufklärung folgen. Das heißt für die Bildungsforschung, dass ihre Erkenntnisse langfristig auch in der Schule oder anderen Bildungsrealitäten ankommen müssen. Empirische Studien können Hinweise liefern und die Aufmerksamkeit auf bedeutsame Einflussfaktoren lenken. Diese können dann wieder Ausgangspunkt für neue Überlegungen und Untersuchungen zur Verbesserung der Praxis sein. Die PISA-Studien haben beispielsweise Erkenntnisse zu Defiziten unseres Schulsystems und gleichzeitig Anhaltspunkte für Ursachen geliefert. Im nächsten Schritt muss es nun darum gehen, diese unter dem Blickpunkt, was verändert werden kann und soll, in ein Konzept umzusetzen, das dann erprobt und wiederum evaluiert werden muss.

Mit unmittelbarer Anwendung von empirischen Erkenntnissen in der Praxis darf das aber nicht verwechselt werden. Zu meinen, aus empirischen Studien direkt Handlungsanleitungen ableiten zu können, ist ein Trugschluss. Politikberatung, wie es der Begriff der ‚Evaluation‘ in einem zentralen Verständnis nahe legt, erscheint mir daher nur schwer möglich – abgesehen von dem generellen Problem der Vermischung von Wertfreiheit im Begründungszusammenhang und legitimatorischen Interessen.

Diese Skepsis hat gerade etwas mit dem Gegenstand von Erziehungswissenschaft zu tun, den Bildungsprozessen. In nur wenigen Ländern ist die Bildungsdebatte so stark politisiert, geradezu von Grabenkämpfen durchzogen wie in Deutschland. Das hängt unter anderem damit zusammen, dass zwar immer die Autonomie der Bildung betont wird, dass diese ihr implizit aber gar nicht zugestanden wird. Durch Bildungsreformen soll immer gleich die ganze Gesellschaft verändert werden oder es wird immer gleich die ganze Gesellschaft in ihrem Bestand bedroht gesehen. Unter diesen Voraussetzungen zu erwarten, dass die Bildungsdebatte mit empirischen Erkenntnissen rational fundiert werden könne, ist erst einmal nur eine Hoffnung. Darüber hinaus bleiben die bei empirischer Forschung gewonnenen ‚harten‘ Fakten immer interpretationsbedürftig. Cochran-Smith formuliert hierzu [Zitat in deutscher Übersetzung]: „Egal welche Schlussfolgerungen letztlich aus Daten gezogen werden, hängen sie zu einem großen Ausmaß von komplexen Vorentscheidungen darüber ab, was als Beleg akzeptiert wird – Entscheidungen, die auf Werten, Zielen und Prioritäten basieren. In diesem Sinne wird die Debatte niemals nur durch empirische Erhebungen entschieden.“

Hinzu kommt, dass Bildungspolitik eine andere Rationalität zugrunde liegt als Wissenschaft. Politik ist notwendigerweise normativ orientiert und interessengeleitet. Wissenschaftliche Erkenntnisse sollten vor diesem Hintergrund zwar berücksichtigt werden, müssen aber nicht unbedingt handlungsleitend sein – ganz abgesehen von dem Problem, dass es ‚die‘ Wissenschaft ja auch nicht gibt. Fragen Sie fünf Erziehungswissenschaftler, wie sie die PISA-Ergebnisse interpretieren, und Sie erhalten mindestens sechs verschiedene Antworten. Dieses Phänomen ist ein Hinweis darauf, dass man Bildungsfragen eben unter ganz verschiedenen normativen Perspektiven betrachten kann.

Vor diesem Hintergrund möchte ich mit Bezug auf die aktuelle Diskussion um Bildungsstandards für Schule und Lehrerbildung abschließend schlaglichtartig auf mögliche gesellschaftliche Folgen von Vergleichsstudien eingehen. Seit PISA ist eine latente Tendenz zu registrieren, von einer input-Steuerung des Bildungssystems – sei es Schule, sei es Lehrerbildung – abzuweichen und output-Erwartungen zu entwickeln. Ihre Erreichung soll dann durch Tests überprüft werden. Ohne die Sinnhaftigkeit dieser Überlegungen grundsätzlich in Frage stellen zu wollen, möchte ich auf drei Probleme aufmerksam machen, die in meinen Augen bisher nicht hinreichend diskutiert werden.

Zum ersten lassen sich in Tests nur messbare Leistungen erfassen. So wichtig es in der Tat ist sicherzustellen, dass alle Kinder und Jugendlichen bestimmte Kenntnisse, Fertigkeiten und Fähigkeiten erwerben, so sehr würde eine ausschließliche Fokussierung der Schule darauf eine Reduzierung bisheriger Bildungsansprüche bedeuten. Die Entwicklung in den USA, wo sich mehr und mehr Schulen im Interesse besserer Testleistungen ihrer Schülerinnen und Schüler auf den Unterricht in Mathematik und Englisch konzentrieren, sollte ein warnendes Beispiel sein. Eine Beibehaltung von – ggf. entschlackten – Curricula in Ergänzung zu Tests scheint mir im Interesse eines weiten Bildungsbegriffs wichtig zu sein.

Ein zweites Problem ist, dass für die Entwicklung der Tests viel Spezialwissen benötigt wird. Dies ist ja auch der Anlass dafür, dass die KMK beschlossen hat, eine nationale Testagentur aufzubauen, deren Sitz die Humboldt-Universität sein wird. Demokratietheoretisch liegt die Gestaltung und Kontrolle des Bildungssystems aber bei Öffentlichkeit und Politik. Wenn es nicht gelingt, eine angemessene Balance zwischen wissenschaftlicher Fundierung und öffentlicher Diskussion zu finden, könnte eine Überprüfung des Bildungssystems durch Tests unseren Vorstellungen von Demokratie widersprechen.

In diesem Zusammenhang stellt sich drittens noch einmal das Problem der Uneinheitlichkeit von ‚Wissenschaft‘. Der Aufbau einer nationalen Testagentur darf nicht damit einhergehen, Vielfalt zu reduzieren. Dass dies aber ggf. eine reale Gefahr ist, wird schon jetzt an den TIMS-, PISA- und IGLU-Studien deutlich: Eine andere Wahl als das dahinter stehende, sehr einheitliche Bildungsverständnis zu akzeptieren, besteht für Öffentlichkeit und Politik kaum. In einer pluralistischen Demokratie ist es aber nicht nur legitim, sondern sogar wünschenswert, dass unterschiedliche Auffassungen von Bildung existieren und sich gegenseitig herausfordern. Von herausragender Bedeutung wird es daher sein, dass bei der Testentwicklung unterschiedliche Bildungsverständnisse die Chance bekommen, sich widerzuspiegeln. Auch in der Wissenschaft sind Monopole nicht unbedingt der beste Weg zur Erkenntnis.